Novel developments in materials informatics - COST Action Session

Thursday 19 September 2024 - Friday 20 September 2024

Warsaw

Narodowe Centrum Badań Jądrowych
National Centre for Nuclear Research
ŚWIERK

NOMATEN
Centre of Excellence in Multifunctional Materials for Industrial and Medical Applications

cost
EUROPEAN COOPERATION IN SCIENCE & TECHNOLOGY

Book of Abstracts

Only accepted abstracts are presented in the book.

Contents

Application of data analysis in Raman spectroscopy 20 1

Application of Machine Learning for Predicting Mechanical Properties and Designing Novel Biocompatible Titanium Alloys 19 1

The effect of proper material model calibration on additively manufactured polymer material 18 2

Recent Advances in Nanostructured TiO_2 Modifications for Enhanced Photocatalytic and Photoelectrochemical Green Hydrogen Production 17 2

Advanced Normal Mode Analysis of Carbon Nanotubes 16 2

Machine learning and workflows for scale bridging in mechanics. 15 3

Automated data integration platforms for materials science and cheminformatics 3 . . . 3

Potassium Molten Salt-Mediated In Situ Structural Reconstruction of a Carbon Nitride Photocatalyst 13 4

Machine Learning Models for Dye Sensitized Solar Cells (DSSCs) 7 4

Computational Exploration of Zeolite Properties Using Neural Networks Potentials 8 . . 5

Synthesis, Characterization, and Density Functional Theory Calculation of Mixed Organic-Inorganic Perovskite Materials 14 5

Novel Developments in Materials Informatics 12 6

Modelling Additive Manufactured Nanocomposites by Bridg-ing Different Computational Schemes 6 6

GrapheNet: A Novel Deep Learning Model for Predicting Physical and Electronic Properties of 2D Materials Using Images 4 7

A Multiscale Computational Approach Integrating Molecular Dynamics and Peridynamics Simulations to Understand Fracture Behaviour in Green Nano Fibrous Network 2 . . 7

Data-Driven Mechanical Property Analysis and Predictions of Steel at High Temperatures. 1 8

ML-Driven Analysis of Self-Healing Properties in Tungsten Grain Boundaries 9 8

The Origins of Formic Acid Electrooxidation on Selected Surfaces of Pt, Pd, and Their Alloys with Sn 11 9

Accelerating the Discovery of Organic Molecules for Photoelectrochemical (PEC) Water Splitting Using Transformer-Based Surrogate Models 10 9

Efficient Workflow Automation for Materials Modeling: Towards Predictive AI Systems Using High Throughput Synthetic Dataset Generation 5 10

20

Application of data analysis in Raman spectroscopy

Kinga Suchorab[1] ; Magdalena Gaweda[1] ; Lukasz Kurpaska[1]

[1] *National Centre for Nuclear Research*

Corresponding Author(s): lukasz.kurpaska@ncbj.gov.pl, magdalena.gaweda@ncbj.gov.pl, kinga.suchorab@ncbj.gov.pl

Raman spectroscopy is widely employed for materials characterisation due to its ability to provide detailed information on chemical structure, phases, and crystallinity. Its sensitivity to subtle structural changes makes it particularly useful for analysing zirconia polymorphs formed during oxidation in nuclear reactors. Zirconium alloys, commonly used for nuclear fuel cladding, are exposed to harsh environment and undergo cyclic corrosion during operation. This process leads to the creation of tetragonal zirconium oxide, which eventually transforms into the monoclinic phase. The monoclinic phase is less protective, and therefore, undesirable. Raman spectroscopy enables precise differentiation between zirconium oxide polymorphs based on their distinct Raman spectra. Additionally, Raman imaging enhances this method by generating a detailed map from numerous measurement points, allowing for a thorough analysis of phases distribution. However, specialised methodologies are required to effectively analyse Raman data [1, 2].
To extract detailed information from the Raman spectra (such as Raman shift, Full Width at High Maximum, and Intensity), every spectrum must be fitted using Voigt functions across all bands simultaneously. In the case of Raman imaging, where results consist of multiple spectra, simultaneous analysis is necessary to save time. Techniques such as K-means clustering (KMC) and multiple spectra fitting can be employed for this purpose [3]. This work compares these methods and discusses the challenges encountered in their application.
References:
[1] A.T. Motta, A. Couet, R.J. Comstock, Corrosion of Zirconium Alloys Used for Nuclear Fuel Cladding, Annual Review of Materials Research 45 (2015) 311-343.
[2] J. Liu, H. Yu, P. Karamched et al., Mechanism of the α-Zr to hexagonal-ZrO transformation and its impact on the corrosion performance of nuclear Zr alloys, Acta Materialia 179 (2019) 328-341
[3] K. Suchorab, M. Gawęda, Ł. Kurpaska, Comparison of Raman imaging assessment methods in phase determination and stress analysis of zirconium oxide layer, Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy 295 (2023) 122625

19

Application of Machine Learning for Predicting Mechanical Properties and Designing Novel Biocompatible Titanium Alloys

Markovic Gordana[1] ; Manojlović Vaso[2] ; Ružić Jovana[3]

[1] *Institute for Technology of Nuclear and Other Mineral Raw Materials*

[2] *University of Belgrade*

[3] *National Institute of the Republic of Serbia*

Corresponding Author(s): g.markovic@itnms.ac.rs

Titanium alloys are renowned for their exceptional strength, low modulus of elasticity, and excellent biocompatibility, making them ideal for orthopedic applications. To optimize these properties and explore new alloy compositions, machine learning was employed to analyze a database of biocompatible alloys. Using the Extra Trees Regressor model, key mechanical properties : Young's modulus, tensile strength, and yield strength were predicted. The correlation coefficients for the test set were 0.793, 0.893, and 0.868, respectively, while the mean absolute errors fell within the acceptable range of values reported in the literature. By designing experiments and constructing relief maps for over thirty alloying element combinations, approximately 16,000 unique compositions of four-component alloy systems were discovered, providing a basis for further experimental research.

18

The effect of proper material model calibration on additively manufactured polymer material

Aleksa Milovanović[1] ; Aleksandar Sedmak[1] ; Miloš Milošević[1]

[1] *University of Belgrade*

Corresponding Author(s): amilovanovic@mas.bg.ac.rs, mmilosevic@mas.bg.ac.rs, asedmak@mas.bg.ac.rs

Additive manufacturing is a well-established way of producing components for home use, prototyping, and engineering purposes. Components are created in a layer-by-layer manner and predefined process parameters have a substantial influence on the final components' properties. Namely, layer height, infill density, build and raster line orientation, process temperature and speed, to name just a few. In mechanical engineering, numerical simulations are a common practice for the estimation of final components' behavior in loading situations. Thus, for every process parameter setting a different material model must be created if one expects valid results in the numerical simulation. In addition, polymer materials are more difficult for material model estimation, due to their anisotropy and viscoplasticity. In practical use, there are software tools that use material data from tensile and compressive tests to calibrate a proper material model. Also, a selection of a material model type (i.e., Three Network, Bergstrom-Boyce, etc.) has an impact on numerical simulation. In this research, the obtained experimental data from the tensile tests were compared with the numerical simulation outputs with ''bad'' and proper material models.

17

Recent Advances in Nanostructured TiO_2 Modifications for Enhanced Photocatalytic and Photoelectrochemical Green Hydrogen Production

Wierzbicka Ewa[1]

[1] *Military University of Technology*

Corresponding Author(s): ewelina.szaniawska@wat.edu.pl

A sustainable approach to hydrogen generation involves the use of photosensitive materials capable of splitting water into hydrogen and oxygen under solar irradiation. Titanium dioxide (TiO_2) is the standard semiconductor material employed in photocatalysis (PC) and photoelectrocatalysis (PEC). TiO_2 can be synthesized into various nanostructures, such as nanopowders and anodic nanotube/nanoporous arrays, which exhibit significantly higher surface areas compared to bulk materials, making them highly suitable for further modifications. The most effective and widely adopted strategy for modifying TiO_2 involves coupling it with noble metal cocatalysts, such as Pt or Au. This modification leads to enhanced PEC and PC efficiencies due to several mechanisms, including Schottky barrier formation, surface plasmon resonance, and co-catalytic hydrogen absorption. Alternatively, noble-metal-free modification approaches, such as sensitization with other semiconductors or self-doping, are gaining prominence. Here, novel strategies of nanostructured TiO_2 will be discussed.

16

Advanced Normal Mode Analysis of Carbon Nanotubes

Farcas Alexandra[1] ; Titus Beu[2]

[1] *National R&D Institute for Isotopic and Molecular Technologies*

[2] *University "Babes-Bolyai", Faculty of Physics, Department of Biomolecular Physics*

Corresponding Author(s): alexandra.farcas@itim-cj.ro

Carbon nanotubes (CNTs) are emerging as promising materials for topical nano-electromechanical applications. This study presents a comprehensive normal mode analysis of both clamped-clamped and clamped-free CNTs utilizing a non-orthogonal tight-binding force field. A novel methodology leveraging geometry-adapted merit functions is introduced to characterize vibrations across high orders. While low harmonics exhibit straightforward behaviors similar to continuum models, high harmonics display intricate mixing patterns. The fundamental frequencies obtained through extrapolation align well with experimental data and previous force constant calculations. Additionally, low bending vibrations show good agreement with continuum theory predictions. Notably, the fundamental frequencies for twisting and stretching demonstrate chirality independence, enhancing their potential utility in mass resonator applications.

15

Machine learning and workflows for scale bridging in mechanics.

Petr Grigorev[1]

[1] *Aix-Marseille University*

Corresponding Author(s): petr.y.grigorev@gmail.com

Mechanical properties such as plasticity and fracture toughness are governed by intrinsically multiscale phenomena. Thus, understanding the underlying mechanisms requires an effective scale bridging strategy. Recent development in machine learning and data driven methods allows tackling long-standing issues in multiscale and multiphysics simulations. In this presentation I will describe how hybrid QM/ML methods can be used to achieve seamless coupling between ab initio and empirical force fields taking the problem of impurity induced embrittlement of metals as an example. I will also discuss the use of machine learning for coarse graining of atomistic simulations in order to move towards experimental length and time scales.

3

Automated data integration platforms for materials science and cheminformatics

Fabio Le Piane[1] ; Francesco Mercuri[None] ; Andrea Lorenzoni[1] ; Tommaso Forni[2] ; Mario Vozza[3]

[1] *CNR-ISMN*

[2] *CNR-ISMN and University Of Turin*

[3] *CNR-ISMN and University of Turin*

Corresponding Author(s): mario.vozza@ismn.cnr.it, tommaso.forni@ismn.cnr.it, andrea.lorenzoni@cnr.it, francesco.mercuri@cnr.it, fabio.lepiane@ismn.cnr.it

In this age defined by the advent of artificial intelligence and deep learning, the significance of ready access to extensive, top-tier datasets cannot be overstated. Within this landscape, neural networks emerge as pivotal tools, possessing the remarkable capacity to discern intricate relationships and patterns within the data they ingest. However, harnessing this potential demands an immense

volume of data. In the realm of materials science, data often exists in a fragmented state across various datasets, each encoded in different formats. This fragmentation poses a considerable challenge for machine learning engineers, who must meticulously navigate and amalgamate these disparate sources. The process of merging these datasets is not only time-consuming but also very prone to errors, due to the myriad of formats and standards endemic to the computational community, further complicating the integration task. Against this backdrop, our endeavor aims to introduce a sophisticated automated data-integration solution tailored specifically to streamline the utilization of machine learning and neural networks within the advanced materials domain. Our approach draws upon a diverse array of technologies to realize this goal. Primarily, we leverage established database technologies as the foundational framework of our solution, ensuring robust and dependable storage for the multitude of files and data at hand. Furthermore, we craft an integration framework characterized by its versatility in assimilating data in multiple formats from different sources. This framework seamlessly transforms this varied data into a unified format, which is then securely stored within the database. In addition, we have incorporated principles from semantic technologies, particularly ontologies, into our methodology. By harnessing the defined terminology and relationships within ontologies, we construct our unified knowledge graph with an emphasis on both expressivity and semantic consistency. This approach ensures that the resulting database maintains coherence across diverse sources. The symbiotic interplay of these disparate elements culminates in a comprehensive solution capable of consistently integrating heterogeneous data sources. Furthermore, thanks to well established query engines like **SPARQL**, it is possible to perform expressive semantic queries on the resulting data structures. Through this concerted effort, we aim to empower researchers and practitioners within the materials science community with a powerful toolset for data-driven exploration and innovation.

13

Potassium Molten Salt-Mediated In Situ Structural Reconstruction of a Carbon Nitride Photocatalyst

Aharon Gedanken[None] ; Dong Jin Kim[None] ; Kedhareswara Sairam Pasupati[None] ; Suresh S Umare[None] ; Toshali Bhoyar[None] ; Vidyasagar Devthade[None] ; Vidyasagar Devthade[None]

Corresponding Author(s): akanksha01orai@gmail.com

Potassium Molten Salt-Mediated In Situ Structural Reconstruction ofa Carbon Nitride Photocatalyst Akanksha Gupta,◆ Toshali Bhoyar,◆ B. Moses Abraham, Dong Jin Kim,Kedhareswara Sairam Pasupuleti, Suresh S. Umare, Devthade Vidyasagar,*and Aharon Gedanken*Cite This: ACS Appl. Mater. Interfaces 2023, 15, 18898−18906 Read OnlineACCESS Metrics & More Article Recommendations *sı Supporting InformationABSTRACT: Metal-free polymeric carbon nitride (PCN) materials are at theforefront of photocatalytic applications. Nevertheless, the overall functionality andperformance of bulk PCN are limited by rapid charge recombination, high chemicalinertness, and inadequate surface-active sites. To address these, here, we employedpotassium molten salts (K+X−, where X− is Cl−, Br−, and I−) as a template for the insitu generation of surface reactive sites in thermal pyrolyzed PCN. Theoreticalcalculations imply that addition of KX salts to PCN-forming monomers causeshalogen ions to be doped into C or N sites of PCN with a relative trend of halogenion doping being Cl < Br < I. The experimental results show that reconstructing Cand N sites in PCN develops newer reactive sites that are beneficial for surfacecatalysis. Interestingly, the photocatalytic H2O2 generation rate of KBr-modified PCNwas 199.0 μmol h−1, about three times that of bulk PCN. Owing to the simple andstraightforward approach, we expect molten salt-assisted synthesis to have wideexploration in modifying PCN photocatalytic activity.

7

Machine Learning Models for Dye Sensitized Solar Cells (DSSCs)

Hanoch Senderowitz[None] ; Paul Clarke[1] ; Hadar Binyamin[1]

[1] *Bar-Ilan University*

Corresponding Author(s): hadarbin1q@gmail.com, hsenderowitz@gmail.com, paulclarke61@gmail.com

DSSCs hold the potential to meet the growing worldwide demand for clean and sustainable energy. Previously we have demonstrated how pharmacophore models could be used to identify new dyes with favorable (predicted) electronic properties for DSSCs. More recently, we have developed the first global ML model based on data compiled in the Dye Sensitized Solar Cell Database. This model combines a newly developed device fingerprint (DFP) that describes the effect of the devices' components with Morgan Fingerprints that describe the dyes' structures into an ANN-based model for the prediction of overall Power Conversion Efficiency (PCE). This global model exhibits high predictive ability on an external test set and has a broad applicability domain, including organic compounds of various scaffolds, metal organic compounds and dye mixes.

8

Computational Exploration of Zeolite Properties Using Neural Networks Potentials

Author(s): Indranil Saha[1]

Co-author(s): Andreas Erlebach [1] ; Christopher Heard [2] ; Lukáš Grajciar [1] ; Petr Nachtigall [1] ; Martin Sipka [1] ; Denisa Novakova [1]

[1] *Charles University*

[2] *Charles Univeristy*

Corresponding Author(s): sahai@natur.cuni.cz

Neural network potentials (NNPs) are emerging as a powerful tool in material science for understanding the structure and properties of complex materials. One such material class is zeolites, widely used in various industrial processes. Understanding the properties and behaviour of these materials is critical to optimising their performance in different applications. In our works, we investigated two topics related to zeolites using in-house-developed neural network potentials. i) aluminosilicate zeolites in interaction with water to quantify the effect of aluminium content, water loading and temperature on water diffusion and proton solvation in H-FAU zeolite. and ii) understanding the role of zeolite topology for the loading and substitution of germanium in germanosilicates.

Overall, with the advantage of NNPs, in both cases, we were able to reduce the computational cost of simulations by several orders of magnitude while keeping a DFT level of accuracy, therefore allowing for the investigation of larger systems for longer timescales enabling rapid evaluation of design strategies under realistic conditions utilising complex models, which was to longer timescales simulation than previously impossible.

14

Synthesis, Characterization, and Density Functional Theory Calculation of Mixed Organic-Inorganic Perovskite Materials

Teoman Ozdal[1] ; Merhan Kilic[1] ; Hamide Kavak[1]

[1] *Cukurova University*

Corresponding Author(s): hkavak@cu.edu.tr, merhankilic@gmail.com, teomanozdal@hotmail.com

Organic-inorganic halide (OIH) perovskites are emerging as promising semiconductors for next-generation photovoltaic (PV) devices due to their exceptional properties, including high absorption coefficients, low exciton binding energy, tunable band gaps, and large carrier lifetimes and diffusion lengths. These crystalline materials have the general formula ABX3, where A and B are cations

(organic/inorganic and inorganic metal, respectively), and X is a halide anion. Achieving high photocurrents involves careful tuning of the stoichiometry between organic and inorganic ions.
The sol-gel spin coating method is employed to grow the methylammonium lead triiodide (MAPbI3) absorber layers. Solvents such as dimethyl sulfoxide (DMSO), dimethylformamide (DMF), and their mixtures are used in this process. The characteristics of these solvents (boiling point, viscosity, vapor pressure, polarity, coordination ability, etc.) significantly influence the growth process and the quality of the deposited perovskite film.
Systematic analyses were conducted to optimize the growth parameters. These involved testing various DMF:DMSO solvent ratios, molarities, annealing temperatures, and annealing times. Toluene is used as an antisolvent, and both the quantity and application method of the antisolvent were investigated. The films were characterized to assess their optical, electrical, structural, and morphological properties.
Density functional theory calculation was performed using the Quantum Espresso software package with the plane-wave pseudopotential approach. The exchange-correlation effects were described by the generalized gradient approximation (GGA) with the Perdew-Burke-Ernzerhof (PBE) functional to determine electro-optical properties, such as band gap, formation energy, dos etc. Since the PBE functional generally underestimates the band gap energy, hybrid functionals were also used to obtain more accurate band gap values.

12

Novel Developments in Materials Informatics

FRIDA ZISKO[1]

[1] *Mediterranean University of Albania*

Corresponding Author(s): frida.zisko@umsh.edu.al

Materials Informatics (MI) has emerged as a transformative approach in materials science, leveraging advancements in machine learning (ML) to accelerate the discovery and development of new materials. Recent developments in MI have harnessed various machine learning models, such as decision trees, support vector machines (SVM), neural networks, and ensemble methods, each offering unique advantages for different applications in materials science. Machine learning models in materials informatics, such as decision trees and random forests, are widely used for their interpretability and ability to handle complex, non-linear relationships in data. They are particularly effective in predicting material properties based on compositional and structural features, offering insights into the factors influencing material behavior. Support vector machines (SVM) are employed in materials informatics for classification tasks, such as categorizing materials based on phase stability or identifying potential candidates for specific applications. Their robustness in handling high-dimensional data makes them suitable for complex datasets commonly encountered in MI. Neural networks and deep learning have revolutionized MI by enabling the analysis of large-scale datasets with high complexity.

6

Modelling Additive Manufactured Nanocomposites by Bridg-ing Different Computational Schemes

Author(s): AbdolMajid Rezaei[1]

Co-author(s): Razieh Izadi [2] ; Nicholas Fantuzzi [3]

[1] *PhD student in Sapienza University of Rome*

[2] *Assistant Professor at Sapienza University of Rome*

[3] *Associate Professor, University of Bologna*

Corresponding Author(s): razie.izadi@uniroma1.it, abdolmajid.rezaei@uniroma1.it

Fused Deposition Modelling (FDM) is a widely used additive manufacturing technique for rapid prototyping in fields such as tissue engineering, aerospace, and electronics. Polylactic Acid (PLA) is a biodegradable polymer used in FDM due to its printability and compatibility with various addi-tives. However, PLA mechanical properties and lack of antibacterial features limit its functionality. Reinforcements like silver nanoparticles, enhance PLA's mechanical strength and antimicrobial properties, making them suitable for biomedical applications. However, FDM introduces microscale voids that affect mechanical properties, though increased porosity can benefit applications such as bone regeneration scaffolds. This study presents a computational approach that bridges molecular dynamics to capture atomistic behaviours and non-classical micropolar theory to account for ma-terial micro-heterogeneities, providing a comprehensive framework for designing additive manu-factured PLA nanocomposites.

4

GrapheNet: A Novel Deep Learning Model for Predicting Physical and Electronic Properties of 2D Materials Using Images

Tommaso Forni[1] ; Fabio Le Piane[2] ; Francesco Mercuri[None] ; Matteo Baldoni[3]

[1] *CNR - ISMN, Polytechnic of Turin*

[2] *CNR-ISMN*

[3] *CNR - ISMN*

Corresponding Author(s): matteo.baldoni@ismn.cnr.it, tommaso.forni@ismn.cnr.it, fabio.lepiane@ismn.cnr.it, francesco.mercuri@cnr.it

The quest to represent complex materials structures and nanostructures has been a persistent challenge in the field of chemistry and materials science. Traditional methods of representing small molecules, such as SMILES, have proven to be insufficient to represent more complex systems. In recent years, the use of 2D images to represent the structure of materials has gained popularity, especially with the development of artificial intelligence (AI) tools and deep learning algorithms. In this study, we present a novel approach to represent the structure of 2D based materials, such as graphene and graphene oxide nanosystems, and an advanced predictive AI-based framework. Our proposed deep learning model, GrapheNet, is based on an Inception-ResNet architecture consisting of multiple blocks of convolutional layers with different kernel sizes. The GrapheNet model can be trained to make predictions about the physical and electronic properties of graphene-based systems using PNG images as structural representations. The efficacy of the approach was tested on datasets of graphene oxide and defected graphene systems, built starting from repositories of computed structure/property data. Structural data of the nanosystems in the dataset, encoded in standard structural formats, are transformed into three-dimensional (graphene oxide) or two-dimensional (defected graphene) tensors, converted into RGB (graphene oxide) or grayscale (defected graphene) PNG images and pre-processed (cropping, resizing, recentering, padding). Upon training, the GrapheNet framework yielded very accurate results in predicting phyisico-chemical properties of graphene oxide and graphene nanostructures, with low mean prediction errors for all target properties considered, also exhibiting a significant computational efficiency. Being based on highly-efficient frameworks borrowed from state-of-art computer vision technologies, the approach proposed demonstrates the potential of using image-like representations of 2D and low-dimensional nanostructures in connection with deep learning predictive models, predicting the chemico-physical properties of nanographenes with great accuracy and outperforming the computational efficiency of current methods.

2

A Multiscale Computational Approach Integrating Molecular Dynamics and Peridynamics Simulations to Understand Fracture Be-

haviour in Green Nano Fibrous Network

Author(s): Razieh Izadi[1]

Co-author(s): Nicholas Fantuzzi [2]

[1] *Assistant Professor at Sapienza University of Rome*

[2] *Associate Professor, University of Bologna*

Corresponding Author(s): razie.izadi@uniroma1.it

We aim to present a hierarchical computational method to understand the fracture of biodegradable three-dimensional nanofibrous networks. Challenges in predicting the mechanical behaviour of these complex structures arise from multiscale features and variability in material properties. Therefore, in this work, a multi-scale approach is adopted that bridges atomistic simulations at the nanoscale and peridynamics at the microscale. At the nano scale, all-atom molecular dynamics simulations are performed on freestanding pristine and silver-doped polylactic acid nanofibres. The results are seamlessly translated to the microscale where non-local continuum peridynamics is used to assess crack propagation and fracture toughness of mode I and mode II in both aligned and randomly oriented fibrous networks. Through this computational approach, we maintain the accuracy of the atomistic description and benefiting from the efficiency of continuum mechanics to address intricate failure phenomena.

1

Data-Driven Mechanical Property Analysis and Predictions of Steel at High Temperatures.

Casim Yazici[1] ; Francisco Javier Dominguez Gutierrez[2]

[1] *Agri Ibrahim Cecen University, Department of Construction, Division of Construction Inspection, Agri, Türkiye*

[2] *NOMATEN Centre of Excellence*

Corresponding Author(s): cyazici@agri.edu.tr, javier.dominguez@ncbj.gov.pl

Traditional experimental methods are widely used to characterize the mechanical properties of steels. However, recent advancements highlight the importance of a data-driven approach. This study presents a machine learning-based approach for analyzing the mechanical properties of S235 grade steel at high temperatures. Experimental data is utilized to train regression models, ensemble methods, and a neural network to predict stress-strain curves for steel samples at specific temperatures and thicknesses. To evaluate model performance, root-mean-square error (RMSE) and the coefficient of determination (R^2) are used. The models, trained on a robust experimental dataset, achieve impressive accuracy with R^2 values exceeding 0.9995 for yield strength and 0.9905 for elongation prediction.

9

ML-Driven Analysis of Self-Healing Properties in Tungsten Grain Boundaries

Jorge Suárez-Recio [1] ; Pablo Piaggi[2] ; Javier Domínguez-Gutiérrez [3] ; Raquel González-Arrabal[1] ; Roberto Iglesias[4]

[1] *Universidad Politécnica de Madrid*

[2] *CIC nanoGUNE BRTA*

[3] *fNOMATEN Centre of Excellence, National Center for Nuclear Research*

[4] *Universidad de Oviedo*

Corresponding Author(s): ppiaggi@princeton.edu, javier.dominguez@ncbj.gov.pl, raquel.gonzalez.arrabal@upm.es, j.srecio@upm.es, roberto@uniovi.es

Nuclear fusion offers a sustainable energy future, but the development of plasma-facing materials that can withstand extreme conditions remains a major challenge. Nanostructured tungsten (NW) shows potential due to its high density of grain boundaries (GBs), which offer potential to enable self-healing properties by facilitating hydrogen diffusion and preventing accumulation. However, the behavior of intrinsic defects at GBs in the presence of helium remains unclear.

To address these challenges, we have developed a machine learning interatomic potential (MLIP) based on extensive DFT data to simulate defect evolution at different densities and temperatures, focusing on the experimentally observed W(110)/W(112) GB. Since traditional methods are often cumbersome and error-prone, we use advanced ML techniques for semi-automated defect analysis to accurately capture defect behavior. This approach streamlines analysis, reduces time, and minimizes human error.

11

The Origins of Formic Acid Electrooxidation on Selected Surfaces of Pt, Pd, and Their Alloys with Sn

Author(s): Radhey Shyam Yadav[1]

Co-author(s): Alex Schechter [1] ; Haya Kornweitz [1] ; Medhanie Gebremedhin Gebru [1]

[1] *Ariel University*

Corresponding Author(s): yadav.radhey137@gmail.com

Bimetallic and trimetallic catalysts enhance the catalytic process of formic acid oxidation (FAO) and its adsorbed CO*main intermediate. A comprehensive study, using DFT calculations and electrochemical experiments, was conducted on a series of five mono-, bi-, and tri-metallic catalysts for FAO: Pt(111), Pd(111), Pt3Sn(111), Pd3Sn(111), and Pt3Pd3Sn2(111)1, considering both pathways: direct oxidation to CO2 and the indirect pathway involving the oxidation of adsorbed CO.* The results show that H–COOH cleavage is thermodynamically predominant on most of the surfaces except Pd3Sn(111). However, kinetically, O–H bond scissoring is preferred on all of the surfaces. The barriers for O–H and C–H cleavage for Pt(111) and Pt3Pd3Sn2(111) are comparable thus the process is governed by thermodynamics. On Pt(111), C–H cleavage is favored, while on Pt3Pd3Sn2(111), both O–H and C–H cleavage are equally viable due to similar thermodynamic profiles. Additionally, Pt-based alloys promote the indirect mechanism, whereas Pd enhances the direct mechanism. Alloying Pt with Sn and the combined effect of Pt, Pd, and Sn in a trimetallic alloy results in the weakening of the HCOOH adsorption and reduces all the activation barriers. The electrochemical findings support the computational results, demonstrating that Pd(111) exhibits oxidation peaks at a low potential, 0.37 V, (indicating a direct mechanism), while Pt-based catalysts display oxidation peaks at around 0.95 V, indicative of the indirect mechanism. Notably, Pt3Pd3Sn2/C shows the highest overall performance towards FAO, with peak current densities of 225 mA mgPGM−1. Thus, Pt3Pd3Sn2 is the most efficient catalyst, providing the lowest energetic pathways for FAO reaction.

10

Accelerating the Discovery of Organic Molecules for Photoelectrochemical (PEC) Water Splitting Using Transformer-Based Surrogate Models

Iman Peivaste[None] ; Francesco Mercuri[None] ; Salim Belouettar[1]

[1] *Louxembourg Institute of Science and Technology (LIST)*

Corresponding Author(s): salim.belouettar@list.lu, francesco.mercuri@cnr.it, iman.peivaste@list.lu

Photoelectrochemical (PEC) water splitting represents a promising technology for sustainable hydrogen production by leveraging solar energy to split water molecules into hydrogen and oxygen. The efficiency of this process heavily depends on the materials used, particularly the photocatalysts. Organic molecules, known for their unique electronic properties, cost-effectiveness, and abundance, have emerged as potential candidates for enhancing PEC performance. However, the vast chemical space of possible organic molecule configurations and the computational intensity of traditional Density Functional Theory (DFT) calculations present significant challenges in identifying optimal materials.

This research outlines a data-driven approach to accelerate the discovery and optimization of organic molecules for PEC water splitting. Transformer-based surrogate models, leveraging machine learning, are employed to predict the properties of organic molecules rapidly and accurately. These models are trained on a dataset generated through extensive DFT calculations, capturing critical properties such as band gap, ionization potential, and electron affinity. The architecture of the transformer model is based on the RoBERTa encoder, specifically modified for this application. Additionally, the tokenization process utilizes byte pair encoding (BPE) technology, as implemented in GPT-2 and GPT-3, to enhance the model's efficiency and accuracy.

The research methodology involves the systematic collection and preprocessing of data, the design and training of transformer models, and the validation and refinement of these models using DFT calculations to ensure high predictive accuracy. By reducing reliance on exhaustive DFT calculations, this approach expedites the high-throughput screening of organic molecules, significantly enhancing the pace of innovation.

The results of this research include a framework capable of accelerating discovery processes and the development of a comprehensive database of organic molecule properties. This study makes considerable contributions to the advancement of PEC water-splitting technology, reduces research and development costs, and promotes the integration of AI / ML techniques into materials science. Ultimately, this research supports the larger goal of sustainable hydrogen production and the transition to renewable energy sources.

5

Efficient Workflow Automation for Materials Modeling: Towards Predictive AI Systems Using High Throughput Synthetic Dataset Generation

Mario Vozza[1] ; Tommaso Forni[2] ; Fabio Le Piane[3] ; Francesco Mercuri[None]

[1] *Politecnico di Torino, CNR-ISMN*

[2] *CNR-ISMN and University Of Turin*

[3] *CNR-ISMN*

Corresponding Author(s): tommaso.forni@ismn.cnr.it, francesco.mercuri@cnr.it, mario.vozza@polito.it, fabio.lepiane@ismn.cnr.it

In the field of materials science, there has been a significant increase in the need for streamlined computational techniques aimed at generating predictive datasets tailored for artificial intelligence (AI) applications. Density Functional Tight Binding (DFTB) simulations serve as a powerful tool for elucidating atomic-scale interactions and material properties. However, the manual preparation of DFTB simulations can be time-consuming, hindering the rapid generation of large-scale datasets necessary for training AI models.
This study presents a comprehensive approach to automating the generation and analysis of materials science datasets, specifically focusing on defected graphene structures. This workflow efficiently handles computations for various material properties, including energy and charge transport. By automating these procedures, we can efficiently generate extensive datasets wherein each structure is correlated with its corresponding properties. This tight coupling between structures and properties

provides a robust foundation for training predictive models. Additionally, beyond the output properties from DFTB, we have augmented the dataset with synthetic Scanning Tunnelling Microscopy (STM) images generated using the Local Density of States (LDOS). This expansion opens ways for correlating experimental measurements directly with the examined structure in future analyses, enhancing the dataset extension.
With the dataset we created, we employed object detection techniques to identify defects within the graphene flakes. Subsequently, we extracted these defects from the structure's image and utilised classical computer vision techniques to derive features from these defects. The aim was to predict material properties based on the defect geometry using eXtreme Gradient Boosting (XGBoost). Moreover, having access to STM images allows us to correlate images and material properties using convolutional neural networks (CNNs). A pivotal element of this study was the robust integration of data, whereby all outputs from the simulations, including the generated STM images, were stored within a NoSQL database like MongoDB. This integrated approach enhances data management capabilities, allowing for easier scalability and ensuring the consistency and reliability of the dataset. By centralising the storage of simulation outputs and images, researchers could seamlessly access and analyse the data, fostering collaboration and accelerating scientific discoveries in materials science.