MATDAT18: Materials and Data Science Hackathon

Team Composition (2 people max.)

Name	Department	Institution	Email
Yanjun Ma	Physics and Astronomy	West Virginia University	yanjun.ma@mail.wvu.edu
Cheng Cen	Physics and Astronomy	West Virginia University	Cheng.Cen@mail.wvu.edu

Project Title

Synthesis of Oxide Heterostructures by Design

Project Synopsis (approx. 100 words)

Oxide heterostructures provide a unique platform to deliver novel science and innovative applications that is not attainable in bulk constituents. The theoretical design of oxides with desired properties have proliferated in the past decade, driven by the rapid advances of computational techniques and deep-learning algorithms. In contrast, the synthesis of oxide heterostructures by design is still in its infancy. The PIs propose to devise a machine-learning algorithm to screen the synthesis parameters for novel oxide heterostructures produced by the pulsed laser deposition method, which will link “material by design” to “synthesis by design” and accelerate the throughput of new functional materials.

Identified Data-Science Collaborative Need (approx. 100 words)

1. The building of a proper text-mining method to identify the synthesis parameters in various publications and transform them into machine-readable data sets.

2. The invention of a machine-learning algorithm that is able to identify key synthesis parameters as well as predict the probable values for these processing variables.

3. The upgrade of the deep-learning algorithm to screen synthesis parameters for materials that have similar composition and/or structure but not synthesized yet.

Data Origin and Access (*data must be available and sharable with data science teams* – please address: data source/origin, access privileges, sharing privileges)

The synthesis parameters for oxide heterostructures are available in the literature from a plethora of publications as one can find from the database such as Google Scholar and Web of Science.

Project Description (approx. 1.5 pages, plus figures and references; please describe data size, form, dimensionality, uncertainties, number of examples, etc.)

The broad spectrum of mechanical, optical and electronic properties exhibited by oxides offers tremendous opportunities for science and technologies[1-3]. Owing to the dependence of properties on crystalline directions, it is often optimal to grow functional oxide thin films in particular directions to maximize film properties for a specific application. Pulsed laser deposition (PLD) technique is widely exploited for producing high quality oxide thin films[4]. The non-equilibrium process involved in the PLD method, however, makes the design of synthesis routes difficult. To further advance the research on functional oxides, a comprehensive approach for computationally screening synthesis parameters via the PLD process across broad categories of oxide materials is in great demand.

In this work, the PIs aim to accomplish this goal by the following approach:

1. The first is to construct a text-mined synthesis database for all interested oxide systems such as ZnO/(MgZn)O, $CaMnO_3$/$CaRuO_3$, $LaAlO_3$/$SrTiO_3$, $LaMnO_3$/$SrMnO_3$, $Y_3Fe_5O_{12}$/$Gd_3Ga_5O_{12}$ and La_2CoMnO_6/$SrTiO_3$. Open access online databases such as Google Scholar, Web of Science and PubMed can be utilized to locate relevant publications. In brief, the experimental method sections of journal articles should be processed automatically to extract synthesis information. For PLD method, the most critical parameters include *laser intensity, substrate temperature, processing gas pressure, laser repetition rate and working distance between target and substrate,* which form a 5-dimensional synthesis space for each material.

   To the best of the PIs' knowledge, journal articles can be programmatically queried and downloaded using CrossRef search Application Programming Interface (API)[5]. Natural learning processing (NLP) algorithms[6-8] can be exploited to convert articles into

machine-readable synthesis parameters. The PIs welcome any creative approach conceived by the data science team to accomplish this step.

2. Once the database is established, statistical distribution of synthesis parameters for various materials can be investigated. For example, one can study the substrate temperature distributions for rock salts like MgO, perovskites like $CaTiO_3$, spinels like $MgAl_2O_4$ and garnets like $Y_3Fe_5O_{12}$. Another example to reveal further relations in the database is to obtain statistical information about synthesis parameters for materials on specific substrates such as perovskites on $SrTiO_3$ and garnets on $Gd_3Ga_5O_{12}$.
3. Based on the statistical knowledge obtained in the above step, one can analyze the situation to identify the key factors that drive synthesis outcomes as well as to formulate a machine-learning model that allows virtual screening of key growth parameters. One possible solution is to inspect the probabilistic model learned by a decision tree and automatically select a predictive set of synthesis parameters that drive the behavior of the model as described in Ref [9]. However, the PIs are willing to work with the data science team to invent a novel route to achieve this goal.
4. One way to benchmark the model established in Step 3 against the known examples is as the following:
    a. By studying the synthesis database for perovskite materials (such as $CaTiO_3$, $SrTiO_3$ and $BaTiO_3$), one can expect to extract information about key factors that influence the synthesis of single crystals of such structure.
    b. A machine-learned phase diagram for crystalline and non-crystalline phase can then be drawn.
    c. The result from Step b can be compared with the literature. For example, for $LaAlO_3/SrTiO_3$ system, both crystalline and amorphous $LaAlO_3$ films have been produced experimentally, which can serve as a reference.
5. After being trained with existing materials, the algorithm can be exploited to screen growth parameters for systems that do not exist in the database. One example can be $LaMnO_3/KTaO_3$: $LaMnO_3$ layer is magnetic, and $KTaO_3$ substrate hosts two-dimensional electron gas at the surface[10]. When these two materials meet each other, two-dimensional spin-polarized electron gas may form at the interface.
6. Based on the parameters predicted by the model, the PIs will carry out the synthesis experiments with the pulsed laser deposition facility at West Virginia University.
7. After the sample is synthesized, the PIs will characterize its structural and physical properties with X-ray diffraction, Atomic Force Microscope and transport measurement facilities at West Virginia University. The investigation results will be shared with the data science team to refine the algorithm until the screened parameters can be confirmed by experiments.

The success of this project will offer 1) a database containing aggregated synthesis information for most of the existing oxide heterostructures; 2) unique data-driven material informatics strategies that will guide the synthesis of new oxide heterostructures with novel functionalities as designed by computational approaches.

References

1 Hwang, H. Y. *et al.* Emergent phenomena at oxide interfaces. *Nat. Mater.* **11**, 103 (2012).

2 Schlom, D. G., Chen, L.-Q., Pan, X., Schmehl, A. & Zurbuchen, M. A. A thin film approach to engineering functionality into oxides. *J. Am. Ceram. Soc.* **91**, 2429 (2008).

3 Mannhart, J. & Schlom, D. G. Oxide interfaces: An opportunity for electronics. *Science* **327**, 1607 (2010).

4 Krebs, H. U. *et al. Pulsed laser deposition (PLD): A versatile thin film technique*. (2003).

5 Lammey, R. CrossRef text and data mining services. *Insights* **28**, 62 (2015).

6 Hawizy, L., Jessop, D. M., Adams, N. & Murray-Rust, P. ChemicalTagger: A tool for semantic text-mining in chemistry. *J. Cheminf.* **3**, 1 (2011).

7 Jones, D. E., Igo, S., Hurdle, J. & Facelli, J. C. Automatic extraction of nanoparticle perperties using natural Language Processing: NanoSifter an application to acqurie PAMAM dendrimer properties. *PLoS One* **9**, e83932 (2014).

8 Swain, M. C. & Cole, J. M. ChemDataExtractor: A toolkit for automated extraction of chemical information from the scientific literature. *J. Chem. Inf. Model* **56**, 1894 (2016).

9 Raccuglia, P. *et al.* Machine-learning-assisted materials discovery using failed experiments. *Nature* **533**, 73 (2016).

10 Zou, K. *et al.* $LaTiO_3$ /$KTaO_3$ interfaces: A new two-dimensional electron gas system. *APL Mater.* **3**, 036104 (2015).