

# Machine Learning for Climate Change: Guiding Discovery of Sorbent Materials for Direct Air Capture of CO<sub>2</sub>

**Diana L. Ortiz-Montalvo**, A. Gilad Kusne, Austin McDannald,  
Daniel W. Siderius, Kamal Choudhary, and Taner Yildirim

[Diana.Ortiz-Montalvo@NIST.gov](mailto:Diana.Ortiz-Montalvo@NIST.gov)

*2021 ICML - CCAI workshop*

# NIST Disclaimer:

Certain commercial equipment, instruments, or materials are identified in this presentation in order to specify the experimental procedure adequately. Such identification is not intended to imply recommendation or endorsement by NIST, nor is it intended to imply that the materials or equipment identified are necessarily the best available for the purpose.

The opinions, recommendations, findings, and conclusions stated here do not necessarily reflect the views or policies of NIST or the United States Government.

## 1 Climate Change – CO<sub>2</sub>

Carbon dioxide (CO<sub>2</sub>) in the atmosphere has contributed to  $\approx 1^\circ\text{C}$  global temperature. Many countries have pledged to reach net-zero emissions of greenhouse gases (e.g., CO<sub>2</sub>) by 2050.

## 3 Material Discovery

The space of potential sorbent materials is vast, growing exponentially with each new element considered for material chemistry. Therefore, historical trial-and-error methods for material discovery are infeasible.

## 2 Direct Air Capture (DAC) of CO<sub>2</sub>

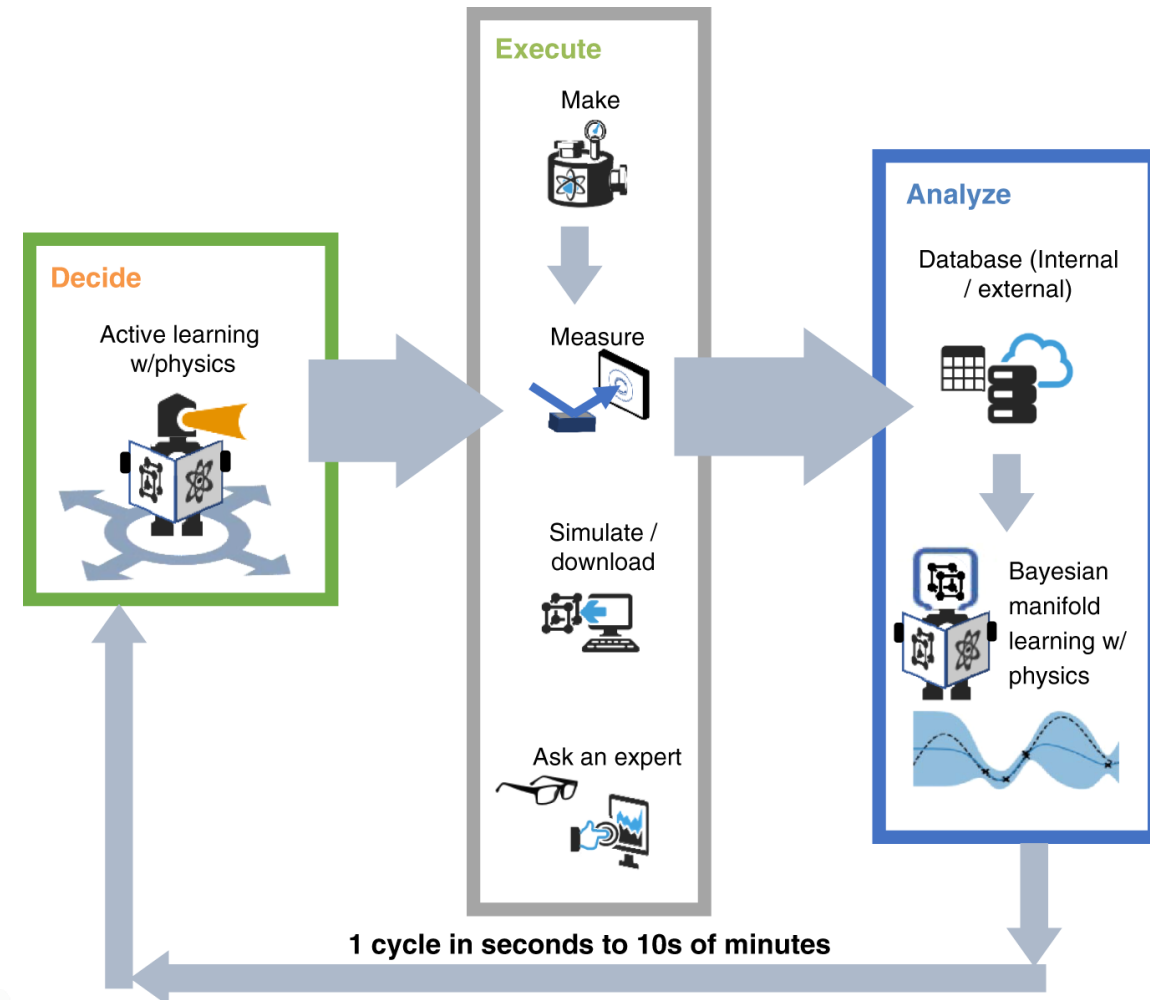
DAC has substantial potential to mitigate climate change. DAC is in an early development stage [1], and novel specialized sorbent materials are needed for DAC viability and success [2].

## 4 Artificial Intelligence (AI)

AI offers a solution to the challenge of discovering next-generation materials. Active learning can help optimize experimental design (lab or simulation) by identifying the most information-rich experiment to perform next.

# Machine Learning & Materials Discovery

- **Machine Learning (ML)** is now regularly used to learn from past experimental and computational data to predict potential novel, advanced materials for different technologies. For example:
  - Discovery of new best-in-class phase change memory material [3]
  - ML used to identify > 30 new candidate superconductors [4]
  - ML guided discovery of a new thermoelectric material [5]
- **Active Learning** has been shown to guide predictions toward the most promising materials, accelerating materials discovery and optimization by orders of magnitude. [3,6-8]
- There are **large open access databases** of experimental and computed material properties.
  - e.g., in-house: NIST/ARPA-E Database of Novel and Emerging Adsorbent Materials (<https://adsorption.nist.gov>)
- **Our interest:** unleash these methods on the open datasets with the goal of identifying potentially better materials for carbon capture, removal, etc. The outputs will be used to guide experimentalists in the lab = COLLABORATIONS!



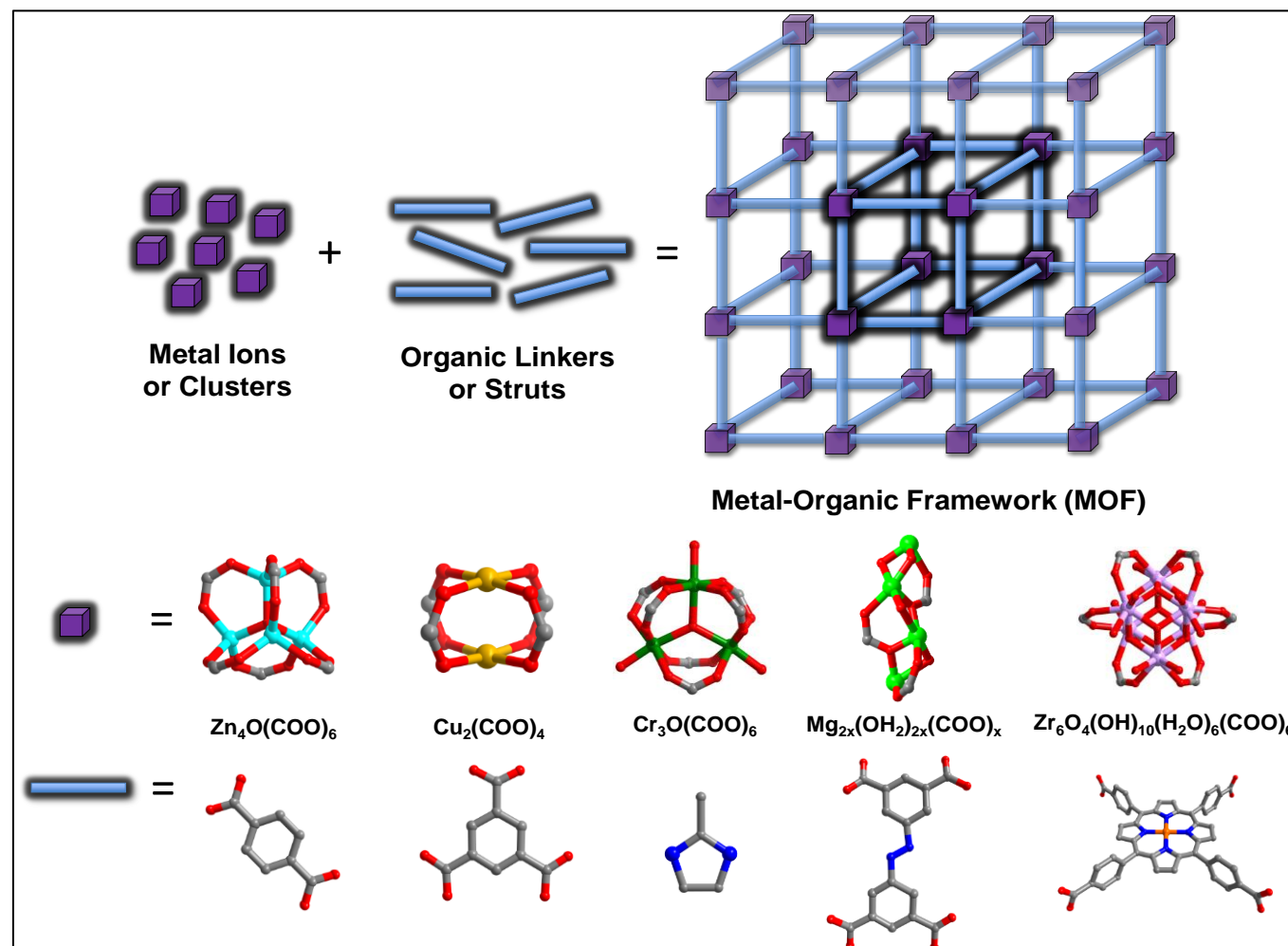
# MOFs = metal-organic frameworks

MOFs are the material-class of focus for this proposal.

They are hybrid materials that are built by assembling metal centers with organic linkers. [9]

Promising properties:

- large internal surface areas
- tunable, but uniform, channels and cavities
- high-capacity gas storage



Schematic representation of the self-assembly of a MOF

The key to the success of machine learning model training is the need for well-curated datasets of both experimental and simulation data.

Experimental and computational databases exist for sorbent materials such as MOFs, but still lack important materials properties, like molecule adsorption energies.

Crystal structure databases are available for experimental MOFs

- Inorganic Crystal Structure Database (ICSD) <https://icsd.products.fiz-karlsruhe.de/>
- Cambridge Structural Database MOF Subset (CSD) [10]
- Crystallography Open Database (COD) [11]

Density functional theory (DFT)-based computational databases are also available

- Computation-Ready, Experimental MOF (CoRE-MOF) [12]
- Quantum MOF (QMOF) [13]

Good resources to do machine learning training on perfect bulk MOFs!

# Proposal Plan

## Step 1

**Expand the NIST adsorbent databases** (<https://adsorption.nist.gov>)- link property measurements to resolved material structures (see next slide; slide 8)

## ...2

**Combine relevant experimental and simulation databases** currently available on MOFs into a new open-access NIST reference database. (slide 9)

## ...3

**Build a machine learning model** to learn from this database and predict the [important sorbent properties] for new materials.

## ...4

**Pair the predictions with active learning** to determine the most knowledge-rich simulations to perform next.

## ...5

**Add generated data** to the open-access database (see slide 9)

## ...6

**Collaborate with experimentalists** to evaluate feasibility of conducting autonomous experiments for the discovery of new DAC sorbent materials



# Step 1: Training MOF atomistic properties

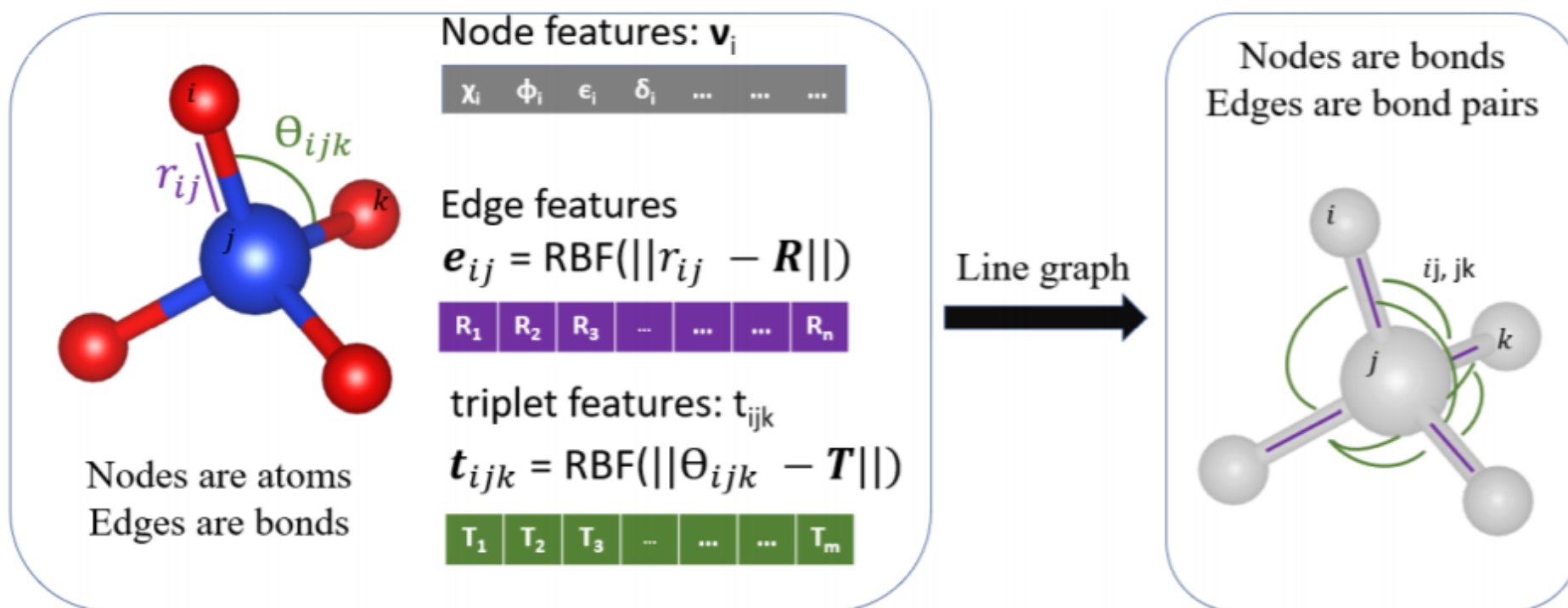
## Current progress:

Use workflow to make a line graph out of an atomistic structure.

Line graphs for MOFs will then be used as ingredients to deep learning models.

RBF = radial basis function

<https://arxiv.org/abs/2106.01829>



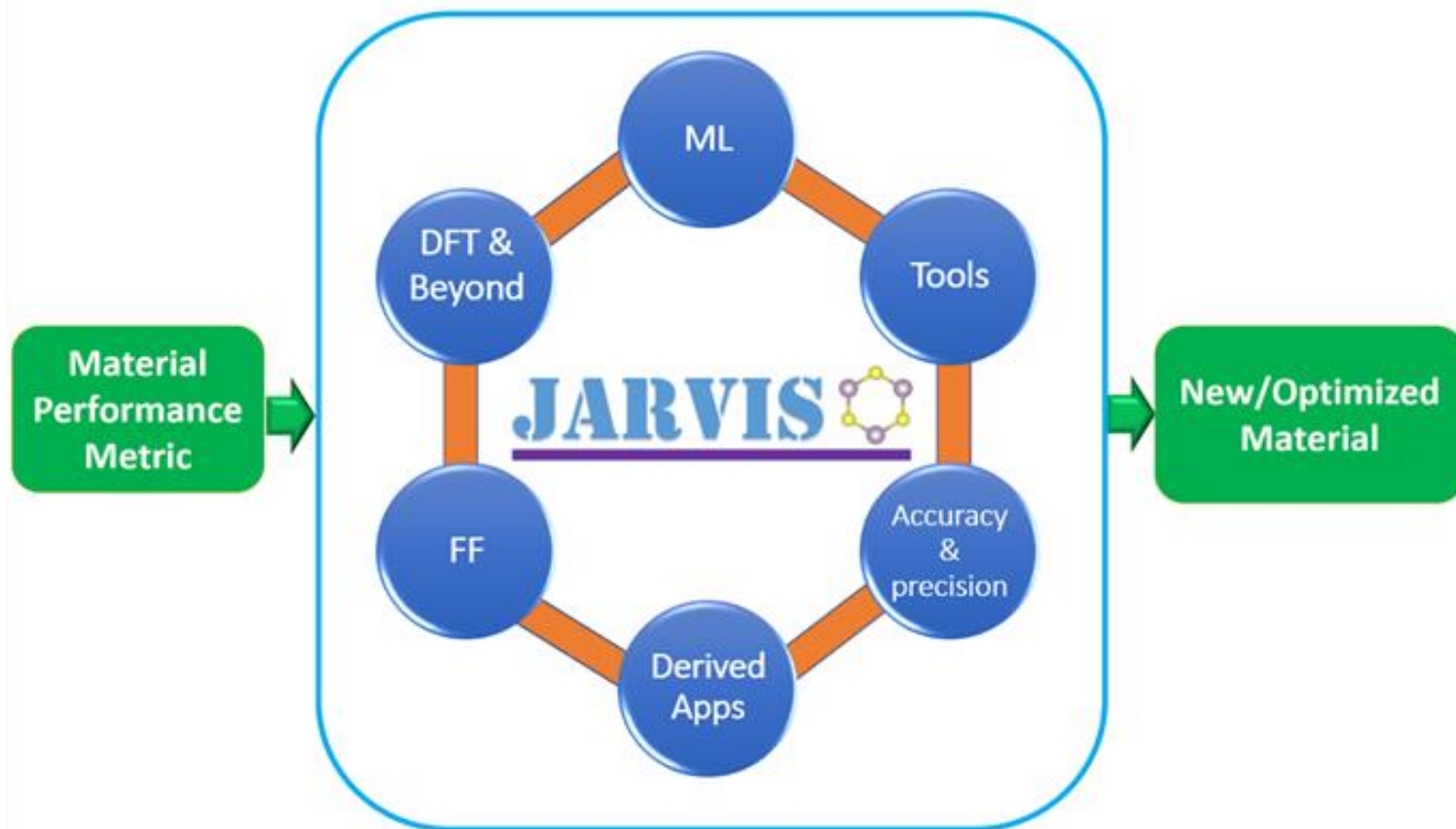


## Step 2: Open-access NIST database

Database development of MOFs properties combining experimental data and DFT-based simulations of pertinent properties (e.g., molecule adsorption energies).

JARVIS is a repository designed to automate materials discovery and optimization using classical force-field, density functional theory, machine learning calculations and experiments. [14]

### JARVIS (Joint Automated Repository for Various Integrated Simulations)



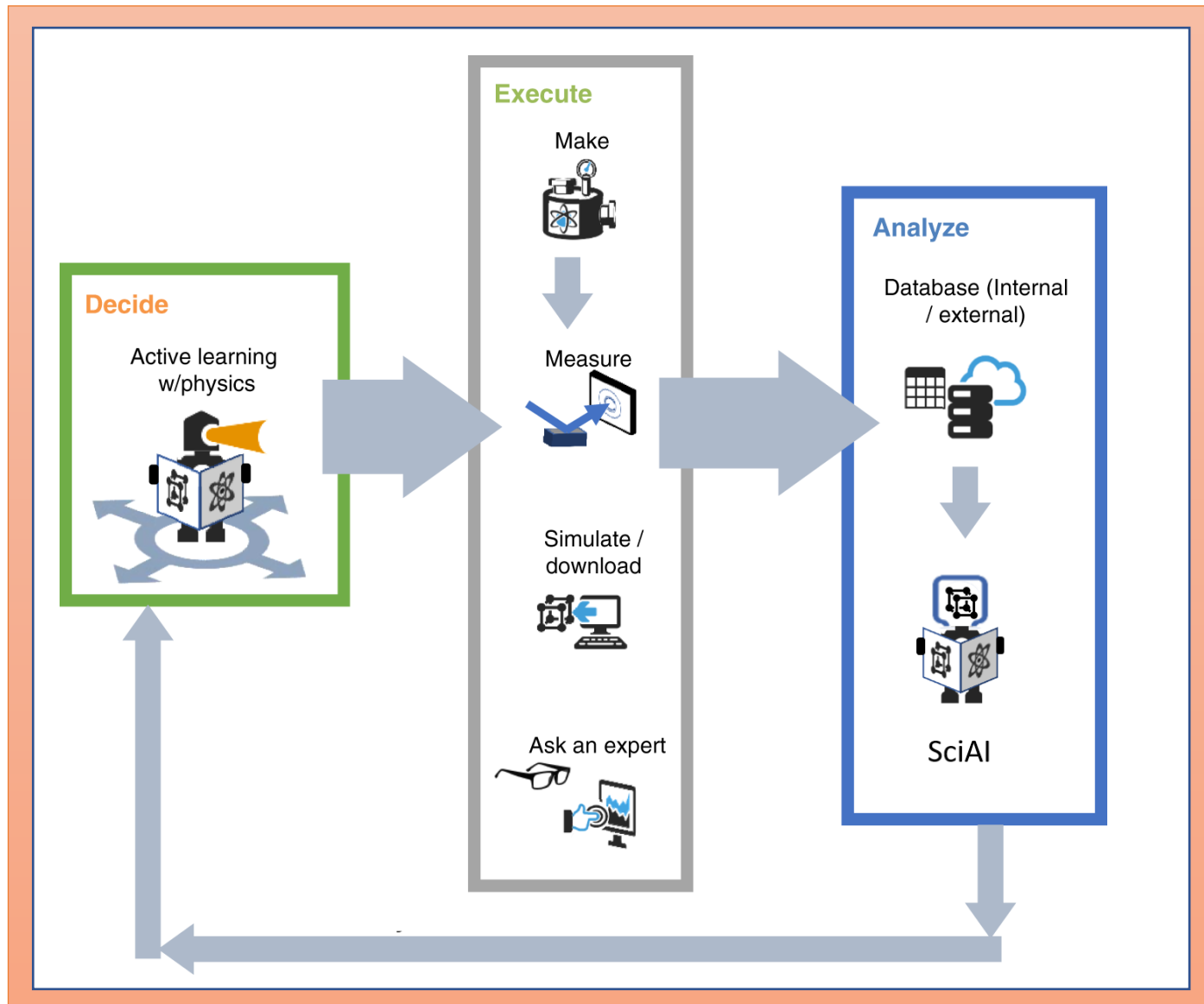
Find more details about JARVIS in:

<https://www.nature.com/articles/s41524-020-00440-1>

Questions? Contact [Kamal Choudhary](mailto:kamal.choudhary@nist.gov) ([kamal.choudhary@nist.gov](mailto:kamal.choudhary@nist.gov))

# End-goal: Discovery of Advanced MOFs...

...using Autonomous Experimentation



Autonomous System



Discovery of Advanced MOFs

Closed-loop autonomous materials exploration and optimization (CAMEO). (Fig 1 from [Kusne et al. 2020](#) [3])

- [1] Azarabadi, H. and K.S. Lackner, *A sorbent-focused techno-economic analysis of direct air capture*. Applied Energy, 2019. **250**: p. 959-975.
- [2] Shi, X., et al., *Sorbents for the Direct Capture of CO<sub>2</sub> from Ambient Air*. Angewandte Chemie International Edition, 2020. **59**(18): p. 6984-7006.
- [3] Kusne, A.G., et al., *On-the-fly closed-loop materials discovery via Bayesian active learning*. Nature Communications, 2020. **11**(1): p. 5966.
- [4] Stanev, V., et al., *Machine learning modeling of superconducting critical temperature*. npj Computational Materials, 2018. **4**(1): p. 29.
- [5] Iwasaki, Y., et al., *Machine-learning guided discovery of a new thermoelectric material*. Scientific Reports, 2019. **9**(1): p. 2751.
- [6] MacLeod, B.P., et al., *Self-driving laboratory for accelerated discovery of thin-film materials*. Science Advances, 2020. **6**(20): p. eaaz8867.
- [7] Nikolaev, P., et al., *Autonomy in materials research: a case study in carbon nanotube growth*. npj Computational Materials, 2016. **2**(1): p. 16031.
- [8] Zhong, M., et al., *Accelerated discovery of CO(2) electrocatalysts using active machine learning*. Nature, 2020. **581**(7807): p. 178-183.
- [9] Li, H., et al., *Recent advances in gas storage and separation using metal–organic frameworks*. Materials Today, 2018. **21**(2): p. 108-121.
- [10] Moghadam, P.Z., et al., *Development of a Cambridge Structural Database Subset: A Collection of Metal–Organic Frameworks for Past, Present, and Future*. Chemistry of Materials, 2017. **29**(7): p. 2618-2625.
- [11] Gražulis, S., A. Merkys, and A. Vaitkus, *Crystallography Open Database (COD)*, in *Handbook of Materials Modeling : Methods: Theory and Modeling*, W. Andreoni and S. Yip, Editors. 2018, Springer International Publishing: Cham. p. 1-19.
- [12] Chung, Y.G., et al., *Advances, Updates, and Analytics for the Computation-Ready, Experimental Metal–Organic Framework Database: CoRE MOF 2019*. Journal of Chemical & Engineering Data, 2019. **64**(12): p. 5985-5998.
- [13] Andrew, R., et al., *QMOF Database*. 2021.
- [14] Choudhary, K., et al., *The joint automated repository for various integrated simulations (JARVIS) for data-driven materials design*. npj Computational Materials, 2020. **6**(1): p. 173.

# Thank You!

[Diana.Ortiz-Montalvo@NIST.gov](mailto:Diana.Ortiz-Montalvo@NIST.gov)