MOSY: *Encoding* Synthesis Data for *Decoding* Process-Property Relations

DATA SCIENCE INSTITUTE
COLUMBIA UNIVERSITY

Yevgeny Rakita[1], Simon J.L. Billinge[1], Ran Gu[1], Paul Todd[2], Jamie Neilson[2]

Colorado State University

[1] Columbia University, Applied Physics and Applied Mathematics ; [2] Colorado State University, Chemistry department

In a world of "big-data", the best way to encode information about chemical synthesis is under debate. Currently, databases typically define materials based on their structure, where for crystalline solids it is standardized in crystallographic forms using CIF - crystallographic information framework. We define a sample as the set of steps that were taken to produce the sample. Using directed acyclic graph (**DAG**) representation, with nodes as actions linked by a sequence that is encoded edges that join the nodes, opens up the possibility to use computational tools. We call this representation **MOSY**, or Movement Motivated Synthesis. Linking characterization data, such as diffraction, calorimetric and optical spectra, to the point in the graph where the characterization was carried out, we will generate machine-learning-based correlative models that link a process with properties. This is hopefully will result possessing process-property dependence, which is currently neglected, and accelerate material discovery.

Question

What will be a standardized encoding of experimental synthesis procedures which facilitates data analysis and machine learning?

Strategy

MOSY : Movement Motivated Synthesis

Independent [x]

Movements

do something

do something

Dependent [f(x)]

?

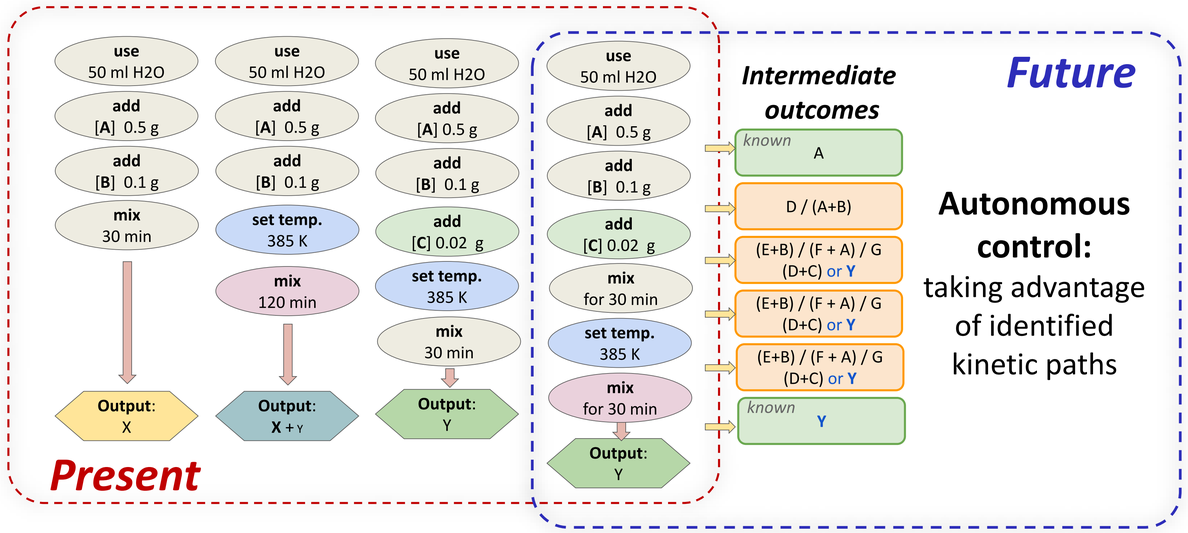
Motivated Synthesis

Product with some promising properties

Goal

- Identify critical paths = greatest potential to make a difference (data source: **lab**)
- Further design of an in-situ experiment (data source: **UV-Vis-IR, NMR, Synchrotron**)
- **Draw kinetic path as a function of "movements"**
- Compare with thermodynamic / kinetic calculations

Challenge

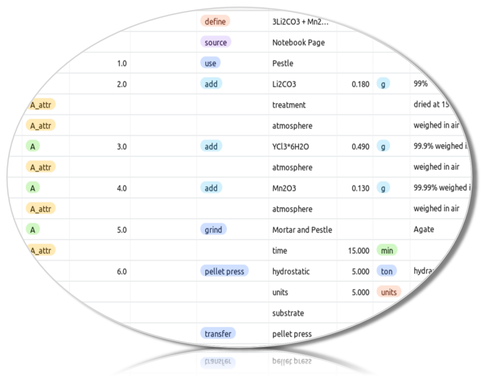


Stage 1

Manual MOSY data collection

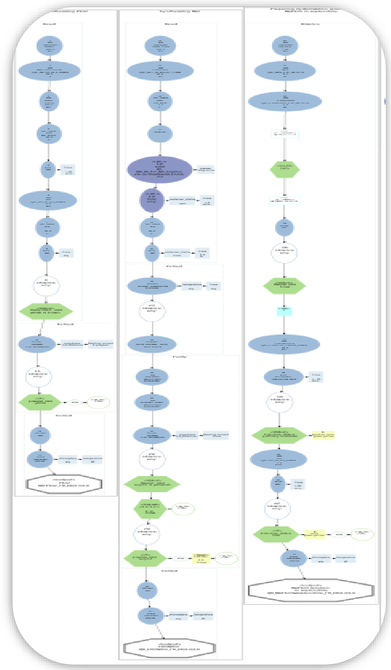
*Thanks to Paul, and still … **WE NEED MORE !***

Platform for inputting data



***Auto*-generation and visualization of directed graphs**

To-be used as a platform for analysis and control



Stage 2

(semi-) **Automatically MOSY data collection**

Looking for collaborators for building infrastructure for automatic data collection in chemical labs . . .

This work was supported as part of GENESIS: A Next Generation Synthesis Center, an Energy Frontier Research Center funded by the U.S. Department of Energy, Office of Science, Basic Energy Sciences under Award Number DE-SC0019212

U.S. DEPARTMENT OF ENERGY | Office of Science