Extended Abstract: Interactive Leaderboard for Requesting and Tracking Expensive Calculations of Optional Properties across a Database of Materials

Donald Winston,* Lawrence Berkeley National Laboratory; Joseph H. Montoya, Lawrence Berkeley National Laboratory; and Kristin A. Persson, Lawrence Berkeley National Laboratory
**Corresponding author address:* Energy Technologies Area, Lawrence Berkeley National Laboratory, Berkeley, CA, 94720, USA; email: dwinston@lbl.gov

Abstract: *Databases of crystalline compounds typically have core properties such as relaxed structure, energy of formation, etc. pre-computed and available to users. However, additional material properties such as full elastic tensor, band structure, etc. may be both not of interest to all users of the database and computationally expensive to determine. Furthermore, researchers interested in such properties may only be interested in their values for specific compounds or chemistries. To crowdsource prioritization of optional-property calculations, we built an interactive leaderboard that serves as a gateway to our cyberinfrastructure (CI) for queuing jobs and dynamically reassigning job priorities. Users browse an ordered list of queued jobs, filter by chemistry for materials that are missing the property of interest, vote on existing or new jobs, and receive email notification when voted-on material properties are online. Here we will detail the use case of full elastic tensor calculations for the Materials Project, though the leaderboard interface can be run as a standalone web service and customized for contexts other than computational materials science.*

1. Introduction

The Materials Project (MP) [l, 2] aims to mitigate guesswork from materials design in a variety of applications by computing properties of known and predicted materials so that experimental research can target the most promising compounds from computational data sets. MP develops open-source codes to ensure that long-running ab-initio simulations can be run in high throughput at supercomputing centers – from defining, managing, and executing complex workflows [3] to domain-specific failure detection and recovery [4]. Thus, in service to its scientific user community, MP is limited less by human time needed to shepherd a job through to completion and more by overall computational budgets. Because materials scientists work on a variety of applications, a given user may be interested in values computed for only a subset of material properties for which we have developed benchmarked workflows, and even then only for particular materials.

Here we introduce the structure and function of a gateway for users to upvote and track calculation of material properties of interest to them. Given a use case of full elastic tensor calculations for crystalline compounds, we discuss:

- data stores and processes for both vote monitoring by the computing service and job status monitoring by the leaderboard and notification services,
- serving a leaderboard of queued jobs with per-user limits on active votes and links to detail pages for each job's workflow progress, and
- filtering and sorting database entries for candidates of interest given declarative configuration both of query structure and of a scoring column secondary to votes.

2. Data Stores and Processes for Voting, Tracking, and Notifying

As detailed in Fig. 1, the interactive leaderboard for property requests and notification (ILPRN) extends existing infrastructure for running calculations by maintaining a collection of votes and notifying users by email when their voted-on properties are online. The subsystem used for submitting, running, and post-processing calculations [5] exposes key database collections and consumes the vote-tracking collection through its submission manager via polling, though because we use MongoDB for database collections, so-

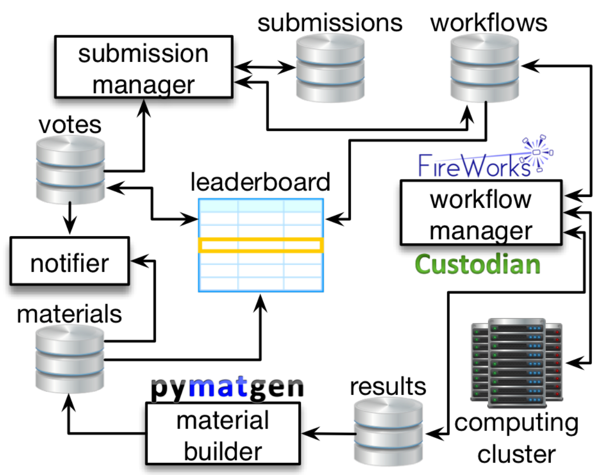

Fig. 1. Data stores and processes. The leaderboard pushes only to a votes database and pulls from existing databases provided by workflow and result-aggregation processes.

called "oplog tailing" on a replica set may be used in future for real-time updates.

3. The Leaderboard Service: Basic Display and Interaction

As shown in Fig. 2, the ILPRN web service provides filtering and sorting customized to the entries and property under review while offering generated links to detail pages provided by external services. Users vote on entries of interest, with a configurable maximum active vote count to incentivize users to consider the scarcity of gateway resources. The view's initial state, without e.g. a filter on chemistry, shows only entries with votes so that a user's first inclination may be to aid a fellow gateway member.

4. Generalized Querying and Sorting via Declarative Configuration

A subset of an example JSON configuration file is shown in Fig. 3. Filtering is by chemical system (e.g. Li-O) or reduced chemical formula (e.g. Fe2O3). An optional pre-filter allows an input chemical system to be sorted by elements so that the database query is covered by a collection index. Sorting is by the aforementioned energy above hull.

5. Conclusion

To be of general interest to the science gateways community, we seek to ensure that ILPRN is both useful outside of MP's existing infrastructure for computational materials science and configurable in a way that requires no modification of source code. We acknowledge that use of MongoDB is currently necessary for integration, if only as glue to connect existing cyberinfrastructure to ILPRN.

6. Acknowledgments

We acknowledge support from the Materials Project Center, under the Department of Energy (USA), Basic Energy Sciences (USA), Grant No. EDCBEE.

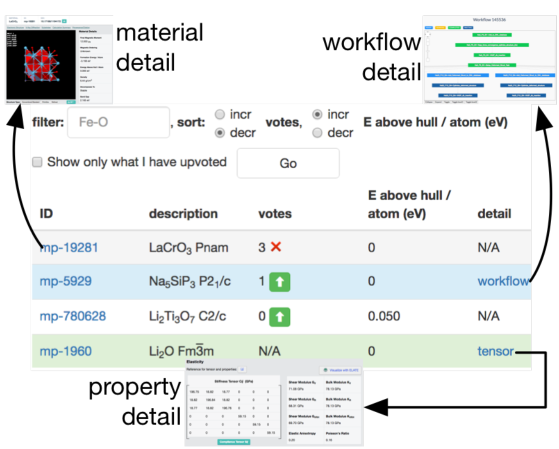

Fig. 2. Leaderboard display of elastic tensor calculations for crystalline materials. One declaratively configures text-input filtering, secondary sorting, and hyperlinking to details.

```
"filter": {
  "placeholder": "Fe-O",
  "matches": [{
    "field": "chemsys",
    "regex": "^[A-Za-z-]+",
    "pre": "prep.sort_elements"
  }, {
    "field": "reduced_formula",
    "regex": "^[A-Za-z2-9]+$"
  }]},
"extrasort": {
  "field": "e_above_hull",
  "label": "E above hull / atom (eV)",
  "default": "incr"}
```

Fig. 3. Configuration for filtering and sorting the *entries* collection. Similar settings specify collection connections, set vote limits, and help generate cell values for the shown table.

7. References

[1] A. Jain, S. P. Ong, G. Hautier, W. Chen, W. D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder, and K. A. Persson, “Commentary: The Materials Project: A materials genome approach to accelerating materials innovation,” APL Materials, vol. 1, no. 1, p. 011002, 2013.

[2] https://materialsproject.org

[3] A. Jain, S. P. Ong, W. Chen, B. Medasani, X. Qu, M. Kocher, M. Brafman, G. Petretto, G.-M. Rignanese, G. Hautier, D. Gunter, and K. A. Persson, “FireWorks: a dynamic workflow system designed for high-throughput applications,” Concurrency Computat.: Pract. Exper., vol. 27, no. 17, pp. 5037–5059, May 2015.

[4] S. P. Ong, W. D. Richards, A. Jain, G. Hautier, M. Kocher, S. Cholia, D. Gunter, V. L. Chevrier, K. A. Persson, and G. Ceder, “Python Materials Genomics (pymatgen): A robust, open-source python library for materials analysis,” Computational Materials Science, vol. 68, pp. 314–319, Feb. 2013.

[5] https://github.com/materialsproject/MPWorks