I recently returned from the first annual Acceleration Conference hosted by the Acceleration Consortium at the University of Toronto in Canada (August 30th to September 2nd, 2022; see Acceleration Consortium). Over the course of 4 days, the event brought together thought leaders in academia, industry, and government to help consolidate the concept of accelerated science—research driven by artificial intelligence (AI), automation, machine learning (ML) and advanced computing—in the area of materials and molecular discovery. The goal was clear: bespoke materials solutions developed to solve society’s most pressing problems, from consumer electronics to drugs, renewable energy, and sustainable plastics. Sessions ranged from technical (e.g., Methods in Machine Learning, Data Analysis, and Modeling) to applied (e.g., Discoveries in Energy), to engaging panel discussions (e.g., How to Commercialize Materials-on-Demand). Also, some time for Paloma cocktails and salsa dancing. It was a lot to take in.

With such a jam-packed and demanding schedule digesting a large variety of information, it was only afterwards I was able to reflect on the event and its potential in the broader scope of materials in general. Indeed, AI, ML, and so-called big data (BD) approaches are the means du jour across modern research and development, from fintech to healthcare to self-driving cars as well as physical sciences such as chemistry and materials. In the past decade, data science (and its related computational approaches) has been labeled the fourth paradigm of scientific discovery, where paradigms one to three are (1) empirical evidence, (2) scientific theory, and (3) computational science. Where data science diverges from traditional computational science is essentially in the control of variables—prior computational methods were more akin to simulated experiments, directed by input control parameters with a desired/predicted response. Modern data science is more open-ended—we don’t know a priori the key variables or the output—the AI/ML algorithms discover new rules, principles, and mechanisms within the data, via variables and patterns either overlooked or undetectable by other means. It’s a kind of simulated microscopy able to explore the multidimensional space of “data.” Spooky stuff.

Clearly there is excitement in the field, and we may very well be at the cusp (or in the midst) of an unprecedented Kuhnian shift. But it is important to reflect on the field before diving headfirst into the hype and adding AI and ML to every research project. Paraphrasing Jason Hattrick-Simpers from the Acceleration Conference, “useful models are useful, but not all models are useful.”

In materials, in simplest terms, the majority of researchers are utilizing AI/ML/BD approaches to efficiently explore the parametric space of materials design and/or discovery. The digital approach enables efficient mapping across variables as well as data generation. However, the parameter space is still intractable. While AI/ML/BD approaches can provide (and “learn”) hidden relationships and mechanisms that experimental approaches may overlook, current projects/efforts are still limited to a narrow definition of the problem and parametric space. For example, there are multiple studies in the literature searching potential metal organic framework (MOF) platforms for carbon capture (with various degrees of rigor; for an early example, see Fernandez et al.). This (1) presumes that MOFs are the solution a priori and (2) limits the search to carbon capture (of course, this is an oversimplification for discussion). While such a study could screen tens of thousands of MOFs (both known and unknown) and produce nice candidate high-performance materials, we are effectively applying AI/ML/BD in a traditional manner: a focused suite of experiments in a defined parameter space. A flashy new technique, but the same problem formulation.

When such studies are written and submitted, I see a lot of Ashby plots showcasing the gains in a property or two. The “newly discovered” material outperforms its predecessors by some metric by maybe 20% or 50%. But there really isn’t anything new. Thinking of other emerging materials systems, perovskites are currently a hot topic, ever pushing the record power conversion efficiency (PCE) of perovskite-based solar cells (around 26% at the time of this writing) closer to industrial viability. However, the beginnings of perovskites for photovoltaic applications were rather meager (PCE on the order of 3.8% in 2009). While current AI/ML/BD methods are being used to explore the “perovskite space,” it begs the question, would (current) digital discovery methods have come up with the perovskite structure (ABX3) and all of its derivatives for solar cells? Would the preliminary 3.8% be sufficient to evolve the system within an algorithm? Only human direction and
guidance has created the “push” for energy conversion efficiency. Indeed, looking at all the systems on the NREL (National Renewable Energy Laboratory) plot, all increases are in “known” materials. I have yet to see “digital” methods produce a completely unprecedented system—something not even on the Ashby plots. Clearly this is the holy grail of digital discovery, more specifically, materials acceleration platforms.

Pondering this, what exactly are we accelerating? With a computational materials and engineering background, I feel sometimes my mind wanders to symbolic representations. Putting in a more rigorous “mathematical framework” to materials acceleration, the math is elementary. We can start by defining the current state (position) of scientific knowledge as $x = \text{science}$. Consequently, the pursuit of new knowledge over time, i.e., research, can be defined as:

$$\frac{dx}{dt} = \frac{d}{dt}(\text{science}) = \text{research}$$

From which it clearly follows that acceleration is the change in research over time:

$$\frac{d^2x}{dt^2} = \frac{d^2}{dt^2}(\text{science}) = \frac{d}{dt}(\text{research}) = \text{acceleration}$$

$$\frac{d}{dt}(\text{acceleration}) = \text{Acceleration Conference}$$

As a scientific editor, I have to somehow place the role of publications here, which easily arise in the integration of research:

$$\int \text{research} \, dt = \text{science} + C$$

where the constant of integration, $C$, is the new publication or series of publication(s). Quite a lovely gedankenexperiment. Linking materials acceleration to kinematic acceleration, however, is more analogous than simple derivatives.

Two things to consider are (1) there are different types of acceleration, and (2) acceleration is a vector. Indeed, acceleration is not always beneficial. Consider a satellite or spaceship in orbit: such an object can be subject to constant centripetal acceleration about the center of mass. For our science analogy, this represents many of the current ML/AI/BD studies—they are accelerating research but frequently lack a truly innovative direction! The pull of “status quo” limits our conceptual advancement (Figure 1). And similar to orbital mechanics, to escape the “gravitational pull” of the status quo, research must reach the necessary escape velocity. Acceleration is necessary but requires both sufficient direction and energy.

Looking at direction first, the problem is nontrivial. Leaning on our spaceflight comparison, consider a flight to the moon. Far from linear, lunar trajectories must consider multiple factors, including launch, escaping the gravitational pull of the Earth, translunar coast, lunar orbit insertion, the motion of the Earth around the sun, the motion of the moon around the Earth, etc. The resulting path is extremely complex, requiring a multitude of accelerations and direction changes (Figure 2). This, of course, is extremely similar to traditional science—areas of incremental change, followed by periods of acceleration and changes in direction. But for both the moon and science, the target is usually set. We know where we want to go, what problem we want to solve (again, like searching MOFs for carbon capture). What happens when we can’t see the moon? How can we plan a complex trajectory when the “target” is unknown?

This is not to say we should aimlessly fly without a target or endgame (even so-called moonshot science has a goal), but merely that sometimes (1) the final destination is unplanned and (2) the path to get there has much more twists, turns, and tangents than anticipated. Tunnel-vision-like research can be useful for incremental improvements but rarely results in transformational concepts.

Moving on from direction, the second requirement is sufficient energy. Keeping with out kinematic analogy, we can define the kinetic energy as:

$$K = \frac{1}{2} m (\frac{dx}{dt})^2 = \frac{1}{2} m(\text{research})^2$$

Now we have introduced mass into the mix of things. How does that fit into our analogy? Mass can be interpreted as the number of researchers undertaking the research in a certain field. Consider a lone researcher, can they make contribution to “science”? Yes, but their research output $(dx/dt)$ must be relatively large. Ten researchers working in the same field provides a larger push. Hundreds provide even more. The more researchers, the more mass involved, the more inertia to drive progress in new directions. Indeed, returning to the example of perovskite-based solar cells, it is precisely because so many researchers are working in the area that
incremental improvements in PCE are occurring relatively frequently.

To escape the pull of “status quo” in a new field such as AI/ML/BD, critical mass much be reached (along with the necessary acceleration to achieve escape velocity). Indeed, this concept of pulling the community together and reaching critical mass is one of the driving motivations of the recent Acceleration Conference as well as a related call for papers. This call, focused on ideas/concepts for materials acceleration platforms (MAPs), we titled “CriticalMASS” to reflect the need for community members to reach a critical size and initiate a chain reaction of acceleration materials discovery. That call is still open, and I invite anyone interested to reach out to myself or any of my co-authors.

I like the idea of position, velocity, and mass and their relationships mapping to the research space. It was a nice framework. As I continued pondering such things and the trajectory of materials discovery, swimming in my head were terms like trajectory, acceleration, journey, path. At the nexus of these thoughts—along with my fondness for pop culture—arose an idea of the “Kessel Run” of digital materials acceleration.

For the non-nerds, the Kessel Run was first introduced in the original Star Wars film, when Han Solo boasts about the speed of the Millennium Falcon stating “You’ve never heard of the Millennium Falcon? ... It’s the ship that made the Kessel Run in less than 12 parsecs.” The Kessel Run was a 20-parsec route used by smugglers to move “glitterstim spice” without getting caught by Imperial ships that were guarding the movement of spice from Kessel’s mines. You may have noticed that parsecs are a measure/unit of distance, not speed or time. What did Han mean by his boast?

In the Star Wars universe of faster-than-light hyperspace travel and “the Force,” one can stretch physical truths a little. While there are a few theories (including Han just being facetious), the one I like (supported by George Lucas) is that the reduction in parsecs (i.e., shorter distance) are due to the Millennium Falcon’s advanced navigational computer rather than its engines; i.e., the navicomputer could calculate much faster routes than other ships could, traversing through hyperspace in the shortest possible distance (parsecs).

Thus, it wasn’t the brute force acceleration that made the Millennium Falcon fast: it was the clever navigation system. Linking back to AI/ML/BD, larger datasets, more variables, and faster computers aren’t necessarily the best means to gain speed. The resulting “acceleration” and attained “speed” can be achieved by clever navigation—i.e., assuming the fixed (problem limiting) distance is variable and changing the trajectory of the Kessel Run. Perhaps Han installed some ML algorithms on the Millennium Falcon.

Clearly, I like to frame topics I am learning about (such as AI/ML/BD) with examples and analogies I know, be they perovskites, MOFs, or Star Wars. It helps me understand the potential and a good platform to ask the “what ifs.” There was great discussion at the Acceleration Conference, but caution must be heeded when everyone is in near-universal agreement conceptually. It becomes difficult to break away from the status quo. We all should be a little like Han Solo every now and then. Shoot first. Don’t be afraid to fly where others would be “crazy” to follow. Never ask the odds.

Comparing the acceleration of materials discovery to literal acceleration, we uncover a few key lessons—lessons I believe are critical to account for to enable true forward progress in the field. In sum:

1. Acceleration is useful in all fields, regardless of the method. But acceleration without direction can lead to simply more results relevant to the status quo. While there are definite benefits to using AI/ML/BD approaches to supplement “traditional” science, there is potential for more.

2. To escape the “pull” of traditional paradigms, the path is highly non-linear and complex. Be wary of tunnel-vision-like objectives. There is value in detours and necessary direction changes.

3. The community needs to achieve critical mass. Open access to data, algorithms, and concepts is necessary, facilitated by efforts such as the Acceleration Conference and our current CriticalMASS call for papers.

4. Acceleration need not be brute force—change the limiting variables, develop smarter algorithms. There is untapped potential in changing the variables, finding novel routes through the multi-dimensional “hyperspace” that AI/ML/BD allows.

While AI/ML/BD is clearly a popular method, I believe it has yet to prove itself deserving of the “fourth paradigm” of scientific discovery, particularly in the materials space. But the potential is there. Efforts arising from the community across all applications are exciting and inspirational. There is an excitement in the field that is palpable. A little more mass, a little more acceleration, and I believe it will reap rewards across a multitude of societal challenges. Join the progress. That contribution could be a white paper for our CriticalMASS call or a research article for Matter or Patterns or the journal of your choice.

Whether artificial intelligence, machine learning, big data, automation, robotics,
self-driving labs, etc., let’s accelerate materials using the “fastest ship in the fleet.”

DECLARATION OF INTERESTS

The author declares no competing interests.

REFERENCES

1. Acceleration Consortium. Home. https://acceleration.utoronto.ca/.
2. Batra, R., Song, L., and Ramprasad, R. (2021). Emerging materials intelligence ecosystems propelled by machine learning. Nat. Rev. Mater. 6, 655–678. https://doi.org/10.1038/s41578-020-00255-y.
3. Butler, K.T., Davies, D.W., Cartwright, H., Isayev, O., and Walsh, A. (2018). Machine learning for molecular and materials science. Nature 559, 547–555. https://doi.org/10.1038/s41586-018-0337-2.
4. Tolle, K.M., Tansley, D.S.W., and Hey, A.J.G. (2011). The fourth paradigm: data-intensive scientific discovery [Point of View]. Proc. IEEE 99, 1334–1337. https://doi.org/10.1109/jproc.2011.2155130.
5. Fernandez, M., Boyd, P.G., Daff, T.D., Aghaji, M.Z., and Woo, T.K. (2014). Rapid and Accurate machine learning Recognition of high performing metal Organic frameworks for CO2 capture. J. Phys. Chem. Lett. 5, 3056–3060. https://doi.org/10.1021/jz501331m.
6. National Renewable Energy Laboratory. Best research-Cell efficiency Chart. https://www.nrel.gov/pv/cell-efficiency.html.
7. Kojima, A., Teshima, K., Shirai, Y., and Miyasaka, T. (2009). Organometal Halide perovskites as Visible-light Sensitizers for photovoltaic cells. J. Am. Chem. Soc. 131, 6050–6051. https://doi.org/10.1021/ja809599r.
8. Seifrid, M., Hattrick-Simpers, J., Aspuru-Guzik, A., Kail, T., and Cranford, S. (2022). Reaching critical MASS: Crowdsourcing designs for the next generation of materials acceleration platforms. Matter 5, 1972–1976. https://doi.org/10.1016/j.matt.2022.05.035.

About the author

A graduate from Memorial University (Newfoundland, Canada), Stanford University (USA), and Massachusetts Institute of Technology (USA), Dr. Cranford was faculty at Northeastern University’s College of Engineering prior to accepting a role as founding editor-in-chief of Matter in 2018. Due to transitional opportunities, he has now assumed the role as interim editor-in-chief of Patterns. Prior to academic publishing, as a researcher he has published over 50 works in the field of computational materials sciences and is fully supportive of the emerging digital and data revolution not only in physical science but across all fields and applications.