Across Dimensions and Scales: How Imaging and Machine Learning Will Help Design Tomorrow's Energy Conversion Devices

Mariana I. Bertoni, Arizona State University

Energy is the single most important factor that impacts the prosperity of any society, underpinning advances on which we all depend. To supply all 7+ billion people on this planet the level of energy that the developed world is accustomed to, we would need to generate 60 terawatts, equivalent to 900 million barrels of oil per day. Where could this astonishing amount of energy come from? The commonly used term "The Terawatt Challenge" describes the sheer magnitude of the endeavor to produce energy at this level in an economically, socially and technologically sustainable way. When one searches for potential sources of energy at the terawatt scale it is striking to find that the biggest resources and most technically exploitable options are the ones that barely make up 10% of the energy mix today – solar, wind and geothermal.

The world energy breakdown will look quite different by 2050 if we indeed manage to solve the Terawatt Challenge and it is not unreasonable to think that renewables can handle a heavy load. Reports have proposed that 100% of the world's energy needs by 2030 (11.5 TW) can be fully achieved with an energy mix roughly composed of: 50% wind, 40% solar (CSP and PV), 4% hydroelectric, 4% geothermal, and 1% tidal turbines [1].

No renewable energy source is as abundant as the Sun and during the past six years we have seen its potential being capitalized to the point where solar has moved from niche generation to reaching grid parity and becoming a mainstream electricity generation source. Thanks to 15 GW_{DC} installed, solar was the number one source of new US capacity additions in 2016 with an unprecedented 39%. That momentum carried into 2017 with solar being 30% of all new electric capacity installed in Q1. Global PV shipments have reached an astonishing 75 GW in 2016, arguably making the solar industry the largest optoelectronic sector in the world, worth $110B/yr [2].

The aspirational goal set by the DoE Sunshot initiative to meet $1/Watt by 2020 initially seemed unrealistic and somewhat comical to many in the industry in early 2011 [3]. However, six years later and three years ahead of schedule, module prices have hit 0.99 $/W for fix-tilt utility size installations [2].

For 20 states the levelized cost of solar energy have fallen below gross electricity bill savings in the first year of a solar PV system's life. This means grid parity under business-as-usual conditions is a reality, with 42 more states expected to follow suit by 2020 [4]. PV reaching 'grid parity' establishes an incredible milestone, but this is just the beginning. Like Kurtz and co-authors suggested, in order to see high market penetration, PV systems costs must drop to cover the additional costs of storage or transmission so that solar generation can be dispatched to cost effectively meet electricity demand more broadly in both time and space [5].

The rapid pace of changes brings its own sets of challenges and opportunities. For example, there are increasing concerns about the maximum penetration that PV can accomplish due to its impact on utility demand, lowering its value as PV penetration increases, and requiring

further cost reductions. In addition, the important metrics of photovoltaics for sustainable energy are expanding to include factors previously not analyzed, such as the impact of capital expenditures on realizing high continued growth rates [6,7].
The technological barriers facing PV have in some ways increased; cost reductions from the

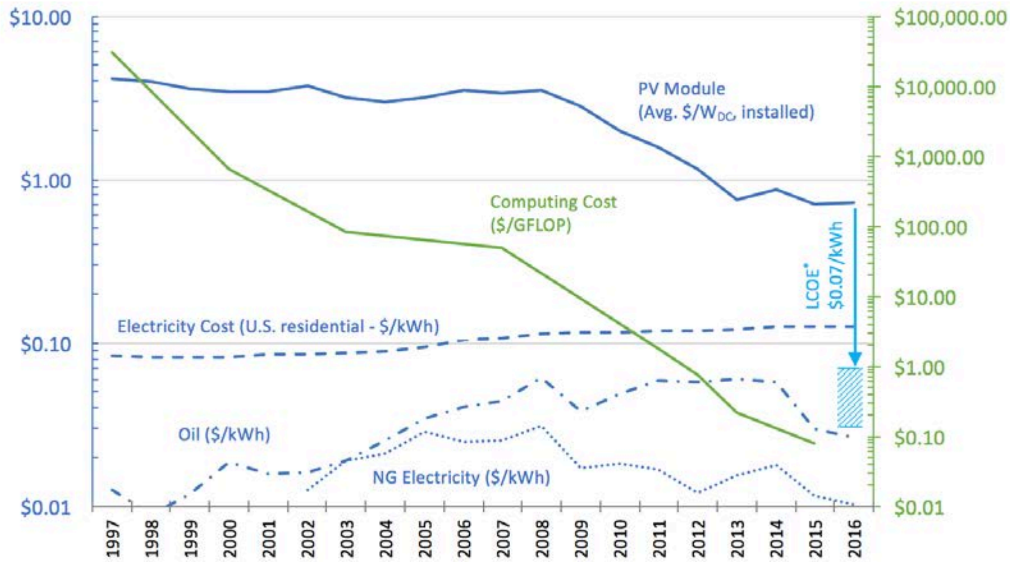

Figure 1. PV module, Natural Gas and Oil Price Trends (blue) compared to Computing Cost Trend (green), 1997-2016 – source from eia.gov
**Levelized Cost of Electricity estimated for 100 MW_{DC}, fixed-tilt systems with 0.5%/yr and financial parameters from [8].*

economies of scale are plateauing, the cost of photovoltaics is a moving target and efficiency from single junction technologies is approaching its technological limits, hampering the ability to use efficiency increases as the lever to overcome the previous barriers.
In this context, just like Moore's Law, there is an underlying law based on fundamental physics that can help make a specific, quantitative prediction, about innovation as a function of time. For semiconductors, the technical parameter has historically been *transistor density*; for photovoltaics the analog is *energy produced per unit volume*.
Figure 1 puts a lot of this discussion in perspective. First, it shows how computing and photovoltaics have seen significant and relentless cost reductions the last 20 years by "packing more in a smaller volume", while oil and natural gas has remained relatively constant despite shorter term price fluctuations. It also depicts how competitive today's solar energy prices are compared to other forms of electricity.

This achieved reality of silicon module prices below US$1/W and projections of US$0.5/W has fundamentally changed solar R&D.
Slim margins have pushed companies into bankruptcy and has yet to make several manufacturers profitable on an annual basis. Cost, intermittency and dispatchability have been major challenges in the pursuit of utility scale solar generation. More recently, degradation and long-term performance R&D has become crucial for the bankability of projects. However, the standard business model of the solar industry with each company eager to outcompete the

next in price has made the industry very risk adverse when it comes to implementation of innovation.

So, where do we go from here? As we move towards an "electric-powered world" and everything around us starts demanding electricity in a clean and efficient way (e.g. electric vehicles, portable electronics, rural electrification), new challenges arise. The first new challenge centers around portability: the use lightweight and flexible modules necessary for implementation into our everyday life. The second is achieving high power in small areas and the use of sustainable materials for device manufacturing. Similar to many consumer applications, solar margins will improve and engineering hurdles centered around aesthetics, customization and functionality will be part of our everyday R&D life. An analogy is the introduction of Ford's Model T car. Photovoltaics is at a stage where it has demonstrated its affordability, impact and potential - from now on we will see a whole new technology taking off.

Following the Moore's Law argument, the path for improved PV is to make cells thinner and more efficient. As the industry matures, costs are increasingly dominated by materials' cost and expensive process changes translate into very small incremental benefits, we come to appreciate the fundamental nature of the scientific breakthroughs necessary to propel this energy source to next generation levels. Higher power cells can be achieved by stacking cells with different bandgaps thus efficiently capturing a wider portion of the solar spectrum. The efficiency limits rise from 33% for a single junction cell to 43% for two junctions under no concentration, 49% for three junctions, and 66% for infinitely many junctions. This approach is not novel, multijunction cells are well known in space applications, where very high quality single crystals are epitaxially grown and cells are engineer to withstand radiation and high levels of illumination [9]. The analogy to the car industry, is having a limited-edition Ferrari in your garage. Although these modules are very expensive, epitaxial lift-off techniques enabling substrate reuse have been demonstrated showing a path to lower costs. The future of solar lies in merging the ubiquitousness of the "model T" solar cells with the performance of the "limited-edition Ferrari" cells.

The first thing we need to realize is that we have to rely on the mass-production low-cost manufacturing lines of the "model T", which most likely means a silicon cell will be our bottom cell and we cannot count on high quality single crystal films for our multilayer stack. Instead we will have to rely on faster deposition methods like evaporation or sputtering and defect engineered top films to achieve the electrical and optical properties desired [10]. The latter point is crucial to the success of next generation solar absorbers, engineered defect tolerant materials is the pathway to enable ultra-low-cost manufacturing technology for high efficiency devices. It has been shown that a top-cell bandgap of 1.7 eV and an efficiency comparable to standard silicon cells today (20%) can enable 32% efficient tandems [11].

The task seems daunting, especially when one considers that the performance of a full device is usually governed by the concentration and distribution of nanoscale inhomogeneities and defects throughout the entire solar cell.

How do we accelerate discovery and defect engineering to facilitate a high-power, portable and economic solar industry? We have to redefine the paradigms for materials discovery, especially for systems with complex functionalities, and move beyond serendipitous discoveries, Edisonian approaches and the classical synthesis-characterization-theory methodologies. The answer lies in highly correlative imaging methods under operating conditions combined with big data analytics.

Understanding the fundamental relationships between composition-structure-properties on a nano-pixel basis, under real operating conditions and *in situ* (controlled temperature and ambient) is fundamental to unraveling the causality and effect of certain defects and their direct impact on performance. Imagining techniques these days do not provide just a mere illustration of the system under study, they actually contain compositional, structural and functional information. As one could imagine the correlation of multiple 2D or 3D mapping

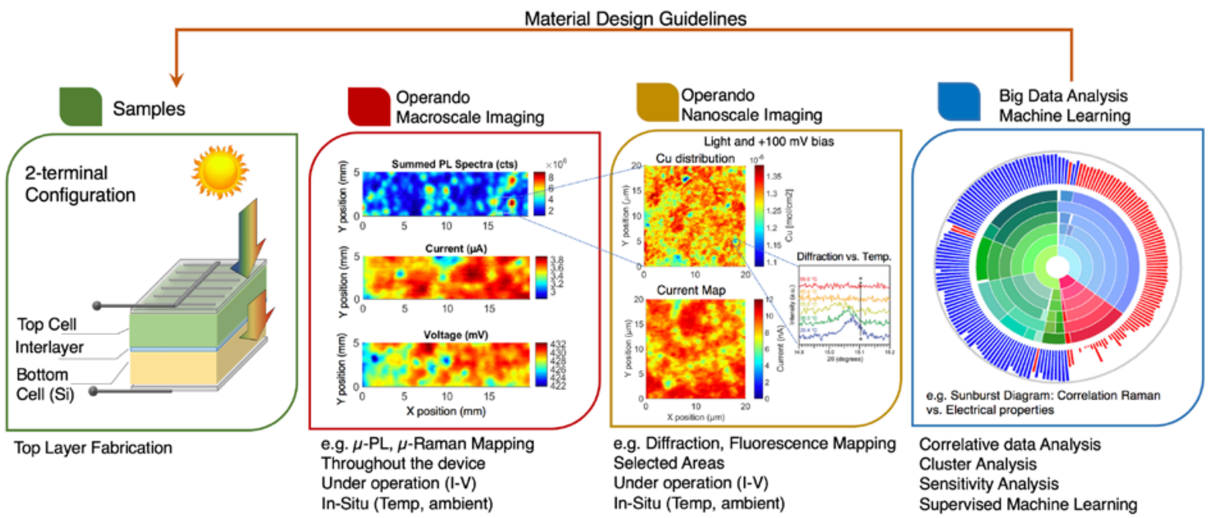

*Figure 2. Schematics of a machine learning driven approach to material's design for discovery and engineering of next generation solar cell materials. Data from multiple imaging modalities at different scales can be analyzed to develop guidelines for growth and/or processing of a material. *Diffraction pattern and sunburst diagram from {Schelhas:2016hc} {Kusne:2015ct} respectively.*

modalities on a pixel-to-pixel basis and the multiple dimensions of these maps given by time, temperature and ambient conditions creates a 'big data' challenge in itself.

In situ and operando measurement techniques combined with nanoscale resolution have proven invaluable in multiple fields of study. I would like to argue that correlative hard X-ray microscopy (HXM) with <100 nm resolution is a unique capability that can radically change the approach for optimizing solar absorbers, interfaces and full devices in solar cell research. Unlike other fields of microscopy, HXM have excellent penetration through layers and entire devices, enabling 3D imaging of buried structures. They can easily go through gases and fluids, enabling studies at pressure, and under process conditions. They allow quantitative studies of sample composition with trace-element sensitivity in structured materials and devices. Chemical-state information of individual atomic species can be obtained using X-ray spectroscopic techniques. X-rays do not interact with external fields, and thereby enable studies in electric or magnetic

fields [12]. As acquisition speeds and resolution increase giving us more density of data points and the functionality of the measurements add more dimensions to be analyzed, the handling, management and analysis of data sets become more and more complicated. Operando measurements as well as in-situ studies pose a new challenge. Finding correlations in the 3+ dimensional data sets that result from many of these measurements is not straight forward, and the possibility of missing connections, relationships and trends is increasingly concerning.
Machine learning techniques including principal component and cluster analyses have been widely used in fields notoriously plagued with tremendous amounts of data [13]. A key benefit of these approaches is the ability to identify trends in highly dimensional data that would be inherently difficult to accomplish by hand.

The first step towards full information recovery from high resolution multifunctional imaging data is the adoption of big-data analytics [14-17]. This means implementation of dimensionality reduction, clustering techniques and statistical unsupervised learning [13]. Unsupervised image analysis tools targeted to high-performance computing platforms have shown the ability to analyze high resolution scanning and electron microscopy data in 2D in real time [18]. The advances in high resolution experimental imaging and high-performance computing [19] will undoubtedly propel materials discovery and ultimately “materials by design” [20].
However, ‘correlation does not imply causation’ meaning that just because we observe the statistical correlation between two observations it does not imply that we understand the underlying physics contained in the multimodal imaging. Transitioning from ‘big data’ to ‘deep data’ is the next step. All the structure-property relationships at the nanoscale retrieved from ‘big data’ can now be visited with real physical models allowing for verification and improvements in predictive modeling [14]. This step allows to close the loop and propose design guidelines to grow or process the material with the desired properties and functionalities.

Materials informatics is ready to lay out the ground for a new paradigm in materials discovery, especially for complex functional systems like solar cells. And it could very well end up being data what ultimately pushes down the cost of solar to fossil fuel levels.

References:

1. Jacobson, M. Z. *et al.* 100% clean and renewable wind, water, and sunlight (WWS) all-sector energy roadmaps for the 50 United States. *Energy Environ. Sci.* **8,** 2093–2117 (2015).
2. Perea, A. *US Solar Market Insight 2017 Q2*. **Q2,** (SEIA, GTM Research).
3. Le, M. $1/W Photovoltaic Systems. 1–28 (2010).
4. Munsell, M. GTM Research: 20 US States at Grid Parity for Residential Solar. (2016). Available at: https://www.greentechmedia.com/articles/read/gtm-research-20-us-states-at-grid-parity-for-residential-solar. (Accessed: 31st August 2017)
5. Kurtz, S. *et al.* Solar research not finished. *Nature Publishing Group* **10,** 141–142 (2016).
6. Powell, D. M. *et al.* The capital intensity of photovoltaics manufacturing: barrier to scale and opportunity for innovation. *Energy Environ. Sci.* **8,** 3395–3408 (2015).
7. Haegel, N. M. *et al.* Terawatt-scale photovoltaics: Trajectories and challenges. *Science*

356, 141–143 (2017).
8. Jones-Albertus, R., Feldman, D., Fu, R., Horowitz, K. & Woodhouse, M. Technology advances needed for photovoltaics to achieve widespread grid price parity. *Prog. Photovolt: Res. Appl.* **24,** 1272–1283 (2016).
9. Takamoto, T., Kaneiwa, M., Imaizumi, M. & Yamaguchi, M. InGaP/GaAs-based multijunction solar cells. *Prog. Photovolt: Res. Appl.* **13,** 495–511 (2005).
10. Bobela, D. C., Gedvilas, L., Woodhouse, M., Horowitz, K. A. W. & Basore, P. A. Economic competitiveness of III-V on silicon tandem one-sun photovoltaic solar modules in favorable future scenarios. *Prog. Photovolt: Res. Appl.* 1–8 (2016). doi:10.1002/pip.2808
11. Yu, Z. J., Leilaeioun, M. & Holman, Z. Selecting tandem partners for silicon solar cells. *Nat. Energy* **1,** 16137–4 (2016).
12. Delongchamp, D. *et al.* Engineering solar cells based on correlative x-ray microscopy. *Journal of Materials Research* **32,** 1825–1854 (2017).
13. Hastie, T., Tibshirani, R. & Friedman, J. *The Elements of Statistical Learning*. (Springer Science & Business Media, 2013).
14. Kalinin, S. V., Sumpter, B. G. & Archibald, R. K. Big–deep–smart data in imaging for guiding materials design. *Nature Publishing Group* **14,** 973–980 (2015).
15. Runkler, T. A. *Data Analytics*. (Springer Fachmedien Wiesbaden, 2016). doi:10.1007/978-3-658-14075-5
16. Rajan, K. Materials informatics. *Materials Today* **15,** 470 (2012).
17. Rajan, K. Materials Informatics: The Materials 'Gene' and Big Data. *Annu. Rev. Mater. Res.* **45,** 153–169 (2015).
18. Belianinov, A. *et al.* Big data and deep data in scanning and electron microscopies: deriving functionality from multidimensional data sets. *Advanced Structural and Chemical Imaging* **1,** 6 (2015).
19. Dongarra, J. et al. e International Exascale So ware Project roadmap. Int. J. High Perform. Comput. Appl. **25**, 3–60 (2011).
20. Materials Genome Initiative; http://go.nature.com/Rkw2mj