Current Opinion in Solid State and Materials Science xxx (2016) xxx–xxx

Contents lists available at ScienceDirect

Current Opinion in Solid State and Materials Science

journal homepage: www.elsevier.com/locate/cossms

Silicon Valley meets the ivory tower: Searchable data repositories for experimental nanomaterials research

Nils Persson, Michael McBride, Martha Grover, Elsa Reichmanis *

Department of Chemical and Biomolecular Engineering, Georgia Institute of Technology, 311 Ferst Dr, Atlanta, GA 30332, United States

ARTICLE INFO

Article history:
Received 15 January 2016
Revised 26 May 2016
Accepted 4 June 2016
Available online xxxx

Keywords:
P3HT
Poly(3-hexylthiophene)
OFET
Field effect transistor
Organic electronics
Solution processing
Solution processed
Data repository
Database
Process-property relationship
Channel length
Molecular weight
Spin coating
Spin-coating
Chloroform
Device fabrication
Mobility
Field effect mobility
Hole mobility

ABSTRACT

Centralized data repositories have proven useful in promoting reproducibility and transparency in many fields of research, particularly the life sciences. Nanomaterials research is frequently difficult to reproduce because of the complexity of the systems under study and a lack of standardization due to the exploratory nature of the field. One system that exemplifies this problem is that of the poly(3-hexylthiophene)-based organic field effect transistor, a platform used to study the process-structure-property relationships in semicrystalline polymeric semiconductors. Performance in this system is judged principally by charge carrier mobility, a fitted model parameter that describes how quickly charge can move through the active layer of a given transistor. Reported values of this electrical property vary by over six orders of magnitude for one material and are impacted by intrinsic material properties, processing conditions, device architectures and calculation methods. A database containing over 200 of these devices from 19 studies was compiled to demonstrate the impact of these parameters on reported performance as well as the utility of having such a database to search and explore. We confirm well-known trends such as the effect of polymer molecular weight, transistor channel length and thin film deposition method before going on to identify a standard device that can be compared across multiple studies. We find that an organized database of process-property information can be used to identify unpublished heuristic knowledge, to help authors standardize their reporting of methods, and to guide hypothesis generation and experimental design. A repository of the data used in this study is available at [http://www.github.com/Imperssonator/OFET-Database].

1. Introduction

Academia is in the midst of a so-called reproducibility crisis [1,2]. While this has been publicized largely in the life sciences, it is nonetheless a concern for research in nanomaterials [3]. Experimental systems have grown in complexity, data has grown in volume and velocity, and analyses are now performed by multiple layers of computing systems, resulting in published figures that are very far removed from the raw data used to generate them. Single words in a manuscript can now represent thousands of lines of computer code or a highly sensitive experimental method developed over multiple studies. Therein lies the heart of the reproducibility crisis: in the amount of information assumed to have been conveyed in the words of a text document.

Journals and research consortia have combatted this problem by setting up centralized data repositories for research methods, data, and analytical codes [4]. For example, in the field of Proteomics, the PRIDE Archive provides an explorable repository of mass spectrometry data from proteomics research, promoting transparency through the entire data analysis process [5]. Organic synthesis has benefitted from searchable databases like Scifinder® or Reaxys®. Material property simulations have benefitted more recently from the establishment of the Materials Genome Initiative and online collaborative organizations like nanoHUB [6,7]. It is recognized that the volume and velocity of materials research data will soon reach a threshold across which large data repositories and tools from data science will be required for management and analysis [8]. The same will undoubtedly be true for nanomaterials research.

Exploratory nanomaterials research data presents a greater challenge due to the variability of types of information presented

---

* Corresponding author.
E-mail address: elsa.reichmanis@chbe.gatech.edu (E. Reichmanis).

http://dx.doi.org/10.1016/j.cossms.2016.06.002
1359-0286/

Please cite this article in press as: N. Persson et al., Silicon Valley meets the ivory tower: Searchable data repositories for experimental nanomaterials research, Curr. Opin. Solid State Mater. Sci. (2016), http://dx.doi.org/10.1016/j.cossms.2016.06.002

and the speed with which reporting practices change. This is not so different from an auction website like eBay, where sellers attempt to provide as detailed a description of their item as possible, but the relevant descriptive features of a given item are determined more or less by the marketplace, which is constantly changing. eBay can give sellers suggestions on what information to provide based on the listings of similar items. For example, if one were to sell a wristwatch, eBay suggests that the seller specify the band material and the movement mechanism based on the listings of similar items. In a similar vein, the Nature Publishing Group recently introduced an initiative to provide authors suggestions of information to include in methods sections [3]. Information like instrument settings and environmental conditions, referred to as metadata, can have a huge influence on measured material performance and is a vital part of any data repository. We aim to demonstrate the value of cataloging experimental metadata using the system of the solution-processed polymeric transistor as a case study.

Solution processable polymeric semiconductors are a promising class of electronic material for a variety of reasons: they can be coated on large areas, are mechanically flexible, and could enable high throughput additive manufacturing techniques [9–13]. Commercial applications include solar energy, display, and sensing technologies [14–18]. For almost two decades, the organic field effect transistor (OFET) has been used as a platform to study the process-structure-property relationships in polymeric semiconductors, because it is relevant to various applications, a facile platform for thin film deposition, and enables the calculation of charge carrier mobility (mobility), an important figure of merit for the performance of a given device [19]. Mobility is essentially a model parameter defined as the velocity of a charge per applied electric field in a material sample; it has units of $cm^2\,V^{-1}\,s^{-1}$ and is very sensitive to starting material, processing, device layout, and calculation method. Poly(3-hexylthiophene) (P3HT) is frequently used as a model material in studies of the effects of processing conditions on OFET mobility. We compile data from over 200 P3HT-based OFET devices published across the literature to envision what an experimental research database would look like for a nanomaterial system and how it could be used to generate physical insight as well as improve the reproducibility of experimental studies.

2. Methods

Publications were selected for the database based on the following criteria: (1) the study must involve OFET devices whose active layer were comprised of neat P3HT, (2) mobility must be reported at room temperature, and (3) a majority of the relevant process and device parameters must be reported [20–38]. The parameters that were recorded in the database are enumerated in Fig. 1. They comprise everything from the molecular parameters of the starting material, to the machine settings of the various deposition techniques, to the analytical methods used to calculate mobility from the current-voltage measurements on the device.

Very few papers contained the complete set of desired processing information; the dataset is not sparse by any means, but missing information is a significant problem. Choi et al. discuss this further in an article on the best practices for reporting on OFET device fabrication [39]. The dataset is also wildly diverse, comprised of text-based designators (e.g., *bottom-gate, bottom-contact architecture* or *spin-coated*...), chemical compounds (*chloroform, chlorobenzene, acetone*...), numeric values (5 mg/mL, 40 kD...), numeric ranges (20–50 nm), and even categories with mixed types (regioregularity is listed variously as a percentage or simply "*highly regioregular*").

Data entry was performed in Excel, with each study receiving its own self-contained spreadsheet to handle the process information. Mobility, typically presented in a graph, was extracted from figures using DataThief [www.datathief.org], a semi-automatic tool that can reliably extract quantitative data from scatter plots up to 3 significant figures. MATLAB scripts were written to compile the database into a structure array, filter and search the devices, and plot the results. For example, a one-line function call can find all devices made with chloroform as a solvent on a bottom-gate, bottom-contact architecture with gold electrodes, and plot the mobility of the filtered devices against their molecular weight. The entire database with accompanying graphical user interface is available for download from a GitHub repository at [http://www.github.com/Imperssonator/OFET-Database]. We now present select examples of the trends and insight that can be extracted from this database.

3. Results and discussion

3.1. The effect of molecular weight distribution

The first widely reported process-property trend in the literature is that of increasing mobility with molecular weight, plotted in Fig. 2**A**. Kline et al., Zen et al., and Verilhac et al. conducted the seminal studies on this parameter, concluding variously that low molecular weights ($<20{,}000\ g\,mol^{-1}$) introduce inherent crystalline disorder that limits their overall performance, while higher molecular weights may introduce chains that are long enough to extend across grain boundaries, providing high-mobility pathways for charge transport [29–31,37,40].

Polydispersity index (PDI) is more difficult to study as a process parameter, yet our dataset contains 146 devices for which PDI has been reported. It displays approximately the same trend as molecular weight, as shown in Fig. 2B, with devices with the lowest PDI showing a wider spread that encompasses very low performance devices ($<10^{-4}\ cm^2\,V^{-1}\,s^{-1}$). It is possible that this is due to a correlation between PDI and molecular weight, however this correlation is quite weak ($R^2 = 0.4$ from a linear fit), so there is some merit to the notion that higher Polydispersity index can benefit device performance. This could be explained on the same morphological grounds as the trend in molecular weight: highly polydisperse samples would contain many chains long enough to connect multiple crystalline grains. Given the difficulty of controlling molecular weight distribution accurately, it is all the more important that it be presented in as quantitative a manner as possible, because it is highly unlikely that authors from different groups will attain the same distribution.

3.2. The effect of channel length

Channel length is a widely investigated device fabrication parameter due to its direct relationship with transistor packing density and power efficiency in electronic applications. Wang et al. and Chang et al. investigated the impact of channel length on mobility and discovered that shorter channels generally have higher mobility [24,41]. Chang associated this trend with a decrease in the number of grain boundaries as the length scale of the channel approaches that of the polymeric crystalline domains. However, contributions from increased contact resistance and deviation from the gradual channel approximation were ignored. Chabinyc et al. demonstrated that field effect transistors below 10 μm in length experience high contact resistance and deviate from the ideal current-voltage relationships [42].

Our database indicates that both of the above referenced analyses are correct. The box plot in Fig. 3 shows mobility versus

Please cite this article in press as: N. Persson et al., Silicon Valley meets the ivory tower: Searchable data repositories for experimental nanomaterials research, Curr. Opin. Solid State Mater. Sci. (2016), http://dx.doi.org/10.1016/j.cossms.2016.06.002