High-Throughput Determination of Band Gap Energy in Combinatorial Thin-Film Libraries

Author: Bin Xiao

Abstract

In this work, we propose a method that allows to make a fully automated band gap energy analysis for thin-film libraries. This method is simple, fast, robust to preprocessing errors and unifies the problems of curvature evaluation, curve decomposition and Tauc segmentation extraction. The main idea is to decompose the given curve by breakpoints and select the fragment which contains the most linear feature that could represent Tauc segment. The breakpoints are divided into cusps and inflection points by testing their significant changes in curvature using adaptive smoothing techniques. Applying this method to experimental data indicated that it could handle all kinds of thin-film library data without prior knowledge.

Introduction

Combinatorial synthesis and high-throughput characterization require the fast acquisition and high-quality analysis of massive individual data in order to map compositions with corresponding structural and functional properties.[1] The produced large number of data sets which beyond manually processing capability make automated data-analysis and visualization a pressing concern. In the field of semiconductor science, the optical band gap energy (E_g) is a basic parameter that plays an important role in predicting material's optoelectronic applicability and performance.[2-4] This value indicates the energy required to excite an electron from the valence band to the conduction band. In 1966, Tauc proposed a method for estimating band gap energy of amorphous semiconductors based on absorption spectroscopic data.[5] Later, this method was further developed by Davis and Mott's general work.[6-7] Currently, due to the widely available and easy-to-use ultra-violet spectrometer, the Tauc method becomes the most commonly used for band gap measurement.

Figure 1 shows a typical analytic procedure. Firstly, the absorbance data from a thin-film sample is collected, which covers a range of energies from below the band gap transition to above it. Then based on Beer-Lambert's law ($\alpha \propto \mathrm{Ln}(I/I_0)$), the absorption coefficient α is calculated. Finally, the band gap energy could be derived from the equation $(\alpha \cdot h\nu)^{1/n} = B(h\nu - E_g)$, where $h\nu$ is the photon energy and B is a constant. The n factor depends on the nature of the electron transition. For direct allowed band gap energy, $n = ½$, and indirect allowed band gap, $n = 2$. In either case, the linear region of the plot is extrapolated, and the interception with the x-axis corresponds to the estimated band gap value.

The effect of above mentioned linear-seeking scheme will directly influence the final experiment result, therefore, how to improve the accuracy and how to achieve high efficiency become an inevitable requirement for the development of spectroscopy analysis.[8-9] Generally, the Tauc method is performed manually, that is, by visual checking of the plot, to decide on a region of linearity, and to draw what seems to be a good linear approximation to the curve over the selected region.[10-11] The result quality is dependent on the accuracy of the linear range selection, as a slight variation can lead to an obvious deviation in the final energy-gap value. In order to accelerate the development of novel and improved optically meaningful materials, automated and accurate extraction of Tauc line segment that are applicable of overcoming the

susceptibility of the manual's bias, error and fatigue is essential. However, only a few computer-assisted automated solutions have been developed up to now.[12-14] Among them, a method based on constrained piecewise linear approximation has been successfully applied for materials with photoelectrochemical water splitting properties.[13] This method enables rapid evaluation of various optical properties, but several predetermined parameters need to be adjusted in order to access these advantages. Two new automatic algorithms, namely Segmentation and MARS, were reported.[15] The advantage of MARS algorithm is that it capable of dealing with large-scale data, allowing fast determination of direct, indirect and alpha band gaps without human intervention. In addition, a five-step algorithm which mimics the judgment of human expert was discussed and applied to obtain band gap energies for a series of pure and mixed powder compounds.[12]

In this work, we introduce a simple approach based on curvature profile to locate and extract the Tauc straight-line segment form a given curve. This method results in accurate band gap values on several thin-film libraries while keeping a fast processing because of the provided types of breakpoints. Besides, the computed bandgap results are immediately visualized so that trends and correlations which might be interesting for further study can be easily exposed.

Band gap energy determination

The objective of band gap determination is to find the optimal straight line which stands for Tauc segment. The first step of the proposed procedure is to get a set of breakpoints for our curve decomposition. Curve decomposition has been an active research filed of image processing, because a segmented curve can represent the image profile in a meaningful and compact form to simplify more advanced visual processing, such as shape analysis and pattern recognition.[16-17] Many algorithms have been developed for this purpose, most of which use the feature points as breakpoints to partition a given curve. For example, *Asada* and *Brady*[18] reported that the curve breakpoints could be divided into two types: the corner and the smooth join, both indicate isolated curvature changed. The former is associated with a discontinuous tangent, whereas the later corresponds to continuous tangent but discontinuous curvature. Although this method is widely used, it is computationally intensive because the breakpoints need to be detected by multiscale convolutions of the curve with the help of high-order Gaussian derivative filters.

In our method, the initial breakpoints between decomposed fragments are defined as either cusps or inflection points. Cusps and inflection points, which are mathematically simple, are the most significant features for straight-line segment identification.[19-20] A cusp stands for a point between two consecutive convex or concave, whereas an inflection point is a point where the curve changes from being convex to concave or vice versa. Their detection is related to local curvature characteristics such as extremes or stationary point, and several examples are shown in **Figure 2**. For example, the local minimum curvature happens when the inflection point preceded by a convex and followed by a concave means, while local maximum occurs when an inflection point experiences the reverse way.

The local changes of a single data point are indicated by the curvature at that point. After the curvature estimation, the tangent angle and the associated geometric properties such as convexity and concavity at the point-of-interest could be obtained. In principle, any small fragment on the curve could be fitted with a continuous function. Here, a cubic polynomial function $\bar{y} = a + b\bar{x} + c\bar{x}^2 + d\bar{x}^3$ is employed to approximate the given data points and obtain local curvature information. The given curve X is composed of data points (x_i, y_i), where x_i and y_i are two functions of the index variable i. X could be denoted by

$\{P_i = (x_i, y_i) \mid i = 0, \ldots, n\}$. The slope of the tangent line at point P_i on the curve X is $\mathrm{d}y_i/\mathrm{d}x_i$, and the corresponding tangent angle θ_i is defined as $\tan^{-1}(\mathrm{d}y_i/\mathrm{d}x_i)$. The curvature of a continuous point is defined as the change of the slope at that point. To estimate the curvature for point P_i, we need to determinate the four unknown coefficients a, b, c and d from the cubic polynomial. Let us consider a small fragment $F(P_j) = \{P_j \mid i - k \leq j \leq i + k\}$ from curve X. $F(P_j)$ is composed of a center point $P_i(x_i, y_i)$ with index length of k between $P_{i-k}(x_{i-k}, y_{i+k})$ and P_{i+k} (x_{i+k}, y_{i+k}). It is desired to approximate $F(P_j)$ such that the Euclidean distances between the point $P_i(x_i, y_i)$ and the corresponding cubic polynomial point $\bar{P}_i(\bar{x}_i, \bar{y}_i)$ is minimized within fragment $F(P_j)$. The the resulting equation has $2(2k+1) + 4$ unknowns and are not solvable directly. Hence, least square fitting approach is employed to estimate the a, b, c and d coefficients over the small fragment $F(P_j)$. The iterative optimization involves complex computation and is extremely time consuming. To improve the performance of fitting process, we could calculate y and x directions separately so that the accumulated vertical or horizontal distances of point $P_i(x_i, y_i)$ are closest to the cubic polynomial point $\bar{P}_i(\bar{x}_i, \bar{y}_i)$. This simplifies the parameter computation and eliminates round-off error. Besides, the error norm is not biased from the true integral squared distance. Under this simplification, the fitting equation is divided as

$$\begin{cases} \sum_{j=i-k}^{j=i+k} [y_j - (a_1 + b_1\bar{x} + c_1\bar{x}^2 + d_1\bar{x}^3)]^2 = \mathrm{Min}\{F_1(a_1, b_1, c_1, d_1)\} \\ \sum_{j=i-k}^{j=i+k} [y_j - (a_2 + b_2\bar{x} + c_2\bar{x}^2 + d_2\bar{x}^3)]^2 = \mathrm{Min}\{F_2(a_2, b_2, c_2, d_2)\} \end{cases} \quad (2)$$

Using Lagrange Multiplier method[21] which involves first and second derivative calculations, the coefficients of equations (2) could be calculated, and the curvature at point P_i is expressed as

$$\kappa_i = 2\frac{(b_1c_2 - b_2c_1)}{({b_1}^2 + {b_2}^2)^{\frac{3}{2}}} \quad (3)$$

After curvature computation, values κ_i are known for each data point P_i. The next step is the selection of the most important points for breakpoints determination. As discussed above, an inflection point is a point where the second derivation is equal to zero, and its two neighbors have opposite signs (positive and negative for convex and concave sides, respectively). A cusp is a local extreme, but the associated tangent is discontinuous. However, the curvature calculation process is sensitive to perturbations of the original data points, and it is not straightforward to obtain the curvature of a discontinuous point because small quantization variations around discrete fragment could lead to large errors. Moreover, our experimental data are comprised of a sequence of points that have many discontinuities near to each other or far apart. Due to discretization or sources of noise, x_i and y_i functions are not always smooth. In this case, smoothing is needed when extracting the significant local features from real-world measurements filled of noisy data. The adaptive smoothing filter method proposed by *Saint-Marc et al*[22] is employed. This method is a smoothing process which allows to preserve local discontinuous information while dealing with smoothing. After t iteration, the smoothed X^t curve is expressed as:

$$X^t = \frac{1}{M}\sum_{-N}^{N} X^{t-1}(P_i + w) \cdot m^{t-1} \quad (4)$$

here $M = \sum_{-N}^{N} X^{t-1}(P_i + w)$, w is convolution weight, m^t is the decreasing function indicating degree of discontinuity at point P_i. m^t could be formulated as $e^{(X'^t)^2/2h^2}$, where X'^t is the first derivation of X^t and h is the parameter which shows the discontinuity preservation status after the iterative convolution. The discontinuities could be determined by applying adaptive smoothing to tangent angle θ_i along the curve:

$$\theta^t = \frac{1}{M}\sum_{-N}^{N} X^{t-1}(P_i + k) \cdot m^{t-1} \quad (5)$$

After a number of iterations, the sharp discontinuities would eventually become the cusps of segmentation. In this step, the breakpoints categorized as by cusps and inflection points are successfully obtained. Segmentation of Tauc line can be achieved based on the identification of types of breakpoints, because line segments could only exist between two sharp cusps. A dynamic technique is applied to merge the adjacent fragment pieces with similar curvature which belongs to a larger linear segment. We first exclude all points with zero curvature because such points locate in the interior of line segments on the curve X, and are not needed for merging. Then a new point $N_i = ((x_i + x_{i+1})/2, (y_i + y_{i+1})/2)$ is interpolated from two consecutive cusps $C_i = (x_i, y_i)$ and $C_{i+1} = (x_{i+1}, y_{i+1})$. Here the new point is the middle point of the region, other interpolation method could also be used. Calculate the curvature κ_{N_i} of N_i and compare with $\kappa_{C_i}$and $\kappa_{C_{i+1}}$of C_i and C_{i+1}, respectively. If $\kappa_{C_i} \leq \kappa_{N_i} \leq \kappa_{C_{i+1}}$, the two consecutive cusps are successfully merged to a new linear segment. A local straight-line fitting algorithm is implemented during merging, and it links adjacent points along process direction and records the coordinates of each merged cusp. Then, the next adjacent cusp is added in merging procedure, and the merge-and-fitting process is iterated on the remaining cusps until the linear fitting process fails. After this, a new candidate cusp is chosen with its beginning at the end of the last failed cusp and the above iteration begins again. The segments with the most significant linear feature are selected for Tauc segment.

Figures and Tables

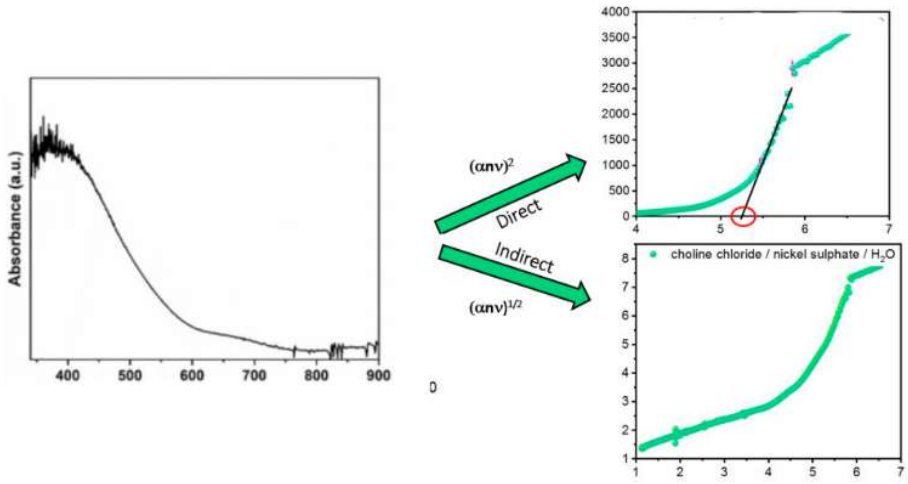

Figure 1. Tauc method allows to determinate the band gap energy (Eg) from absorption spectrum. UV-VIS spectrum and Tauc plots for band gap determination. (This figure will be re-drawn)

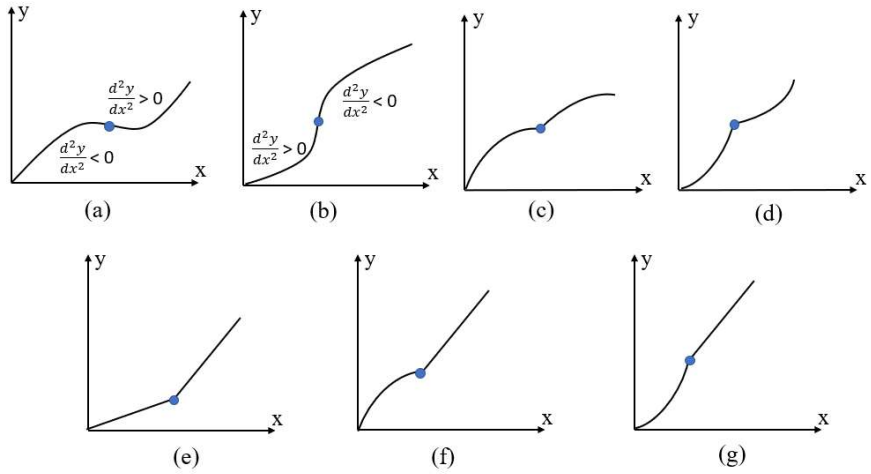

Figure 2. Possible types of cusps and inflection points for curve decomposition. A cusp stands for a point between two consecutive convex or concave, whereas an inflection point is a point where the curve changes from being convex to concave or vice versa.

References

1. Ludwig, A., Discovery of new materials using combinatorial synthesis and high-throughput characterization of thin-film materials libraries combined with computational methods. *npj Computational Materials* **2019,** *5* (1), 70-70.
2. Saravanan, R.; Sacari, E.; Gracia, F.; Khan, M. M.; Mosquera, E.; Gupta, V. K., Conducting PANI stimulated ZnO system for visible light photocatalytic degradation of coloured dyes. *J. Mol. Liq.* **2016,** *221*, 1029-1033.
3. Wang, X.; Rogalla, D.; Ludwig, A., Influences of W Content on the Phase Transformation Properties and the Associated Stress Change in Thin Film Substrate Combinations Studied by Fabrication and Characterization of Thin Film V1-xWxO2 Materials Libraries. *Acs Combinatorial Science* **2018,** *20* (4), 229–236-229–236.
4. Kumari, S.; Gutkowski, R.; Junqueira, J. R. C.; Kostka, A.; Hengge, K.; Scheu, C.; Schuhmann, W.; Ludwig, A., Combinatorial Synthesis and High-Throughput Characterization of Fe-V-O Thin-Film Materials Libraries for Solar Water Splitting. *Acs Combinatorial Science* **2018,** *20* (9), 544–553-544–553.
5. Tauc, J.; Grigorovici, R.; Vancu, A., Optical Properties and Electronic Structure of Amorphous Germanium. *physica status solidi (b)* **1966,** *15* (2), 627-637.
6. Mott, N. F.; Davis, E. A., *Electronic processes in non-crystalline materials*. Oxford university press: 2012.
7. Davis, E. A.; Mott, N. F., Conduction in non-crystalline systems V. Conductivity, optical absorption and photoconductivity in amorphous semiconductors. *The Philosophical Magazine: A Journal of Theoretical Experimental and Applied Physics* **1970,** *22* (179), 0903-0922.
8. Mok, T. M.; O'Leary, S. K., The dependence of the Tauc and Cody optical gaps associated with hydrogenated amorphous silicon on the film thickness: α l Experimental limitations and the impact of curvature in the Tauc and Cody plots. *J. Appl. Phys.* **2007,** *102* (11), 113525.
9. Coulter, J. B.; Birnie III, D. P., Assessing Tauc plot slope quantification: ZnO thin films as a model system. *physica status solidi (b)* **2018,** *255* (3), 1700393.
10. Murphy, A., Band-gap determination from diffuse reflectance measurements of semiconductor films, and application to photoelectrochemical water-splitting. *Sol. Energy Mater. Sol. Cells* **2007,** *91* (14), 1326-1337.
11. Tumuluri, A.; Naidu, K. L.; Raju, K. J., Band gap determination using Tauc's plot for LiNbO3 thin films. *Int. J. Chem. Tech. Res* **2014,** *6* (6), 3353-3356.
12. Escobedo-Morales, A.; Ruiz-López, I. I.; Ruiz-Peralta, M. d.; Tepech-Carrillo, L.; Sánchez-Cantú, M.; Moreno-Orea, J. E., Automated method for the determination of the band gap energy of pure and mixed powder samples using diffuse reflectance spectroscopy. *Heliyon* **2019,** *5* (4), e01505-e01505.
13. Suram, S. K.; Newhouse, P. F.; Gregoire, J. M., High Throughput Light Absorber Discovery, Part 1: An Algorithm for Automated Tauc Analysis. *ACS Comb. Sci.* **2016,** *18* (11), 673–681-673–681.
14. Zhao, Z.; Jin, Y.; Shi, P.; Xue, Y.; Zhao, B.; Zhang, Y.; Huang, F.; Bi, P.; Wang, Q., An Improved High-Throughput Data Processing Based on Combinatorial Materials Chip Approach for Rapid Construction of Fe-Cr-Ni Composition-Phase Map. *ACS Comb. Sci.* **2019,** *21* (12), 833–842-833–842.
15. Schwarting, M.; Siol, S.; Talley, K.; Zakutayev, A.; Phillips, C., Automated algorithms for band gap analysis from optical absorption spectra. *Materials Discovery* **2017,** *10*, 43–52-43–52.
16. Tsai, A.; Yezzi, A.; Willsky, A. S., Curve evolution implementation of the Mumford-Shah functional for image segmentation, denoising, interpolation, and magnification. *IEEE transactions on Image Processing* **2001,** *10* (8), 1169-1186.
17. Arrebola, F.; Sandoval, F., Corner detection and curve segmentation by multiresolution chain-code linking. *Pattern Recognition* **2005,** *38* (10), 1596-1614.

18. Asada, H.; Brady, M., The Curvature Primal Sketch. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **1986,** *PAMI-8* (1), 2-14.
19. Lowe, D. G., Organization of smooth image curves at multiple scales. *International Journal of Computer Vision* **1989,** *3* (2), 119-130.
20. Smeulders, A.; Dorst, L., Decomposition of discrete curves into piecewise straight segments in linear time. *Contemporary Mathematics* **1991,** *119*, 169-195.
21. Bertsekas, D. P., *Constrained optimization and Lagrange multiplier methods*. Academic press: 2014.
22. Saint-Marc, P.; Chen, J.; Medioni, G., Adaptive smoothing: a general tool for early vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **1991,** *13* (6), 514-529.