Automated method for the determination of the band gap energy of pure and mixed powder samples using diffuse reflectance spectroscopy

A. Escobedo-Morales a,∗, I. I. Ruiz-Lópex a, M.deL. Ruiz-Peralta a, L. Tepech-Carrillo b,c, M. Sánchez-Cantú a, J. E. Moreno-Orea a

a Facultad de Ingeniería Química, Benemérita Universidad Autónoma de Puebla, Av. San Claudio y 18 Sur, C.P. 72570, Ciudad Universitaria, Puebla, Puebla, Mexico
b Universidad Autónoma Benito Juárez de Oaxaca, Escuela de Ciencias, C.P. 68120, Oaxaca de Juárez, Oaxaca, Mexico
c Facultad de Ingeniería, Benemérita Universidad Autónoma de Puebla, Av. San Claudio y 18 Sur, C.P. 72570, Ciudad Universitaria, Puebla, Puebla, Mexico

∗ Corresponding author.
E-mail address: alejandro.escobedo@correo.buap.mx (A. Escobedo-Morales).

Abstract

An automated method to determine the band gap energy (Eg) of pure and mixed powder compounds using diffuse reflectance spectroscopy is presented. This method is based on a five-step algorithm that mimics the judgment made by an expert analyst in identifying the linear segments in Tauc plots and subsequent estimation of the Eg value. It is demonstrated that the method to estimate Eg by intersecting the straight-line fit of the Tauc segment with the photon energy axis is not appropriate for those samples containing more than one optical absorbing phase because systematic underestimation of the Eg value results. The automated method accounts for such cases by introducing a base line function. The robustness of the implemented algorithm was tested using three model systems, ZnO-Al2O3, ZnO-CoO and ZnO-CdO. The estimated Eg’s using the automated method differ in less than 1% than those obtained by its manual counterpart.
1. Introduction

The band gap energy \( E_g \) is a pivotal characteristic of the electronic structure of materials, and frequently, it determines their potential applications. However, like other physical and chemical properties of materials, the value of \( E_g \) can be notoriously influenced by the synthesis route and/or processing method. Therefore, precise determination of its value is a concern in materials science and engineering. In this sense, optical characterization techniques in combination with the well-known Tauc method is by far the most widely used procedure to determine the \( E_g \) of supported materials \([1, 2]\). The Tauc method is based on the relationship between \( E_g \) and the optical absorption coefficient \( \alpha \), according to Eq. (1):

\[
\alpha h\nu = C_1 (h\nu - E_g)^n
\]  

(1)

where \( C_1 \) is a proportionality constant, \( h\nu \) is the energy of the incident photon and \( n \) is a coefficient that depends on the kind of electronic transition, being, \( n = 1/2 \) for direct allowed transition, \( n = 3/2 \) for direct forbidden transition, \( n = 2 \) for indirect allowed transition, and \( n = 3 \) for indirect forbidden transition. Basically, in the Tauc method the value of \( E_g \) is found at the intersection of the straight-line fit of the region associated to the optical absorption edge (i.e., the Tauc segment; TS) and the \( h\nu \)-axis in the corresponding \( (\alpha h\nu)^{1/n} \) vs \( h\nu \) plot (Tauc plot). It has been proved that, when appropriately executed, this method gives accurate values of \( E_g \) \([3]\). However, because \( \alpha \) is used to construct the Tauc plot, negligible light scattering is required to apply the method successfully.

In the case of powders samples, the scattering component cannot be neglected; thus, optical absorption spectroscopy is not an appropriate technique to determine the value of \( E_g \). In its place, diffuse reflectance spectroscopy (DRS) is a better choice as Escobedo \textit{et al.} demonstrated \([4]\). For further details, the reader is referred to excellent reviews concerning the Kubelka-Munk theory of diffuse reflectance (DR) phenomenon \([5, 6, 7, 8, 9]\) and DRS \([10, 11]\).

Using DRS, the analogous Tauc plots can be obtained from calculating the Kubelka-Munk or reemission function, \( F(R_\infty) \), according to Eqs. (2), (3), and (4):

\[
R_\infty = \frac{R_{\text{sample}}}{R_{\text{standard}}}
\]  

(2)

\[
F(R_\infty) = \frac{K}{S} = \frac{(1 - R_\infty)^2}{2R_\infty}
\]  

(3)
\[ F(R_{\infty})h\nu = C_2(h\nu - E_g)^n \]  \hspace{1cm} (4)

where \( R_{\infty} \) is the reflectance of the sample with “infinite thickness”, hence, there is no contribution of the supporting material, \( K \) and \( S \) are the absorption and scattering K-M coefficients, respectively, and \( C_2 \) is a proportionality constant. In practice, thicknesses \( > 2 \text{ mm} \) are enough to avoid some contribution from the supporting material (the sample holder), and the value of \( R_{\infty} \) is equal to the ratio of the reflectance of the sample (\( R_{\text{sample}} \)) and a diffuse reflectance standard material (\( R_{\text{standard}} \)); commonly, polytetrafluoroethylene (PTFE) or BaSO\(_4\) pellets. In correspondence with optical absorption spectroscopy, the value of \( E_g \) is found in an analogous Tauc plot, \( (F(R_{\infty})h\nu)^{1/n} \) vs \( h\nu \), at the intersection of the TS and the \( h\nu \)-axis [12, 13, 14].

Although the bases of the Tauc method are well-established, frequently, the success in the determination of \( E_g \) depends on the analyst’s experience. It is because the method requires manual identification of the TS to be fitted. Moreover, the \( xy \)-aspect ratio of the Tauc plot might induce subjectivity in recognizing the range of data considered in the straight-linear fitting (see Fig. 1). This process complicates further when more than one absorption edge is observed, and, as we will show, straightforward use of the Tauc method in polyphasic samples could lead to systematic underestimation of the \( E_g \) of those constituents with wider band gaps [15, 16].

![Fig. 1. Tauc plot of a powder mixture sample presented in two different \( xy \)-aspect ratios. (a) 1:1 and (b) 2:1.](https://doi.org/10.1016/j.heliyon.2019.e01505)

Recently, Suram et al. [17] developed an algorithm for automated Tauc analysis. It is based on a piecewise linear fit of the Tauc plot. Although the estimated \( E_g \)'s values agree with those of experienced scientists, the algorithm requires many predetermined process parameters; not all them intuitive. Additionally, since the slope of the linear segments is used frequently as input data in the decision boxes, this approach makes the algorithm vulnerable to signal noise, especially in those cases when it cannot be reduced by traying hard filtering, because it significantly distorts the original spectrum.
Here, an alternative algorithm for automated determination of the $E_g$ of pure and mixed powder samples using DRS is presented. This algorithm proceeds in a more human-like way to find the linear segments in Tauc plots. Additionally, it accounts for the influence of additional phases in the determination of $E_g$.

2. Materials and methods

2.1. Preparation of samples

ZnO (#96479, Sigma—Aldrich, St. Louis, MO), Al$_2$O$_3$ (#11028, Sigma—Aldrich, St. Louis, MO), CoO (#343153, Sigma—Aldrich, St. Louis, MO) and CdO (#244783, Sigma—Aldrich, St. Louis, MO) powders were purchased and used without further purification. Using these compounds, three samples sets were prepared by mixing ZnO with a second oxide powder: Al$_2$O$_3$, CoO or CdO. Each set consists of five mechanical mixtures with different ZnO content: 100, 75, 50, 25 and 0 wt.%. While ZnO was selected as model compound, Al$_2$O$_3$ and CoO were used as whitening and darkening constituents, respectively; CdO was selected to incorporate a second TS in the Tauc plots.

The procedure to prepare the samples was as follows. First, the corresponding amount of each constituent was put into an agate mortar and mixed gently during 30 min. The total mass of the mixture was fixed in 1.0 g and acetone droplets were added to avoid loss of powders. Then, the mixture was poured into a baker containing 10 ml of acetone and sonicated during 1 min. Finally, the suspension was verted in a PTFE crucible and dried in air at 60 °C during 2 h.

2.2. Characterization of materials

The samples were characterized by X-ray diffraction using Cu-K$_\alpha$ radiation (XRD; D8 Discover X-ray diffractometer, Bruker, Billerica, MA), scanning electron microscopy (SEM; JSM-6610LV, Jeol USA, Inc., Peabody, MA) and diffuse reflectance spectroscopy (DRS; Cary 5000 UV-Vis-NIR spectrophotometer, Agilent, Santa Clara, CA). A pre-packed PTFE cell was used as a reflectance standard (#9910111400, Agilent). The DRS measurements were carried out in triplicate.

2.3. Automated method

The automated method used to determine the $E_g$ of the powder samples from their DR spectra is based on a five-step algorithm, namely, preprocessing, segmentation, merging, identification of Tauc segment (TS)-base line segment (BLS) pairs, and estimation of $E_g$’s. Each step is described in the following sections. The stages segmentation, merging and identification of TS-BLS pairs use filtered and scaled spectra, whereas the estimation of $E_g$’s use the original raw data.
2.3.1. Preprocessing

The algorithm starts by applying the Savitzky-Golay filtering method to the DR spectrum to remove excessive signal noise. Then the filtered spectrum is used to calculate the corresponding Tauc plot ($TP_f$ vs $h\nu$). Finally, a scaled plot ($TP_s$ vs $h\nu$) is obtained by using Eqs. (5) and (6):

$$TP_s = \phi \frac{TP_f}{\max(TP_f)}$$

$$\phi = [\max(h\nu) - \min(h\nu)]$$

Suram et al. [17] used Eq. (5) with $\phi = 1$ to normalize their data, thus making $0 \leq TP_s \leq 1$. However, by using Eq. (6), the length of $TP_s$ and $h\nu$ intervals are brought equal, i.e., a square $TP_s$ vs $h\nu$ plot is obtained. The idea behind this procedure is to avoid distortion of perceived inclination and/or length of the linear segments in Tauc plots as might occur while manual estimation of $E_g$ is performed using an arbitrary $xy$-aspect ratio.

2.3.2. Segmentation

This stage involves the recursive bisection of the $TP_s$ vs $h\nu$ plot. First, a straight-line is fitted to the whole spectrum. Then its determination coefficient $R^2$ is compared with the parameter $R_{tol^2}$, which establishes the segmentation sensitivity. If $R^2 < R_{tol^2}$ the spectrum is bisected, and new straight-line fits are performed to each resulting data group. Again, the value of $R^2$ of each straight-line fit is calculated and compared against $R_{tol^2}$. Iterated bisections continue in those data groups until the inequality $R^2 \geq R_{tol^2}$ is satisfied. Fig. 2 illustrates the initial bisection iterations of a scaled Tauc plot.

![Fig. 2. Iterative segmentation process of a scaled Tauc plot. Solid circles correspond to bisection points.](https://doi.org/10.1016/j.heliyon.2019.e01505)
2.3.3. Merging

In this step, the adjacent data groups whose fitted lines have similar inclination are recursively merged. The similitude criterion is based on a predefined inclination angle difference \( \Delta \theta_{\text{tol}} \). The procedure runs as follows. First, the inclination angle \( \theta_i \) of each \( i \)-segment is estimated from its fitted slope. Then, the consecutive forward angle difference for the \( i \)-segments is estimated as \( \Delta \theta_i = \theta_{i+1} - \theta_i \), and the segment pair with the minimum angle difference \( |\Delta \theta_{\text{min}}| \) is identified. If \( |\Delta \theta_{\text{min}}| < \Delta \theta_{\text{tol}} \), they are merged and a new straight-line is fitted to the merged data group. Next, the \( \Delta \theta_i \)'s set is updated and that segment pair having the new \( |\Delta \theta_{\text{min}}| \) is located to proceed merging process. These operations are repeated until \( |\Delta \theta_{\text{min}}| \) is no longer smaller than \( \Delta \theta_{\text{tol}} \). Notice that merging propagates from anywhere within the spectrum, i.e., the data are not swept from left to right. Suram et al. [17] used a merging criterion based on slopes \((m)\), but it has some drawbacks. When \( m \) is used to measure the inclination of the fitted lines, its value lies in the open interval \((-\infty, \infty)\). Conversely, a closed interval \([-90^\circ, 90^\circ]\) is obtained if \( \theta \) is chosen instead. As a result, inclination changes are always perceived the same for a fixed \( \Delta \theta \). Moreover, small \( \Delta m \) values can be associated to large \( \Delta \theta \)'s when \( \theta \rightarrow 0^\circ \) or large \( \Delta m \) values might be associated to negligible \( \Delta \theta \)'s when \( \theta \rightarrow \pm 90^\circ \). Fig. 3 exemplifies the iterative merging process of a segmented Tauc plot.

![Iterative merging process of a segmented Tauc plot](image)

**Fig. 3.** Iterative merging process of a segmented Tauc plot. Enclosed segments correspond to those where \( |\Delta \theta_{\text{min}}| \) is identified in successive iterations.

2.3.4. Identification of Tauc segment-base line segment pairs

The proposed algorithm initially identifies potential TS-BLS pairs and then examines their validity. Let us consider \( T \) and \( B \) are the indices of some potential TS and BLS pair, respectively. The algorithm first verifies whether the \( T \)-segment qualifies as a valid TS by using the following rules: (1) the number of elements of the
data group $N_T$ from which it was generated is more than two ($N_T > 2$); (2) it is neither the first nor the last $i$-segment in the spectrum (being $i = 1, \ldots, s; T \neq 1, s$); (3) it has a positive inclination angle $\theta_T$ ($\theta_T > 0$); (4) its $\theta_T$ is larger than those of the adjacent segments ($\theta_T > \theta_{T \pm 1}$). Next, the potential BLS for each TS is identified by using the following rules: (1) the number of elements of the data group $N_B$ from which it was generated is more than two ($N_B > 2$); (2) the ratio of its length $L_B$ and that of the corresponding TS, $L_T$, is larger than the minimum ratio $w_L$ ($L_B/L_T > w_L; 0 \leq w_L \leq 1$); (3) its inclination angle, $\theta_B$, is lower than a half of its corresponding TS ($\theta_B/\theta_T < 0.5$); (4) it is located at the left side of its corresponding TS, but at the right of the next TS downward ($T_j > B_j > T_{j-1}$). Up to this point, several scenarios are possible for each potential TS. Table 1 summarizes them along with the action taken by the algorithm. Fig. 4 illustrates the geometrical properties of TS-BLS pairs considered for subsequent analysis by the algorithm.

Table 1. Action taken by the algorithm under the different scenarios in identifying Tauc segment (TS)-base line segment (BLS) pairs.

| Scenario                        | Action                                                                 |
|---------------------------------|------------------------------------------------------------------------|
| None segment qualifies as BLS   | The potential TS is discarded.                                          |
| A single segment qualifies as BLS| The TS-BLS pair is stored for subsequent analysis.                     |
| Multiple segments qualify as BLS| The segment with the minimum absolute inclination angle $|\theta_B|$ is chosen as BLS; the TS-BLS pair is stored for subsequent analysis. |

Fig. 4. Geometrical properties of a valid Tauc segment (TS)-base line segment (BLS) pair.
2.3.5. Estimation of $E_g$’s

In this step, the initial and end points of the identified TS-BLS pairs in the TPs vs $hv$ spectrum are mapped into the original Tauc plot. Next, straight-lines are fitted to raw data to obtain the linear functions corresponding to the TS and BLS, $f_{TS}$ and $f_{BLS}$, respectively. Thus, using the slope ($m$) and intercept ($b$) values of $f_{TS}$ and $f_{BLS}$, the coordinates of their intersection point are determined according to Eqs. (7) and (8):

$$x_{int} = \frac{b_{TS} - b_{BLS}}{m_{BLS} - m_{TS}}$$  \hspace{1cm} (7)

$$y_{int} = \frac{b_{BLS}m_{TS} - b_{TS}m_{BLS}}{m_{TS} - m_{BLS}}$$  \hspace{1cm} (8)

After, the length of the $hv$-axis standard error interval, $\sigma_{tol}$, is calculated from the corresponding standard errors ($\sigma$) of $m$ and $b$ of $f_{TS}$ and $f_{BLS}$ using Eqs. (9) and (10):

$$\sigma_{m_{TS},BLS} = \sqrt{\frac{\sum_{k=1}^{N} (y_k - (b_{TS,BLS} + m_{TS,BLS}x_k))^2}{N - 2}}$$  \hspace{1cm} (9)

$$\sigma_{b_{TS},BLS} = \sqrt{\frac{\sum_{k=1}^{N} (y_k - (b_{TS,BLS} + m_{TS,BLS}x_k))^2}{N - 2}} \left( \frac{1}{N} + \frac{\bar{x}^2}{\sum_{k=1}^{N} (x_k - \bar{x})^2} \right)$$  \hspace{1cm} (10)

Uncertainty of slope and intercept estimation can be reported as $m \pm \sigma_m$ and $b \pm \sigma_b$, respectively. Thus, from Eqs. (11) and (12), the set of uncertainty linear functions corresponding to $f_{TS}$ and $f_{BLS}$ ($g_{TS}$ and $g_{BLS}$, respectively) may be calculated as:

$$g_{TS} = (b_{TS} \pm \sigma_{b_{TS}}) + (m_{TS} \pm \sigma_{m_{TS}})x$$  \hspace{1cm} (11)

$$g_{BLS} = (b_{BLS} \pm \sigma_{b_{BLS}}) + (m_{BLS} \pm \sigma_{m_{BLS}})x$$  \hspace{1cm} (12)

Finally, the length of the $hv$-axis standard error interval, $\sigma_{tol}$, is determined as the difference between the minimum and maximum intersections of $g_{TS}$ and $g_{BLS}$ (Eq. (13)):

$$\sigma_{tol} = \max \left[ \frac{(b_{TS} \pm \sigma_{b_{TS}}) - (b_{BLS} \pm \sigma_{b_{BLS}})}{(m_{BLS} \pm \sigma_{m_{BLS}}) - (m_{TS} \pm \sigma_{m_{TS}})} \right] - \min \left[ \frac{(b_{TS} \pm \sigma_{b_{TS}}) - (b_{BLS} \pm \sigma_{b_{BLS}})}{(m_{BLS} \pm \sigma_{m_{BLS}}) - (m_{TS} \pm \sigma_{m_{TS}})} \right]$$  \hspace{1cm} (13)
The final step in the algorithm is to assign the value of $x_{int}$ to $E_g$. Prior, the intersection point $(x_{int}, y_{int})$ is evaluated to discriminate it from the spurious ones introduced by signal noise. A valid intersection point of a TS-BLS pair must satisfy the following rules: (1) $x_{int}$ is not located outside the $h\nu$-range established by the TS-BLS pair and also is larger than $\sigma_{tol}$ ($x_{B, min} \leq x_{int} \leq x_{T,max}$; $x_{int} > \sigma_{tol}$); (2) $y_{int}$ is not located above a preestablished height $y_{tol}$ associated with the TS ($y_{int} < y_{tol}$; $y_{tol} = y_{T,min} + 0.2 (y_{T,max} - y_{T,min})$). Fig. 5 shows the geometrical requirements that intersection points $(x_{int}, y_{int})$ must satisfy to assign $x_{int}$ as $E_g$.

![Diagram of a Tauc segment (TS)-base line segment (BLS) pair with geometrical requirements.]

**Fig. 5.** Geometrical requirements of a valid intersection point $(x_{int}, y_{int})$ of a Tauc segment (TS)-base line segment (BLS) pair.

The automated estimation of the $E_g$’s of the pure and mixed powder samples was performed using the following control parameters. 3 and 69 were chosen as the polynomial degree and window size in the Savitzky-Golay filtering routine to avoid obtaining noisy and/or distorted spectra [18]. The $R_{tol}^2$ parameter was set in 0.990. This value allows to identify satisfactorily the limits of the TS and BLS while the algorithm performance against signal noise holds robust. An excessive segmentation may result when $R_{tol}^2 \rightarrow 1$, especially in those noisy spectra. The $\Delta \theta_{tol}$ was 3°. Higher inclination angle differences give curvy TS and/or BLS, whereas lower angles might result in short segments. Finally, TS/BLS length ratio $w_L$ was established in 0.2. This value provides reliable results when two TS’s are close and, at the same time, prevents to assign as BLS some tiny segment located at the absorption band tail.

The algorithm described above has been implemented in the GapExtractor© software. The software and raw data used in this study are available for downloading at the Mendeley Data repository [19].
3. Results and discussion

Fig. 6 presents SEM micrographs of the pure oxide powders. As can be seen, they are constituted by aggregates of sub-micrometer particles. Although faceted crystals are observed, neither a definite habit nor characteristic size are recognized.

![Representative SEM micrographs of the pure (a) ZnO, (b) Al₂O₃, (c) CoO, and (d) CdO powders.](image)

The X-ray diffractograms of the pure oxide powders are presented in Fig. 7a. In each case, the observed X-ray reflections could be assigned to a single phase, namely, wurtzite ZnO (JCPDS #36-1451), corundum Al₂O₃ (JCPDS #46-1212), cubic CoO (JCPDS #48-1719) and cubic CdO (JCPDS #05-0640). In general, the X-ray peaks are sharp and, for some samples, the Kα doublet is resolved even at low diffraction angles, indicating high crystallinity. Fig. 7b shows the X-ray patterns of selected ZnO-Al₂O₃, -CoO and -CdO mixture samples. As expected, the diffractograms of the mixtures present the reflections corresponding to both constituents. Again, the X-ray peaks are sharp and no additional signal attributable to a third phase
was detected. It is concluded that the mixing method does not lead to formation of a new phase or degrades the crystallinity of the materials.

The DR spectra of the powder mixtures are shown in Fig. 8. In the case of the ZnO-Al$_2$O$_3$ set, the spectra of the samples containing ZnO show an abrupt decrease of the reflectance around 380 nm. It is associated to the optical adsorption edge of ZnO, whose $E_g$ at 300 K is $\sim 3.3$ eV ($\sim 375$ nm) [20, 21, 22]. As can be noted, the use of Al$_2$O$_3$ as diluent (≤75 wt.%) has negligible effect in the DR spectrum. It is because multiple light reflections within the agglomerate give ZnO the opportunity to absorb the fraction of light reflected by Al$_2$O$_3$ grains (Al$_2$O$_3$–$E_g$ $\sim 5$ eV at 300 K) [20, 22, 23]. Oppositely, addition of CoO reduces significantly the reflectance of visible light. This effect is more evident as the CoO wt.% increases. Since the $E_g$ of CoO at 300 K is $\sim 1$ eV ($\sim 1240$ nm) [20, 22], no additional adsorption edge associated to this compound appears in the UV-Vis range. Like CoO, addition of CdO powder decreases the value of $R\%$ in the visible region. However, in the DR spectra of the ZnO-CdO mixtures, a second absorption edge extending from 550 nm to 800 nm appears besides that attributed to ZnO. It is associated to the optical band gap absorption of CdO, whose $E_g$ at 300 K is $\sim 2.2$ eV ($\sim 565$ nm) [20, 22, 24].
The DR spectra of ZnO-CoO and CdO mixtures demonstrate that $R\%$ is sensitive to the relative amounts of the optical absorbing materials. In this regard, several models to determine the optical response of powder mixtures have been proposed [10, 25]. These models consider that the reemission function of a mixture of $N$ constituents, $F_m (R_{\infty})$, is just the ratio of the summation of their $K$- and $S$-coefficients, each weighted by the mass fraction, $w_i$, of the $i$-constituent:

$$F_m (R_{\infty}) = \frac{\sum_{i=1}^{N} w_i K_i}{\sum_{i=1}^{N} w_i S_i}$$  \hspace{1cm} (14)

Where $\sum_{i=1}^{N} w_i = 1$  \hspace{1cm} (15)

The main limitation to calculate $F_m (R_{\infty})$ using Eqs. (14) and (15) is the requirement of knowing both K-M coefficients of each constituent; not the $K_i/S_i$ ratios. Although several methods to achieve this goal have been developed, in practice, they are difficult and tedious [26, 27].

Fig. 8. Diffuse reflectance spectra of the (a) ZnO-Al$_2$O$_3$, (b) ZnO-CoO and (c) ZnO-CdO powder mixtures; the 2nd and 3rd measurement replicates are shown in gray. (d) Dependence of the reemission function, $F (R_{\infty})$, on the reflectance percent, $R\%$. 

[https://doi.org/10.1016/j.heliyon.2019.e01505](https://doi.org/10.1016/j.heliyon.2019.e01505)
Since the reemission function of powder samples can be directly calculated from their DR spectra, knowledge of the K-M coefficients is not necessary to determine the $E_g$ of pristine samples using DRS and the Tauc method. The success of this methodology lies in considering the $S$-coefficient as a constant. In the case of homogeneous mixtures, and because the DR reflection of powder samples is a collective phenomenon, the $S$-coefficient of each constituent, $S_i$, can be approached with relative confidence to an average value $S_m$. Therefore, $F_m(R_\infty)$ can be approximated to

$$F_m(R_\infty) = \frac{1}{S_m} \sum_{i=1}^{N} w_i K_i = \sum_{i=1}^{N} w_i F_i(R_\infty)$$  \hspace{1cm} (16)$$

Eq. (16) states that the reemission function of mixtures is determined by the major constituent and/or that with the largest $K$-coefficient. Conversely, constituents with large optical reflectance ($R\% \geq 50\%$) have negligible effect in the reemission function of mixtures. This is because $F(R_\infty)$ decays sharply as $R\%$ increases (see Fig. 8). Once having $F_m(R_\infty)$, it can be used to obtain the corresponding Tauc plot to determine the different $E_g$’s. However, as described below, additional considerations to that of pristine samples must be attended to estimate their values appropriately.

Using Eq. (16), the analogous expression to Eq. (4) for sample mixtures is derived:

$$[F_m(R_\infty)hv]^{1/n} = \left[ \sum_{i=1}^{N} w_i F_i(R_\infty) \right]^{1/n}$$  \hspace{1cm} (17)$$

Eq. (17) implies that the value of $E_g$ of the $i$-constituent cannot simply be obtained by intersecting the straight-line fit of the corresponding TS ($f_{TS}$) with the $hv$-axis. To illustrate this, consider a mixture of two direct band gap semiconductors, 1 and 2, with mass fractions $w_1$ and $w_2$, and reemission functions $F_1(R_\infty)$ and $F_2(R_\infty)$, respectively. According to Eq. (16), the reemission function of this mixture, $F_{1,2}(R_\infty)$, is given by Eq. (18):

$$F_{1,2}(R_\infty) = w_1 F_1(R_\infty) + w_2 F_2(R_\infty)$$  \hspace{1cm} (18)$$

and from Eq. (17), the Tauc plot corresponds to

$$[F_{1,2}(R_\infty)hv]^2 = [w_1 F_1(R_\infty)hv + w_2 F_2(R_\infty)hv]^2$$  \hspace{1cm} (19)$$

Defining a new function $f_{BLS}$ as

$$f_{BLS} = [w_2 F_2(R_\infty)hv]^2 + 2w_1 w_2 F_1(R_\infty) F_2(R_\infty)(hv)^2$$  \hspace{1cm} (20)$$

and after combining Eqs. (19) and (20) gives
\[ F_{1,2}(R_w)hv = |w_1 F_1(R_w)hv| + f_{BLS} \]

As expected, if \( w_2 = 0 \) then \( f_{BLS} \) vanishes, and Eq. (21) reduces to the pristine case. Note that \( f_{BLS} \) also approaches to zero in the \( hv \)-range where \( F_2(R_w) \) is small, which corresponds to large \( R\% \) values of the constituent 2, otherwise \( f_{BLS} \) cannot be neglected. Commonly, \( F(R_w) \) is a smooth function with low curvature below and above the absorption edge. This allows with confidence to model \( f_{BLS} \) as a linear function of \( hv \). Under this scheme, \( f_{BLS} \) can be interpreted as the BLS whose intersection with the TS provides the corresponding \( E_g \) value (see Fig. 9a). As discussed above, for pristine samples \( f_{BLS} = 0 \), thus the \( E_g \) value is found at the intersection of the \( f_{TS} \) with the \( h v \)-axis, as is usually done. Nonetheless, a BLS must be considered for those compounds whose TS lies within the \( hv \)-range where the reemission function of other constituents is not negligible. Under this condition, the intersection point in the Tauc plot coinciding with the correct \( E_g \) value is shifted above the \( h v \)-axis by an amount equal to \( \Delta_{BLS} = f_{BLS}(E_g) \) (see Fig. 9b).

The validity of this methodology and the performance of the algorithm described in Section 2.3 were tested using the ZnO-Al\(_2\)O\(_3\), -CoO and -CdO mixture sets. For comparison, the \( E_g \) values were determined using three different methods: (1) manual determination of the \( E_g \) without considering a BLS (\( h v \)-axis method), (2) manual determination of the \( E_g \) considering a BLS (base line method), and (3) determination of the \( E_g \) using the proposed algorithm (automated method). The estimated \( E_g \) of each semiconductor phase contained in the different sample mixtures are summarized in Table S1 (see Supplementary material). Fig. 10 presents the statistical analysis of the dependence of \( E_g \) with the mixture composition determined by using each of the above-mentioned methods.
In the case of the ZnO-Al₂O₃ mixtures, the estimated \( E_g \) of the semiconductor phase (ZnO) is not sensitive to the Al₂O₃ content. It was determined in 3.25 eV using either the \( h\nu \)-axis or the base line method; the automated method gives the same value. Since the \( R\% \) of Al₂O₃ is high within the wavelength range where the absorption edge of ZnO appears, the value of the \( f_{BLS} \approx 0 \). Therefore, in this case, the \( h\nu \)-axis and base line methods are equivalent. The automated method succeeded in identifying and fitting the TS and BLS, thus the estimated \( E_g \) matches with those obtained by manual methods (see Figs. S1-S13 and Tables S2-S14 in the Supplementary material).

According to the \( h\nu \)-axis method, when the ZnO powders are mixed with CoO, its \( E_g \) changes with CoO content. It decreases monotonically from 3.25 eV (0 wt.% CoO) to 3.15 eV (75 wt.% CoO). The \( h\nu \)-axis method fails in determining the \( E_g \) value of ZnO because, in this case, \( f_{BLS} \) cannot be neglected. Thus, by intersecting the TS with the \( h\nu \)-axis instead of the corresponding BLS, the value of \( E_g \) is underestimated as much as the \( \Delta_{BLS} \) magnitude. Conversely, the base line and automated methods give values of \( \approx 3.24 \) eV, regardless the CoO wt.%.

Like ZnO-CoO mixtures, the \( h\nu \)-axis method underestimates the \( E_g \) of ZnO as the CdO mass fraction increases. Again, it is because the \( E_g \) of CdO is smaller than...
that of ZnO, thus its optical absorption contributes to the reemission function in the $h\nu$-range where the TS associated to the ZnO is located. Oppositely, the base line and automated methods give consistent values for the $E_g$ of ZnO ($\approx 3.24$ eV) regardless of CdO content. Unlike ZnO, the estimation of the $E_g$ of CdO is $\sim 2.10$ eV whatever method is used. It is because ZnO acts a diluent for CdO, like Al$_2$O$_3$ does for ZnO.

The above results agree with the equations to model the DR of powder mixtures. The main implication for this kind of samples is the requirement to identify in the Tauc plots the linear segment associated to both TS and BLS. It is because in the samples containing several optical absorbing materials the intersection point containing the true $E_g$ value is displaced above the $h\nu$-axis. Conversely, if the reemission function of the other constituents is negligible around the TS of certain phase, determining a BLS is not mandatory. Whatever the scenario, pure or mixed powder samples, the proposed automated method is robust and determines accurately the $E_g$ of each constituent.

The method described above can be extended to analyze blends of more than two constituents and to determine the $E_g$ of diluted phases (see Figs. S14-S15 and Tables S15-S16 in the Supplementary material). However, it is worth noting that the $E_g$ of an absorbing optical compound can be estimated using DRS only if those constituents with narrower band gap do not absorb almost completely the incident light, and/or its absorption band edge does not overlap another one.

### 4. Conclusions

The analysis of the obtained results demonstrates that DRS along with the Tauc method can be used with confidence to estimate the band gap energy ($E_g$) of pure and mixed powder samples. Nevertheless, several concerns must be taken to obtain accurate results. In general, once the appropriate Tauc plot has been calculated, the $E_g$ value of a certain compound is found at the intersection of the straight-line fits of its corresponding Tauc segment (TS)- base line segment (BLS) pair. This methodology accounts for those samples containing several optical absorbing materials. Otherwise, systematic underestimation of the $E_g$ values might be induced. In the case of pure samples, the base line function can be considered equal to zero. Thus, the $E_g$ value is found by just intersecting the straight-line fit of the TS with the $h\nu$-axis. The latter methodology gives accurate results meanwhile the $R\%$ at the low-energy side of the absorption edge be high ($>50\%$). Finally, the analysis of the obtained $E_g$ values demonstrates that the automated method successes in identifying TS-BLS pairs in the Tauc plots of either pure or mixed powder samples. Moreover, the calculated $E_g$’s differ in less than 1% than those estimated manually by an expert analyst.
Declarations

Author contribution statement

A. Escobedo-Morales: Conceived and designed the experiments; Performed the experiments; Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data; Wrote the paper.

I. I. Ruiz-López: Conceived and designed the experiments; Analyzed and interpreted the data; Wrote the paper.

M. de L. Ruiz-Peralta, L. Tepech-Carrillo, M. Sánchez-Cantú: Performed the experiments; Analyzed and interpreted the data.

J. E. Moreno-Orea: Performed the experiments.

Funding statement

This work was supported by the Vicerrectoría de Investigación y Estudios de Posgrado, Benemérita Universidad Autónoma de Puebla (Grant #100473300-VIEP2018).

Competing interest statement

The authors declare no conflict of interest.

Additional information

Supplementary content related to this article has been published online at https://doi.org/10.1016/j.heliyon.2019.e01505.

References

[1] J. Tauc, R. Grigorovici, A. Vancu, Optical properties and electronic structure of amorphous germanium, Phys. Status Solidi. 15 (1966) 627–637.

[2] J.I. Pankove, Optical Processes in Semiconductors, Dover Publications, Inc., New York, NY, 1971.

[3] B.D. Viezbicke, S. Patel, B.E. Davis, D.P. Birnie, Evaluation of the Tauc method for optical absorption edge determination: ZnO thin films as a model system, Phys. Status Solidi. 252 (2015) 1700–1710.

[4] A. Escobedo Morales, E. Sanchez Mora, U. Pal, Use of diffuse reflectance spectroscopy for optical characterization of un-supported nanostructures, Rev. Mexic. Fisica 53S (2007) 18–22.
[5] W.E. Vargas, G.A. Niklasson, Applicability conditions of the kubelka–munk theory, Appl. Opt. 36 (1997) 5580.

[6] B. Philips-Invernizzi, Bibliographical review for reflectance of diffusing media, Opt. Eng. 40 (2001) 1082.

[7] L. Yang, B. Kruse, Revised kubelka–munk theory I. Theory and application, J. Opt. Soc. Am. A 21 (2004) 1933.

[8] L. Yang, B. Kruse, S.J. Miklavcic, Revised Kubelka–Munk theory II Unified framework for homogeneous and inhomogeneous optical media, J. Opt. Soc. Am. A 21 (2004) 1942.

[9] L. Yang, S.J. Miklavcic, Revised Kubelka–Munk theory III A general theory of light propagation in scattering and absorptive media, J. Opt. Soc. Am. A 22 (2005) 1866.

[10] R.W. Frei, H. Zeitlin, Diffuse reflectance spectroscopy, CRC Crit. Rev. Anal. Chem. 2 (1971) 179–246.

[11] E.L. Simmons, Diffuse reflectance spectroscopy: a comparison of the theories, Appl. Opt. 14 (1975) 1380–1386.

[12] M. Nowak, B. Kauch, P. Szperlich, Determination of energy band gap of nanocrystalline SbSI using diffuse reflectance spectroscopy, Rev. Sci. Instrum. 80 (2009), 046107.

[13] R. López, R. Gómez, Band-gap energy estimation from diffuse reflectance measurements on sol–gel and commercial TiO2: a comparative study, J. Sol-Gel Sci. Technol. 61 (2012) 1–7.

[14] N. Sangiorgi, L. Aversa, R. Tatti, R. Verucchi, A. Sanson, Spectrophotometric method for optical band gap and electronic transitions determination of semiconductor materials, Opt. Mater. (Amst). 64 (2017) 18–25.

[15] A.A. Baqer, K.A. Matori, N.M. Al-Hada, A.H. Shaari, H.M. Kamari, E. Saion, J.L.Y. Chyi, C.A.C. Abdullah, Synthesis and characterization of of binary (CuO) 0.6 (CeO2) 0.4 nanoparticles via a simple heat treatment method, Results Phys. 6 (2018).

[16] F.E. Loranca-Ramos, C.J. Diliegros-Godines, R. Silva González, M. Pal, Structural, optical and electrical properties of copper antimony sulfide thin films grown by a citrate-assisted single chemical bath deposition, Appl. Surf. Sci. 427 (2018) 1099–1106.
[17] S.K. Suram, P.F. Newhouse, J.M. Gregoire, High throughput light absorber discovery, Part 1: an algorithm for automated Tauc analysis, ACS Comb. Sci. 18 (2016) 673–681.

[18] G. Vivó-Truyols, P.J. Schoenmakers, Automatic selection of optimal Savitzky-Golay smoothing, Anal. Chem. 78 (2006) 4598–4608.

[19] A. Escobedo-Morales, I.I. Ruiz-López, GapExtractor v1.0, 2018. https://data.mendeley.com/datasets/j9ypzmdx5n/draft?a=bfc2457b-5b5c-427e-9120-5c9c455f5057.

[20] W.H. Strehlow, E.L. Cook, Compilation of energy band gaps in elemental and binary compound semiconductors and insulators, J. Phys. Chem. Ref. Data. 2 (1973) 163–200.

[21] V. Srikant, D.R. Clarke, On the optical band gap of zinc oxide, J. Appl. Phys. 83 (1998) 5447–5451.

[22] W.M. Haynes, CRC Handbook of Chemistry and Physics, ninety fourth ed., CRC Press, Boca Raton, 2013.

[23] Z.K. Heiba, M.B. Mohamed, A.M. Wahba, N.G. Imam, Structural, optical, and electronic characterization of Fe-doped alumina nanoparticles, J. Electron. Mater. 47 (2018) 711–720.

[24] S.K. Vasheghani Farahani, V. Muñoz-Sanjosé, J. Zúñiga-Pérez, C.F. McConville, T.D. Veal, Temperature dependence of the direct bandgap and transport properties of CdO, Appl. Phys. Lett. 102 (2013) 1–5.

[25] D.R. Duncan, The colour of pigment mixtures, Proc. Phys. Soc. 52 (1940) 390–401.

[26] E. Walowit, C.J. McCarthy, R.S. Berns, An algorithm for the optimization of kubelka-munk absorption and scattering coefficients, Color Res. Appl. 12 (1987) 340–343.

[27] R. Molenaar, J.J. ten Bosch, J.R. Zijp, Determination of Kubelka–Munk scattering and absorption coefficients by diffuse illumination, Appl. Opt. 38 (1999) 2068–2077.