



# Reducing or eliminating noise in ellipsometric spectra

Long V. Le<sup>1</sup> · Tae Jung Kim<sup>2,3</sup> · Young Dong Kim<sup>2</sup> · David E. Aspnes<sup>4</sup>

Received: 6 May 2022 / Accepted: 9 May 2022 / Published online: 4 August 2022  
© The Korean Physical Society 2022

## Abstract

In the last several years, considerable progress has been made in reducing or eliminating noise in ellipsometric spectra, including the development of algorithms for eliminating endpoint-discontinuity artifacts, a reciprocal-space expression for quantifying the effectiveness of linear filters, and the corrected maximum-entropy (CME) approach for eliminating apodization and its associated errors. These lead to new capabilities, together with additional opportunities for extracting information from spectra. These results enhance the utility of ellipsometry for analyzing surfaces, interfaces, materials, and structures. Examples are provided.

**Keywords** Noise reduction · Gauss-Hermite kernels · Maximum-entropy · Ellipsometric spectra · Endpoint-discontinuity artifacts

## 1 Introduction

The nondistortional removal of noise from spectra has been an ongoing challenge in all forms of spectroscopy, ellipsometric and otherwise. The classic approach to noise reduction is direct- (spectral-) space (DS) convolution, as exemplified by the extensive tables of Savitzky and Golay (SG), published in 1964 [1]. Results are historically assessed by inspection, although concerns have been raised. In 1997 Kaiser and Reed [2] noted that noise-reduction methods tend to be applied blindly, with results simply inspected to see if they “look good.” In 1995 Barak [3] pointed out that too many approaches were “hit and miss.” While noise reduction

is a worthwhile goal for cosmetic reasons, additional advantages include minimizing uncertainties in parameters determined by lineshape analysis and in interpolation, whether simply to change the scale or more generally to convert spectra obtained linear in wavelength to spectra linear in energy [4].

The challenge is not simply to remove noise but to do it in ways that leave the underlying spectrum—the information—unchanged. Commonly known as filtering, noise reduction is typically done by linear methods, either by DS convolution, as mentioned above [1], or by reciprocal- (Fourier-) space (RS) processing [4–7]. Both take advantage of the separation of information and noise into low- and high-index Fourier coefficients, respectively, because structure follows from point-to-point correlations whereas noise is due to point-to-point fluctuations. This separation opens the path to optimizing filtering either directly in RS, through attenuation or replacement of unwanted coefficients, or indirectly in DS, by convolving the data with a set of coefficients. Filtering in RS can be assessed qualitatively by comparing the Fourier transform of the data with the transfer function of the filter [7]. In DS similar assessments require first calculating the Fourier coefficients of the data and convolving function, then comparing them in RS.

In principle, the ideal filter is that which leaves the information-containing coefficients unchanged while suppressing the noise-dominated coefficients completely. However, results obtained by applying this “brick-wall” (BW) filter

---

Long V. Le and Tae Jung Kim have contributed equally.

✉ Young Dong Kim  
ydkim@khu.ac.kr

✉ David E. Aspnes  
aspnes@ncsu.edu

<sup>1</sup> Institute of Materials Science, Vietnam Academy of Science and Technology, Hanoi 100000, Vietnam

<sup>2</sup> Department of Physics, Kyung Hee University, Seoul 02447, Republic of Korea

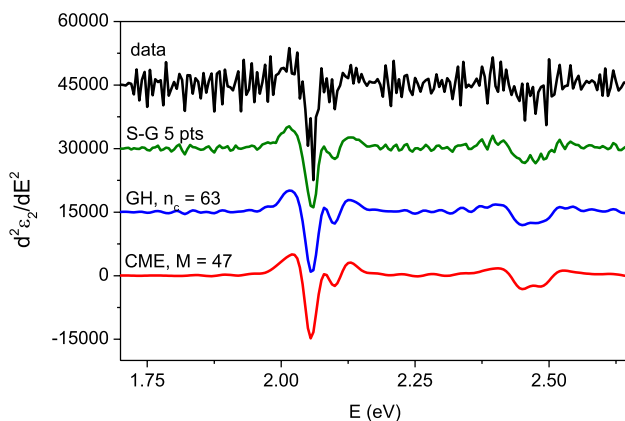
<sup>3</sup> Center for Converging Humanities, Kyung Hee University, Seoul 02447, Republic of Korea

<sup>4</sup> Department of Physics, North Carolina State University, Raleigh, NC 27695-8202, USA

are generally unacceptable because its abrupt cutoff generates Gibbs oscillations (ringing) in reconstructed spectra [8]. Because all linear filters incorporate apodization (cutoff), all require a compromise to be made among distortion of information, ringing, and leakage of noise.

In the absence of a quantitative RS measure to assess performance, a wide variety of transfer functions have been developed, largely empirically. However, Le et al. recently capitalized on Parseval's Theorem to develop an informative way of calculating mean-square deviations in RS. This expression allows the major sources of filtering errors to be recognized directly and assessed quantitatively. This led to the demonstration that the Gauss-Hermite (GH) is the best previous linear filter, and the development of an improved version, the cosine-terminated brick-wall (CT) [7].

The need for apodization—along with its associated errors—was recently eliminated with the development of the corrected maximum-entropy (CME) approach [9]. Based on calculations of Burg [10], this method projects, in a model-independent way, trends established by low-order coefficients into the white-noise region. By replacing high-order coefficients with values obtained analytically, this makes possible operations like constructing essentially noise-free second-derivative spectra, as seen in Fig. 1. Here, the second derivatives of  $\epsilon_2$  data obtained by spectroscopic ellipsometry (SE) on a monolayer  $\text{WS}_2$  sample at 41 K [11] are calculated by multiplying the data or filtered RS coefficients  $C_n$  by  $n^2$ , except that the 5-point second-derivative SG DS convolution was used directly. While the SG result is already a significant improvement relative to differentiating the data directly, the GH result is better. The best result is obtained with the CME filter. Further details are provided below. The figure is presented here to introduce capabilities.



**Fig. 1** Second derivatives of  $\epsilon_2$  of  $\text{WS}_2$  at 41 K obtained by multiplying the original, GH-filtered, and CME-filtered Fourier coefficients  $C_n$  by  $n^2$  as described below. The SG result is obtained by using the 5-point second-derivative DS convolution coefficients directly

## 2 Theory

### 2.1 Fourier analysis

CME is based on the coefficients of complex-exponential expansions rather than those of cosines and sines [9]. As these are less common, we provide basic equations here. We assume that a spectrum consists of  $(2N + 1)$  real, positive-definite data  $\{f_j\}$  where  $-N \leq j \leq N$ . Although Fourier analysis does not require the number of data to be odd, CME does. This restriction also simplifies mathematics.

Define the Fourier coefficients  $R_n$  according to

$$f_j = \sum_{n=-N}^N R_n e^{in\theta_j}, \quad (1a)$$

where

$$\theta_j = \frac{2\pi}{2N + 1} j. \quad (1b)$$

Because the  $f_j$  are real, the  $R_n$  satisfy the reality condition  $R_j^* = R_{-j}$ . From Eq. (1b) it follows that  $j = (-N - 1/2)$  at  $\theta = -\pi$  and  $j = (N + 1/2)$  at  $\theta = \pi$ , so all  $f_j$ ,  $j = -N$  to  $N$ , are interior points, an advantage. The inverse transformation is

$$R_n = \frac{1}{2N + 1} \sum_{j=-N}^N f_j e^{-in\theta_j}. \quad (2)$$

We assign the normalization factor  $(2N + 1)^{-1}$  to Eq. (2) rather than Eq. (1a) for reasons described below. The  $R_n$  are related to the more familiar cosine and sine coefficients  $A_n$  and  $B_n$  as

$$A_0 = R_0 = C_0; \quad (3a)$$

$$A_n - iB_n = 2R_n, \quad n \geq 1. \quad C_n = \sqrt{A_n^2 + B_n^2} = 2|R_n| \quad (3b)$$

Basing all calculations on a common range  $(-\pi \leq \theta \leq \pi)$  is also an advantage, but data are never obtained as a function of  $\theta$ . Hence projection is necessary. Using energy as an example, let  $E_i$  correspond to the first point  $j = -N$  and  $E_f$  to the last point  $j = +N$ . Then

$$\theta = \frac{4\pi}{(2N + 1)(E_f - E_i)} \left( E - \frac{1}{2}(E_f + E_i) \right). \quad (4a)$$

The inverse transformation is

$$E = \frac{1}{2}(E_i + E_f) + \frac{(2N + 1)(E_f - E_i)}{4\pi} \theta. \quad (4b)$$

### 2.2 Linear filtering

Linear filtering is defined most conveniently as a convolution in DS:

$$\bar{f}_j = \sum_{\kappa=-N}^N f_{j-\kappa} b_\kappa, \tag{5}$$

where the  $\{\bar{f}_j\}$  are the filtered data and  $\{b_\kappa\}$  are the filter coefficients. In Eq. (5), values of  $(j - \kappa)$  that lie outside the  $(-N, N)$  range are wrapped mod  $(2N + 1)$ . Following the normalization convention of Eqs. (1a, 1b) and (2), the Fourier coefficients associated with the  $b_\kappa$  are

$$B_n = \frac{1}{2N + 1} \sum_{\kappa=-N}^N b_\kappa e^{in\theta_\kappa}. \tag{6}$$

The reason for this normalization is now clear: because the  $\{b_\kappa\}$  form a unitary set, then  $B_0 = 1$ , as appropriate for a low-pass filter. The classic examples are the sets of coefficients published in 1964 by Savitzky and Golay [1]. Typically only a few  $b_\kappa$  are nonzero; for values that are not specified, the corresponding  $b_\kappa$  in Eqs. (5) and (6) are set equal to zero. For example, for the 3-point running-average filter,  $b_\kappa = 1/3$  for  $\kappa = -1, 0, 1$  and zero otherwise.

As noted in the Introduction, the effect of convolution in DS is represented in RS by the convolution theorem

$$\bar{F}_n = F_n B_n, \tag{7}$$

where  $\bar{F}_n, F_n,$  and  $B_n$  are the Fourier coefficients of  $\bar{f}_j, f_j,$  and  $b_\kappa$ , respectively. The above equation shows explicitly that the operation of the linear filter defined by the  $b_\kappa$  is independent of the data being processed. Capitalizing on Parseval’s Theorem, Le et al. [7] recently showed that the mean-square deviation  $\delta_{MSE}^2$  between  $\{f_j\}$  and  $\{\bar{f}_j\}$  is given in RS by

$$\delta_{MSE}^2 = \sum_{n=-N}^N |\bar{f}_j - f_j|^2; \tag{8a}$$

$$= \sum_{n=-N}^N |F_n|^2 |1 - B_n|^2. \tag{8b}$$

With  $B_0 = 1$  for a low-pass filter, Eq. (8b) provides clear justification for the Butterworth criterion of eliminating as many derivatives as possible in a Taylor-series expansion of  $B(n)$  about  $n = 0$ , treating the  $B_n$  as a continuous function  $B(n)$ .

Of the various linear filters that have been proposed, Eq. (8b) reduces the list to 2: the Gauss-Hermite filter mentioned previously [4, 12], and more recent development, the cosine-terminated brick-wall filter [7]. The

Gauss-Hermite filter can be written either as a sum of Hermite polynomials of increasing even order multiplying a Gaussian  $e^{-n^2/\Delta n^2}$ , as originally done by Hoffman et al., or as the product of the Gaussian and an  $M^{\text{th}}$  partial sum of  $e^{+n^2/\Delta n^2}$ , which more directly illustrates the removal of low-order derivatives in the Taylor-series expansion.. In the latter case

$$B(n) = \left( 1 + \zeta + \frac{1}{2}\zeta^2 + \dots + \frac{1}{M!}\zeta^M \right) e^{-\zeta}, \tag{9a}$$

where

$$\zeta = \frac{n^2}{\Delta n^2}. \tag{9b}$$

It is straightforward to show that the first nonvanishing term in a full Taylor-series expansion of Eq. (9a) is proportional to  $\zeta^{M+1}$ , illustrating that the GH filter satisfies the Butterworth criterion explicitly.

The cosine-terminated brick-wall filter is defined by

$$B(n) = 1; \quad \text{for } 0 \leq n \leq n_1; \tag{10a}$$

$$= a \cos((n - n_1)/\Delta n) \quad \text{for } n_1 \leq n \leq n_2; \tag{10b}$$

$$= 0 \quad \text{for } n > n_2, \tag{10c}$$

where  $a, n_1,$  and  $n_2$  are parameters. In applications to discrete data, the variable  $n$  in Eqs. (9a, 9b) and (10a, 10b, 10c) is an integer.

GH filters are shown in Fig. 2 for a range of  $M$  and compared to the CT filter for high  $M$  in Fig. 3. Figure 2 shows apodizations for a range of orders from zero (Gaussian line) to  $M = 100$ . As can be appreciated from the figure, GH filters are most effective at relatively high orders, for

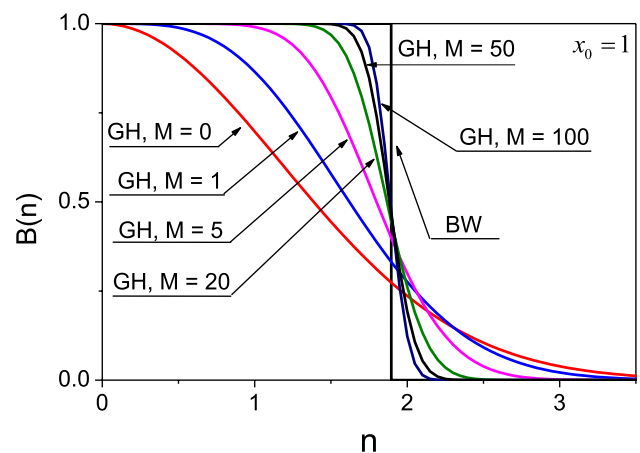
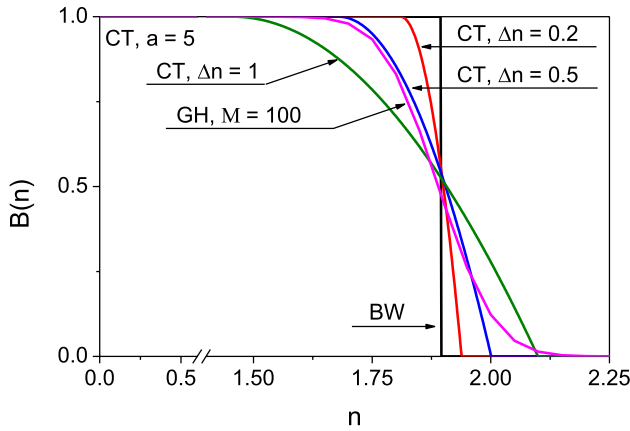


Fig. 2 Transfer functions of the GH filter for  $M=0, 1, 5, 20, 50,$  and  $100$



**Fig. 3** Transfer functions of the CT filter for  $x_0 = 1$ ,  $a = 5$ , and  $\Delta n = 0.2, 0.5$ , and  $1$ . The transfer functions of the BW and GH,  $M = 100$  filters are shown for comparison

example  $M \sim 50$  and above. This makes them particularly computationally intensive.

Figure 3 compares the  $M = 100$  GH filter to several CT filters for  $a = 5$  to illustrate its apodizations. As can be seen, the main difference between the GH filter and the CT filter for  $\Delta n = 0.5$  occurs at the upper cutoff. The distortions caused by the abrupt cutoff of the CT filter at the high end is more than compensated by its lack of distortion at low indices.

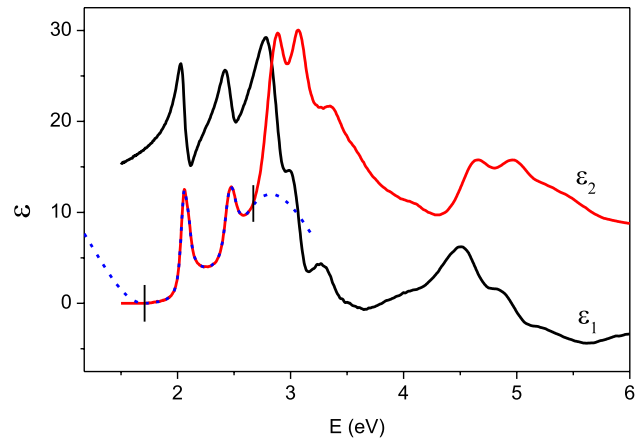
### 2.3 Nonlinear filters

The history of the corrected maximum-entropy filter is complicated, involving investigations of phase noise by Walker and Yule [13, 14], extraction of harmonics buried in noise in stationary-time-sequence data by many workers, and the forward-prediction theory of Kolmogorov and Wiener [15, 16]. Burg used a maximum-entropy approach [10] to develop a deconvolution (sharpening or “whitening”) procedure to better define the frequencies of weak signals buried in noise. The CME filter was identified by Le et al. [9] as an alternative solution in the analysis. We summarize the CME filter below.

Let  $P(\theta)$  be a continuous function with the Fourier representation

$$P(\theta) = \sum_{n=-\infty}^{\infty} R_n e^{in\theta}, \tag{11}$$

where the Fourier coefficients  $R_n$  are available from  $-N \leq n \leq N$ . However, because of noise only the coefficients



**Fig. 4** Real ( $\epsilon_1$ ) and imaginary ( $\epsilon_2$ ) parts of the dielectric function of monolayer  $WS_2$  at 41 K [11]. The dashed curve shows the ERED extension in the excitonic region, as described in the text

for  $-M \leq n \leq M$  are useful. Obviously,  $M$  is less, possibly much less, than  $N$ . However, we can recover the most-probable missing  $R_n$  by maximizing the entropy rate

$$S = \sum_{-\infty}^{\infty} \ln(P(\theta)) = \sum_{-\infty}^{\infty} \ln \left( \sum_{n=-\infty}^{\infty} R_n e^{in\theta} \right) \tag{12}$$

with respect to the  $R_n$  that are unavailable. After numerous steps the result is found to be given by

$$P(\theta) = \frac{1}{|\sum_{n=0}^M a_n e^{in\theta}|^2}, \tag{13}$$

where the coefficients  $a_n$  are given by

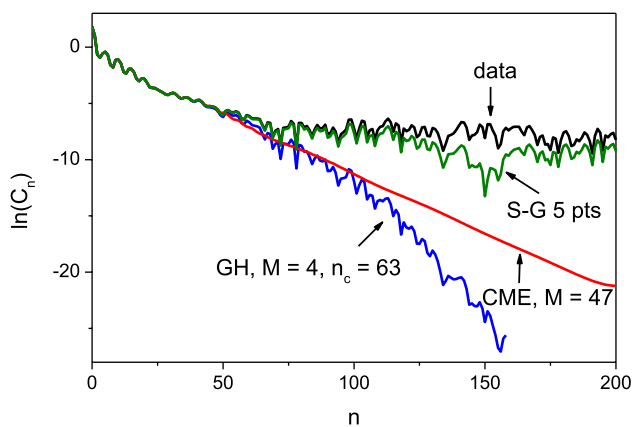
$$\begin{pmatrix} R_0 & R_1 & \dots & R_M \\ R_1^* & R_0 & \dots & R_{M-1} \\ \dots & \dots & \dots & \dots \\ R_M^* & R_{M-1}^* & \dots & R_0 \end{pmatrix} \begin{pmatrix} a_0^* \\ a_1^* \\ \dots \\ a_M^* \end{pmatrix} = \begin{pmatrix} 1/a_0 \\ 0 \\ \dots \\ 0 \end{pmatrix}. \tag{14}$$

Several points can be mentioned. First, because  $P(\theta)$  appears as the argument of a logarithmic function, it must be positive definite. Second, the infinite series Eq. (11) has been replaced by the reciprocal of a finite series, Eq. (13). Third, it is easily shown that the features generated by Eq. (13) are pseudo-Lorentzians, being periodic in  $\theta$  rather than decreasing monotonically to zero, as true Lorentzians. Consequently, being a modified spectral representation, the CME filter is particularly appropriate for Lorentzian lines generated by first-order decay processes. Finally, Eqs. (12) and (13) show that the CME procedure, or maximum-entropy more generally, intrinsically performs an inverse Fourier transformation.

### 3 Application

As an example, Fig. 4 presents the  $\text{WS}_2$  data used to generate Fig. 1. The sample temperature was 41 K. Data were obtained on a JAWoollam RC2 spectroscopic ellipsometer. The data, obtained linear in wavelength, were converted to linear in energy by linear interpolation between points. We are interested in the structure of the fundamental absorption edge giving rise to the features at approximately 2.1 and 2.5 eV. At issue is whether these are simple or composite features including one or more singularities. As indicated by the dashed line in the figure, the data of interest were augmented by the extended removal of endpoint discontinuity (ERED) process [17] to eliminate endpoint-discontinuity artifacts and, as a byproduct, double the number of Fourier coefficients available for analysis.

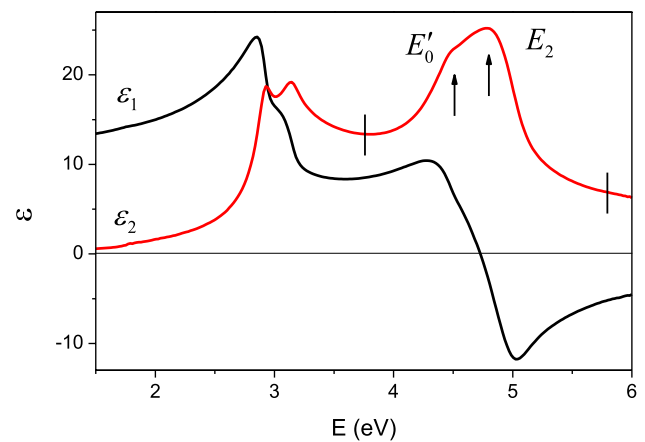
The Fourier coefficients of the segment of interest in Fig. 4 were obtained to double precision using MATLAB. These are displayed in Fig. 5 as  $\ln(C_n)$  before the second-derivative processing. The data exhibit interference, indicating the presence of multiple features with amplitudes decreasing approximately linearly on a log scale. The white-noise onset occurs near  $n \sim 60$ . To assess filtering the red and blue traces exhibit the results of GH and CME processing before multiplying by  $n^2$ . For SG processing the order was set at 5 because higher orders noticeably attenuate the structure. For the GH calculation filtering was done with  $M = 4$  and  $n_c = 63$  to best match the white-noise onset. Because the SG DS convolution already multiplies the RS data by  $n^2$ , its filtering equivalent is obtained by dividing the resulting  $C_n$  by  $n^2$ . Consistent with Fig. 1,



**Fig. 5** (Black) Values of  $\ln(C_n)$  of the highlighted data in Fig. 4. (Blue) same data after passing through a GH filter of order  $M = 4$  set with a cutoff index  $n_c = 63$ . (Red) same data after passing through a CME filter of order 47. (Green) same data after passing through the SG 5-point second-derivative convolution converted to a filter by multiplying the  $C_n$  by  $n^2$

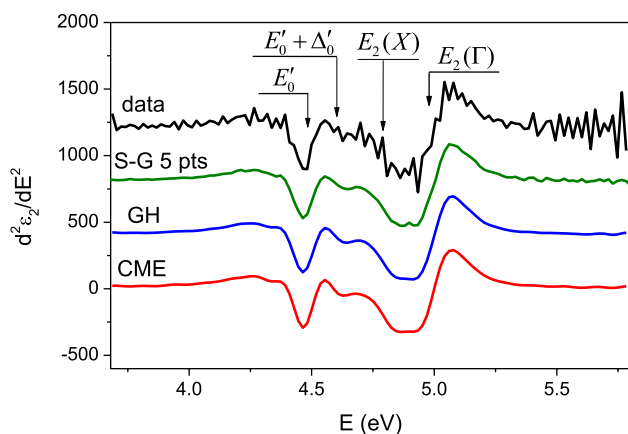
Fig. 5 shows that the SG filter is ineffective at suppressing noise except near its node at  $n = 150$ . The GH filter is significantly better but still allows a substantial amount of noise to leak through.

More striking is the CME replacement of the coefficients  $C_n$  in the white-noise region. Its extrapolation of the trend established in the low-index range is obviously appropriate. With 47 coefficients retained for the calculation, the extrapolation is solidly based. Because it is also based on analytic functions, the calculation of the second derivative generates no noise beyond the noise cutoff. Thus noise in the CME reconstruction is essentially non-existent. It can be noted that the CME shows unambiguously that the higher-energy peak is a doublet. As a second example, we consider archival room-temperature  $\varepsilon_2$  data for GaAs obtained by rotating-analyzer ellipsometry as 250 data points equally spaced in energy from 1.5 to 6.0 eV [18]. Integration time per point was 1 s. The data are shown in Fig. 6. The question here is whether the four  $E'_0$  and  $E_2$  features reported by Lautenschlager et al. [19] at a sample temperature of 22 K can be resolved at room temperature, given that these appear to collapse into two at 80 K. The second derivatives with respect to energy in the  $E'_0 - E_2$  spectral region, along with the second derivatives calculated after GH and CME processing, are shown in Fig. 7. The SG computation exhibits noticeable noise, but in this case there is little difference between the GH and CME results. Both the GH and CME calculations show the presence of the  $E'_0$ ,  $E'_0 + \Delta'_0$ ,  $E_2(X)$ , and  $E_2(\Gamma)$  critical points in the room-temperature data. The  $E_2(X)$  structure in the SG result is marginal. Thus progress has been made since the data were obtained.



**Fig. 6** Room-temperature data for GaAs was obtained spectroscopically as 250 points equally spaced in energy from 1.5 to 6.0 eV [18]. Our interest is in the  $E'_0 - E_2$  spectral range indicated by the arrows





**Fig. 7** Second energy derivatives of the  $\epsilon_2$  spectrum of Fig. 7 in the  $E'_0$ ,  $E_2$  spectral range for the data (black), after SG DS 5-point convolution (green), after GH filtering (blue), and after CME filtering (red). The locations of critical points giving rise to features in these spectra are indicated by arrows

## 4 Discussion

While linear filtering is well known, the CME is relatively new, and hence comments on its use are in order. The requirement that the original spectrum be positive definite follows from Eq. (12). The procedure is based on maximizing the entropy rate, which is expressed initially as a logarithm. For spectra that have negative excursions, for example, the real part of the dielectric function, this requires the use of an additive constant. It can be shown that in the lowest order the effect of modifying  $R_{00}$  is to change the rate of decrease of the projection of  $\ln(C_n)$  (the log of the broadening parameter) into the white-noise region, which for cosmetic purposes has little effect.

It can also be shown that Eq. (13) is a form of spectral representation, and hence the CME is most efficient when used with Lorentzian lineshapes. The characteristic of the spectral representation is its linear decrease of  $\ln(C_n)$  with  $n$ . This is inconsistent with Gaussian lineshapes, where the decrease of  $\ln(C_n)$  with  $n$  is quadratic. Nevertheless, model calculations show that when used for its intended purpose, neither the positive-definite nor linear-decrease characteristic appears to be significant.

## 5 Conclusion

In this work, we discuss different methods of minimizing noise in ellipsometric and other spectra by linear and non-linear methods. The field has advanced rapidly in the last

several years, and our objective is to provide an overview of current best practices, along with some examples of their use.

**Acknowledgements** This research was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIP) (NRF-2020R1A2C1009041, NRF2021R1A2C1005359, and NRF-2020R1A5A1016518).

## References

1. A. Savitzky, M.J.E. Golay, *Anal. Chem.* **36**, 1627 (1964)
2. J.F. Kaiser, W.A. Reed, *Rev. Sci. Instrum.* **48**, 1447 (1977)
3. P. Barak, *Anal. Chem.* **67**, 2758 (1995)
4. V.L. Le, T.J. Kim, Y.D. Kim, D.E. Aspnes, *J. Vac. Sci. Technol. B* **37**, 52903 (2019)
5. D.E. Aspnes, V.L. Le, Y.D. Kim, *J. Vac. Sci. Technol. B* **37**, 51205 (2019)
6. V.L. Le, T.J. Kim, Y.D. Kim, D.E. Aspnes, *J. Korean Phys. Soc.* **77**, 819 (2020)
7. L.V. Le, Y.D. Kim, D.E. Aspnes, *Opt. Express* **28**, 38917 (2020)
8. E. Hewitt, R.E. Hewitt, *Arch. Hist. Exact Sci.* **21**, 129 (1979)
9. L.V. Le, T.J. Kim, Y.D. Kim, D.E. Aspnes, *J. Appl. Phys.* **129**, 224902 (2021)
10. J.P. Burg, *Maximum Entropy Spectral Analysis* (Stanford University, 1975).
11. H.T. Nguyen, T.J. Kim, H.G. Park, V.L. Le, X.A. Nguyen, D. Koo, C.-H. Lee, D.D. Cuong, S.C. Hong, Y.D. Kim, *Appl. Surf. Sci.* **511**, 145503 (2020)
12. D.K. Hoffman, D.J. Kouri, E. Pollak, *Comput. Phys. Commun.* **147**, 759 (2002)
13. G.T. Walker, *Proc. R. Soc. London. Ser. A Contain. Pap. Math. Phys. Character* **131**, 518 (1931)
14. G.U. Yule, *Philos. Trans. R. Soc. London. Ser. A Contain. Pap. a Math. or Phys. Character* **226**, 267 (1927)
15. A. Kolmogorov, *Izv. Acad. Sci. CCCP, Ser. Math.* **05**, 3 (1941)
16. N. Wiener, *Rep. Serv. 19, Res. DIC-60-37 MIT* (1942)
17. V.L. Le, T.J. Kim, Y.D. Kim, D.E. Aspnes, *Curr. Appl. Phys.* **20**, 232 (2020)
18. D.E. Aspnes, A.A. Studna, *Phys. Rev. B* **27**, 985 (1983)
19. P. Lautenschlager, M. Garriga, S. Logothetidis, M. Cardona, *Phys. Rev. B* **35**, 9174 (1987)

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.