TECHNICAL RESEARCH REPORT

A Wavelet Approach to Wafer Temperature Measurement via Diffuse Reflectance Spectroscopy

by P.S. Krishnaprasad,
T. Kugarajah and
W.P. Dayawansa

T.R. 96-61

ISR
INSTITUTE FOR SYSTEMS RESEARCH

Sponsored by
the National Science Foundation
Engineering Research Center Program,
the University of Maryland,
Harvard University,
and Industry

A Wavelet Approach to Wafer Temperature Measurement via Diffuse Reflectance Spectroscopy

P.S. Krishnaprasad, T. Kugarajah and W.P. Dayawansa
Institute for Systems Research and Department of Electrical Engg
University of Maryland, College Park, MD 20742

May 7, 1996; This version: June 18, 1996

Abstract

A methodology for the determination of wafer temperature in Molecular Beam Epitaxy via diffuse reflectance measurements is developed. Approximate physical principles are not used, instead, patterns in the data (reflectance versus wavelength) are exploited via wavelet decomposition and Principal Component Analysis.

1 Introduction

Progress made in optical techniques for *in situ* monitoring of semiconductor film properties (thickness, temperature etc.) in Molecular Beam Epitaxy (MBE) has made Diffuse Reflectance Sensors (DRS) an attractive possibility in the semiconductor industry. A brief explanation of the principle of DRS follows. During measurement, white light is focused onto the semiconductor substrate, which is polished on the front surface and textured on the back. The measurement is such that only the light diffusely reflected from the back surface is collected. The diffuse reflectance as a function of the wavelength is measured by a spectrometer. Since at short wavelengths light does not penetrate the front surface, the onset of transparency corresponds to a sharp increase in the diffuse reflectance at a particular (knee) wavelength (λ_{knee}). Though such a knee wavelength is not always clear-cut, recent research has focused on using physical principles to relate the temperature to band gap, film thickness, etc., (their effect on the onset of transparency and hence the knee wavelength), and the temperature is obtained approximately as a function of knee wavelength and substrate thickness. Johnson et al [8] obtain a function that is quadratic in λ_{knee} and linear in substrate thickness. However the observation that the reflectance measurements are made available over an entire range of wavelengths motivates us to explore the possibility of finding the temperature solely from analysis of patterns contained in the data without relying on approximate physical principles.

In order to bring out such patterns, we use 2-dimensional (2-D) spectra defined on an appropriate phase plane, that we construct from a wavelet analysis of the DRS spectra. Then such 2-D spectra (which we call *DRS faces*) are further analysed with a view to finding important features or patterns in them, localized in the phase plane. Our approach to finding such patterns uses what is widely known in the signal processing context as Karhunen-Loeve decomposition and in the statistics literature as Principal Component Analysis (PCA). We show that, when applied to our data-set of DRS faces, PCA reveals a small number of characteristic patterns (or *eigen*-faces), and every DRS face in the data set is representable as a linear combination of these patterns. The associated coefficients, known as *modal coefficients* or principal components, constitute a compact code for a DRS face (equivalently the DRS spectrum). Then the problem reduces to finding a mapping that connects the modal coefficients to the temperature. There are various ways to do this and, in section 5 below, we demonstrate the use of a linear regression model.

2 Wavelet Analysis

Wavelet theory has emerged as an elegant tool in the analysis of data with interesting time-frequency concentrations/localizations. In essence, one seeks to decompose a given function via special bases (or their generalized form called frames), which show such localizations clearly. Such special bases include orthogonal or bi-orthogonal wavelets and wavelet packets [1]. One aspect of this theory leads to a type of coarse-to-fine scale decomposition known as Multi-Resolution

Analysis (MRA) [3]. In the one dimensional case, MRA leads to an efficient decomposition via orthonormal basis elements, and a fast algorithm for evaluating the resulting coefficients. We give a brief exposition of MRA; more details can be found in [3].

2.1 MRA

Given a function $\psi \in L^2(\mathbf{R})$, the space of square-integrable functions on the real line, the translated and dilated family $\{\psi_{j,k} = 2^{j/2}\psi(2^j x - k), j, k \in \mathbf{Z}\}$, under certain conditions, forms an orthonormal basis for $L^2(\mathbf{R})$. Consider the closure of the span $\{\psi_{j,k}, k \in \mathbf{Z}\}$ for each $j \in \mathbf{Z}$, and notice that these subspaces denoted W_j have the property $\overline{\bigcup_{j\in\mathbf{Z}} W_j} = L^2(\mathbf{R})$. With increasing J, $\overline{\bigcup_{-\infty}^{J} W_j}$ forms a nested sequence of subspaces, and such a sequence is called multi-resolution analysis. One may define functions whose translations span such subspaces. Such functions are known as *scaling functions*, and are essential to MRA.

Definition: A function ϕ is a scaling function if

1. The translated and dilated family $\{\phi_{j,k} | \phi_{j,k}(x) = 2^{j/2}\phi(2^j x - k), j, k, \in \mathbf{Z}\}$ generates a sequence of sub-spaces of the form $\cdots V_{-1} \subset V_0 \subset V_1 \cdots$.
2. $\overline{\bigcup_{j\in\mathbf{Z}} V_j} = L^2(\mathbf{R})$.
3. $\{\phi(. - k) : k \in \mathbf{Z}\}$ is an orthonormal basis (or at least a Riesz basis) for V_0.
4. $f(x) \in V_0 \Leftrightarrow f(2^j x) \in V_j$.

The above properties also imply that

1. $\bigcap_{j\in\mathbf{Z}} V_j = \{0\}$.
2. $\phi(x) = \sqrt{2}\sum_{n\in\mathbf{Z}} h_n \phi(2x - n)$ since $\phi \in V_0 \subset V_1$. This can be written as $\widehat{\phi(\omega)} = H\widehat{(\omega/2)}\phi\widehat{(\omega/2)}$ where $H\widehat{(\omega)} = \sqrt{2}\sum_{n\in\mathbf{Z}} h_n e^{-i\omega n}$.
3. one can introduce W_j as $V_{j+1} = V_j \oplus W_j; j \in \mathbf{Z}$.

Using the above properties, one can choose suitable scaling functions ϕ and find corresponding orthonormal wavelets ψ. Moreover a fast algorithm for calculating the wavelet coefficients can be derived as follows. Let $f \in L^2(\mathbf{R})$ and $c_{j,k} =< f, \phi_{j,k} >$, and $d_{j,k} =< f, \psi_{j,k} >$. Then it can be shown that

$$c_{j-1,k} = \sum_n \overline{h_{n-2k}}\, c_{j,n}$$

and

$$d_{j-1,k} = \sum_n \overline{g_{n-2k}}\, c_{j,n}$$

where $g_n = (-1)^n h_{-n+1}$.

Hence we can think of the decomposition as a filtering scheme with filters H, G (corresponding to the g_n) and the algorithm is illustrated in Figure 1. The symbol $2 \downarrow$ denotes that every other sample is discarded, also known as *downsampling* or *decimation.* The sampled version of the function is an N-element (N is assumed to be a power of 2) vector and corresponds to the dilation level $j = 0$. When the algorithm is applied to this vector (i.e., $c_0 = f$) once, one gets a coarser level ($j = -1$) information in the coefficients $(d_{-1,k})_{k=1}^{N/2}$, and L successive applications of the algorithm give information on all coarser levels $j = -1, \cdots, -L$, in the coefficients $(d_{j,k})_{k=1}^{\frac{N}{2^j}}$. For the current discussion, it suffices to note that when $-L$ is the coarest level of resolution selected, f can be written as

$$f = \sum_{j=-L}^{-1} \sum_k d_{j,k}\psi_{j,k} + \sum_k c_{-L,k}\phi_{-L,k}.$$

The second term above represents the residual obtained when only levels finer than $-L$ are analysed. Thus one can see that given an N-element vector for the sampled function, the coarsest level that can be analysed is constrained by $\log_2(N)$.

The orthonormal bases resulting from the above analysis (starting from the scaling function) often have the drawback of having infinite support, which makes their computer implementation difficult; however, starting with the filtering scheme rather than scaling functions, Daubechies [3] has derived several compactly supported orthonormal bases, and in this paper, we use what is known widely as the (Daubechies) D4 orthonormal wavelet.

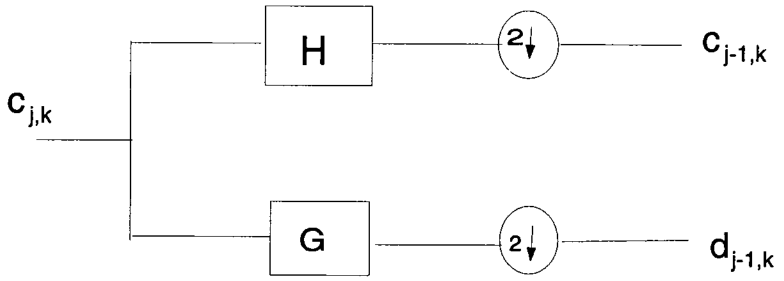

Figure 1: The fast algorithm

3 Feature Extraction via PCA

Principal component analysis is a standard method used in diverse applications for dimensionality reduction [6, 9, 10]. The idea is to find a transformation of the data space into a feature space, which is of the same dimension as the data space but is such that a small number of these characteristic features are adequate to capture the information in the patterns. These characteristic (eigen-) features are then used in pattern classification problems.

Let the N-dimensional vectors, $\{\theta_1 \cdots \theta_M\}$ be the set of 2-D spectra (patterns) obtained via wavelet analysis of the DRS data-set available for M temperatures. The average pattern is formed as $\tilde{\phi} = \frac{1}{M}\sum_{k=1}^{M}\theta_k$. Deviations from this average pattern will be used efficiently in characterising the patterns. We shall deal with the ensemble $\phi_k = \theta_k - \tilde{\phi}, k = 1, \cdots, M$, and we will denote by ϕ the random variable corresponding to these features .

Let $\{u_i\}_1^M$ be an orthonormal basis for the set of patterns.

Then we can write

$$\phi_k = \sum_{i=1}^{M} a_i^k u_i \tag{1}$$

where $\{a_i^k\}_{i=1}^M$ are the principal components or modal coefficients, given by $a_i^k =< u_i, \phi_k >$.
The discrete Karhunen-Loeve expansion theorem states that the vectors $\{u_i\}_1^M$ that optimally represent the patterns are the eigenvectors of the covariance matrix $R = E[\phi\phi^T]$. (Here and later, the superscript

T is used to denote transpose of a vector or a matrix.) Proof of this theorem in different forms can be found in [5, 6, 9]. We shall therefore refer to members of this orthonormal basis set as "eigen-patterns" or eigen-faces. Since we have only a set of M elements, the co-variance matrix R is approximated by

$$C = \frac{1}{M} \sum_{k=1}^{M} \phi_k \phi_k^T. \tag{2}$$

Hence we can find the orthonormal basis by solving for u_k in

$$C u_k = \lambda_k u_k, \quad k = 1, \cdots, M. \tag{3}$$

Since C is an $N \times N$ matrix, directly solving the above problem is not advisable. Instead observe that equations (1), (2) ,(3) combined with the orthonormality of $\{u_k\}_1^M$ imply that the eigen-patterns are admixtures of patterns in the ensemble; precisely,

$$\begin{aligned} C u_k &= \frac{1}{M} \sum_{i=1}^{M} \phi_i {\phi_i}^T u_k \\ &= \frac{1}{M} \sum_{i=1}^{M} \phi_i \{\sum_{j=1}^{M} a_j^i u_j\}^T u_k \\ &= \frac{1}{M} \sum_{i=1}^{M} a_i^k \phi_i \end{aligned} \tag{4}$$

and hence

$$\begin{aligned} u_k &= \frac{1}{M\lambda_k} \sum_{i=1}^{M} a_i^k \phi_i \\ &= \sum_{i=1}^{M} b_i^k \phi_i. \end{aligned} \tag{5}$$

Hence upon substituting in equation (3), we have a simplified (i.e., when $M << N$) problem, whereby the coefficents b_i^k are obtained as components of the eigenvectors of an $M \times M$ matrix L as shown below:

$$\frac{1}{M} \sum_{i=1}^{M} \phi_i \phi_i^T \sum_{j=1}^{M} b_j^k \phi_j \quad = \quad \lambda_k \sum_{m=1}^{M} b_m^k \phi_m$$

$$\sum_{i=1}^{M} \phi_i (\sum_{j=1}^{M} \phi_i^T \phi_j b_j^k) = \lambda_k \sum_{m=1}^{M} b_m^k \phi_m \tag{6}$$

Equation (6) implies that by letting $L = [L_{ij}]$ with $L_{ij} = \frac{1}{M} < \phi_i, \phi_j >$, and $b^k = [b_j^k]^T$, the problem is essentially solving for b^k and λ_k in the eigenvalue problem:

$$Lb^k = \lambda_k b^k. \tag{7}$$

4 Data Analysis

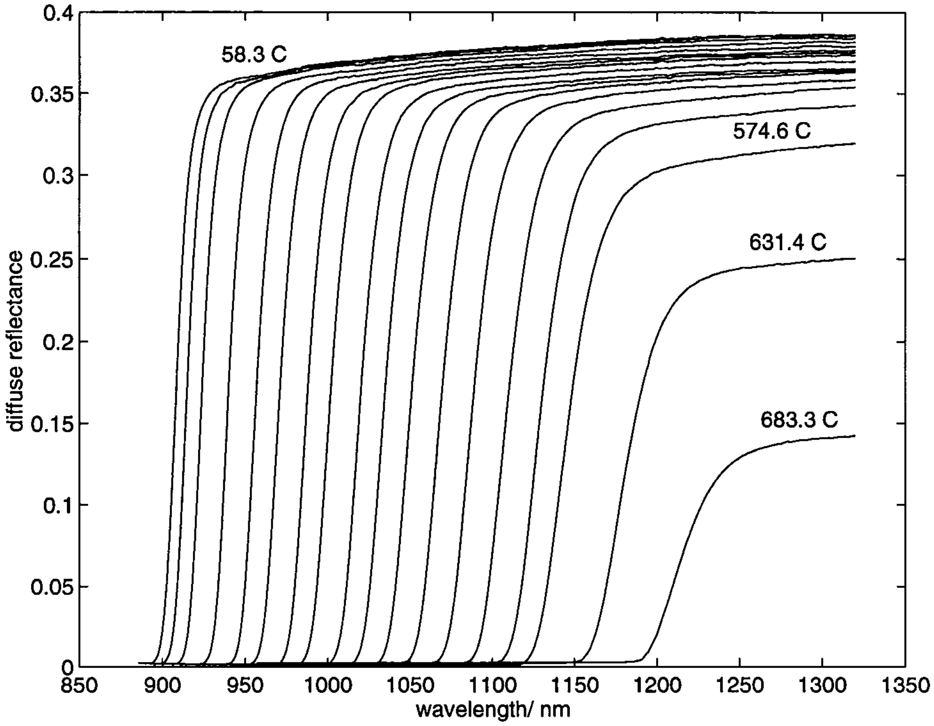

Figure 2: The data (from S.R. Johnson)

We received [7] a set of data from Shane Johnson (then of the University of British Columbia) representing measurements of diffuse reflectance of a wafer for eighteen wafer temperatures. The plot of data is shown in Figure 2. For this data-set, we used the fast wavelet algorithm available in WaveLab [2] and obtained a set of DRS faces alluded to in the introduction. Each such DRS face is simply a plot of the wavelet coefficient $d_{j,k} =< f, \psi_{j,k} >$ versus j, k, where f denotes the measured DRS spectrum (function of wavelength) and $\psi_{j,k}$ arise from the chosen multi-resolution. A sample DRS face is shown in Figure 3. Observe that $j = 0$ corresponds to the finest scale at the sampling interval of $0.5nm$ wavelength. Further, reversing the notation used in section 2, increasing values of j correspond to coarser features in the face. The procedure outlined in section 3 was carried out for the spectra with $M = 18, N = 1024, L = 10$. Examining the normalized eigenvalues in Figure 4, we see that only four or five of the principal components are necessary to represent the patterns adequately, thus achieving significant dimensionality reduction.

5 Modeling for Temperature

Utilizing the small number of principal components as input, our task is to find a mapping that gives temperature. Once the mapping is established, for any new measurement, one would compute the DRS face and then the principal components; these components would then be plugged into the mapping to find the temperature.(*This is the essence of extrapolation of temperature from a fresh DRS spectrum.*) We attempted linear regression of the form $T = W^t a + b$ where $W = (W_1, \cdots, W_l)^t$ is the weight vector and $a = (a_1, \cdots, a_l)^t$ is the principal component vector. (Here the superscript t is used to denote transpose of a vector.) Using $l = 5$, a fit was obtained as shown in Figure 5, which gives error upto 7 degrees C (Figure 6). The resulting weight values are $W_1 = 2.9590, W_2 = 3.2220, W_3 = -2.5669, W_4 = -1.8820, W_5 = 4.7129$ and $b = 348.3999$. The resulting model fit error is significant, but our experience with fitting procedures indicates that this is a result of being constrained by the limited number of observations available to us. With more data, better fit may be obtained for the linear regression, or better models can be constructed via methods such as splines or neural networks. We await more data in order to use the proposed method toward the goal of achieving an error less than 1 deg C.

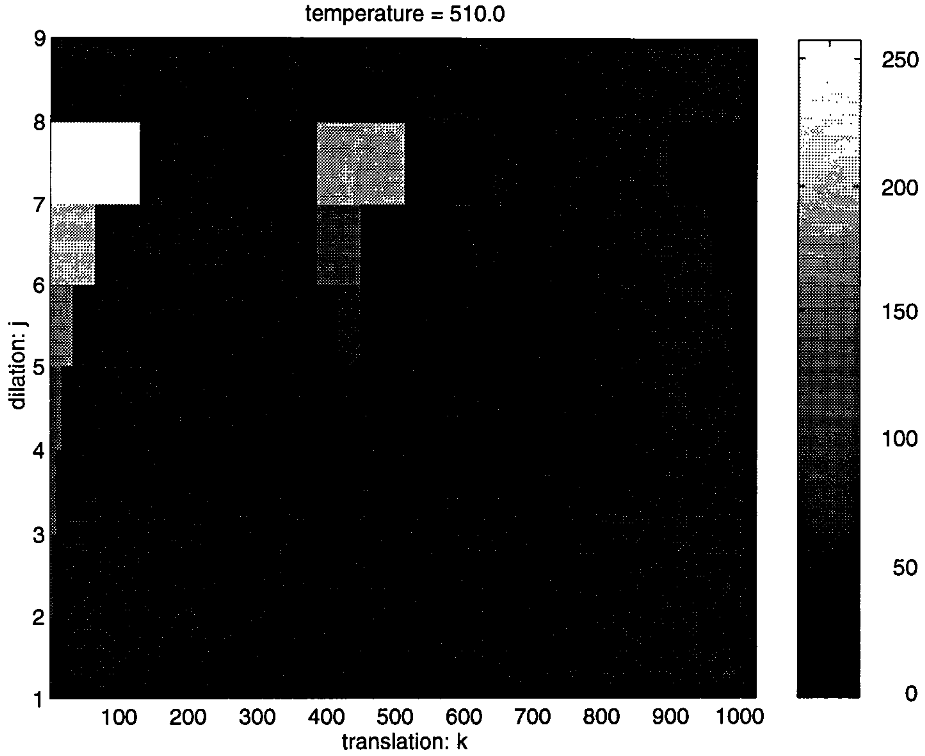

Figure 3: Sample DRS face corresponding to 510 Celsius.

6 Conclusions

Additional experimental data is needed to construct a more accurate mapping of the temperature T from modal coefficients. Our results suggest that affine mappings (as in section 5) may do an adequate job, and as such can lead to an effective, automatic, fast temperature extrapolation procedure. Compared to existing methods, we expect that our approach will require more data for fitting an accurate model, but this is done off-line only once, and does not seem to pose any real disadvantage. There exist fast methods for performing both PCA and final model fitting using regressions or neural networks. Moreover, our method can be extended to other measurements of complex phenomena for which physical models are hard to obtain otherwise. For instance, in many chemical processes, such complex phenomena often require interpretation of patterns by human experts even when visu-

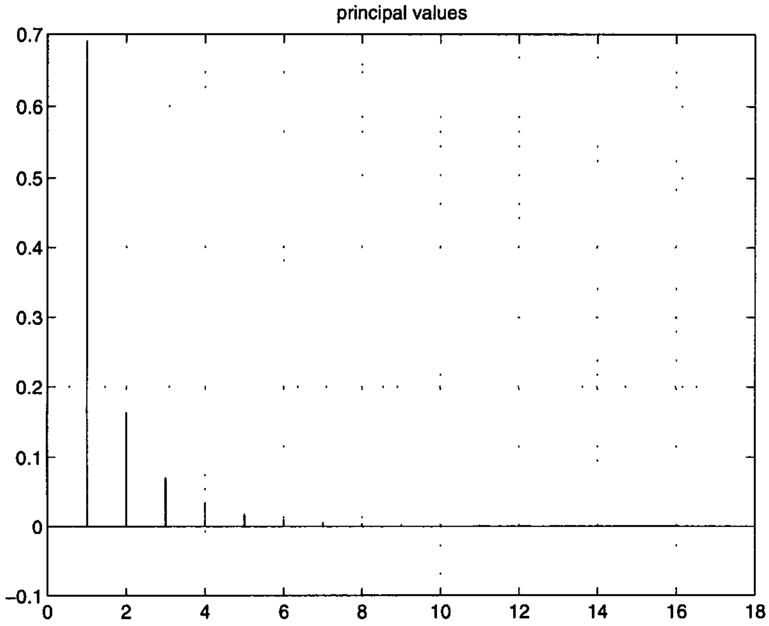

Figure 4: The principal values: normalization is such that the sum of these values is one

alization tools such as wavelet packets are used [4]. We believe our methods could be extended to such processes and render the active intervention of users in carrying out interpretation unnecessary. We hope to continue to explore this further.

7 Acknowledgements

This work was stimulated by interactions that one of us (P.S.K) had during the Engineering Foundation Conference on Monitoring and Control Techniques for Intelligent Epitaxy held June 11-16, 1995 in Banff, Canada. In particular, the difficulties in obtaining a satisfactory determination of the knee wavelength led us to consider the alternative approach presented here. Shane Johnson kindly provided us the DRS spectral data. This research was supported in part by a grant from the National Science Foundation's Engineering Research Centers Program: NSFD CDR 8803012, and by the NSF Grant EEC-9527576.

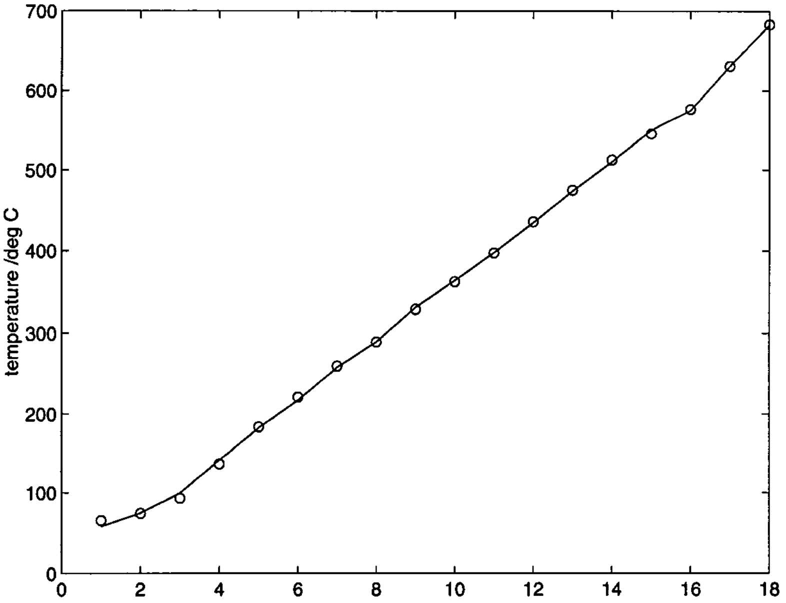

Figure 5: Modelled temperatures (ooo) and actual temperatures (—)

References

[1] C. K. Chui. *An Introduction to Wavelets.* Academic Press, 1992.

[2] D. Donoho et.al. The WaveLab Software, ftp.playfair.stanford.edu, 1995.

[3] I. Daubechies. *Ten Lectures on Wavelets.* CBMS-NSF regional conference series in applied mathematics. SIAM, 1992.

[4] Xue dong Dai, B. Joseph, and R. L. Motard. Process signal feature analysis. In R. L. Motard and B. Joseph, eds., *Wavelets in Chemical Engineering*, pages 115 –137. Kluwer Academic Publishers, Boston, 1994.

[5] K. Fukunaga. *Introduction to Statistical Pattern Recognition.* Computer Science and Scientific Computing, Academic Press, second edition, 1990.

[6] S. Haykin. *Neural Networks: A Comprehensive Foundation.* MacMillan College Publishing, 1994.

[7] S. R. Johnson, June 1995. Private Communication.

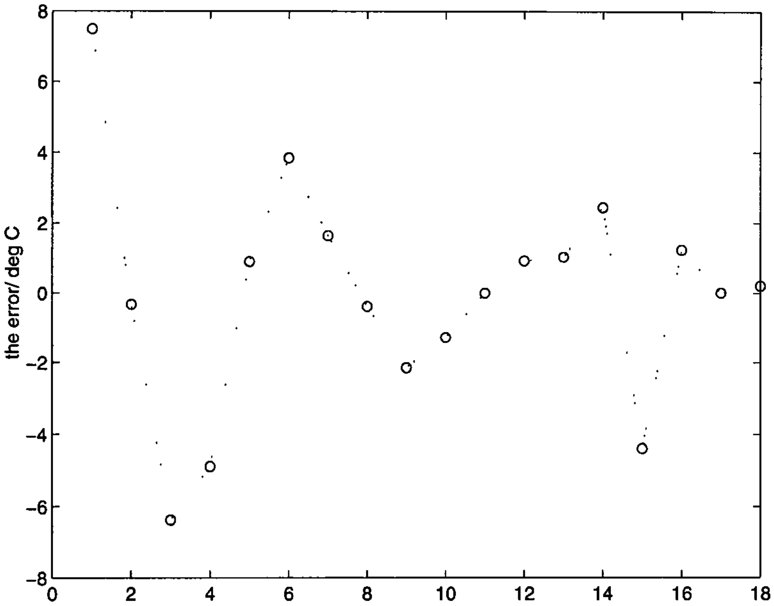

Figure 6: Model fit error versus data-set index

[8] S. R. Johnson, C. Lavoie, T.Tiedje, and J. A. Mackenzie. Semiconductor substrate temperature measurement by diffuse reflectance spectroscopy in molecular beam epitaxy. *J. Vac. Sci. Technol. B*, 11(3):1007 -1010, May 1993.

[9] A. J. Newman. Model reduction via the Karhunen-Loeve expansion, part 1: An exposition. Technical Report T.R.96-32, Institute for Systems Research, University of Maryland, 1996.

[10] L. Sirovich and M. Kirby. Low-dimensional procedure for the characterization of human faces. *J. Opt. Soc. Am. A*, 4(3):519 -524, March 1987.