PCA- Wavelet Coefficients for T^2 chart to detect endpoint in CMP process

Sihem Ben Zakour[1] and Hassen Taleb[2]

[1]Departement of Quantitative methods, Higher institute of Management of Tunis, Tunisia
SihemBenZakour@yahoo.com
[2] Departement of Quantitative methods, Higher institute of business administration of gafsa, Tunisia
hassen.taleb@isg.rnu.tn

ABSTRACT

The development of the semiconductor industry, with advances in sensors oblige us to deal with large datasets do not stop increasing, while monitoring devices are becoming more and more complexes and sophisticated. As the measurement points become closer. Among the complex monitoring process, the detection of the end of polishing (EPD) during the chemical mechanical planarization (CMP) process is considered as a critical task in semiconductor manufacturing. In this paper, a sequence of an Acoustical emission (AE) waveform signals are collected during the progression of the CMP process will be monitored using PCA- Wavelet daubechie Coefficients based on T^2 chart. In order to detect the endpoint, we should not only to remove the noise from the obtained acoustical waveform signal, but also reduce dimensionality of monitored coefficient, by employing discrete wavelet algorithm for cleaning signals , Principal component analysis (PCA) for reducing dimensionality . Also, a comparative study is presented to show the out-performance of PCA-Wavelet Hotelling chart in detecting the endpoint.

Keywords

Hotelling chart (T^2 chart) , End point detection (EPD), Chemical mechanical planarization (CMP), Acoustic emission (AE), Wavelet Analysis (WA), Principal component analysis (PCA), Digital signal processing, monitoring process.

1. INTRODUCTION

The semiconductor wafer, composed by a tranche of silicon on which one comes to deposit successive layers of metallization, has been a subject of interest these last years. Indeed, the reduction of its size, the introduction of new materials like the copper and low-k dielectrics and the superposition of different metals create an enormous challenge in terms of accomplishment of wafer production and checking its quality. The CMP process, given by figure1, combines mechanical and chemical properties. The chemical agents react to the surface of wafer and transform it. The abrasive particles come to remove this layer of transformed matter. The CMP process enables us: 1) To maximize lithographic performance. 2) To improve the cleaning ability. 3) To reduce yield limiting defects from other processes. 4) To employ difficult -etch metals like Cu. During the polishing, it seems necessary, even crucial to determine the right moment of stopping this process which remain to have a planar wafer surface with the desired pre-established thickness, as shown in figure 2 and 3.

DOI : 10.5121/ijcsit.2012.4402

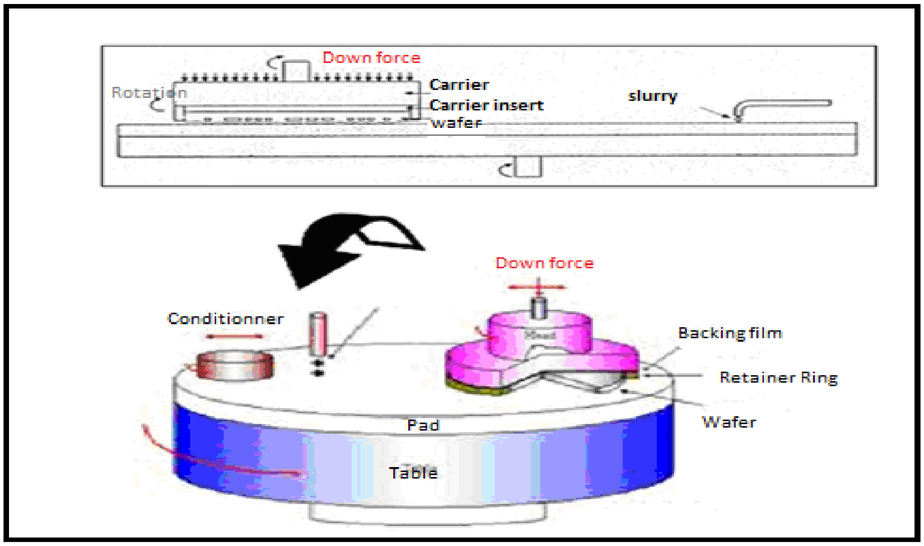

Figure 1. Chemical mechanical planarization (CMP) process.

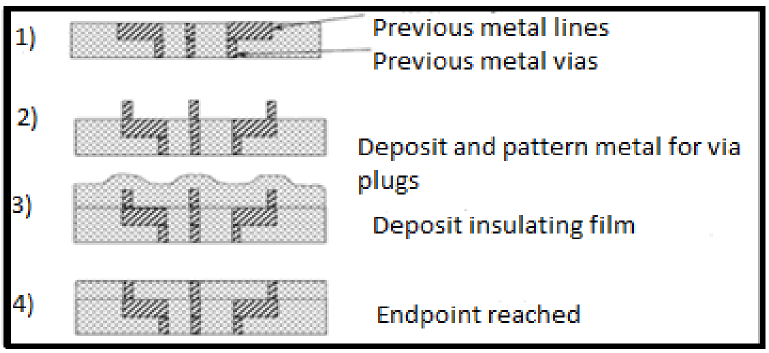

Figure 2. Endpoint reached in 4-rth step of metal deposition and planarization process.

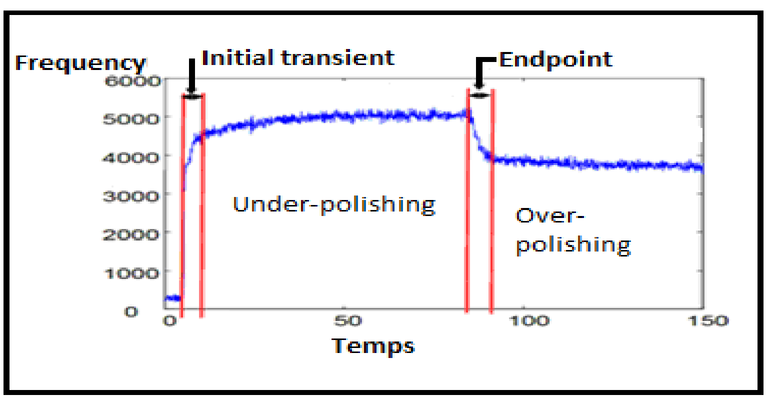

Figure 3. Plot of an endpoint trace showing Delineations of the CMP process including: initial transient, main CMP (starting of polishing and the wafer is under-polished if we stop the CMP process), endpoint, and over polishing (over-planarization).

2. PRINCIPAL COMPONENT ANALYSIS (DATA REDUCTION TECHNIQUE)

The most famous used Multivariate statistical analysis and data reduction is principal component analysis (PCA). PCA is used for data compression and information extraction. The idea behind, is to reduce the dimensionality of the original data by forming a new set of latent variables which are a linear combination of the original data, without losing essential information. PCA explains the amount of variability in the data. The first PC explains most of the variance in the data, and each subsequent one accounts for the largest proportion of variability that has not been accounted for by its predecessors [1, 2, and 3]. A rotational algorithm such as 'varimax' rotation is frequently calculated to obtain the rotated factor loadings that represent the contribution of each variable to a specific PC in clean and obvious way. The Multivariate endpoint data can be expressed in the form of a 2-D matrix representation:

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1j} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2j} & \cdots & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{i1} & x_{i2} & \cdots & x_{ij} & \cdots & x_{in} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mj} & \cdots & x_{mn} \end{bmatrix}$$

With **X** is a matrix of intensity data containing *m* time samples and *n* acoustical wavelength channels. Each sample *xij* corresponds to the signal intensity for the *i*-th time sample and the *j*-th wavelength channel. The basis for PCA modeling comes from the decomposition of the **X** matrix as follow:

$$X=TV^T= (XV)*V^T$$
$$V= [v_1...v_n]$$

With, **V** is an orthonormal set of vectors and **T** is the projection of **X** onto **V**. The $\mathbf{t}_i$ values, are called the Principal component scores, and the $\mathbf{v}_i$ values represent the principal component loadings. The scores are related to the samples, which means that, the first principal component ($\mathbf{t}_1$) scores is a vector of values for each sample point, see figure 4. The loadings are related to the variables, e.g. the first principal component loadings ($\mathbf{v}_1$) are a vector of weightings for each variable. It should be noted that, Often, X can be decomposed into a few number (1-5) of principal components that are able to capture most of the variation in **X**, even for large numbers of variables (>1000).

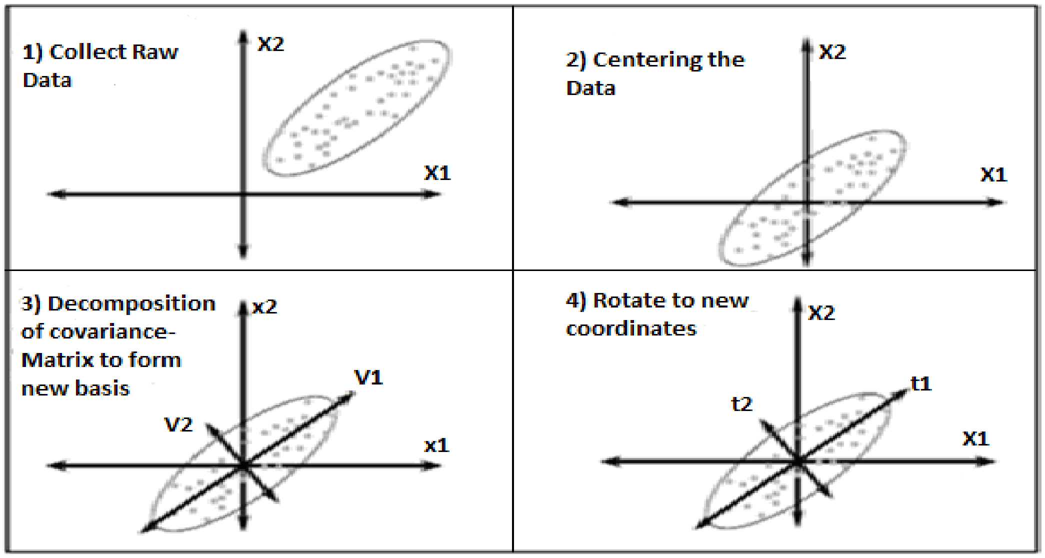

Figure 4. PCA for 2-D case (x_1 and x_2). The mean-centered data, X, are rotated to a new set of coordinates {v1, v2} that are statistically independent for each other. The projections onto these directions are called scores and are noted as t1 and t2.

3. WAVELET ANALYSIS (DENOISING PROCEDURE)

Many papers have treated and employed wavelet analysis in many fields [4, 5 and 6]. The wavelet analysis uses linear combinations of basic functions (wavelets), localized both in time and frequency, to represent any function in the $L^2(R)$ space. It provides a tool for time-frequency localization.

$$f(t) = \sum_{-\infty}^{\infty} \sum_{-\infty}^{\infty} b_{j,k} w_{j,k}(t)$$

Where j is the dilation (or scale) index and k is translation index. $W_{j,k}$ denotes a collection of basic functions, and $b_{j,k}$ designates the coefficients of these functions. Also the wavelet basis functions can be derived from the dilation and translation of scaling functions (ϕ) that span the subspace $L^2(R)$. By the combination of the scaling and the wavelet functions, any class of signal will be represented as follows:

$$f(t) = \sum_{k=-\infty}^{\infty} c_{j_0,k}\,\phi(t-k) + \sum_{k=-\infty}^{\infty} \sum_{j=j_0}^{\infty} d_{j,k}\, w(2^j t - k)$$

Where $c_{j_0,k}$ and $d_{j,k}$ are coefficients for the scaling and wavelet functions, respectively. They are also called the discrete wavelet transform (DWT) of the function and it is customary to start with $j_0=0$. If the wavelet system is orthogonal, then the coefficients can be calculated by:

$$c_{j_0,k} = \left\langle f(t), \phi_{j_0,k}(t) \right\rangle = \int f(t)\phi_{j_0,k}(t)dt$$

$$d_{j,k} = \left\langle f(t), w_{j,k}(t) \right\rangle = \int f(t) w_{j,k}(t)dt$$

The de-noising objective is to suppress the noise part of the signal S and to recover f.

$$S = f + noise$$

3.1 Decomposition

There are different types of wavelet, in figure 5, and the differences between different mother wavelet functions (e.g. Haar, Daubechies, Coiflets, Symlet and etc.) show the manner of these scaling signals and the wavelets are defined. One must select the appropriate wavelet is selected according to its associated properties on which we want to focus on [7]. Hence, making the choice of a wavelet could be made according to these type classifications continuous versus discrete wavelets, orthogonal versus redundant decompositions. Briefly, the continuous wavelets present a disadvantage concerning redundancy in decomposition, but they are more robust to noise compared with other decomposition schemes. Discrete wavelets have fast implementation but generally, the number of scales and the time invariant property depends closely on the data length.

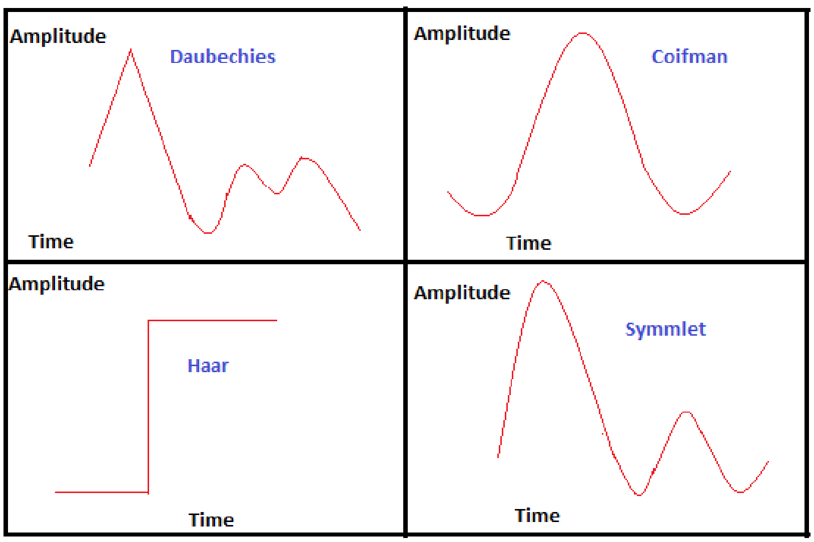

Figure 5. The most famous wavelet types

There are several methods for choosing the higher level of decomposition [4] , among them, after decomposing the signal of length (n) to the maximum allowed level of decomposition j=n/2, we can note

1) Apply threshold value and observing the significant coefficients
2) Calculate the Energy function and choose the level after which there is a significant drop in the energy content
3) Making test to the noise at each level of decomposition and stopping the test when the noise is removed
4) Define the de-noising level and use an iterative scheme in which the relative difference of the energy is consider as a stopping criterion.

Finally, Calculating the coefficients $c_{j_0,k}$ and $d_{j,k}$ from the scaling and wavelet functions.

3.2 Thresholding

The second step of de-noising procedure is thresholding in which the thresholded value is calculated using Visushrink method (or Donohos universal threshold rule proposed by Donoho et al.[5, 6]and is given as follows:

$$t_j = \sigma_j \sqrt{2\log(n)}$$

Where n: the length of signal and σ_j is the standard deviation of the noise at scale j.

$$\sigma_j = \frac{1}{0.675} median(|d_{j,k}|)$$

Where $d_{j,k}$ are the wavelet coefficients. The σ_j is the standard deviation *(j)* estimated from the median of the absolute deviation of the wavelet coefficients at the same scale *j*. Only the significant wavelet coefficients situated outside of the threshold limits are extracted by applying soft or hard thresholding.

The thresholded coefficients are given as bellow:

1) Hard thresholding:

$$\tilde{d}_{j,k} = \begin{cases} 0 if \ |d_{j,k}| < t \\ d_{j,k} if \ |d_{j,k}| \geq t \end{cases}$$

2) Soft thresholding:

$$\tilde{d}_{j,k} = \begin{cases} 0 if \ |d_{j,k}| < t \\ sign(d_{j,k})(|d_{j,k}| - 1) if \ |d_{j,k}| \geq t \end{cases}$$

Where sign ($d_{j,k}$) is the positive or negative sign of the wavelet coefficient $d_{j,k}$. For more details about soft and hard thresholding read [8, 9, 10, and 11].

3.3 Reconstruction

The last step of wavelet analysis is reconstruction. The signal *f*t is reconstructed from the thresholded wavelet coefficients through the use of inverse wavelet transforms. After the determination of the thresholded details and approximations at level J, they will be used as inputs to calculate the coefficients at level *(j-1)* until getting the original signal with the noise eliminated. The reconstructed signal at the level j is as follows:

$$f_j(t) = \sum_{k=-\infty}^{\infty} \tilde{d}_{j,k} w(2^j t - k)$$

4. HOTELLING CHART

The Hotelling chart was proposed firstly by Harold Hotelling in 1947, [12]. The Hotelling T^2, represent a scalar that combines information from the dispersion and mean of several variables, and measure the covariance structure of a multivariate normal distribution. The value T^2 is obtained by multiplying the following three quantities: 1) The vector of deviations between the observations and the mean **m,** which is expressed by **(X-m)'**, 2)The inverse of the covariance matrix, $\mathbf{S}^{-1}$, 3)The vector of deviations, **(X-m**).

It should be mentioned that for independent variables, the covariance matrix is a diagonal matrix and T^2 becomes proportional to the sum of squared standardized variables. In general, the higher the T^2 value, the more distant is the observation from the mean. The formula for computing the T^2 is:

$$T^2=c(X-\mathbf{m})S^{-1}(X-\mathbf{m})$$

\The constant c is the sample size from which the covariance matrix was estimated. For $\mathbf{X}_1,\ldots\mathbf{X}_n$ be n* p-dimensional vectors of observations are sampled independently and follow $N_p(\mathbf{m}, \Sigma)$ with $p < n$-1, with Σ the covariance matrix of **X**. The observed mean vector $\bar{\mathbf{X}}$and the sample dispersion matrix

$$\mathbf{S} = \frac{1}{n-1}\sum_{i=1}^{n}(\mathbf{X_i} - \bar{\mathbf{X}})(\mathbf{X_i} - \bar{\mathbf{X}})'$$

are the unbiased estimators of **m** and Σ, respectively.

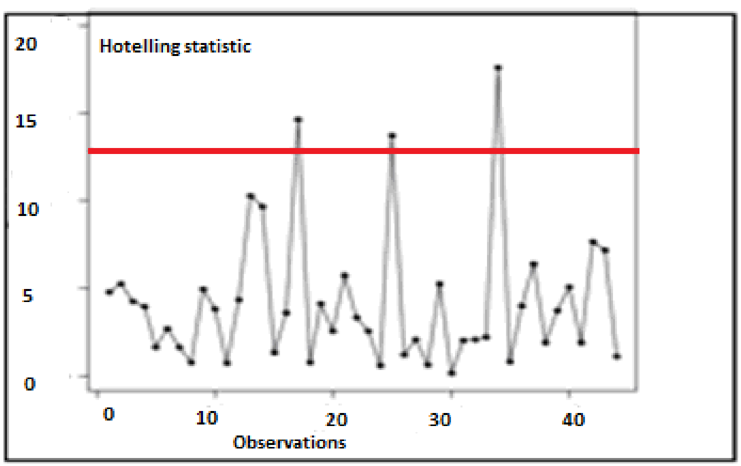

Figure 6. Example of Hotelling T^2 individual chart

In figure 6, Hotelling values are plotted against control limits given as follow:

$$UCL= [(m-1)(m+1) / m(m-n)]* F(1-\alpha, n, m-n)$$

$$LCL=0$$

With m represents the number of samples from which the mean and covariance is calculated, n exhibits the number of variables, and UCL and LCL refer to the upper and lower control limits, respectively.

5. EXPERIMENTAL SETUP AND RESULTS:

The test bad should be equipped with an AE sensor in order to collect the data during CMP. We have generated multi-normal normal distribution with noise; of sequence signals (450 times) are non-stationary since the mean value change during the progression of CMP process (from 7.10^{-6} to $5.10^{-6)}$. The planarization of the wafer is done under a specific combination of rotational speed (rpm) and downward pressure (psi). In this paper, the studied data are simulated. A sample of these data, plotted in figure 7, shows the presence of noise and huge number a sample channel per signal. And figure 8, shows the non-stationary of coefficient 798 (taken arbitrary).

Therefore, one should not only eliminate the noise and decrease the number of data per signature (waveform signal), to have a robust statistical tool used in monitoring the CMP process.

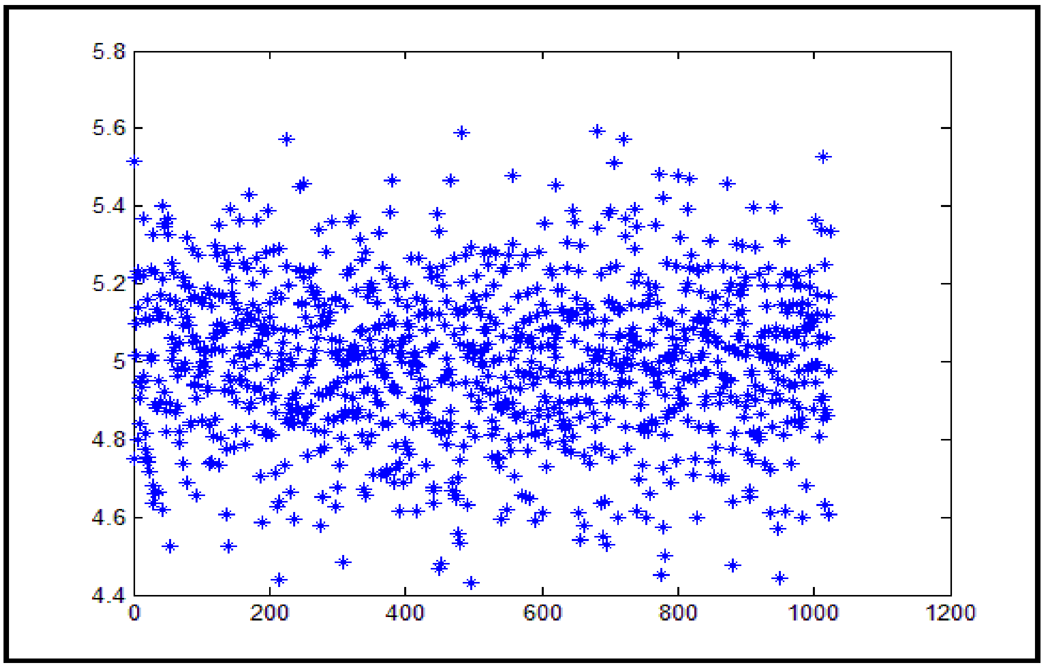

Figure 7. Signature obtained from acoustical sensor at t=150s, containing 1024 coefficients (points).

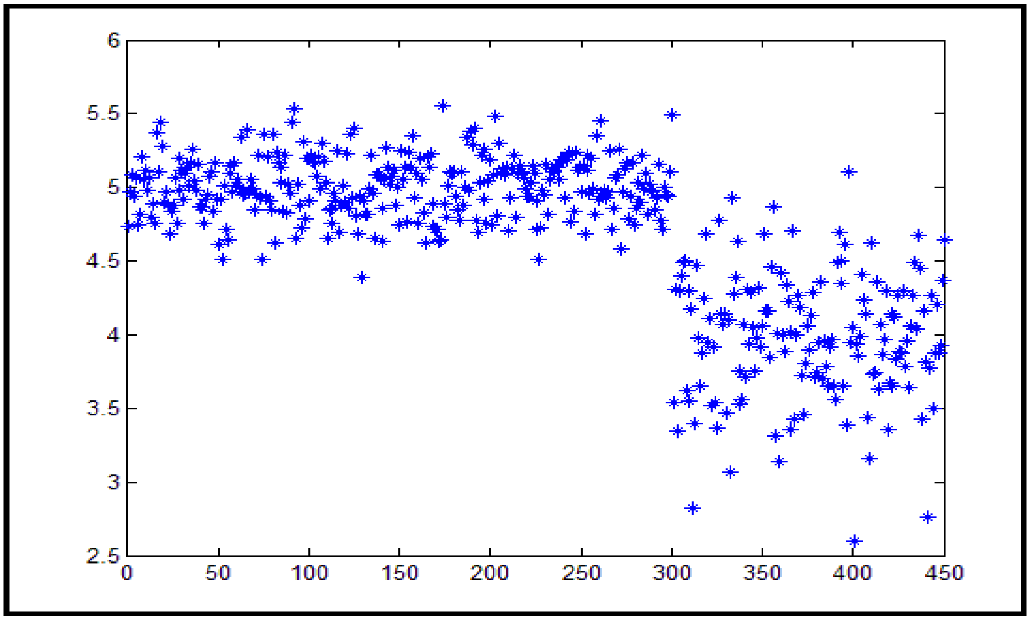

Figure 8. The intensity of coefficient 798, plotted during CMP process, from the starting 1s to 450s.

The amplitude of the AE signal tends to decrease during the progression of polishing procedure allowing to better situating the status of wafer (under -polished, desired level of polishing, over -polished). The implementation of acoustical sensor allows to measure and to convert the output of data and the environmental factors into signal. Subsequently, to transmit the obtained signal to data acquisition system. The PC based DAQ system is charged of the acquisition of electrical signals and digitized it. To make interface between data acquisition system (DAS) and MATLAB, the use of appropriate scientific engineering software will be unavoidable allowing the transfer of the data digitized by DAS at a particular dyadic length, to a (.m) file in MATLAB. The obtained Matlab file is considered is executed automatically to analyze digitized data. The non-stationary mentioned above, the presence of white noise and the autocorrelation in AE signal have justify the use of wavelet analysis. The execution of multi-resolution analysis will be as follow: Firstly, the decomposition, in this step, selecting the level of wavelet decomposition to optimize the representation of the filtered signal.

To illustrate the details and approximation coefficients, we have used multi-signal Analysis (in 'wavemu' Matlab) and to compute the $c_{j_0,k}$ and $d_{j,k}$ coefficients (for the scaling and wavelet functions), dec = mdwtdec('r',X,6,'db4'). With 'r' direction of decomposition, X the matrix of sequence of acoustical signals from the beginning to the end of CMP process, 6 the level of decomposition, and finally, db4 is the type of wavelet. We will interest only to approximation coefficients noted in matlab (ca) at level 6, to detect the endpoint. After, wavelet analysis, we have obtained only 10 coefficients from the original 1024 points, without noise and having significant informations about the CMP process. In figure 9, at each scale of decomposition, the high frequency is filtered out, represented by the detail coefficients. Secondly, the application of soft thresholding to the resulting wavelet coefficients at each scale of decomposition. The reconstruction of details in time domain is done from the thresholded wavelet coefficients. The wavelet chosen is Daubechies fourth-order wavelet basis functions (db4) since it is more widespread and practical in the discrete wavelet analysis. The multiresolution analysis is restricted to eight levels of decomposition. At each scale of decomposition, the high frequency filtered out shows by the detail coefficient. The original signal will be approximated through the use of the set of the basic functions (wavelets) of the first scale in order to give the approximation coefficient a_1 and the coefficient detail d1 represents the difference between the first approximation signal a_1 and the original signal. To determinate approximation at level two, a_1 is taken and approximated by a set of basis of the second scale to obtain a_2 with d_2 the difference between a_1 and a_2. The same work is done until the last level of decomposition 8. After that we will apply PCA- Hotelling chart to detect the endpoint.

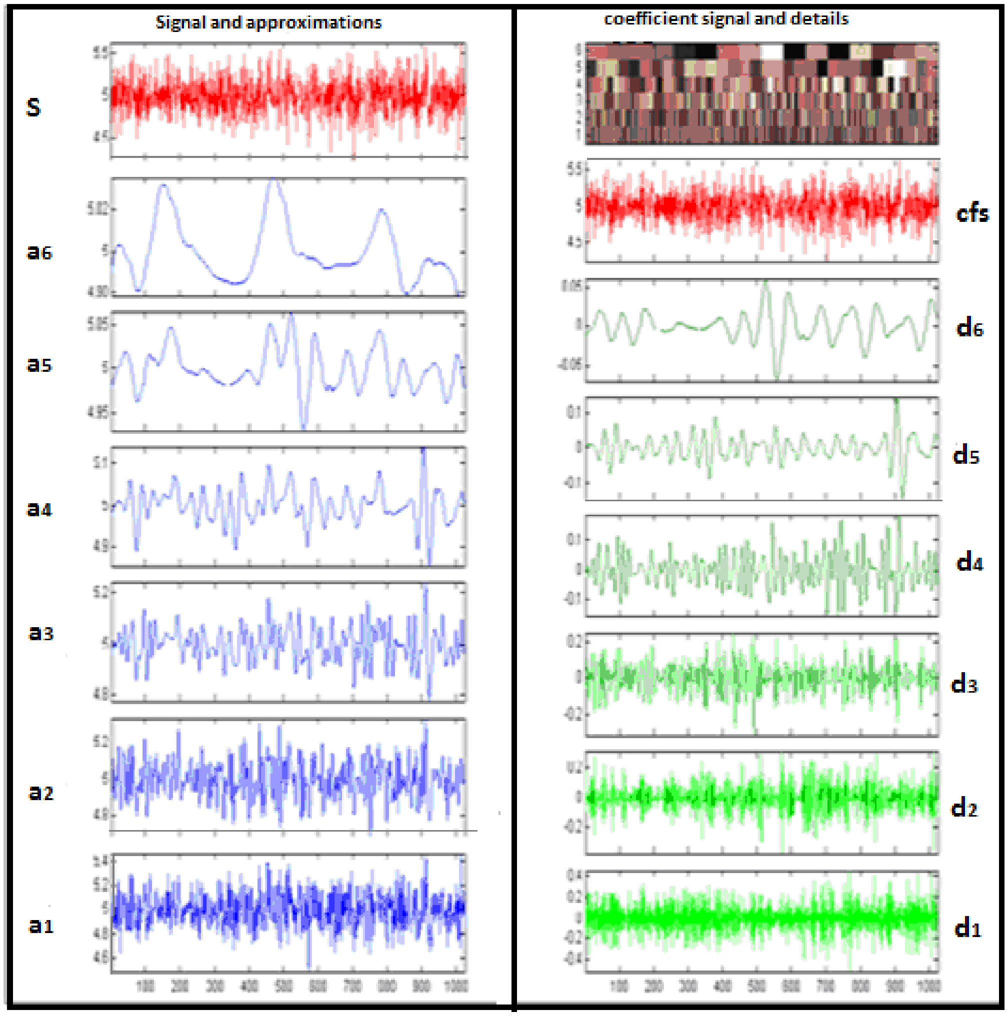

Figure 9. The application of multiresolution to the first signal (t=1s).

Table1. Variable importance after PCA

Variable	Power	Importance
1	0,999999	1
4	0,999998	2
3	0,999997	3
2	0,999997	4
10	0,999988	5
5	0,999919	6
8	0,999839	7
9	0,999704	8
6	0,999236	9
7	0,999040	10

The number of approximation coefficients retained after PCA are 5 from 10, and they are (v_1, v_4, v_3, v_2, v_{10}) given by table1. These 5 variables to construct Hotelling T2 chart, shown in table1, in order to detect the end point. The control chart is plotted in figure 10. It should be noted, with the prior knowledge of the process, based on previous experiments (Phase I), the mean of the process in control (endpoint is not detect) is equal to 7. 10-6 and the endpoint is generally suspected between the 370s and 380s. And the endpoint is detected at 378, if we stop the process before 370s, the wafer is not well polished and if we continue the polishing after 380s, we will damage the wafer (wafer over-polished).

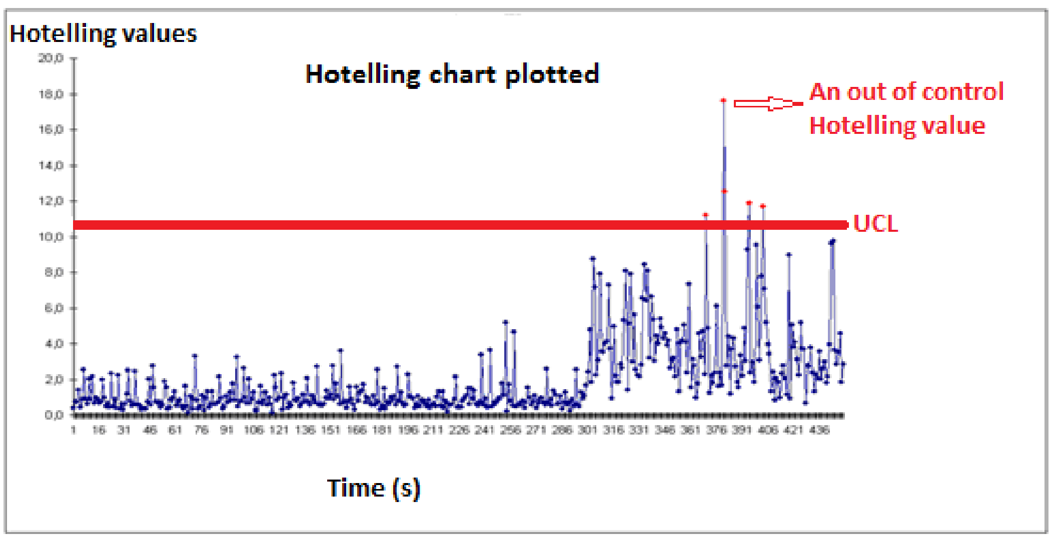

Figure10. Monitoring Endpoint detection based on Hotelling chart (in phase II) and PCA-wavelet approximation coefficients.

6. COMPARATIVE STUDY:

In dimension reduction techniques, there are two problem statements considered: The first one, called unsupervised dimension reduction methods and the second named supervised dimension reduction methods. In the supervised dimension reduction techniques, we have a set of observations $S=\{Xi\}Ni=1$ and we need to find a reduced representation for it, by applying,

- **Linear models** We have few techniques Like Multidimensional scaling (MDS) technique and *PCA* which is the most popular technique.
- **Nonlinear models** We have a lot of different techniques, like Local tangent space alignment (LTSA), Isomap, Locally Linear Embedding (LLE) and many others.

For the supervised dimension reduction, we have a set of observations expressed by both regressors and responses $S = \{Xi, Yi\}Ni=1$, $Xi \in Rp$, $Yi \in Rq$ and we need to find a reduced representation for *X*s which preserves information of the dependent variable *Y*s. And we assume that the following model holds: $Y=f(X)+\varepsilon=g(\tilde{X})+\varepsilon$, $\tilde{X}=XB$, $B \in Rp\times d$, $d<p$. Which mean that, the regression $Y=f(X)$ to lye on some hyper-plane of lower dimensionality. This problem statement considered in the case when we finally need a regression to be constructed.

- **Linear models.** PLS is the most popular (and probably the only linear) technique.
- **Nonlinear models.** Like MAVE, Outer Product of Gradients (OPG), Sliced Inverse Regression (SIR) and many others.

As noticed above, the introduction of wavelet analysis becomes inevitable with the presence of noise, nonlinearity and huge amount of data. Also, due to the hardness of defining the dependent variable in CMP process, the use of unsupervised learning will be attractive candidate for dimension reduction techniques because these types of algorithms try to find correlations without any external inputs other than the raw data. The transformation of nonlinear data to linear combination through wavelet analysis will drive us to exclusively concentrate on only to the linear technique. In our study, we will use Wavelet analysis, PCA-Wavelet, MDS-Wavelet to construct Hotelling chart and detect the endpoint. Usually, the MDS methods are efficient when the data is highly non-metric or sparse [14]. As shown is figure 11, the endpoint is detect at the 320s for the reconstructed Hotelling chart directly to the wavelet coefficients, and at 329s for MDS-Wavelet Hotelling chart. These two values (320; 329) translate a changement in the mean value (from 7.10^{-6} to 5.10^{-6}), but at the same time, the endpoint is reached earlier compared to the usual interval of the end polishing [370s-380s], reflecting an under-polished wafer.

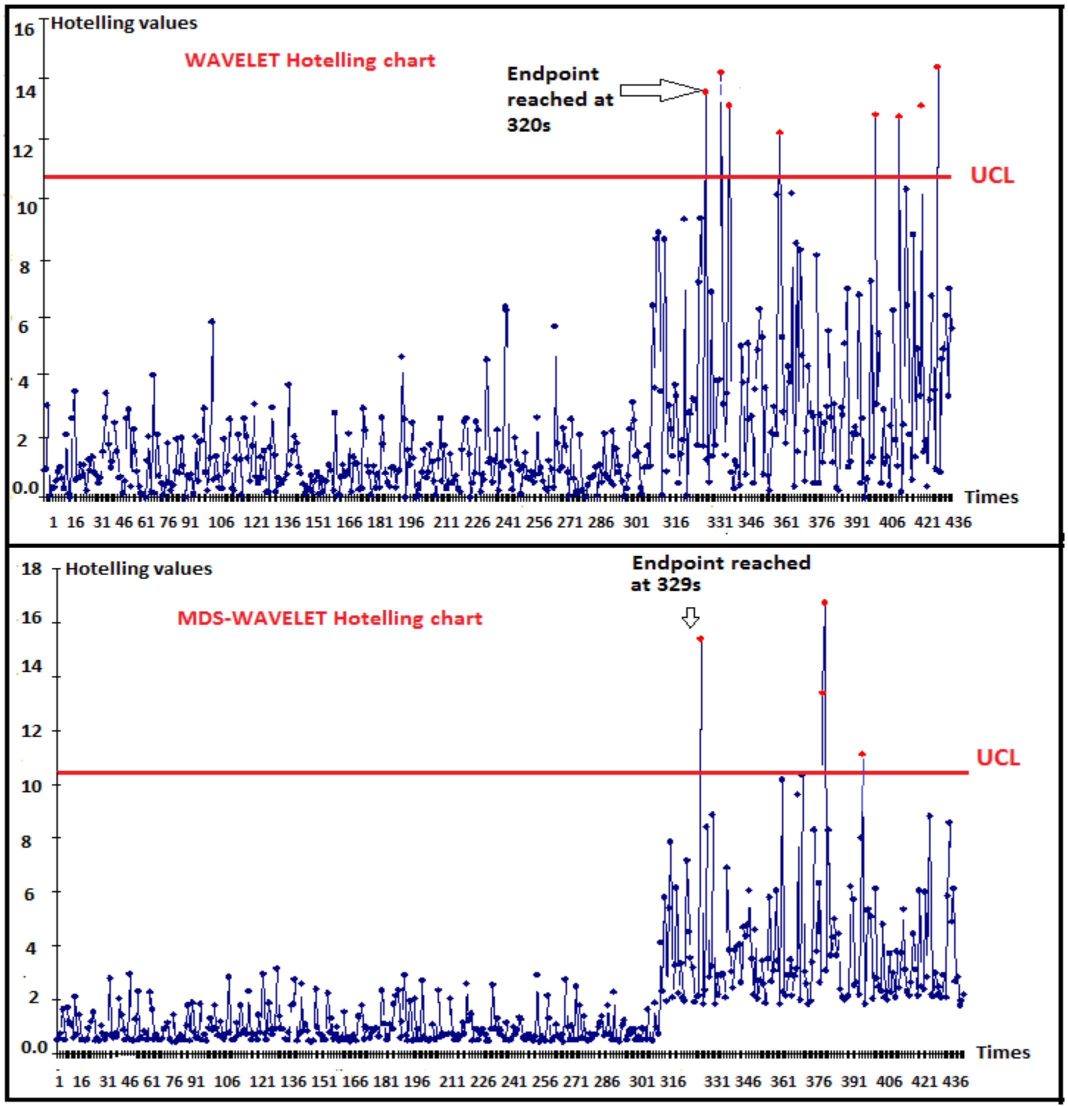

Figure11. The Hotelling statistic values plotted in Wavelet and MDS-Wavelet Hotelling chart.

7. CONCLUSION AND FUTURE WORK

The most industrial processes are represented by data situated at multiple scales due to the occurrence of various events in a process at different time and frequency localizations, with huge data acquisition system. Hence, the application of wavelet analysis in the context of system identification has become ineluctable to clean the signal and reduce the dimensionality. To overcome the problem of increasing false alarm, due to the brobdingnagian collected data for each signature, we have employ two famous unsupervised dimension reduction techniques, PCA and MDS to the wavelet coefficients. After that, Hotelling chart is constructed to detect the endpoint. This latter is reached if the Hotelling T^2 value exceeds the upper limit. Matlab software is used for Signal processing and Statistica is used for Data reduction (PCA and MDS). The obtained results show the out-performance of PCA- wavelet chart compared to MDS-Wavelet and Direct wavelet Hotelling charts. The suggested researches in the future are examination of other charts using wavelet-based multi-scale to different defect in CMP process, using wavelet analysis to identify the type of defect according to the obtained values of decomposed coefficients (studying local changes using wavelet analysis to differentiate the type of defect), employing different types of wavelet to study the robustness of results.

REFERENCES

[1] Abdul-Wahab SA, Bakheit CS, Al-Alawi SM. (2005) Principal component and multiple regression analysis in modelling of ground-level ozone and factors affecting its concentrations. Environ Modell Softw.

[2] Viana M, Querol X, Alastuey A, Gil JI, Menéndez M. (2006) Identification of PM sources by principal component analysis (PCA) coupled with wind direction data. Chemosphere.

[3] Sousa SIV, Martins FG, Alvim-Ferraz MCM, Pereira MC. 2007 Multiple linear regression and artificial neural networks based on principal components to predict ozone concentrations. Environ Modell Softw.

[4] Youssef S, Elfarag A A, Raouf R (2011) A robust steganography model using walete-based block-partition modification, International Journal of Computer Science & Information Technology, Vol 3, No 4.

[5] AbuBaker A (2012) Mass lesion detection using wavelet decomposition transform and support vector machine, International Journal of Computer Science & Information Technology, Vol 4, No 2.

[6] Karthik D, Manikandan N, Shri Hari R R (2011) Wavelet based spectrum sensing techniques for cognitive Radio- A survey, International Journal of Computer Science & Information Technology, Vol 3, No 2.

[7] Ben Zakour S. and Taleb H. (2012), Using Discrete Wavelet analysis and Sequential test to detect the endpoint in CMP process, international journal of computer applications.

[8] Donoho, D.L. and Johnstone, I.M. , (1994), Ideal spatial adaptation by wavelet shrinkage, Biometrica, 81, 425455.

[9] Donoho,D.L. and Johnstone, I.M. , (1995), Adapting to unknownsmoothness via wavelet shrinkage, Journal of the American Statistical Association, 90(432), 12001224.

[10] Donoho,D.L., Johnstone, I.M.,Kerkyacharian,G. and Picard,D. , (1995), Wavelet shrinkage: asymptopia (with discussion), Journal of the Royal Statistical Society, 57(2), 301369.

[11] Donoho, D.L. and Johnstone, I.M , (1998), Minimax estimation via wavelet shrinkage, The Annals of Statistics, 26, 879921.

[12] Hotelling, H. 1947. Letter of June 18, (1947), to Newton B. Drury. Included in the report The Economics of Public Recreation: An Economic Study of the Monetary Evaluation of Recreation in the National Parks, 1949. Mimeographed. Washington, D.C.: Land and Recreational Planning Division, National Park Service.

[13] Shiyu Zhou, Baocheng Sun, and Jianjun Shi (2006), An SPC Monitoring System for Cycle-Based Waveform Signals Using Haar Transform, IEEE TRANSACTIONS ON AUTOMATION SCIENCE AND ENGINEERING, VOL. 3.

[14] Flora S. Tsai (2011), Dimensionality reduction techniques for blog visualization, Expert Systems with Applications 38 (2011) 2766–2773.

Authors

Sihem Ben Zakour

Associate researcher in LARODEC and PHD student in statistic,
Departement of quantitative methods, Higher institute of management, Tunis, Tunisia

Hassen Taleb

Associate professor, associate researcher in LARODEC and Head (Director) of Higher institute of Business administration, Gafsa, Tunisia.
Department of quantitative methods Higher institute of Business administration, Gafsa, Tunisia