201st Meeting of the Electrochemical Society, International Symposium
on Plasma Processing XIV, Abs. 415, Philadelphia, PA, May 2002.

QUANTITATIVE ANALYSIS AND COMPARISON OF ENDPOINT DETECTION BASED ON MULTIPLE WAVELENGTH ANALYSIS

Brian E. Goodlin, Duane S. Boning, Herbert H. Sawin

Massachusetts Institute of Technology

Cambridge, MA 02139, USA

Traditional methods of endpoint detection for plasma etching processes, where the intensity of a single spectral line from optical emission spectroscopy (OES) is monitored, have proven unreliable for low open area (<1%) etching processes such as contact and via etch. In this paper we show that with the proper use of CCD-array based optical emission spectrometers, multi-wavelength endpoint detection can improve the endpoint detection sensitivity by a factor of 5-6 over the traditional single wavelength methods. A quantitative method for determining endpoint detection sensitivities of various algorithms using signal-to-noise ratio is developed. A new multi-wavelength method for "optimal" improvement in sensitivity is developed that uses prior knowledge of endpoint for improved sensitivity. This method called the *MSN statistic* is compared and found to give superior sensitivity to other multi-wavelength algorithms commonly used in commercial products and to the traditional single wavelength method.

Noise sources have been characterized as either uncorrelated noise primarily arising at the sensor or correlated process variations (drifts), each of which needs to be removed for sensitive endpoint detection. The mechanism for removing noise is different for uncorrelated and correlated noise; multi-wavelength algorithms are classified according to the type of noise that can be removed with that algorithm. Frequency-based filters (smoothing and differentiating) can also be used for removing each of these types of noise. For our experimental data, the best results were obtained when using a high pass filter to remove signal drift (correlated noise) combined with the MSN statistic to remove photon shot noise (uncorrelated noise). A flowchart with an overall strategy for multi-wavelength endpoint detection for any general data set has been developed.

INTRODUCTION

Real-time endpoint detection has been utilized since the late 1970s as a means of achieving better control over plasma etching processes in semiconductor manufacturing [1]. Nevertheless, for the critical contact and via etching steps, endpoint detection has been difficult to achieve, and recent work has focused on improving the sensitivity of endpoint detection for these low open area etches (<1%) [2-17]. In particular, the use of array-based optical emission spectroscopy (OES) sensors has shown promise for improved sensitivity, with several companies now offering multi-wavelength OES endpoint detection sensors.

Many sensors have been explored for use in endpoint detection in the past twenty-five years. These sensor systems include: 1) optical emission spectroscopy [4-8,10,12,16-24], 2) mass spectrometry [9,21,22,25-27], 3) RF impedance monitoring [28,29] 4) reflectometry (in-situ ellipsometry, laser interferometry [30], optical emission interferometry [31-33], or interferometry with an external light source [34]), 5) DC bias or current [11] and ESC bias compensation [3], 6) thermal imaging [35], 7) absorption spectroscopy (tunable diode laser [15] and FTIR absorption spectroscopy). Comparisons of some of these techniques are given in [22,30,36]. Many of these techniques are not practical in a manufacturing environment, or have limited applicability. The primary sensor technology used has been optical emission spectroscopy (OES), because it is both non-invasive and sensitive to changes in the plasma.

Optical emission spectroscopy (OES) has historically been used for endpoint detection by monitoring the emission intensity from a single wavelength corresponding to a product or reactant species in the plasma etcher during the etch resulting in an endpoint trace, like those shown in Figure 1B. Endpoint is detected when there is some shift in the endpoint trace corresponding to the clearing of the film being etched to the underlying film or stop layer. The emission intensity is sometimes taken as a ratio to some inert species emission intensity in an attempt to remove drift in the emission signal over time. For example, see Litvak [12]. Unfortunately, these one or two wavelength methods are unreliable for etching processes with a low open area fraction (<1%) on the wafer being etched. With the introduction of array-based OES sensors (CCD array or diode array), one is no longer limited to monitoring only one or two wavelengths, and several companies are now offering multi-wavelength OES endpoint detection systems with the expectation of improved sensitivity for endpoint detection.

Much of the recent work on endpoint detection has focused on methods for improving the sensitivity to endpoint detection using multi-wavelength OES [4-8,16,17,19,23,24]. Although a number of different strategies have been proposed and/or demonstrated, there has been a lack of true understanding as to why these algorithms actually work or don't work. There has been little quantification of the results for improvement of endpoint detection sensitivity over the traditional single wavelength methods. There has also been no comparison between different multi-wavelength endpoint strategies.

This leaves many questions unanswered such as: how much improvement is possible with a multi-wavelength OES system and why? What is the best multi-wavelength

strategy to utilize and why? Under what conditions will a multi-wavelength strategy work and not work? These are the questions that we will attempt to address in this paper.

In the process of answering these questions, we will develop and understanding for the nature of multivariate noise sources in OES and how to remove these sources of noise. We will develop a method to quantify the sensitivity of various endpoint detection strategies for comparison. Using this method, we will develop a new multi-wavelength endpoint detection strategy that provides the maximum improvement in sensitivity. We will compare the results of our strategy with several other strategies used in commercial products, revealing that our strategy gives the best improvement in sensitivity. Finally, we will develop an overall strategy for multi-wavelength endpoint detection for any given etch process.

This paper is a summary of the results from a thesis, which should be referred to for more specific details [5]. We are planning to publish several papers based on the thesis in the Journal of the Electrochemical Society in the near future.

MULTI-WAVELENGTH ENDPOINT DETECTION

Multi-wavelength OES Data

Typical multi-wavelength OES systems use a dispersion grating to separate light that is collected from the plasma onto a linear CCD (or diode) array that can often have over 1000 pixels. The overall wavelength range of the CCD array depends on the spectrometer with typical ranges extending from 200 to 1000 nm, leading to a resolution of 5-10 Å per pixel. Some OES systems use a grating with higher dispersion to achieve better resolution over a smaller wavelength range, at the expense of lower overall throughput of light.

A modern CCD array spectrometer can process signals rapidly, with typical integration times on the order of several milliseconds possible. Nevertheless, the light collection efficiency generally results in integration times that are longer (100 ms is typical). Since most etching processes are a minute or two, this results in over 1000 wavelength channels collected for over 1000 time samples, as depicted in Figure 1A. The goal of a multi-wavelength endpoint detection algorithm is to extract a couple of pieces of information (beginning, middle, and end of endpoint), from over one million data points. This is often done by first combining the wavelength data together to form a single multi-wavelength parameter or statistic that can be monitored over time and forms an endpoint trace similar to the traditional single wavelength endpoint traces.

There are many ways in which the multi-wavelength data can be visualized. The *time-dependent multi-wavelength data view* is a way to visualize the evolution of the spectra over time as shown in Figure 1A. Looking at the data for different wavelength channel results in a series of *endpoint traces* that correspond to different chemical species from the plasma, as seen in Figure 1B. One can see from this figure that a number of different wavelength channels have signal changes at endpoint. In fact, due to the highly coupled nature of the plasma, the majority of the wavelength signals from a typical

process exhibit a change at endpoint that can be used to improve the sensitivity to endpoint. Another way to view the data is to look at different time slices of the data, which results in a *snapshot spectrum view* of the data, as shown in Figure 1C. The snapshot spectrum represents a fingerprint of the plasma chemistry at any given point in time, and is useful for identifying specific chemical species in the plasma. Unfortunately, none of these views of the data gives any intuition about how to combine the data together in a way that will enhance the sensitivity of endpoint. For this we need a different way of depicting the data – by using a *multivariate parameter space view*.

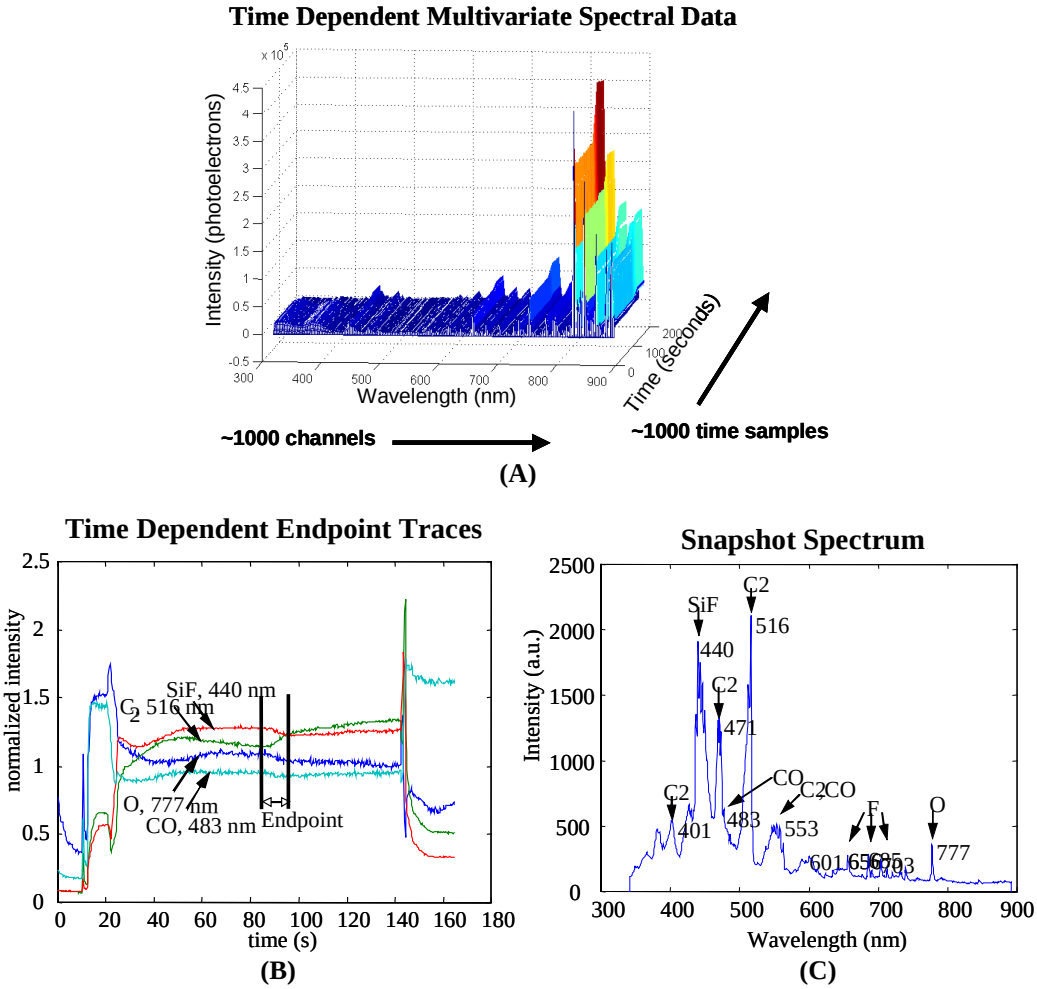


Figure 1. (A) Plot of time dependent multi-wavelength spectral data, with over 1000 discrete wavelength channels from 300-900 nm, collecting data at a rate of 10 Hz, leading to over 1000 time samples for a typical etch. (B) Each wavelength channel gives a unique endpoint trace, from which endpoint may or may not be detected. (C) During each time slice, a snapshot spectrum reveals the nature of the emission process at that given time. Unfortunately, none of these views is particularly useful in terms of creating an algorithm for combining the data together for endpoint detection; for this, we need a multivariate parameter space view, depicted in Figure 2.

The *multivariate parameter space view* of the OES data plots each snapshot spectrum as a single time point in a multi-dimensional space, where each axis is the intensity of a

particular wavelength channel. This means that if there are 1024 wavelength channels, then the data is plotted in a 1024-dimensional space, which is pretty hard for us to picture. Nevertheless, a lot of intuition can be gained by looking at a more simplified three dimensional example, as shown in Figure 2, and extending our intuition out to the higher-dimensional cases.

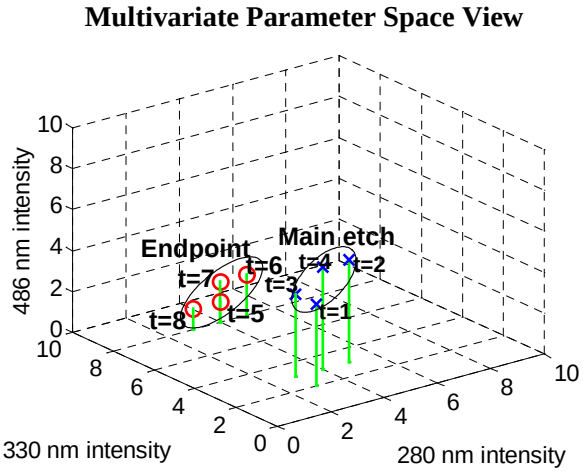


Figure 2. Multivariate Parameter Space view of optical emission spectroscopy (OES) data has a unique dimension for every wavelength channel, which in this example is limited to three dimensions, but in practice will result in a parameter space with over 1000 dimensions. The wavelength spectrum at each time sample are plotted as single data points in the multi-dimensional space. In general, the data points tend to cluster in some region in this space during the main etch, and upon clearing to the underlying film, the data points shift to some other region in the multivariate space. We can draw a hyper-ellipse to bound the main etch and post-endpoint data, which can be used to define when endpoint has occurred.

In this example, we have eight time samples and three wavelength channels. The first four time samples correspond to data during the main etch and the last four time samples correspond to the data after endpoint. The data points during the main etch and after endpoint tend to cluster in different regions in the three-dimensional space. These regions can be bounded hyper-ellipses which can be used to detect when endpoint has occurred, i.e. when a point falls outside of the main etch hyper-ellipse, this indicates that endpoint has started.

Now what remains is to map this multi-dimensional picture to a single multi-wavelength parameter, e.g. a linear combination of the wavelengths is one simple way to combine the variables. In fact, there are a number of different mappings that we can use, and this is what separates the different multi-wavelength strategies that are available. These strategies are discussed next.

Multivariate Algorithms

Broadly speaking, we can classify multivariate algorithms into one of two categories, which are easily envisioned using the multivariate parameter space view as depicted in Figure 3. The first category looks for a *mean shift* in the data between the main etch and

endpoint data. The second looks for changes in the variance structure or noise in the data, i.e. a *covariance shift*, between the main etch and endpoint data – this covariance shift can be due to either rotation or stretching and expanding of the hyper-ellipse bounding the data from each subset. In Figure 3, we have summarized a number of multivariate algorithms that are commonly available and classified them according to each of these categories. In our work, we focus mostly on mean shift algorithms, since they are generally more sensitive [5]. Commercial multi-wavelength OES systems utilize principal component analysis (PCA), Hotelling's T^2, evolving window factor analysis (EWFA) or some combination of them to achieve endpoint, so these will be compared against the single wavelength methods later in the paper. Application of Hotelling's T^2 to endpoint detection is described in Le [23] revealing qualitative improvements in endpoint detection sensitivity. The use of T^2 statistics based on principal component analysis (PCA) scores along with Q-statistic residuals is demonstrated in White et al. demonstrating difficulties associated with "drift" noise in endpoint traces as well as qualitative improvements in endpoint detection sensitivity. Yue et al. [16] describe the use of individual PC scores for endpoint detection with the advantage of removing much of the drift by not using the first principal component, with some qualitative advantages demonstrated. Hosch et al. [4] describe the use of a neural network based PCA approach with qualitative demonstration of low open area endpoint detection sensitivity. Branagh et al. [24] describe the use of evolving window factor analysis (EWFA) for endpoint detection to capture changes in the variance of the data. Chen [19] describes many of the other covariance shift algorithms with application to endpoint detection.

In our work, we have developed a new way of combining the endpoint data together using a new statistic called the MSN statistic and described as a nearly equivalent method using discriminant analysis [5,7]. Both of these algorithms use prior knowledge of the direction of endpoint in the multivariate parameter space to enhance the endpoint detection sensitivity. Similar methods are now being developed by [17] as well.

Mean Shift					Covariance Shift		
Multivariate Quality Control			Discriminant Analyses		Magnitude	Direction	
T^2	PCA	Modelled Outputs	Linear Discriminants	MSN*	Generalized Variance	EWFA / Eigen-function	PCO

Focus of current research is shaded * MSN statistic developed from this work

Figure 3. Multivariate endpoint detection algorithms can be classified according to the type of shift that one is trying to observe, either a shift in the position of the mean of the data or a shift in the covariance structure of the data by either rotation or stretching of the hyper-ellipse bounding the data. The focus of this paper is on mean shift algorithms, although we do compare one commonly used covariance shift algorithm called evolving window factor analysis (EWFA) in this paper. Some commercial products rely on T^2 or PCA, while others are based on EWFA. The MSN statistic and linear discriminant methodologies are proposed in this paper as superior alternatives for multi-wavelength endpoint detection.

Quantifying Endpoint Sensitivity

With different multi-wavelength algorithms available, some method is required to compare the endpoint detection sensitivity quantitatively. To quantify endpoint detection sensitivity, we developed a simple method, which calculates the signal-to-noise ratio (SNR) for an endpoint trace by dividing the change in signal intensity at endpoint by the standard deviation of the noise during the main etch. This method is depicted in Figure 4, for a single wavelength endpoint trace. The signal-to-noise ratio calculation can be used for single wavelength traces as well as for most multi-wavelength endpoint traces.

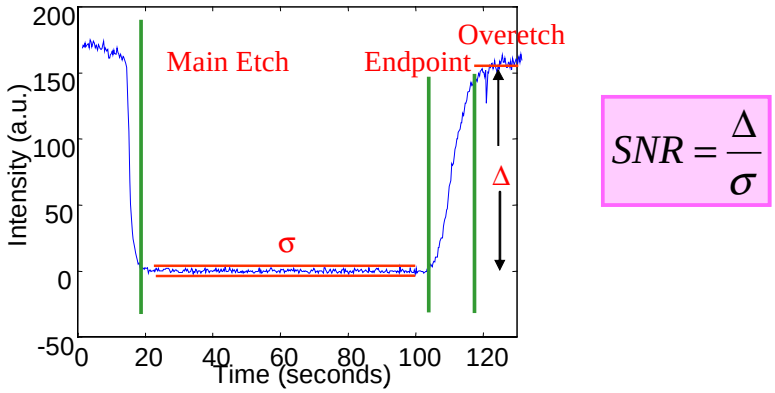


Figure 4. Quantification of endpoint sensitivity. To quantify endpoint detection sensitivity, we calculate a signal to noise ratio, by taking the change in the signal intensity at endpoint and dividing by the standard deviation of the noise during the main etch. This same methodology can be applied to single wavelength endpoint traces, and most multi-wavelength endpoint traces, provided that they were formed by linear combinations of the incoming wavelength channels.

Once the SNR metric has been defined, we can define an objective for multi-wavelength OES clearly:

Objective of multi-wavelength endpoint detection: To maximize the endpoint signal-to-noise ratio

We can also use the SNR to look at the data in an important way; the SNR can be calculated for each of the wavelength channels and a SNR spectrum can be plotted to reveal which wavelengths are useful for endpoint and which wavelengths are not. The SNR spectrum are plotted in Figure 5 for OES data collected during the etching of three different open area fractions of polysilicon using a Cl_2/HBr chemistry. These spectrum reveal that there is a large variation in the SNRs for different wavelength channels as expected, and that the SNR spectra roughly scale as the fraction of open area decreases. If one wished to determine the optimum single wavelength with the highest endpoint detection sensitivity, one would use the 656.2 nm line corresponding to an α-H emission peak, since it has the highest SNR.

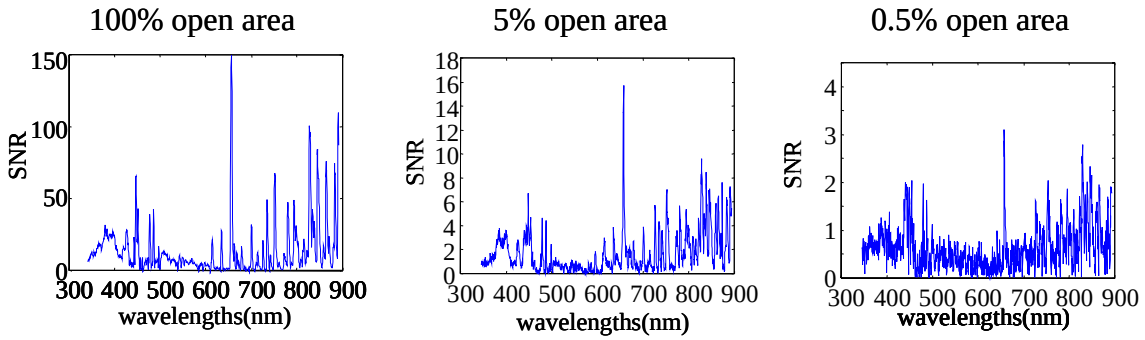


Figure 5. SNR spectra for 3 different levels of open area during polysilicon etching with HBr/Cl_2 reveal which wavelengths are useful for endpoint detection. Notice that the SNR spectra have similar characteristics for the 3 open area fractions, but not identical. The SNR spectra scale approximately with the open area fraction as we would expect. The largest single wavelength SNR is for the α-H peak at 656.2 nm, with a SNR that decreases from 150 for the blanket polysilicon film down to less than 3 for the 0.5% open area wafer.

The SNR can also be used in a more powerful way; we can use it to create an optimal weighting of wavelength channels to maximize the SNR improvement. This is explored in the next section.

OPTIMAL MULTI-WAVELENGTH STRATEGY

Now that an objective has been defined to achieve maximum improvement in SNR, we can determine an optimal multi-wavelength strategy given certain assumptions. The first assumption we make is that the noise in the different wavelength channels are statistically independent of one another. This assumption will hold true if photon shot noise is dominant; multivariate noise sources will be discussed later. The second assumption requires the multi-wavelength parameter to be formed as some linear combination of the different wavelength variables.

Following these assumptions, consider the following two wavelength variable example where the second wavelength ($\mathbf{x}_1$) has an endpoint change r times that of the first wavelength ($\mathbf{x}_2$) and the variables have been scaled to have noise during the main etch with unit variance:

Variable	Noise	Change at Endpoint
$\mathbf{x}_1$	N(0,1)	Δ_1
$\mathbf{x}_2$	N(0,1)	$\Delta_2=r\,\Delta_1$ $\quad 0<r<1$
$\mathbf{y}=a_1\mathbf{x}_1+a_2\mathbf{x}_2$	$N(0,a_1^2+a_2^2)$	$(a_1+ra_2)\,\Delta_1$

where $\mathbf{y}$ is an arbitrary linear combination of the two variables, with constants a_1 and a_2. If we arbitrarily restrict $a_1+a_2=1$, then we can calculate the general formula for SNR_y, the multi-wavelength signal-to-noise ratio as:

$$SNR_y = \frac{[a_1(1-r)+r]\Delta_1}{(2a_1^2-2a_1+1)^{1/2}} \tag{1}$$

We then take the derivative of this equation with respect to the weighting a_1 and set it equal to zero to find the best value of a_1, which is found to be:

$$a_1 = \frac{1}{1+r} \tag{2}$$

Plugging into a_2 and finding the ratio to a_1 reveals that:

$$a_2 = ra_1 \tag{3}$$

In other words, the optimum weighting of variables requires scaling them by their signal to noise ratios. Geometrically speaking this is easy to see as shown in Figure 6; the best direction in which to search is the direction in which the change is occurring. Extending the picture out to n dimensions, we still expect that the optimum signal to noise is obtainable by looking in the direction of the actual signal change. A sub-optimal case such as the projection to only one of the variables will always result in lower signal to noise.

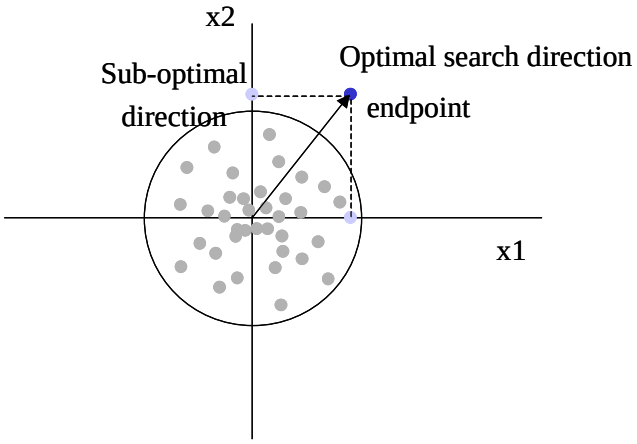


Figure 6. Optimal search direction for a 2-D case which maximizes SNR occurs in the direction that the change occurs. Any other direction requires a projection of the endpoint change that results in a lower SNR.

On the basis of this observation, we define a new metric called the *MSN* or *Multivariable statistic weighted by the Signal to Noise ratio* to be a weighted sum of the signal changes at endpoint where the weightings are based on the SNR of the individual components. The MSN statistic is given by:

$$\mathbf{y}_{MSN} = \sum_{j=1}^{n} \left(SNR_j \right) \frac{\mathbf{x}_j - \overline{\mathbf{x}}_j}{\sigma_j} \tag{4}$$

where *j* is the wavelength variable index for the n different wavelengths collected by the spectrometer, $\mathbf{x}_j$ are the raw signal endpoint traces which are mean-centered and scaled to unit variance with the mean and standard deviation during the main etch. These scaled signals are then weighted by their SNR's before summing them together into the MSN statistic. It is important to note that this form of the MSN statistic assumes that the noise is statistically independent from one wavelength to the next. This form of the MSN statistic is called the *MSN with no rotation* or *MSN_NR* for short.

The theoretical SNR for the MSN statistic is given by:

$$SNR_{y_{MSN,theoretical}} = \sqrt{\sum_{j} \left(SNR_j \right)^2} \tag{5}$$

which is always greater than the largest single wavelength SNR. If all of the single wavelength SNRs are equal, the improvement over the single wavelength signal SNR is maximum and equal to $\sqrt{n}$ where *n* is the number of single wavelength channels. If there is only one wavelength with a signal change at endpoint, then there is no improvement in SNR for the multivariate signal.

Pictorially, we can represent the MSN_NR statistic as shown in Figure 7, as consisting of a scaling to unit variance and projecting onto the optimal direction.

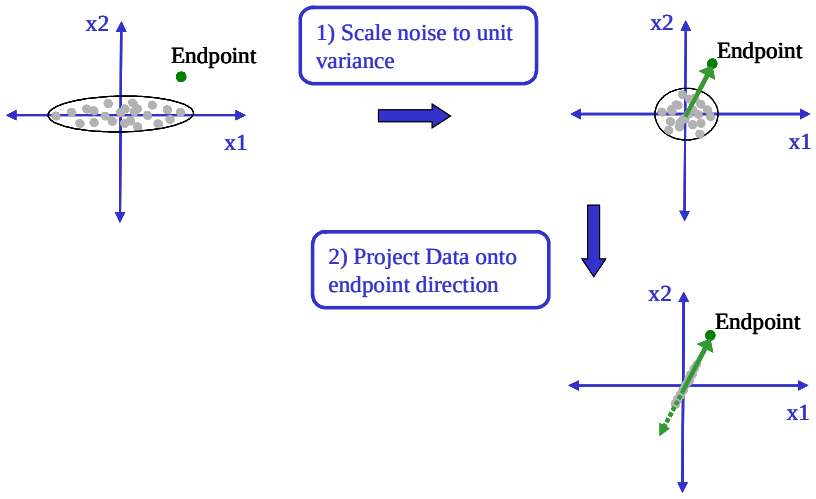


Figure 7. Pictorial Representation of MSN statistic as applied to uncorrelated data (MSN_NR) in 2 dimensions. The endpoint data are mean centered and scaled to unit variance. The data is then projected in the optimal direction, towards the direction of the mean shift at endpoint. The MSN statistic data are the projections onto this direction.

MSN statistic with rotation – MSN_R

It is possible to also formulate an MSN statistic that is used with noise that is not statistically independent, i.e. noise that is correlated from one wavelength to the next. In such a case, the original data set is rotated to create a statistically independent set of directions, from which the MSN statistic is calculated. This form of the MSN statistic is called *MSN with rotation* or *MSN_R* for short. To better understand the algorithm with rotation, see Figure 7, which pictorially represents an example with only two wavelengths. The original data set is highly correlated during the main etch, and can be bounded by an ellipse as shown. The major and minor axes of the ellipse represent a new basis set of directions in which the data is not correlated. The data is rotated to this new basis set and then is scaled to unit variance in all directions. Finally the data is projected onto the endpoint direction, which gives the MSN statistic.

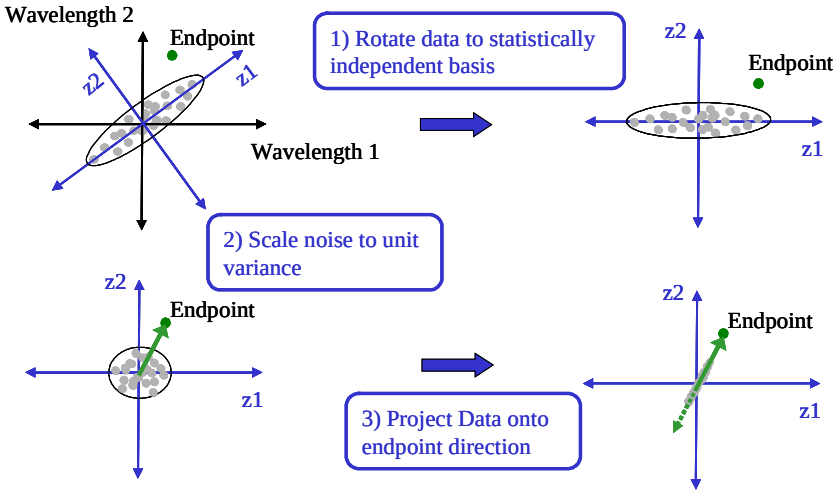
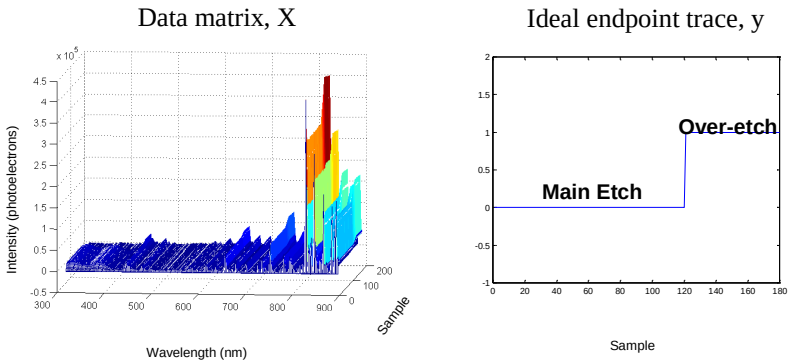


Figure 8. Pictorial representation of MSN statistic as applied to correlated data (MSN_R) in 2 dimensions. The data is first rotated to a statistically independent basis set. The data is then scaled to unit variance, before projecting the data onto the direction pointing to endpoint (mean of overetch). The projections onto this direction are the MSN statistic data.

MSN statistic same as linear discriminant (MLR_DA)

We have found that the MSN statistic with rotation is nearly equivalent to using a Fischer linear discriminant [37], so we will briefly describe the linear discriminant methodology as well. The application of linear discriminant analysis to endpoint detection is depicted in Figure 9. The discriminant is found by regressing the data onto an ideal endpoint trace and using the regression coefficients as weights for a linear combination of the individual wavelength channels. The discriminant analysis provides a way to maximize the ability to discriminate between two distinct classes, which in our case are the data before and after endpoint. For the endpoint data we have used multiple linear regression (MLR), also known as ordinary least squares (OLS), for the regression calculation, and hence the abbreviation MLR_DA. If the data are collinear, partial least squares (PLS), principal components regression (PCR), or ridge regression (RR) can alternatively be used for the discriminant analysis. In this paper we focus on the results using only MLR_DA.



$$y = X \cdot b$$

Find: $b = X^{+} y$

Figure 9. Linear discriminant analysis (MLR_DA) applied to endpoint detection. The data matrix of the OES data is regressed against an ideal endpoint trace, with the resulting regression coefficients (the discriminant) giving the weighting for the linear combination of wavelengths that best discriminates between the main etch and over-etch data. MLR_DA gives similar results to the MSN statistic with rotation (MSN_R).

Using Prior Knowledge of Endpoint

In the development of the MSN statistic, we show that projecting along the endpoint direction, i.e. weighting by the SNRs, gives the best endpoint detection sensitivity. For real-time implementation of the MSN statistic, we must calculate the direction of endpoint from some prior wafer(s) and presume that the direction of endpoint will be the same for the current wafer, i.e. prior knowledge of endpoint must be known for the MSN statistic to be applied. This of course requires that the direction of the endpoint change does not vary much from wafer to wafer, which we assume to be the case provided the etching chemistry and wafers being processed are the same from one wafer to the next.

By effectively using the prior knowledge of endpoint, we can limit the space (out of the entire multivariate parameter space) in which we look for endpoint, thus enhancing our sensitivity to endpoint. Algorithms such as Hotelling's T^2, do not use prior knowledge, and therefore require us to use the entire multivariate parameter space to search for faults, which decreases the inherent sensitivity of the algorithm. In some of our previous work, we showed that by using a subset of the multivariate space by projecting the data onto only a few principal component directions, which are then combined to form T^2, we can enhance the sensitivity towards detecting endpoint [8]. Yue et al. have shown that using principal component analysis (PCA) where endpoint is included in training also helps to reduce the space to a subset where endpoint is more likely to appear [16], and picking only select principal components and wavelengths can further enhance the sensitivity. The MSN statistic provides a means to use the prior knowledge of endpoint in the optimal way to reduce the space one has to search for endpoint to achieve the highest endpoint detection sensitivity.

NOISE IN MULTI-WAVELENGTH OES

In developing the optimal multi-wavelength statistic we made some assumptions regarding the nature of the noise. In this section we will look at various noise sources that commonly arise in multi-wavelength OES systems, and characterize the noise into one of two categories: 1) uncorrelated noise – noise that is independent from one wavelength channel to the next, and 2) correlated noise – noise that is similar from one wavelength channel to the next. We will show that the ability to remove the noise source using a multivariate algorithm is different depending on which category of noise is most dominant.

Uncorrelated Noise – Sensor noise

Noise sources occurring in a CCD array spectrometer have been characterized [5,38] and consist of the following:

1) Photon shot noise – noise arising from quantum detection of photons
2) Dark current and noise – spontaneous photoelectron generation in the absence of photons
3) Readout noise – Arising from electronic processing of signal, not a problem for high quality CCD array.
4) Thermal drift – Variations in the dark current and noise due to heating of detector. This is removable by thermoelectric cooling.
5) Pixel shift and drift – Change in the location of the pixel as a function of time, possibly due to subtle mechanical vibrations.

Of these noise sources, the readout noise, thermal drift, and pixel shift and drift can be significantly reduced by proper design of the CCD array. This leaves photon shot noise and dark noise as the primary sources of sensor noise. Shot noise is the most dominant noise source, but dark noise becomes more significant for weaker signals. The magnitude of the dark current and noise can be significantly reduced by thermoelectric cooling.

Since photon shot noise is the most dominant sensor noise, we will now focus on it. The quantum detection of photoelectrons is a Poissonian process, resulting in noise that is proportional to the square root of the signal intensity. More specifically, the number of photoelectrons detected on any given pixel of the CCD array over a given amount of time (integration time) is poisson-distributed, such that the mean value and variance of the number of photoelectrons detected are equivalent and can be approximated by a normal distribution centered at the mean. Further, the shot noise that arises in each pixel of the CCD array will be independent from the shot noise in the neighboring pixel (ignoring for quantum tunneling, pixel shift, and other sources of cross-talk between adjacent pixels). This type of noise that is independent from one wavelength to the next is called *uncorrelated noise*. The origin of uncorrelated noise from the OES sensor is illustrated in Figure 10.

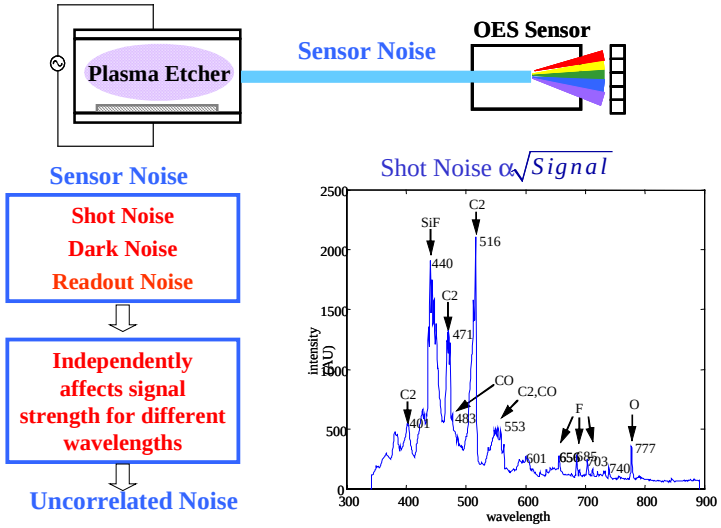


Figure 10. Noise arising at the CCD array sensor is usually dominated by photon shot noise, which is proportional to the square root of the signal intensity. This type of noise independently affects the signal strength for different wavelengths, and therefore it is uncorrelated noise.

Now that the uncorrelated noise has been characterized, it is important to understand how to remove this source of noise. It is well known that adding together signals with equal noise results in a reduction of the noise by the square root of the number of signals for uncorrelated or independent noise sources. This reduction in noise comes as a result of partial cancellation of the noise, hence we will call the mechanism to remove uncorrelated noise, *noise reduction*.

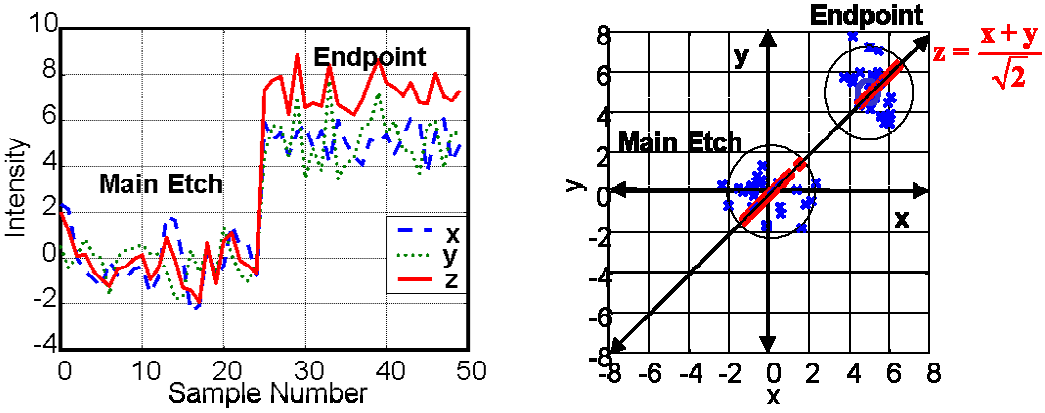


Figure 11. Demonstration of noise reduction of uncorrelated noise. For two endpoint signals (x,y) with uncorrelated noise, but similar changes at endpoint, the noise is removed by summing the signals together to get a partial cancellation of the uncorrelated noise. In the multivariate parameter space, the optimal improvement is achieved by projection along the endpoint direction, as is done with the MSN statistic.

In the left side of Figure 11, we show an example of two wavelength channels (**x** and **y**) with identical signal changes at endpoint and uncorrelated noise. Noise reduction is achieved by the addition of the two channels resulting in **z**, which has an improvement in signal to noise by the square root of two. If *n* such signals are added together, this would

result in an improvement by $\sqrt{n}$. Unfortunately, the signal change at endpoint relative to the noise, i.e. the endpoint SNRs, are generally not of equal magnitude for the different wavelengths, so the improvement is often much less than the by $\sqrt{n}$, where n the number of wavelength channels collected by the CCD array. In fact the maximum improvement is found by applying the MSN statistic to the data.

The right side of Figure 11 shows how the signal is improved in the multivariate parameter space view, by projecting the data along the direction of the signal change.

Correlated Noise – Process variations

We have seen how sources of noise that arise in the OES sensor system result in uncorrelated noise, but what about sources of noise that arise within the plasma etching system? Process noise can arise from a number of variations in the process including:

1) Variations in the input parameters such as power, pressure, flow rate, etc.
2) Variations in the chamber conditions – temperature of walls, deposition on walls, deposition or sputter of glass to reduce the transmission.
3) Normal process variation – rotating magnetic field resulting in oscillatory behavior of plasma at a certain frequency.
4) Reflections from wafer – interferometry effect [39].
5) Exposed area at edge of wafer – wafer edge effect [39].

Ultimately, these sources of variation, will lead to changes in the plasma including changes in the plasma density, plasma temperature, and plasma chemistry. These changes in turn result in changes in emission spectra that are highly correlated, that is the intensity of various signals tend to go up and down in a similar fashion. As a result, this type of variation is called *correlated noise*; the evolution of which is shown in Figure 12.

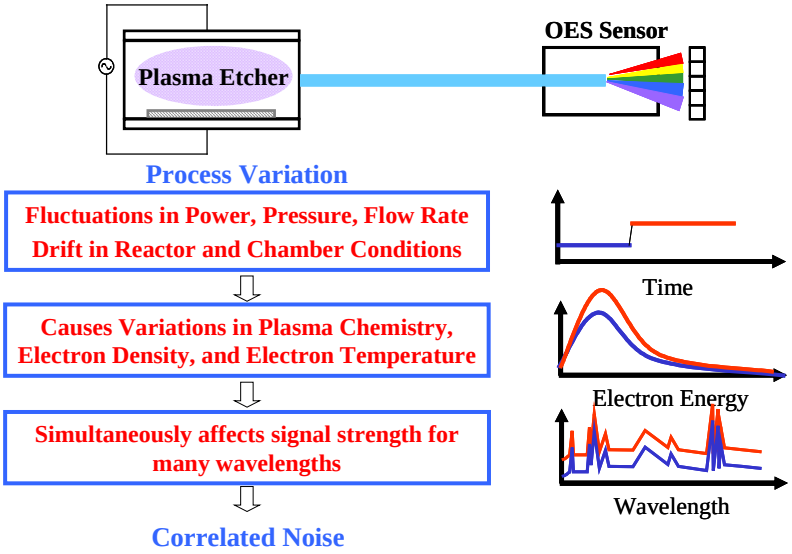


Figure 12. Variations in the plasma etcher lead to variations in the plasma, which in turn causes the signal strengths for many wavelengths to be affected in a similar manner. Since the variation is the same for many wavelengths, this is called correlated noise.

The removal of correlated noise is accomplished by different means than the noise reduction that was used to remove the uncorrelated noise. Since the noise is correlated, adding the noise together does not result in a partial cancellation of the noise. In fact for a two signal example where the noise is perfectly correlated as shown in Figure 13, the noise can be completely removed by subtracting the signals, and the resulting signal will have an infinite improvement in signal to noise. The right side of Figure 13 depicts the signals in the multivariate parameter space, revealing that the correlated noise (disturbance) falls along a given direction in that space, and can therefore be rejected along that direction. Hence we term this method for removing correlated noise, *disturbance rejection*. By looking in a direction orthogonal to this direction, infinite signal to noise is obtainable. For multi-wavelength signals with over 1000 wavelengths (resulting in 1000 dimensions in the multivariate parameter space), the noise may be correlated along many directions and must be rejected along each.

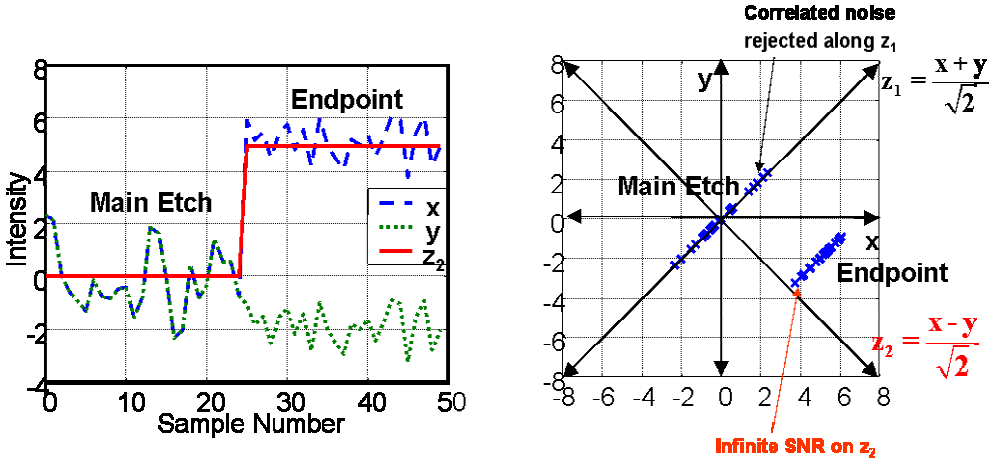


Figure 13. For two endpoint signals (x,y) that have perfectly correlated noise, the correlated noise or disturbance can be rejected by subtracting the two signals. In the multivariate parameter space view, the noise is rejected along a particular direction (z_1). By projecting along the direction orthogonal to z_1, namely z_2, infinite SNR is possible. For data that is not perfectly correlated the improvement in SNR is not infinite and is limited by the relative orientation of the endpoint direction and the correlated disturbance direction.

In practice, one will never obtain signals that are perfectly correlated, and hence one can only hope to reject as much of the correlated disturbance as possible. Further, the shift at endpoint will often be in a similar direction(s) in the multivariate space as the disturbance, such that the disturbance cannot be rejected without simultaneously removing the endpoint signal. For example, see the signal in Figure 14; we will show how frequency-based filters can be used to remove correlated noise in certain cases where disturbance rejection is ineffective.

Frequency-based filters for noise removal

We have shown that multivariate algorithms can be used for disturbance rejection, but there is another, perhaps more effective means to reject correlated disturbances for many

endpoint signals, by using frequency-based filters. To illustrate how this is accomplished, consider the example signals shown in Figure 14, in which two signals **x** and **y** are drifting over time during the main etch. The resulting signal in the multivariate parameter space reveals a high degree of cross-correlation. Unfortunately, in this example (as is often the case with experimental data), the endpoint signal change aligns closely with the disturbance, so that the disturbance cannot be rejected effectively. If we apply a high pass filter to this data for each of the signals, however, then the disturbance is removed and the residual data is uncorrelated, as shown in the right side of Figure 14. Such an approach will only work if the time scale (frequency) for the endpoint change is different than the time scale (frequency) for the drift. Since the drift has a low frequency, a high pass filter (derivative filter) is effective at removing the drift. Similarly, a low pass filter (box moving average) can be used to remove high frequency shot noise if the endpoint signal change occurs slowly over time. As a result frequency-based filters can be used to remove both uncorrelated and correlated noise sources. One must be careful, however, in the application of these filters so as not to remove the endpoint signal change that we are trying to detect.

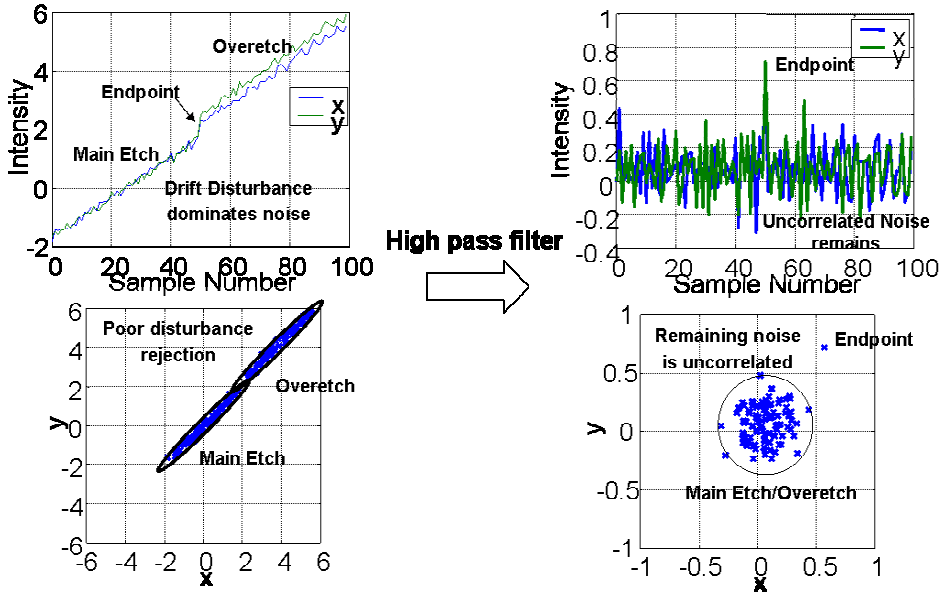


Figure 14. Frequency based disturbance rejection. Left: Two signals with low frequency drift as the dominant noise. Bottom plot shows that disturbance rejection by rotation is not effective for this data because the endpoint signal is in the same direction as the disturbance. Right: Application of high pass filter (derivative) to the data on the left results in rejection of the correlated drift noise resulting in a signal having noise that is uncorrelated.

The use of time-series analysis applied to endpoint data is a more simplified statistical approach that is often used to remove the non-stationary, or drifting behavior of signals by modeling them as auto-regressive integrative moving average (ARIMA) processes [40]. Statistical limits can then placed on the residuals, as demonstrated for tool fault detection in Spanos et al. [41], to look for endpoint. Unfortunately, the statistical time-series modeling approach does not take into consideration the nature of the signal that

one is trying to detect, but instead focuses on creating a non-drifting (stationary) signal first and then looking for faults second. This can lead to the removal of the fault (endpoint change) when modeling the drifting signal, by applying what is effectively a high frequency filter that is too aggressive. As a result, the author suggests using the frequency-based signal processing approach that was described in the previous paragraph, with the primary goal of minimizing the correlated drift noise without substantially affecting the endpoint signal.

Comparison of multi-wavelength methods

Now that we have determined that noise can be characterized into two separate categories – uncorrelated and correlated noise – it is useful to classify multivariate endpoint detection algorithms according to the action they perform on the data set. For example, the MSN statistic without rotation only reduces the uncorrelated shot noise component of data by noise reduction. The Hotelling T^2 statistic applies a rotation to try to reject disturbances, but does not benefit from any noise reduction in the uncorrelated noise. The MSN statistic with rotation applies a rotation to reject disturbances and also benefits from noise reduction along the rotated directions. Discriminant analysis also utilizes both noise reduction and disturbance rejection. Principal component analysis (PCA) generally provides the rotation to reject disturbances if only the main etch data is used for calculating the principal components (PCA_ME), but can provide both disturbance rejection and noise reduction if the endpoint data is also included in calculating the principal components (PCA_EP) depending on the circumstances. The EWFA algorithm does not utilize either of these types of noise rejection, but instead relies on a second order effect of new variance directions arising. The multivariate algorithms that are compared in this paper and their mechanism for improvement over single wavelength signals are summarized in Table 1, along with abbreviations that are used throughout the paper for each algorithm.

Multivariate Algorithm (abbrev.)	**Mechanism for improvement / Noise Removal**	**Prior knowledge of endpoint**
Hotelling's T^2 (T^2)	Disturbance rejection only	No
MSN with no rotation (MSN_NR)	Noise reduction only	Yes
MSN with rotation (MSN_R)	Disturbance rejection and noise reduction	Yes
Discriminant Analysis (MLR_DA)	Disturbance rejection and noise reduction	Yes
Principal Component Analysis – main etch data only (PCA_ME)	Disturbance rejection	No
Principal Component Analysis – main etch and endpoint data (PCA_EP)	Noise reduction for large endpoint changes Disturbance rejection for small endpoint changes	Yes
Evolving Window Factor Analysis (EWFA)	Origin of new variance direction (2^{nd} order effect)	No

Table 1. Comparison of multivariate endpoint algorithms by mechanism for improvement / noise reduction and whether prior knowledge of endpoint is used to enhance the sensitivity.

We have also described how the MSN statistic uses prior knowledge of endpoint to improve its endpoint detection sensitivity. We can characterize each of the various multivariate algorithms according to whether or not prior knowledge of endpoint is used. This is also shown in Table 1.

By examining Table 1, one might conclude that the MSN with rotation and discriminant analysis approaches should provide the best ability to reduce the noise in the multivariate data, provided that the endpoint direction can be adequately predicted from prior wafer data. In the next section, each of these multivariate algorithms will be applied to a set of experimental data and compared for endpoint detection sensitivity.

EXPERIMENTAL

Etch experiments were performed in an inductively coupled plasma etcher, shown in Figure 15. RF power is supplied at 13.56 MHz with a Comdel CPS-500A power supply to generate the plasma, with plasma densities of around 10^{10} cm^3 and plasma temperatures of around 2-3 eV for typical operating conditions. Ions are accelerated to the wafer which sits on a bottom electrode and is biased negatively relative to the plasma by an independent rf bias power supply, a Comdel CX-2000, operated at 4 MHz, to avoid crosstalk with the top power supply. Impedance matching networks are used to couple maximum power to the plasma for both the top and bottom power supplies.

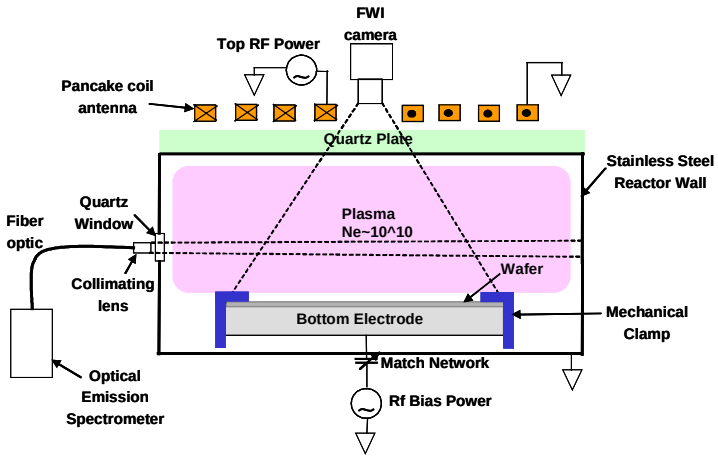


Figure 15. Inductively coupled plasma etcher with full wafer interferometry (FWI) and optical emission spectroscopy (OES) diagnostics.

The wafers used in this reactor are four inches in diameter and are clamped to the bottom electrode with a mechanical clamp that applies a force of 10 lbs. to the outer 3/8" perimeter of the wafer. Backside helium pressure of 10-20 torr is used to help cool the wafer during processing; the bottom electrode is water-cooled at 15°C. Wafer temperatures during processing are typically around 60°C, and undergo a rise in temperature of around 45°C during the first 10-20 seconds of processing.

Gas is fed to the chamber through a main feedthrough at the side of the chamber near the top. The chamber is pumped by a Leybold TMP 361C turbopump backed by a Roots

blower in-line with a mechanical pump to facilitate larger gas flow rates. The pressure in the chamber is controlled by a VAT adaptive pressure control valve.

The etching process used in this study is the etching of polysilicon with HBr and Cl_2 gases. HBr and Cl_2 is the primary chemistry used in industry for polysilicon etching, and the process has been studied extensively in the literature. The nominal operating conditions including the polysilicon film properties are given in Table 2.

Parameter	Nominal operation
Top power (W)	300
Bottom power (W)	50
Total flow rate (sccm)	60
HBr:Cl_2	2:1
Chamber pressure (mtorr)	20
He backside cooling (torr)	15-20
Polysilicon etch rate (Å/min)	3500
Polysilicon thickness (Å)	5000
Oxide stop layer thickness (Å)	1000
Oxide etch rate (Å/min)	<100

Table 2. Nominal operating conditions for HBr and Cl_2 etching of polysilicon endpoint experiments

The etch rate for this chemistry was typically 3500 Å/minute for the low open area wafers, but for the blanket wafers, a substantial loading effect was observed that drops the etch rate to around 2500 Å/min. For 5000Å thick polysilicon, this led to endpoint clearing times of around 1-2 minutes in duration.

Low open area samples were prepared by mounting smaller polysilicon wafer samples, of a specified size, to blanket oxide wafers with silver paint. The silver paint ensured that the samples were well adhered and had good thermal contact to the wafer. The sample sizes that were used, along with the polysilicon area and area fraction, are given in Table 3.

Sample	**Polysilicon area (cm^2)**	**Polysilicon open area (%)**
Blanket	8.29	100%
5% open area	0.41	5%
0.5% open area	0.04	0.5%

Table 3. Sample sizes for polysilicon with open area fractions indicated. 5% and 0.5% open area samples are mounted onto blanket oxide wafers.

For each of the three open area levels, ten consecutive wafers were processed and optical emission spectroscopy and FWI data were collected. The OES data were collected using a Jarrell-Ash Monospec 27 monochromator with a Hamamatsu S7031-1007 CCD array detector mounted at the exit focal plane of the monochromator. The Jarrell Ash Monospec 27 monochromator has a focal length of 275 nm, with an f/3.8, and an entrance slit that 25 μm wide. The spectrometer is configured in a crossed Czerny-Turner configuration, with a holographic grating having 150 grooves/mm, 24 nm/mm dispersion, and a 590 nm wavelength range from 310-900 nm. The resolution with this grating was typically around 2 nm, when used with the Hamamatsu CCD array detector.

The Hamamatsu S7031-1007 is a 2D CCD array detector that is back-thinned and consists of 1024x58 pixels; the active area of the CCD is ~25 mm x 3 mm. Full-well capacity of this CCD array is 600,000 electrons when operating in line-binning mode, which is used to reduce the data to a 1D array of 1024 channels. The CCD was thermoelectrically cooled to -10°C to minimize dark current and thermal drift. A National Instruments AT-AI-16XE-10 DAQ board was used to control the CCD array and collect the data. The A/D conversion was 16 bit and was done at clock speeds of 500 kHz. Integration times of 5 ms to 10 seconds were possible with this detector, but OES data were collected at a rate of 10 Hz during the etch. FWI data were collected simultaneously using an LES-1200 system to validate the endpoint time.

Data Pre-treatment

So that calculations were easier, only one-quarter of the original set of 1024 wavelength channels were kept for analysis; this was accomplished by keeping only every fourth wavelength channel for each run, leaving only 256 wavelength channels for analysis. The raw data were then processed using frequency filtering in the form of simple box moving averages and derivatives to generate data that was dominated by either a low frequency correlated drift, or by high frequency shot noise. The specific treatments that were used are summarized in Table 4.

Case	Box Moving Average Width (seconds)	Derivative Width (seconds)
Correlated Noise – drift limited	1	N/A
Uncorrelated Noise – shot noise limited	1	2

Table 4. Smoothing and derivative parameters

Multivariate Endpoint Implementation

After pretreatment of the data, the multivariate algorithms listed in Table 1 were applied to the data. The first wafer for each open area level was used to "train" for the covariance of the main etch data for the Hotelling's T^2, the first five principal component directions for both of the PCA analyses (PCA_ME and PCA_EP), and the endpoint change direction and regression coefficients for the MSN statistics (MSN_NR and MSN_R) and linear discriminant analysis (MLR_DA). Data from the remaining nine wafers were then used as "test" samples which were projected onto the covariance matrix, principal component directions, and projection directions, to calculate the corresponding multivariate endpoint traces for each algorithm. See Yue et al. [16] and Goodlin [5] for details on PCA implementation and Le [23] and White et al. for details on T^2 implementation. For the EWFA algorithm, no training wafers are required; the first five singular values are calculated using a window width of five seconds for each wafer. See Branagh et al. for implementation of EWFA to endpoint. Each of the 1024 wavelengths collected was also individually considered for endpoint detection.

After creating all of the single wavelength and multi-wavelength endpoint traces, signal-to-noise ratios (SNRs) were calculated for each algorithm (except Hotelling's T^2)

for each of the 9 test wafers at each of the three open area levels. The mean and standard deviation of the SNRs were then tabulated for the three open area levels for each algorithm. For the PCA and EWFA algorithms, only the principal component or singular value with the best SNR is retained in the results for each algorithm. For the single wavelengths, only the best single wavelength SNR is retained for comparison. For Hotelling's T^2, the SNR calculation could not be computed properly, but analysis of the qualitative results (described elsewhere in Goodlin [5]) indicates that Hotelling's T^2 generally has no better endpoint detection sensitivity than the best single wavelength result.

RESULTS

Quantitative Results – Drift limited noise

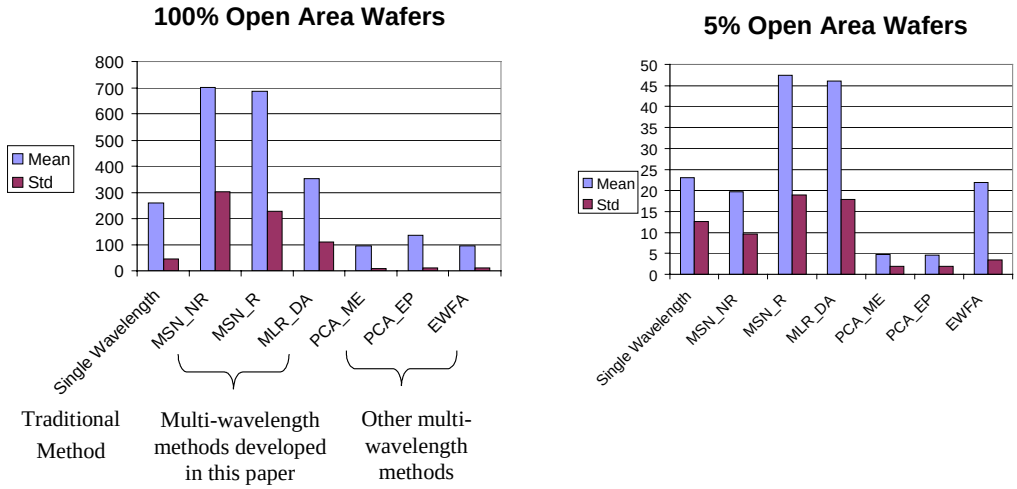


Figure 16. Bar plots showing the Mean and Standard Deviation of the SNRs calculated for the 100% and 5% open area wafers. The multi-wavelength methods developed in this paper (MSN_R, MSN_NR, and MLR_DA) improve the sensitivity over the single wavelength result by a factor of 2 to 3 for the best method (MSN_R). These methods also outperform the other multi-wavelength methods that are often used in commercial products. The Hotelling's T^2 result is not shown, since its SNR cannot be properly calculated. For all methods, The SNR decreases between 10 and 20 times in going from the 100% open area results to the 5% open area results.

The results from the SNR calculation for the drift limited case are shown in Figure 16 for the 100% and 5% open area fractions. The 0.5% open area results were not tabulated because the drift limitation led to an erroneous calculation of the SNR for the 0.5% open area results. On this figure we have indicated the traditional single wavelength method, the multi-wavelength methods developed in this paper (MSN_NR, MSN_R, and MLR_DA) and other multi-wavelength methods commonly used in commercial products (PCA_ME, PCA_EP, and EWFA). As mentioned in the experimental section, the T^2 results could not be displayed, but have been shown to have less sensitivity to endpoint than the multi-wavelength methods developed in this paper [6].

The results indicate that for drift limited noise, the multi-wavelength methods developed in this work outperform both the traditional method for endpoint detection and the multi-wavelength methods used in other products. The MSN_R algorithm provides the best result with a factor of 2-3 improvement in SNR over the best single wavelength. Somewhat surprisingly, the other multi-wavelength methods have less sensitivity to endpoint than the best single wavelength result. However, the methods described in this paper have a large standard deviation of the SNR (see Figure 16), indicating that the disturbance rejection is not robust, and these techniques will not be able to provide sensitive endpoint detection without retraining the data often. Looking at the 5% open area wafer results, we can see that the EWFA algorithm does not provide improved sensitivity over the single wavelength result, but may lead to a more robust endpoint signal for drift limited OES data, as demonstrated by a low standard deviation of the SNR relative to the mean value.

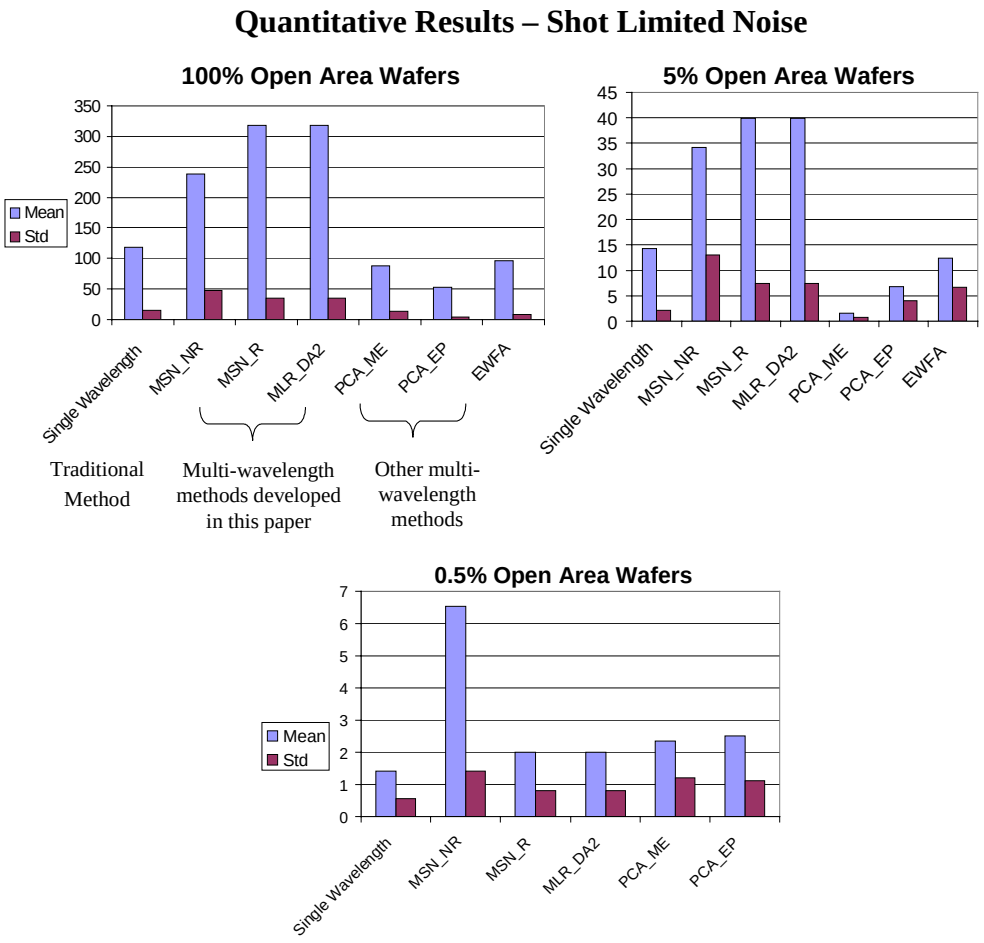


Figure 17. Bar plots showing the Mean and Standard Deviation of the SNR calculated for wafers 2-10 for the 3 levels of open area wafers (100%, 5%, 0.5%). The MSN_NR clearly gives the best results for endpoint detection sensitivity for the 0.5% open area wafers. For larger open areas, linear discriminant (MLR_DA2) and MSN_R give the best result. SNR improvements of between 2 and 3 are gained in going from the best single wavelength result to optimum multi-wavelength result. The SNR decreases between 50 and 200 times in going from the 100% open area results to the 0.5% open area results, which is dependent on the nature of the noise at each open area fraction. Robustness issues will be discussed in the text.

The results from the SNR calculation for the shot noise limited derivative case are plotted in Figure 17 for the 100%, 5%, and 0.5% open area fraction results. Recall that in this case, the drift has been removed by frequency filtering, so that only uncorrelated shot noise remains. The results indicate that for shot limited noise (uncorrelated noise), the multi-wavelength methods developed in this paper again outperform the single wavelength method by a signal to noise ratio of 2 to 3. The other multi-wavelength methods are again less sensitive to endpoint than the best single wavelength method. In this case, the standard deviations of the SNR for the multi-wavelength methods are smaller relative to the mean values of the SNRs than in the drift limited case. This indicates that these algorithms are more robust after the frequency filtering has removed

the drift. In this case, the EWFA does not provide a more robust method of endpoint detection than the single wavelength method.

For the lowest open area case (0.5%), the results indicate that only the MSN_NR or MSN statistic without rotation provides sensitive endpoint detection with a SNR greater than 3. The reason the MSN_NR provides the best result in this case is due to the nature of the training of the endpoint direction for the 0.5% open area wafers. For these wafers, the 5% open area wafers are used for training in determining the endpoint direction for the MSN_R, MSN_NR and MLR_DA algorithms, because the SNRs for the individual wavelengths are much larger and therefore the certainty in the weighting of the wavelength variables (which is by the SNR) is much higher. This assumes that the SNRs will scale approximately linearly with open area fraction, resulting in a similar direction for endpoint for the 5% and 0.5% open area wafers. However, for the MSN_R and MLR_DA algorithms, a rotation is also applied, which requires the covariance of the 5% and 0.5% wafers to be similar. In fact the covariance of these open area levels does change so that training on the 5% open area wafers is not effective, except in the case where only the SNRs are scaled. The MSN_NR, however, can only reduce uncorrelated noise and has no ability to reject disturbances, so this method is only effective if the data is dominated by uncorrelated noise. This requires that the frequency filter used to remove the drift must be effective, otherwise the MSN_NR will be limited by any drift or other correlated noise that remains.

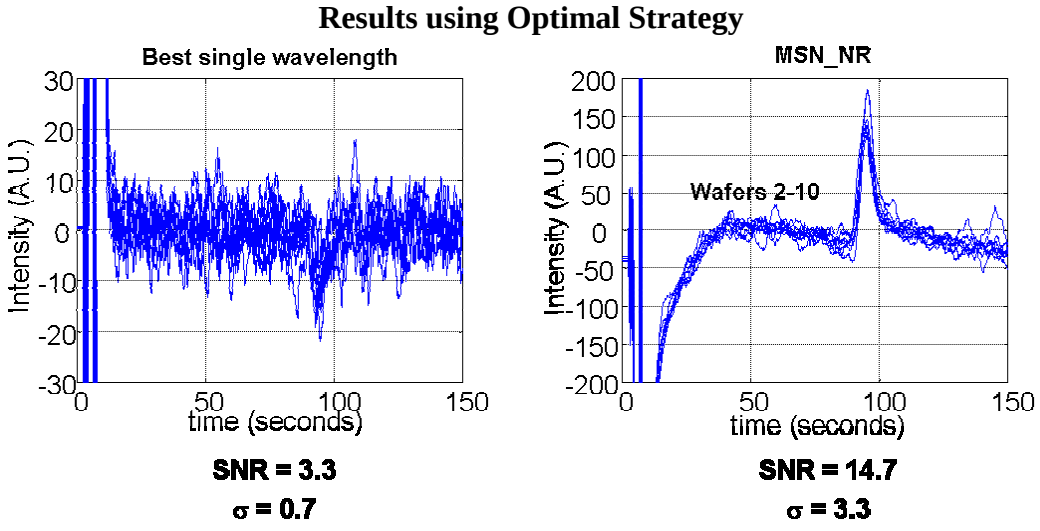


Figure 18. Comparison of best single wavelength endpoint detection endpoint traces to the optimum multi-wavelength strategy for 0.5% open area wafers, when utilizing the entire 1024 wavelength channels available from the CCD array spectrometer. Endpoint is not detectable for the best single wavelength, while it is easily detectable using the optimum multi-wavelength methodology. This methodology provides an improvement in SNR of 5.3.

Based on the results from the previous section, we have shown that the strategy for achieving the best endpoint detection sensitivity utilized a combination of a bandpass filter to remove the correlated noise (drift), together with the MSN statistic without

rotation to provide maximum noise reduction of the uncorrelated shot noise. In the previous section, however, only one-quarter of the available wavelengths were used, so the maximum potential of this strategy was not realized. Here, we apply this strategy to the entire set of 1024 wavelengths and demonstrate the results. The results are shown qualitatively for the 0.5% open area wafers in Figure 18, revealing a substantial improvement in the ability to detect endpoint for the multi-wavelength methodology over the best single wavelength endpoint trace. Quantitatively the improvement in SNR is calculated to be 5.3.

OVERALL STRATEGY FOR MULTI-WAVELENGTH ENDPOINT DETECTION

We have shown for this specific etch example how one can improve endpoint detection sensitivity by a factor of 5-6. For a more general etch process, the amount of improvement that can be obtained will depend on both the nature of the plasma and the sources of noise that are most prevalent in that particular etch process and OES sensor system. One can characterize the nature of the noise as being uncorrelated or correlated and then apply frequency filtering combined with multivariate endpoint detection to remove these noise sources accordingly. We have developed a flow chart for determining how to achieve the optimal multi-wavelength endpoint detection depending on which noise source is most dominant as shown in Figure 19. We would expect that for many etching processes improvements in endpoint detection sensitivity using multi-wavelength analysis should be similar to that demonstrated in this paper.

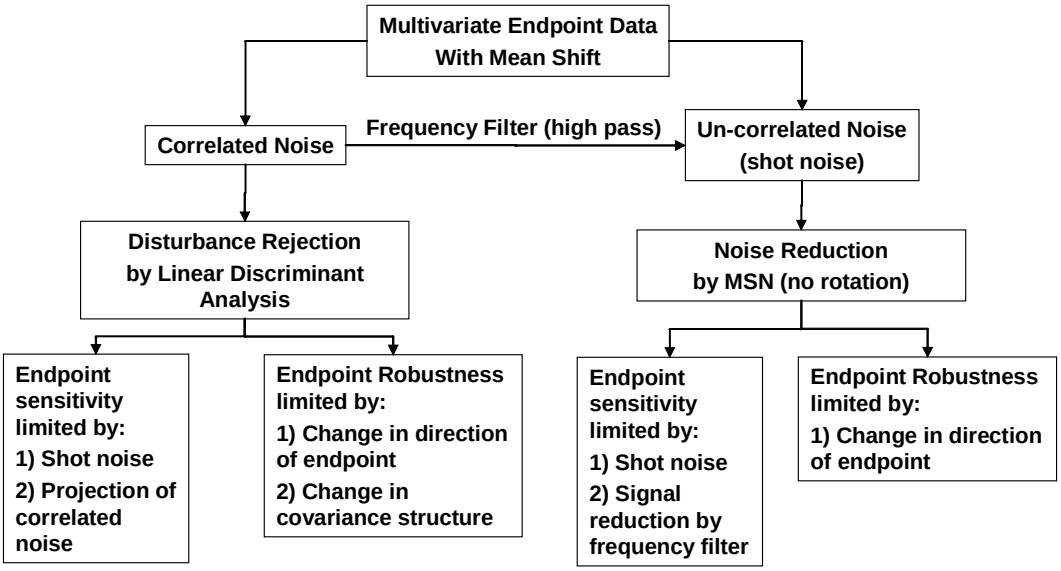


Figure 19. Summary of endpoint detection results / scheme for optimally detecting endpoint.

CONCLUSIONS

In this paper, we have demonstrated how multi-wavelength endpoint detection can be used to improve the sensitivity of endpoint detection over traditional methods by a factor of 5-6 using a quantitative signal-to-noise ratio for determining the sensitivity. Such an

improvement is obtainable when properly combining together frequency filtering and an optimal multi-wavelength statistic that uses prior knowledge of endpoint to enhance the detection sensitivity. Understanding the nature of the noise sources is key to proper implementation of both frequency filters and multi-wavelength strategies.

Multivariate noise sources have been characterized as either correlated noise which arises from variations in the process or uncorrelated noise which arises in the OES sensor system. These two noise sources are fundamentally different and must be removed by different mechanisms: namely, disturbance rejection to remove correlated noise and noise reduction to remove uncorrelated noise. The ability of multi-wavelength algorithms to improve the endpoint detection sensitivity depends heavily on the ability to remove these various noise sources. We have characterized several of the commonly used algorithms according to their ability to remove both uncorrelated and correlated noise, revealing that the MSN statistic developed in this work and the discriminant analysis approach, first described in this work, provide the only means to remove both noise sources. Nevertheless, frequency filters can be used to remove both correlated and uncorrelated noise as well, and can be used in conjunction with multi-wavelength strategies to enhance the overall sensitivity to endpoint detection.

Several strategies for endpoint were compared and the methods developed in this paper proved superior to both the single wavelength traditional method for endpoint detection and other multi-wavelength strategies suggested in the literature and used in commercial products. In fact, several other methods, such as Hotelling's T^2, EWFA, and PCA based methods often provided worse sensitivity to endpoint detection than the traditional method with a single wavelength, provided that the best single wavelength is chosen. For our experimental data, the dominant noise sources were a slow drift, correlated noise associated with heating of the reactor wall during the etching process and photon shot noise (uncorrelated noise) at the detector. The best result was obtained when using a high-pass filter to remove the drift combined with the MSN statistic without rotation to optimally remove the noise by using prior knowledge of endpoint obtained from a training wafer. This resulted in the improvement by a factor of 5-6. In a general plasma process, one must carefully consider the dominant noise sources that are present and then appropriately apply the multi-wavelength strategy according to the flowchart given in Figure 19. We believe that similar improvements in endpoint detection sensitivity can be obtained for most plasma processes, if the algorithms are applied properly.

ACKNOWLEDGEMENTS

This work was completed under Semiconductor Research Corporation Task 704.001.

REFERENCES

[1] R. G. Poulsen, J. Vac. Sci. Technol., **14**(1), 266 (1977).

[2] B. Wangmaneerat, T. M. Niemczyk, G. G. Barna, and D. M. Haaland, *Plasma Processing (9th International Symposium)*, G. S. Mathad, and D. W. Hess, Eds. p. 115, (1992).

[3] E. A. Hudson, and F. C. Dassapa, *Plasma Etching Processes for Sub-Quarter Micron Devices - 196th Meeting of the ECS*, Honolulu (1999).

[4] J. Hosch, F. G. Celii, C. Huffman, and K. Harvey, *AEC/APC XIII Symposium*, p. 825, Banff, Canada (2001).

[5] B. E. Goodlin, Ph.D. Thesis, MIT, Cambridge, MA (2002).

[6] B. E. Goodlin, D. S. Boning, and H. H. Sawin, *Presented at 46th International Symposium of the American Vacuum Society*, Seattle, Washington (2000).

[7] B. E. Goodlin, H. H. Sawin, and D. S. Boning, *AEC/APC Symposium XIII*, p. 767, Banff, Canada. (2001).

[8] D. White, B. Goodlin, A. Gower, D. Boning, H. Chen, H. Sawin, and T. J. Dalton, IEEE Trans. Semicond. Manufact., **13**(2), 193 (2000).

[9] M. E. Welch, P. E. Luscher, Z. Sui, and B. Lee, Semiconductor International, **19**(8), 269 (1996).

[10] P. Biolsi, D. Morvay, L. Drachnik, and S. Ellinger, Solid State Techol., **39**(12), 61 (1996).

[11] G. Chang, IEEE Electron Device Letters, **5**(12), 514 (1984).

[12] H. E. Litvak, J. Vac. Sci. Technol. B, **14**(1), 516 (1996).

[13] S. Limanond, J. Si, and Y.-L. Tseng, J. Vac. Sci. Technol. B, **16**(5), 2707 (1998).

[14] E. A. Rietman, J. T.-C. Lee, and N. Layadi, J. Vac. Sci. Technol. A, **16**(3), 1449 (1998).

[15] H. C. Sun, V. Patel, B. Singh, C. K. Ng, and E. A. Whittaker, Applied Physics Letters, **64**(21), 2779 (1994).

[16] H. Yue, S. Qin, J. Wiseman, and A. Toprac, J. Vac. Sci. Technol. A, **19**(1), 66 (2001).

[17] J. Zimpel, K. Voigtlander, and D. Knobloch, *AEC/APC XIII Symposium*, p. 1573, Banff, Canada (2001).

[18] R. l. Allen, R. Moore, and M. Whelan, J. Vac. Sci. Technol. B, **14**(1), 498 (1996).

[19] K. H. Chen, Thesis, MIT, Cambridge, MA (2001).

[20] C. K. Hanish, J. W. Grizzle, H. H. Chen, L. I. Kamlet, S. I. Thomas, F. L. Terry, Jr., and S. W. Pang, J. Electron. Matls., **26**(12), 1401 (1997).

[21] S. I. Thomas, H. H. Chen, C. K. Hanish, J. W. Grizzle, and S. W. Pang, J. Vac. Sci. Technol. B, **14**(4), 2531 (1996).

[22] A. D. Weiss, Semiconductor International, **6**(9), 98 (1983).

[23] M. S. Le, M.S. Thesis, MIT, Cambridge (1997).

[24] W. Branagh, J. Rivers, and R. Fry, *SPIE Conference on Advanced Sensors and Monitors for Process Industries and the Environment*, p. 317, Boston, MA (1998).

[25] J. J. Chambers, K. Min, and G. N. Parsons, J. Vac. Sci. Technol. B, **16**(6), 2996 (1998).
[26] A. P. Day, D. Field, D. F. Klemperer, and Y. P. Song, Semiconductor International, **12**(12), 110 (1989).
[27] X. Li, M. Schaepkens, G. S. Oehrlein, R. E. Ellefson, L. C. Frees, N. Mueller, and N. Komer, J. Vac. Sci. Technol. A, **17**(5), 2438 (1999).
[28] C. Almgren, Semiconductor International, **20**(9), 99 (1997).
[29] H. L. Maynard, E. A. Rietman, J. T. C. Lee, and D. E. Ibbotson, J. Electrochem. Soc., **143**(6), 2029 (1996).
[30] J. P. Roland, P. J. Marcoux, G. W. Ray, and G. H. Rankin, J. Vac. Sci. Technol. A, **3**(3), 631 (1984).
[31] D. Angell, and G. S. Oehrlein, Applied Physics Letters, **58**(3), 240 (1991).
[32] F. Heinrich, H. P. Stoll, H. C. Scheer, and P. Hoffman, *Multichamber and In-Situ Processing of Electronic Materials*, p. 185, Santa Clara, CA (1989).
[33] K. Wong, D. S. Boning, H. H. Sawin, S. W. Butler, and E. M. Sachs, J. Vac. Sci. Technol. A, **15**(3), 1403 (1997).
[34] N. Layadi, S. J. Molloy, T. C. Esry, T. Lill, J. Trevor, M. N. Grimbergen, and J. Chinn, J. Vac. Sci. Technol. B, **17**(6), 2630 (1999).
[35] V. Patel, M. Patel, S. Ayyagari, W. F. Kosonocky, D. Misra, and B. Singh, Applied Physics Letters, **59**(11), 1299 (1991).
[36] P. J. Marcoux, and P. D. Foo, Solid State Techol., 115 (1981).
[37] P. A. Lachenbruch, *Discriminant Analysis*, Hafner Press, New York (1975).
[38] *ISA's: Guide for Spectroscopy*, (1994).
[39] B. E. Goodlin, D. S. Boning, H. H. Sawin, and M. Yang, *Plasma Processing XII*, G. S. Mathad, and D. W. Hess, Eds. p. 100, Toronto (2000).
[40] G. E. P. Box, G. M. Jenkins, and G. C. Reinsel, *Time Series Analysis: Forecasting and Control*, 3rd ed., Prentice-Hall Inc., Englewood Cliffs, NJ (1994).
[41] C. J. Spanos, H. F. Guo, A. Miller, and J. Levine-Parrill, IEEE Trans. Semicond. Manufact., **5**(4), 308 (1992).