Abstract

This paper discusses the development of a high-accuracy endpointing algorithm for the emitter etch of a heterojunction bipolar transistor (HBT). Fabrication of high-performance HBT's using self-aligned base-emitter processes requires etching through the emitter layer and stopping with very high accuracy on the base layer. The lack of selectivity in dry etching coupled with the high etch rates possible in high density plasmas render the use of a standard timed overetch impractical, especially as device layers continue to become thinner. The etch process under study requires the complete removal of an AlInAs emitter while etching no more than 5 nm of the underlying GaInAs base layer. Etch products are monitored using optical emission spectroscopy (OES) to determine etch endpoint. The process under study relies on the intensity of the 417.2 nm Ga emission line. The detection of the Ga line indicates that the etch has reached the GaInAs layer. However, the presence of a time-varying Ga baseline signal before endpoint and significant noise in the OES signal necessitate more than a simple threshold scheme for critical endpoint detection. The algorithm presented here is based on a generalized likelihood ratio (GLR) with a signature function. This algorithm is robust to variance in the optical gains of the measurement equipment and is applicable to other etch processes. Experimental results of automated endpointing using this algorithm are presented in the form of pre- and post-etch ex situ film thickness measurements.

Modeling and Algorithm Development for Automated Optical Endpointing of an HBT Emitter Etch

C. K. Hanish   J. W. Grizzle   H.-H. Chen   L. I. Kamlet
S. Thomas III   F. L. Terry, Jr.   S. W. Pang

June 21, 1997

1 Introduction

As device layers become thinner, the lack of selectivity in dry etching, coupled with the high etch rates possible in high density plasmas, are rendering the use of a standard timed overetch impractical. An endpoint detection scheme is necessary to stop the etch at the correct time without unacceptable amounts of overetch of the underlying layer.

Several etch endpoint detection schemes are documented in the literature. Reflectometry [1], tunable diode laser spectroscopy [2], and neural nets [3] have been used successfully. A slope-sensitive endpoint detector has also been used on a ratio of etch product and background optical emission signals [4]. The present work differs from these studies in that no extensive modeling of the material is required, the single optical emission measurement employed requires no complex alignment, the algorithm is based on a simple model of the system, and the free parameters of the model can be tuned with a few experimental runs.

The goal of this research was to successfully detect the endpoint of the emitter etch for a heterojunction bipolar transistor (HBT). During this etch, which takes about 4.5 minutes, the unmasked portion of the AlInAs emitter layer must be completely removed while etching no more than 5 nm of the underlying GaInAs base layer. These layers top an InP substrate. Because both layers contain indium and arsenic and because the selectivity between layers is nearly unity, the only appropriate emission signals to monitor for endpoint detection are aluminum and gallium. Because the Ga emission rise at 417.2 nm is much more apparent than the drop in the Al signal at 396.1 nm, the gallium signal is monitored for endpoint detection using a monochromator, photomultiplier tube, and transimpedance amplifier.

During the first part of the etch, the emission at 417.2 nm represents background noise, which we have observed to be fairly linear during the course of an etch. The slope of the background trend, including its direction, changes with the state of the reactor walls. When the etch reaches completion, the emission at 417.2 nm rises from the background level due to the presence of Ga in the plasma, as shown in Figure 1. When the amount of excited gallium reaches steady state, the signal flattens.

2   Experimental Setup

These etches were performed in an ECR source etcher with an rf-powered stage described previously [5] using 50 W microwave power, 100 W rf power, 3 sccm Cl_2, 27 sccm Ar, and at 2 mTorr. The movable stage was set at 13 cm from the ECR source. The optical emission signal was passed through an optical fiber to a SPEX 500 1/2-meter monochromator, where the 417.2 nm emissions were isolated. A Hamamatsu R636 photomultiplier tube converted the light to current pulses which were collected by a transimpedance amplifier for conversion to an analog voltage. A PC running LabVIEW and equipped with a National Instruments MIO board collected the voltage signal at 2 Hz and ran the data through the algorithm in real time. When the etch endpoint was detected, a digital signal was sent to the sequential process controller, a Techware system, to automatically stop the etch.

Because the wafers required for these experiments were quite expensive, they were broken into small squares, about 1.2 cm on each side, and one sample was used for each run. Due to slight variations in sample size and run-to-run variations in optical gains of the measurement

system, the absolute signal intensity cannot be used in the design of an endpoint algorithm. In other words, an algorithm which stops the etch when the intensity reaches a preset level would be inappropriate.

3   State-Space Model for Endpoint Detection

As shown in Figure 1, the typical emission signal during the etch consists of a linear background trend with a rise when the interface is reached. When the background trend is subtracted from the data, the endpoint detection problem can be described with a simple discrete-time state-space model:

$$Ga(k+1) \;=\; (1 - \gamma\tau)Ga(k) + W(k) + v\,\Upsilon_x(k, t_0)\tau \tag{1}$$

$$Y(k) \;=\; Ga(k) + V(k), \tag{2}$$

where $Ga(k)$ is the amount of gallium in the plasma, in the same arbitrary units as the measurement, at time k, γ is the instantaneous rate of loss of Ga from the plasma, τ is the sample time, $W(k)$ and $V(k)$ are independent Gaussian white noise sequences with variances Q_0 and R_0, respectively, and $Y(k)$ is the emission measurement at time k. The variances of the noise sequences are unknown and must be estimated during the etch from the data. The change time, at which the interface is reached and gallium first starts coming from the surface, is denoted t_0. The function $v\Upsilon_x$ is the instantaneous rate of gallium being etched

from the surface and entering the plasma; it can be described as

$$
v \Upsilon_x(k, t_0) = \begin{cases} 0 \text{ for } k < t_0 \\ v \frac{k - t_0}{t_s} \text{ for } t_0 \leq k < t_0 + t_s \\ v \text{ for } k \geq t_0 \end{cases}, \tag{3}
$$

where v is an unknown parameter depending on the exposed area, etch rate, and optical gain of the system, and t_s is an unknown parameter indicating the time needed to clear the AlInAs layer once the GaInAs layer is reached. This "slope time," during which the amount of Ga coming from the surface linearly increases, exists due to layer thickness or etch nonuniformity or surface roughness which causes both layers to be etched simultaneously.

The loss coefficient, γ, was experimentally determined to be 0.125 at the operating conditions used for the etch. The background trend can be subtracted from the data in real time by monitoring the emission signal for the first few minutes of the etch (not including the time it takes for the emission signal to settle once the etch is started), fitting a line through this data well before the endpoint is expected, and subtracting this line from subsequent data points. An estimate $\hat{R}_N$ of the measurement noise variance can be formed by summing the squares of the differences between the data points and the estimated trend and dividing this sum by the number of data points minus two [6]. If the data is filtered to eliminate measurement noise, the maximum error between the filtered data and the background trend will approximately represent a 3-sigma level for the state noise; the state noise variance can thus be roughly estimated by dividing the maximum error by three and squaring the result.

Note that the gallium emission signal will start to rise before the AlInAs layer completely

clears, as shown in Figure 2. Care must be taken not to stop the etch too soon.

4   Model Validation

Three emission signals are shown in Figure 3 along with their best fits from the model. For the first signal, the etch was stopped manually approximately three seconds after the rise in the Ga emission signal was visually detected. In the second plot, the etch was stopped after about 6 seconds, and in the third plot, 12 seconds. Table 1 shows the model parameters used for the fits, the times at which the etches ended according to the model, and the times at which the etches were actually stopped. Note that for the first sample the etch was stopped before the end of the slope time determined by the model, and it is therefore expected that some of the AlInAs layer will remain on the surface. The other two samples should have no aluminum left on the surface. X-ray photoelectron spectroscopy (XPS) was performed on the post-etch samples to analyze the elements present on the surface, and the percentage of aluminum detected is given in the table. As expected, there is a significant amount of Al on the first sample, whereas the amount measured on the other two samples is small enough to be attributable to measurement noise. The plots in Figure 3 along with the data in Table 1 suggest that the model describes the process well enough for endpoint detection.

Uncorrelated Gaussian white noise on the measurement is a valid assumption under certain conditions. Figure 4 shows the frequency components of the noise on a typical Ga signal, collected at 500 Hz, while etching GaInAs. The drop near 20 Hz is due to the presence of anti-aliasing filters. Thus, if data is collected at a frequency sufficiently lower than this,

the noise can indeed be considered white. Figure 5 shows a histogram of measured emission levels while monitoring a GaInAs etch. The near-Gaussian shape of the histogram suggests that Gaussian noise is a valid assumption. Figure 6 shows the autocorrelation of a detrended background emission signal sampled at 4 Hz. Each sample is only correlated with itself, and therefore uncorrelated noise is also a valid assumption.

5   Endpoint Algorithm Development

The endpoint algorithm discussed here assumes that the data forms an independent sequence. To ensure the validity of this assumption despite possible detrending errors, the detrended data is processed by a Kalman filter to form a sequence of innovations [7]. The innovation $\epsilon(k)$ at time k is the difference between the detrended measurement and the estimate of the state obtained from past data:

$$\epsilon(k) = Y(k) - \hat{G}a(k|k-1) \tag{4}$$

The estimate of the state is in this case initialized to zero because no gallium is expected in the chamber at the start of the etch. That is, $\hat{G}a(1|0) = 0$. Future estimates are obtained by using the variance of the innovation, $\Sigma(k)$, and the state estimation error variance, $P(k|k-1)$, to calculate a Kalman gain, $K(k)$. The Kalman gain, innovation, and past state estimate are used to calculate a current estimate of the state, which is in turn used to predict what the state will be on the next iteration. The state estimation error variance is also initialized

to zero and updated at each iteration. The relevant equations are:

$$\Sigma_k \;=\; P_{k|k-1} + \hat{R}_N \tag{5}$$

$$K_k \;=\; \frac{P_{k|k-1}}{\Sigma_k} \tag{6}$$

$$P_{k|k} \;=\; (1 - K_k)P_{k|k-1} \tag{7}$$

$$P_{k+1|k} \;=\; (1 - \gamma\tau)^2 P_{k|k} + \hat{Q}_{N,f} \tag{8}$$

$$\epsilon_k \;=\; Y_k - \hat{G}a_{k|k-1} \tag{9}$$

$$\hat{G}a_{k|k} \;=\; \hat{G}a_{k|k-1} + K_k\epsilon_k \tag{10}$$

$$\hat{G}a_{k+1|k} \;=\; (1 - \gamma\tau)\hat{G}a_{k|k}, \tag{11}$$

where the time, k, has been moved into the subscript for easier notation.

A generalized likelihood ratio (GLR) algorithm with a signature function is used to detect the initial increase in the innovations sequence which reflects the Ga emissions signal rise caused by the change $v\Upsilon_x(k,t_0)\tau$ on the state. The log-likelihood ratio is defined as [7]

$$s_i = \ln\frac{p_{\theta_1}(\epsilon_i)}{p_{\theta_0}(\epsilon_i)}, \tag{12}$$

where θ_i is the mean of the signal before or after the change, the parameter after change is $\theta_1 = \delta$, the parameter before change is $\theta_0 = 0$, and ϵ_i is the innovation. Before the change, the expected value of the log-likelihood ratio is negative, while after the change the expected value is positive. For the general case of detecting a change in the mean of an independent Gaussian signal, the probability density is

$$p_\theta(\epsilon) = \frac{1}{\sqrt{2\pi\Sigma}}e^{-\frac{(\epsilon-\theta)^2}{2\Sigma}}, \tag{13}$$

7

yielding, for the present model,

$$s_i = \frac{\theta_1 - \theta_0}{\Sigma_i}\left(\epsilon_i - \frac{\theta_1 + \theta_0}{2}\right) \tag{14}$$

$$= \frac{\delta}{\Sigma_i}\left(\epsilon_i - \frac{\delta}{2}\right). \tag{15}$$

If a cumulative sum $S_1^k = \sum_{i=1}^{k} s_i$ is calculated, the sum will drift downward before the change and back upward after the change. The current minimum is given as

$$m_k = \min_{1 \le j \le k} S_1^j, \tag{16}$$

and thus a decision rule, g_k, can be calculated as

$$g_k = S_1^k - m_k \tag{17}$$

$$= \max_{1 \le j \le k}\left(S_1^k - S_1^j\right) \tag{18}$$

$$= \max_{1 \le j \le k} S_j^k, \tag{19}$$

where $S_j^k = \sum_{i=j}^{k} s_i$. When g_k exceeds a certain threshold h, it is likely that the change has occurred and the etch should be stopped. Appropriate choices for h depend on the signal-to-noise ratio, allowable mean time between the change and detection, and other parameters discussed by Basseville and Nikiforov [7].

Of course, because of the system and Kalman filter dynamics, a sudden step on the state will generally not result in a step change on the innovations. In addition, even the state does not change by a step in this endpoint application. To account for this, the $\theta_1 - \theta_0$ and $\theta_1 + \theta_0$ terms in (14) are replaced by a time-varying signature term, $\rho(k, t_0)$, describing the effect of the change on the innovations. The state, its estimate from the Kalman filter, and

8

the innovation can be decomposed as [7]

$$Ga_k = Ga_k^0 + \alpha(k, t_0), \tag{20}$$

$$\hat{Ga}_{k|k} = \hat{Ga}_{k|k}^0 + \beta(k, t_0), \tag{21}$$

$$\epsilon_k = \epsilon_k^0 + \rho(k, t_0), \tag{22}$$

where the superscript denotes the values of the unchanged model and the α, β, and ρ terms represent the effect of the change $v\Upsilon_x(k, t_0)\tau$ at time $t_0 \leq k$. Using the initial conditions $\alpha(t_0, t_0) = 0$ and $\beta(t_0 - 1, t_0) = 0$, the functions are computed recursively as

$$\alpha(k, t_0) = (1 - \gamma\tau)\alpha(k - 1, t_0) + v\Upsilon_x(k - 1, t_0)\tau, \tag{23}$$

$$\beta(k, t_0) = (1 - K_k)(1 - \gamma\tau)\beta(k - 1, t_0) + K_k\alpha(k, t_0), \tag{24}$$

$$\rho(k, t_0) = \alpha(k, t_0) - (1 - \gamma\tau)\beta(k - 1, t_0). \tag{25}$$

Because the signature takes into account the model and the Kalman filter gains, the change on the state, v, is used, rather than the change on the innovations, δ.

The log-likelihood function computes whether it is likely that the mean shifted to θ_1 at some time $t_0 \leq k$. The idea now is to have the log-likelihood function compute whether it is likely that the innovations started tracking the signature function $\rho(k, t_0)$ at some time $t_0 \leq k$. The function $\rho(k, t_0)$ is incorporated into the log-likelihood function in place of θ_1. With $\theta_0 = 0$, the decision function thus becomes

$$g(k) = \max_{1 \leq j \leq k} S_j^k, \tag{26}$$

$$S_j^k = \sum_{i=j}^{k} \frac{\rho(i, j)\epsilon_i}{\Sigma_i} - \frac{1}{2}\sum_{i=j}^{k} \frac{\rho(i, j)^2}{\Sigma_i}. \tag{27}$$

9

Because the change magnitude is not known, a double maximization must be performed to calculate the decision function g_k:

$$g_k = \max_{1 \le j \le k} \sup_{\theta_1} S_j^k(\theta_1). \tag{28}$$

Because this algorithm considers the most likely value of θ_1, it is often referred to as the maximum likelihood ratio (MLR) or generalized likelihood ratio (GLR). With the unknown change magnitude, the signature function is decomposed as $v\tilde{\rho}(k, t_0)$, where $\tilde{\rho}(k, t_0)$ is the signature resulting from a unit change on the state (i.e. $v = 1$), yielding

$$\sup_v S_j^k = \hat{v}_k(j) \sum_{i=j}^{k} \frac{\tilde{\rho}(i,j)\epsilon_i}{\Sigma_i} - \frac{\hat{v}_k^2(j)}{2} \sum_{i=j}^{k} \frac{\tilde{\rho}^2(i,j)}{\Sigma_i}, \tag{29}$$

$$\hat{v}_k(j) = \text{sign}\left[\frac{\sum_{i=j}^{k} \tilde{\rho}(i,j)\epsilon_i/\Sigma_i}{\sum_{i=j}^{k} \tilde{\rho}^2(i,j)/\Sigma_i}\right] \left(\sup\left[0, \left|\frac{\sum_{i=j}^{k} \tilde{\rho}(i,j)\epsilon_i/\Sigma_i}{\sum_{i=j}^{k} \tilde{\rho}^2(i,j)/\Sigma_i}\right| - v_m\right] + v_m\right). \tag{30}$$

6 Monte Carlo Analysis

Because of the expense involved in etching the samples, it would be prohibitively expensive to tune all of the parameters in the algorithm on the basis of data alone. Therefore, simulated data sequences were created using the model and uniformly distributed parameters whose upper and lower bounds were determined on the basis of a few etches; see Table 2. The ranges were chosen to be somewhat wider than values observed in the actual data sequences, and a sampling rate of 4 Hz was assumed. Each data sequence was formed by summing the state value obtained from the model for the chosen parameters, a Gaussian noise sequence formed by choosing normally distributed numbers with zero mean and chosen measurement noise variance, and a background trend created from the chosen slope and intercept.

The GLR algorithm with the signature function was tested on 25 simulated sequences using a slope time of 8 seconds to calculate the signature function and a minimum change magnitude of 0.005. The resulting decision functions are shown in Figure 7. The decision functions have been shifted in time so that in each sequence, the change on the state occurs at a time of zero seconds. The appropriate time to stop the etches thus occurs at about ten seconds, the maximum slope time. For the sequences with the minimum slope time, six seconds, this would result in an overetch of about 4 nm, assuming the fastest observed etch rate of 48 nm/min. This overetch is within the 5 nm limit.

To choose the threshold, h, for a particular algorithm, one must draw a horizontal line through the decision functions in such a way that most of the decision functions cross the horizontal line at approximately ten seconds. This is obviously impossible because the decision functions have widely varying slopes. Fortunately, there are a few seconds between the change time and when the etch must be stopped. The change time can be estimated as [7]

$$\hat{t}_{0_k} = \arg \max_{1 \le j \le k} S_j^k, \tag{31}$$

where $\arg \max$ is the value of j (the sample index) which maximizes S_j^k. Therefore a low threshold can be set, and whenever the decision function exceeds the threshold, the change time is estimated. If the change occurred approximately ten seconds ago (or some other selected "lag time"), the etch is stopped.

For a given threshold and lag time, the etch stop time can be determined for each decision function. The overetch time is then calculated as the difference between the stop time and the sum of the change and slope times for each sequence. A range of overetch depths is then

11

calculated from the overetch time and the minimum and maximum etch rates. The optimum threshold and lag time can be determined by calculating a cost function for various thresholds and lag times and minimizing this cost. The cost for each sequence was calculated twice, once using the maximum overetch depth and once using the minimum, from the following, somewhat arbitrary, cost function:

$$\text{cost} = \begin{cases} 0 & \text{if} & 0 \text{ nm} \leq \text{depth} < 3 \text{ nm} \\ \frac{5(\text{depth}-3)}{3} & \text{if} & 3 \text{ nm} \leq \text{depth} < 6 \text{ nm} \\ 5 & \text{if} & \text{depth} < 0 \text{ nm or depth} \geq 6 \text{ nm} \end{cases} \quad . \tag{32}$$

The cost for a given threshold and lag time is obtained by summing the costs of each of the sequences, and the optimum parameters will minimize this cost. For the simulated sequences, the optimum threshold was 1.1 and the optimum lag time was 11.5 seconds.

Further testing was done on 250 new simulated sequences using the same threshold and lag. The overetch depths are shown in Figure 8. This algorithm has a failure rate of only 5.2%, where a failure is defined as any etch where either the slowest or fastest etch rate result in an overetch not between 0 and 5 nm. Note that there are several etches very close to the 5 nm limit, and these devices may meet the performance specifications, so the failure rate may be closer to 3.6%. When the data collection rate is slowed to 2 Hz, the maximum rate at which the computer attached to the ECR can process the data using this algorithm in real time, the failure rate rises to 10%.

12

7   Experimental Results

The GLR algorithm using the precomputed signature function was tested during three AlInAs/ GaInAs etches. The measurement was filtered at 1 Hz using a low-pass filter in the transimpedance amplifier. This gave a higher signal-to-noise ratio [8], but simulations showed no significant effect on the cost. Due to the time needed to calculate the signature function, several of these functions were precomputed for various noise levels, and once the noise level was determined for a particular run, the correct file was read from disk. Simulations showed that this also had no significant effect on the cost.

The day-to-day variation in change and slope times for the etch process under study is currently too great to use the same endpoint algorithm parameters on every batch of runs. Therefore, to test this algorithm, two calibration runs were done at the beginning of the day. Monitoring the gallium emission line during the complete removal of the AlInAs layer and fitting the model parameters to these two runs showed that the change times were 371.6 seconds and 377.2 seconds, the slope times were 13.6 seconds and 11.8 seconds, and the noise variances were 0.005 and 0.003. Both runs had change magnitudes of about 0.12 and took about 150 seconds for the signal to settle before data could be collected to determine the background trend. The average etch rates for the AlInAs layer during these runs were 30.8 and 30.4 nm/min.

The threshold and lag time were optimized using 25 sequences at 2 Hz with random noise variances between 0.002 and 0.005, change times between 359.2 and 389.2 seconds, change magnitudes between 0.06 and 0.24, and slope times between 10.5 and 14.5 seconds. When

calculating the cost, the minimum and maximum etch rates were assumed to be 25 nm/min and 35 nm/min. The optimum lag time was found to be 15.5 seconds, and the cost was minimized when the threshold was between 4.0 and 8.5.

Three endpoint experiments were performed with a lag time of 15.5 seconds and threshold of 6.0. For the first 150 seconds of the etch, no action was taken, and data taken between 150 seconds and 359 seconds was used for the background trend calculation. Based on the calculated noise levels, an appropriate signature function matrix was read from a file. The size of the matrix allowed for the endpoint to be detected between 359 and 404 seconds, and if the endpoint did not occur in this window, the etch would not automatically stop.

The algorithm stopped the three etches at 398.0 seconds, 398.5 seconds, and 391.0 seconds, respectively. In the third run, the argon flow was set to 8 sccm instead of 27 sccm for the first 31 seconds of the run. This disturbance was corrected at 31 seconds and the gallium emission signal settled by 150 seconds into the run.

For each of the three samples, *ex situ* spectroscopic reflectometry measurements were taken at five positions on the sample before and after the etch so that the thickness of the GaInAs layer could be calculated. With spectroscopic reflectometry, white light is shone onto the surface of the wafer, usually at normal incidence. The reflectance is measured as a function of wavelength, and if a model relating these measurements to the optical properties of the layers is known, the thicknesses of the layers can be calculated [9]. The post-etch samples also have some surface roughness, calculated using a Beckman-Kirchoff model. The thickness of the roughness layer can be added to the thickness of the GaInAs layer to give

an indication of the total thickness. The results are shown in Table 3.

For all three samples, the nonuniformity before the etch was less than 2 nm and the nonuniformity after the etch was greater than 4 nm. Considering the 5 nm overetch specification, the post-etch nonuniformity is quite high. The roughness layer after the etch also exceeded the overetch specification in most cases, and therefore, although some of this roughness can be removed by rinsing the etched sample in water, the roughness during the etch should be reduced if possible. Considering these problems, the endpoint algorithm performed very well, with overetches ranging from -0.9 to 4.7 nm, where negative overetches probably indicate that the AlInAs was not thoroughly cleared.

After the samples were rinsed in DI water to remove soluble chloride compounds, the thicknesses were again measured and are tabulated in Table 4. Except where there are severe uniformity problems, the GaInAs overetch amounts are well within the 5 nm specification.

The average etch time required in the calibration runs was 387.1 seconds. Had a timed etch based on the calibration been used for the other three samples, the samples would have been etched 10.9, 11.4, and 3.9 seconds less, respectively. This would have resulted in average underetches of 0.4 nm and 1.8 nm for the first two runs and an average overetch of 1.9 nm for the third run. That is, two of the three etches would have failed due to insufficient clearing of the AlInAs layer.

8   Summary

An endpoint detection algorithm has been developed for etching AlInAs on GaInAs using intensity measurements of a gallium emission line. Preliminary etch results show that this algorithm is capable of stopping the etch with less than 5 nm overetch into the GaInAs layer when the etch does not have uniformity problems. Overetches well within the 5 nm specification were obtained on three different runs despite a disturbance on the argon flow in the third run. If the surface roughness and etch nonuniformity can be reduced, this algorithm shows great promise of being able to meet the overetch specifications.

9   Acknowledgements

Addresses of authors

C. K. Hanish is with Texas Instruments Inc., 13536 N. Central Expressway, MS 944, Dallas, TX 75243; J.W. Grizzle, H.-H. Chen, F.L. Terry, Jr. and S.W. Pang are with the Electrical Engineering and Computer Science Department, University of Michigan, Ann Arbor, MI 48109-2122; L.I. Kamlet is with Medtronic Micro-Rel, 2343 W. 10th Place, Tempe, AZ. 85281;

References

[1] Sonny Maung, Somnath Banerjee, Dennis Draheim, Steven Henck, and Stephanie Watts Butler. Integration of in situ spectroscopic ellipsometry with MMST machine control. *IEEE Transactions on Semiconductor Manufacturing*, 1994.

[2] H. C. Sun, V. Patel, B. Singh, C. K. Ng, and E. A. Whittaker. Sensitive plasma etching endpoint detection using tunable diode laser absorption spectroscopy. *Applied Physics Letters*, 1994.

[3] Ronald L. Allen, Randy Moore, and Mike Whelan. Application of neural networks to plasma etch end point detection. *Journal of Vacuum Science and Technology, B*, 14(1):498–503, January/February 1996.

[4] Herbert E. Litvak. End point control via optical emission spectroscopy. *Journal of Vacuum Science and Technology B*, 1996.

[5] S. W. Pang, K. T. Sung, and K. K. Ko. Etching of photoresist using oxygen plasma generated by a multipolar electron cyclotron resonance source. *Journal of Vacuum Science and Technology, B*, 1992.

[6] L. Ljung. *System Identification: Theory for the User*. Prentice-Hall, 1987.

[7] Michèle Basseville and Igor V. Nikiforov. *Detection of Abrupt Changes*. System Sciences Series. Prentice Hall, Inc., 1993.

[8] S. Thomas III, H. H. Chen, and S. W. Pang. Response time for optical emission and mass spectrometric signals during etching of heterostructures. In *Materials Research Society*

Symposium Proceedings, volume 406, pages 27–32. Materials Research Society, 1996.

[9] Leonard I. Kamlet. *Phenomenological modeling of the optical properties of semiconductors for process characterization and control.* PhD thesis, University of Michigan, 1995.
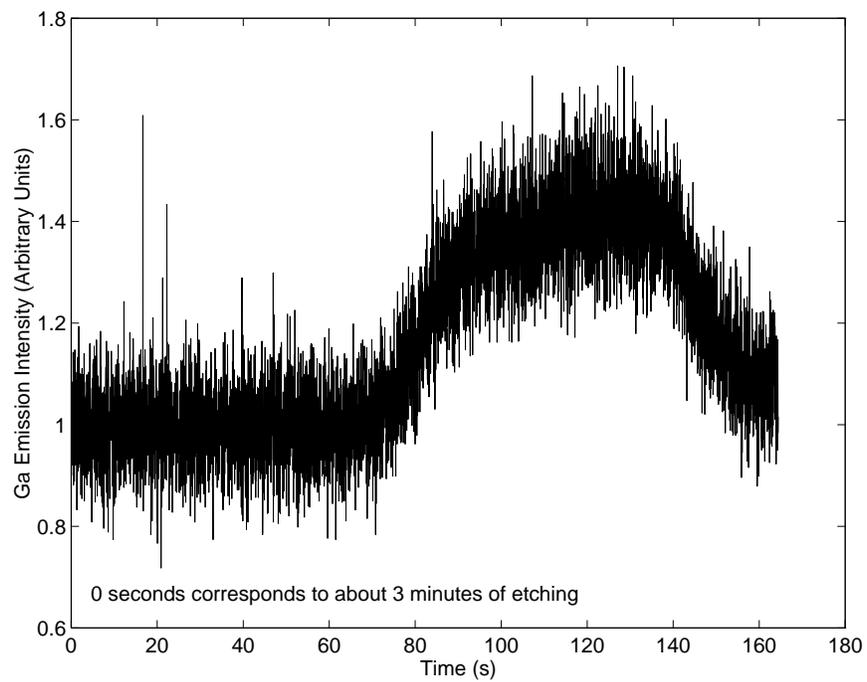
Figure 1: Typical Ga emission signal while etching through the emitter and base. Data collection was started approximately 3 minutes into the etch.
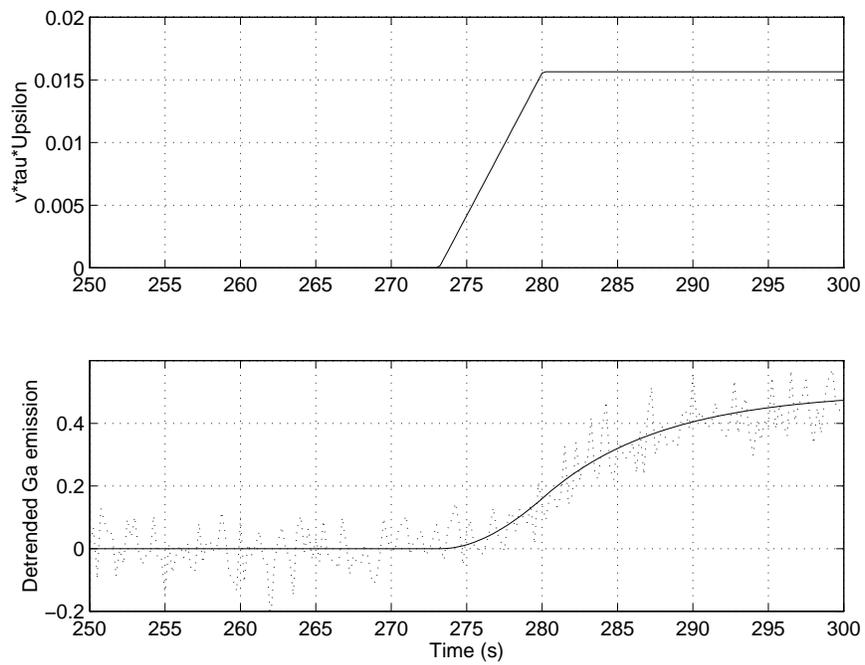
Figure 2: Gallium coming from the surface and the resulting emission intensity signal.
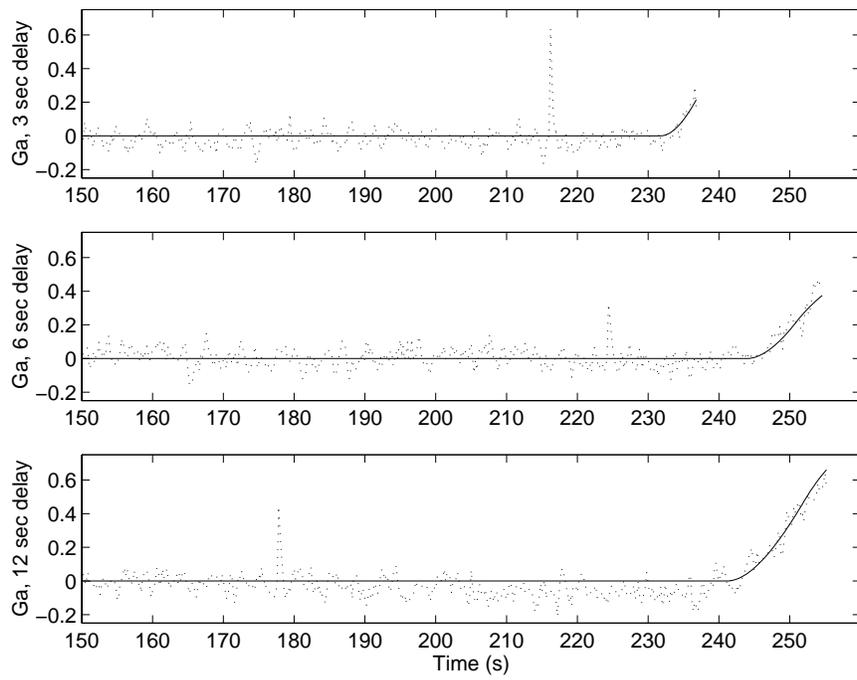
Figure 3: Actual data and fits with modified model for etches stopped approximately 3, 6, and 12 seconds after the rise in Ga

Table 1: Comparison of model fit to actual data with XPS results. For each data sequence, the change magnitude, change time, and slope time are chosen to minimize the least squared error between the model and actual signals. The etch end time is the sum of the change and slope times, and the etch stop time is the time at which the etch was manually stopped. The % aluminum on the surface was measured with XPS. When the etch was stopped before the etch end time, the % Al is higher.

Approx. Delay	v	t_0	t_s	Etch End Time	Etch Stop Time	% Al
3	0.12	281.6	6.0	287.6	286.8	7.31
6	0.11	293.4	9.9	303.3	304.6	3.52
12	0.11	290.4	10.6	301.0	305.0	3.17
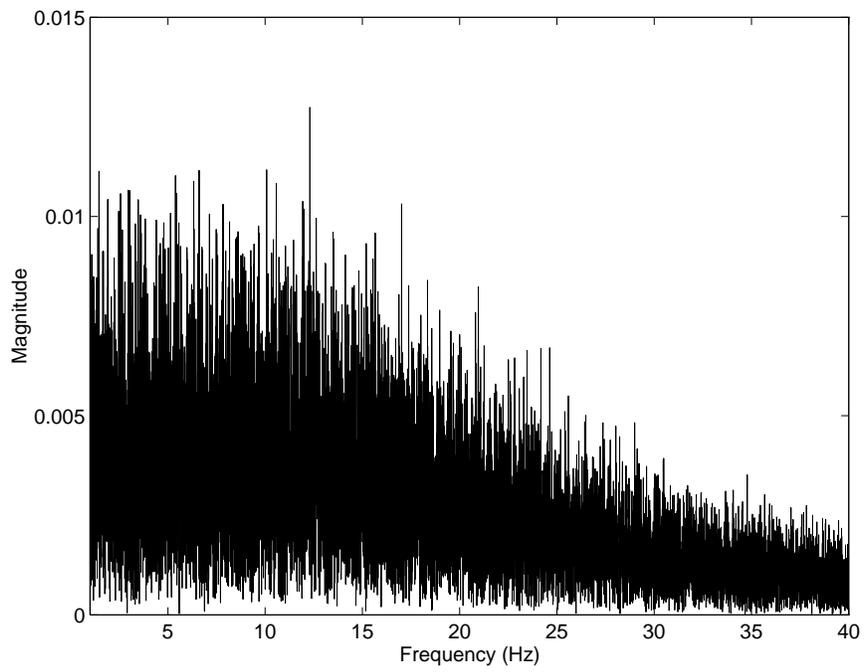
Figure 4: Fast Fourier Transform of the background noise on the Ga emission signal
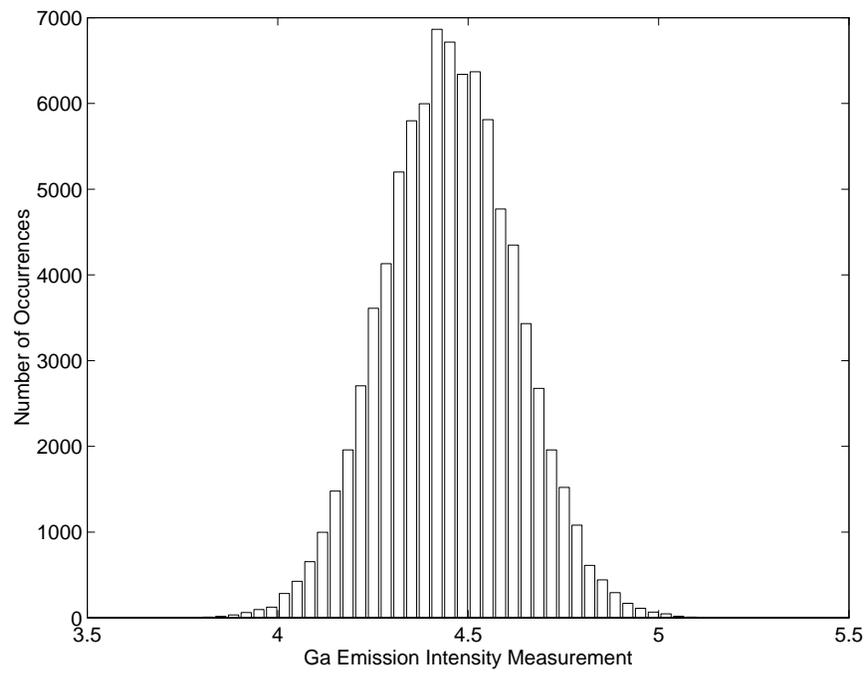
Figure 5: Histogram of values of the background noise signal. The values have an approximately Gaussian distribution.
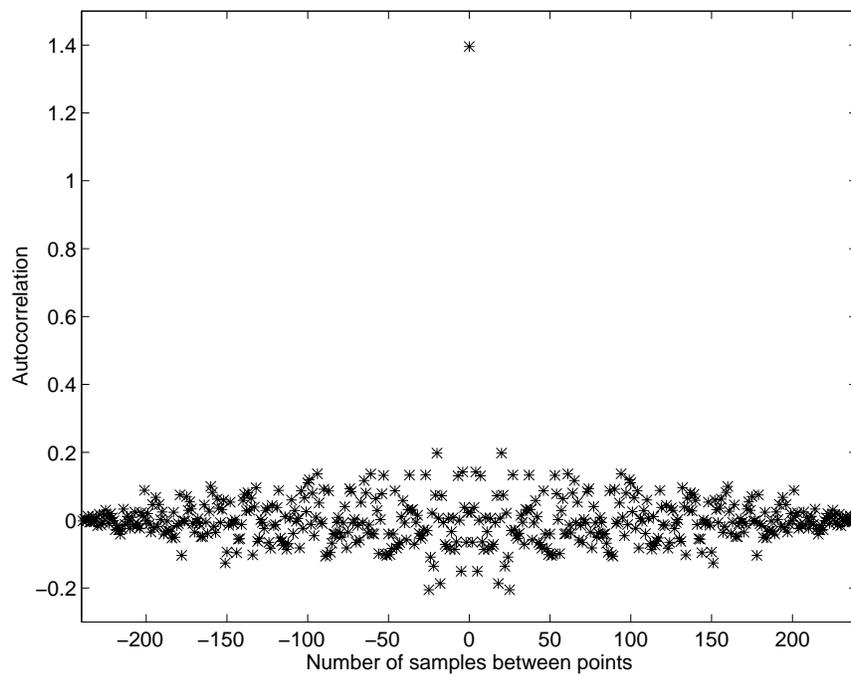
Figure 6: Autocorrelation of detrended background emission signal

Table 2: Parameters and ranges used to create simulated data sequences

Parameter	Minimum	Maximum
Noise variance, R_n	2×10^{-3}	8×10^{-3}
Background trend slope	-5×10^{-4}	5×10^4
Background trend intercept	0.5	5
Change magnitude, v	0.01	.2
Change time, t_0 (s)	255	285
Slope time, t_s (s)	6	10
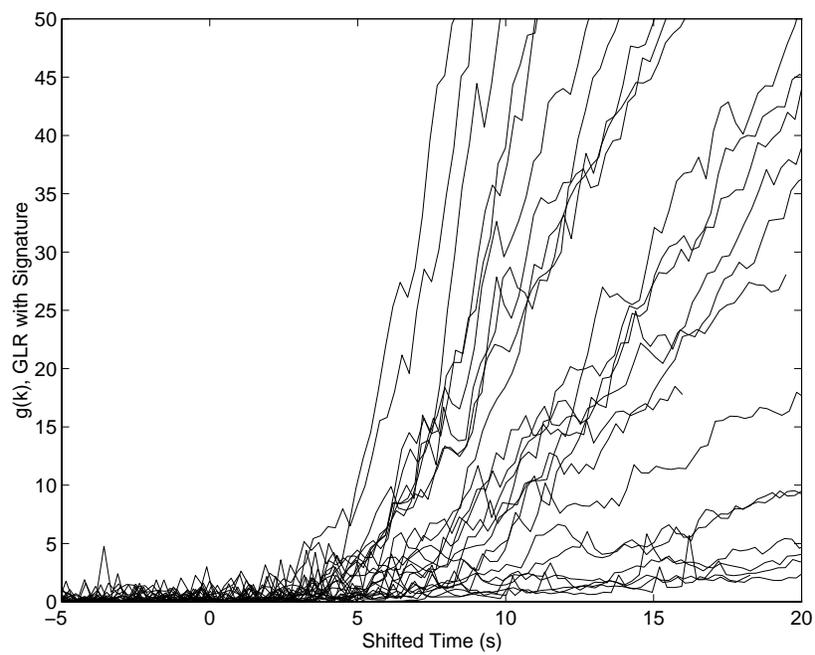
Figure 7: Decision function using GLR algorithm with signature function on all 25 data sequences
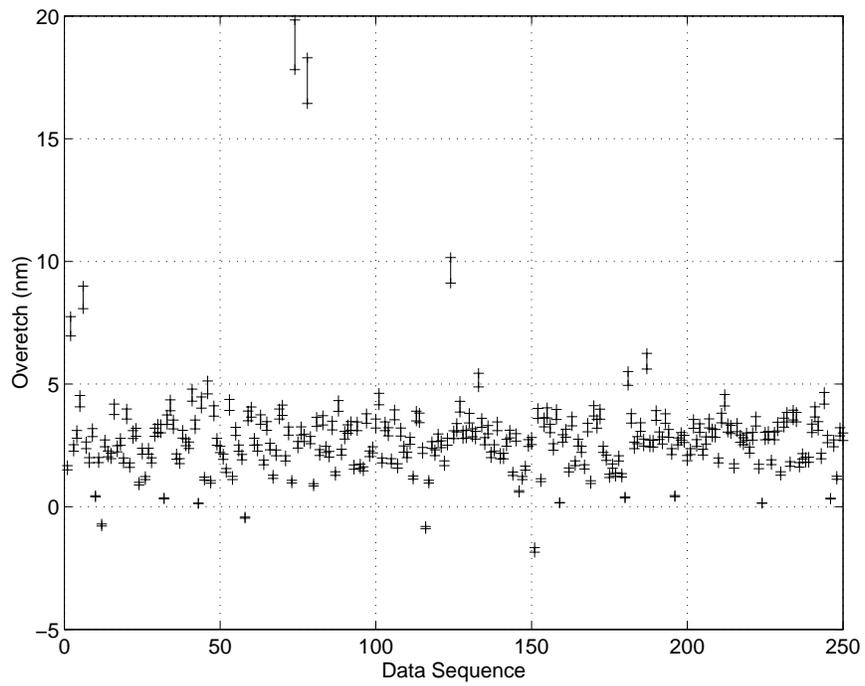
Figure 8: Overetch depths for 250 data sequences using the GLR algorithm with a signature function.

Sample	Position	Pre-etch GaInAs	Post-etch GaInAs	Post-etch Roughness	Total Post-etch Thickness	GaInAs Overetch
1	1	62.0	57.1	5.8	62.9	−0.9
1	2	62.6	57.2	6.0	63.2	−0.6
1	3	63.2	58.1	5.7	63.8	−0.6
1	4	63.1	54.6	4.7	59.3	3.8
1	5	63.5	55.2	5.8	61.0	2.5
2	1	62.6	56.8	6.7	63.5	−0.9
2	2	63.0	56.4	6.7	63.1	−0.1
2	3	63.0	56.4	5.8	62.2	0.8
2	4	62.1	53.6	5.0	58.6	3.5
2	5	62.7	54.9	4.8	59.7	3.0
3	1	61.9	55.4	3.0	58.4	3.5
3	2	62.8	55.1	3.0	58.1	4.7
3	3	63.7	55.8	5.0	60.8	2.9
3	4	62.3	58.0	4.7	62.7	−0.4
3	5	62.9	55.5	5.8	61.3	1.6

Table 3: Pre- and post-etch GaInAs thicknesses obtained using the endpoint algorithm. All thicknesses are given in nm and are accurate to within 1 nm.

Sample	Position	Pre-etch GaInAs	Post-rinse GaInAs	Post-rinse Roughness	Total Post-rinse Thickness	GaInAs Overetch	Average
1	1	62.0	56.9	3.8	60.7	1.3	
1	2	62.6	57.9	2.6	60.5	2.1	
1	3	63.2	55.2	3.0	58.2	5.0	4.5
1	4	63.1	52.8	2.9	55.7	7.4	
1	5	63.5	54.0	2.9	56.9	6.6	
2	1	62.6	56.0	3.7	59.7	2.9	
2	2	63.0	54.6	4.0	58.6	4.4	
2	3	63.0	54.5	3.7	58.2	4.8	3.8
2	4	62.1	55.0	3.6	58.6	3.5	
2	5	62.7	56.0	3.5	59.5	3.2	
3	1	61.9	57.1	2.8	59.9	2.0	
3	2	62.8	54.7	3.0	57.7	5.1	
3	3	63.7	57.6	2.8	60.4	3.3	3.2
3	4	62.3	57.1	2.8	59.9	2.4	
3	5	62.9	57.0	2.7	59.7	3.2	

Table 4: Pre-etch and post-rinse GaInAs thicknesses obtained using the endpoint algorithm. All thicknesses are given in nm and are accurate to within 1 nm.