Advances in Broadband RF Sensing for Real-time Control of Plasma-Based Semiconductor Processing

Craig Garvin, Dennis S. Grimard, and Jessy W. Grizzle

University of Michigan Electronics Manufacturing Laboratory,

3300 Plymouth Road, Ann Arbor, MI 48105-2551

garv@eecs.umich.edu, dgrimard@eecs.umich.edu, grizzle@eecs.umich.edu

Abstract

A novel sensing system based on plasma impedance spectroscopy is compared to standard RF metrology. The system uses an antenna in the glow discharge to excite the bulk plasma at a frequency range of $27.5 MHz$ to $2.75 GHz$. Standard RF metrology is implemented by measuring 1000 points of the RF power signal using a digital oscilloscope sampling at 1 GHz. An experiment varying power, pressure, Ar and O_2 is constructed. Using a subset of the data to regress a model, standard RF sensing reconstructs the experimental variables with a best average R^2 of 0.49 , whereas the novel sensing system results in a best average R^2 of 0.876. A nearest neighbor algorithm is used which results in 70% correct identification of process conditions for standard RF sensing, and 99.5% correct identification of process conditions for the novel sensing system.

I. INTRODUCTION

Ever shrinking geometries and larger substrates are mandating improvements in sensor systems for control and diagnostics of plasma processing. Experts in industry and academia have recognized the RF signal ($13.56 MHz$) and its harmonics as a potential source of process information. In

addition to using the existing RF signal as an excitation source, there has been considerable work done on plasma impedance analysis via dedicated high frequency test system, often referred to as plasma impedance spectroscopy. This paper continues the work of Garvin *et al.* [1] and presents a more thorough investigation of a unique interpretation of plasma impedance spectroscopy, referred to as 'broad band' sensing.

A number of researchers have used RF metrology as a tool for plasma diagnostics. Maynard *et al.* [2] have used RF metrology for end-pointing of an industrial etch process. Spanos *et al.* [3, 4] have extensively used RF metrology in their plasma diagnostic and control work. A number of researchers [5–8] have used ion flux information obtained via RF metrology to characterize etching. Researchers at the Adolph Slaby Institute [9, 10] have developed a diagnostic system that uses plasma physics models to infer process - relevant information from the fundamental frequency exciting the discharge and multiple harmonics resulting from plasma non linearities. Attempts to model the relationship between measured RF parameters and plasma physics have been presented in [11–13]. This list is by no means complete, but the common element is the attempt to relate measurement of the $13.56 MHz$ signal (and in some cases its harmonics) to plasma, wafer, and tool conditions.

An alternative to relying on the RF power as a source of excitation for diagnostic and sensing purposes is to use a low power high frequency source to excite the plasma over a frequency range typically far above the standard $13.56 MHz$ range. Typically, the high frequency source is coupled to the discharge by means of an antenna. It is possible to measure either reflection from this antenna or transmission to another antenna in the plasma [14]. The technique has been pursued both in contained [15] and atmospheric [16] plasmas. More recently, results with more modern equipment on more process relevant plasmas have been published by Lieberman *et al.* [17] and de Vries *et al.* [18]. All these approaches to plasma impedance spectroscopy share a common feature. Regardless of the amount of data taken at an operating point, only one parameter, typically the frequency of peak power absorption by the plasma, is considered relevant.

This single parameter approach contrasts with the multi parameter approach of plasma harmonic sensing. In principle, it is possible to infer multiple plasma parameters from the multi-

ple measured harmonics. Modern network analyzers and computer aided data acquisition methods make it relatively straight forward to collect the entire frequency signal generated in plasma impedance spectroscopy as well. Accordingly, it is possible to pursue a form of multi parameter plasma impedance spectroscopy. In fact, initial results suggest that the frequency range over which typical processing plasmas respond to excitation is on the order of several gigahertz. This range is much broader than the several hundred megahertz over which the plasma appears to produce significant harmonics. To underscore the frequency range of the modified plasma impedance spectroscopy, it is referred to as *broad band sensing* whereas methods relying on the plasma's RF signal and its harmonics are referred to as *narrow band sensing*.

This article investigates the use of broad band sensing in a micro-electronics processing plasma. The primary goal is to evaluate the relative observability of the state of the plasma under standard RF metrology (narrow band) and modified plasma impedance spectroscopy (broad band). More precisely, the aim of most plasma diagnostics is to *detect* a change in the plasma state and to *isolate* the source of this change. A first step towards this goal was presented in Ref. [1]. As in Ref. [1], a simple experiment is conducted to compare the abilities of the two measurement techniques to detect and isolate plasma changes due to the variation of generator power, chamber pressure, and gas chemistry. Here a more direct experiment is performed, with more complete analysis and more conclusive results.

II. EXPERIMENTAL SETUP

An experiment was designed in order to simultaneously collect broad band and narrow band data as a function of widely varying plasma conditions, as shown in Figure 1. Experiments were performed on a GEC research reactor, described in Ref. [19]. Power is delivered using an ENI generator and matching network, and generator power is measured using the built in ENI power meter. Pressure is measured using an MKS barratron, and gas flow rates are measured with MKS flow meters with MKS gas correction factors.

Narrow band sensing is implemented using a Werlatone D5281 $2.0MHz$ to $250MHz$ direc-

tional coupler rated at 2 kWatts power with a nominal $-50dB$ coupling between main line and sensor ports received by a Tektronix TDS 420 digital storage oscilloscope. Groups of 500 points of both forward and reverse waves are sampled at 1 Gs/sec, resulting in a total of about 7 periods of each wave. The entire waveform is logged for subsequent processing.

Broad band sensing is achieved by a mono-pole antenna constructed of a length of RG402u stainless steel rigid coaxial cable. Approximately one inch of center conductor is exposed to the plasma to act as an antenna. The probe is inserted in the bulk plasma using an O-ring compression sealed vacuum port. A Hewlett Packard 8753B vector network analyzer drives the resonance probe over a range of $27.5MHz$ to $2.75\,GHz$ at a power level of 0 dBM. A Mini Circuits $25MHz$ high pass filter is used to isolate the vector network analyzer from the discharge. After calibration, the complex reflection coefficient (Γ) is recorded at 201 frequency points linearly uniformly spaced between $27.5MHz$ and $2.75\,GHz$. The set-points of power, pressure and flow rate for the GEC as well as data acquisition are controlled with a PC running LabVIEW data acquisition and control software. All data is logged and written to file automatically.

III. EXPERIMENT AND INITIAL RESULTS

The goal of the experiment is to evaluate the ability of broad band and narrow band sensing to isolate basic plasma perturbations due to changes in power [1], pressure and chemistry. Accordingly, a full factorial experiment is performed, as summarized in Table I. For this investigation, chemistry is limited to Ar and O_2 in order to eliminate variability due to polymer build up in the chamber. Statistical methods are then used to construct a model to predict the variables listed in Table I based on measured broad band and narrow band data, respectively.

It is informative to consider the response of the two sensing systems to a variation in pressure. Figure 2 shows both the broad band and narrow band magnitude response to a change in pressure from $90\ mTorr$ to $100\ mTorr$ with power at a constant 100 W, and a chemistry of Ar at $10\ sccm$

[1] Power level refers to a generator set point and not a calculated delivered power to the plasma.

flow rate and O_2 at $5\ sccm$ flow rate. In the broad band signal, there is a clear and distinct trend differentiating the two pressure conditions. As a result of a relatively small change in pressure, the magnitude, location and even number of absorption peaks changes substantially. Though it remains to be seen whether this qualitative observation will be elicited in the statistical analysis, the structure of the response using the broad band sensor suggests that substantial information about fundamental plasma physics may be embedded in the sensor data. The narrow band signal is also reported in Figure 2. Because the difference between the two signals is so small, the difference itself is plotted. It is more difficult to discern a pattern in the narrow band response, but this by no means indicates that structure is lacking.

IV. EXPERIMENTAL ANALYSIS

The experiment described in Table I results in 16 different treatment combinations, each consisting of a specific level of the independent variables. For each treatment combination, 1000 points of narrow band and 201 complex points of broad band data are collected. The goal of the experimental analysis is to generate a model from each of the sensor systems to the plasma state as represented by power, pressure and chemistry. It is assumed that if the same methods are used in both cases, differences in fit can be attributed to fundamental differences in observability between the two systems. To present a direct comparison of methods, no additional transformations (such as using impedance or standing wave ratio representations) are used. The narrow band data is considered as time domain $\mathbf{V}^+$ and $\mathbf{V}^-$. The broad band data is considered as real numbers expressed as magnitude and phase of the reflection coefficient.

The experiment is repeated 6 times, for a total of 96 treatment combinations. For both narrow band and broad band response, a 'round robbin' approach is used for model evaluation. A *modeling set* composed of 5 repetitions is used to develop a model relating measured signal to factor levels. The remaining repetition forms the *testing set* which is used to evaluate the model. The process is repeated 6 times so that each repetition is used as the *testing set* and results are reported as an average of all 6 testing sets.

Two methods are used to evaluate the information content of the respective measured signal. One approach is a standard linear regression, as described in [20] or [21]. Since the dimension of response is larger than the number of experiments, subset selection must be used. Stepwise regression, as described in [22] is performed. Additionally, the data is analyzed using a 'nearest neighbor' algorithm, as described in [23]. Linear regression results in a globally defined function relating measurements to inputs. Given the non linear nature of plasma processing, the regression model can only be expected to be valid over a reasonably small subset of the operating space near the operating points. The nearest neighbor algorithm has a much more limited goal. It assumes that only the 16 plasma conditions used in the experiment are possible and given a new measurement determines which of the 16 'standard responses' it most closely approximates.

Stepwise Regression is a rigorous method for adjusting model size, using the partial F test [20] to evaluate whether the change in performance associated with the change in model size is statistically relevant. The specific implementation is as follows. The goal of the experiment is to build the largest model that produces a statistically significant performance improvement over the next largest model. Because the experiment is over determined, the *modeling set* is *itself* divided into a 4 repetition *model modeling set*, and 1 repetition *model training set*. A model of dimension $n + 1$ is regressed on the *model modeling set*, and its predictive performance is evaluated on the *model training set*. If its performance on the *model training set* is sufficiently better than the best n dimensional model, then the larger model is accepted. Failure to subdivide the *modeling set* results in a model of dimension equal to the size of the *modeling set* being chosen by the stepwise regression algorithm. This model has zero error on the *modeling set*, but is very unlikely to perform well on any other data, as it is likely to be extremely ill-conditioned and overly optimized for a single data set.

It is common to use an orthogonal projection, such as a singular value decomposition (SVD), to reduce the dimension of and over determined problem such as this one. The problem with such an approach is that if variance in the data is distributed non-uniformly, a standard projection will group 'good' points (points that are strongly correlated to the input factors and have low variability) and 'bad' points (points that are weekly correlated to the input factors and have high

variability) indiscriminately. In the broad band signal, for example, some frequency ranges show more variability than others. Clearly, the best model will rely on points from a low variability frequency range and ignore points from high variability regions. Such a choice is impossible if the points lie in the same basis and are thus projected onto the same coefficient of the new reduced dimensional space. The approach proposed here is to build a model using individual response points, where each point is selected based on best error reduction from the existing model. If the number of points used becomes larger than the test set, then an SVD is used to reduce the problem dimension.

At its most basic, the nearest neighbor algorithm is a way of determining which vector in the set $[z_{\mathrm{mdl}\,1} \ldots z_{\mathrm{mdl}\,n}]$ a vector z_{tst} most closely resembles. Especially if n is large, the technique may be very slow, and the question of which norm serves as most effective metric is a difficult one. Speed is not an issue at this point, and since our goal is to compare relative performance, the standard L_2 norm is sufficient. A transformation seeking to reduce the variation in the data is used prior to implementing the algorithm, as described as follows.

The problem of reducing the model order is addressed as follows. A simple regression is performed from factors to measurements, as shown:

$$\widehat{Y} = X \cdot B, \tag{4.1}$$

$$B = (X^{'} \cdot X)^{-1} \cdot X^{'} \cdot Y. \tag{4.2}$$

In equation 4.1, X is:

$$X = [1, x_1, \ldots, x_4, x_1 \cdot x_2, \ldots, x_3 \cdot x_4, \ldots, x_1 \cdot x_2 \cdot x_3 \cdot x_4]. \tag{4.3}$$

where x_i is the i^{th} factor and 1 is the mean term. Non linear interaction terms are included in the standard experimental design fashion [24] as polynomials in the factors. Y is a matrix composed of row vector, where each row vector, y_n is the total number of response points for the measurement being evaluated (1000 or 402) and n is the experimental run being considered. $\widehat{Y}$ is the least squares estimate of Y. We observe that $\widehat{Y}$ is not simply the best estimate of the measured response Y, but can also be considered a partition of the measurement as shown in Eq. 4.4.

$$Y_{\mathrm{mdl}} = \widehat{Y_{\mathrm{mdl}}} + (Y_{\mathrm{mdl}} - \widehat{Y_{\mathrm{mdl}}}), \tag{4.4}$$

$$\Rightarrow Y_{\mathrm{mdl}} = y_{\mathrm{mdl}}^{\mathrm{inp}} + Y_{\mathrm{mdl}}^{\mathrm{dist}}\,. \tag{4.5}$$

In Eq. 4.5, $Y_{\mathrm{mdl}}^{\mathrm{inp}}$ is the least squares estimate of that part of the measurement in the *modeling set* which is due to the input and $Y_{\mathrm{mdl}}^{\mathrm{dist}}$ is that part of the measurement due to disturbances. It is straight forward to obtain a linear (projection[2]) operator, L^{inp} to achieve this decomposition directly:

$$Y_{\mathrm{mdl}}^{\mathrm{inp}} = L^{\mathrm{inp}} \cdot Y_{\mathrm{mdl}}. \tag{4.6}$$

When we apply the projection operator to the measurement data in the *testing set*, we obtain an *estimate* of $Y_{\mathrm{tst}}^{\mathrm{inp}}$:

$$\widehat{Y_{\mathrm{tst}}^{\mathrm{inp}}} = L^{\mathrm{inp}} \cdot Y_{\mathrm{tst}}. \tag{4.7}$$

Qualitatively, the accuracy of $\widehat{y_{\mathrm{tst}}^{\mathrm{inp}}}$ is a function of the degree to which:

$$y_{\mathrm{tst}}^{\mathrm{inp}} \in \mathrm{span}(Y_{\mathrm{mdl}}^{\mathrm{inp}}). \tag{4.8}$$

Likewise, the degree to which errors in $\widehat{y_{\mathrm{tst}}^{\mathrm{inp}}}$ affect the accuracy of $\widehat{x}$, the estimate of the input levels, is a function of the conditioning of the relationship between x and y^{inp}. If a change in factor level results in a small change in response, then small errors in $\widehat{y_{\mathrm{tst}}^{\mathrm{inp}}}$ are very likely to result in large errors in the estimate of x.

V. EXPERIMENTAL RESULTS

Table II summarizes R^2 performance averaged over all six permutation of experiments where one repetition of the experiment is used as the *testing set* and the other five are used as the *modeling set*. Overall broad band performance is borderline. Two factor, power and pressure, are predicted accurately, while prediction of the chemical factors is less reliable. The difference in performance between broad band and narrow band performance, however, is very clear. Of the four factors,

[2] L is the orthogonal projection of span$\{\widehat{Y_{\mathrm{mdl}}}\}$ onto span$\{Y_{\mathrm{mdl}}\}$. L can be computed with a standard SVD.

only power is predicted with reasonable accuracy. The limitations of the power fit at first seem surprising, but serve to exemplify the limitations of the narrow band sensing. While the magnitude of the forward voltage signal is proportional to the generator power level, it is also affected by other factors. Thus the power prediction is unreliable with narrow band sensing. In the case of the broad band signal, the generator power results in a sufficiently unique response to accurately infer that factor.

The results in Table II summarize performance using a nearest neighbor algorithm. Differences here are even more pronounced. Of 96 total experimental runs, only one broad band run was was mis-identified. In the narrow band case, only power level was correctly identified in all cases. Pressure and Oxygen levels were correctly identified about 65% of the time, and Argon level was correctly identified *less* than 50% of the time.

The results of this experiment suggest that there is a fundamental difference in the way that standard plasma inputs affect the narrow band compared to the way they affect the broad band measurement. With the factors and levels used in this experiment and modeling methods used in the analysis, the broad band sensing approach appears to provide substantially more information about the plasma than the narrow band approach. It should be noted that the response of the broad band signal, although clear and distinct, is extremely complex. Achieving a more rigorous relationship between input factors and broad band response is a daunting task.

VI. FUTURE WORK

The promise shown by this experiment motivates several areas of future work. Although commonly used, oscilloscopes have severe resolution limitations. It is possible that the weak narrow band results presented here are due not to fundamental limitations of the approach but rather due to instrument resolution. Such limitations are not present in a spectrum analyzer. Accordingly, it would be of value to repeat the experiment with a spectrum analyzer as well as an oscilloscope.

Although results using an intrusive probe are promising, the potential of an intrusive diagnostic is clearly limited. In order to be practical, a non-intrusive application of broad band sensing is

needed, as proposed in Figure 3. In this approach, referred to as 'direct injection', a high power high pass filter is used upstream of the matching network to allow a high frequency signal to pass through to the plasma without allowing the high power low frequency signal to damage the network analyzer. An implementation of 'direct injection' has been implemented on our Lam TCP and its sensitivity to process relevant variables is currently being evaluated.

Additional theoretical work is warranted both in the design of non intrusive and minimally intrusive sensors. The design of an antenna to couple to a waveguide is a relatively standard problem in microwave engineering and a waveguide or resonant cavity may be a reasonably accurate model for a plasma chamber. Likewise, inferring the relative permittivity of the medium from the frequency response of the system may be feasible. Finally, a relationship between relative permittivity and relevant plasma parameters may be determined as well.

VII. CONCLUSION

Power, pressure, Ar and O_2 levels relevant to micro-electronics processing were varied using a full factorial experiment performed on a GEC reference cell. Standard RF sensing (referred to as 'narrow band sensing') was compared to a novel sensing technique based on resonance probes used in ionospheric research (referred to as 'broad band sensing'). Standard statistical techniques were used to regress a linear model against narrow band and broad band data respectively. A much better fit to the data was obtained using broad band sensing. Suggestions for further work include using a spectrum analyzer for the narrow band data and methods for designing a non intrusive broad band system.

ACKNOWLEDGMENTS

The authors sincerely thank Steven C. Shannon, and Professor Mary L. Brake for use of and assistance with the GEC cell. Sven G. Bilén constructed the resonance probe used in this research. Hyun-Mog Park designed and wrote the automated data collection algorithm. Dr. Helen Maynard and Dr. Oliver D. Patterson provided many useful discussions.

This work was supported in part by AFOSR/ARPA MURI Center under grant # F49620-95-1-0524 and The Semiconductor Research Company under contract #97-FC-085.

REFERENCES

[1] C. Garvin, D. S. Grimard, and J. W. Grizzle, in *RF Sensing and Calibration for Real Time Control of Plasma-Based Deposition and Etching* (AIP Press, New York, NY, 1998).

[2] H. L. Maynard, E. A. Reitman, J. T. C. Lee, and D. E. Ibbotson, J. Echem. Soc. **143**, 2029 (1996).

[3] A. J. Miranda and C. J. Spanos, J. Vac. Sci. Technol. A **13**, 1888 (1996).

[4] A. Ison, W. Li, and C. J. Spanos, in *1997 IEEE International Symposium on Semiconductor Manufacturing* (The Institute for Electrical and Electronics Engineers, San Francisco, California, 1997), Vol. 1997.

[5] A. J. van Roosmalen, W. G. M. van den Hoek, and H. Kalter, J. Appl. Phys. **58**, 653 (1985).

[6] N. Hershkowitz, J. Ding, and J. Jenk, in *Proceedings of the IEEE International Conference on Plasma Science* (The Institute for Electrical and Electronics Engineers, Santa Fe, New Mexico, 1994), Vol. 1994, p. 110.

[7] S. Bushman, T. F. Edgar, I. Trachtenberg, and N. Williams, in *Proceedings of SPIE* (International Society for Optical Engineering, Bellingham, Washington, 1994), Vol. 2336.

[8] M. A. Sobolewski, IEEE Transactions on Plasma Sciences **23**, 1006 (1995).

[9] M. Klick, J. Appl. Phys. **79**, 3445 (1996).

[10] M. Klick, W. Rehak, and M. Kammeyer, Jpn. J. Appl. Phys. **36**, 4625 (1997).

[11] P. A. Miller, L. A. Romero, and P. D. Pochan, Phys. Rev. Letts. **71**, 863 (1993).

[12] V. A. Godyak, R. B. Piejak, and B. M. Alexandrovich, J. Appl. Phys. **69**, 3455 (1991).

[13] N. Hershkowitz, IEEE Transactions on Plasma Sciences **22**, 11 (1994).

[14] M. A. Heald and C. B. Wharton, *Plasma diagnostics with microwaves* (John Wiley and Sons, New York, NY, 1965).

[15] C. B. Wharton, R. F. Post, and T. Prosser, Lawrence Radiation Lab Report **5**, 238 (1955).

[16] K. Takayama, H. Ikegami, and S. Miyasake, Phys. Rev. Letters **5**, 238 (1960).

[17] R. M. Moroney, A. J. Lichtenberg, and M. A. Lieberman, J. Appl. Phys. **66**, 1618 (1989).

[18] C. A. M. de Vries, A. J. van Roosmalen, and G. C. C. Puylaert, J. Appl. Phys. **57**, 4386 (1985).

[19] P. J. Hargis Jr. *et al.*, Rev. Sci. Inst. **65**, 140 (1994).

[20] J. Neter, W. Wasserman, and M. H. Kutner, *Applied Linear Statistical Models* (Richard D. Irwin, Inc, New York, New York, 1985).

[21] A. Sen and M. Srivastava, *Regression Analysis: Theory, Methods, and Applications* (Springer-Verlag, New York, New York, 1990).

[22] A. J. Miller, *Subset Selection in Regression* (Chapman and Hall, New York, New York, 1990).

[23] S. A. Nene and S. K. Nayar, IEEE Transactions on Pattern Analysis and Machine Intelligence **19**, 989 (1997).

[24] T. P. Barker, *Quality by Experimental Design* (Marcel Dekker, inc, New York, New York, 1985).

FIGURES

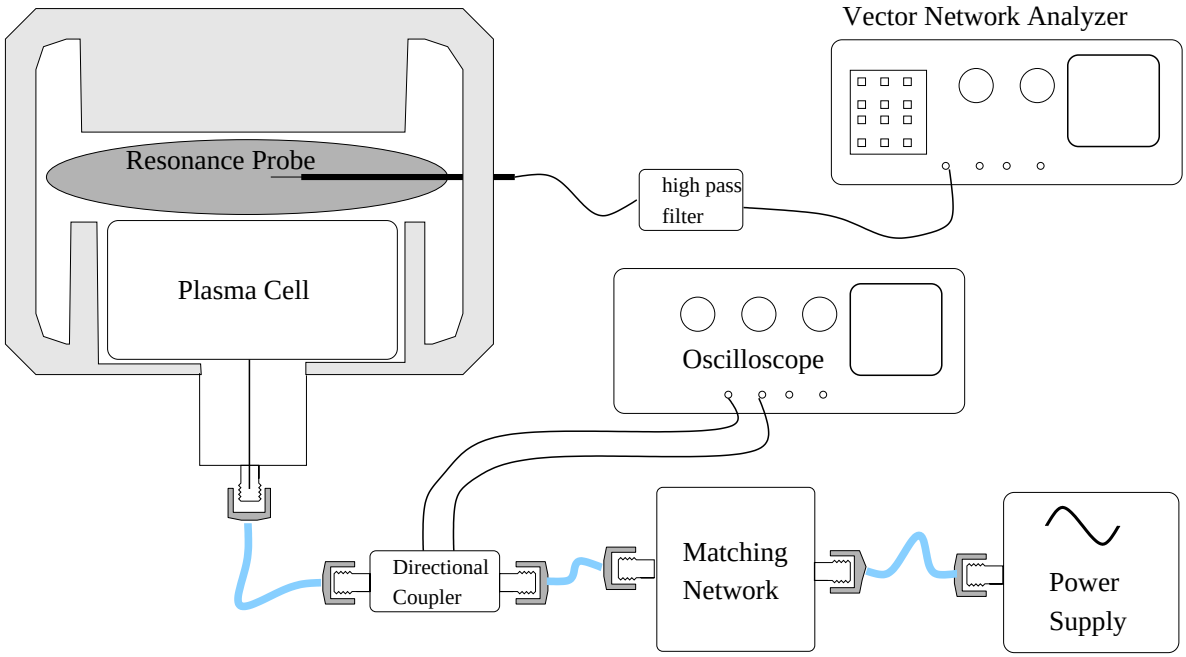

FIG. 1. Experimental Setup for Comparing Broad Band and Narrow Band Sensing

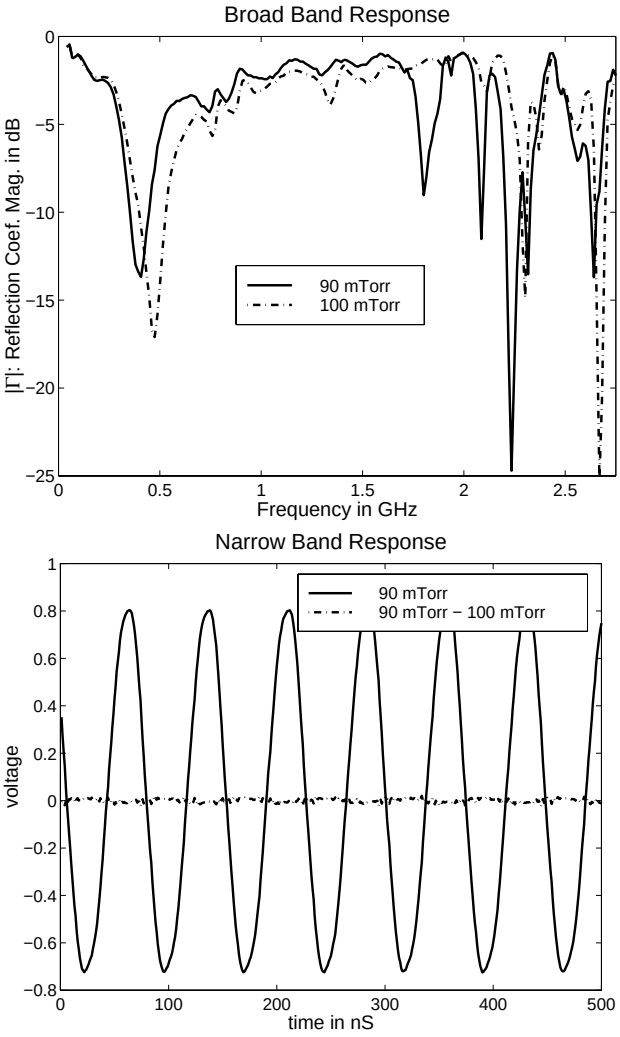

FIG. 2. Broad Band and Narrow Band Response for 90 and 100 mTorr Plasma Pressures

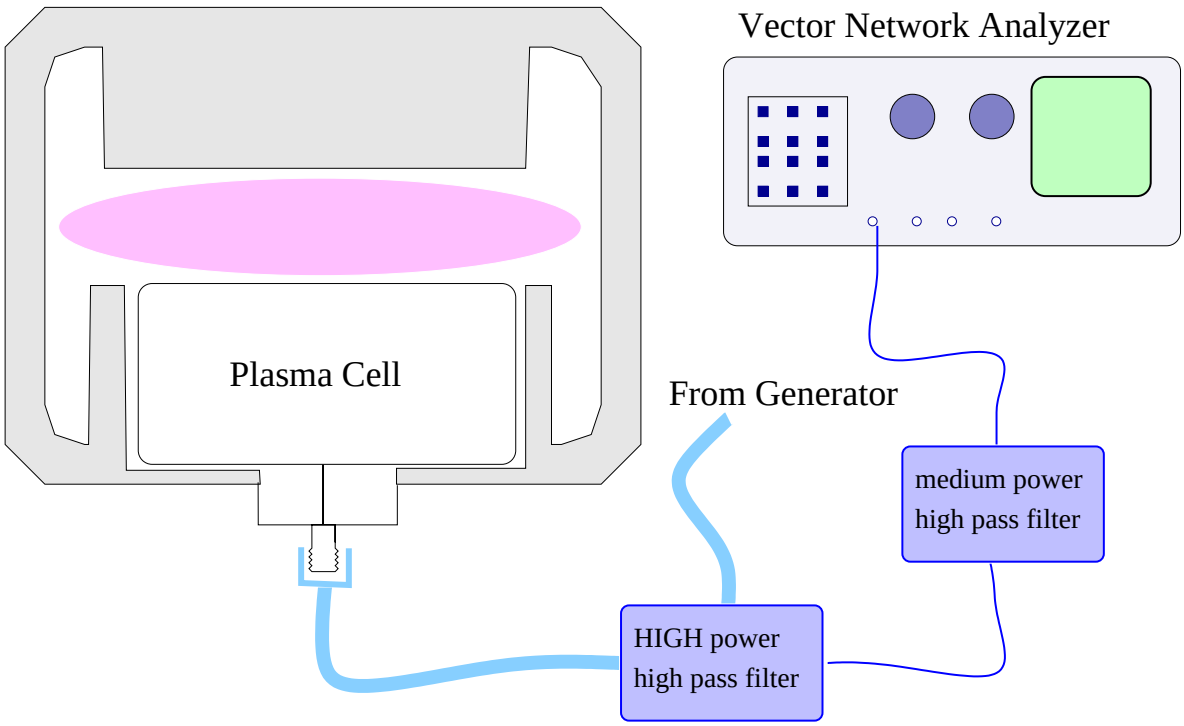

FIG. 3. Potential Non-intrusive Setup for Broad Band Sensing

TABLES

TABLE I. Experimental Variables and Levels

	Ar	O_2	Pressure	Power
Level 1	$10 sccm$	$5 sccm$	$90 mTorr$	$90 W$
Level 2	$20 sccm$	$15 sccm$	$100 mTorr$	$100 W$

TABLE II. R^2 of Broad and Narrow Band Sensor vs Factor Levels

method	Ar	O_2	Pressure	Power
Broad Band	0.8207	0.8240	0.9261	0.9330
Narrow Band	0.1809	0.2838	0.4323	0.8201

TABLE III. Nearest Neighbor Success of Broad and Narrow Band Sensor vs Factor Levels

method	Ar	O_2	Pressure	Power
Broad Band	96/96	96/96	95/96	95/96
Narrow Band	47/96	62/96	64/96	96/96