Dynamic Matching System for Radio-Frequency Plasma Generation

Anas Al Bastami, Alexander Jurkov, Parker Gould, Mitchell Hsing, Martin Schmidt, and David J. Perreault
Department of Electrical Engineering and Computer Science
Massachusetts Institute of Technology
Cambridge, Massachusetts
Email: abastami@mit.edu

Abstract—Plasma generation systems represent a particularly challenging load for radio-frequency power amplifiers owing to the combination of high operating frequency (e.g., 13.56 MHz) and highly variable load parameters. We introduce a dynamic matching system for Inductively Coupled Plasma (ICP) generation that losslessly maintains near-constant driving point impedance (minimal reflected power) across the entire plasma operating range. This new system utilizes a Resistance Compression Network (RCN), an impedance transformation stage, and a specially-configured set of plasma drive coils to achieve rapid adjustment to plasma load variations. As compared to conventional matching techniques for plasma systems, the proposed approach has the benefit of relatively low cost and fast response, and does not require any moving components. We describe suitable coil geometries for the proposed system, and treat the design of the RCN and matching stages, including design options and tradeoffs. A prototype system is implemented and its operation is demonstrated with low pressure ICP discharges with O\textsubscript{2}, C\textsubscript{4}F\textsubscript{8}, and SF\textsubscript{6} gases at 13.56 MHz and over the entire plasma operating range of up to 250 W.

I. INTRODUCTION

The use of plasma in the processing of materials has become prevalent in a wide range of industries [1]–[3]. A commonly used means of generating these plasmas is to inductively couple energy from a radio-frequency (RF) power amplifier into the chamber containing the gas to be ionized, (e.g., by driving RF current through a coil wound around the chamber). Key challenges in RF plasma generation include efficiently generating and controlling the RF power delivered into the plasma, while maintaining acceptable loading of the associated RF power amplifier under the highly-variable conditions in a plasma system. ICP loads represent a dynamically-variable load impedance that depends on gas type and pressure, operating mode, power level and other features. For example, in processes such as the Bosch process in reactive ion etching, there is a frequent change of the gas being ionized, which can contribute to the variation in the impedance presented by the coil [4], [5]. The effective load impedance can vary substantially in both its real and reactive components, making matching challenging.

Power amplifiers (including switched-mode amplifiers or inverters) used to generate the plasmas are typically designed to operate efficiently into a fixed load resistance (e.g., a 50\,\Omega termination). Moreover, the performance of many high-frequency (HF, 3-30 MHz) switched-mode power amplifiers / inverters degrades quickly with variations in load impedance, e.g., [6]–[8]. Because the effective loading provided by the plasma coil varies greatly across operating conditions, a tunable matching network (TMN) or “Antenna Tuning Unit” (ATU) is typically utilized between the power amplifier and the plasma coil. A TMN typically comprises a matching network having adjustable passive components that are dynamically tuned, such as by using servo motors to mechanically adjust a set of variable capacitors and/or inductors, by using switched capacitor and/or inductor banks, or by using high-power varactors. Such TMN units are very effective, and can present a constant (e.g., 50\,\Omega) input impedance across the entire range of the plasma operation [9]–[11], but tend to be costly, bulky, and exhibit a slow response to load changes. Inverters or power amplifiers that can directly drive the wide impedance range in ICP applications at HF frequencies may be possible, but are presently in only a very nascent stage of development [12].

This paper introduces a new approach to driving inductively-coupled plasmas that provides narrow-range loading of the power amplifier with lossless, high-bandwidth adjustment for plasma load variations. It also has relatively small size and cost, and doesn’t require moving or switchable components. The proposed approach comprises a set of specially-configured plasma drive coils along with an impedance matching system based on a resistance compression network [6], [13]–[15]. Section II of the paper describes the overall structure and operation of the proposed system, including the coil design and resistance compression network. Section III describes the design of a prototype system demonstrating the approach, and Section IV presents experimental results. Lastly, Section V concludes the paper.

II. SYSTEM ARCHITECTURE

Figure 1 illustrates the proposed plasma drive and matching system. The system comprises multiple plasma coils for driving the plasma which have nearly identical but dynamically-varying driving-point impedances, a tuning and impedance transformation stage which cancels the reactive impedance components of the matched plasma coils, and a resistance compression network that losslessly transforms the multiple matched but variable resistive loads into a single, nearly constant input impedance across the operating range. We describe each of these stages in turn.
A. Plasma Drive Coils

A key element of the proposed approach is two (or more) plasma drive coils that ideally act as independent and matched loads that vary in impedance together across the operating conditions (e.g., as the state of the plasma changes).

In a plasma system, the total impedance seen at the terminals of a drive coil is a function of the power absorbed by the plasma as well as plasma composition, pressure, temperature and other conditions [5]. The system seen looking into a plasma drive coil can be modeled as a transformer, with the drive coil in the primary side, and the plasma load in the secondary [2], as shown in Figure 2(a). The voltage-current relationships for this network can be written as follows:

\[
\begin{bmatrix}
\tilde{V}_{rf} \\
\tilde{V}_p 
\end{bmatrix} = j\omega \begin{bmatrix}
L_{11} & L_M \\
L_M & L_{22}
\end{bmatrix} \begin{bmatrix}
\tilde{I}_{rf} \\
\tilde{I}_p
\end{bmatrix},
\]

(1)

\[
\tilde{V}_p = -R_p \tilde{I}_p,
\]

(2)

where \(L_{11}\) and \(L_{22}\) denote the drive coil and the plasma load self-inductances, respectively, \(L_M\) denotes the mutual inductance between the drive coil and the plasma, and \(R_p\) denotes the plasma resistance. Equations (1) and (2) can be combined to obtain an expression for the effective input impedance \(Z_S\) as seen from the terminals of the plasma coil:

\[
Z_S = \frac{\tilde{V}_{rf}}{\tilde{I}_{rf}} = \frac{\omega^2 L_M^2 R_p}{R_p^2 + \omega^2 L_{22}^2} + j\omega \left( L_{11} - \frac{\omega^2 L_M^2 L_{22}}{R_p^2 + \omega^2 L_{22}^2} \right)
\]

(3)

Equation (3) can be further simplified by using the fact that in ICP discharge systems, the skin depth is typically much smaller than the radius of the plasma chamber, which results in \(\omega^2 L_{22}^2 \gg R_p^2\), yielding:

\[
Z_S \approx \left( \frac{L_M}{L_{22}} \right)^2 R_p + j\omega \left( L_{11} - \frac{L_M^2}{L_{22}} \right)
\]

(4)

Therefore, the effective impedance seen at the coil terminals can be represented as shown in Figure 2(b). The effective series resistance \(R_s\) depends on the plasma resistance \(R_p\) as well as the inductive coupling between the coil and the plasma. In practice, \(R_s\) also depends on the coil’s resistance. Similarly, the reactive component of the load impedance \(L_s\) depends on the coil’s self-inductance, mutual inductance, as well as the plasma inductance. The effective inductance \(L_s\) of each coil should be compensated so that the RCN stage is ideally resistively loaded.

To realize the multiple loads which vary in impedance together, we utilize multiple plasma drive coils that should ideally fulfill two conditions: (a) the coils should have similar self-inductances and provide low magnetic coupling to each of the other coils (ideally uncoupled); and (b) each coil should drive plasma having substantially the same conditions, such that the coils have similar driving-point impedances. This is
ideally provided by having the multiple coils drive the same physical region of plasma. Together, these conditions lead to a set of coils providing identically varying load impedances.

There are many coil geometries that can provide one or both of the aforementioned conditions. A preferred construction for the multiple coils is an implementation in which the coils are magnetically uncoupled (magnetically “orthogonal”) but spatially close (such that they drive similar or identical portions of the plasma). This can be accomplished with coils that physically overlap the same space and drive similar (or the same) plasma volume, but are wound such that net flux from each coil does not link the other coil(s). One way to implement this is to make the windings overlapped but geometrically orthogonal, as illustrated in Figure 3. (The desired requirements can be achieved, for example, by placing the coils with the same center but rotated by 90 degrees).

B. Tuning and Impedance Transformation

The next stage of the system in Figure 1 provides tuning and impedance transformation. In order for a resistance compression network to provide the most effective transformation and compression of the matched (but variable) load impedances presented by the coils into a narrow-range resistive input impedance, the networks loading the RCN should appear resistive in nature.

A first function of the intermediate stage is to tune the coil impedances to appear resistive at a desired operating frequency. This is done with a tuning network for each coil by, for example, using a trimmer capacitor in series with each coil to null the reactive component of impedance. In such an implementation, the trimmer capacitor ought to be physically adjusted only once for a particular set of operating conditions, unlike conventional TMNs where adjustment of passive elements is performed continuously by a feedback loop. A more sophisticated network such as a series tank could also be used to enable dynamic reactance cancellation adjustment through frequency control, as described below. The reactive impedance can vary somewhat across operating conditions (e.g., owing to the screening out of the RF fields from the interior of the plasma chamber as the plasma density increases [5]). While this can be neglected, one can optionally address this by utilizing small variations in frequency to adaptively tune for matched resistive loading by the coils. (The coil tuning can be made far more frequency selective than other elements of the system, so that the small frequency adjustments for nulling coil reactance do not substantially disturb other system aspects).

A second function of this stage is to transform the reactance-compensated plasma load impedance to a desired range over which the RCN is designed to operate (e.g., utilizing transformers, transmission-line transformers [16], immittance converters [17] or other transformation networks that function well with variations in loading impedance). In addition to this, it is possible to add another tuning element or network after the transformation if the reactive component of the transformed-impedance is non-zero. The result of this is a matched pair of resistive load impedances that can be compressed into a narrow range by the RCN.

C. Resistance Compression Network

The final component of the plasma matching system is the resistance compression network (RCN). As described in [6], and shown in Figure 4, an RCN is a special single-input multi-output matching network. It utilizes resonant impedance transformation to losslessly transfer energy from a single input port to a pair of output ports loaded with identical resistances (which we refer to as a matched pair of loads), such that the resistance looking into the input port varies much less than the matched load resistances [6], [13]. Thus, the variation of the input resistance appears “compressed” as compared to the resistance variation of the matched loads. The matched loads ideally see equal portions of the input power, and the relative phases of the load voltages vary as the matched load impedances vary. Together, these elements enable energy transfer into the plasma while presenting a narrow-range input impedance for the RF power amplifier.

Note that it is possible to switch the order of the RCN and the impedance transformation stages in Figure 1 to achieve the desired match. Moreover, it is also possible to add another impedance transformation stage to step-up or step-down the RCN-compressed resistive input impedance as required. This function may also be incorporated into the resistance compression network if desired [6]. Likewise, as will be illustrated in the next section, the functional elements of the different blocks
Fig. 4. Structure of two basic RCN topologies. The input impedance $Z_{in}$ is resistive and is given by: (a) $Z_{in} = \frac{X_2^2}{2\pi R} \left[ 1 + \left( \frac{R}{X_2} \right)^2 \right]$, and (b) $Z_{in} = \frac{2R}{1+(R/X)^2}$.

in Figure 1 can in some cases be combined together, providing reduced component count and increased power efficiency.

It is important to note that the loads seen by each branch of the RCN should (ideally) be matched and uncoupled. If the coils are coupled, then the circuit may no longer compress the plasma resistance in the expected manner. Likewise, not only is mutual inductance relevant, but mutual resistance among coupled coils would be of concern [18].

III. Matching Network Design and Implementation

In order to validate the proposed approach, a prototype system was implemented and tested in a low pressure ICP system. The design target is to deliver up to 250 W at a nominal frequency of 13.56 MHz into a 3.25 inch-diameter plasma chamber. Figure 5 shows the circuit schematic of the matching network.

The plasma drive coil subsystem comprises two coils wound on a fixture using the orthogonal structure illustrated in Figure 3. The fixture is shown in Figure 6, and is designed to mount around the cylinder of the plasma chamber. The Delrin base of the fixture has a 3.25 inch hole that allows the fixture to slide onto the plasma chamber, and a set of grooves machined to provide space for the pair of orthogonal two-turn windings. Two Teflon strips threaded with 100 mil wide and 50 mil deep T-slot grooves (which were made with a micro-slitting saw with a cutting diameter of 93 mil and 30 mil saw thickness) are fixed on the Delrin base; these allow the two-turn coils to be wound tightly and consistently around the chamber, and provide thermal insulation between the wires and the Delrin fixture as well as electrical insulation between the turns. The two two-turn coils are each wound with about 17 inches of 92 mil × 25 mil rectangular cross-section wire; a layer of Kapton tape and a turn-to-turn spacing of 100 mil are used to prevent any arcing between the turns. Under this configuration, the self-inductance of each coil was around 1.2 µH, and the mutual inductance was negligible.

The network in Figure 5(a) is a conceptual illustration showing the different circuit blocks in the high-level structure in Figure 1. $C_3$ and $C_5$ are used to tune out the coils reactances.

Fig. 5. Design of the matching network. (a): A conceptual illustration showing the different circuit blocks of the structure in Figure 1, (b): The actual implementation of the matching system in which a number of the conceptual elements of (a) are combined together.

Fig. 6. Fixture used to realize the orthogonal coil structure. This makes the wires easily accessible and holds them in place. The Teflon strips (white) help thermally insulate the wires from the Delrin fixture (black).
and the impedance transformation is implemented using a T-network immittance converter. This network is designed to transform the range of plasma resistance into a range suitable for use by the RCN. Inductors \(L_{s}/L_{i}\) and the trimmer capacitors \(C_{T}\) form a pair of high-Q series-resonant tanks in series with the plasma coils. Each tank is resonant at a frequency close to the nominal RF operating frequency of the system thus allowing one to control its effective reactance (capacitively or inductively) either by narrow-band frequency modulation or by direct adjustment of the trimmer capacitors. This feature of the design allows for cancellation of any reactive mismatch between both plasma loads due to network implementation non-idealities and small differences in the coil implementation and geometry. The RCN stage is designed to compress the resistance variation about a nominal value of 50 \(\Omega\). In this system, the quality factor of the output resonant tank (e.g., \(C_{3}, L_{s}, R_{s}\) in Figure 5(a)) is much larger than that of the RCN tank. This makes this narrowband output network highly sensitive to any mismatch in output capacitance (and to any parasitic capacitance at the output nodes) as well as to any variation in the operating frequency. Since the coil reactance slightly varies across the operating range due to the plasma conductivity, the high sensitivity of the output narrowband tanks to the operating frequency (compared to the RCN and transformation tanks) can be exploited to adjust the tuning of the coils across the plasma regions of operation by making small variations to the operating frequency of the power amplifier. This can be implemented in a closed-loop control strategy that dynamically selects the optimal power amplifier operating frequency for which the reflection is minimal.

In the actual implementation, the conceptual elements of Figure 5(a) are combined together, resulting in the circuit of Figure 5(b). A photograph of the implemented circuit is shown in Figure 7.

IV. EXPERIMENTAL RESULTS

An ENI 3100LA RF power amplifier was used to drive the ICP system at a nominal frequency of 13.56 MHz (with frequency variations over a range of 13.56-14.04 MHz to allow dynamic tuning cancellation of the coil reactances). The etching system used for these experiments was a custom-built, ICP-Reactive Ion Etching (RIE) reactor designed for etching small substrates (1-2 inch diameter). The reactor was built as part of an effort to produce high performance yet ultra-low capital cost semiconductor equipment for low throughput applications found often in academia, early-stage businesses, and corporate R&D. The reactor chamber consists of a 3 inch ID alumina \((\text{Al}_2\text{O}_3)\) cylinder sealed on top and bottom to standard Conflat-style vacuum flanges. The top flange includes a gas showerhead for providing a spatially uniform distribution of the desired process gas to the chamber. This showerhead is fed from an external gas block assembly containing a bank of mass flow controllers and an electronically-controllable valve manifold. The bottom flange mates to a lower chamber assembly that contains a port for the vacuum foreline and a differentially-pumped vacuum feedthrough for the substrate chuck assembly (which serves as the RIE electrode). The system is pumped down to a base pressure of ~5 mTorr using a 2-stage rotary vane pump. Due to the small overall size of the reactor (<2 L) and the small gas flows used (<10 standard cubic centimeters per minute), this rotary vane pump can easily maintain the 20-50 mTorr operating pressures needed for the dry etching processes performed. The substrate chuck assembly contains connections for the RIE bias power (0.5 W at 13.56 MHz), substrate cooling, temperature measurement, and a clamping system to register and confine the wafer being etched. The entire chuck assembly can be translated up and down within the reactor chamber via the differentially pumped feedthrough to achieve the desired spacing between the plasma generation region (the region spanned by the orthogonal coils, see Figure 3) and the substrate. The etching results shown below used a spacing of 7.5 inches between the center of the coils and the substrate.

The results shown here are for generating plasmas using \(\text{O}_2, \text{C}_4\text{F}_8, \text{and SF}_6\) gases, which are commonly used in plasma ashing, passivation, etching, among other processes, at a low pressure of about 20 mTorr. Power was gradually increased from zero to about 230 W (280 W for \(\text{SF}_6\)), and a Bird 4421 RF power meter was used to record the forward and reflected powers at the output of the RF power amplifier, in order to gauge the degree to which desired performance is achieved. Our metric in this experiment is the ratio of power reflected from the load to forward RF power. This power reflectance would ideally be zero. When the input impedance of the dynamic impedance matching system is 50 \(\Omega\), it is matched to the power amplifier’s output impedance, and thus ideally no power should be reflected, and the power capability of the power amplifier is used effectively. In this experiment, the aforementioned frequency tuning strategy was
utilized to adjust the match to the plasma. At each power level, the operating frequency was manually adjusted to obtain the lowest measured reflected power. For each gas, the ratio of the reflected power to the forward power was recorded, as shown in Figure 8. It can be seen that less than 7% of the power was reflected back in all three gases, indicating an acceptable matching behavior across the entire plasma range of operation, which included both E-mode and H-mode drive of the plasmas. Figure 9 shows the plasma system being powered by the orthogonal coils.

The next experiment is used to demonstrate that the plasmas generated using the prototype matching system at these power levels have enough energy and are dense enough to perform useful processes on a silicon wafer. Figure 10(a) shows a 650-675 µm thick silicon wafer obtained for this experiment. A 10 µm thick photoresist layer was patterned on top of it using photolithography techniques, with a horizontal gap width of 100 µm, shown in Figure 10(b). The experiment starts by flowing 5 standard cubic centimeters per minute (sccm) of O₂ at 200 W for about 10 minutes to clean the chamber from any impurities that might affect the wafer. Next, the wafer in Figure 10(b) is placed and 6 sccm of SF₆, the primary gas used for etching, is allowed to flow into the chamber for 10 minutes at 250 W. The result of this step is the isotropically-etched wafer shown in Figure 10(c). The depth of the etch after the 10 minutes was measured to be 37 µm, giving a time-averaged etch rate of 3.7 µm/min. This etch rate is comparable to results obtained in the same reactor [4] using traditional plasma generation and impedance matching systems, and is more than sufficient for many applications.

The etching step was followed by three rounds of 15-minute runs of O₂ ashing, using 5 sccm of O₂ at 200 W of power. Figures 10(d)-(f) show the result after each ashing run. It can be seen that the photoresist gradually gets thinner, until no noticeable traces of it are observed. It can be concluded that the proposed matched coil system can be effective in systems designed for semiconductor processing with plasmas.

V. Conclusion

A dynamic matching system for radio-frequency inductively-coupled plasma generation has been introduced. This system can provide efficient power delivery into plasma loads with good matching to the power amplifier, while achieving small size and low cost, and eliminating mechanical or switched components. The proposed approach utilizes a special set of matched but orthogonal drive coils along with a resistance compression network to compress the variable plasma impedance to an almost fixed value that results in minimal reflected power to the power amplifier. By arranging the plasma drive coils in special geometries such that they are largely magnetically uncoupled, the compression networks can be used to provide passive, high-bandwidth matching with no mechanically-controlled or moving parts. This can permit the design of efficient, compact, fast-response, and cost-effective plasma generation systems. An ICP discharge system utilizing this technique has been constructed and the results show good matching across the entire range of plasma operation for three gases commonly used in various semiconductor manufacturing processes: O₂, C₄F₈, and SF₆. The results also demonstrate the ability of the plasma generated using the proposed matching system to perform etching and ashing on a silicon wafer.

ACKNOWLEDGMENT

The authors would like to thank the Qatar Foundation Research Division for their financial support of author Anas Al Bastami.

REFERENCES

Fig. 9. The proposed plasma drive system with O$_2$ being ionized during a plasma ashing process.


Fig. 10. Photograph of the wafer cross section (a) before processing it; (b) after adding a patterned photoresist layer; (c) after etching with SF$_6$ for 10 minutes; (d) after ashing for 15 minutes; (e) after ashing for 30 minutes; and (f) after ashing for 45 minutes.