Neural Network Tuning of Radio Frequency Plasma Sources

by

Robert W. Jones

Abstract

A neural network has been trained to perform the difficult task of optimal tuning of a radio frequency plasma source.

Introduction

Radio frequency plasma generators are one of the most common types of plasma source. In RF plasma sources, gas is ionized when electrons accelerated in the RF electric fields collide with neutral gas atoms. In our present device a 10 turn coil wound around the plasma serves as antenna when coupled to an RF oscillator tuned to a few megahertz frequency, figure 1.

The antenna coil L forms a part of a resonant circuit which is driven by the crystal controlled RF oscillator, figure 2. The matching circuit is adjusted by tuning 3 variable capacitors C_1, C_2, and C_3. Unfortunately the plasma antenna loading is poorly understood and neither analytical nor computational models are sufficiently well developed to accurately describe and optimize the real resonant circuit. (Petty and Smith, 1986.) We turned instead to artificial neural networks which are capable of modeling complex input-output relationships learned on the basis of empirical examples alone.

Theory/ Methods

The neural network consists of 3 input units [1], 1 for neutral gas pressure, 1 for RF power, and 1 for RF

frequency. These are fully connected to 10 hidden units. The hidden units are, in turn, fully connected to 3 output units, one for each tuning capacitor, C_1, C_2, and C_3. All variables are normalized so as to fall within the unit interval.

For each vector I the hidden unit j receives an input:

$$h_j = \Sigma\, w_{jk} I_k + w_o \qquad (1)$$

and produces an output:

$$v_j = \frac{1}{1+\exp(-h_j)} \qquad (2)$$

w_o is an adjustable bias.

Output unit i then receives:

$$h_i = \Sigma w_{ij} v_j + w_o \qquad (3)$$

and computes the output vector:

$$C_i = \frac{1}{1+\exp(-h_i)} \qquad (4)$$

The activation function of equation 4 is known theoretically to make the network a universal approximator. The initial values of the weights w_{jk} and w_{ij} are random numbers in the range between 0 and 1. The feed forward network is shown schematically in figure 3 and is trained by the back propagation algorithm, the most common sort of supervised network training. 7 hidden units are shown in figure 3. Various numbers have been tried in practice.

The computed theoretical output vector (C_i, C_2, C_3) is compared with the actual measured vector C_i and the error C_p-C_i is back propagated in order to correct the weights w_{ij} and w_{jk} (Bishop, 1994).

The weights wij are corrected by amounts:

$$\Delta w_{ij} = L(C_p - C_i)\left(\frac{1}{1+\exp(-h_i)}\right)\left(1 - \frac{1}{1+\exp(-h_i)}\right) V_j \qquad (5)$$

and the weights wjk are corrected by amounts:

$$\Delta w_{jk} = I\, L(C_p - C_i)\left(\frac{1}{1+\exp(-h_i)}\right)\left(1 - \frac{1}{1+\exp(-h_i)}\right) \qquad (6)$$

$$w_{ij}\left(\frac{1}{1+\exp(-h_e)}\right)\left(1 - \frac{1}{1+\exp(-h_e)}\right) I_k$$

where L is a constant, typically 0.2.

The weights are adjusted (corrected) at each calculation step by:

$$\text{new } w_{ij} = \text{old } w_{ij} + \Delta w_{ij} \qquad (7)$$

$$\text{new } w_{jk} = \text{old } w_{jk} + \Delta w_{jk}$$

and the process is repeated, using the <u>same single</u> I, C_p vector pair, until C_p-C_i is smaller than some desired error. This whole process is then repeated with <u>each additional</u> available training pair (I, C_p).

Training on subsequent I, C_p data sets reduces the network's ability to reproduce previous I, C_p pairs. For this reason, repeated training "passes" are made throughout the complete training set until C_p-C_i is acceptably small for <u>all</u> data sets employed. (The acceptable error C_p-C_i is chosen based on a knowledge of the typical <u>experimental</u> accuracy obtainable for C_p). A separate network is trained for different fill gases.

Results/Discussion

Once trained, the neural network model was then tested by inputting a number of data sets which had not been employed during the training process. The required number of hidden units depends upon the complexity of the function to be learned. If too few hidden units are employed a highly non-linear function cannot be learned. If too many hiddens are employed the network will not generalize well. No theory currently exists with which to determine how many hiddens to use. We tried various numbers and chose the best network empirically.

A trained neural network can then be used to

predict the tuning conditions for optimal source operation. Optimal operation is determined by maximization of the ion saturation current drawn by a Langmuir probe inserted into the plasma. In table 1 we compare the tuning capacitor values C_1, C_2, and C_3 suggested by the neural network compared with those obtained by careful manual adjustment (requiring 5-15 minutes in the laboratory). The neural network substantially speeds up the tuning process every time the source conditions are changed.

Literature Cited

Bishop, C.M. 1994. Neural networks and their application. *Rev. Sci. Instrum.* 65(6): 1803-1832.

Petty, C.C., and D.K. Smith. 1986. High-power radio-frequency plasma source. *Rev. Sci. Instrum.* 57(10): 2409-2414.

Robert W. Jones is a full professor in the Departments of Physical Sciences at Emporia State University.

Figure 1: RF plasma device.

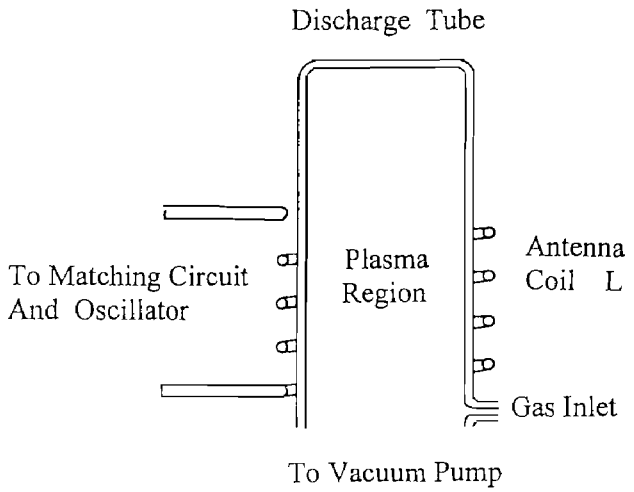

Figure 2: Antenna coupling circuit with tuning capacitors C_1, C_2, and C_3. C is typically 150 pF.

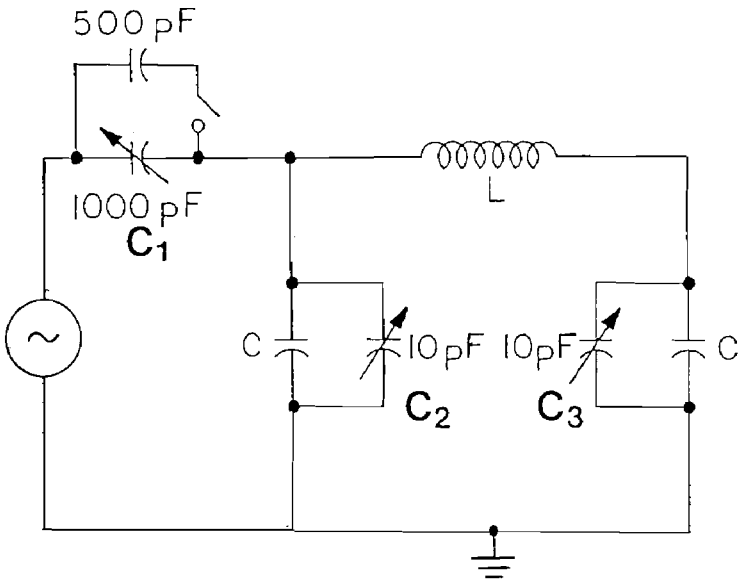

Figure 3: Neural network composed of inputs I, hidden units h, and outputs, c.

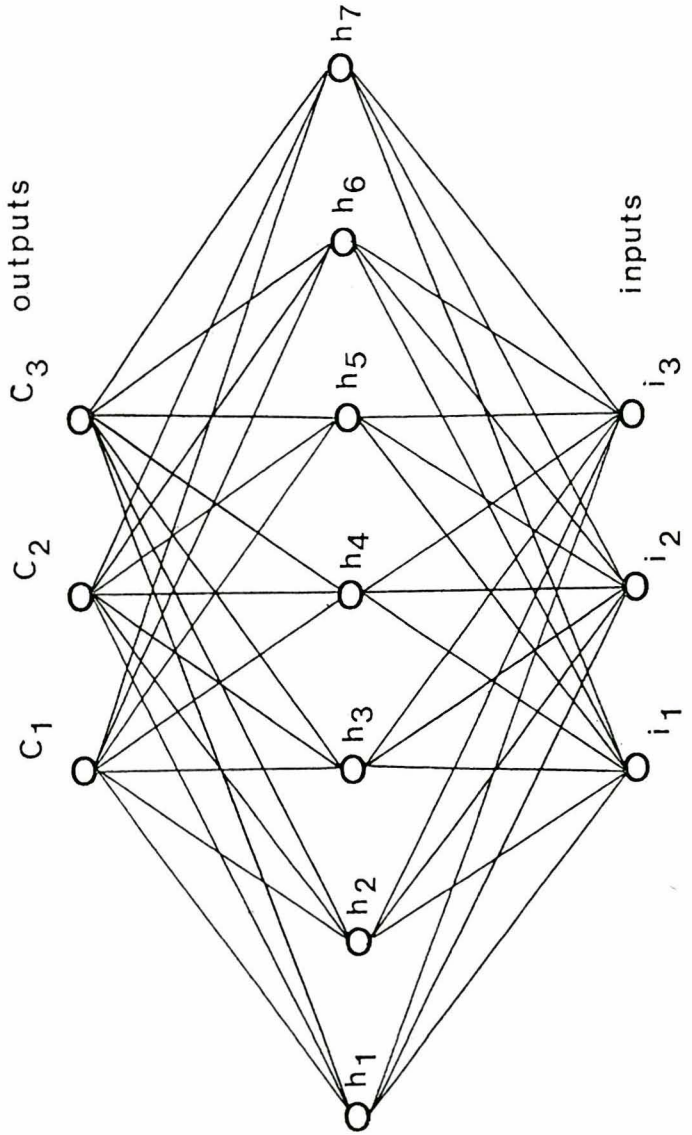

Table 1: Tuning conditions found by a human operator compared with those predicted by the neural network.

Gas Pressure	RF Frequency	RF Power	Neural Network Prediction	Hand Tuning
N_n 10^{-3}	f (MHz)	P (Watts)	C_1 C_2 C_3 (pF) (pF) (pF)	C_1 C_2 C_3 (pF) (pF) (pF)
10^{-3}	9	5	8.5 5.5 2.5	8.3 5.5 2.0
10^{-3}	9	50	2.5 3.5 2.0	2.2 3.4 2.0
10^{-3}	8	50	9.0 10 9.5	9.4 10 9.0