Application of Neural Fuzzy Network to Pyrometer Correction and Temperature Control in Rapid Thermal Processing

Jiun-Hong Lai and Chin-Teng Lin, *Member, IEEE*

***Abstract*— Temperature measurement and control are two difficult problems in the rapid thermal processing (RTP) system. For many applications such as rapid thermal processing chemical vapor deposition (RTCVD) [1] and rapid thermal oxidation (RTO) [2], large changes in wafer emissivity can occur during film growing, leading to erroneous temperature measurements with a single wavelength pyrometer. The error in the inferred temperature will affect the temperature control of the RTP system. In order to correct the temperature reading of the pyrometer, a neural fuzzy network is used to predict the emissivity changes for the compensation of measured temperature. As for the temperature control, to overcome ill performance of the temperature tracking system due to the inaccuracy of the identified model, another neural fuzzy network is used in the RTP system for learning inverse control simultaneously. The key advantage of neural fuzzy approach over traditional ones lies on that the approach does not require a mathematical description of the system while performing pyrometer correction and temperature control. Simulation results show that the adopted neural fuzzy networks can not only correct the pyrometer reading accurately, but also be able to track a temperature trajectory very well.**

***Index Terms*— Emissivity, feedforward learning, inverse control, temperature measurement, TSK fuzzy rules, wafer.**

I. Introduction

In recent years, semiconductor manufacturing has been moving toward single-wafer processing in individual or cluster tools. The manufacture of high-density integrated circuits on wafer with increasingly larger diameter and submicron design rules requires a minimization of the thermal budget. Rapid thermal processing (RTP) technology inherently possesses these features and provides several advantages over traditional batch furnaces. One advantage of RTP is that it eliminates the long ramp-up and ramp-down time associated with furnaces, enabling a significant reduction in the thermal budget. Another advantage of RTP is that it allows better control over the processing environment (e.g., the amount of oxygen present), which is becoming critical in some applications. Today, RTP is in production used for annealing (RTA) [4], cleaning (RTC) [5], oxidation (RTO) [6], chemical vapor deposition (RTCVD) [7], and nitridation (RTN) [5]. Wafer temperature measurement and control are two critical issues for RTP. Many recent papers have been done to identify and compensate for the sources of the two problems [8]–[10].

At present, the temperature sensor used for RTP is typically a single-wavelength pyrometer, which is a noncontact technique. The temperature is determined from the emissivity of the wafer at the specified wavelength of the pyrometer. The optical pyrometer has been successfully used for annealing, cleaning, and nitridation where the emissivity of the wafer surface does not change during processing. However, for applications where the characteristics of the surface are being changed (such as rapid thermal oxidation or rapid thermal chemical vapor deposition), the emissivity is varying as a function of the type of film structure, the thickness of the film, and the composition of the surface it is being deposited on. This change can lead to temperature errors in excess of 50°C for the cases described above. Some papers deal with this problem and various techniques for improving temperature measure are proposed, e.g., dual-wavelength pyrometer [11] and models of the optical properties to predict the emissivity change [14] that are based on measuring emitted radiation. The dual-wavelength technique needs two different instruments to measure emitted radiation directly and emissivity indirectly. These testing results show difficulties with both implementations of the technique. The approach using the model-based emissivity correction is only suitable for thin films and short-time processing. Reference [14] showed this was due to differences between expected and actual deposition rates. An alternative approach has been the use of the temperature dependence of acoustic wave speed [12]. There are also a variety of approaches based on measuring the thermal expansion of wafer [13]. A survey of a broad range of RTP temperature sensing techniques was given in [32].

In this paper, the simulation technique is based on the RTP global model which includes the temperature sensor, control loop, and lamp system. The process simulated is the rapid thermal oxidation (RTO). Emissivity changes during oxidation are calculated according to reflection and refraction within thin dielectric films on a silicon substrate. The oxide thickness as a function of oxidation time at various temperatures is simulated by a linear parabolic model. Based on the basic heat transfer law, we shall derive a pyrometer model to simulate the temperature sensor in the RTP system. Then we use a

Manuscript received December 29, 1997; revised October 15, 1998. This work was supported by the National Science Council, R.O.C., under Grant NSC 87-2213-E-009-146.

The authors are with the Department of Electrical and Control Engineering, National Chiao-Tung University, Hsinchu, Taiwan, R.O.C.

Publisher Item Identifier S 1063-6706(99)02797-6.

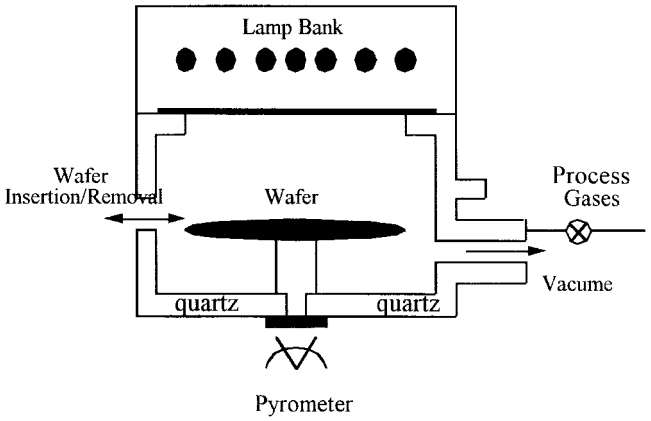


Fig. 1. Schematic of the RTP system.

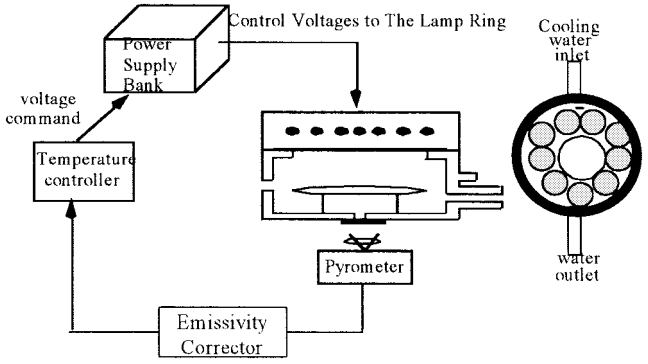


Fig. 2. Schematic of the closed-loop rapid thermal processing system.

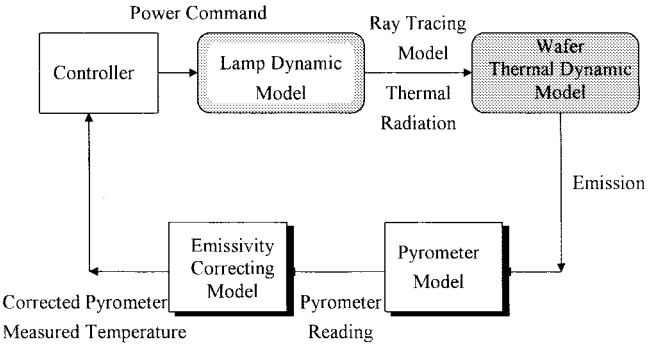


Fig. 3. Global model components of the closed-loop RTP system.

neural fuzzy network to learn and predict the variations of film growing as a function of oxidation time under different process temperatures, and use the optical model to calculate the emissivity of the wafer. The good prediction capability of the neural fuzzy network can predict the emissivity changes on the wafer surface and convert the pyrometer reading to a correct number. Another neural fuzzy network is used to control the temperature of a RTP system to achieve two control objectives: temperature trajectory following and temperature uniformity on the wafer.

This paper is organized as follows. In Section II, physical modeling of the RTP system is performed. In Section III, the optical pyrometer model is discussed. In Section IV, the configuration of the adopted neural fuzzy network, training process, and pyrometer measure correcting method are introduced. In Section V, simulation studies on temperature control combined temperature measurement of the RTP system using the neural fuzzy networks are presented. Conclusions are made in Section VI.

II. Modeling of the RTP System

There have been a number of papers [15]–[17], [33] concerning the analysis or modeling of the wafer temperature distribution during RTP. However, these papers always neglect some heat transform on the wafer or the temperature sensor model. The importance of the interface (lamp dynamic, sampling, analog-to-digital, and digital-to-analog conversions) between controlling computer and RTP processor when implementing the software on the actual equipment is also ignored. The lamps transfer function that we propose will take this into account and a global modeling of the RTP system will be used for off-line simulation.

The RTP system considered in this paper is shown in Fig. 1. In Fig. 1, a bank of tungsten-halogen lamps mounted below a diffusely reflection ceiling constitutes the heat source. Cooling air is forced over the lamps to prevent the quartz sheaths from overheating. Two quartz plates separate the lamps from the lower half of the oven. The wafer rests on three quartz pins above the black water cooled oven floor. The side walls of the bottom half of the oven are partially reflective and are at an angle to the vertical. A pyrometer views the bottom surface of the wafer through a central hole in the floor.

Mathematical model of the closed-loop RTP system is described here. The model is called a global model because it simulates all the components in the RTP system and can thus be used to investigate the interplay of the wafer itself. A simplified schematic of the closed-loop system is shown in Fig. 2. The system uses one bank of lamps which is arranged in orthogonal directions. The lamps are placed outside the reaction chamber's quartz windows. A flat reflector is located behind the bank of lamps. The system is controlled by a feedback control loop that utilizes the difference between the converted temperature T_c and the set temperature T_s to control the lamp power. The constituent components of the global model are shown in Fig. 3. The components include a wafer thermal-dynamic model (in particular, the heat transfer to and from the wafer) and a lamp dynamic and ray tracing model for the dynamics of lamp power to the wafer. A power supplier used to provide the power to lamps and an emissivity corrector for correcting the pyrometer reading are also included. In the following subsections, the mathematics model used for each component is described separately. These models are then integrated into a global model.

The present application is to the growth of silicon oxide for temperature above 600 °C and, hence, the wafer is opaque to lamp radiation [18]. When compared with conventional furnace oxidation, RTO has the advantages of single-wafer processing that customizes the process for each wafer and

reduces the risk of scrapping a large number of wafers of low-thermal-budget processing to form shallow junction, which RTO achieves with short oxidation time and ease of integration into cluster tools for automation—*in situ* fabrication of integrated circuits.

A. Wafer Thermal Dynamic and Lamp Dynamic Models

The approach employed to modeling is analytical/numerical in that the heat transfer to, from, and within the wafer is calculated. Included in the calculation is the radiation heat transfer to the wafer, the heat conduction within the wafer, and the heat convection and the heat loss emitted from the wafer surfaces. For the radiation, the heat from the lamps is absorbed at the wafer surface and the radiation heat loss occurs at the surface. As shown in Fig. 2, the controller sends a voltage command to the power amplifier after receiving measured temperature signal from the pyrometer. The power reaches steady voltage level after receiving control voltage command in the ideal case. But in the actual situation, it is ramp up/down to reach steady-state level. The lamp dynamics describes the power from the lamps after receiving power supply voltage. In most published papers, the dynamics of the lamp power intensity to the control voltage command was neglected and the power from the lamps was assumed to be directly proportional to the power supply voltage. For this cause, we present a simple dynamic model between the command voltage and lamp power. The presented lamp dynamics has the following form:

$$\begin{aligned} V(t) &= V(t_{k-1}) + (\mathrm{com}_k - V(t_{k-1})) \\ &\quad \times \left(1 - \exp\left(-\frac{t - t_{k-1}}{\tau}\right)\right) \\ P_{\mathrm{lamp}} &= f(V(t), T(x,t)) \cdot V(t)^2 \end{aligned} \tag{1}$$

where $V(t)$ is the power supply temporary voltage, com_k is the present command sent by the controller, t_{k-1} the last time step, t the time constant, P_{lamp} is the lamp power, and the function $f(V, T)$ is varied by $V(t)$ and temperature $T(x, t)$ at position x. The present lamp dynamics will raise the complexity of the RTP simulation, and match up the overall RTP system to actual RTP dynamics.

In our RTP system, electrical energy is supplied to a "ring" cylindrical arrangement of tungsten-halogen bulbs of which more will be mentioned later. Energy is radiated through a quartz window onto a thin semiconductor wafer. A model of the heat transfer for such a system is developed in cylindrical coordinates, where the origin of the coordinate system is the center of the wafer bottom surface, and the z axis of the coordinate system coincides with the central axis of the wafer. The model is based on the assumption that the temperature distribution is axisymmetric and that the wafer is thin enough such that axial (z axis) thermal gradients can be neglected. Furthermore, the wafer is discretized into annular zones in each of which the temperature is assume to be uniform. Such an approach is often used in radiative heat transfer applications and has been used for RTP systems and for furnaces in [9].

The heat transfer model of the wafer takes into account convective, conduction, and radiative energy transport mechanisms. The model is written as

$$\dot{T} = -A_{\mathrm{rad}}T^4 - A_{\mathrm{conv}}(T - T_a) - A_{\mathrm{cond}}T + BP \tag{2}$$

where T_a is the ambient temperature expressed as an $N \times 1$ vector (the ambient temperature is assumed to be constant in the chamber), T is the $N \times 1$ temperature vector of the wafer elements, B is the $N \times M$ radiation-energy matrix from the lamps to the wafer, and P is the $M \times 1$ lamp-power vector, where N is the number of the wafer segments and M is the number of lamps [15]. The matrices A_{rad}, A_{conv}, and A_{cond} represent the radiative, convective, and conduction heat transfer, respectively. A complete description of these matrices can found in [15]. The capacitive effects of the thick windows are neglected here since the associated time constant is two-order magnitude larger than that of the wafer. Instead, the windows heating model is considered as a slowly varying disturbance for the purpose of system identification and controller design. Physical parameters used in the RTP model are the same as those used in [8].

As discussed above, the wafer temperature is measured by a pyrometer and thus requires accurate knowledge of the emissivity of the region of the wafer viewed by the pyrometer. The sensitivity of this emissivity with respect to layer structure and to pyrometer wavelength can be modeled. This gives a guide to correctly choose the emissivity value, evaluate the wafer temperature response, and thus predict overall furnace operation. This approach will be illustrated in Section II-B.

B. Wafer Emissivity Model

Thermal radiation is a part of electromagnetic radiation. Assume that a radiant heat flux Φ (a heat quantity in a unit time) is incident on the surface of a wafer. Of this heat flux, the portion Φ_α is absorbed, Φ_ρ is reflected, and Φ_τ is transmitted. The following definitions are introduced:

$$\begin{aligned} &\text{absorptivity} && \alpha = \Phi_\alpha/\Phi \\ &\text{reflectivity} && \rho = \Phi_\rho/\Phi \\ &\text{transmissivity} && \tau = \Phi_\tau/\Phi. \end{aligned} \tag{3}$$

For the wafer, applying the principle of energy conservation gives

$$\alpha + \rho + \tau = 1. \tag{4}$$

This relation is also called Kirchhoff's law. Similar to (4), which is valid for the total radiation, the spectral components of α, ρ, τ at wavelength λ, denoted by α_λ, ρ_λ, and τ_λ, respectively, also meet the following relation:

$$\alpha_\lambda + \rho_\lambda + \tau_\lambda = 1. \tag{5}$$

The values of α, ρ, and τ depend upon the material, its surface state and temperature, while α_λ, ρ_λ, and τ_λ additionally depend upon the wavelength λ.

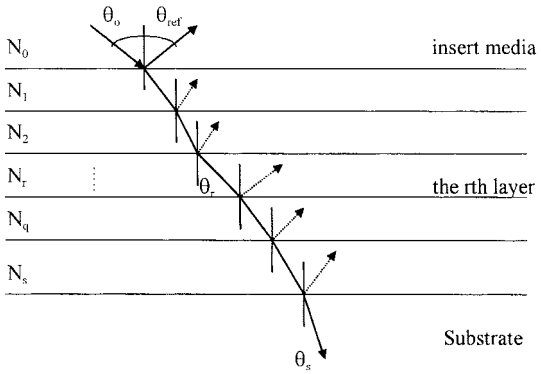


Fig. 4. Propagation of a radiation wave through multilayer films.

By the Kirchhoff's law as demonstrated above, the emissivity of the wafer surface can be determined immediately from the reflectivity for the case of an opaque wafer (transmissivity, τ, equals to zero). With the reflectivity known, the emissivity is equated to (1 − reflectivity). To determine the reflectivity of multiple dielectric layers for a radiation wave at a given wavelength and inserting angle, we need to consider the multireflection phenomena on the adjacent surfaces in multiple thin layers. According to the basic theory for the optics in [19]–[22], each layer's optical property can be described by a characteristic matrix that can be used to describe the optics of radiation through a planar N-layer structure. Fig. 4 shows the propagation of a radiation wave through a multilayer film. For a given layer structure on top of a single crystal silicon substrate, the characteristic matrix is

$$\begin{bmatrix} \cos\delta_r & \dfrac{i}{\eta_r}\sin\delta_r \\ i\eta_r \sin\delta_r & \cos\delta_r \end{bmatrix}. \tag{6}$$

For N dielectric layers, the characteristic matrix is a product sequence as follows:

$$\prod_{r=1}^{N} \begin{bmatrix} \cos\delta_r & \dfrac{i}{\eta_r}\sin\delta_r \\ i\eta_r \sin\delta_r & \cos\delta_r \end{bmatrix}. \tag{7}$$

With the characteristic matrix available, the reflectivity can be determined [21] by

$$R = \left(\frac{\eta_o B - C}{\eta_o B + C}\right)^2 \tag{8}$$

where

$$\begin{bmatrix} B \\ C \end{bmatrix} = \prod_{r=1}^{N} \begin{bmatrix} \cos\delta_r & \dfrac{i}{\eta_r}\sin\delta_r \\ i\eta_r \sin\delta_r & \cos\delta_r \end{bmatrix} \begin{bmatrix} 1 \\ \eta_s \end{bmatrix} \tag{9}$$

where

$$\delta_r = \frac{2\pi}{\lambda} N_r d_r \cos\theta_r: \quad \text{phase thickness}$$

$$\eta_o = \begin{cases} N_o \cos\theta_o, & S \text{ polarization} \\ N_o/\cos\theta_o, & P \text{ polarization} \end{cases}$$

$$\eta_r = \begin{cases} N_r \cos\theta_r, & S \text{ polarization} \\ N_r/\cos\theta_r, & P \text{ polarization} \end{cases}$$

$$\eta_s = \begin{cases} N_s \cos\theta_s, & S \text{ polarization} \\ N_s/\cos\theta_s, & P \text{ polarization} \end{cases}$$

TABLE I
COMPARISONS BETWEEN EXPERIMENTAL MEASUREMENTS OF EMISSIVITIES AND THE DESIRED THEORETIC RESULTS FOR DIFFERENT OXIDE THICKNESSES

Oxide thickness(Å)	Emissivity (Ref. [30])	Emissivity (Eq. (8)) (N_i=1.43, N_s=3.43+i0.1)	Error(%)
0	0.71	0.6988	1.6027
500	0.72	0.7023	2.5202
1000	0.73	0.7117	2.5713
2000	0.76	0.7467	1.7811
4000	0.87	0.8613	1.0101
80000	0.87	0.8686	0.1611

where

λ	wavelength;
$N_r d_r \cos\theta_r$	optical thickness (at a specific wavelength λ_0);
$N_r = n_r - iK_r$	complex refractive index;
n_r	real part of refractive index;
K_r	image part of refractive index, $K_r = 0$ for nonabsorption material;
d_r	layer thickness;
θ_r	the rth layer's insert angle determined by Snell's law;
$N_o \sin\theta_o = N_r \sin$	$\theta_r = N_s \sin\theta_s$;
N_o	insert material's complex refractive index;
θ_o	insert angle;
N_s	complex refractive index of the substrate;
θ_s	refractive angle of the substract.

Table I lists the comparisons between the actual experimental measurements of emissivities and the above theoretic results for different oxide thicknesses. It is obvious that the theoretic results and actual measurements are very close. In other words, if the oxide thickness is known, the emissivity of the wafer can be inferred precisely by this optical model. The experimental measurements in the second column of the table were done in [30].

C. Silicon Oxidation Growing Model

In this paper, the application of the RTP system is on thermal oxidation on a silicon wafer. The temperature of this formation process is exceeding 600 °C, where the wafer is opaque to the wavelength of concern here. The thickness of the SiO_2 on the top of Si x_0 can be obtained from the equation derived by Deal and Grove [23] in their linear-parabolic model of oxidation

$$x_o^2 + Ax_o = B(t+\tau) \tag{10}$$

where A and B are rate constants, and τ is the time displacement needed to account for the initial oxide layer x_i at $t = 0$

$$\tau = (x_i^2 + Ax_i)/B. \tag{11}$$

Both B and (B/A) may be well represented at the singly activated processes [24]

$$B = C_1 \exp(-E_1/kT) \tag{12}$$

and

$$(B/A) = C_2 \exp(-E_2/kT) \tag{13}$$

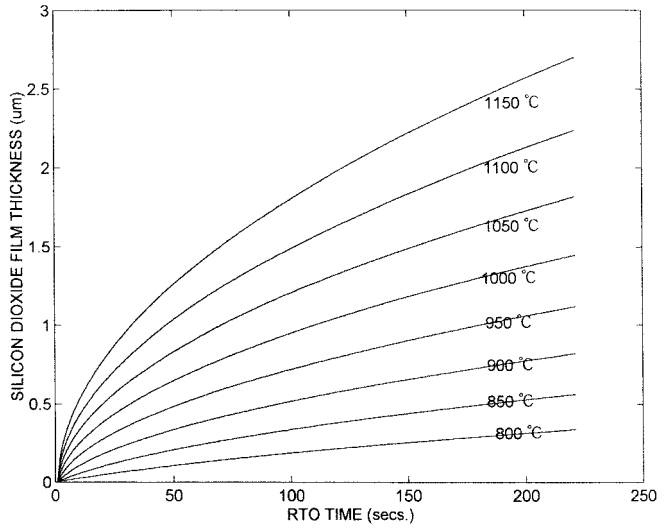


Fig. 5. Silicon oxide thickness as a function of RTO time under different temperatures (simulation results).

where $C_1 = 7.72 \times 10^2$ μm^2/h, $E_1 = 1.23$ eV, $C_2 = 6.23 \times 10^6$ μm/h, and $E_2 = 2.0$ eV. Details of the thermal oxidation process can be found in [23].

Solving the quadratic equation in (10) gives the oxide thickness as a function of time

$$x_o = \frac{A}{2}\left\{\left[1 + \frac{(t+\tau)}{A^2/4B}\right]^{1/2} - 1\right\}. \tag{14}$$

Fig. 5 is a plot of oxide thickness as a function of oxidation time at various temperatures. The curves in this figure are quite close to those obtained by practical experiments obtained in [23] showing that the model derived in the above is very accurate.

III. Optical-Pyrometer Temperature Measurement

A single wavelength pyrometer is widely used to infer wafer temperature in today's RTP systems. Pyrometers can measure the temperature of an object without physically touching the object. The temperature is determined from the radiated intensity of the wafer at the specified wavelength of the pyrometer. In Section III-A, we shall introduce the basic physics of energy transfer in thermally radiating bodies and then develop the pyrometer model according to this basic heat transfer law in Section III-B.

A. Emission and Absorption of Radiation

The radiant intensity W or the radiation exitance is the heat flux per unit area expressed as the ratio of the heat flux $d\Phi$ emitted from the infinitesimal element of the surface dA to the surface area dA itself

$$W = \frac{d\Phi}{dA} \text{ w/m}^2. \tag{15}$$

In the same units as the radiant intensity, the heat flux density q of the incident radiation is given by

$$q = \frac{d\Phi}{dA} \text{ w/m}^2. \tag{16}$$

This also takes account of the conduction and convection heat flux in addition to the radiation heat flux.

The spectral radiation intensity W_λ is defined as

$$W_\lambda = \frac{dW}{d\lambda} \text{ w/m}^2\ \mu\text{m}. \tag{17}$$

Planck's law gives the radiant flux distribution of a black body as a function of the wavelength and of the body's temperature by the relation

$$W_{o\lambda} = \frac{c_1 \lambda^{-5}}{e^{c_2/\lambda T} - 1} \tag{18}$$

where $W_{o\lambda}$ is the spectral radiant intensity of a black body, w/m^2 μm (the suffix "0" will be used in the future to indicate a black body), λ is the wavelength, μm, T is the absolute temperature of the thermal radiator, K, c_1 is the first radiation constant $c_1 = 3.7415 \times 10^8$ w $\cdot$ μm^4/m^2, and c_2 is the second radiation constant $c_2 = 1.4388 \times 10^4$ μm $\cdot K$.

For a given wavelength range from λ_1 to λ_2, (18) can be evaluated as

$$W_{o,\lambda_1-\lambda_2} = \int_{\lambda_1}^{\lambda_2} \frac{c_1 \lambda^{-5}}{e^{c_2/\lambda T} - 1}\, d\lambda \tag{19}$$

where $W_{o,\lambda_1-\lambda_2}$ is the band radiant intensity of a black body.

The ratio of the spectral radiant intensity W_λ at the wavelength λ of a nonblack body to the spectral radiant intensity of a black body $W_{o\lambda}$ at the same temperature is called the spectral emissivity ε_λ, where

$$\varepsilon_\lambda = \frac{W_\lambda}{W_{o\lambda}}. \tag{20}$$

If the spectral emissivity ε_λ of a given body is constant for each wavelength (i.e., ε_λ = constant), such a body is called a gray body. Similar to (20), if all the wavelengths from zero to ∞, are taken into consideration, the term "total emissivity" ε is used

$$\varepsilon = \frac{W}{W_o} \tag{21}$$

where W is the radiant intensity of any given body and W_o is the radiant intensity of a black body.

Following Kirchhoff's law, the spectral absorptivity α_λ of all opaque bodies equals their emissivity ε_λ so that we have

$$\alpha_\lambda = \varepsilon_\lambda. \tag{22}$$

For a given wavelength band from λ_1 to λ_2, Kirchhoff's law is expressed by

$$\alpha_{\lambda_1-\lambda_2} = \varepsilon_{\lambda_1-\lambda_2} \tag{23}$$

where $\alpha_{\lambda_1-\lambda_2}$ is the band absorptivity and $\varepsilon_{\lambda_1-\lambda_2}$ is the band emissivity.

When all the wavelengths from $\lambda_1 \to 0$ to $\lambda_2 \to \infty$ are taken into consideration, the corresponding form for (22), which is also valid, then becomes

$$\alpha = \varepsilon \tag{24}$$

where α is the total absorptivity and ε is the total emissivity.

B. Pyrometer Model

A narrow-band pyrometer senses the radiation intensity emitted by the center of the wafer at a specified wavelength according to the heat balance principle, i.e.,

$$W_w(\lambda, T_w, \varepsilon_w) = W_p(\lambda, T_p, \varepsilon_p) \tag{25}$$

where the subscript w denotes the wafer and p the pyrometer. Then, by the Planck's law described in (18), we have

$$\frac{\varepsilon_w C_1 \lambda^{-5}}{e^{(C_2/\lambda T_w)} - 1} = \frac{\varepsilon_p C_1 \lambda^{-5}}{e^{(C_2/\lambda T_p)} - 1}. \tag{26}$$

The actual temperature T_w can be obtained by solving (26)

$$T_w = \frac{1}{\dfrac{1}{T_p} - \dfrac{\lambda}{C_2} \ln \dfrac{\varepsilon_p}{\varepsilon_w}} \tag{27}$$

where ε_p is the emissivity of the pyrometer, ε_w is the wafer surface emissivity at the wavelength λ, and T_p is the pyrometer reading. We also can obtain T_p by solving (26)

$$T_p = \frac{1}{\dfrac{1}{T_w} + \dfrac{\lambda}{C_2} \ln \dfrac{\varepsilon_p}{\varepsilon_w}}. \tag{28}$$

In (27), we can see that if the emissivity of the wafer and that of the pyrometer are the same, the actual temperature and inferred temperature by the pyrometer are identical. It is evident from (28) that the temperature inferred from the pyrometer T_p is only equal to the actual wafer temperature when the assumed emissivity is equal to the actual wafer emissivity. Hence, the variations in wafer emissivity, ε_w, can lead to errors in temperature measurement. We shall propose a method to predict the different emissivity between wafer and pyrometer and compensate the measurement errors of the pyrometer in the next section.

IV. Pyrometer Measurement-Correcting Method

A. Methodology

In this section, we propose a method of using a neural fuzzy network to correct the pyrometer reading. Fig. 6 shows the block diagram of this method. In Fig. 6, the previous corrected temperature value $T_c(k)$ and the current processing time k are used as the inputs of the neural fuzzy network. The network will then predict the current film thickness, which is used to compute the emissivity of the wafer $\varepsilon'_w(k+1)$ according to the wafer optical model described in Section II-B. The converter in Fig. 6 is used to correct the pyrometer reading value $T_p(k+1)$ to $T_c(k+1)$, which is the predicted current wafer temperature. In Fig. 6, we use Z^{-1} block to represent the operator of unit-time delay. We expect that the corrected temperature value $T_c(k+1)$ is equal to the actual wafer temperature $T_w(k+1)$. We shall describe the details of the block diagram in the following subsections.

In [26], Sorrel *et al.* numerically integrated the Arrhenius growth-rate equation over time to linearly approximate the actual film thickness and then calculated the wafer emissivity by the approximated film thickness. The major source of errors

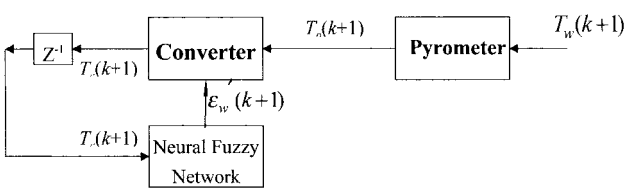


Fig. 6. Block diagram of the pyrometer measurement-correcting method.

of this method is due to the assumption of constant growth rate of the film thickness as a function of temperature. Once the wafer temperature cannot be kept at the setting value exactly, the growth rate will vary from time to time. If the estimated growth rate changes ±10% versus the actual growth rate, a large estimated temperature error will be observed at later time [26]. This method only can be applied to short-time and thin-films processes, because the actual growth rate is always not well known. In this paper, we use a separate neural fuzzy network to off-line learn the actual samples of the oxide thickness growth as a function of oxidation time at several different process temperatures. The neural fuzzy network estimates the growth behavior as associative memories without requiring a mathematical description of how the output functionally depends on the input, i.e., it learns from numerical samples. This method is more precise in estimating the actual oxide thickness than the mathematical function approximate approach. With the film thickness being predicted precisely by the neural fuzzy network, the pyrometer reading can be corrected accurately to the actual wafer temperature.

B. Self-Constucting Neural Fuzzy Inference Network

In this subsection, we shall introduce the neural fuzzy network used in our approach called self-constucting neural fuzzy inference network (SONFIN), modified from the one we proposed previously in [3]. The SONFIN is a fuzzy rule-based network possessing neural learning ability. As compared to other existing neural fuzzy networks [31], a major characteristic of the network is that no preassignment and design of fuzzy rules are required. The rules are constructed automatically during the training process. The SONFIN can overcome both the difficulty of finding a number of proper rules for the fuzzy logic controllers (FLC) and the over-tuned and slow-convergence phenomena of the backpropagation neural networks [25]. Another feature of the SONFIN is that it can optimally determine the consequent of fuzzy IF–THEN rules during the structure learning phase. We have shown that the SONFIN outperforms the pure neural networks greatly both in learning speed and accuracy [3].

A. Structure of the SONFIN

The structure of the SONFIN is shown in Fig. 7. Let $u^{(k)}$ and $a^{(k)}$ denote the input and output of a node in layer k, respectively. The functions of the nodes in each of the five layers of the SONFIN are described as follows.

Layer 1: No computation is done in this layer. Each node in this layer, which corresponds to one input variable, only

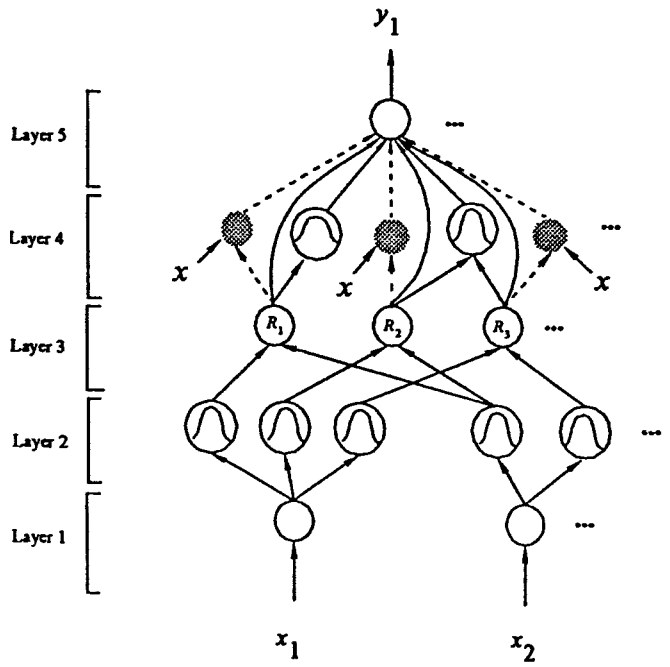


Fig. 7. Structure of the proposed self-constructing neural fuzzy inference network (SONFIN).

transmits input values to the next layer directly, i.e.,

$$a^{(1)} = u_i^{(1)} = x_i. \tag{29}$$

Layer 2: Each node in this layer corresponds to one linguistic label (small, large, etc.) of one of the input variables in layer 1. In other words, the membership value which specifies the degree to which an input value belongs a fuzzy set is calculated in layer 2. With the use of Gaussian membership function, the operations performed in this layer is

$$a^{(2)} = e^{-(u_{ij}^{(2)} - m_{ij})^2/\sigma_{ij}^2} \tag{30}$$

where m_{ij} and σ_{ij} are, respectively, the center (or mean) and the width (or variance) of the Gaussian membership function of the jth term of the ith input variable x_i. Unlike other clustering-based partitioning methods, where each input variable has the same number of fuzzy sets, the number of fuzzy sets of each input variable is not necessarily identical in the SONFIN.

Layer 3: A node in this layer represents one fuzzy logic rule and performs precondition matching of a rule. Here, we use the following AND operation for each layer 3 node

$$a^{(3)} = \prod_i u_i^{(3)} \tag{31}$$

where the summation is over the layer 2 nodes participating in the IF part of the rule.

Layer 4: This layer is called the consequent layer. Two types of nodes are used in this layer and they are denoted as blank and shaded circles in Fig. 7, respectively. The node denoted by a blank circle (blank node) is the essential node representing a fuzzy set (described by a Gaussian membership function) of the output variable. Only the center of each Gaussian membership function is delivered to the next layer for the LMOM (local mean of maximum) defuzzification operation [27] and the width is used for output clustering only. Different nodes in layer 3 may be connected to a same blank node in layer 4, meaning that the same consequent fuzzy set is specified for different rules. The function of the blank node is

$$a^{(4)} = \sum_j u_j^{(4)} \cdot a_{0i} \tag{32}$$

where $a_{0i} = m_{0i}$—the center of a Gaussian membership function. As to the shaded node, it is generated only when necessary. Each node in layer 3 has its own corresponding shaded node in layer 4. One of the inputs to a shaded node is the output delivered from layer 3 and the other possible inputs (terms) are the input variables from layer 1. The shaded node function is

$$a^{(4)} = \sum_j a_{ji} x_j \cdot u_i^{(4)} \tag{33}$$

where the summation is over all the inputs and a_{ji} is the corresponding parameter. Combining these two types of nodes in layer 5, we obtain the whole function performed by this layer for each rule as

$$a^{(4)} = \left(\sum_j a_{ji} x_j + a_{oi} \right) u_i^{(4)}. \tag{34}$$

Layer 5: Each node in this layer corresponds to one output variable. The node integrates all the actions recommended by Layers 3 and 4 and acts as a defuzzifier with

$$a^{(5)} = \sum_i a_i^{(4)} \Big/ \sum_i a_i^{(3)}. \tag{35}$$

B. Learning Algorithms for the SONFIN

Two types of learning—structure and parameter learning—are used concurrently for constructing the SONFIN. The structure learning includes both the precondition and consequent structure identification of a fuzzy IF–THEN rule. There are no rules (i.e., no nodes in the network except the input/output nodes) in the SONFIN initially. They are created dynamically as learning proceeds upon receiving on-line incoming training data by performing the following learning processes simultaneously: 1) input/output space partitioning; 2) construction of fuzzy rules; 3) consequent structure identification; and 4) parameter identification. In the above, processes 1, 2, and 3 belong to the structure learning phase and process 4 belongs to the parameter learning phase. The details of these learning processes are described in the rest of this subsection.

1) Input/Output Space Partitioning: The way the input space is partitioned determines the number of rules extracted from training data as well as the number of fuzzy sets on the universal of discourse of each input variable. For each incoming pattern $\boldsymbol{x}$, the strength a rule is fired can be interpreted as the degree the incoming pattern belongs to the corresponding cluster. For computational efficiency, we can

use the firing strength given in (31) directly as this degree measure

$$F^j(\boldsymbol{x}) = \prod_i u_i^{(3)} = e^{-[D(\boldsymbol{x}-\boldsymbol{m})]^T[D(\boldsymbol{x}-\boldsymbol{m})]} \tag{36}$$

where $F^j \in [0, 1]$, $D = \text{diag}(1/\sigma_{i1}, 1/\sigma_{i2}, \cdots, 1/\sigma_{\text{in}})$, and $\boldsymbol{m} = (m_{i1}, m_{i2}, \cdots, m_{\text{in}})^T$.

Using this measure, we can obtain the following criterion for the generation of a new fuzzy rule. Let $\boldsymbol{x}(t)$ be the newly incoming pattern. Find

$$J = \arg \max_{1 \le j \le c(t)} F^j(\boldsymbol{x}) \tag{37}$$

where $c(t)$ is the number of existing rules at time t. If $F^J \le \overline{F}(t)$, then a new rule is generated where $\overline{F}(t) \in (0, 1)$ is a prespecified threshold that decays during the learning process. Once a new rule is generated, the initial centers and widths are set as

$$\boldsymbol{m}_{(c(t)+1)} = \boldsymbol{x} \tag{38}$$

$$D_{(c(t)+1)} = -\frac{1}{\beta} \cdot \text{diag}(1/\ln(F^J), \cdots, 1/\ln(F^J)) \tag{39}$$

according to the first-nearest-neighbor heuristic [27], where $\beta \ge 0$ decides the overlap degree between two clusters.

After a rule is generated, the next step is to decompose the multidimensional membership function formed in (38) and (39) to the corresponding one-dimensional membership function for each input variable. For the Gaussian membership function used in the SONFIN, the task can be easily done as

$$e^{-[D_i(\boldsymbol{x}-\boldsymbol{m}_i)]^T[D_i(\boldsymbol{x}-\boldsymbol{m}_i)]} = \prod_j e^{-((x_j-m_{ij})^2/\sigma_{ij}^2)} \tag{40}$$

where m_{ij} and σ_{ij} are, respectively, the projected center and width of the membership function in each input dimension. To reduce the number of fuzzy sets of each input variable and to avoid the existence of redundant ones, we should check the similarities between the newly projected membership function and the existing ones in each input dimension. Since the Gaussian membership functions are used in the SONFIN, we use the formula of the similarity measure $E(A, B)$ of two fuzzy sets A and B derived previously (see [28] for details), where $0 \le E(A, B) \le 1$ and the larger $E(A, B)$ is, the more fuzzy set A is similar to B. Let $\mu(m_i, \sigma_i)$ represent the Gaussian membership function with center m_i and width σ_i. The whole algorithm for the generation of new fuzzy rules as well as fuzzy sets in each input dimension is as follows. Suppose no rules are existent initially. IF $\boldsymbol{x}$ is the first incoming pattern THEN do

PART 1. { Generate a new rule
with center $\boldsymbol{m_1} = \boldsymbol{x}$, width
$D_1 = \text{diag}(1/\sigma_{init}, \cdots, 1/\sigma_{init})$
where σ_{init} is a prespecified constant.
After decomposition, we have n
one-dimensional membership functions
with $m_{1i} = x_i$ and $\sigma_{1i} = \sigma_{init}$
$i = 1 \cdots n$.
}
ELSE for each newly incoming $\boldsymbol{x}$, do

PART 2. { find $J = \arg \max_{1 \le j \le c(t)} F^j(\boldsymbol{x})$
IF $F^J \ge \overline{F}_{in}(t)$
do nothing
ELSE
$\{c(t+1) = c(t) + 1$
generate a new fuzzy rule, with
$\boldsymbol{m}_{c(t+1)} = \boldsymbol{x}$, $D_{c(t+1)} = -(1/\beta)$
$\cdot\text{diag}(1/\ln(F^J), \cdots, 1/\ln(F^J))$.
After decomposition, we have
$m_{new-i} = x_i$, $\sigma_{new-i} = -\beta \cdot \ln(F^J)$
$i = 1 \cdots n$.
Do the following fuzzy measure for
each input variable i:
{ $degree(i, t) \equiv \max_{1 \le j \le k_i}$
$E[\mu(m_{new-i}, \sigma_{new-i}), \mu(m_{ji}, \sigma_{ji})]$
where k_i is the number of partitions of
the ith input variable.
IF $degree(i, t) \le \alpha(t)$
THEN adopt this new membership
function, and set $k_i = k_i + 1$
ELSE set the projected membership
function as the closest one. }
}

In the above algorithm, $\alpha(t)$ is a scalar similarity criterion that is monotonically decreasing such that higher similarity between two fuzzy sets is allowed in the initial stage of learning. For the output space partitioning, the same measure in (37) is used. Since the criterion for the generation of a new output cluster is related to the construction of a rule, we shall describe it together with the rule construction process in Process 2 below.

2) Construction of Fuzzy Rules: As mentioned in learning process 1, the generation of a new input cluster corresponds to the generation of a new fuzzy rule, with its precondition part constructed by the learning algorithm in Process 1. At the same time, we have to decide the consequent part of the generated rule. Suppose a new input cluster is formed after the presentation of the current input/output training pair ($\boldsymbol{x}$, $\boldsymbol{d}$), then the consequent part is constructed by the following algorithm:

IF there are no output clusters
do { *PART 1* in Process 1, with $\boldsymbol{x}$ replaced by $\boldsymbol{d}$ }
ELSE
do {
find $J = \arg \max_j F^j(\boldsymbol{d})$.
IF $F^J \ge \overline{F}_{\text{out}}(t)$
connect input cluster $c(t+1)$ to the existing
output cluster J
ELSE
generate a new output cluster
do the decomposition process in PART 2 of
Process 1
connect input cluster $c(t+1)$ to the newly
generated output
cluster.
}.

The algorithm is based on the fact that the preconditions of different rules may be mapped to the same consequent fuzzy set. Compared to the general fuzzy rule-based models with singleton output where each rule has its own individual singleton value [9], [11], fewer parameters are needed in the consequent part of the SONFIN, especially for the case with a large number of rules.

3) Consequent Structure Identification: Up to now, the SONFIN contains fuzzy rules in the form as

$$\begin{aligned}&\text{Rule } j\text{: IF } x_1 \text{ is } A_{i1} \text{ and } \cdots \text{ and } x_n \text{ is } A_{\mathbf{in}}\\ &\qquad\text{THEN } y_i \text{ is } m_i\end{aligned} \tag{41}$$

where x_i and y_i are the input and output variables, respectively, A_{ij} is a fuzzy set, and m_i is the position of a symmetric membership function of the output variable with its width neglected during the defuzzification process. Even though such a basic SONFIN can be used directly for system modeling, a large number of rules are necessary for modeling sophisticated systems under a tolerable modeling accuracy. To cope with this problem, we adopt the spirit of the TSK model [29] into the SONFIN. In the TSK model, each consequent part is represented by a linear equation of the input variables. It is reported in [29] that the TSK model can model a sophisticated system using a few rules. Even so, if the number of input and output variables is large, the consequent parts used in the output are quite considerable, some of which may be superfluous. To cope with the dilemma between the number of rules and the number of consequent terms, instead of using the linear equation of all the input variables (terms) in each rule, we add these additional terms only to some rules when necessary. The idea is based on the fact that for different input clusters, the corresponding output mapping may be simple or complex. For simple mapping, a rule with a singleton output is enough, while for complex mapping, a rule with a linear equation in the consequent part is needed. The criterion to deciding which type of consequent part should be used for each rule is based on computing

$$RE(i) = \sum_t \frac{a_i^{(3)}}{\sum_{k=1}^{c} a_k^{(3)}} (y(t) - y^d(t))^2 \tag{42}$$

where $a_i^{(3)}$ is the firing strength of rule i, c is the number of rules, $y^d(t)$ is the desired output, $y(t)$ is the current output, and $RE(i)$ is the accumulated error caused by rule i. By monitoring the error curve, if the error doesn't diminish over a period of time and the error is still too large, we shall add linear combinations of input variables to the rules whose $RE(i)$ values are larger than a predefined threshold value. The process may be done repeatedly after a period of time until a satisfactory result occurs.

4) Parameter Identification: The parameter identification process is done concurrently with the structure identification process. The idea of backpropagation is used for this supervised learning. Considering the single-output case for clarity, our goal is to minimize the error function

$$E = \tfrac{1}{2}(y(t) - y^d(t))^2 \tag{43}$$

where $y^d(t)$ is the desired output and $y(t)$ is the current output. The parameters, a_{ji}, in layer 4 are tuned by RLS algorithm as

$$\boldsymbol{a}(t+1) = \boldsymbol{a}(t) + P(t+1)\boldsymbol{u}(t+1)(y^d(t) - y(t)) \tag{44}$$

$$P(t+1) = \frac{1}{\lambda}\left[P(t) - \frac{P(t)\boldsymbol{u}^T(t+1)\boldsymbol{u}(t+1)P(t)}{\lambda + \boldsymbol{u}^T(t+1)P(t)\boldsymbol{u}(t+1)}\right] \tag{45}$$

where $0 < \lambda \leq 1$ is the forgetting factor, $\boldsymbol{u}$ is the current input vector, $\boldsymbol{a}$ is the corresponding parameter vector, and P is the covariance matrix. The initial parameter vector $\boldsymbol{a}(0)$ is determined in the structure learning phase and $P(0) = \sigma I$, where σ is a large positive constant. As to the free parameters m_{ij} and σ_{ij} of the input membership functions in layer 2, they are updated by the backpropagation algorithm. Using the chain rule, we have

$$\begin{aligned} m_{ij}^{(2)}(t+1) &= m_{ij}^{(2)}(t) - \eta \frac{\partial E}{\partial m_{ij}^{(2)}} \\ &= m_{ij}^{(2)}(t) - \eta \frac{\partial E}{\partial y} \sum_k \frac{\partial y}{\partial a_k^{(3)}} \frac{\partial a_k^{(3)}}{\partial m_{ij}^{(2)}} \end{aligned} \tag{46}$$

where

$$\begin{aligned} \frac{\partial E}{\partial y} &= y(t) - y^d \\ \frac{\partial y}{\partial a_k^{(3)}} &= \frac{a_k^{(4)} - y}{\left(\sum_i a_i^{(3)}\right)^2} \end{aligned} \tag{47}$$

$$\frac{\partial a_k^{(3)}}{\partial m_{ij}^{(2)}} = \begin{cases} a_k^{(3)} \dfrac{2(x_i - m_{ij})}{\sigma_{ij}^2}, & \text{if term node } j \text{ is} \\ & \text{connected to rule node } k \\ 0, & \text{otherwise.} \end{cases} \tag{48}$$

Similarly, we have

$$\begin{aligned} \sigma_{ij}^{(2)}(t+1) &= \sigma_{ij}^{(2)}(t) - \eta \frac{\partial E}{\partial \sigma_{ij}^{(2)}} \\ &= \sigma_{ij}^{(2)}(t) - \eta \frac{\partial E}{\partial y} \sum_k \frac{\partial y}{\partial a_k^{(3)}} \frac{\partial a_k^{(3)}}{\partial \sigma_{ij}^{(2)}} \end{aligned} \tag{49}$$

where

$$\frac{\partial a_k^{(3)}}{\partial \sigma_{ij}^{(2)}} = \begin{cases} a_k^{(3)} \dfrac{2(x_i - m_{ij})^2}{\sigma_{ij}^3}, & \text{if term node } j \text{ is} \\ & \text{connected to rule node } k \\ 0, & \text{otherwise.} \end{cases} \tag{50}$$

C. Feedforward Learning of Oxide Film Growing

The feedforward learning of the oxide film growing rate is an off-line training process as shown in Fig. 8. The inputs to the SONFIN are the processing temperature (T) and the sampling time (t), and the desired output of the SONFIN is the oxide thickness (X) at the corresponding sampling time and processing temperature. Due to the high nonlinearity of the oxide thickness growth rate at the beginning of processing, we take more training samples in this period. The SONFIN is

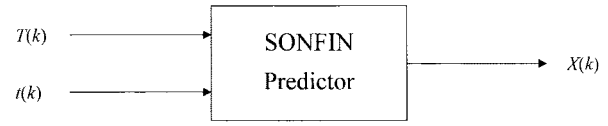


Fig. 8. Feedforward learning of the oxidation growing.

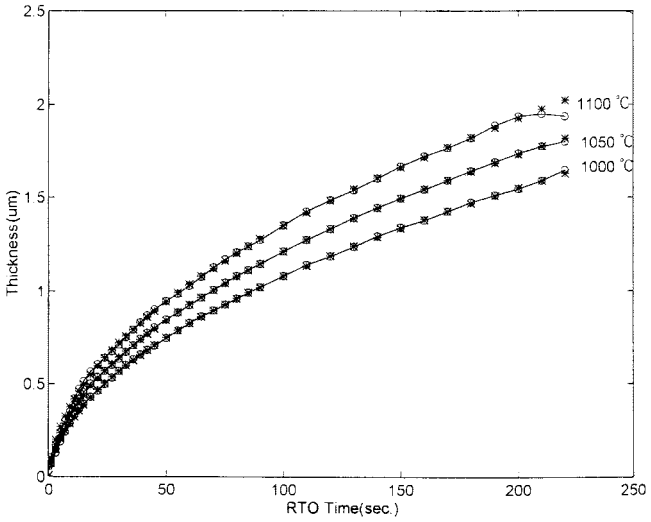


Fig. 9. The result of learning three oxidation growing profiles at 1000, 1050, and 1100 °C simultaneously.

updated supervisedly to minimize an error function E defined by $E = \sum_{k=1}^{n}(1/2)[X_{rd}(k) - \hat{X}(k)]^2$, where $X_{rd}(k)$ is the desired film thickness, $\hat{X}(k)$ is the actual output (predicted film thickness) of the training network and n is the number of training patterns. After some epochs of training, the SONFIN has learned the oxide film growth behavior and can be used to predict the film thickness at given processing temperature and time. Fig. 9 shows the result of learning three oxidation growing profiles at 1000, 1050, and 1100 °C simultaneously, where "$*$" denotes the actual film thickness and "o" denotes the estimated ones by the SONFIN. The average prediction error is smaller than 0.012 μm. Notice that the learned results in Fig. 9 are very close to those obtained by the practical experiment performed in [23].

D. Pyrometer Measurement Correcting Rule

The block diagram of emissivity prediction and pyrometer reading correction is show in Fig. 6, where the wafer temperature is measured by a pyrometer. It is observed from (27) that if we do not predict the emissivity change (ε_w) during processing, then the difference between the pyrometer's measured temperature value and the wafer's actual temperature value can be more than ± 50 °C in few seconds [14]. Hence, a converter is employed to convert the pyrometer reading to a correct temperature value by predicting the wafer's emissivity changes. The SONFIN is used here to predict the current oxide film thickness and then infer the wafer emissivity by the optical model described previously.

According to the temperature-emissivity relation of the pyrometer described by (27), we set the emissivity correction rule for the converter in Fig. 11 as

$$T_c = \frac{1}{\dfrac{1}{T_p} - \dfrac{\lambda}{C_2} \ln \dfrac{\varepsilon_p}{\varepsilon'_w}} \tag{51}$$

where T_c is the corrected temperature value, T_p is the temperature measured by the pyrometer, and ε'_w is the predicted emissivity of the wafer. It can be seen that if the predicted emissivity of the wafer ε'_w is equal to the actual emissivity of the wafer ε_w, then (27) will be the same as (51), i.e., $T_c = T_w$, and, thus, the corrected temperature will be equal to the actual wafer temperature.

In (51), T_p is given by the pyrometer and can be calculated by the pyrometer model in (28) in simulations and ε'_w can be determined immediately from the reflectivity of the opaque wafer, i.e., $\varepsilon'_w = 1-$ reflectivity. According to (8), the current film thickness is required to evaluate the reflectivity R. Because the film thickness cannot be measured directly, it is estimated by the SONFIN predictor, which has been off-line trained before processing. To predict the current film thickness, the previous corrected temperature value $T_c(k)$ and the current processing time k are used as the inputs of the SONFIN. We assume the emissivity of the pyrometer in (51) is a constant value that is identical to the bare silicon wafer (i.e., $\varepsilon_p = 0.7$). Substituting these values described above into (51), the approximate current temperature $T_c(k+1)$ can be inferred and then used in the control loop for wafer temperature control. The temperature control is achieved by another SONFIN. The design of this SONFIN controller is introduced in the next section.

V. Inverse Learning for Neural Fuzzy Controller Design

Of the various techniques for controlling the temperature of the RTP system, model-based control has the greatest potential for attaining the best performance in the field of conventional control when the model is accurate. There have been a number of identification methods presented to help obtaining more accurate RTP models. Nevertheless, it is often difficult to identify the models accurately due to some complex and highly nonlinear situations. Recently, neural networks have been shown to possess good capability to adaptively control a nonlinear model. To overcome the drawback of the model-based control, we apply the inverse learning method, originally reported in the neural network literature, to the design of the neural fuzzy controller for the temperature control of the RTP system. The key advantage of the neural fuzzy approach over traditional ones lies on it does not require a mathematical description of the system while controlling. Again, we shall use the SONFIN as the neural fuzzy controller here. In [31], we have compared the temperature control performance of several kinds of controllers on the RTP system, including the backpropagation (BP) neural network controller, model-reference adaptive controller (MRAC), proportional-derivative (PD) controller, and our SONFIN. With the performance indexes of the nonuniformity, maximum positive error, maximum negative error, and mean-square tracking error, we

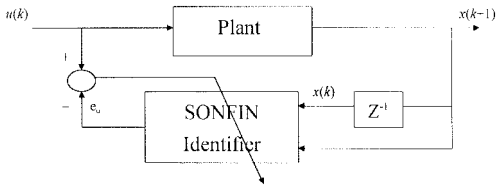


Fig. 10. Block diagram of the learning phase of the inverse learning control method.

found that the SONFIN had the best performance. Hence, in this paper we adopt the SONFIN controller in the RTP system to check the effectiveness of the SONFIN prediction for pyrometer-reading correction. After a while we shall use a SONFIN as a pyrometer-reading corrector and another SONFIN as the temperature control for the RTP system in this section. The simulation results will be compared to those of other approaches.

A. Inverse Learning of the RTP System

The direct inverse learning configuration shown in Fig. 10 is adopted in our approach, where off-line training is used for the design of the controller. The inverse learning of the SONFIN involves two phases. In the learning phase, the SONFIN is trained to model the inverse dynamics of the plant. The obtained neural fuzzy model (which represents the inverse dynamics of the plant) is then used to generate control actions in the application phase.

Suppose the plant of RTP model is described by

$$y(k) = plant(y(k-1), u(k)) \tag{52}$$

where $y(k)$ and $y(k-1)$ are the wafer temperatures at time k and $k-1$, respectively, and $u(k)$ is the control signal at time k. Here we assume that the dynamics of the plant is unknown and we are going to build a SONFIN that maps a given pair $(y(k), y(k+1))$ to a desired control action $u(k)$. This mapping is not easily expressed as analytic formula because the plant is nonlinear and system parameters vary with current temperature.

For off-line training, we have to collect a set of training data pairs and then train the SONFIN in the batch mode. A sequence of random input signals $u(k)$ under the magnitude limits of the plant input is injected directly to the plant and then an open-loop input/output characteristic of the plant is obtained. According to the input/output characteristic of the plant, proper training patterns are selected to cover the entire reference output space. Using the collected training patterns with the values of the selected input variables as the input pattern and the corresponding control signal $u_{rd}(k)$ as the target pattern, the SONFIN can be updated supervisedly to minimize an error function E defined by $E = \sum_{k=1}^{kn}(1/2)[u_{rd}(k) - \hat{u}(k)]^2$ according to the learning algorithm developed in Section IV-B, where k_n is the number of training patterns and $\hat{u}(k)$ is the actual output of the training network.

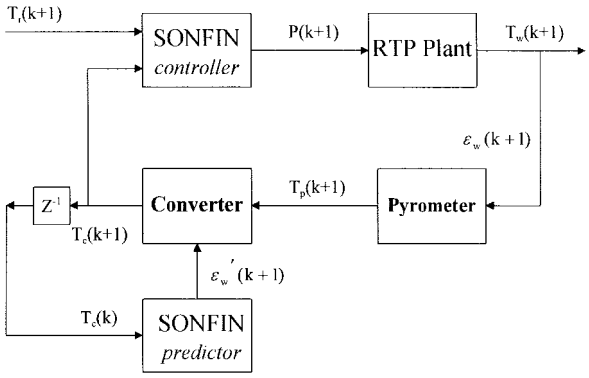


Fig. 11. Block diagram of the overall RTP temperature control system.

B. Simulation Results and Performance Comparisons

Fig. 11 shows the block diagram of close-loop temperature control of the RTP system. The whole system consists of a SONFIN controller, which has been trained in the inverse learning process described in Section V-A, and a SONFIN predictor which has been trained through feedforward learning described in Section IV-C.

The SONFIN controller is to control the wafer temperature to follow a desired profile. The SONFIN predictor is to predict the film thickness of the wafer at any processing time such that the actual emissivity of the wafer can be calculated for correcting the pyrometer reading. Several experimental simulations have been done with this structure and the results are presented in this section. We also replace the SONFIN predictor with a normal linear predictor in the structure shown in Fig. 11 for performance comparisons.

In the first simulation, we try to control the wafer temperature to follow the desired temperature profile shown in Fig. 12(a), where the desired final process temperature (operating point) is 1050 °C. For this purpose, the SONFIN predictor for correcting the pyrometer reading is trained to learn three oxidation profiles obtained by setting the wafer temperature at 1000, 1050, 1100 °C, respectively. Learning all the three profiles around the operating point 1050 °C can enhance the generalization capability of the SONFIN predictor. The control performance of this simulation is shown in Fig. 12. The actual wafer temperature (T_w) under the SONFIN control as well as the desired temperature profile are shown in Fig. 12(a). It is observed that the actual wafer temperature profile follows the desired one closely. In this figure, we also show the SONFIN controller's output signals. To see the control performance more closely, Fig. 12(b) shows the difference between the actual and desired wafer temperatures, where two measure points are set, the center of the wafer (solid line) and the margin of the wafer (dashed line). To see the effect of the SONFIN predictor in correcting the pyrometer reading, the difference between the pyrometer reading T_c and the actual wafer temperature T_w are shown in Fig. 12(c) as the solid line. In the same figure, the dashed line shows the difference between the uncorrected pyrometer reading T_p and T_w. We can see that the SONFIN predictor provides the controller with more accurate wafer temperature information.

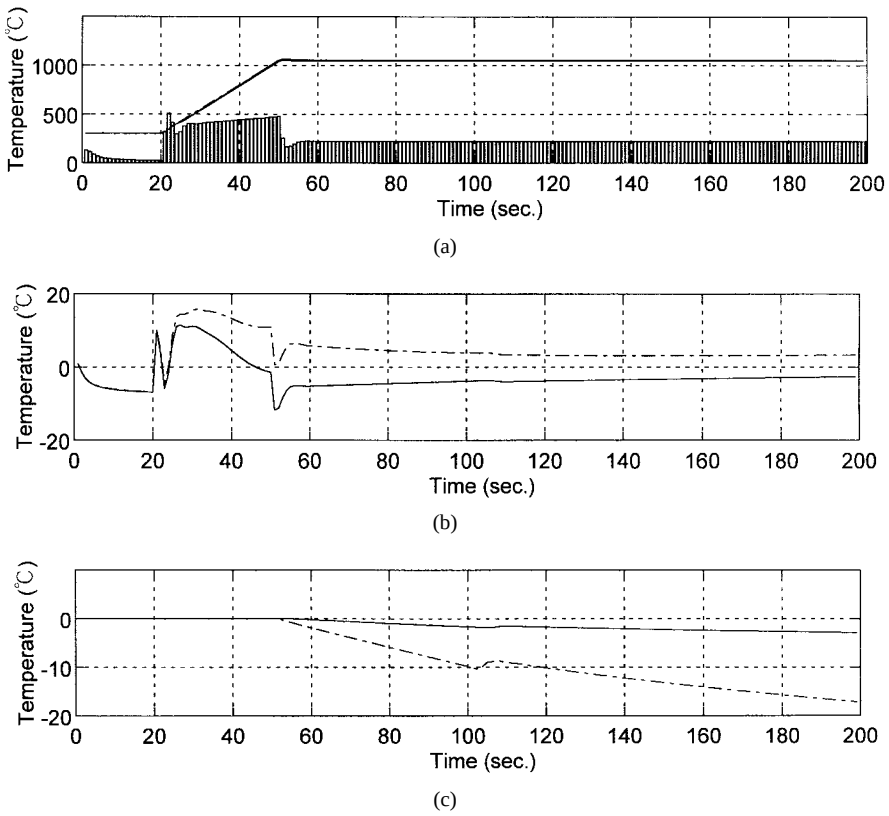


Fig. 12. Simulation results of using SONFIN controller and SONFIN predictor with the wafer temperature set at 1050 °C. (a) The desired temperature profile (dashed line), the actual wafer temperature (T_w) under control (solid line), and the SONFIN controller's output signals (bar lines). (b) The difference between the desired temperature and the actual wafer temperature T_w where the solid line denotes the temperature measured at the center of the wafer and the dashed line the margin of the wafer. (c) The difference of the actual wafer temperature T_w and the corrected measured temperature value T_c (solid line); the difference of the actual wafer temperature T_w and the pyrometer-measured temperature T_p (dashed line).

To test the generalization capability of the SONFIN predictor, we set the operating point of the RTP system at 900 °C in the second simulation; the desired temperature profile is shown in Fig. 13(a) (dashed line). In this simulation, we use the same SONFIN predictor and controller as those in the first simulation. Notice that since the SONFIN controller is trained by the inverse learning scheme stated in the last section—it is independent of the operation point. On the contrary, the prediction power of the SONFIN predictor focuses only on the temperature region in which it is trained. The results of this simulation, corresponding to those of the first simulation, are show in Fig. 13. The figure shows that our control system still achieves quite good performance, although the operation point lies outside the region, [1000 °C, 1100 °C], in which the SONFIN predictor was trained.

To see the role played by the SONFIN predictor, we replace it by a normal linear predictor described by

$$X_{k+1} = X_k + p \times X'(T_{ck}) \tag{53}$$

where X_{k+1} is the oxide thickness at time $k+1$, X_k is the oxide thickness at time k, p is the sampling period, and $X'(T_{ck})$ is the growth rate at temperature T_{ck}. The corrected temperature from the previous control sequence T_{ck} is used to predict the current film thickness which is in turn used to approximate the current temperature T_{ck+1}. The linear predictor assumes that the oxide has a constant growth rate. The corresponding control performance is shown in Fig. 14. Fig. 14 shows that the temperature errors are larger than those of the previous two simulations. Especially, the errors are increasing as processing time is running. This is obviously due to the big temperature measurement errors shown in Fig. 14(c). It is understood that the bigger temperature measurement errors are due to the lower prediction accuracy of the linear predictor on the oxide thickness as compared to that of the SONFIN predictor [see Figs. 12(c) and 14(c)].

As a contrast, in the fourth simulation we take off the predictor and converter in Fig. 11 and feed the pyrometer's reading value into the SONFIN controller directly. In other words, we want to see what happens if we don't correct the pyrometer reading. The simulation results are shown in Fig. 15. Due to the inaccurate temperature measurement shown in Fig. 15(b), the SONFIN controller fails to track the desired temperature profile [see Fig. 15(a)].

Finally, to compare the performance of a traditional approach, we replace the SONFIN controller by the proportional integral derivative (PID) controller and the SONFIN predictor by the linear predictor [see (53)] in our control system in Fig. 11. We have tried our best to tune the PID controller to achieve its best performance and the simulation results are shown in Fig. 16. The curves in the figure obviously show the inferior performance of this traditional control architecture. The control system cannot track the desired temperature profile

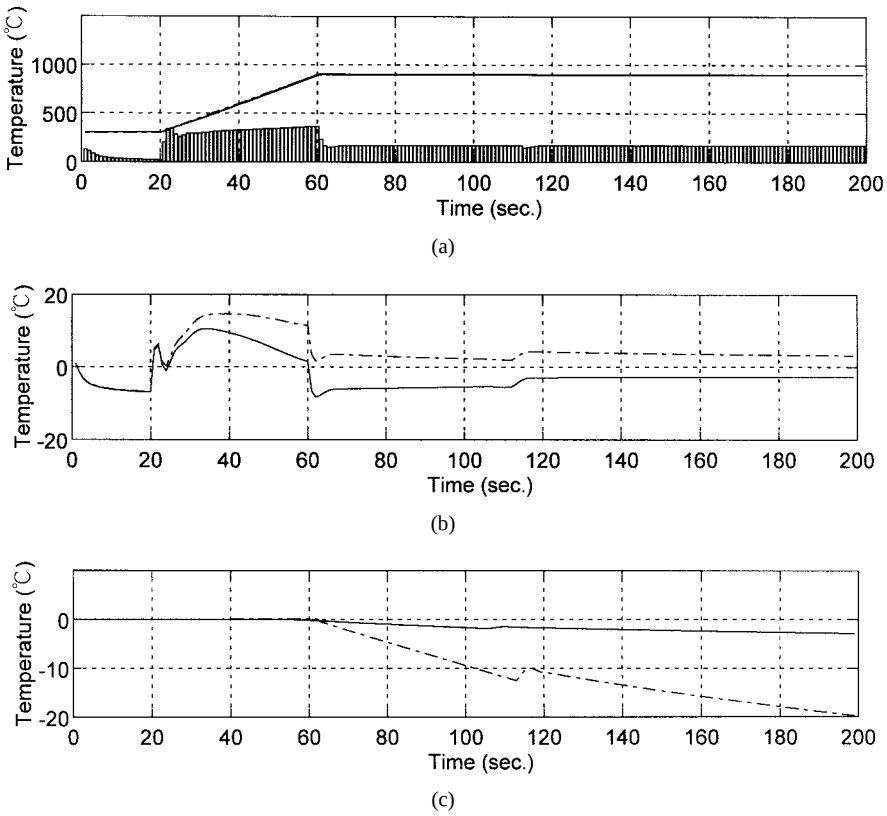


Fig. 13. Simulation results of using SONFIN controller and SONFIN predictor with the wafer temperature set at 900 °C. (a) The desired temperature profile (dashed line), the actual wafer temperature (T_w) under control (solid line), and the SONFIN controller's output signals (bar lines). (b) The difference between the desired temperature and the actual wafer temperature T_w where the solid line denotes the temperature measured at the center of the wafer and the dashed line the margin of the wafer. (c) The difference of the actual wafer temperature T_w and the corrected measured temperature value T_c (solid line); the difference of the actual wafer temperature T_w and the pyrometer-measured temperature T_p (dashed line).

accurately and stablely. It was found in our trials that the oscillations in Fig. 16 were unavoidable. Two major reasons account for this phenomenon; first, we performed long-time control; second, we performed temperature-trajectory tracking control instead of set-point control. The oscillation phenomenon of PID control was also observed by others in [1], [14], and [17]. Fig. 16 also shows increasing tracking errors as processing time is running due to the low prediction accuracy of the linear predictor. It is observed that the simulation results shown in Figs. 12–16 are very similar to those obtained by others' practical experiments [1], [14], [26].

The results of the above simulations are summarized in Table II. This table lists the maximum/minimum tracking errors, the mean square tracking error (tracking MSE), and the temperature nonuniformity of different control structures, where the nonuniformity is given by the difference of the two controlled temperatures measured at the center and the margin of the wafer. From the table we see that the SONFIN controller with SONFIN predictor achieves the smallest tracking error and keeps a low nonuniformity. Across wafer uniformity is one major issue in RTP processing. To obtain uniform processing across the wafer surface and to prevent the creation of slip defects due to the thermal stress, the temperature must be nearly uniform across the wafer at all time. The simulations in this section have shown that the SONFIN-based scheme can keep better temperature uniformity than the traditional PID scheme. However, we observe from the controlled results that the temperature gradient difference between the inner and outer rings of wafer is not small enough. In fact, to further reduce the nonuniformity, various means have been suggested, as for example, modification of reflector characteristics, special lamp arrangements, individual lamp powering, or mechanical movement of the wafer [31]. Thermal gradients may also be reduced by using either guard rings or suspector, i.e., by virtually extending the wafer edge. The edge loss radiation energy of the wafer will be reflected by the guard ring and the difference between the outer and edge energy will be reduced. Another approach to solving the nonuniformity problem thoroughly is to use a multioutput controller. In this approach, a circular bank of lamps is added over the wafer border to compensate the edge-loss effect of the wafer. The lamps emphasize the incident radiation energy on the wafer edge and also add the energy at the center. A multi-output controller is required to control the power of different lamps individually. In any approach of improving uniformity mentioned in the above, the proposed SONFIN-based scheme can be applied to get even better performance than the traditional PID scheme.

VI. Conclusions

This paper proposes the use of the neural fuzzy network (called SONFIN) for solving the two difficult problems in the RTP system, i.e., temperature measurement and control.

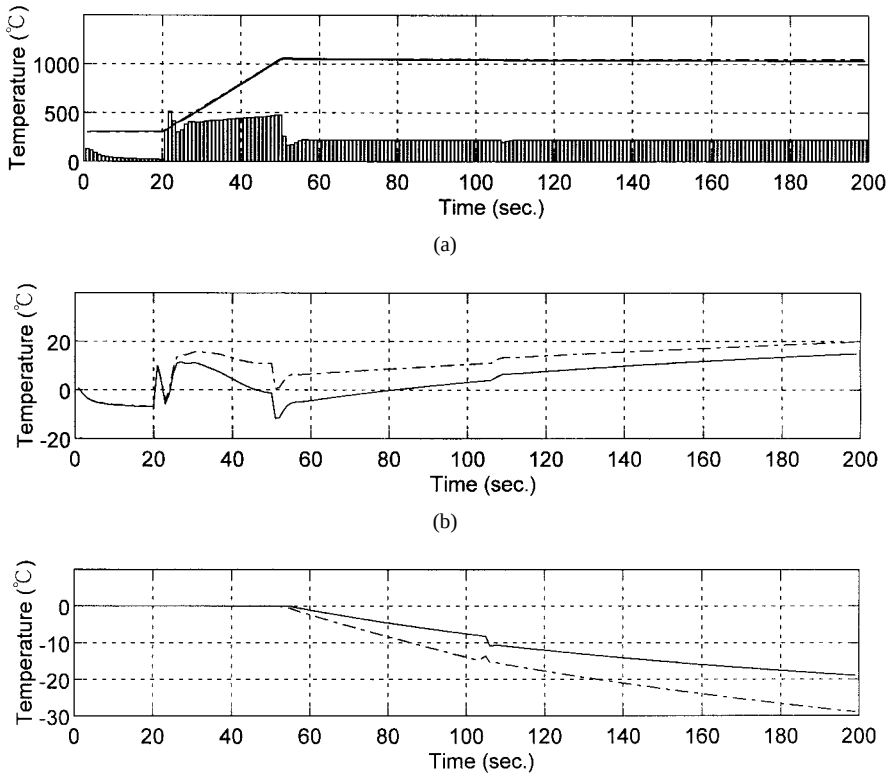


Fig. 14. Simulation results of using SONFIN controller and SONFIN predictor with the wafer temperature set at 1050 °C. (a) The desired temperature profile (dashed line), the actual wafer temperature (T_w) under control (solid line), and the SONFIN controller's output signals (bar lines). (b) The difference between the desired temperature and the actual wafer temperature T_w where the solid line denotes the temperature measured at the center of the wafer and the dashed line the margin of the wafer. (c) The difference of the actual wafer temperature T_w and the corrected measured temperature value T_c (solid line); the difference of the actual wafer temperature T_w and the pyrometer-measured temperature T_p (dashed line).

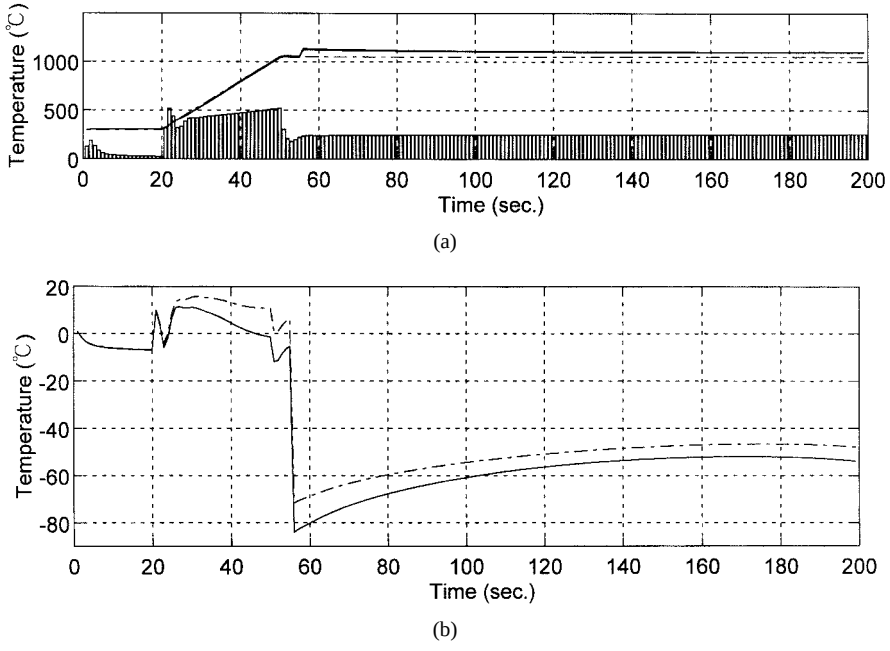


Fig. 15. Simulation results of using SONFIN controller and SONFIN predictor with the wafer temperature set at 1050 °C. (a) The desired temperature profile (dashed line), the actual wafer temperature (T_w) under control (solid line), and the SONFIN controller's output signals (bar lines). (b) The difference between the desired temperature and the actual wafer temperature T_w where the solid line denotes the temperature measured at the center of the wafer and the dashed line the margin of the wafer.

The SONFIN is a general connectionist model of a fuzzy inference system, which can find its optimal structure and parameters automatically. Two SONFIN's are used in the proposed control structure. One is to learn the oxidant growth rate as a function of oxidation time at various processing temperatures for correcting the measurement error of the

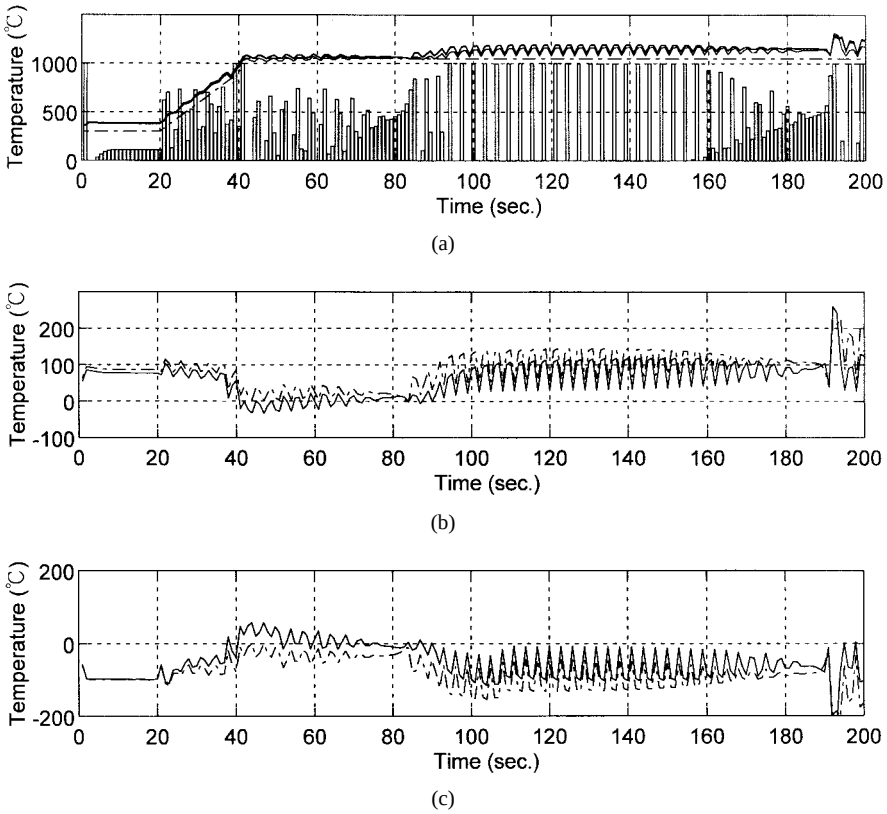


Fig. 16. Simulation results of using SONFIN controller and SONFIN predictor with the wafer temperature set at 1050 °C. (a) The desired temperature profile (dashed line), the actual wafer temperature (T_w) under control (solid line), and the SONFIN controller's output signals (bar lines). (b) The difference between the desired temperature and the actual wafer temperature T_w, where the solid line denotes the temperature measured at the center of the wafer and the dashed line the margin of the wafer. (c) The difference of the actual wafer temperature T_w and the corrected measured temperature value T_c (solid line); the difference of the actual wafer temperature T_w and the pyrometer-measured temperature T_p (dashed line).

TABLE II
SUMMARY TABLE OF PERFORMANCE INDEX

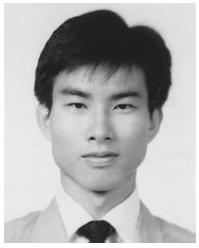
	SONFIN controller with SONFIN predictor	SONFIN controller with linear predictor	SONFIN controller without predictor	PID controller with linear predictor
Max. tracking error	11.4098	14.8810	11.5012	200.9889
Min. tracking error	-6.7468	-11.7439	-80.7679	-55.0237
Nonuniformity	5.5033	4.7881	4.4996	177.7
Tracking MSE	34.3811	73.4889	259.80	2037.0

pyrometer due to emissivity variations during film growing. The other SONFIN is to model the inverse of the RTP system for temperature control. To compensate the pyrometer errors, we also derive the wafer emissivity model, silicon oxidation growing model and pyrometer model. Several experimental simulations have been done and the performance comparisons to other control structures are also presented in this paper. The simulation technique is based on the desired accurate RTP global model, which includes the temperature sensor, control loop, and lamp system. From the simulation results, we see that the SONFIN controller with SONFIN predictor has the best performance. Although we test the performance of the proposed scheme on the RTO process only in this paper due to the shortage of available experimental data, we can also apply our scheme to other process such as RTCVD. However, since the RTCVD process has a higher emissivity change rate than the RTO process, the prediction accuracy and control performance of the proposed SONFIN-based scheme needs further investigation and improvement in the future.

REFERENCES

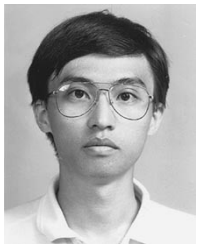
[1] S. Yu, F. Y. Sorrell, and W. J. Keither, "Temperature control strategies for RTP systems," in *Proc. 10th Biennial Univ./Govt./Indus. Microelectron. Symp.*, Research Triangle Park, NC, May 1993, pp. 229–234.
[2] R. Deaton and H. Z. Massoud, "Manufacturability of rapid-thermal oxidation of silicon: Oxide thickness, oxide thickness variation, and system dependency," *IEEE Trans. Semiconduct. Manuf.*, vol. 5, pp. 347–358, Nov. 1992.
[3] C. F. Juang and C. T. Lin, "An on-line self-constructing neural fuzzy inference network for system modeling," *IEEE Trans. Fuzzy Syst.*, vol. 6, pp. 12–32, Feb. 1998.
[4] J. C. Gelpey, P. O. Stump, and J. W. Smith, "Process control for a rapid optical annealing system," in *Proc. Mater. Res. Soc. Symp.*, Boston, MA, Dec. 1985, vol. 52, pp. 199–207.
[5] J. Nulman, "*In-situ* processing of silicon dielectrics by rapid thermal processing: Cleaning, growth, and annealing," in *Proc. Mater. Res. Soc.*

Symp., 1987, vol. 92, pp. 141–146.
[6] J. Nulman, J. P. Krusius, and A. Gat, "Rapid thermal processing of thin gate dielectrics: Oxidation of silicon," *IEEE Electron Device Lett.*, vol. EDL-6, pp. 205–207, May 1985.
[7] M. C. Ozturk, J. J. Wortman, Z. Yu-Lin, R. Xiao-Wei, R. M. Miller, J. F. Scott, D. T. Grider, and D. A. Abercrombie, "Low-pressure chemical vapor deposition of polycrystalline silicon and silicon dioxide by rapid thermal processing," North Carolina State Univ., Tech. Rep., May 1989.
[8] T. Breedijk, "Model identification and nonlinear predictive control of rapid thermal processing systems," Ph.D. dissertation, Univ. Texas, Austin.
[9] S. A. Norman, "Optimization of transient temperature uniformity in RTP system," *IEEE Trans. Electron Devices*, vol. 39, pp. 205–207, Jan. 1992.
[10] F. Roozeboom and N. Parekh, "Rapid thermal processing systems: A review with emphasis on temperature control," *J. Vacuum Sci. Technol.*, vol. B8, pp. 1249–1259, 1990.
[11] C. C. Gibson. D. P. Dewitt, and F. Y. Sorrell. "In-process temperature measurement of silicon wafers," in *7th Symp. Temp., Measurement, Contr. Sci. Indust.*, Toronto, ON, Canada, Apr. 28–May 1, 1992.
[12] K. C. Saraswat, P. P. Apte, L. Booth, Y. Chen, P. C. P. Dankoski, F. L. Degertekin, G. F. Franklin, B. T. Khuri-Yakub, Y. J. Lee, J. Pei, and S. C. Wood, "Rapid thermal multiprocessing for a programmable factory for adaptable manufacturing of IC's," *IEEE Trans. Semiconduct. Manuf.*, vol. 7, pp. 159–175, May 1994.
[13] V. M. Donnellu and J. A. McCaulley, "Infrared-laser interferometric thermometry: A nonintrusive technique for measuring semiconductor wafer temperatures," *J. Vac. Sci. Technol. A.*, vol. 8, no. 1, pp. 84–92, 1990.
[14] F. Y. Sorrell and R. S. Gyurcsik, "Model-based emissivity correction in pyrometer temperature control of rapid thermal processing systems," *IEEE Trans. Semiconduct. Manuf.*, vol. 6, p. 273, Aug. 1993.
[15] Y. M. Cho and T. Kailath, "Model identification in rapid thermal processing systems," *IEEE Trans. Semiconduct. Manuf.*, vol. 6, pp. 233–245, Aug. 1993.
[16] H. A. Lord, "Thermal and stress analysis of semiconductor wafers in a rapid thermal processing oven," *IEEE Trans. Semiconduct. Manuf.*, vol. 1, pp. 105–114, Aug. 1988.
[17] F. Y. Sorrell, "Temperature uniformity in RTP furnace," *IEEE Trans. Semiconduct. Manuf.*, vol. 39, pp. 75–80, Jan. 1992.
[18] T. Sato, "Spectral emissivity of silicon," *Japan J. Appl. Phys.*, vol. 6, no. 3, pp. 339–347, 1991.
[19] C. Hill, S. Jones, and D. Boys, "Rapid thermal annealing-theory and practice," in *Reduced Thermal Processing for ULSI*, R. A. Levy, Ed. New York: Plenum, 1988.
[20] P. Vandebabeele and K. Maex, "Temperature nonuniformities during rapid thermal processing of patterned wafers," in *Rapid Isothermal Processing*, R. Singh, Ed. Bellingham, WA: SPIE, 1990, vol. 1189, pp. 89–103.
[21] M. Born and E. Wolf, *Principles of Optics.* Elmsford, NY: Pergamon, 1980.
[22] J. M. Stone, *Radiation and Optics.* New York: McGraw-Hill, 1963.
[23] B. E. Deal and A. S. Grove, "General relationship for the thermal oxidation of silicon," *J. Appl. Phys.*, vol. 36, p. 3770, 1965.
[24] B. E. Deal, A. Hurrle, and M. J. Schulz, "Chlorine concentration profiles in O2/HCI and H2O/HCI thermal silicon oxides using SIMS measurements," *J. Electrochem. Soc.*, vol. 125, no. 12, pp. 2024–2027, Dec. 1978.
[25] D. Psalitis, A. Sideris, and A. Yamamura, "A multilayered neural network controller," *IEEE Contr. Syst. Mag.*, vol. 10, pp. 44–48, Apr. 1989.
[26] F. Y. Sorrell, S. Yu, and W. J. Kiether, "Applied RTP optical modeling: An argument for model-based control," *IEEE Trans. Semiconduct. Manuf.*, vol. 7, pp. 454–459, Nov. 1994.
[27] C. T. Lin and C. S. G. Lee, "Neural-network-based fuzzy logic control and decision system," *IEEE Trans. Comput.*, vol. 40, pp. 1320–1336, Dec. 1991.
[28] ———, *Neural Fuzzy Systems: A Neural-Fuzzy Synergism to Intelligent Systems*. Englewood Cliffs, NJ: Prentice-Hall, 1996 (with disk).
[29] T. Talagi and M. Sugeno, "Fuzzy identification of systems and its applications to modeling and control," *IEEE Trans. Syst., Man, Cybern.*, vol. 15, pp. 116–132, Jan. 1985.
[30] D. W. Pettibone, J. R. Surarez, and A. Gat, "The effect of thin dielectric films on the accuracy of pyrometer temperature measurement," in *Proc. Mater. Res. Soc. Symp.*, Boston, MA, Dec. 1985, vol. 52.
[31] C. T. Lin, C. F. Juang, and J. C. Huang, "Temperature control of rapid thermal processing system using adaptive fuzzy network," *Fuzzy Sets Syst.*, to be published.
[32] D. H. Peyton, H. Kinoshita, G. Q. Lo, and D. L. Kwong, "Systems oriented survey of noncontact temperature measurement techniques for rapid thermal processing," in *Rapid Thermal and Related Processing Techniques*, R. Singh and M. M. Moslehi, Eds. Bellinham, WA: SPIE, 1991, vol. 1393, pp. 295–309.
[33] P. J. Timans, "The role of thermal radiative properties of semiconductor wafers in rapid thermal processing, " in *Proc. Mater. Res. Soc. Symp.*, San Francisco, CA, Apr. 1996, vol. 429, pp. 3–14.



Jiun-Hong Lai received the B.S. degree in electrical engineering from the Chung-Hua University, Hsinchu, Taiwan, R.O.C., in 1996, and the M.S. degree in electrical and control engineering from the National Chiao-Tung University, Hsinchu, Taiwan, R.O.C., in 1997.

From December 1997 to October 1998, he was with the College of Agriculture, National Taiwan University, Taipei, where he was a Research Assistant of agricultural machinery engineering of neural simulation and the effect of the environmental conditions on seedling quality. His research interests are in fuzzy systems, neural networks, thermal modeling and temperature control of RTP systems.



Chin-Teng Lin (S'88–M'91) received the B.S. degree in control engineering from the National Chiao-Tung University, Hsinchu, Taiwan, R.O.C., in 1986, and the M.S.E.E. and Ph.D. degrees in electrical engineering from Purdue University, West Lafayette, IN, in 1989 and 1992, respectively.

Since August 1992, he has been with the College of Electrical Engineering and Computer Science, National Chiao-Tung University, Hsinchu, Taiwan, ROC, where he is currently a Professor of electrical and control engineering. He also serves as the Deputy Dean of the Research and Development Office of the National Chiao-Tung University since 1998. His current research interests are fuzzy systems, neural networks, intelligent control, human-machine interface, and video and audio processing. He is the coauthor of *Neural Fuzzy Systems—A Neuro-Fuzzy Synergism to Intelligent Systems* (Englewood, NJ: Prentice-Hall, 1996) and the author of *Neural Fuzzy Control Systems with Structure and Parameter Learning* (Singapore: World Scientific, 1994). He has published over 30 journal papers in the areas of neural networks and fuzzy systems.

Dr. Lin is a member of Tau Beta PI and Eta Kappa Nu. He is also a member of the IEEE Computer Society, the IEEE Robotics and Automation Society, and the IEEE Systems, Man and Cybernetics Society. He has been the Executive Council Member of the Chinese Fuzzy System Association (CFSA) since 1995 and the Supervisor of Chinese Automation Association since 1998. He was the Vice Chairman of IEEE Robotics and Automation Taipei Chapter in 1996 and 1997. He won the Outstanding Research Award granted by the National Science Council (NSC), Taiwan, in 1997, and the Outstanding Electrical Engineering Professor Award granted by the Chinese Institute of Electrical Engineering (CIEE) in 1997.