to appear, 1992 American Control Conference, Chicago

Multivariable Feedback Control of Semiconductor Wafer Temperature

Stephen A. Norman and Stephen P. Boyd*
Information Systems Lab., Dept. of Electrical Engineering
Stanford University, Stanford CA 94305

Abstract

Rapid thermal processing (RTP) of semiconductor wafers involves processing single wafers in small reaction chambers. Powerful lamp arrays are used to heat wafers quickly. The nature of RTP makes control of wafer temperature a difficult problem; ensuring near-uniform temperature across the wafer is critical.

After a brief introduction to RTP and the RTP temperature control problem, this paper presents a novel approach to the analysis of candidate sensor configurations for multi-actuator, multi-sensor feedback control systems needed to ensure small temperature error in the face of substantial disturbances.

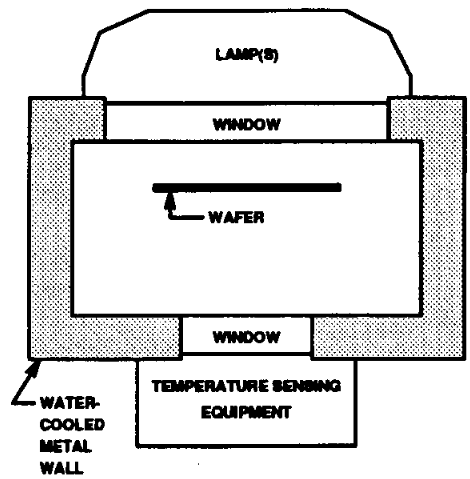

Figure 1: Cross-section of a generic RTP system. Not shown in the diagram: gas inlets and outlets; quartz pins to support the wafer.

1 Introduction

1.1 Rapid Thermal Processing

In rapid thermal processing (RTP) of semiconductor wafers, single wafers are heated, processed, and cooled in small reaction chambers. RTP is a promising new technology for a wide variety of integrated circuit (IC) fabrication steps, including cleaning, annealing, oxidation, nitridation, chemical vapor deposition and others [1, 2, 3].

Heat treatment of wafers in current mainstream IC processing is achieved by means of large, hot-wall ovens in which many wafers are processed simultaneously. Processing times for wafer batches range from tens of minutes to many hours for a single process step; speed is limited by the large thermal masses of the oven walls.

In RTP, only the relatively small thermal mass of the wafer itself is heated to and cooled from processing temperature; the walls of the reaction chamber are water-cooled and remain at room temperature. Consequently, process steps may require only tens of seconds for completion. A diagram of a generic RTP system is given in Figure 1. The wafer is heated by visible and infrared radiation supplied by a powerful lamp or lamp array, which is separated from the chamber by a transparent window. Wafer processing temperatures may be as high as 1200°C. Power requirements for the lamp array are typically several tens of kW; the exact power requirements are determined by wafer size, chamber geometry, and process specifications.

During heat treatment of wafers, dopant ions will undesirably diffuse from their intended locations into neighboring regions; problems caused by this diffusion become more acute as circuit feature size shrinks. The main advantage of RTP over batch heating, then, from a processing point of view, is the shorter time spent by the wafer at high temperature in RTP. From a manufacturing viewpoint, RTP has the additional advantage that it fits naturally into multiprocessing or cluster-tool systems which promise to provide IC fabrication lines that are smaller, more flexible, and less expensive than current IC fabs [4].

1.2 Temperature Control in RTP

A major impediment to acceptance of RTP in mainstream IC fabrication is the current lack of adequate control of wafer temperature during processing. There are two main (and related) requirements for RTP wafer

*This research was supported by the Advanced Research Projects Agency of the Department of Defense and was monitored by the Air Force Office of Scientific Research under Contract No. F49620-90-C-0014. This manuscript was submitted for publication with the understanding that the U. S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation hereon. The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either express or implied, of the Advanced Research Projects Agency or the U. S. Government.

temperature control: (1) the wafer temperature must closely follow a pre-specified process trajectory and (2) the spatial temperature variation over the wafer must be small at all times. The former requirement must be met to ensure wafer-to-wafer process repeatability; the latter must be satisfied to ensure both that processing is uniform across the wafer surface and that defects due to thermal stress are not created in the wafer.

Because in RTP the wafer is very far from being in thermal equilibrium with its surroundings, the problem of temperature nonuniformity over the wafer is much more acute in RTP than in hot-wall oven processing. Heat conduction within the wafer is much too small to equalize temperature across the wafer, so it is critical to ensure that at all times during the RTP process cycle the spatial distribution over the wafer of incident radiative energy closely matches the distribution required for perfect trajectory following. The ideal distribution of incident flux over the wafer varies substantially with wafer temperature, the rate of temperature change and process gas composition and pressure. Failure to properly distribute radiative energy over the wafer surface can result in temperature differences across the wafer of over 50°C; such temperature differences during processing will render a wafer useless. Exact requirements for uniformity depend on the particular process, but typical bounds on uniformity range from ±1°C to ±5°C. This problem of spatial temperature nonuniformity has been widely studied and is explained well in more than one paper [2, 5, 6].

1.3 Temperature Control Actuators

With most currently available RTP systems it is impossible to achieve the wafer temperature uniformity that IC fabrication processes require. From a control engineer's perspective this failure can be regarded as the consequence of poor actuator selection. Most RTP systems have a single wafer temperature control actuator — one big lamp or an array of lamps with a single lamp power controller — and with this single actuator the trajectory of a distributed state must be controlled. Use of a lamp array with multiple, independently-controlled lamp zones to provide a dynamically adjustable flux pattern over the wafer has been proposed, analysed, and experimentally studied by the authors and colleagues [7, 8, 9, 10, 11]. It has been shown that some such multiple-actuator arrangements can deliver satisfactorily small temperature nonuniformity over a wide range of process conditions during both steady-state hold and fast transients.

1.4 Temperature Sensors

Temperature measurements at various points on the wafer during RTP are necessary for identification of system parameters and for feedback control of the temperature profile. At the time of writing, accurate multi-point temperature sensing methods for generic use in RTP have not been publicly demonstrated. Wafer temperature measurement is difficult because sensors placed near the wafer will contaminate the wafer and/or be destroyed by the harsh environment of the processing chamber.

The most popular approach to temperature sensing is the use of one or more pyrometers located outside the chamber. The most serious problem with pyrometry is the fact that the emissivity of the wafer surface will change from one wafer to the next and, more seriously, may change drastically during processing of a single wafer. See [1] for a survey of proposed temperature measurement techniques.

For RTP temperature control system experiments it is possible to use special wafers to which thermocouples have been attached as temperature sensors. Such wafers may be used only in when the chamber contains a vacuum or an inert gas such as pure nitrogen. Because of cost and contamination problems thermocouples can not be attached to wafers that are actually being processed.

It is assumed here that accurate measurements of wafer temperature at discrete points will be available in the future, so that it is valuable to investigate the problem of how best to use these measurements for feedback control.

2 RTP Temperature Dynamics

This section presents a brief summary of the thermal model used in this and other work by the authors on RTP temperature control system analysis. The modeling approach taken here is an extension of that used in [5]. A complete description of the model along with many references to work on RTP system thermal modeling can be found in [8].

We consider here RTP systems in which the reaction chamber and lamp array are effectively axisymmetric and coaxial with the wafer. A cylindrical coordinate system is used with the z-axis coincident with the wafer axis. Due to symmetry, the temperature at a point (r, θ, z) within the wafer will be depend only on r and z. The small temperature variation through the thickness of the wafer can be neglected, meaning that spatially the temperature depends only on radial position r. The wafer temperature dynamics can thus be modeled using a system of first-order ordinary differential equations (ODE's); the state of this system is a vector of temperatures over a grid of radial positions.

In the model, the modes of heat flow into and out of a wafer element are: radiative loss proportional to the fourth power of the absolute temperature of the element and to the exposed surface area of the element; convective loss proportional to the difference between the element temperature and the ambient gas temperature and to the exposed surface area of the element;

heat conduction to neighboring elements; and absorption of lamp radiation and wafer radiation reflected by chamber walls.

Let $\mathbf{T}(t)$ denote the $I \times 1$ vector of temperatures at time t over the grid of radial positions. The system of ODE's relating $\mathbf{T}(t)$ to the $J \times 1$ vector of lamp power settings $\mathbf{P}(t)$ can be written as

$$\dot{\mathbf{T}}(t) = C(\mathbf{T})^{-1} \Big(\mathbf{f}(\mathbf{T}(t)) + L\mathbf{P}(t)\Big) \qquad (1)$$

where the i,i-entry of the $I \times I$ diagonal matrix $C(\mathbf{T})$ is $m_i C_P(T_i)$, the heat capacity of wafer element i; m_i is the mass of element i; and C_P is the specific heat of silicon as a function of temperature. The smooth, nonlinear function $\mathbf{f}$ gives the net heat flux into each wafer element due to radiative and convective loss from the wafer surface, radial heat conduction within the wafer, and radiation emitted by the wafer and reflected back to the wafer. The product $L\mathbf{P}(t)$, where L is an $I \times J$ matrix, represents heat flux into the wafer due to the lamps; the implicit assumption of linearity is valid if the total reflectivities and emissivities of the chamber and wafer surfaces are close to constant with respect to temperature and wavelength; this is often a reasonable approximation for RTP of Si with the wafer temperature over 600°C. Let $\mathbf{P}^{\min}$ and $\mathbf{P}^{\max}$ denote the vectors of minimum and maximum values for the lamp powers; physics dictates that all of the entries of $\mathbf{P}^{\min}$ will be nonnegative.

3 Feedback/Feedforward Temperature Control

Open-loop wafer temperature control in RTP can be designed using a thermal model or by trial and error with a test wafer with embedded thermocouples. Methods for using a thermal model to design lamp power trajectories for wafer temperature tracking with small spatial temperature nonuniformities are presented and discussed in [8] and [9]; in particular, these papers present methods for minimizing worst-case temperature error across the wafer in steady-state and across the wafer and over time during during transients. Results of successful experiments using trial and error to design lamp power trajectories for open-loop wafer temperature control can be found in [10].

However, it is unlikely that in practical applications an RTP system will be so free of disturbances that purely open-loop control schemes will work well. Feedback from temperature sensors must be used to ensure small temperature errors in the face of substantial disturbances and/or model errors.

Figure 2 is a block diagram of an RTP temperature control scheme that incorporates both open-loop transient control and feedback control. The lamp power signal vector sent to the lamp power supplies is the sum

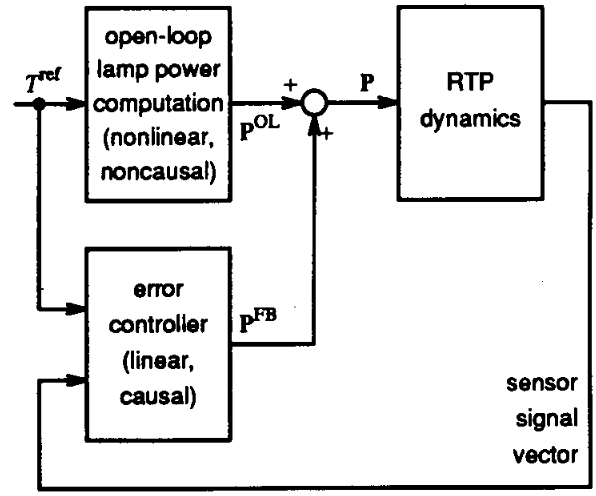

Figure 2: Feedback added to system for disturbance rejection.

of $\mathbf{P}^{\rm OL}$, the lamp power signal vector generated by the open-loop controller, and $\mathbf{P}^{\rm FB}$, the lamp power signal vector generated by the error controller. The intended effect of adding the error controller is to nearly eliminate the drift of the wafer temperature profile away from $T^{\rm ref}$ that would probably occur if open-loop control only was used.

Information about the wafer temperature profile increases with the number of sensors until there are enough sensors to infer the exact profile; because sensors may be expensive, determination of the relation between the number and locations of sensors and achievable system performance is worth the effort.

4 Limits of Performance for Sensor Configurations

In this section a method is presented for numerically determining the best static disturbance rejection achievable in an RTP system with any linear error controller, given a particular sensor configuration — by *sensor configuration* we mean a list of radial positions on the wafer at which temperature is being measured. Two main assumptions are made: the most significant disturbances are at such low frequencies that they can be taken as constant in time; the goal in disturbance rejection should be to minimize the worst error in temperature over all radial positions on the wafer.

4.1 Problem Formulation

Figure 3 is a block diagram showing the framework used here for error controller performance analysis. Discussion of this framework and explanation of its use along with Q-parametrization and convex optimization for determining limits of performance for linear control sys-

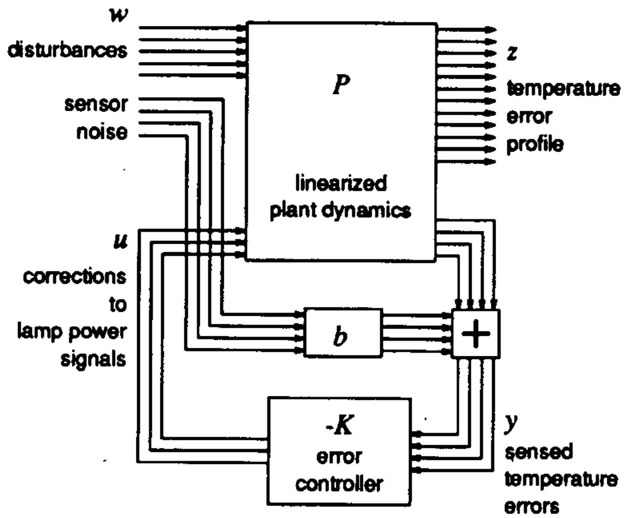

Figure 3: Model linearized about a nominal wafer temperature.

tems can be found in [12] or [13].

The system P is a linearization of the model (1) in Section 2 about a reference temperature T^{ref}. For convenience let $\mathbf{T}^{\text{ref}}$ denote the $I \times 1$ vector $T^{\text{ref}} \begin{bmatrix} 1 & \cdots & 1 \end{bmatrix}^T$. The state x of the system P is the vector of deviations from T^{ref}, that is, $x = \mathbf{T} - \mathbf{T}^{\text{ref}}$.

The exogenous input w is the concatenation $\begin{bmatrix} w_{\text{dist}} \\ w_{\text{sb}} \end{bmatrix}$ of a vector w_{dist} of disturbances and a vector w_{sb} of sensor biases, and the actuator input u is the vector of corrections made by the error controller to the vector of feedforward-generated lamp power commands. The state equation is

$$\dot{x} = C(\mathbf{T}^{\text{ref}})^{-1}\left(\frac{\partial \mathbf{f}}{\partial \mathbf{T}}(\mathbf{T}^{\text{ref}})x + Lu + Mw_{\text{dist}}\right), \quad (2)$$

where C, $\mathbf{f}$ and L are as in (1) and where each column of the matrix M is a different distribution of power in Watts to the wafer elements. The matrix M is selected to match anticipated system disturbances.

The regulated output z is identical to x. The sensed output y is a vector of measurements of temperature error at some radial positions. If the radial grid is fine enough then these measurements at radial positions might as well correspond to measurements of wafer element temperatures. Let C_{sens} be the matrix formed from the $I \times I$ identity matrix by selecting rows corresponding to elements whose temperatures are being measured. The measurement is assumed to be corrupted only by a vector of small sensor biases; the largest possible absolute value of a sensor bias in °C is the scalar b. So the output equations for the linearized system are

$$\begin{aligned} z &= x \\ y &= C_{\text{sens}}x + bw_{\text{sb}}. \end{aligned} \quad (3)$$

Let $P_{zw}(s)$, $P_{zu}(s)$, $P_{yw}(s)$, and $P_{yu}(s)$ be the open-loop transfer matrices between pairs of input and output vectors — these matrices can be determined from the above state and output equations — and let $H_{zw}(s)$ be the closed-loop transfer matrix from z to w. For our RTP model, all of the open-loop transfer matrices are stable, so the set of stable closed-loop transfer matrices achievable with a linear controller $K(s)$ can be written as

$$\{P_{zw}(s) + P_{zu}(s)Q(s)P_{yw}(s) \mid Q(s) \text{ is stable}\}; \quad (4)$$

the set of achievable DC values of H_{zw} is then

$$\{P_{zw}(0) + P_{zu}(0)Q(0)P_{yw}(0) \mid Q(0) \in \mathbf{R}^{n_u \times n_y}\}, \quad (5)$$

where n_u and n_y are the numbers of actuators and sensors.

If the columns of M in the model (2) are scaled appropriately, we may assume that each component of the vector w lies in the interval $[-1, 1]$. Then the quantity $\|H_{zw}(0)\|_{\infty i}$, the infinity-induced (max-row-sum) norm of $H_{zw}(0)$ is the largest possible absolute value of steady-state temperature error when w is constant. The parametrization (5) allows us to pose the problem of minimizing $\|H_{zw}(0)\|_{\infty i}$ as a finite-dimensional convex optimization problem whose variables are the entries of the matrix $Q(0)$ and which can be easily solved by computer.

If there is concern that the error controller might cause one or more actuators to saturate, the vector of regulated inputs can be expanded to include the actuator signals. A new convex optimization problem can be formulated and solved: that of minimizing worst-case temperature error subject to bounds on the worst-case values of the signals in u.

4.2 Example

The problem formulation outlined above was applied to a model of a fictitious RTP system described fully and referred to as System S in [8]. A schematic diagram of this system is shown in Figure 4. The lamp array consists of a flat gray ceiling into which three ideal blackbodies have been set; the temperatures of the blackbodies are assumed to be independently manipulated by the actuator signals. The temperature setpoint T^{ref} was chosen to be 1100°C and the gas environment was taken to be nitrogen at 1 atm at a low flow rate.

The vector w_{dist} was taken to be 6×1. The columns of the matrix M were selected as listed in Table 1. The first three columns are chosen to represent the largest possible expected differences between actual distributions of wafer heat loss through three different modes and the corresponding distributions predicted by the nonlinear RTP system model. For example, the third column of M is one-half of the model-predicted vector of convective heat fluxes from wafer elements when the wafer is at 1100°C. Columns four through six account for the possibility that the actual power radiated by a lamp zone may be as much as 5% smaller or larger than

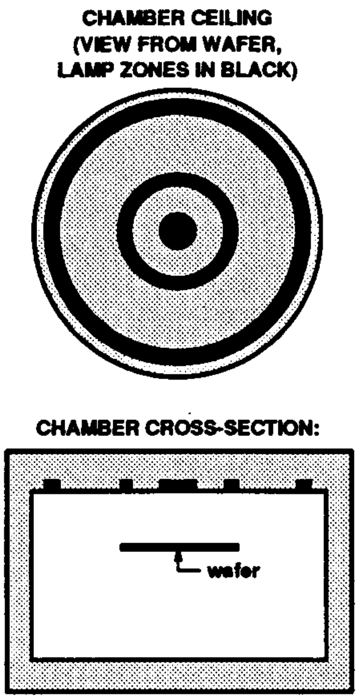

Figure 4: Idealized RTP system used for example. (Approximately to scale. Wafer diameter is 100 mm.)

its nominal value. The disturbance distributions were chosen to be close in size and shape to those expected in actual RTP systems. The value of b was chosen to be 0.1°C, to match the error in the best wafer temperature measurement scheme that can reasonably be hoped for.

Table 2 shows the minimal values of $\|H_{zw}(0)\|_{\infty i}$ achievable for selected sensor configurations with the models given above for the system and disturbances. Worst-case actuator levels were not constrained but those resulting from unconstrained optimization were observed to be acceptable. Inspection of the table reveals that performance varies considerably both with number of sensors and with sensor positions. The data in the table suggests the following conclusions, which are supported by exhaustive study:

- Optimal performance with the best two-sensor configuration is much better than that achievable with any single-sensor configuration, but only 50% worse than performance with best three-sensor arrangement.
- The best three-sensor configuration is essentially as good as any configuration with four or more sensors.
- The performance with multi-sensor configurations is very sensitive to the location of the sensor nearest the edge and much less sensitive to the locations of the other sensors.

Table 1: Columns of the disturbance power distribution matrix M for the example problem.

column index	distribution of power to wafer elements	scale factor
1	modeled radiative loss from wafer edge at 1100°C	0.2
2	modeled radiative loss from wafer bottom surface at 1100°C	0.1
3	modeled convective losses from wafer top, bottom, and edge surfaces at 1100°C	0.5
4	distribution of power from lamp 1 used for steady-state hold at 1100°C	0.05
5	distribution of power from lamp 2 used for steady-state hold at 1100°C	0.05
6	distribution of power from lamp 3 used for steady-state hold at 1100°C	0.05

Table 2: Results for example problem: Optimal DC disturbance rejection for various sensor configurations. Sensor positions are given as fractions of R, the wafer radius. Optimal 1-, 2-, and 3-sensor configurations are marked with asterisks.

sensor configuration	minimal $\|H_{zw}(0)\|_{\infty i}$
$0.10R$	14.1°C
* $0.84R$	7.0°C
$0.96R$	10.9°C
* $0.47R$, $0.96R$	1.5°C
$0.10R$, $0.96R$	2.0°C
$0.10R$, $0.84R$	4.0°C
$0.10R$, $0.53R$	9.9°C
* $0.28R$, $0.78R$, $0.96R$	0.98°C
$0.28R$, $0.78R$, $0.92R$	1.6°C
$0.10R$, $0.78R$, $0.96R$	1.02°C
$0.10R$, $0.47R$, $0.78R$	4.4°C
many 6-sensor configurations	always $\geq$ 0.96°C

4.3 Practical Applications

Suppose a control system designer is given an RTP system, a good model for likely disturbances, and a good estimate of the DC gain matrix from small changes in lamp settings to changes in the wafer temperature. The designer can use the technique outlined above to apply to a candidate sensor configuration the following test: With this sensor configuration, is there *any* linear error controller that can be designed to meet the DC disturbance rejection specifications? This test can be used to immediately rule out many candidates. If the test is failed by sensor configurations that accurately measure the entire wafer temperature profile, the designer can conclude that the actuator configuration is inadequate.

Once the designer has chosen a sensor configuration that is adequate in the sense that with it DC disturbance rejection specifications can be met, she or he must come up with an error controller design with acceptable static and dynamic performance. To measure the static performance obtained with the designer's controller, it can be compared with the optimal static performance for the chosen sensor configuration.

Determining the optimal $Q(0)$ could also be a step in a method of error controller synthesis. One possible method involves the modification of estimated-state-feedback controllers to achieve optimal static performance; another involves designing multivariable PID controllers with an equality contraint on $H_{zw}(0)$.

5 Summary

An introduction was given to the problem of wafer temperature control in rapid thermal processing (RTP) of semiconductor wafers. A method was presented to find the best possible static disturbance rejection performance achievable with any linear feedback controller using a given temperature sensor configuration. The method allows minimization of worst-case temperature error — a quantity of direct engineering importance — in response to disturbances described by a realistic model. An example RTP system was analyzed; results obtained with the example suggest that when feedback control of temperature error is implemented, overall system performance will depend critically on sensor configuration.

References

[1] F. Roozeboom and N. Parekh. Rapid thermal processing systems: A review with emphasis on temperature control. *J. Vac. Sci. Technol. B*, 8(6):1249–1259, November 1990.

[2] C. Hill, S. Jones, and D. Boys. Rapid thermal annealing — theory and practice. In R. A. Levy, editor, *Reduced Thermal Processing for ULSI*, pages 143–180. Plenum Press, 1988.

[3] M. M. Moslehi. Single-wafer optical processing of semiconductors: Thin insulator growth for integrated electronic device applications. *Appl. Phys. A*, 46:255–273, 1988.

[4] S. C. Wood, K. C. Saraswat, and J. M. Harrison. The economic impact of single wafer multiprocessors. In R. Singh and M. Moslehi, editors, *Rapid Thermal and Related Processing Techniques*, volume 1393 of *SPIE Proceedings*, pages 36–48. SPIE, 1991.

[5] H. A. Lord. Thermal and stress analysis of semiconductor wafers in a rapid thermal processing oven. *IEEE Trans. Semicond. Manufact.*, 1(3):105–114, August 1988.

[6] R. Kakoschke, E. Bußmann, and H. Föll. The appearance of spatially nonuniform temperature distributions during rapid thermal processing. *Appl. Phys. A*, 52(1):52–59, January 1991.

[7] S. A. Norman, C. D. Schaper, and S. P. Boyd. Improvement of temperature uniformity in rapid thermal processing systems using multivariable control. In *Mater. Res. Soc. Proc.*, volume 224, pages 177–183. Materials Research Society, 1991.

[8] S. A. Norman. Optimization of wafer temperature uniformity in rapid thermal processing systems. Technical Report 91–SAN–1, Information Systems Lab., Stanford Univ. Dept. of Elec. Eng., June 1991.

[9] S. A. Norman. Optimization of transient temperature uniformity in RTP systems. *IEEE Trans. Electron Dev.*, 39(1):205–207, January 1992.

[10] P. P. Apte and K. C. Saraswat. Rapid thermal processing uniformity using multivariable control of a circularly symmetric three zone lamp. *IEEE Trans. Semicond. Manufact.*, submitted for publication, 1991.

[11] P. P. Apte, S. Wood, K. C. Saraswat, and M. M. Moslehi. Temperature uniformity optimization using three-zone lamp and dynamic control in a rapid thermal multiprocessor. In *Mater. Res. Soc. Proc.*, volume 224, pages 209–214. Materials Research Society, 1991.

[12] S. Boyd, C. Barratt, and S. Norman. Linear controller design: Limits of performance via convex optimization. *Proc. IEEE*, 78(3):529–574, March 1990.

[13] S. Boyd and C. Barratt. *Linear Controller Design: Limits of Performance*. Prentice-Hall, 1991.