Optimization of Transient Temperature Uniformity in RTP Systems

Stephen A. Norman†

Information Systems Laboratory
Dept. of Electrical Engineering
Stanford University, Stanford CA 94305

PREPRINT

Accepted for publication in
IEEE Transactions on Electron Devices,
Special Issue on Rapid Isothermal Processing, January 1992.

Abstract

Independent, dynamic control of multiple lamps in a rapid thermal processing (RTP) system is a potential solution to the problem of transient wafer temperature nonuniformity. This brief presents a technique based on linear programming for minimization of worst-case error during temperature trajectory following, given a model of an axisymmetric, multi-lamp RTP system.

---

†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.

1 Introduction

In rapid thermal processing (RTP) of semiconductor wafers precise control of wafer temperature is required throughout the process cycle. To obtain uniform processing across the wafer surface and to prevent the creation of slip defects due to thermal stress, the temperature must be nearly uniform across the wafer at all times.

It is known that the distribution of energy from an RTP system lamp or lamp array must be nonuniform over the wafer to obtain uniform temperature; the reasons for this are radiative loss by the wafer edge and nonuniform convective cooling [1, 2, 3]. If the distribution of lamp energy used during transients is simply a scaled version of that which provides steady-state temperature uniformity, serious temperature nonuniformity will occur during transients [1, 2, 4]. One potential solution to this problem is independent dynamic control of individual lamps or zones of lamps within a lamp array [5, 6, 3]. This brief paper addresses the problem of minimizing temperature error during transients when independent control of lamp power settings is available.

2 RTP System Thermal Model

If the small temperature variation through the thickness of the wafer is neglected, the wafer temperature dynamics in an axisymmetric RTP system can be modeled using a finite difference approximation over a grid of radial positions; this is the approach used in [2] and much other research. Let $\mathbf{T}(t)$ denote the $I \times 1$ vector of temperatures at these radial positions at time t. A dynamic model relating $\mathbf{T}(t)$ to the $J \times 1$ vector of lamp power settings $\mathbf{P}(t)$ is:

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

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 the element and C_P is the specific heat of Si 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 Problem Formulation and Solution

3.1 Minimax error approximation

Two pieces of notation possibly unfamiliar to some readers will be used: first, $\mathbf{x} \leq \mathbf{y}$, where $\mathbf{x}$ and $\mathbf{y}$ are vectors, means that each component of $\mathbf{x}$ is less than the corresponding component of $\mathbf{y}$; and second, $\|\mathbf{x}\|_\infty$, the *max-norm* of the vector $\mathbf{x}$, is the largest of the absolute values of $\mathbf{x}$.

Consider the system of linear equations represented by

$$A\mathbf{x} = \mathbf{b}, \tag{2}$$

where A is an $m \times n$ matrix, $\mathbf{x}$ is an $n \times 1$ column vector, $\mathbf{b}$ is an $m \times 1$ column vector, and $m > n$. The system is overdetermined: In general, for no value of $\mathbf{x}$ is (2) exactly satisfied. The most familiar approximate solution to (2) is the least-squares solution. An alternate approximate solution to (2) is what will be called in this paper the *minimax-error* solution. The minimax-error solution $\hat{\mathbf{x}}_{\mathrm{MME}}$ is the value of $\mathbf{x}$ for which $\|A\mathbf{x} - \mathbf{b}\|_\infty$ is minimal. A relative of the minimax-error problem is what will be called here the *constrained minimax-error* problem:

$$\text{Find } \mathbf{x} \text{ to minimize } \|A\mathbf{x} - \mathbf{b}\|_\infty \text{ subject to } \mathbf{l} \le \mathbf{x} \le \mathbf{u}; \tag{3}$$

here there are lower and upper bounds placed on each of the entries of $\mathbf{x}$. Both the minimax-error and constrained-minimax-error problems can be formulated as linear programming problems (LP's) and solved by computer.

A reference for this material is [7, Ch. 14].

3.2 Minimax error in trajectory following

Consider the problem of designing $\mathbf{P}(t)$ so that the entire wafer closely tracks a reference trajectory $T^{\mathrm{ref}}(t)$ over the time interval $t_o \le t \le t_f$. For notational convenience, let $\mathbf{T}^{\mathrm{ref}}(t)$ be an $I \times 1$ vector all of whose entries are $T^{\mathrm{ref}}(t)$. A suitable goal in designing $\mathbf{P}(t)$ over $t_o \le t \le t_f$ is to find the one that minimizes the worst error from T^{ref} at any time between t_o and t_f and at any position on the wafer, that is to minimize

$$\max_{t_o \le t \le t_f,\ i=1,\ldots,I} \left| T_i(t) - T^{\mathrm{ref}}(t) \right| = \max_{t_o \le t \le t_f} \left\| \mathbf{T}(t) - \mathbf{T}^{\mathrm{ref}}(t) \right\|_\infty . \tag{4}$$

Assume that the RTP temperature control system is discrete-time with sample period Δt such that $t_f = t_o + N\Delta t$ where N is an integer and that $\mathbf{P}(t)$ is constrained to be constant between sample instants. (One of many references on sampled-data control is [8].) Let $\mathbf{P}_n$ be the value of $\mathbf{P}(t)$ for t between $t_o + n\Delta t$ and $t_o + (n+1)\Delta t$, let $\mathbf{T}_n = \mathbf{T}(t_o + n\Delta t)$, and let $\mathbf{T}_n^{\mathrm{ref}} = \mathbf{T}^{\mathrm{ref}}(t_o + n\Delta t)$. Then if Δt is small enough ($\Delta t = 0.1\,\mathrm{s}$ is easily adequate) the problem of finding $\mathbf{P}(t)$ to minimize (4) is accurately approximated by the following:

$$\begin{gathered} \text{Given } \mathbf{T}_0, \text{ find } \mathbf{P}_0, \mathbf{P}_1, \ldots, \mathbf{P}_{N-1} \text{ to minimize} \\ \max_{n=1,2,\ldots,N} \left\| \mathbf{T}_n - \mathbf{T}_n^{\mathrm{ref}} \right\|_\infty \\ \text{subject to, for } n = 0, 1, \ldots, N-1, \\ \mathbf{P}^{\min} \le \mathbf{P}_n \le \mathbf{P}^{\max}. \end{gathered} \tag{5}$$

Define $\mathbf{T}_n^{\mathrm{zero}}(t)$ to be the value at time t of the solution to (1) with initial condition $\mathbf{T}(t_o + n\Delta t) = \mathbf{T}_n^{\mathrm{ref}}$ and with $\mathbf{P}(t)$ equal to zero for t between $t_o + n\Delta t$ and $t_o + (n+1)\Delta t$. If $\mathbf{T}(t_o + n\Delta t)$ is close to $\mathbf{T}_n^{\mathrm{ref}}$ then (1) can be linearized as follows:

$$\frac{d}{dt}(\mathbf{T} - \mathbf{T}_n^{\mathrm{zero}}) \approx C(\mathbf{T}_n^{\mathrm{ref}})^{-1} \left(F_n \cdot (\mathbf{T} - \mathbf{T}_n^{\mathrm{zero}}) + L\mathbf{P}_n \right), \tag{6}$$

where the matrix F_n is the derivative at $\mathbf{T}_n^{\mathrm{ref}}$ of $\mathbf{f}$ with respect to $\mathbf{T}$. Define the error vector $\mathbf{E}_n$ at sample time n as $\mathbf{E}_n = \mathbf{T}_n - \mathbf{T}_n^{\mathrm{ref}}$ and the vector $\mathbf{T}_n^{\mathrm{cool}}$ as $\mathbf{T}_n^{\mathrm{cool}} = \mathbf{T}_n^{\mathrm{zero}}(t_o + (n+1)\Delta t) - \mathbf{T}_n^{\mathrm{ref}}$. Then solution of (6) yields

$$\mathbf{E}_{n+1} = \Phi_n \mathbf{E}_n + \Gamma_n \mathbf{P}_n + \mathbf{T}_n^{\mathrm{cool}} + \mathbf{T}_n^{\mathrm{ref}} - \mathbf{T}_{n+1}^{\mathrm{ref}}, \tag{7}$$

where $\Phi_n = \exp\left(C(\mathbf{T}_n^{\mathrm{ref}})^{-1}F_n\Delta t\right)$ and $\Gamma_n = \int_0^{\Delta t}\exp\left(C(\mathbf{T}_n^{\mathrm{ref}})^{-1}F_n\tau\right)L\,d\tau$, with $\exp(\cdot)$ being the matrix exponential function. Expansion of (7) for $n = 0,\ldots,N-1$ gives an expression of the form

$$\begin{bmatrix}\mathbf{E}_1\\ \vdots\\ \mathbf{E}_N\end{bmatrix} = A\begin{bmatrix}\mathbf{P}_0\\ \vdots\\ \mathbf{P}_{N-1}\end{bmatrix} + \mathbf{b}, \tag{8}$$

where the matrix A and the vector $\mathbf{b}$ are determined by the vector $\mathbf{E}_0$ and the sequences T_n^{ref}, Φ_n, Γ_n, and $\mathbf{T}_n^{\mathrm{cool}}$. Minimizing the max-norm of the right-hand side of (8) subject to the constraints $\mathbf{P}_{\mathrm{min}} \leq \mathbf{P}_n \leq \mathbf{P}_{\mathrm{max}}$ for $n = 0,\ldots,N-1$ is a constrained minimax-error problem; if the error made in linearization is small then it is equivalent to (5).

4 Example Application

Figure 1 shows a hypothetical axisymmetric RTP system with independently controlled lamp zones. The wafer diameter is four inches and the diagram is to scale. The wafer and chamber surfaces are modeled as gray and diffuse, so radiative heat transfer can be computed using angle factors [9]. The wafer emissivity is taken to be 0.6. The chamber wall surfaces are assumed to be at 27°C and to have an emissivity of 0.5, except for the black lamp zones, the temperatures of which are directly manipulable. This hypothetical system is not a practical design, but its behavior is qualitatively similar to that of some real systems; it is used here to illustrate the application of the optimization technique outlined above.

The first example trajectory considered here is a ramp-down of wafer temperature from 1100°C to 800°C at 40 °C/s, with a sample rate of 10 Hz. The upper plot in Figure 2 shows the lamp power trajectory designed by minimizing (5), and the lower plot shows the temperature error plotted against time and radial position using this lamp power trajectory. The minimal worst-case nonuniformity predicted by the LP solution was 3.0°C and that obtained from nonlinear simulation was 3.1°C, indicating that the error in the LP formulation due to linearization of the model is indeed small. Two points about the solution are worth making here: first, that independent control of three lamp settings has clearly been exploited, and second, that explicit specification of the constraint that lamp powers be positive allows the design of a lamp power trajectory that is "pushed against" this constraint.

A second example trajectory studied was a ramp-up from 600°C to 1100°C at 100 °C/s; minimal worst-case nonuniformity over this trajectory was found to be 3.2°C. It was determined using an LP formulation related to the one given here that worst-case steady-state temperature error across the wafer for setpoints between 600°C and 1100°C can be made less than 0.5°C for this system. The ramp-down, ramp-up, and steady-state results together suggest that multi-lamp RTP systems can be designed to deliver excellent temperature uniformity both during transients and during steady-state hold.

5 Discussion

For proof that dynamic independent control of lamp zones can solve the problem of transient temperature nonuniformity in RTP, experimental evidence is required. But analysis and careful simulation has demonstrated that this approach has great potential.

Optimization techniques such as that outlined in this paper are potentially valuable not only in control system design of given RTP systems but also in design of new RTP systems. If a reasonably accurate thermal model can be developed for a system before it is built, it can be determined whether the system can provide adequate uniformity for specified process trajectories. Development of RTP systems incorporating such analysis of candidate designs would be more efficient than a purely trial-and-error approach.

A detailed discussion of the material in this paper and related ideas can be found in [10].

References

[1] 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.

[2] 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.

[3] S. A. Campbell, K.-H. Ahn, K. L. Knutson, B. Y. H. Liu, and J. D. Leighton. Steady-state thermal uniformity and gas flow patterns in a rapid thermal processing chamber. *IEEE Trans. Semicond. Manufact.*, 4(1):14–20, February 1991.

[4] 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.

[5] 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. Materials Research Society, 1991.

[6] 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. Materials Research Society, 1991.

[7] V. Chvátal. *Linear Programming.* W. H. Freeman, 1983.

[8] G. F. Franklin, J. D. Powell, and M. L. Workman. *Digital Control of Dynamic Systems.* Addison Wesley, 2nd edition, 1990.

[9] E. M. Sparrow and R. D. Cess. *Radiation Heat Transfer.* Hemisphere Publishing, 1978.

[10] 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.

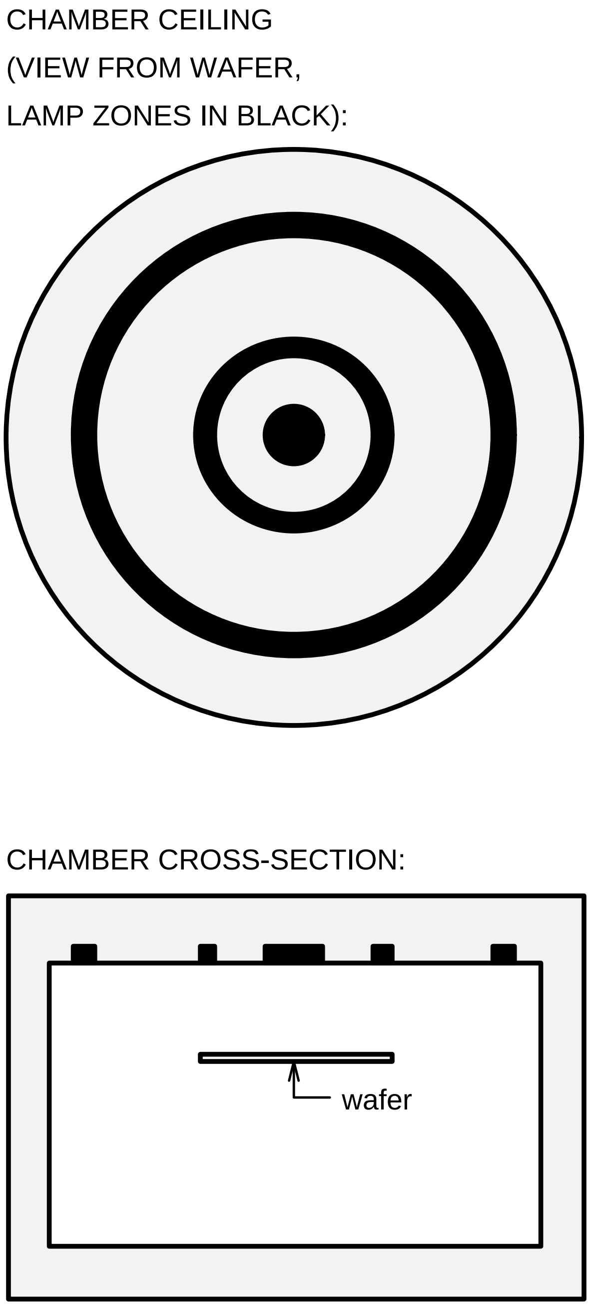
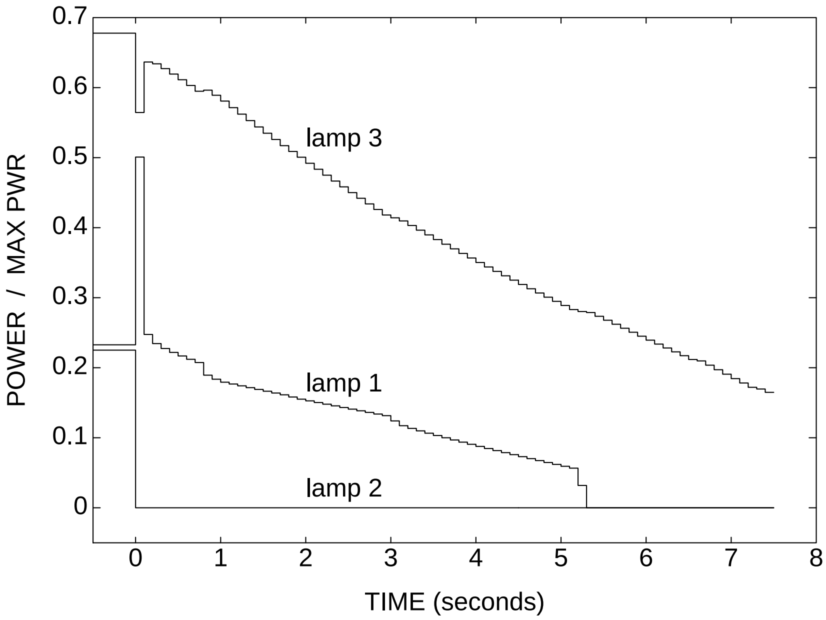


Figure 1: Schematic for hypothetical axisymmetric RTP system.



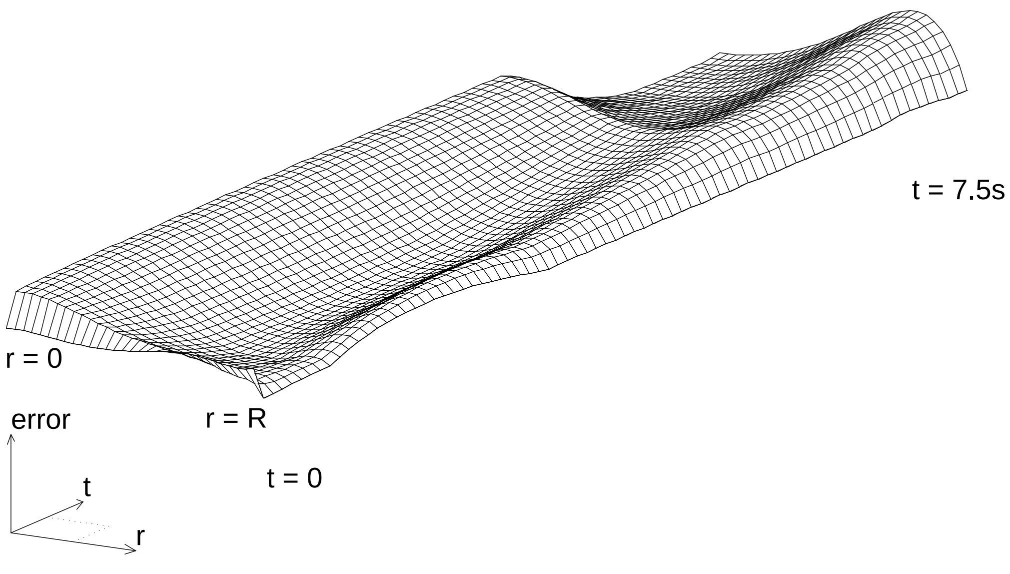


Figure 2: Optimal cool-down from 1100°C to 800°C at 40 °C/s. *Above*: Lamp settings — lamp 1 is innermost and lamp 3 is outermost. *Below*: Mesh plot of temperature error over time and radial position — worst-case nonuniformity is 3.2°C.