Model-Based Control of Fast-Ramp RTP Systems*

D. DE ROOVER, A. EMAMI-NAEINI, J. L. EBERT, S. GHOSAL, G. W. VAN DER LINDEN

SC Solutions Inc.
3211 Scott Boulevard
Santa Clara, CA 95054
USA
roover@scsolutions.com

Abstract

This paper analyses control of RTP systems for ramps with high ramp rates such as 250° C per second. The limitations on bandwidth and accuracy for a feedback control system are considered. It is shown that feedforward control can be used in conjunction with feedback control to improve the temperature uniformity of RTP systems for very fast ramp rates.

1 Introduction

RTP is now becoming part of the mainstream semiconductor manufacturing technology after years of being a niche technology. A typical fabrication process may consist of 26 different steps of RTP oxidation, annealing (RTA), nitridation and CVD. Submicron critical dimensions place stringent demands on thermal processing of wafers. In RTA, both temperature ramp up and ramp down rates are important. To minimize diffusion lengths, the amount of time that the wafer is at the processing temperature must be minimized and high temperature ramp rates are desirable. For example, an RTA ramp rate of 400° C/sec resulted in minimum junction depth [1].

However, from a control point of view, fast ramping is a completely new challenge with problems that have not been encountered before: dynamic behavior — which did not play a role in tracking ramps with low ramp rates — can seriously limit the achievable performance when tracking ramps with high ramp rates. Important issues you have to deal with are: overshoot, actuator saturation, wafer temperature non-uniformity, and robustness against high-frequency modeling errors and disturbances. This paper will address these issues in order to explore the limitations of tracking fast ramps.

In our earlier work [4,6], model-based feedback controllers were developed for a generic RTP chamber without considering sensor disturbances due to wafer rotation, which can be of considerable magnitude. This paper will address the design of new feedback controllers that take these disturbances into account.

The outline of the paper is as follows. The plant properties and the physical model are described in Section 2. The control system specifications are

*This work is supported by the Advanced Defense Projects Agency (DARPA) under contract No. N00014-98-C-0201.

contained in Section 3. In Section 4, the physical limits of the system are established through an open-loop analysis of the system. The control problem is then analyzed, and a suitable controller structure introduced in Section 5. New feedback controller designs and the limits of feedback are discussed in Section 6. The advantages of adding feedforward control are discussed in Section 7 and the concluding remarks appear in Section 8.

2 Plant Properties and Physical Model

The controller design in this paper is based on a previously derived physical model of the generic RTP system [2-8]. The generic RTP system geometry is shown in Figure 1, which is representative of commercial RTP systems. The system consists of five independently powered lamps near

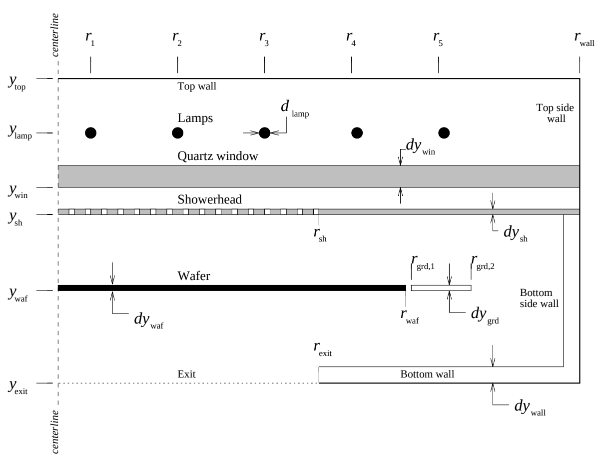

Figure 1: Schematic of the Generic RTP Chamber.

the top wall that form axi-symmetric rings at radii $r_1, \ldots, r_5$. The walls of the chamber are highly reflective (95%) and water-cooled. A thick quartz window (6.35 mm) and a thinner quartz showerhead (1mm) transmit radiation from the hot lamps at wavelengths shorter than approximately 4μm, but are opaque to radiation at longer wavelength. The silicon wafer and guard ring are heated by this short wavelength lamp radiation. A physical model of this nonlinear system was constructed that predicts the dynamic temperature response. Details of this model are described in Ebert et al. [2].

Due to the significance of thermal radiation as a mechanism for transfer of energy in RTP systems and the large temperature range of operation, the system's behavior is highly nonlinear. In addition, these systems are multivariable having multiple temperature sensors (outputs) and multiple lamp groups for actuation (inputs). The generic RTP system has five inputs and five outputs. Since the individual channels are strongly coupled, it is required that the controller handles this appropriately, and therefore must have a truly multivariable structure rather than controlling single loops.

An important property of rapid thermal processes is the repetition of standard process runs, with a fixed reference temperature trajectory. This means that the controller can be fine-tuned towards these references, rather than trying to achieve good performance for a large class of references. The temperature profiles are typically piecewise linear.

The dynamic range of an RTP system is fairly large due to the large variation in thermal mass of the various components. The lamp filaments are the smallest elements, and hence the fastest. On the other hand, a quartz window has a relatively large thermal mass, and hence is relatively slow. It is very difficult to predict the open-loop response of an RTP chamber within a required accuracy of about 1° C, which is one of the main reasons for using closed-loop control.

All the feedback controllers presented here are linear, and rely on a linear model that is derived from the simplified nonlinear physical model of the generic RTP chamber [2], denoted by:

$$\left\{ \begin{array}{rcl} \dot{x} & = & \underbrace{A_r x^4 + A_c x + C_1}_{f(x)} + B_1 u \\ y & = & h(x) \end{array} \right.$$

By selecting a suitable linearization (operating) point (x_o, u_o) a linear model can be found by computing:

$$\begin{aligned} A &= \left.\frac{\partial f}{\partial x}\right|_{x_o} \\ B &= B_1 \\ C &= \left.\frac{\partial h}{\partial x}\right|_{x_o} \end{aligned}$$

resulting in the linear model:

$$\begin{cases} \dot{\tilde{x}} = A\tilde{x} + B\tilde{u} \\ \tilde{y} = C\tilde{x} \end{cases} \tag{1}$$

For notational simplicity the tildes will be dropped in the sequel. An important issue in the linearization is the selection of the linearization point (temperature).

3 Control Problem Formulation

To be able to design temperature controllers that achieve the desired wafer quality, it is important to consider the performance specifications in terms of temperature control quality. The temperature control problem in an RTP system typically has the following demands to ensure good and uniform wafer properties:

1. Good steady-state tracking, better than 1° C, preferably zero error;

2. Good wafer uniformity during ramp, with little (only a few degrees Celsius) or no overshoot for temperature changes up to 600° C, varying ramp rates (50° C/sec to 250° C/sec), and setpoints up to 1100° C;

3. Insensitivity to sensor noise, process disturbance and variations, such as wafer-to-wafer variations (*e.g.*, variation in wafer emissivity), changes in temperature setpoints, etc.

These demands pose a serious challenge for the controller design. In this paper the focus is on item 2 above. We wish to explore the maximum achievable ramp rates while attempting to meet the other performance specifications.

4 Physical Limits to Fast-Ramping

There are physical limits to the rate at which the wafer can be heated based on chamber geometry, physical properties of the components, and available lamp power. Consider the temperature rise of the wafer induced by open-loop heating with all lamp powers set to their maximum levels instantaneously (*i.e.*, step response). The results are shown in Figure 2 for maximum lamp power of 65 kW. The plots show the temperatures of twenty-one points on the wafer from the center to the outer edge that are equally spaced along the radius. The trajectories at higher temperatures are shown as dashed lines because silicon melts at 1410° C, and temperatures much higher than 1100° C are irrelevant. The lamps heat up in the first half-second, when the wafer temperature increase is small. The wafer temperature then rises to the process temperature of 1100° C at ramp rates in the range 200–325° C (see Figure 3). The simulation shows that the maximum achievable ramp rate for this chamber is ≈300° C/sec. As expected, the open-loop temperature gradients are large, and the maximum temperature difference between any two points on the wafer was found to reach 65° C at about the time the process temperature is reached.

5 Controller Structure and LQG Control Design

In our previous work [6], we proposed a suitable controller structure as shown in Figure 4. We now briefly review this structure. The *feedforward* controller takes advantage of the known reference temperatures to compute a suitable control signal that

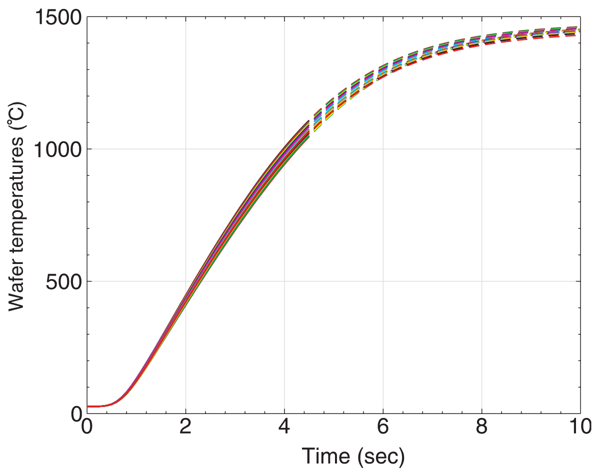

Figure 2: Wafer temperature profile during heat-up for lamp power of 65 kW. Each line corresponds to a point on the wafer. The solid lines show trajectories up to process temperature.

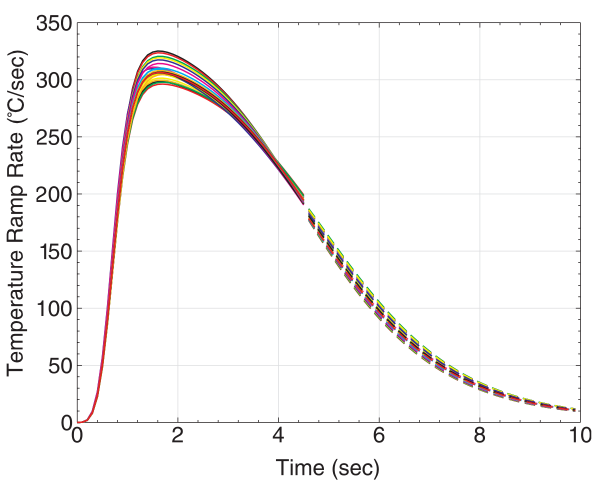

Figure 3: Temperature ramp rates for twenty-one points on the wafer. The solid lines show rates up to process temperature of 1100° C.

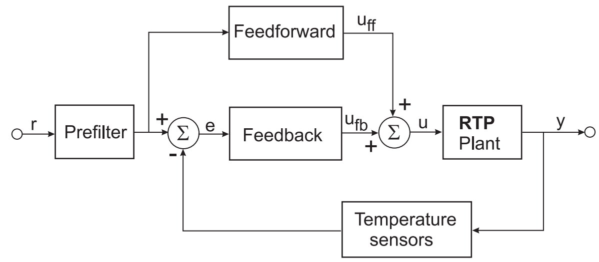

Figure 4: Controller structure.

is injected in the closed-loop. Due to the straightforward structure of a feedforward controller, it can be nonlinear and can be based directly on the nonlinear RTP model. An important practical consideration is whether the reference is known *a priori* to the feedforward controller, or if the reference is provided in real-time. The latter case is the most common in practice, but the first option allows a global optimization of the trajectory rather than point-to-point optimal commands. It is assumed here that the reference will be provided in real-time.

The *feedback* controller is based on a linear design, as dynamic output feedback is required, which is not adequately addressed by current nonlinear controller design methods. Its task is to address any mismatch that arises from the limited fidelity of the feedforward controller, and to deal with the process disturbances.

The *prefilter* smoothes the temperature reference, which is piecewise linear and thereby has discontinuities in the rate of change. If the "raw" reference is tracked closely by the controller, it will inevitably result in overshoot, because finite lamp dynamics introduce delays between the feedback signal and the actuator (i.e., the system is at least second order). In addition, this reduces excessive control action due to the sudden changes in rate.

In the following section the focus will be on the feedback part of the RTP controller. Results on the feedforward design will be considered later in Section 7.

For feedback control design we use Linear Quadratic Gaussian (LQG) control extended with Loop Transfer Recovery (LTR). LQG-LTR is a standard controller design method that has been successfully applied to Multiple-Input Multiple-Output (MIMO) RTP control problems, (see [6]).

The LQG controller structure for the generic RTP system has been explained in [6]. The basis for LQG control is a linear model of the RTP system as represented by Equation 1. To be able to enforce zero steady-state tracking error, this model is augmented with integrators on the plant output, such that the resulting model becomes:

$$\begin{cases} \dot{z} = \begin{bmatrix} 0 & C \\ 0 & A \end{bmatrix} z + \begin{bmatrix} 0 \\ B \end{bmatrix} \tilde{u} \\ y = \begin{bmatrix} I & 0 \end{bmatrix} z \end{cases}$$

with $z = [\xi^\top \; x^\top]^\top$. The design of an LQG controller is separated into the design of the state feedback gain K, and the design of the estimator gain L. The state feedback gain is found by minimizing the quadratic cost function:

$$J_K = 1/2 \int_0^\infty \left(z^\top Q z + u^\top R u \right) dt$$

where the symmetric positive-(semi-) definite matrices Q and R are the key design parameters. R is used to penalize the control effort, whereas Q is used to penalize tracking error. The resulting gain K can be partitioned according to:

$$K = [K_I \; K_P]$$

where K_I is the integral gain associated with the integral states ξ, and K_P with the plant state x. The design of the estimator gain is similar, which is found by minimizing:

$$J_L = 1/2 \int_0^\infty \left(z^\top R_w z + y^\top R_v y \right) dt$$

where the symmetric positive-definite matrices R_w and R_v are the design parameters. Typically, R_w and R_v are used to characterize the statistical properties of Gaussian noise at state z and output y. We used these parameters for LTR purposes, *i.e.*, we tuned these parameters such that state feedback performance was recovered at the plant input.

By including the integral states in the controller, the control law now has the following structure

$$\begin{aligned} \dot{\gamma} &= e \\ \dot{\hat{x}} &= A\hat{x} + Bu + L(e - \hat{e}) \\ \hat{e} &= C\hat{x} \\ u &= -K_I \gamma - K_P \hat{x} \end{aligned}$$

where $e = r - y$, with r the reference, such that the resulting controller state-space realization becomes:

$$\begin{cases} \dot{q} = \begin{bmatrix} 0 & 0 \\ -BK_I & A - LC - BK_P \end{bmatrix} q + \begin{bmatrix} I \\ L \end{bmatrix} e \\ u = -\begin{bmatrix} K_I & K_P \end{bmatrix} q \end{cases}$$

with $q = [\gamma^\top \; \hat{x}^\top]^\top$. Note that the resulting controller is totally based on the error e, rather than using y for the estimator.

6 Performance Limits Using Feedback Only

The goal of the controller design in this paper is to track ramps with high ramp rates. Currently, typical ramp rates in RTP systems range from 25° C/sec to 75° C/sec. In a previous paper we showed good tracking control for a 50° C/sec ramp rate, see [6]. Four different model-based feedback controllers were designed (LQG, LQG with Implicit Model-Following, H_∞ and a μ-controller), all showing good ramp tracking, fast settling, and less than 1° Celsius overshoot. The resulting performance of the LQG controller as tested on the full nonlinear generic RTP model is presented in Figure 5. This figure shows the tracking response for a 50° C/sec ramp rate using feedback only; no prefilter or feedforward were used. The tracking performance is good: the tracking error is small, the overshoot is limited to 1° C, and hence settling is achieved as soon as the response enters the band of ±1° C around the final temperature. Also, wafer temperature non-uniformity is limited to approximately 3° C during ramp-up, which is acceptable. The largest deviation in temperature uniformity occurs at the wafer edge, which is difficult to control because of the difference in thermal mass between wafer and edge. The peak at 10 sec and the drop at 20 sec are both due to the sudden change in the reference ramp.

From a tracking point of view this controller is good. However, it is likely to be sensitive to sensor noise, disturbances, and/or model uncertainty at high frequencies. To investigate this shortcoming, consider a representative 2 Hz frequency periodic measurement disturbance induced by a 120 rpm wafer rotation. We modeled this disturbance as a sinusoidal signal of frequency 2 Hz with random phase, and amplitude linearly increasing from 1° C at the center temperature measurement to 5° C at the edge measurement. Figure 6 shows the tracking response using the same controller used in Figure 5. Its performance is now very poor. The superb tracking performance is overshadowed by the effect of the periodic disturbance: larger overshoot, no settling, and increased temperature non-uniformity. The reason for this is the high sensitivity of the power input to the disturbance at frequency 2 Hz, which is displayed in Figure 6(d); the high controller gain at high frequencies, which provided good tracking, also amplified the measurement disturbance. Note that we displayed the *actual* wafer temperature at a measured node instead of the *measured* wafer temperature in order to see how the disturbance affects the actual wafer temperature, and in order to make the tracking properties more visible.

To decrease the controller sensitivity to measurement disturbances, a frequency shaping filter was added for LQG control design to improve high frequency rolloff, especially at 2 Hz. Figure 7 shows the same simulation as in Figure 6, but now with the improved controller. Clearly, this controller is much more insensitive to the 2 Hz measurement disturbance. The good tracking properties shown in Figure 5 are partly recovered. However, the overshoot has increased to approximately 2° Celsius, and consequently settling has increased. Also, wafer temperature non-uniformity has become slightly worse. However, for a 50° C/sec ramp rate, this performance is acceptable according to the requirements in Section 3.

Figures 8 and 9 show the tracking response with the same controller for ramp rates 150° C/sec and 250° C/sec, respectively. These figures show the limits of fast-ramp tracking performance using feedback only: excessive overshoot, increased temperature non-uniformity and no improvement in settling result from increasing the ramp rate. Also, lamp saturation occurs for ramp rates of 250° C/sec and higher, which results in degradation of performance. This poor performance seems to be a fundamental limit of feedback control. To track a fast ramp, the controller should have high gain at high frequencies, which is in conflict with the requirement of being insensitive to measurement disturbances.

7 Performance Improvement Using Feedforward Control

It is well-known that it is very difficult to independently achieve both good tracking, disturbance suppression, robustness to unmodeled dynamics, and stabilization with a single-degree-of-freedom (feedback) controller [9,10]. By adding a *feedforward* controller, as shown in Figure 4, advantage is taken of the *a priori* known reference temperatures by computing a suitable control signal that is injected in the closed-loop. Since we wish to move the system from one operating point to another along a specified trajectory, we can approximately determine the input that is required for this movement. Consequently, we can apply this input directly to the system instead of letting the feedback controller compute it based on the tracking error e.

It is obvious that a fairly accurate system model is required for computing the feedforward input signal. As a matter of fact, the closer the (dynamic) model is to the actual system behavior, the more accurate the computed input will be, leading to a smaller tracking error. Typically, dynamic models are quite accurate at low frequencies and uncertain at high frequencies because of the low system gain at high frequencies, and because of the limited bandwidth of the actuators and sensors, *i.e.* the system does not provide much information at high frequencies, and the provided information is corrupted by measurements. The advantage of feedforward is the fact that the frequency range for which we have high-fidelity in the modeled dynamics, is usually larger than the closed-loop bandwidth attained with feedback control only. The disadvantage of feedforward control is the fact that its performance is worse compared to feedback control only, for those frequencies where model uncertainty is too large. This is the reason why feedforward control needs to be cut off after a certain frequency. This can be done effectively using a *prefilter* as shown in Figure 4. Theoretically, the prefilter and the feedforward filter can be merged into one filter, but for implementation reasons it makes sense to leave them as two separate filters. This has to do with the signals which coming out of the filters, and the place where these signals enter the feedback loop, see [10].

In the control structure of Figure 4, the feedforward filter should approximate the *inverse dynamics* of the RTP plant. We computed this inverse for the high-order 116-state linear plant model and used that as feedforward; inverting a low-order approximate plant model resulted in unstable feedforward filters because of the presence of non-minimum phase transmission zeros in the low-order approximation. The high-order inverse model can be successfully reduced, if necessary, but we did not attempt to do this for our simulations.

Figures 10 to 12 show the simulation results for tracking ramps with rates 50° C/sec, 150° C/sec and 250° C/sec, respectively, for a controller including feedback, feedforward and prefilter. The prefilter consisted of a second-order lowpass filter for each measured output channel. These figures show the merits of using prefilter and feedforward. The prefilter suppresses the overshoot shown in Figures 7 to 9, but also delays the response, whereas the feedforward speeds up the response. By exploiting the full freedom in the controller design, we are now able to achieve tight tracking, fast settling, very little overshoot, and robustness against high frequency model errors and measurement disturbances. Note that the increase in ramp rate results in a corresponding faster settling.

It is interesting to investigate the effect of the prefilter on the *actually achieved* ramp rates. Figure 13(a) and (b) show the actually achieved ramp rates computed from differentiating the measured outputs, without and with prefilter, respectively. The effect of the prefilter is clearly seen in Figure 13(b): instead of an instantaneous jump from zero to maximum ramp rate, the prefilter smoothes the discontinuous reference profile. As a consequence, the filtered reference trajectory never reaches the commanded maximum ramp rate, because the unfiltered reference ramp rate drops to zero before the actual ramp rate reaches the commanded ramp rate. As the ramp rate increases, this effect plays a more dominant role, as can be seen from Figure 13(b). Although the maximum commanded ramp rate is never reached, the positive effect of the prefilter is evident from Figures 10 to 12: the settling time is fast, and there is very little overshoot. As a matter of fact, the prefilter is a means to trade-off overshoot *vs* speed of response. The faster the response, the larger the overshoot, or stated differently: for a specific allowed amount of overshoot, the prefilter can be tuned to maximize speed of response.

8 Conclusions

In this paper, we have addressed the performance limitations of RTP systems in tracking high ramp rates. Model-based LQG feedback controllers were derived for temperature tracking and (periodic) disturbance rejection. It was shown that feedback control has a fundamental limitation in tracking fast ramps: to track a fast ramp, the controller should have high gain at high frequencies, which is in conflict with the requirement of being insensitive to high-frequency measurement disturbances and modeling errors.

It was shown that a combination of feedforward and feedback control design allows tracking of fast ramps with fast speed of response and good temperature uniformity, while providing robustness against unmodeled dynamics and disturbances. To successfully implement feedforward control, use was made of a prefilter that cuts off the high-frequency content of the reference ramp trajectory. In fact, the prefilter is a means of trading off overshoot *vs* speed of response. For a specified overshoot, the prefilter can be tuned to maximize speed of response.

References

[1] S. Shishiguchi, A. Mineji, T. Hayashi, and S. Saito, “Boron Implanted Shallow Junction Formation by High-Temperature /Short-Time/High-Ramping-Rate (400°C/sec) RTA,” *Symposium on VLSI Technology Digest of Technical Papers*, pp. 89-90, 1997.

[2] J. L. Ebert, A. Emami-Naeini, R. L. Kosut, “Thermal Modeling of Rapid Thermal Processing Systems,” *3rd International Rapid Thermal Processing Conference*, R. B. Fair, B. Lojek (eds), pp. 343-355, Amsterdam, The Netherlands, 1995.

[3] H. Aling, J. Abedor, J. L. Ebert, A. Emami-Naeini, R. L. Kosut, “Application of feedback linearization to models of rapid thermal processing (RTP) reactors,” *3rd International Rapid Thermal Processing Conference*, R. B. Fair, B. Lojek (eds), pp. 356-366, Amsterdam, The Netherlands, 1995.

[4] J. L. Ebert, A. Emami-Naeini, R. L. Kosut, “Thermal Modeling and Control of Rapid Thermal Processing Systems,” in *34th IEEE Conf. Decision and Control*, December 1995.

[5] M. Refai, J. L. Ebert, A. Emami-Naeini, and R. L. Kosut, “RTP Robust Control Design, Part I: Sensitivity Analysis,” *4th International Rapid Thermal Processing Conference*, Boise, Idaho,

September 1996.

[6] G. W. van der Linden, J. L. Ebert, A. Emami-Naeini, and R. L. Kosut, “RTP Robust Control Design : Part II Controller Synthesis,” *4th International Rapid Thermal Processing Conference*, Boise, Idaho, September 1996.

[7] H. Aling, R. L. Kosut, A. Emami-Naeini, J. L. Ebert, “Nonlinear Model Reduction with Application to Rapid Thermal Processing,” *in Proc. 35th IEEE Conf. Decision and Control*, pp. 4305-4310, Kobe, Japan, December 1996.

[8] H. Aling, S. Banerjee, A. K. Bangia, V. Cole, J. Ebert, A. Emami-Naeini, K. F. Jensen, I. G. Kevrekidis, S. Shvartsman, “Nonlinear Model Reduction for Simulation and Control of Rapid Thermal Processing Systems,” *in Proc. American Control Conference*, pp. 2233-2238, June 1997.

[9] Horowitz, I., M., *Synthesis of Feedback Systems*, Academic Press, 1963.

[10] Roover, D. de, “Motion Control of a Wafer Stage - A Design Approach for Speeding Up IC Production,” Ph.D. Thesis, Delft University of Technology, The Netherlands, 1997.

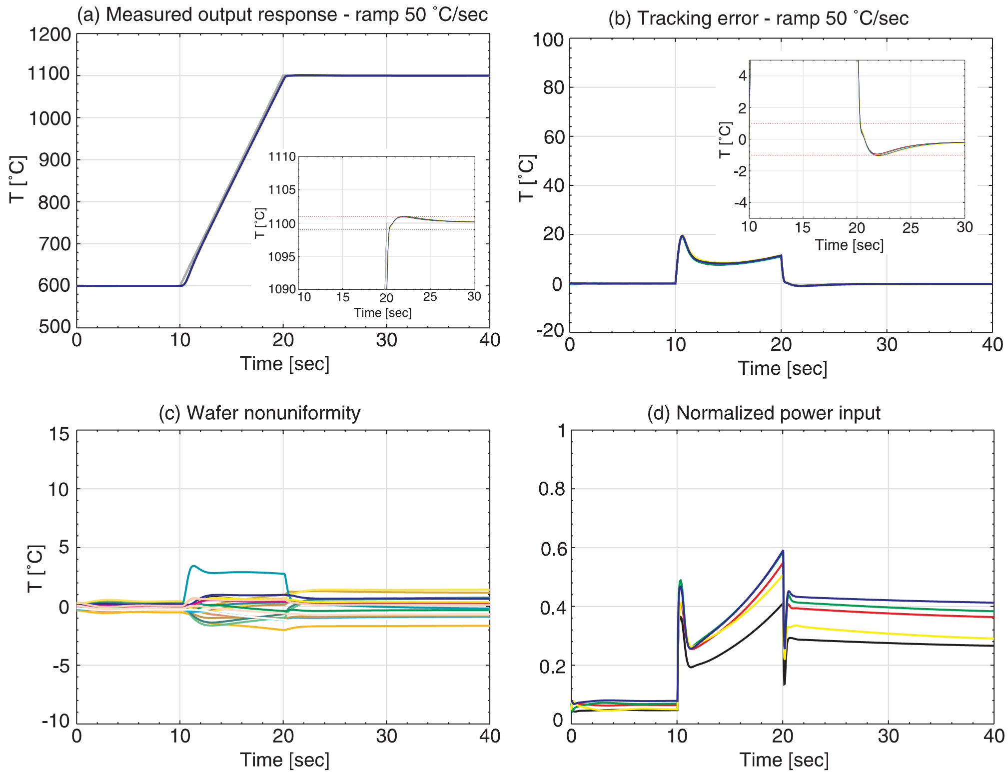

Figure 5: Simulated tracking response LQG controller for ramp rate 50^o C/sec; (a) reference ramp r and measured wafer temperature y; (b) tracking error $e \doteq r - y$; (c) wafer temperature non-uniformity for 21 nodes on wafer; each line represents the distance from the average wafer temperature; (d) power input u to RTP system, normalized to 1, *i.e.* u equals 1 represents the maximum power of 65kW.

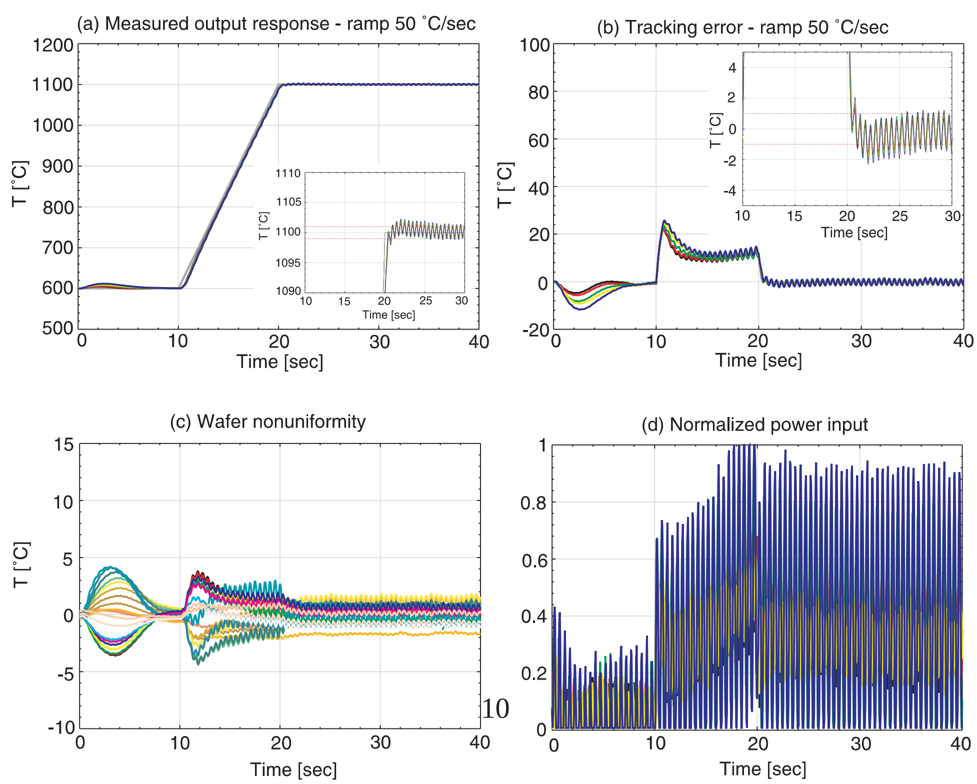

Figure 6: Simulated tracking response LQG controller for ramp rate 50^o C/sec with 2 Hz periodic disturbance added to the measurements; see Figure 5 for an explanation of (a)$\cdots$(d).

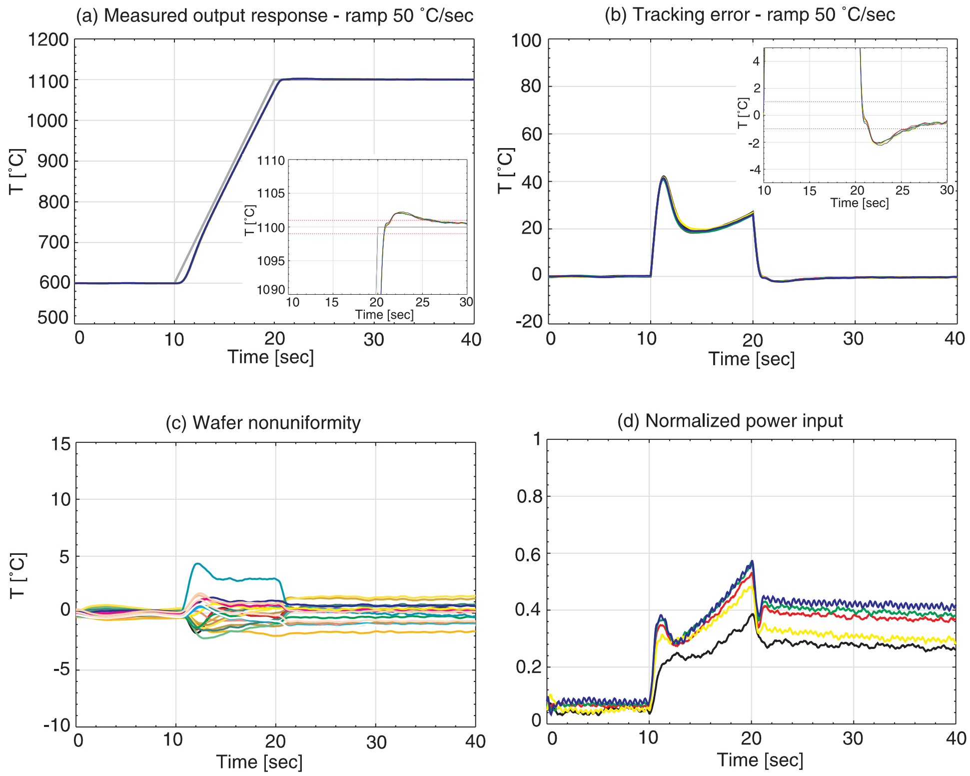

Figure 7: Simulated tracking response for controller with improved high frequency rolloff for ramp rate 50^{o} C/sec with 2 Hz periodic disturbance added to the measurements; see Figure 5 for an explanation of (a)$\cdots$(d).

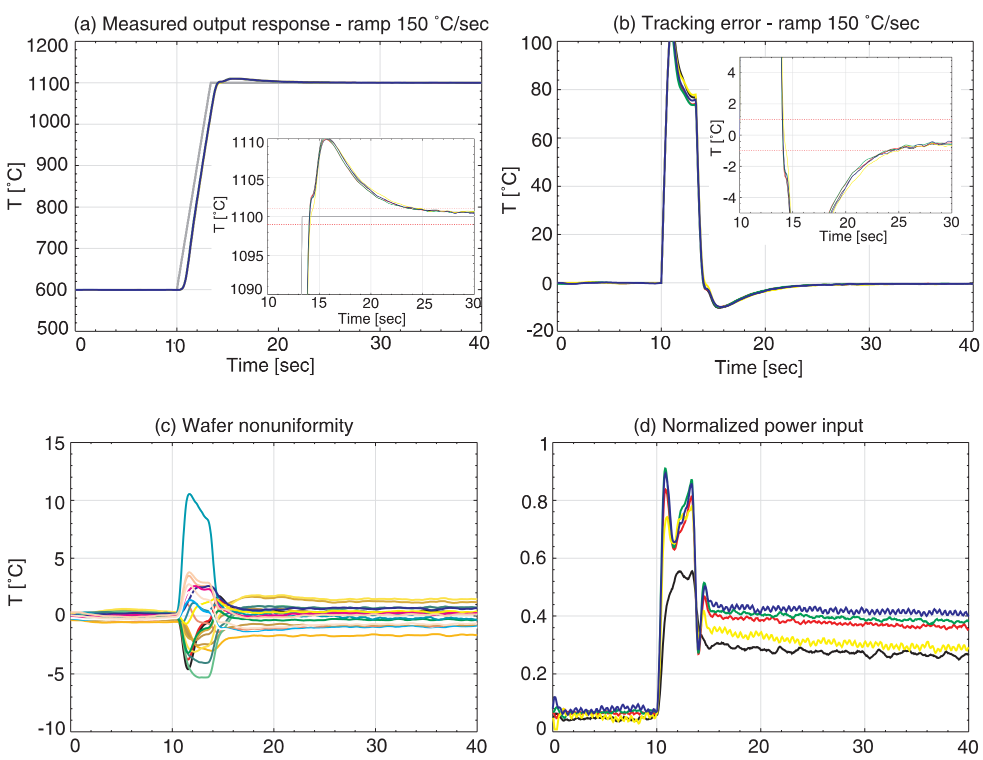

Figure 8: Simulated tracking response for controller with improved high frequency rolloff for ramp rate 150^{o} C/sec with 2 Hz periodic disturbance added to the measurements; see Figure 5 for an explanation of (a)$\cdots$(d).

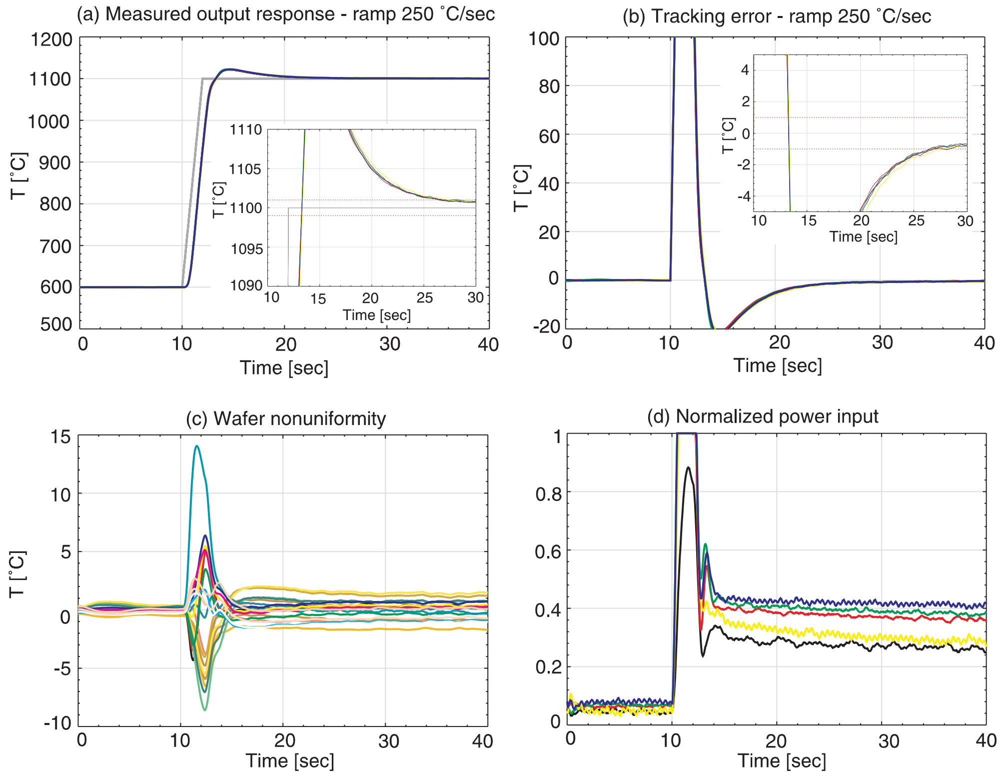

Figure 9: Simulated tracking response for controller with improved high frequency rolloff for ramp rate 250^o C/sec with 2 Hz periodic disturbance added to the measurements; see Figure 5 for an explanation of (a)$\cdots$(d).

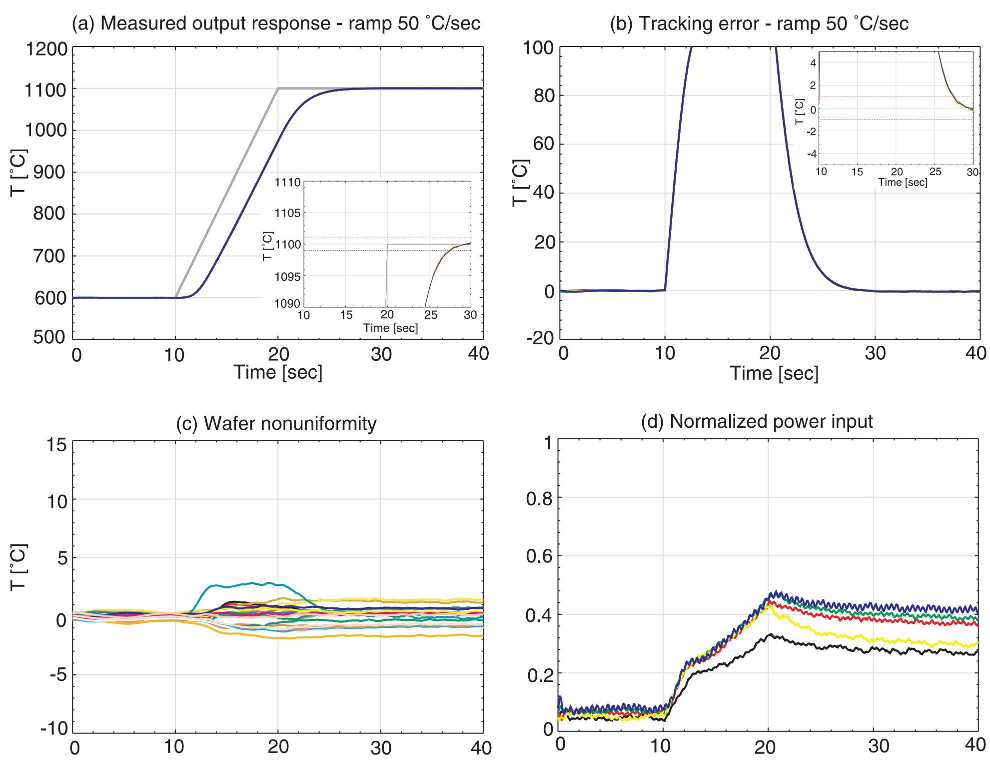

Figure 10: Simulated tracking response for controller including feedback, feedforward and prefilter for ramp rate 50^o C/sec; see Figure 5 for an explanation of (a)$\cdots$(d).

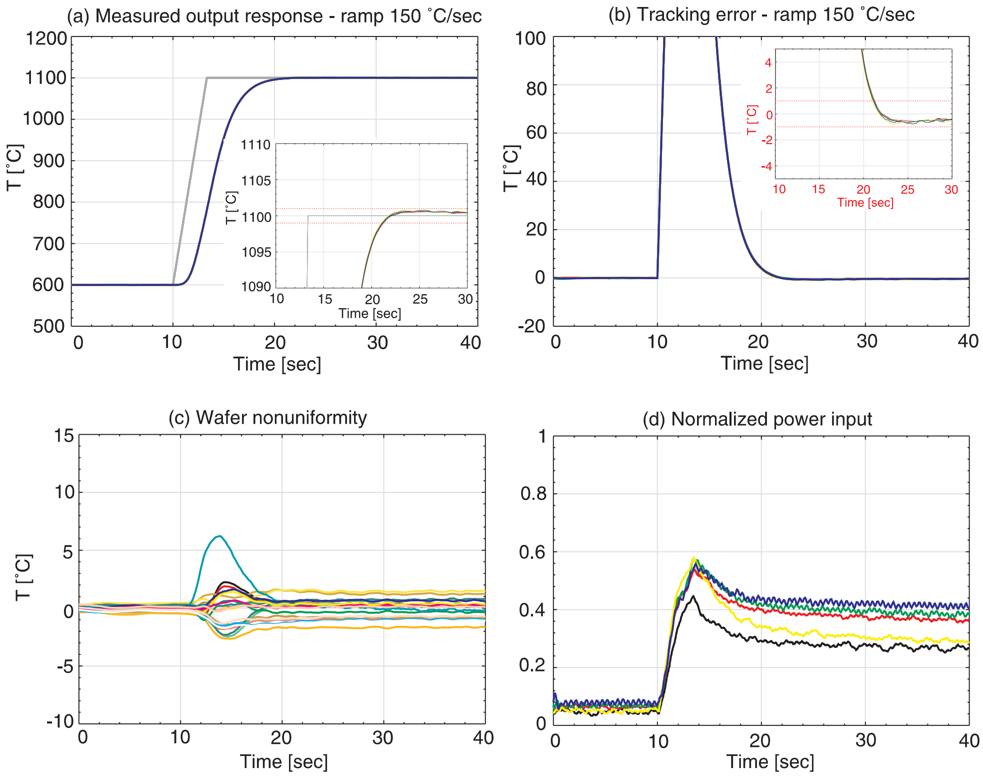

Figure 11: Simulated tracking response for controller including feedback, feedforward and prefilter for ramp rate 150^{o} C/sec; see Figure 5 for an explanation of (a)$\cdots$(d).

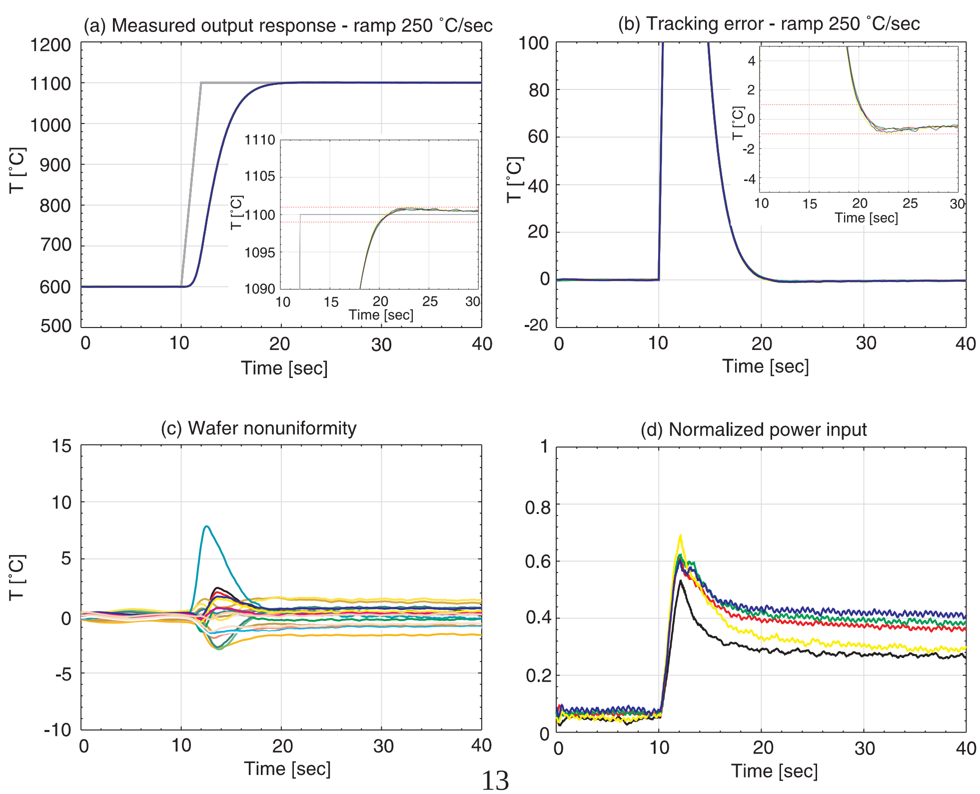

Figure 12: Simulated tracking response for controller including feedback, feedforward and prefilter for ramp rate 250^{o} C/sec; see Figure 5 for an explanation of (a)$\cdots$(d).

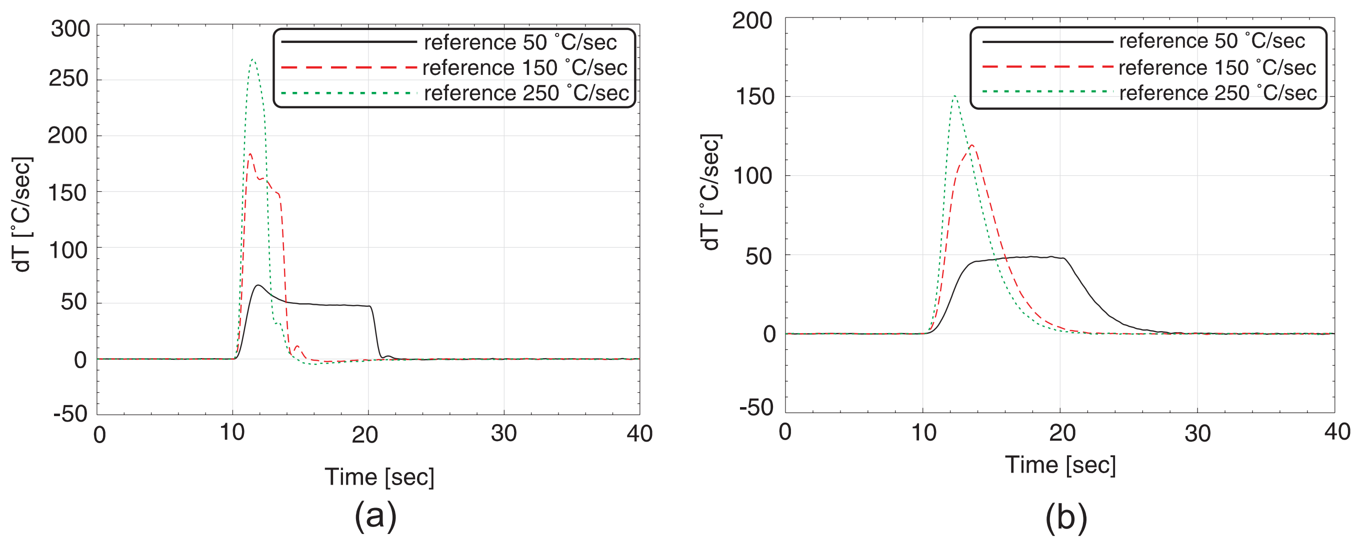

Figure 13: (a)Actual ramp rates for center outputs for different ramp rates using feedback only. (b) Actual ramp rates for center outputs for different ramp rates using feedback, feedforward, and prefilter.