INTERACTIVE CONTROL ENGINEERING
Principles and Applications

Volume 14 Control in Semiconductor Manufacturing

CHAPTER 5

A Hundred-Zone Bake Plate Simulator for Decentralized Control and Spatial Bandwidth

Pages 109–136

Edited by John D. Hedengren

How to cite this chapter. "A Hundred-Zone Bake Plate Simulator for Decentralized Control and Spatial Bandwidth," in *Control in Semiconductor Manufacturing*, John D. Hedengren, Ed., vol. 14 of *Interactive Control Engineering: Principles and Applications*. Advanced Process Solutions, LLC, 2026, ch. 5, pp. 109–136.

The page numbers in this file are those of the printed volume.

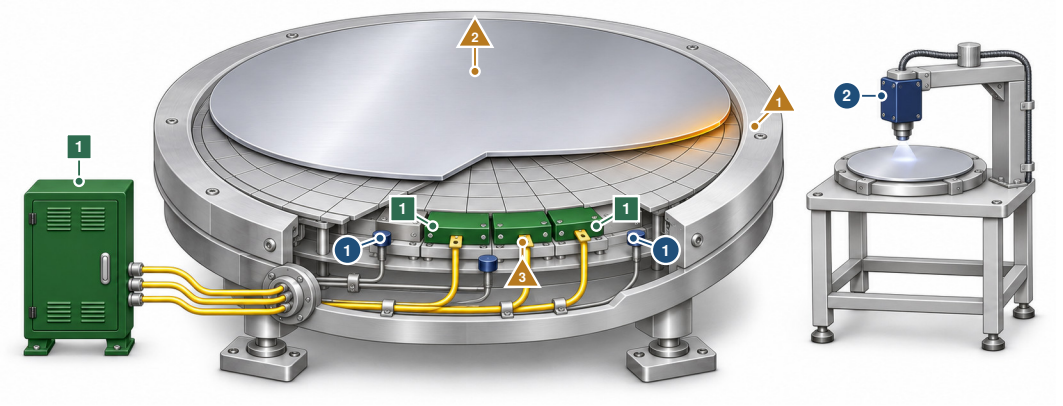
The app for this chapter
https://apmonitor.com/apps/peb-bake-zones/

The volume online: contents, other chapters and citation records
https://apmonitor.com/apps/books/v14-semiconductor/#peb-bake-zones

Copyright © 2026 Advanced Process Solutions, LLC. All rights reserved. Produced by Advanced Process Solutions, LLC.

CHAPTER 5

A Hundred-Zone Bake Plate Simulator for Decentralized Control and Spatial Bandwidth



The controller moves 1 a heater under each of the hundred zones, switched from a power cabinet

The controller reads 1 a thermocouple in each zone of the plate

2 a station that measures the printed linewidth on every field

What disturbs it 1 a bowed wafer that touches the plate badly at its edge 2 a cold wafer landing on the hot plate

3 a heater that loses most of its power

Notice that the thermocouples sit in the plate, not in the wafer: only the station at the right sees what the bake did to the printed lines. Illustration, not to scale.

IN PLAIN TERMS A Hot Plate Cut Into a Hundred Squares

The main idea. Light draws a pattern on a thin disk of silicon. Then the disk is baked on a hot plate, so a reaction in its coating can finish. The plate is cut into a hundred heated squares, and each has its own thermometer and controller.

The challenge. Each square shares heat with its neighbors. So a controller set up on one square overreacts when all of them change at once. The plate can also read perfectly while the disk bakes unevenly, because a bent disk touches the plate badly at its edge.

Why it matters. A tenth of a degree changes the width of the printed lines, so the disk itself has to be measured.

THREE WAYS TO READ THIS CHAPTER

Explore	11 min	The first page, the history box in Section 5.2, the try-it and fact boxes of Section 5.6, and the key points
Study	60 min	The whole chapter with its pause points and the twelve exercises
Practice	25 min	The first page, Section 5.6, Section 5.7, Section 5.8, and the open challenge at the end of Section 5.9

The control question A hundred zone loops on one conducting plate can each be well tuned and still hold a temperature profile nobody asked for, while the wafer on top prints a linewidth no thermocouple sees. Which of the plate's disturbances does feedback answer, and which need a model, a timed feedforward or a measurement of the product?

Learning objectives

- Derive the steady gain of the plate against the spatial wavenumber of a power pattern and explain why the zone count does not set the correction the plate can deliver.
- Explain why a gain tuned on one zone overshoots when every zone moves, and compute the factor from the conductances.
- Compute the forward-model power that holds a flat profile at the edge of the plate, and state what it buys that feedback does not.
- Size a wafer-drop pulse on the linewidth rather than on the plate dip, and explain the integral hold.
- Judge a hundred-channel run-to-run correction on its linewidth range and its spatial roughness, with and without a spatial filter.

Prerequisites Sections 1.4.5 (multizone thermal control), 1.4.3 (run-to-run correction), 1.6.2 (velocity form and loop action) and 1.6.4 (scenarios, scorecard and CSV).

The app https://apmonitor.com/apps/peb-bake-zones/

See also Volume 6, Chapter 2 and Volume 6, Chapter 3 for pairing and static decoupling. Volume 8, Chapter 10 and Volume 8, Chapter 9 for corrections learned from late measurements. Volume 3, Chapter 6 for feedforward that leaves feedback only a trim.

5.1 Introduction

One hundred heater zones are cut into one piece of aluminum, each with its own thermocouple, its own power supply and its own loop. Every zone conducts heat to the four around it. That single sentence contains the whole difficulty, because a hundred loops that are individually well tuned can still produce a temperature profile that nobody asked for, and the mechanism is not interaction in the usual pairing sense. The relative gain array of this plant is close to the identity, so the pairing question that a multivariable course asks first is already answered [1]. What is not answered is what happens when the hundred loops move in the same direction at the same time.

The reason they are there at all is chemistry. A chemically amplified resist does not finish reacting when the exposure stops, because the photoacid generated by the light deprotects the polymer during a bake that follows, and the reaction is Arrhenius [2, 3]. What the resist integrates is a thermal dose rather than a temperature, and the linewidth that results is sharply sensitive to the bake, on the order of several nanometers per degree in published measurements [4]. A plate specification written to a tenth of a degree follows directly from that number, and a plate with many independently driven zones follows from the specification.

The class of plant is larger than the application. A diffusive medium carrying many collocated actuators and sensors describes a multizone rapid thermal processing chamber [5, 6] (Chapter 9), a chemical vapor deposition susceptor, an extruder die, an injection mold and a battery electrode dryer. Every one of them has the property that matters here. The steady-state map from a power pattern to a temperature pattern is not one gain but a family of gains indexed by spatial wavenumber, which is the defining feature of a spatially invariant distributed system [7]. A plate with more zones cut into it resolves a finer correction and does not necessarily produce one. However, a reader handed a 100×100 gain matrix and told to design a decentralized controller learns the linear algebra and never meets the consequence, which is that a correction the control system can compute may be one the actuator cannot deliver without an unreasonable power bill. Section 1.4.5 of Chapter 1 sets out the same argument on a plate of five zones in a row.

This chapter works with a hundred-zone bake plate simulator on which the spatial frequency response, the boundary behavior, the wafer-drop transient and the per-zone run-to-run loop can each

be measured, and it prices four different disturbances in nanometers against the four different control elements that answer them. The simulator carries a 201-state thermal and chemical model with its instrument chain, three control architectures and two feedforward paths, and eleven scenarios that run from one power setting for a hundred zones to a graded lot on which the temperature map is perfect and the wafers are not. The thermal model, the spectral argument and the run-to-run arithmetic are standard. Section 5.2 traces how bake plates became a control problem. Section 5.3 gathers every equation and constant of the plate, the wafer and the resist, and Section 5.4 reads the spectrum of the plate and the three different gains a single loop can be said to have. Section 5.5 describes the zone loops, the three architectures, the wafer-drop pulse with its integral hold and the hundred run-to-run channels. Section 5.6 reports what each scenario measures, from the uniform actuator to the graded challenge. Section 5.7 names the thermocouples, heaters and controllers behind the loops. Section 5.8 states the limitations, Section 5.9 discusses plate design and the methods the model admits, and Section 5.10 summarizes.

5.2 Background

Resist bake was not always a control problem. A diazonaphthoquinone novolac resist prints a linewidth set mainly by exposure dose and develop time, and the post-exposure bake that follows exposure serves to relax standing waves rather than to drive the reaction that defines the feature. Bake plates of that era were single-zone aluminum slabs with one thermocouple and one cartridge heater, and the across-plate uniformity they achieved was a property of the metal and of the insulation around it rather than of any controller. Nothing in the tool was arranged to produce a deliberate temperature profile, because no profile was wanted.

Chemical amplification changed the arithmetic. Designing a resist so that one photogenerated acid catalyzes many deprotection events buys the sensitivity that short-wavelength exposure tools require [2], and it also moves the reaction that sets the linewidth out of the exposure step and into the bake [3]. Linewidth then acquires a bake-temperature sensitivity of several nanometers per degree, a coefficient later measured directly on extreme ultraviolet resists by holding a plate at a sequence of temperatures and reading the resulting critical dimension [4]. A single-zone plate that is flat to a degree is a plate that prints a range of several nanometers.

> ⧖ HISTORY **From Laboratory Curiosity to Production Resist (1980s)**
>
> Chemically amplified resists were proposed in the early 1980s, when exposure tools were moving into the deep ultraviolet and their lamps gave too little light there for the resists of the day. The first patent disclosure was rated an interesting laboratory curiosity and not worth filing. Yet the resist went quickly into the production of one-megabit memory chips, and it was so sensitive that the tool’s stage could not move fast enough: a filter was put in the beam to slow the exposure down. A serious problem followed. Airborne bases, at less than ten parts per billion, neutralized the acid while exposed wafers waited for their bake, and a skin formed on the developed lines. Filtering the air around the wafers through activated carbon became common practice, and a later resist that could be baked near 150 °C was barely affected by a long delay between exposure and bake [8].

The engineering answer arrived from two directions at once. Multizone temperature control had already been developed for rapid thermal processing, where lamp banks heat a wafer that has no thermal contact with anything, and where the modeling, identification and multivariable control of the resulting array were worked out in the early 1990s [5]. Applying the same idea to a bake plate is easier in the dynamics and harder in the specification, and the decentralized question of whether independent single-loop controllers suffice on such an array was posed in that literature [6]. Programmable thermal processing modules with many independently driven zones under a photoresist wafer followed [9], and zone-controlled post-exposure bake was demonstrated as a way of removing a global critical dimension signature on production mask blanks [10]. Later work closed the loop

Table 5.1 Variables of the Hundred-Zone Plate. Nominal Entries Describe the Design Point Loaded by Every Scenario.

Symbol	Role	Nominal	Range
T_i, $i = 1..100$	CV, measured	110 °C	90–130 °C SP
CD mean	CV, measured	62 nm	52–74 nm SP
P_i, $i = 1..100$	MV	30.80 W	0–120 W
ΔT_i^{SP}	MV, outer loop	0 °C	±4 °C
T_a ambient	DV, measured	22 °C	15–30 °C
T_{in} wafer in	DV, measured	22 °C	15–60 °C
c_d contact	DV, measured	1×	0.3–2×
b wafer bow	DV, unmeasured	0	0–0.8
s_x zone coupling	DV, measured	1×	0.2–3×
s_ℓ plate loss	DV, measured	1×	0.5–2×
g_h heater supply	DV, measured	1×	0.7–1.2×
g_f zone 34 heater	DV, unmeasured	1×	0–1×
β_0 TE bias	DV, unmeasured	0 °C	±2 °C
T_i	state (100)	110 °C	—
$T_{w,i}$	state (100)	22 °C	—
T_{rim}	state	98.42 °C	—

on a measurement of the wafer itself rather than of the plate, using in-situ sensing to control the steady-state spatial temperature profile [11], and multiplexed model predictive control was applied to the same multizone thermal plant when the computational cost of solving all zones at once became the binding constraint [12].

The correction that closes on the product came last, because it had to wait for the data infrastructure. Run-to-run control adjusts a recipe between wafers or lots using measurements taken after the step (Section 1.4.3), and it became standard practice in semiconductor manufacturing once advanced process control systems existed to carry a measurement from a metrology station back to the tool [13]. Extending a single run-to-run channel to an across-wafer map, so that a spatial signature rather than a mean is corrected, is a distinct step and has its own literature [14]. A hundred-zone plate is what an across-wafer run-to-run loop needs on the actuator side, and the reason zone counts rose is that lithographic scaling tightened the linewidth budget faster than the plates improved [15].

5.3 Process Model

Plate BP-301 is a 300 mm post-exposure bake plate cut into $N = 100$ heater zones, one under each of the hundred exposure fields of the wafer it bakes. Each field is 22 mm by 24 mm and the outermost field center sits at 130.2 mm from the wafer center, inside the 150 mm wafer radius. Zones are numbered on the field lattice, so a zone has four nearest neighbors unless it is on the boundary of the pattern, in which case the missing links attach instead to a passive plate rim that carries its own heat capacity and its own loss to ambient. Sixty-nine zones have four neighbors, eighteen have three and thirteen have two, and the forty-four missing links are rim links. Figure 5.1 draws the arrangement, the thermal paths and each symbol carried by the equations below.

The quantities the control system touches are sorted in Table 5.1. The plate carries 201 states, one temperature for each zone, one for the wafer element above it, and one for the rim. Working units are seconds, degrees Celsius, watts and nanometers, and the Arrhenius expressions convert to kelvin internally. Two clocks run above the physics, a zone loop that executes every 0.1 s and a run-to-run loop that moves once per wafer, which is every 110 s.

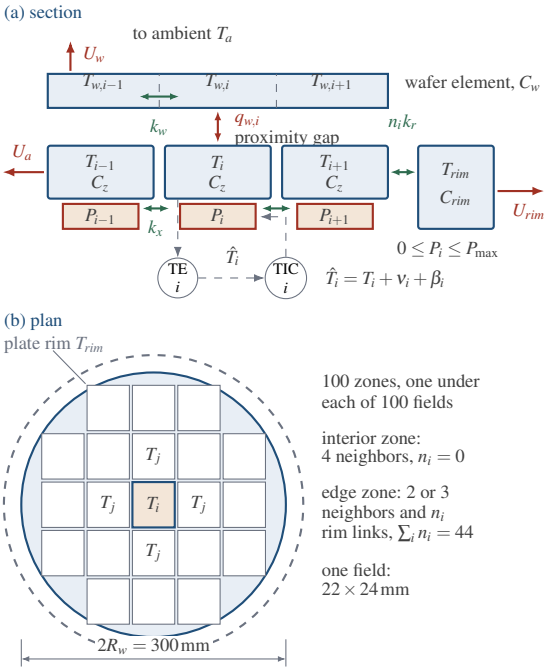


Figure 5.1 Bake plate BP-301. Section (a) shows three of the hundred cells, each a heater delivering P_i into a plate zone of heat capacity C_z at T_i, losing U_a to ambient, exchanging k_x with each neighbor and $n_i k_r$ with the rim, and passing $q_{w,i}$ across the proximity gap into the wafer element of heat capacity C_w at $T_{w,i}$, which conducts k_w sideways and loses U_w from its top face. Thermocouple TE-i reports $\hat{T}_i$ and TIC-i trims the feedforward power. Plan (b) shows the hundred fields on the wafer and the rim that the boundary zones are attached to.

5.3.1 Plate Zones, Rim and the Graph Laplacian

Each zone is a single lumped capacity driven by its heater and coupled to its neighbors, its rim links and the wafer above it, as shown in Equation (5.1).

$$\begin{aligned} C_z \frac{\mathrm{d}T_i}{\mathrm{d}t} &= P_i - U_a s_\ell (T_i - T_a) - q_{w,i} \\ &\quad - \sum_{j \in \mathcal{N}(i)} k_x s_x (T_i - T_j) \\ &\quad - n_i k_r s_x (T_i - T_{rim}) \end{aligned} \tag{5.1}$$

where C_z is the heat capacity of one zone in J/K, P_i is the delivered heater power in W, U_a is the loss to ambient, k_x is the conductance to one neighboring zone, $\mathcal{N}(i)$ is the set of lattice neighbors that exist, n_i is the number that do not, k_r is the conductance of one rim link, $q_{w,i}$ is the heat drawn by the wafer, and s_ℓ and s_x are operator multipliers on loss and on zone coupling. Both sides carry watts. The rim is one more lumped capacity collecting every rim link, as shown in Equation (5.2).

$$C_{rim} \frac{\mathrm{d}T_{rim}}{\mathrm{d}t} = \sum_{i=1}^{N} n_i k_r s_x (T_i - T_{rim}) - U_{rim} (T_{rim} - T_a) \tag{5.2}$$

With every zone at 110 °C and ambient at 22 °C, setting the left side of (5.2) to zero gives $T_{rim} = 98.42$ °C, and every scenario starts the rim state there so that the plate opens at a genuine equilibrium rather than climbing into one.

Collecting (5.1) at steady state over all hundred zones writes the plate as a matrix, as shown in Equation (5.3).

$$\mathbf{A}\,\theta = \mathbf{p}, \qquad \mathbf{A} = U_a\mathbf{I} + k_x\mathbf{L} + k_r\mathbf{D}_n \tag{5.3}$$

where θ holds the zone temperatures above ambient, $\mathbf{p}$ holds the powers, $\mathbf{L}$ is the graph Laplacian of the field lattice and $\mathbf{D}_n$ is the diagonal matrix of rim-link counts. Equation (5.3) is the entire control-relevant description of the plate. It is the two-dimensional form of the conductance matrix of Equation (1.8), with a rim in place of insulated ends, and Section 5.4 reads its spectrum.

5.3.2 The Wafer and Its Contact

A wafer is present only during the bake. While it is on the plate each wafer element exchanges with the zone below it, conducts to the wafer elements beside it and loses from its top face, as shown in Equation (5.4).

$$C_w \frac{\mathrm{d}T_{w,i}}{\mathrm{d}t} = q_{w,i} - U_w(T_{w,i} - T_a) - \sum_{j \in \mathcal{N}(i)} k_w(T_{w,i} - T_{w,j}) \tag{5.4}$$

$$\begin{aligned} q_{w,i} &= h_w\, c_d\, m_i\, (T_i - T_{w,i}) \\ m_i &= \mathrm{clip}\big(1 - b\, r_i^2 / R_w^2,\ 0.05,\ 1.5\big) \end{aligned} \tag{5.5}$$

where C_w is the heat capacity of one wafer element, h_w is the contact conductance of one cell, c_d is an operator contact multiplier, m_i is the contact map produced by a wafer bow b at field radius r_i, k_w is the lateral conductance across the silicon, and U_w is the loss from the free surface. Equation (5.5) is the only path by which a wafer defect enters, and it is deliberately radial, because a bowed wafer touches well near the center and badly at the edge. Off the plate the wafer element is not physically present, and the state is held inert at the incoming wafer temperature by $\frac{\mathrm{d}T_{w,i}}{\mathrm{d}t} = (T_{in} - T_{w,i})/\tau_h$ with $\tau_h = 2\,\mathrm{s}$, so that the state stays defined between bakes and is re-initialized to T_{in} the instant the next wafer lands.

Two numbers from Table 5.2 set the size of every result that follows. The hundred wafer elements total 89 J/K against 800 J/K for the hundred plate zones, and the wafer time constant $C_w/h_w = 2.54\,\mathrm{s}$ is half the plate zone time constant $C_z/(U_a + 4k_x) = 5.16\,\mathrm{s}$. In scenario 5 of Section 5.6 the wafer therefore covers 95 % of the 88 °C step from ambient to the plate setpoint 8.85 s after it lands and then rides 1.48 °C below the plate for the rest of the bake, which is the algebraic balance $U_w(T_i - T_a)/(h_w + U_w) = 1.483\,^{\circ}\mathrm{C}$ between the top-face loss and the contact conductance.

5.3.3 Thermal Dose and Linewidth

The resist integrates a reaction rate rather than a temperature, and because the rate is exponential in inverse absolute temperature, a cold interval and a hot interval of equal size do not cancel. The accumulated dose and the linewidth it produces are shown in Equations (5.6) and (5.7).

$$\Theta_i = \frac{1}{t_b}\int_0^{t_b} \exp\left[-\frac{E_a}{R}\left(\frac{1}{T_{w,i}} - \frac{1}{T_0}\right)\right] dt \tag{5.6}$$

$$CD_i = CD_0 - \kappa \ln\left(\frac{\Theta_i}{\Theta_0}\right) \tag{5.7}$$

where t_b is the bake time, E_a/R is the activation temperature of the deprotection in kelvin, $T_{w,i}$ and T_0 are absolute temperatures, κ is the linewidth change per unit of log dose, CD_0 is the linewidth of the calibrated bake and Θ_0 is the dose that calibrated bake delivers. The integral in (5.6) is accumulated at the 0.05 s physics step rather than at the controller scan, because it is the product.

The reference Θ_0 deserves its own sentence, because it is a calibration constant and not a derived one. A bake recipe is calibrated by printing wafers and measuring them, so the warm-up transient at the start of the bake, the landing pulse of Section 5.5 and the 1.48 °C by which the wafer sits below the plate are all inside the reference. Here $\Theta_0 = 0.917514$ is the retained historical reference for a

bake calibrated with the landing pulse running. It corresponds to holding the wafer isothermally at 108.95 °C for the whole 90 s. Writing (5.7) against Θ_0 rather than against unity is what maps a dose equal to the reference to 62.00 nm by construction, and it is also what makes every subsequent result a departure from a calibration rather than from an ideal. The reference is not recalibrated after correcting the six-second pulse clock. The shipped pulse prints 62.110 nm. A wafer held at its steady 108.52 °C for the whole bake, with no landing transient at all, would print 63.39 nm, so a bake without the pulse cannot reach the target without a temperature offset.

Differentiating (5.6) and (5.7) for a small uniform temperature offset held through the bake gives the sensitivity the whole plate specification is written from, as shown in Equation (5.8).

$$\frac{\mathrm{d}CD}{\mathrm{d}T} = -\frac{\kappa E_a / R}{T^2} = -\frac{39.1 \times 12000}{383.15^2} = -3.196\,\mathrm{nm/K} \tag{5.8}$$

A tenth of a degree is therefore 0.32 nm, which is a quarter of the 1.2 nm across-wafer range this application allows and a sixth of the 2.0 nm specification half-width on the mean.

5.3.4 Instruments

Each zone carries one thermocouple, and the controller sees $\hat{T}_i = T_i + \nu_i + \beta_i$, where ν_i is a Gaussian draw of standard deviation σ_T redrawn every controller scan and β_i is a fixed bias applied to zone 0 only. Every scenario except the graded challenge sets σ_T to zero. Critical dimension is measured on every field of every wafer at the end of the bake as $\widehat{CD}_i = CD_i + \sigma_{CD}\,\varepsilon_i$ with ε_i an independent standard normal draw. The instruments carry no dead time and no filter, and the run-to-run loop sees the map of the wafer that has just finished with no metrology queue in front of it.

5.3.5 Constants, Initial Condition and Numerics

Table 5.2 holds the constants. All but one of them are visible and adjustable in the model panel. The reference bake temperature T_0 in (5.6) is hidden, because a reader who can read it can read the answer to the dose calibration straight off the panel.

Every scenario opens with all hundred zones at the plate setpoint, the wafer elements at ambient, the rim at the equilibrium of (5.2), and each zone loop already holding the power the forward model of Section 5.5 computes. A startup test confirms that no state derivative at that condition exceeds 10^{-9} K/s. Fourth-order Runge-Kutta advances the 201 states with the step held at 0.05 s, the zone loops fire on a 0.1 s scan, and every logged channel reaches the historian five times per simulated second. The wafer cycle is counted in zone-loop scans, with its running-sum clock compared against its limits to within a microsecond, so the first wafer lands at $t = 20$ s, each bake lasts 90 s, the cycle repeats every 110 s and the shipped 6 s landing pulse runs for 60 scans. Each random draw comes from a mulberry32 stream seeded at 20260917, and the engine, its clocks and its seeding are those of Section 1.6.1, so a noisy scenario repeats draw for draw at any playback speed.

5.3.6 Assumptions

Detail was spent where a control decision depends on it and withheld everywhere else. Every simplification listed here has a condition under which it ceases to hold.

Two lumped layers per cell. A production plate has a heater layer, a plate body and a proximity air gap held by pins, and the gap contributes most of the dynamics as well as most of the sensitivity to wafer shape. Collapsing the gap into the single conductance h_w of (5.5) makes the landing transient first order when a real one is closer to second order, and it removes the difference between a wafer that is bowed and a wafer that is sitting on a particle.

Nearest-neighbor conduction with one isotropic coefficient. Production plates are cut with structured slots whose conductance is a deliberate design variable and is usually different along the two lattice directions [9]. The isotropic k_x in (5.1) makes the spectral gain of Section 5.4 separable, which is convenient and is not what a real plate does.

Table 5.2 Model Constants. All but the Reference Bake Temperature Are Shown and Editable in the Model Panel of the Application.

Symbol	Meaning	Value	Unit	Visibility
C_z	heat capacity, one zone	8	J/K	adjustable
U_a	zone loss to ambient	0.35	W/K	adjustable
k_x	conductance to a neighbor	0.30	W/K	adjustable
k_r	conductance of one rim link	0.15	W/K	adjustable
C_{rim}	heat capacity of the rim	240	J/K	adjustable
U_{rim}	rim loss to ambient	1.0	W/K	adjustable
$P_{\max}$	heater power per zone	120	W	adjustable
C_w	heat capacity, wafer element	0.89	J/K	adjustable
h_w	wafer contact conductance	0.350	W/K	adjustable
k_w	conductance across the wafer	0.10	W/K	adjustable
U_w	wafer top-face loss	0.006	W/K	adjustable
E_a/R	activation temperature	12000	K	adjustable
κ	CD per unit log dose	39.1	nm	adjustable
CD_0	CD of the calibrated bake	62	nm	adjustable
Θ_0	dose of the calibrated bake	0.917514	—	adjustable
—	CD specification half-width	2.0	nm	adjustable
—	across-wafer range allowed	1.2	nm	adjustable
t_b	bake time	90	s	adjustable
—	wafer exchange time	20	s	adjustable
σ_T	thermocouple noise, 1σ	0.02	°C	adjustable
σ_{CD}	CD-SEM repeatability, 1σ	0.40	nm	adjustable
T_0	reference bake temperature	110	°C	hidden

The wafer is in contact from the instant it lands. A real wafer is lowered onto proximity pins over roughly a second and the gap closes progressively, so the true landing transient is softer at its start than the step in (5.5) makes it. The assumption overstates how fast the first seconds of dose accumulate and therefore how sharp the feedforward pulse has to be.

One Arrhenius reaction stands for the resist. There is no acid diffusion length in (5.6), and acid diffusion is the physical effect that ultimately limits how spatially local any linewidth correction can be [3]. The model will therefore reproduce a zone-to-zone linewidth step that a real resist would blur.

The rim is one lumped state with no guard heater. Production plates carry a guard heater on the rim, which is itself a control loop and would remove most of the edge effect this application uses to make its point [9]. Raising U_{rim} or lowering k_r in Table 5.2 moves the plate toward that case.

Every field of every wafer is measured, with no queue. Production sampling plans measure a handful of sites per wafer and the measurement arrives one or more lots later, both of which make the run-to-run noise problem worse and the spatial filter of Section 5.5 more valuable rather than less [13]. Section 1.4.4 treats the metrology delay this assumption leaves out, and Chapter 6 runs a run-to-run loop through such a queue.

Heater power is continuous and instantaneous. Real zones are driven by solid-state relays with a cycle time of a second or more, which quantizes the correction and interacts with the loop gain.

5.4 The Spectrum of the Plate

5.4.1 Gain as a Function of Spatial Wavenumber

Away from the boundary the lattice in (5.1) is spatially invariant, so its steady-state response to a power pattern is diagonal in spatial frequency [7]. Take the interior of the plate with the wafer absent, write $\theta_i = T_i - T_a$, and substitute a plane wave $\theta_i = \hat{\theta}\, e^{\jmath \mathbf{q}\cdot\mathbf{x}_i}$ driven by $P_i = \hat{P}\, e^{\jmath \mathbf{q}\cdot\mathbf{x}_i}$ on an integer lattice

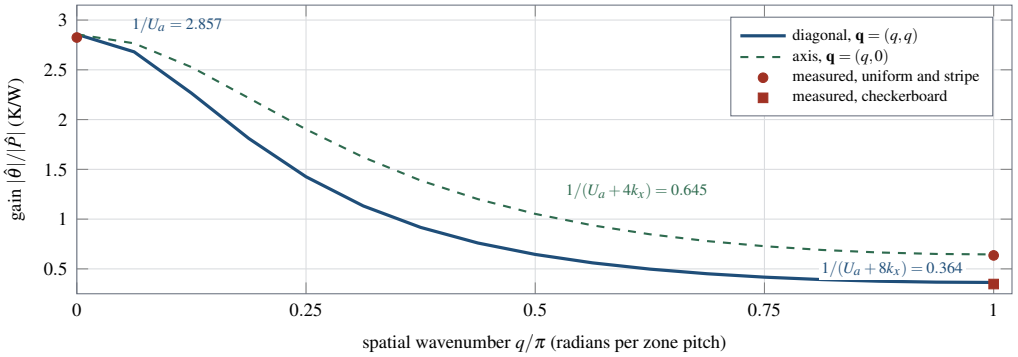


Figure 5.2 Steady-state gain of the plate against the spatial wavenumber of the power pattern, from Equation (5.9) with $U_a = 0.35\,\mathrm{W/K}$ and $k_x = 0.30\,\mathrm{W/K}$. The thick solid curve runs along the diagonal and the dashed curve along a lattice axis. The markers are steady states measured on the hundred-zone plate for a uniform 1 W addition, a striped $\pm 1\,\mathrm{W}$ pattern and a checkerboard $\pm 1\,\mathrm{W}$ pattern, averaged over the sixty-nine interior zones, filled circles for the uniform and striped patterns and a filled square for the checkerboard. The measured points sit slightly below the infinite-lattice curve because the finite plate loses heat to its rim.

$\mathbf{x}_i$. The neighbor sum collapses because $\sum_{j \in \mathcal{N}(i)} (\theta_i - \theta_j) = \theta_i\,(4 - 2\cos q_x - 2\cos q_y)$, and the result is Equation (5.9).

$$G(\mathbf{q}) = \frac{\hat{\theta}}{\hat{P}} = \frac{1}{U_a + 2k_x\,(2 - \cos q_x - \cos q_y)} \tag{5.9}$$

where q_x and q_y are spatial wavenumbers in radians per zone pitch and $\jmath$ is the imaginary unit. It is the two-dimensional counterpart of the mode gains of Equation (1.9). The number of zones does not appear in (5.9). What appears is the ratio of the diagonal loss to the coupling, and the three values a reader needs are the uniform pattern at $\mathbf{q} = 0$, the striped pattern at $\mathbf{q} = (\pi, 0)$ and the checkerboard at $\mathbf{q} = (\pi, \pi)$, as shown in Equation (5.10).

$$\begin{aligned} G(0,0) &= \frac{1}{U_a} = 2.857\,\mathrm{K/W} \\ G(\pi,0) &= \frac{1}{U_a + 4k_x} = 0.6452\,\mathrm{K/W} \\ G(\pi,\pi) &= \frac{1}{U_a + 8k_x} = 0.3636\,\mathrm{K/W} \end{aligned} \tag{5.10}$$

The ratio $G(0,0)/G(\pi,\pi) = 7.857$ is the spatial frequency response of the plate reduced to one number. It is a ratio of steady-state gains and not a bandwidth in time, so nothing about it changes when the loops are made faster. Figure 5.2 plots (5.9) along the two principal directions with the corresponding steady states measured on the simulator.

Adding 1 W to every zone of the simulated plate raises the sixty-nine interior zones by 2.8245 K on average (the whole plate by 2.7900 K) and the center zone by 2.8535 K, against 2.857 K from (5.10). A striped $\pm 1\,\mathrm{W}$ pattern produces 0.6353 K of amplitude against a predicted 0.6452 K, and a checkerboard produces 0.3476 K against 0.3636 K. The residual in each case is the boundary, which (5.9) does not model and the simulator does.

Pause. Try Exercise 5.1 before reading on.

5.4.2 Three Different Single-Loop Gains

The plate offers a reader three different answers to the question of what one loop's process gain is, and they differ by a factor of four. Holding the other ninety-nine zone temperatures fixed and adding

1 W to zone i moves only T_i, and the gain is the reciprocal of the diagonal of $\mathbf{A}$ in (5.3), which for an interior zone is $1/(U_a + 4k_x) = 0.6452\,\mathrm{K/W}$. That is the gain a loop sees when its neighbors are in perfect regulation, and it is the correct gain for decentralized tuning. Adding 1 W to zone i with every other heater left alone gives the lattice Green's function, measured on the simulator as 0.8005 K/W on the zone itself and 0.2006 K/W on each of its four neighbors. Adding 1 W to all hundred gives 2.857 K/W.

The first two numbers are exactly what the relative gain array is built from [1]. Since $\mathbf{A} = \mathbf{G}^{-1}$, the diagonal element is $\lambda_{ii} = G_{ii}A_{ii} = 0.8005 \times 1.55 = 1.241$ for an interior zone, which is close enough to unity that the array recommends the natural one-to-one pairing without reservation. That recommendation is correct and it is also not the whole story, because λ_{ii} compares the single-loop gain against the gain with every other loop closed, and it says nothing about the direction in which every loop moves together. The ratio of the common-mode gain to the single-loop gain is $G(0,0)/G(\pi,0) = 4.43$, and Section 5.6.2 measures what that factor does to a controller tuned on one loop. The value 1.241 belongs to an interior zone. A boundary zone has a smaller A_{ii} because it has fewer neighbors and a rim link instead, so its relative gain differs, and the array is not uniform across the plate even though every diagonal element stays near unity.

5.5 Controller Design

5.5.1 One Hundred Zone Loops

Each zone runs the velocity-form PID algorithm of Section 1.6.2. The derivative gain is fixed at zero in this app, so each zone loop is a PI controller, because the measurement is the plate itself and the useful anticipation arrives from the feedforward instead. In the zone's own symbols the increments are Equation (5.11), and the trim is assembled and limited as shown in Equation (5.12).

$$
\begin{aligned}
e_{i,k} &= a\,(T_{i,k}^{\mathrm{SP}} - \hat{T}_{i,k}), \qquad a = +1 \\
\Delta u_{i,k}^{P} &= K_c e_{i,k} - (K_c e_i)_{k-1}\,, \qquad \Delta u_{i,k}^{I} = \frac{K_c}{\tau_I}\, e_{i,k}\,\Delta t_c
\end{aligned}
\tag{5.11}
$$

$$
\begin{aligned}
s_{i,k} &= s_{i,k-1} + \Delta u_{i,k}^{P} \\
c_{i,k} &= s_{i,k} + I_{i,k-1} \\
u_{i,k} &= \mathrm{clip}\Big(c_{i,k},\ -P_{i,k}^{ff},\ P_{\max} - P_{i,k}^{ff}\Big) \\
I_{i,k} &= I_{i,k-1} + \Delta u_{i,k}^{I} + \frac{\Delta t_c}{T_t}\,\big(u_{i,k} - c_{i,k}\big)
\end{aligned}
\tag{5.12}
$$

where T_i^{SP} is the zone setpoint, $\hat{T}_i$ is the thermocouple reading, s_i is the running sum S_k of Equation (1.13), c_i is the trim the algorithm demands (v_k of Equation (1.14)), u_i is the trim the loop contributes in watts, I_i is the integral state of Equation (1.15), P_i^{ff} is the feedforward power described below, $\Delta t_c = 0.1\,\mathrm{s}$ is the scan period, and $T_t = \tau_I$ is the tracking time of the back-calculation, the routine's default because the app passes none. A zone reading above setpoint receives less power, so $a = +1$ in (5.11) is a reverse-acting loop in the convention of Section 1.6.2.

Two details of (5.12) are specific to this plant. The trim limits are the real heater limits less whatever the feedforward is already asking for, so the unwinding happens against the physical limit rather than a notional one. The back-calculation removes $\Delta t_c/T_t$ of the excess per scan (1 % at the default $\tau_I = 10\,\mathrm{s}$), and the running sum $s_{i,k}$ is never truncated. In Manual mode, and while the landing pulse of Section 5.5.3 runs, the routine is handed the trim actually applied and sets $I_{i,k} = u_{i,k} - s_{i,k}$, so the integral tracks rather than integrates.

Manual mode applies one power to all hundred zones, which is the arrangement scenario 1 exists to discredit. The model term is not applied in Manual, but each loop tracks the operator power less the feedforward its architecture will add in Auto, so a transfer to either closed architecture hands

over exactly the operator power. After 30 s in Manual in scenario 2 the center zone moves from 31.56 W to 31.55 W at the first Auto scan, the edge zone to 31.57 W, and the plate mean peaks at 110.16 °C. Open loop carries no trim and applies its own uniform power. The reverse transfer applies the mean delivered power to every zone, which is bumpless in the mean only. The tuning constants are bounded to $0 \le K_c \le 60\,\mathrm{W/^\circ C}$ and $0.5\,\mathrm{s} \le \tau_I \le 200\,\mathrm{s}$, and the output carries no rate limit. The scorecard scores the two displayed zone loops against a 0.15 °C band and the linewidth loop against a 0.4 nm band.

5.5.2 Three Architectures and a Forward Model

The feedforward term in (5.12) is what distinguishes the three architectures the application offers. Open loop supplies the power an interior zone needs and nothing else, $P_i^{ff} = U_a(T_i^{\mathrm{SP}} - T_a)$, with the loops tracking rather than acting. One hundred independent PI loops set $P_i^{ff} = 0$ and leave the whole power to the trim. The third architecture computes the power the entire setpoint profile requires from the conductances of (5.1), as shown in Equation (5.13).

$$\begin{aligned} P_i^{ff} = {} & U_a s_\ell (T_i^{\mathrm{SP}} - T_a) \\ & + \sum_{j \in \mathcal{N}(i)} k_x s_x (T_i^{\mathrm{SP}} - T_j^{\mathrm{SP}}) \\ & + n_i k_r s_x (T_i^{\mathrm{SP}} - T_{rim}) \end{aligned} \tag{5.13}$$

with an additional $U_w(T_i^{\mathrm{SP}} - T_a)$ while a wafer is present. Equation (5.13) is the row of $\mathbf{A}$ belonging to zone i applied to the setpoint vector, which is to say it is $\mathbf{A}\boldsymbol{\theta}^{\mathrm{SP}}$ evaluated one row at a time, the feedforward of Section 1.4.5. It is a static decoupler, like the one Volume 6, Chapter 3 places on a two-by-two plant, and the point worth making about it is the direction in which it runs. Asking what temperature a power pattern produces requires $\mathbf{A}^{-1}$, a dense inverse of a hundred-by-hundred matrix that has to be recomputed whenever a conductance changes. Asking what power a temperature profile needs requires $\mathbf{A}$ itself, which is sparse with at most five nonzero entries per row, and the answer is one matrix-vector product with no inverse in it at all. A controller only ever asks the second question. The rim temperature in (5.13) is read from the plant rather than predicted, which is legitimate because the rim is a measured state and not a manipulated one.

5.5.3 The Wafer-Drop Pulse and the Integral Hold

A cold wafer arriving on a hot plate is a disturbance whose time of arrival and size are both known before it happens, which makes it an accounting problem rather than a control problem (Section 1.4.7 treats feedforward on a known event). The app answers it with a rectangular power pulse added to every zone, triggered by the wafer landing and not by the error it causes, as shown in Equation (5.14).

$$P_i^{pulse}(t) = \begin{cases} P_p, & 0 \le t - t_{land} < t_p \\ 0, & \text{otherwise} \end{cases} \tag{5.14}$$

with defaults $P_p = 26\,\mathrm{W}$ and $t_p = 6\,\mathrm{s}$. The delivered power is then assembled and limited as shown in Equation (5.15).

$$\begin{aligned} P_i^{cmd} &= \mathrm{clip}\Big(P_i^{ff} + u_i + P_i^{pulse},\ 0,\ P_{\max}\Big) \\ P_i &= \mathrm{clip}\Big(P_i^{cmd}\, g_{f,i}\, g_h,\ 0,\ P_{\max}\Big) \end{aligned} \tag{5.15}$$

where P_i^{cmd} is the power the control system asks the heater for, $g_{f,i}$ is the zone heater health, unity everywhere except on the faulted zone, and g_h is a supply multiplier. The two clips in (5.15) are separate on purpose. A control system cannot command a heater above its rating whatever the arithmetic in front of it produces, and a heater that has lost most of its element delivers less than it

was asked for. The lower limit of zero is the one-sided actuator. A resistive heater cannot remove heat, so a zone that needs to come down has no actuator at all and the whole profile has to be lifted to make room for it. The saturation flag the scorecard reports is raised when P_i^{cmd} sits at either limit or the trim in (5.12) is clipped, which is a statement about the command rather than about the power that reached the plate. On a zone whose heater has failed the two differ by the heater health, and it is the command that says the loop has run out of authority.

While the pulse of (5.14) runs, the zone loops stop integrating and track their last trim instead, using the same $I_{i,k}$ assignment that manual mode uses. That hold is not an implementation convenience. Without it each loop sees the pulse as an error of its own making, drives its integrator down during the six seconds the pulse is applied, and then holds the plate below setpoint for the remainder of the bake while the integrator recovers, which undoes more than the pulse gained (Section 5.6.4 measures it). A feedforward term the feedback fights is worth less than no feedforward at all, and holding the integrators through a transient whose size is already known is the standard remedy.

🕮 **Pause.** Try Exercise 5.3 before reading on.

5.5.4 One Hundred Run-to-Run Channels

Above the zone loops sits an integral controller with one channel per zone, updated once per wafer from the measured linewidth map, as shown in Equation (5.16).

$$\begin{aligned} \delta_i &= \lambda_z \left(\widehat{CD}_i - CD^{\mathrm{SP}} \right) / s_{CD} \\ \Delta T_i^{\mathrm{SP}} &\leftarrow \mathrm{clip}\left(\Delta T_i^{\mathrm{SP}} + \delta_i,\ -\Delta T_{\max},\ \Delta T_{\max} \right) \end{aligned} \tag{5.16}$$

where $\lambda_z = 0.50$ is the update weight, $s_{CD} = 3.196\,\mathrm{nm/K}$ is the magnitude of the sensitivity in (5.8) and converts a linewidth error into a temperature correction, and $\Delta T_{\max} = 4\,^{\circ}\mathrm{C}$ bounds the accumulated correction. While the loop is on, the zone setpoints in (5.11) are $T_i^{\mathrm{SP}} = T^{\mathrm{SP}} + \Delta T_i^{\mathrm{SP}}$. Equation (5.16) is one hundred independent discrete integrators, each the incremental form of Equation (1.3) with the process gain estimated as $-s_{CD}$, and every one of them is driven by its own independent measurement noise. A field printed wide raises its zone setpoint, so each channel is direct acting in the convention of Section 1.6.2.

An optional spatial filter averages each update with those of its lattice neighbors before it is accumulated, replacing δ_i in (5.16) by $\bar{\delta}_i$ as shown in Equation (5.17).

$$\bar{\delta}_i = (1 - w)\,\delta_i + \frac{w}{1 + |\mathcal{N}(i)|} \left(\delta_i + \sum_{j \in \mathcal{N}(i)} \delta_j \right) \tag{5.17}$$

with $w = 0.60$ by default. The accumulated map is never filtered, so the filter shapes each increment and not the correction already in place. The app reports the spatial roughness of the correction map as a scalar, defined in Equation (5.18) as the root-mean-square of the discrete Laplacian of the offset map divided by the root-mean-square of the map itself.

$$\begin{aligned} \mu_i &= \frac{1}{|\mathcal{N}(i)|} \sum_{j \in \mathcal{N}(i)} \Delta T_j^{\mathrm{SP}} \\ c^2 &= \sum_i \left(\Delta T_i^{\mathrm{SP}} - \mu_i \right)^2 \Big/ \sum_i \left(\Delta T_i^{\mathrm{SP}} \right)^2 \end{aligned} \tag{5.18}$$

A smooth radial correction scores near zero on (5.18) and a correction that is independent from zone to zone scores near one. Equations (5.9) and (5.18) together say what the filter is for. A checkerboard offset of 1 °C amplitude costs $1/G(\pi,\pi) = 2.750\,\mathrm{W}$ per zone while a uniform offset of the same amplitude costs $1/G(0,0) = 0.350\,\mathrm{W}$, so every unit of roughness in the correction is paid for at 7.86 times the watts per degree of the part of the correction that was in the product to begin with.

Table 5.3 Scenarios and Completion Requirements

#	Scenario and kind	What it shows	Required to complete
1	One power for a hundred zones, Manual *experiment*	a uniform actuator on a plant with a boundary	In Manual, apply the uniform setpoint power (30.80 W) until the plate range exceeds 5.2 °C, then raise every zone to at least 31.5 W for 60 s and compare the shape of the map with its level.
2	Holding the profile *experiment*	common-mode against single-loop gain	After at least 2 s of running, step the plate setpoint to 110.5 °C with $K_c < 10\,\mathrm{W/°C}$ until the center and edge zones are within 0.05 °C of it, then reload the case, raise K_c to 20 W/°C, run 2 s at 110 °C and step again until both are within 0.05 °C.[b]
3	Open loop and the edge *experiment*	the forward model against feedback	Settle in open loop to a plate range above 5.2 °C and switch to a hundred PI loops until it is below 0.05 °C, then reload the case, settle again and reach the same range under Model plus PI.
4	A hundred independent loops, zone 34 heater fails at $t = 120\,\mathrm{s}$ *demonstration*	one fault read off two maps	Run a hundred PI loops at $K_c \geq 10\,\mathrm{W/°C}$ through the failure until a CD range above 10 nm, then Reset with $K_c \leq 5\,\mathrm{W/°C}$ and run through the failure until a wafer baked after it again shows more than 10 nm.
5	A wafer lands *demonstration*	a disturbance no loop is fast enough for	With the landing pulse off, see the plate mean fall more than 2 °C below setpoint during a bake, then finish that wafer.
6	The feedforward pulse *experiment*	timed feedforward and the integral hold	Finish a wafer with a pulse of 24 to 28 W, then Reset and bake a wafer at 16 W or less that prints above 64 nm and one at 38 W or more that prints below 60 nm.
7	A bowed wafer *demonstration*	a signature no plate instrument can see	See the plate range exceed 0.8 °C at a landing, then with run-to-run off and bow at least 0.3 bake three wafers until the CD range exceeds 1.2 nm while the plate range is below 0.3 °C.
8	Run-to-run, one channel per zone *experiment*	a hundred integrators and their noise	With run-to-run on, reach a CD range below 2.5 nm once three wafers are baked and, with the correction per zone, below 1.2 nm once six are.
9	The frequencies the plate cannot produce *experiment*	spatial bandwidth as a cost	With the spatial filter on, reach a CD range below 1.2 nm once seven wafers are baked, then set $w \geq 0.95$, fail the zone 34 heater and hold a CD range above 10 nm for 220 s.
10	A thermocouple reading high, $+0.5\,°\mathrm{C}$ at $t = 200\,\mathrm{s}$ *demonstration*	a loop that satisfies a bad instrument	After the bias, bake to the third wafer with a CD range above 1.0 nm, then switch run-to-run on and bake to the fifth with a zone 0 offset above 0.15 °C.
11	Challenge, a bowed lot *challenge*	four disturbances, four answers, in order	Over 700 s (six wafers), keep the mean CD within 0.35 nm of 62 nm from $t = 462\,\mathrm{s}$ (wafers 4 to 6), finish with an across-wafer range below 1.2 nm and a mean zone offset within $\pm 1\,°\mathrm{C}$, with the model parameters and disturbances as given.[a]

[a] As issued the worst mean-CD error over wafers 4 to 6 is 5.69 nm, sixteen times the tolerance, with a 3.12 nm range and all six wafers out of specification. The suggested set (Model plus PI, the pulse, $K_c = 6\,\mathrm{W/°C}$, $\tau_I = 12\,\mathrm{s}$, filtered run-to-run at $\lambda_z = 0.6$) gives 0.09 nm, 0.83 nm and 0.40 °C, passing with margins of 0.26 nm, 0.37 nm and 0.60 °C, identically after a Reset and on eight other noise seeds. [b] A setpoint chosen before the run starts is where the plate starts, so both steps are made on a running plate. Both zones settle within 0.05 °C 9.9 s after the step at $K_c = 4\,\mathrm{W/°C}$ and 0.75 s after it at 20 W/°C.

5.6 Results

Table 5.3 sets out the eleven scenarios in the order the app lists them, with what each one shows and what completes it (Section 1.6.4 describes scenarios and completion rules). Each one pins the mode, the architecture, the tuning, the disturbances and any timed event, and each puts its question to the reader ahead of the run rather than after it. Every result below was measured on the app's own engine with the case selected and any stated change made before the run starts, unless the text says otherwise. Table 5.4 holds what was measured, and Figure 5.3 redraws its wafer rows. Figure 5.4 shows the app during scenario 9. The exported file keeps the true temperature of the center zone beside its thermocouple reading, the mean and the extremes of the linewidth map, the spread of the zone offsets and the spatial roughness of the correction.

Table 5.4 Measured Outcomes Read From the App's Own Scorecard. Plate Range Is the Spread Across the Hundred Zones at the End of a Bake and CD Figures Are Those of the Last Completed Wafer. Rows 8 and 9 Report the Mean Over Wafers 10 to 20 of the Shipped Noise Seed and Challenge Rows the Sixth Wafer.

#	Configuration	Plate range (°C)	Mean *CD* error (nm)	*CD* range (nm)	Roughness c (—)	Note
1	MANUAL, 30.80 W every zone	5.437	—	—	—	17.38 nm of linewidth in the plate map
3	Open loop, same power	5.437	—	—	—	corner zones 104.45 °C, interior 109.89 °C
3	Forward model, all feedback removed*	1.1×10^{-12}	—	—	—	$K_c = 0$, no trim, from open-loop profile
2	Model plus PI, $K_c = 4$, $\tau_I = 10$ s	0.000	—	—	—	settles from the open-loop profile in 52.8 s
4	100 PI loops, heater fault, wafers baking	12.089	+4.609	23.635	—	faulted field 23.63 nm off the median
5	Model plus PI, wafer landing, no pulse	0.000	+4.015	0.010	—	plate mean dips 2.777 °C
6	Pulse 26 W for 6 s	0.001	+0.110	0.266	—	dip 0.096, overshoot 9.290 °C
6	Pulse 26 W for 2 s	0.000	+3.869	0.004	—	dip 0.287, overshoot 1.238 °C
6	Pulse 14 W for 6 s	0.001	+3.592	0.020	—	dip 1.539 °C
6	Pulse 40 W for 6 s	0.003	−4.928	0.798	—	dip 0.000 °C
7	Bowed wafers $b = 0.5$, no run-to-run	0.002	+1.996	3.765	—	radial, center 62.19 nm to edge 65.95 nm
8	Run-to-run per zone, $\lambda_z = 0.5$	1.702	+0.011	0.745	0.254	offset map RMS 0.785 °C, spread 1.71 °C
9	Run-to-run spatially filtered, $w = 0.6$	1.599	+0.014	0.595	0.189	offset map RMS 0.778 °C, spread 1.61 °C
10	Thermocouple on zone 0 reads +0.5 °C	0.500	+0.126	1.298	—	zone 0 held at 109.501 °C
11	Challenge, tuning as issued	0.008	+5.693	3.120	—	6 of 6 wafers out of specification
11	Challenge, suggested configuration	1.223	+0.009	0.832	0.234	1 of 6 wafers out, mean offset 0.40 °C

* A model run outside the app's controls. In the app, switching architecture carries the open-loop power into the trims, so with $K_c = 0$ the plate keeps its 5.437 °C profile, and scenario 2 at $K_c = 0$ holds its flat start exactly.

5.6.1 A Uniform Actuator on a Plant With a Boundary

Scenario 1 gives every zone the power an interior zone needs. That power is $U_a(T^{\mathrm{SP}} - T_a) = 0.35 \times 88 = 30.80$ W, and the interior zones farthest from the edge reach 109.89 °C because their neighbors are at nearly their own temperature and little heat flows sideways. The sixty-nine four-neighbor zones average 108.99 °C, the eighteen three-neighbor zones 106.18 °C and the thirteen two-neighbor corner zones 104.72 °C, the coldest at 104.45 °C, against a rim that has fallen to 94.36 °C. The plate range is 5.437 °C, which at the 3.196 nm/K of (5.8) is 17.38 nm of linewidth against a 1.2 nm across-wafer allowance. None of that is a control failure. It is what a uniform actuator does to a plant with a boundary, and it is why the zones exist. Raising every zone to 31.8 W lifts the plate mean from 107.93 °C to 110.72 °C and scales every temperature rise above ambient by the same factor, so the level of the map moves and its shape does not.

The power the edge zones need is not a mystery either. Evaluating (5.13) on a flat 110 °C profile returns 30.800 W for a four-neighbor zone, 32.537 W for a three-neighbor zone and 34.274 W for a two-neighbor corner, a total of 3156.4 W against the 3080 W the uniform recipe delivers. In a model run outside the application's controls, loading those powers with no feedback at all (loop gain zero, trims zero) brings the plate from the 5.437 °C profile to a range of 1.1×10^{-12} °C, which is the

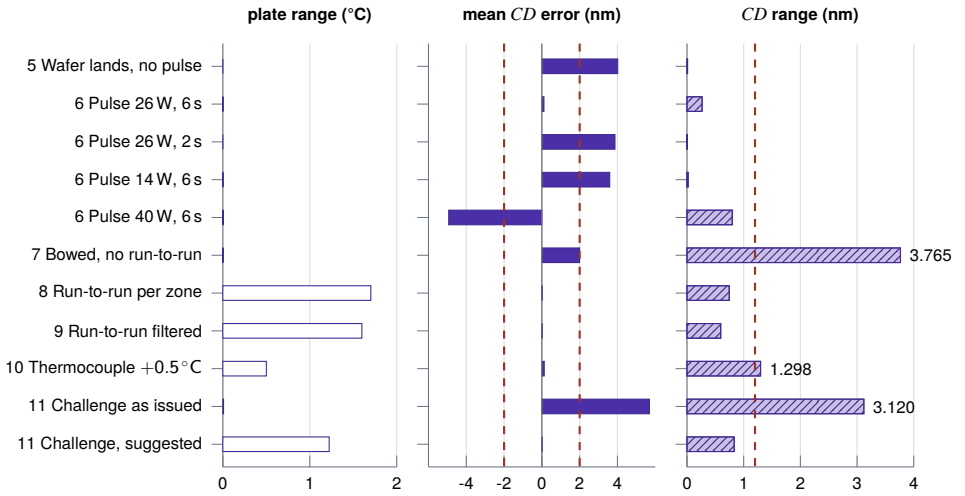


Figure 5.3 The wafer rows of Table 5.4 side by side: plate range at the end of a bake (open bars), mean linewidth error (solid bars, the 2.0 nm specification half-width dashed) and across-wafer linewidth range (hatched bars, the 1.2 nm allowance dashed). The rows of scenarios 1 to 3 carry no linewidth, and the heater fault of scenario 4 is left out because its 23.635 nm range would set the scale. Notice that the bowed wafers of scenarios 7 to 9 and 11 come into specification only on the three plates left least flat.

residual of the integrator and not of the model. In the application, scenario 2 at $K_c = 0$ holds its flat start exactly. The approach takes 56.9 s to a range below 0.05 °C, because the edge zones wait on the rim, whose time constant $C_{rim}/(U_{rim} + \sum_i n_i k_r) = 31.6\,\text{s}$ is six times that of a zone. Selecting either closed architecture from the settled open-loop plate carries each zone's command into its new trim and keeps the present proportional term in the loop's running sum, so the transfer is bumpless. The coldest corner zone holds 30.80 W through the scan that applies the change and moves 0.22 W on the next, and the plate mean never falls below its open-loop value. The plate then reaches a range below 0.05 °C in 44.7 s under a hundred PI loops and 52.8 s under model plus PI. The hundred PI loops are the faster of the two, which is the opposite of what a reader expects a decoupler to buy. At the instant of a bumpless switch the trims cancel the extra edge power the model asks for, and they release it only as fast as they integrate. Both beat the feedforward alone because the loops drive the corner zones above their final 34.32 W (35.24 W after 10 s) while the rim warms. Speed is not what the model buys. What it buys is a power profile computed once from the plate drawing rather than searched for by a hundred integrators, which leaves the authority those integrators would have spent available for the disturbances nobody wrote down.

5.6.2 A Gain Tuned on One Loop

Scenario 2 makes the spectral argument of Section 5.4 operational. In the application the plate runs Model plus PI and the common-mode step of 0.5 °C brings the center and edge zones within 0.05 °C of the new setpoint 9.9 s after a step made on the running plate at $K_c = 4\,\text{W}/^\circ\text{C}$, with a 10.6 % center overshoot, and 0.75 s after it at 20 W/°C, because the forward model carries the new power at once and the loops only trim it. A reader who tunes one zone by stepping that zone's setpoint alone measures the single-loop gain of 0.6452 K/W and a time constant of 5.16 s, and a SIMC tuning that asks for a closed loop no faster than the plant itself returns $K_c = 1/0.6452 = 1.55\,\text{W}/^\circ\text{C}$ and $\tau_I = 5.16\,\text{s}$ [16]. The same procedure run on the common mode measures 2.857 K/W and 22.86 s and returns $K_c = 0.35\,\text{W}/^\circ\text{C}$ and $\tau_I = 22.86\,\text{s}$. The two gains differ by exactly $G(0,0)/G(\pi,0) = 4.43$.

Table 5.5 The Same Two Tunings on Two Different Spatial Directions, Measured on One Hundred PI Loops With No Feedforward. A 0.5 °C Setpoint Move Is Applied to One Zone or to All Hundred, and the Center Zone Is Read. Settling Is Timed From the Scan at Which the Move Takes Effect.

	single-loop tuning $K_c = 1.55$, $\tau_I = 5.16$		common-mode tuning $K_c = 0.35$, $\tau_I = 22.86$	
Move	one zone	all zones	one zone	all zones
Overshoot (%)	0.0	18.8	0.0	0.0
Peak rise (°C)	0.500	0.594	0.421	0.500
Settling to ±0.02 (s)	15.35	23.55	> 200	73.35
Loop gain $K_c G$ (—)	1.00	4.43	0.226	1.00

Table 5.5 applies each tuning to both kinds of move on the same plate with a hundred PI loops and no feedforward. It is a model experiment rather than a run in the app, because the app offers no setpoint for a single zone and its slider grid does not hold these exact constants. The nearest slider settings, $K_c = 1.6\,\mathrm{W/°C}$ and $\tau_I = 5\,\mathrm{s}$, overshoot 19.1 % on the all-zone move and settle in 22.8 s. The single-loop tuning is textbook on the move it was designed for and overshoots 18.8 % when all hundred zones step together, because the loop gain product on that direction is $K_c\,G(0,0) = 4.43$ instead of the unity the rule intended. The common-mode tuning reverses the failure exactly, reaching only 84 % of a single-zone step within 200 s because its loop gain product on that direction is 0.226. Neither controller is mistuned in any sense a single step test can reveal.

5.6.3 A Local Fault With a Distributed Signature

Scenario 4 bakes wafers with a hundred independent PI loops and removes 88 % of one heater's authority at $t = 120\,\mathrm{s}$, between the first and second wafers of the lot. It is worth reading twice, because what the plate does and what the wafer does are different, and the scenario is arranged so that the same fault can be read off both maps at once. On the plate the fault is one zone wide. Sixty seconds into any bake from the third wafer on, the faulted zone sits 12.072 °C below its setpoint while its three neighbors are within 0.003 °C of theirs, and exactly one zone of the hundred is more than 0.003 °C off. The neighbors succeed completely. Each of them sees heat leaving sideways, raises its own power, and holds its own zone, so the temperature map shows one cold square and nothing around it. The instant has to be named because the neighbor deviation is still decaying at that point, from 0.043 °C at 30 s into the bake through 0.0022 °C at 60 s to 0.0001 °C at the end of it. What the neighbors are finishing off is the wafer landing and not the fault, because the same decay repeats to the fourth decimal on every wafer from the third onward. What is diagnostic on the plate is not the temperature map but the power map and the saturation flag, because the faulted loop is commanded the full 120 W rating and delivers 14.40 W through the damaged element. The flag is raised at the scan 2.1 s after the fault lands and stays raised for the rest of the run, and it is clear in every other scenario as shipped. On the wafer the same fault is five zones wide. Silicon conducts sideways with nothing regulating it, so the wafer above the faulted zone runs 6.97 °C below the median of the eighty-six far-field elements and the wafer above the three neighbors runs 1.03 °C, 1.14 °C and 1.31 °C below it. Those three differ from each other because the three zones carry different nominal powers, 30.80 W, 32.54 W and 34.27 W, so the edge effect of Section 5.6.1 reappears inside the fault. The linewidth map that results puts the faulted field 23.63 nm off the median, its three neighboring fields 3.43 nm to 4.39 nm off and the second ring 0.52 nm to 1.08 nm off, with 11 of the hundred fields more than 0.2 nm from the median and 8 more than 0.5 nm. The spreading is the k_w term of (5.4) and not conduction in the plate. The width of the signature is set by the wafer and not by the actuator that caused it, which is what makes reading the map backward to find the broken heater a skill rather than a lookup. The first wafer of the lot is baked before the fault and is worth keeping. It comes out at 66.210 nm with an across-wafer range of 0.001 nm, which is the wafer-drop error of Section 5.6.4 arriving on its own

because this scenario runs with the feedforward pulse switched off. Every later wafer carries the same 66.21 nm median with the fault signature on top of it. Two faults of completely different kinds are therefore visible in the same map, one uniform and one local, and only the local one moves the range.

5.6.4 A Disturbance With a Known Arrival Time

Scenario 5 drops a 22 °C wafer onto a 110 °C plate with every loop healthy and the forward model in service. The plate mean falls 2.777 °C and is back within 0.1 °C 33.7 s after the landing, and the linewidth comes out 4.015 nm wide of target on every field of the wafer, which is twice the 2.0 nm specification from one event on a tool whose loops are all working. The across-wafer range is 0.010 nm, which is the point. Nothing about this failure is a uniformity failure.

> **TRY IT IN THE APP** **5 · A Wafer Lands**
>
> Load the scenario and press Run. Watch the landing strip as the first wafer lands: the plate mean falls 2.777 °C and is back within 0.1 °C 33.7 s later. When the bake ends, read the wafer note against the 62 nm target. The linewidth comes out 4.015 nm wide on every field, although every loop on the plate is working.

Most of that cost is not the dip. The wafer heats by taking heat out of the plate, so the plate is coldest exactly while the wafer is climbing, but the wafer is then far below the plate and the Arrhenius rate it integrates is small. Integrating the dose of (5.6) through one bake shows where the deficit lives. The first 10 s of the bake contribute 4.31 % of the final dose against the 11.11 % they would contribute at a constant rate, and by 30 s the accumulated dose has recovered to 28.53 % against 33.33 %. A 26 W pulse held for only 2 s removes the dip almost entirely (0.287 °C, with a 1.238 °C overshoot) and the line still prints 65.869 nm. The deficit is the calibration of Section 5.3.3, whose reference dose belongs to a bake with the pulse running. A simple mixing estimate $C_w^{tot}(T^{\mathrm{SP}} - T_{in})/(C_w^{tot} + C_z^{tot}) = 8.81\,^{\circ}\mathrm{C}$ badly overstates the dip, because the heaters keep running and the plate zone time constant is twice the wafer time constant, but it gives the right order.

> **DID YOU KNOW?** **The First Seconds on the Plate**
>
> In a 2005 experiment a wafer set on a bake plate climbed from 24 °C to 110 °C in 3.1 s. A thin polyimide sheet under the wafer stretched the climb to 30.5 s, and the printed lines came out wider by about 0.3 nm for each second of ramp [4].

Feedback cannot answer it. The shortfall is over in the time a filtered thermocouple takes to decide that it happened, and it has a known time of arrival and a known size. Scenario 6 announces it instead. Table 5.4 reports a sweep of the pulse amplitude at fixed length. At 26 W for 6 s the dip falls from 2.777 °C to 0.096 °C and the mean linewidth returns to 62.110 nm, an error of +0.110 nm. The plate does not stay flat to deliver it. The pulse puts 15.6 kJ into the plate (26 W on a hundred zones for 6 s) where the wafers need about 7.8 kJ, so the plate mean overshoots the setpoint by 9.290 °C and the wafer peaks at 114.11 °C. At 14 W the pulse leaves 3.592 nm on the table and at 40 W it overshoots by 4.928 nm, so the optimum is findable and it is not wide. The same values result when the amplitude is changed between wafers. The length sweep is sharper still, because at 26 W a pulse of 4 s leaves 2.490 nm and one of 8 s overshoots by 3.130 nm. Both curves are energy balances in dose rather than control problems, which is the lesson.

The part of the pulse that is not the pulse is the integral hold described in Section 5.5.3. It is switched on whenever the pulse is running and cannot be switched off in the application, so its value was measured on a copy of the controller with the hold removed. Without the hold each zone loop reads the 26 W as an error of its own making, and by the end of the pulse the center zone trim has fallen to −17.4 W against −0.8 W with the hold. The plate mean then peaks at only 113.72 °C, sags to 108.98 °C while the integrators climb back, and the wafer prints 66.450 nm, wider than the 66.015 nm of no pulse at all.

5.6.5 The Plate Is Not the Product

Scenario 7 uses wafers with $b = 0.5$ in (5.5), which touch the plate well at the center and badly at the edge. The linewidth map that results is a clean radial signature running from 62.19 nm at the wafer center to 65.95 nm at the outermost field, a range of 3.765 nm and a three-sigma spread of 3.253 nm. The plate range at the end of the bake is 0.0016 °C, which is 0.005 nm of equivalent linewidth. Every thermocouple on the tool reads its setpoint and the tool is printing three nanometers of across-wafer range.

> **TRY IT IN THE APP** **7 · A Bowed Wafer**
>
> Press Run and bake three wafers with Run-to-run zone offsets left Off. Watch the CD map build a ring pattern while the plate map stays flat. The linewidth runs from 62.19 nm at the center to 65.95 nm at the outermost field, a range of 3.765 nm, while the Plate range badge reads about 0.0016 °C at the end of each bake.

Only a measurement of the wafer can see that, which is the argument for run-to-run correction and not merely a preference for it [11, 14]. Scenario 8 turns Equation (5.16) on. The radial signature is removed in four or five wafers, with the across-wafer range falling from 3.764 nm to 1.074 nm by the fourth wafer and to a mean of 0.745 nm over wafers 10 to 20. The mean linewidth returns to 62.011 nm.

Pause. Try Exercise 5.2 before reading on.

The plate is no longer flat, and it is not supposed to be. Over the same wafers the zone offsets span 1.71 °C and the plate range at the end of a bake is 1.702 °C against the 0.002 °C of scenario 7. A temperature map read on its own would rank that plate below every healthy closed-loop case, and it prints wafers with a fifth of the range that the flat plate of scenario 7 prints.

5.6.6 Noise Filtered in Space Rather Than in Time

What also comes out of scenario 8 is roughness. Every channel of (5.16) is driven by its own independent 0.40 nm measurement draw, no two neighboring zones ever see the same number, and the offset map accumulates a grain that is a random walk with nothing to stop it. The roughness of (5.18) climbs from 0.213 after the first wafer to a mean of 0.254 over wafers 10 to 20, and over the same wafers the discrete Laplacian of the offset map has a root-mean-square of 0.1996 °C against a map root-mean-square of 0.785 °C.

The spatial bandwidth limit does not appear here as an inability. The forward model of (5.13) inverts the plate exactly and will ask for a checkerboard without complaint, and the plate will produce one. It appears as a cost, at 7.86 times the watts per degree that a smooth correction of the same amplitude would need, spent on a pattern that was measurement noise before it was a correction and that is printed on every wafer thereafter.

Scenario 9 applies the neighbor average of (5.17) before the correction reaches the setpoints. Over wafers 10 to 20 the roughness falls from 0.254 to 0.189, the Laplacian root-mean-square of the offset map from 0.1996 °C to 0.1473 °C, and the across-wafer linewidth range from 0.745 nm to 0.595 nm, with the three-sigma spread falling from 0.446 nm to 0.358 nm. The comparison holds across noise seeds. On five further seeds the mean range runs from 0.685 to 0.848 nm per zone and from 0.545 to 0.661 nm filtered, and the filtered run is lower on every seed by 0.11 to 0.24 nm. The magnitude of the correction itself is nearly untouched, at 0.785 °C and 0.778 °C root-mean-square respectively, and so is the speed of the initial convergence (1.074 nm against 1.196 nm at the fourth wafer and 0.984 nm against 0.943 nm at the fifth), because a radial signature carries little energy at the wavenumbers (5.17) removes. The filter costs nothing on the disturbance and takes a quarter out of the roughness, which is the whole argument for it. The Q-filter of the iterative learning controller in Volume 8, Chapter 10 plays the same part for a correction stored along a scan path.

The limit is equally real, and scenario 9 completes on it. When the zone 34 heater fails after the seventh wafer, no setting removes its field, because the zone is commanded at its 120 W rating and its offset reaches the 4 °C bound within one to three wafers. Six wafers later the CD range is 17.46 nm at $w = 1$, 17.60 nm at $w = 0.6$ and 18.33 nm per zone, against 24.29 nm with the run-to-run loop off. What the filter changes is where the rest of the correction lands. At $w = 1$ the three neighboring fields print 1.01 nm to 1.56 nm narrow and six fields besides the faulted one sit more than 0.5 nm from the median, against three fields, none more than 0.69 nm off, per zone. Spatial filtering trades resolution against noise in exactly the way temporal filtering trades speed against noise, and for exactly the same reason.

5.6.7 An Instrument That Lies

Scenario 10 biases the thermocouple on zone 0 by $+0.5\,^{\circ}\mathrm{C}$ at $t = 200\,\mathrm{s}$. The loop reduces the power of that zone at the end of a bake from 34.71 W to 34.03 W until the plate really is half a degree cold at 109.501 °C, and then reports that it is exactly on setpoint, a measurement that is wrong in the sense of Section 1.4.2. The linewidth on that field moves from 62.281 nm to 63.330 nm and the across-wafer range from 0.266 nm to 1.298 nm. The 1.050 nm of linewidth is smaller than the 1.598 nm that (5.8) predicts for half a degree, because the silicon above a single cold zone is being warmed from the side through k_w, which is the same mechanism that spread the heater fault. Turning the run-to-run loop on after the third wafer removes the error, because it measures the product rather than the plate, and from the eighth wafer on the field stays within 0.26 nm of the median. What it leaves behind is a permanent zone offset, 0.51 °C to 0.67 °C on zone 0 from the seventh wafer to the twenty-first, that is a record of an instrument fault nobody has diagnosed.

5.6.8 The Graded Challenge

The challenge presents four disturbances at once. The wafers are bowed at $b = 0.45$ and arrive at 19 °C, the forward model is switched off, the pulse is off, the run-to-run loop is off and the loops are slow at $K_c = 2\,\mathrm{W/^{\circ}C}$ and $\tau_I = 40\,\mathrm{s}$, with thermocouple noise of 0.02 °C. The target has three parts over the 700 s lot of six wafers: a mean linewidth within 0.35 nm of 62 nm over the last third (wafers 4 to 6), an across-wafer range under 1.2 nm on the last wafer, and a mean zone offset within $\pm 1\,^{\circ}\mathrm{C}$ at the end. As issued the configuration prints 67.693 nm with a range of 3.120 nm, and all six wafers of the lot carry fields outside the specification. The suggested configuration turns on the model, the pulse, the run-to-run loop at $\lambda_z = 0.6$ and the spatial filter, and raises the gain to $K_c = 6\,\mathrm{W/^{\circ}C}$ with $\tau_I = 12\,\mathrm{s}$. It prints 62.009 nm with a range of 0.832 nm by the sixth wafer and a mean zone offset of 0.398 °C, the worst mean error over wafers 4 to 6 is 0.087 nm, and only the first wafer of the lot (63.20 nm) is out of specification. On eight other noise seeds the same settings pass with worst errors of 0.08 to 0.14 nm and ranges of 0.59 to 0.85 nm.

Read over the same 700 s window, the linewidth loop integrated absolute error falls from 3395.3 nm s to 222.3 nm s and the peak linewidth error from 5.998 nm to 1.198 nm. The zone-loop numbers move the other way, because the center zone integrated absolute error rises from 349.5 °C s to 862.6 °C s and its total variation from 252.3 W to 852.3 W. The configuration that prints good wafers is the configuration whose temperature scorecard looks worse, since the pulse and the per-zone offsets both move the plate on purpose. Ordering matters as well. The run-to-run loop alone ($\lambda_z = 0.6$, filtered) brings the sixth wafer to 62.040 nm with a 0.76 nm range, but only by writing the edge and the landing into the zone offsets, which end at a mean of 1.815 °C, and the offset part of the target rejects it on every seed tried. Switching the run-to-run loop on first and the pulse two wafers later counts the landing twice, so the next wafer prints 58.52 nm, and the integrators of (5.16) hand the excess back only one wafer at a time (60.52, 61.34 and 61.725 nm), too late for the last third of the lot. The pulse itself is sized for tuned loops. On the slow loops as issued it over-bakes the line to 55.27 nm, and with the model at $K_c = 6\,\mathrm{W/^{\circ}C}$, $\tau_I = 12\,\mathrm{s}$ it gives 63.2 nm, which leaves the last nanometer and the bow to the run-to-run loop.



Figure 5.4 The hundred-channel run-to-run loop of scenario 9 at $t = 21.67$ min, correcting bowed wafers with the correction spatially filtered, cropped to the plant view and the trends. The left map (1) is the plate temperature about its setpoint and the right map (2) is the linewidth the resist produced from it, drawn on the same hundred cells, with the power of every zone as a bar field underneath (3). A twelfth wafer is 70 s into its 90 s bake and the map on the right belongs to wafer 11, whose across-wafer linewidth range has fallen to 0.56 nm against a 1.20 nm allowance at a mean of 62.01 nm. The plate is deliberately 1.517 °C out of flat because that is the profile the linewidth map asked for, the mean zone offset has reached 0.652 °C and carried the rim with it to 100.0 °C, and the spatial roughness of the correction is holding at 0.188. The trends show the whole run on a 30 min axis, with one heater pulse per bake on the two zone traces (4) and a linewidth mean stepping down toward its target once per wafer. The lowest strip in frame (5) collects the across-wafer numbers on one axis, the linewidth range against its allowance, the plate range, the spread of the zone offsets and the spatial roughness. An eighth strip carrying the wafer-landing detail sits below the visible area.

✦ DID YOU KNOW? A Flat Plate and a Failed Lot

From the simulator: in Table 5.4 the challenge as issued ends with a plate range of 0.008 °C and six of six wafers out of specification. The suggested configuration ends with a plate range of 1.223 °C and one of six out. The pulse and the zone offsets move the plate on purpose.

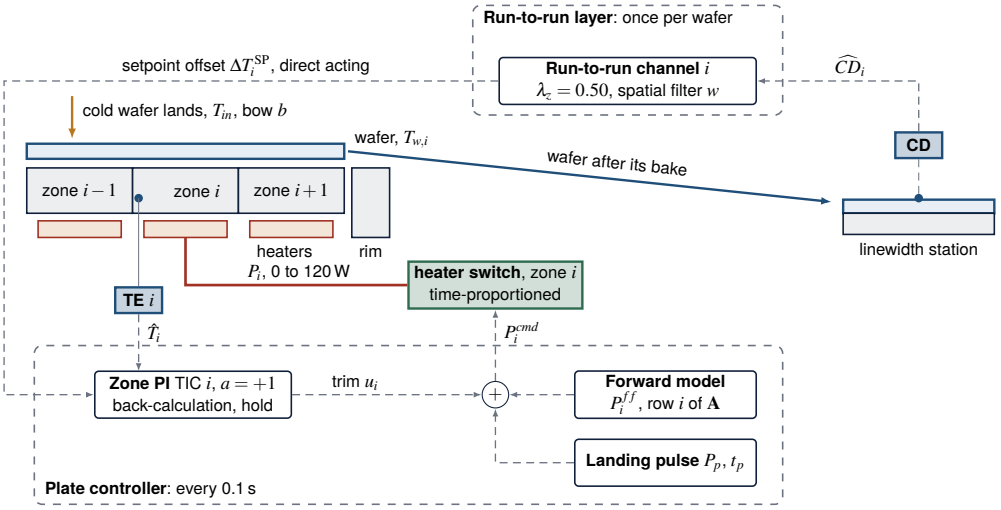


Figure 5.5 One of the hundred zone loops of the simulated plate drawn with its hardware, and the run-to-run loop above it. The thermocouple TE-i in the plate feeds the zone controller TIC-i, a velocity-form PI loop with $a = +1$ every 0.1 s. Its trim is added to the forward-model power and to the timed landing pulse, and the sum drives the heater of zone i. Once per wafer the linewidth map from the measurement station moves the zone setpoint by ΔT_i^{SP}, a direct-acting channel. Notice that no instrument of the inner loop sees the wafer.

5.7 Technology in Practice

The loops of Section 5.5 run on a hundred thermocouples, a hundred heaters, the plate's own controller and, one level up, a linewidth measurement and the run-to-run system that reads it, drawn together in Figure 5.5.

What Measures It

Each zone reads one thermocouple, which the model of Section 5.3 reduces to the plate temperature plus a noise of 0.02 °C in the graded challenge and a fixed bias on zone 0.

> **TECHNOLOGY** **Sensor: A Thermocouple in the Plate**
>
> Two wires of different metals joined at one end produce a voltage at the other end that is roughly proportional to the temperature of the junction. Wherever the wires meet the instrument a second junction forms, so every precision instrument compensates for the temperature of that reference junction. A thermocouple is usually less accurate than a resistance thermometer, but it is rugged and spans a wider range [17]. Its most common failure is to break open, called burnout. An open input picks up electrical noise that reads as a wildly varying temperature, so the instrument is set to drive its reading to the top or the bottom of the scale instead, whichever is safer. On a furnace, where too much heat is the danger, the reading is sent high and the controller then cuts the heat [17].

The model's thermocouples never burn out. They fail only by the bias of scenario 10, the fault that a loop satisfies by holding the plate wrong. The product is measured on every field of every wafer with a repeatability of 0.40 nm and no queue, the most optimistic assumption of the chapter.

What Moves It

Each heater is modeled as a power between zero and 120 W, delivered at once, scaled by the zone's heater health and the supply multiplier of Equation (5.15).

TECHNOLOGY Final Element: A Heater Switched On and Off

An electric heater is a final control element, like a control valve or a variable-speed motor. Where a heating element can only be turned fully on or fully off, the controller sends a time-proportioned output instead of a continuous one: a pulse train whose duty cycle, the share of each period spent on, runs from 0 to 100 % [17]. The switching is done without moving parts. In an output channel for alternating current the controller lights a diode inside an optocoupler, which turns on a light-sensitive semiconductor switch, a TRIAC, and power flows to the load [17]. In the heater fault of this chapter the command sits at the full rating while the zone stays cold, and the saturation flag, not the temperature map, names the zone.

The continuous power of the model stands for the average that such a pulse train delivers, and Section 5.3.6 names the switching it leaves out.

Where the Controller Runs

The zone loops run in the plate's own controller, and the run-to-run channels run above it in the fab's advanced process control system. Control loops in operational technology run with cycle times from milliseconds to minutes, many on real-time operating systems, and what counts as real time depends on the application and has to be stated for each [18]. The app executes the zone loops every 0.1 s against a physics step of 0.05 s, and the run-to-run update once per wafer, every 110 s. In the controller the structure is three kinds of block: a PI block per zone with back-calculation and the integral hold, a sum that adds the forward-model power of Equation (5.13) and the timed pulse, and on the upper layer the hundred-channel integrator with its spatial filter.

IN PRACTICE Run-to-Run Control in a Fab: Filters and Threads

In a fab, run-to-run control adjusts a recipe lot by lot or wafer by wafer, and most systems combine it with fault detection and classification under the name advanced process control. The exponentially weighted moving average is the most common filter, chosen for its simplicity and robustness. Lots are grouped into control threads by context, such as product, layer and tool, and only lots of one thread update that thread's estimate. The danger is the number of threads: a product that runs rarely leaves its thread without data while the tool drifts [19]. Each of this chapter's hundred channels is one such filter, with a weight of 0.50, on a single thread.

What Is Modernizing It

Two developments bear on this tool. The first measures the wafer's view of the plate. A wireless sensor wafer carries temperature sensors across its face and records the spatial and temporal uniformity of a bake plate through a bake. One such measurement found a laboratory plate uniform near its center with a rapid drop toward its edges, the shape of the uniform-power map of scenario 1. Such checks matter because the international roadmap asked, for the 32 nm node, for a linewidth sensitivity to bake temperature of 1 nm per degree, well below the deep ultraviolet resists of the time [4]. The resist of this chapter has 3.196 nm/K. The second is state estimation that shares information across contexts. Kalman filter methods have been proposed for linewidth control in lithography, and non-threaded estimators attribute a disturbance to the sum of tool, product and layer states with regression or a Kalman filter, at the price of an observability problem in the matrix that assigns runs to contexts [19].

5.8 Limitations

The model is not intended for plate design. The thermal network of (5.1) has no heater layer, no plate body and no proximity gap as separate capacities, so a real plate's step response is higher order than this one and the identification exercise a designer would run does not transfer. The chemistry is a single Arrhenius reaction with no acid diffusion length, so the model reproduces zone-to-zone linewidth steps that a real resist would blur, and it cannot be used to argue about how many zones are worth cutting [3]. Metrology arrives with no queue and every field is measured, which is the most

optimistic assumption in the application and makes the run-to-run loop of (5.16) converge faster than a production loop does.

One limit is worth stating because a reader may look for it and not find it. The lower bound of zero in (5.15) is documented and is never reached in any supplied scenario. Run as shipped for twenty wafers, the smallest power any healthy zone is commanded across all eleven is 28.54 W, in the per-zone run-to-run case, and the 40 W pulse that completes scenario 6 takes it to 26.21 W. Reaching the floor would take a negative offset of 19.9 °C on one zone against its neighbors, or a checkerboard of 11.2 °C, against a bound of 4 °C. The one smaller number in the collection is a delivery rather than a command, the 3.90 W that leaves the damaged heater in the tenth of a second after the fault lands and before its loop winds up. As shipped, the upper bound is reached in scenario 4 only, where the faulted zone is commanded the full 120 W rating for the rest of the run against an unclipped demand that reaches 394.9 W, and delivers 14.40 W through the damaged element. The saturation flag is raised 2.1 s after the fault and stays raised. The same heater failure fired to complete scenario 9 raises it there as well, and it is clear in every other case.

Every disturbance enters either as a commanded step or as a scheduled event, and never as the slow wander a tool shows in production. Perfect repeatability is itself a fiction, because holding the integration step and the scan period constant returns one trajectory on every machine, which no real control system does. Four properties were kept through every simplification, the spatial frequency response, the boundary, the disturbance with a known arrival time and the gap between the plate and the product, and each was kept because a control decision rests on it. Section 5.3.6 carries the condition under which each assumption fails.

Two disturbances are stylized for clarity. A heater that loses 88 % of its authority in one step is not how a cartridge heater usually fails, although it does isolate the difference between the plate signature and the wafer signature of the same fault. A thermocouple bias that appears at once at half a degree is equally abrupt, and it is there so that the moment of onset is unambiguous on the trend.

5.9 Discussion

The spectral view in Equation (5.9) turns a plate specification into arithmetic. Halving k_x by cutting deeper slots lowers $G(0,0)/G(\pi,\pi)$ from 7.86 to 4.43 and buys spatial authority, and it also makes the bare plate less uniform and moves more of the burden onto the control system. Doubling k_x takes the ratio to 14.71, which the application offers as a disturbance so that the trade can be measured rather than asserted. What the calculation shows is that the useful design variable is the ratio of coupling to loss and not the zone count, so a plate with four hundred zones and the same slots resolves a finer correction and delivers it at the same eight-to-one penalty. That conclusion generalizes to any distributed actuator array whose plant is diffusive [7], and Chapter 11 meets it in pressure, where a ragged radial target is out of reach rather than badly tuned.

The model as written also carries control methods the application does not implement. The 201-state balance with hard power limits on every zone is a natural model predictive control problem, and the literature on multizone thermal processing has addressed exactly the computational difficulty that arises when all hundred zones are solved at once [12]. A moving horizon estimator on (5.1) and (5.4) could estimate the contact map m_i of (5.5) from the plate temperatures during the first seconds of a bake, which would convert the wafer bow from an unmeasured disturbance corrected run to run into a measured one corrected within the wafer. An economic formulation could trade the power cost of a rough correction against the yield value of the linewidth range it buys, using the two gains of (5.10) as the exchange rate.

The wider consequence is that linewidth budgets have tightened faster than plates have improved [15], so a share of the uniformity budget at every advanced node is now spent on thermal control rather than on optics. A bake plate is one of the cheapest pieces of a lithography cell and one of the larger contributors to across-wafer linewidth variation, which is why zone counts rose and why

per-zone correction from a measured map became standard [10, 13]. The argument this application makes is that the next increment of performance is not bought with more zones.

OPEN CHALLENGE Randomness Below the Reach of a Heater

During the bake the photoacid diffuses while it catalyzes the deprotection. Photon shot noise, the blur of secondary electrons and that diffusion make pattern formation random at the scale of single features, roughening the lines and sometimes breaking or bridging them. One exposure now prints trillions of features, so for logic chips the failure rate must stay below a few parts per trillion. Moving away from chemically amplified resists may be required, but the metal oxide resists that would replace them raise new questions of stability and reproducibility [20]. Progress would be a resist and bake that keep their speed with fewer random failures. The plate of this chapter sets the dose of a field and nothing finer.

5.10 Conclusion

This chapter studied decentralized control of a coupled actuator array on a hundred-zone post-exposure bake plate. The simulator integrates 201 states, one temperature per zone, one per wafer element and one for the rim, with the resist linewidth taken from an Arrhenius thermal dose referenced to a calibrated bake at a sensitivity of −3.196 nm/K. The steady-state plate is a graph Laplacian plus an ambient term whose gain runs from 2.857 K/W for a uniform power pattern to 0.3636 K/W for a checkerboard, a ratio of 7.86 fixed by the conductance design rather than the zone count.

A gain tuned against a single loop overshoots 18.8 % when all hundred zones move together, because the two spatial gains differ by 4.43. Uniform power leaves the plate 5.437 °C out corner to center, and the forward conductance model holds the flat profile with the feedback switched off entirely.

A cold wafer landing costs 4.015 nm on every field, mostly in dose rather than in the plate dip, and a 26 W pulse for 6 s with the integrators held recovers all but 0.110 nm of it by overshooting the plate 9.3 °C. Bowed wafers print a 3.765 nm radial signature that a hundred run-to-run channels reduce to 0.745 nm while injecting a spatial roughness of 0.254, which a neighbor average takes to 0.189 and 0.595 nm.

A fuller model would add a separate heater layer and proximity gap, an acid diffusion length, a sampled metrology queue and a rim guard heater. The plate as it stands answers the question of what a distributed actuator can be asked to produce.

KEY POINTS

- Every zone shares heat with its neighbors, so the plate answers a change made everywhere far more strongly than a fine pattern, and more zones do not change that.
- A gain tuned on one loop overshoots 18.8 % when all hundred zones move together, because the two spatial gains differ by 4.43.
- A cold wafer landing costs 4.015 nm on every field, mostly in dose, and a timed 26 W pulse with the integrators held recovers all but 0.110 nm of it.
- A bowed wafer prints a 3.765 nm radial signature that no thermocouple sees, and a hundred run-to-run channels on the measured linewidth reduce it to 0.745 nm.

Where else this applies. A spatial bandwidth limit on an actuator array in Chapter 11 and Chapter 9, and a correction learned from late measurements and filtered in space or time in Volume 8, Chapter 10.

References

[1] E. Bristol, “On a new measure of interaction for multivariable process control,” *IEEE Transactions on Automatic Control*, vol. 11, no. 1, pp. 133–134, 1966.

[2] H. Ito and C. G. Willson, “Chemical amplification in the design of dry developing resist materials,” *Polymer Engineering & Science*, vol. 23, no. 18, pp. 1012–1018, 1983.

[3] M. D. Smith and C. A. Mack, "Examination of a simplified reaction-diffusion model for post-exposure bake of chemically amplified resists," in *Advances in Resist Technology and Processing XVIII*, ser. Proceedings of SPIE, vol. 4345. SPIE, 2001, p. 1022.

[4] J. P. Cain, P. Naulleau, and C. J. Spanos, "Critical dimension sensitivity to post-exposure bake temperature variations in EUV photoresists," in *Emerging Lithographic Technologies IX*, ser. Proceedings of SPIE, vol. 5751. SPIE, 2005, p. 1092.

[5] C. D. Schaper, M. M. Moslehi, K. C. Saraswat, and T. Kailath, "Modeling, identification, and control of rapid thermal processing systems," *Journal of The Electrochemical Society*, vol. 141, no. 11, pp. 3200–3209, 1994.

[6] C. D. Schaper, T. Kailath, and Y. J. Lee, "Decentralized control of wafer temperature for multizone rapid thermal processing systems," *IEEE Transactions on Semiconductor Manufacturing*, vol. 12, no. 2, pp. 193–199, 1999.

[7] B. Bamieh, F. Paganini, and M. A. Dahleh, "Distributed control of spatially invariant systems," *IEEE Transactions on Automatic Control*, vol. 47, no. 7, pp. 1091–1107, 2002.

[8] H. Ito, "Chemical amplification resists: Laboratory curiosity to paradigm," *Journal of Photopolymer Science and Technology*, vol. 20, no. 3, pp. 319–332, 2007.

[9] K. El-Awady, C. D. Schaper, and T. Kailath, "Programmable thermal processing module for semiconductor substrates," *IEEE Transactions on Control Systems Technology*, vol. 12, no. 4, pp. 493–509, 2004.

[10] L. Berger, "Global critical dimension uniformity improvement for mask fabrication with negative-tone chemically amplified resists by zone-controlled postexposure bake," *Journal of Micro/Nanolithography, MEMS, and MOEMS*, vol. 3, no. 2, p. 203, 2004.

[11] A. Tay, W. K. Ho, and N. Hu, "An in situ approach to real-time spatial control of steady-state wafer temperature during thermal processing in microlithography," *IEEE Transactions on Semiconductor Manufacturing*, vol. 20, no. 1, pp. 5–12, 2007.

[12] K. V. Ling, W. K. Ho, B. F. Wu, A. Lo, and H. Yan, "Multiplexed MPC for multizone thermal processing in semiconductor manufacturing," *IEEE Transactions on Control Systems Technology*, vol. 18, no. 6, pp. 1371–1380, 2010.

[13] J. Moyne, E. del Castillo, and A. M. Hurwitz, Eds., *Run-to-Run Control in Semiconductor Manufacturing*. Boca Raton, FL: CRC Press, 2018.

[14] Q. Zhang, K. Poolla, and C. J. Spanos, "One step forward from run-to-run critical dimension control: Across-wafer level critical dimension control through lithography and etch process," *Journal of Process Control*, vol. 18, no. 10, pp. 937–945, 2008.

[15] M. van den Brink, "Continued scaling in semiconductor manufacturing enabled by advances in lithography," in *2019 IEEE International Electron Devices Meeting (IEDM)*. IEEE, 2019, pp. 1.2.1–1.2.5.

[16] S. Skogestad, "Simple analytic rules for model reduction and PID controller tuning," *Journal of Process Control*, vol. 13, no. 4, pp. 291–309, 2003.

[17] T. R. Kuphaldt, *Lessons in Industrial Instrumentation*, version 2.33 ed. Open textbook, Creative Commons Attribution 4.0 license, ibiblio.org/kuphaldt/socratic/sinst, 2024.

[18] K. Stouffer, M. Pease, C. Tang, T. Zimmerman, V. Pillitteri, S. Lightman, A. Hahn, S. Saravia, A. Sherule, and M. Thompson, "Guide to operational technology (OT) security," National Institute of Standards and Technology, Gaithersburg, MD, Special Publication 800-82r3, 2023.

[19] A. V. Prabhu, "Performance monitoring of run-to-run control systems used in semiconductor manufacturing," Ph.D. dissertation, The University of Texas at Austin, 2008. [Online]. Available: http://hdl.handle.net/2152/17721

[20] D. Kazazis, J. Garcia Santaclara, J. van Schoot, I. Mochi, and Y. Ekinci, "Extreme ultraviolet lithography," *Nature Reviews Methods Primers*, vol. 4, p. 84, 2024.

5.11 Exercises

5.1 ⌕ Concept check · 2 min

Which change to plate BP-301 lowers the ratio $G(0,0)/G(\pi,\pi)$ of its uniform-pattern gain to its checkerboard gain?

(a) Halving the zone-to-zone conductance k_x by cutting deeper slots

(b) Cutting four hundred zones instead of a hundred, with the same slots

(c) Raising the loop gain K_c of all hundred zone loops together

(d) Doubling the heat capacity C_z of every zone of the plate

5.2 Concept check · 1 min

In scenario 7 every thermocouple reads its setpoint, yet the wafers print a 3.765 nm radial range. Why can no zone loop correct it?

(a) The zone loops are too slow to follow the radial signature
(b) The bow changes only the plate-to-wafer contact, which no plate sensor sees
(c) Thermocouple noise is larger than the temperature error at the edge
(d) The edge heaters are already at their power limit during the bake

5.3 Concept check · 2 min

The zone loops stop integrating while the wafer-drop pulse runs. What would happen if they kept integrating through the pulse?

(a) The plate would overshoot further and the line would print narrower
(b) Nothing measurable, because the pulse is shorter than τ_I
(c) The integrators would wind down and sag the plate after the pulse, printing wider
(d) The trims would saturate at the heater rating and trip the flag

5.4 Theory and calculation · 6 min

Use the constants of Table 5.2 and Equation (5.9).

(a) Compute the plate gain for the uniform pattern, the striped pattern $\mathbf{q} = (\pi, 0)$ and the checkerboard $\mathbf{q} = (\pi, \pi)$, to four significant figures. *Checkpoint:* 2.857 K/W, 0.6452 K/W and 0.3636 K/W.

(b) The disturbance Zone-to-zone conduction multiplies k_x. Find the uniform-to-checkerboard gain ratio at 1, 2 and 0.5 times the nominal coupling. *Checkpoint:* 7.857, 14.71 and 4.429.

(c) Find the heater power per zone that holds a 1 °C checkerboard offset and a 1 °C uniform offset, and the ratio of the two. *Checkpoint:* 2.750 W and 0.350 W, a ratio of 7.86.

5.5 Theory and calculation · 7 min

The plate is held flat at 110 °C with no wafer, ambient 22 °C. Use Table 5.2 and the lattice counts of Section 5.3 (69 zones with four neighbors, 18 with three, 13 with two, 44 rim links in all).

(a) Find the rim temperature that sets Equation (5.2) to zero. *Checkpoint:* 98.42 °C.

(b) With Equation (5.13), find the power a four-, a three- and a two-neighbor zone needs, and the total for the plate. How much more is that than the uniform 30.80 W of scenario 1 delivers? *Checkpoint:* 30.800 W, 32.537 W, 34.274 W; 3156.4 W, 76.4 W more than 3080 W.

(c) Find the rim time constant $C_{rim}/(U_{rim} + \sum_i n_i k_r)$ and the zone time constant $C_z/(U_a + 4k_x)$, and their ratio. *Checkpoint:* 31.6 s and 5.16 s, about six.

5.6 Theory and calculation · 7 min

The wafer settles below the plate during a bake, and the linewidth is read against the calibrated dose $\Theta_0 = 0.917514$ of Equation (5.7).

(a) Evaluate Equation (5.8) at 110 °C, and the linewidth worth of a tenth of a degree. *Checkpoint:* −3.196 nm/K and 0.32 nm.

(b) Find the steady gap $U_w(T_i - T_a)/(h_w + U_w)$ between a plate at 110 °C and the wafer on it, to three decimals. *Checkpoint:* 1.483 °C.

(c) A wafer held isothermally at $110 - 1.483$ degrees for the whole bake has a dose equal to its Arrhenius rate. Find its linewidth to 0.01 nm, and the isothermal temperature that prints exactly 62.00 nm. *Checkpoint:* 63.39 nm; 108.95 °C.

5.7 App challenge · 9 min

Load scenario 6, "The feedforward pulse" (default seed), and run it as loaded at the default 2 s/s until the clock passes 110 s, when the first wafer finishes. Export the CSV (its time column is Time (min)). Then press Reset, set Wafer-drop pulse to 14 W, and run to the same time. Repeat once more from Reset with 40 W.

(a) Read the Mean CD and CD range badges after each run, to 0.01 nm. *Checkpoint:* 62.11 nm and 0.27 nm; 65.59 nm and 0.02 nm; 57.07 nm and 0.80 nm.

(b) In the first file, find the highest value of `plate_mean_C` during the bake and the overshoot above setpoint, to 0.01 °C. *Checkpoint:* 119.29 °C, an overshoot of 9.29 °C.

(c) The shipped pulse holds the plate far from flat. Explain why the setting that prints 62 nm is not the one that flattens the plate dip, and what the pulse is sized on. *Checkpoint:* The pulse is sized on dose, because the dip happens while the wafer is cold and costs little dose.

5.8 App challenge · 8 min

Load scenario 3, "Open loop, and the edge" (default seed). Run at 20 s/s until the clock passes 10 min, then Pause.

(a) Read the Plate range badge and convert it to linewidth with Equation (5.8), to 0.1 nm. *Checkpoint:* 5.437 °C, 17.4 nm.

(b) Set Zone control to 100 PI loops, press Run, and let it run 3 min more. Export the CSV. Find the time from the switch (the first row whose `arch` reads `pi`) to the first row with `plate_range_C` below 0.05 °C, to 0.3 s. Read the Edge zone heater value at the end. *Checkpoint:* 44.7 s; about 34.3 W.

(c) Reload the scenario with the reload button beside the scenario list, repeat (b) with Zone control on Model + PI, and explain why the model is not the faster of the two. *Checkpoint:* 52.8 s; the bumpless switch hands the model's edge power to the trims, which release it only as fast as they integrate.

5.9 App challenge · 8 min

Load scenario 8, "Run-to-run, one channel per zone" (default seed), and run it as loaded at 20 s/s until the clock passes 13 min, when seven wafers are done. Export the CSV. Then load scenario 9, "The frequencies the plate cannot produce", and do the same.

(a) Read the CD range badge of each run, and the last value of `corr_checker` (the spatial roughness c of Equation (5.18)) in each file, to ± 0.01. *Checkpoint:* Scenario 8: 0.87 nm and 0.22. Scenario 9: 0.76 nm and 0.16.

(b) Explain why the spatial filter lowers both numbers without slowing the removal of the radial bow signature. *Checkpoint:* The bow signature is smooth and carries little energy at the wavenumbers the filter removes, while the measurement noise is independent from field to field.

5.10 Application area · 3 min

A production version of BP-301 measures linewidth on nine fields per wafer, and each map reaches the run-to-run loop two wafers late ($m = 2$).

(a) What do most of the hundred channels of Equation (5.16) now have to work from, and what does that do to the case for the spatial filter? *Checkpoint:* No measurement of their own: the correction must be spread from nine sites, so a spatial filter or a smooth basis is needed.

(b) With the gain matched, each channel is the newest-reading loop of Equation (1.6) with $K = \lambda_z$. Is $\lambda_z = 0.5$ inside the stability boundary of Equation (1.7) for $m = 2$? Is the challenge's 0.6? *Checkpoint:* $K_{\max} = 0.618$: 0.5 is stable, 0.6 only just.

5.11 Application area · 3 min

In scenario 4 the zone 34 heater keeps 12 % of its authority from $t = 120\,\mathrm{s}$.

(a) Why does the plate map show one cold zone while the linewidth map shows a defect five fields wide? *Checkpoint:* The plate neighbors are regulated; the wafer above them is not, and conducts sideways through k_w.

(b) Which logged signal should a maintenance system alarm on to find the heater, and why not the CD range? *Checkpoint:* The saturation flag with the zone command at its rating; the CD map mixes the fault with every other signature.

5.12 Review the agent · 5 min

An AI assistant tunes BP-301 and writes: “A step test on one zone, with its neighbors held by their own loops, gives 0.645 K/W and 5.16 s. SIMC with a closed loop as fast as the plant gives $K_c = 1.55\,\mathrm{W/°C}$ and $\tau_I = 5.16\,\mathrm{s}$. The relative gain is 1.24, so the loops barely interact, and a plate-wide setpoint change will respond like the single-zone test, without overshoot.” Find the flaw, and say how you would check it in the app.