INTERACTIVE CONTROL ENGINEERING
Principles and Applications

Volume 14 Control in Semiconductor Manufacturing

CHAPTER 7

A Plasma Etch Simulator for Recipe Trades and Endpoint Detection

Pages 163–187

Edited by John D. Hedengren

How to cite this chapter. "A Plasma Etch Simulator for Recipe Trades and Endpoint Detection," in *Control in Semiconductor Manufacturing*, John D. Hedengren, Ed., vol. 14 of *Interactive Control Engineering: Principles and Applications*. Advanced Process Solutions, LLC, 2026, ch. 7, pp. 163–187.

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

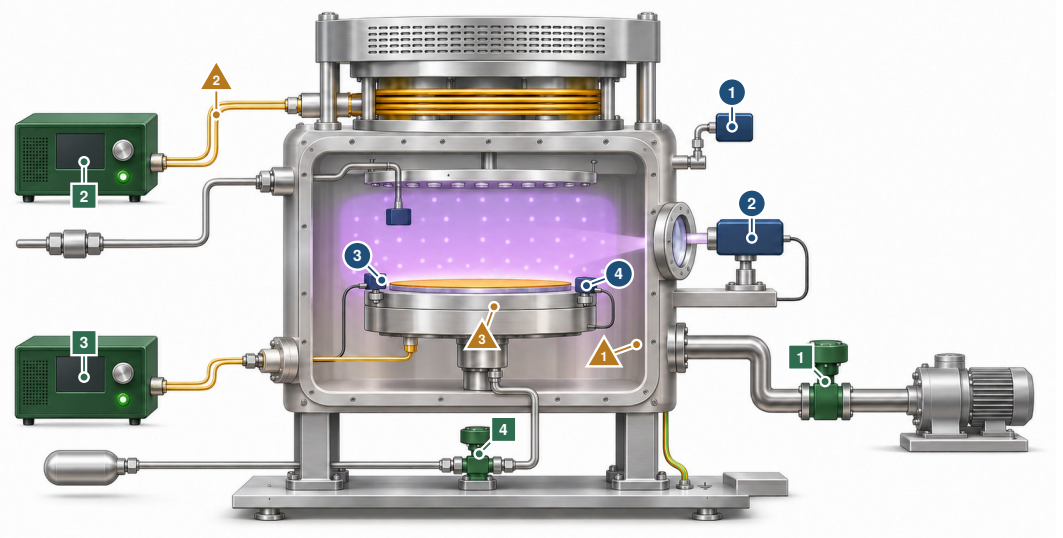
The app for this chapter
https://apmonitor.com/apps/plasma-etch/

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

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

CHAPTER 7

A Plasma Etch Simulator for Recipe Trades and Endpoint Detection



The controller moves 1 throttle valve on the pump line 2 source radio-frequency generator, wired to the coil 3 bias radio-frequency generator, wired to the chuck 4 backside helium valve

The controller reads 1 chamber pressure gauge 2 optical emission head at the window: density and endpoint 3 self-bias pickup at the chuck 4 wafer temperature probe

What disturbs it 1 polymer film building on the walls after a wet clean 2 drift of the radio-frequency match 3 leaking chuck helium seal

Notice the bare chamber walls: after a wet clean they slow the etch of the first wafers while the plasma loops hold their setpoints, and only the optical head, which watches the glow change as the film clears, knows when to stop. Illustration, not to scale.

IN PLAIN TERMS **Knowing When to Stop Etching**

The main idea. Chips get their tiny patterns by etching. A glowing gas eats away a thin film wherever a protective mask leaves it bare, and four automatic controls keep the gas, the glow and the wafer temperature steady.

The challenge. After the chamber is cleaned, its bare walls slow the etching, and none of the four controls can see that. An etch stopped by a timer then leaves the film only partly removed on the first wafers.

Why it matters. A light sensor that watches the glow change as the film runs out knows when to stop, and it saves wafers that better tuning cannot.

THREE WAYS TO READ THIS CHAPTER

Explore	12 min	The first page, the fact box of Section 7.1, the history box of Section 7.2, the try-it boxes and the fact box of Section 7.6, and the key points
Study	50 min	The whole chapter with its pause points and the twelve exercises
Practice	20 min	The first page, Section 7.6, Section 7.7, Section 7.8 and the open challenge of Section 7.9

Table 7.1 Measured open-loop step response of each variable to its own actuator from the design point, with the four loops out of service, for steps of 5 W of bias, 10 W of source power, 0.3 percentage points of throttle and 0.1 Torr of helium, timed from the controller scan that applies the step. The main etch step lasts 86 s.

Actuator	Variable	τ_{63} (s)	τ_{95} (s)
Bias generator	self-bias V_{dc}	0.063	0.147
Source generator	plasma density n_e	0.104	0.285
Throttle valve	pressure P	0.480	1.115
Backside helium	wafer temperature	15.3	45.0

The control question The four loops of an etch chamber settle in a small fraction of the time a wafer spends in it and barely interact. What then decides whether a cassette survives a chamber clean: the loop tuning, the recipe setpoints, a model predictive controller, or a detector that decides when to stop?

Learning objectives

- Justify the pairing of the four chamber loops from the measured gain matrix and its relative gain array, and name the couplings the array cannot see.
- Explain from the rate, sputter and uniformity laws why ion energy and pressure trade etch rate, mask selectivity, uniformity and sidewall angle.
- Compute the over-etch a wafer needs from the clearing transition and the endpoint threshold.
- Explain why a timed recipe loses wafers after a wet clean while the loops hold their setpoints, and why an emission endpoint recovers them.
- Attribute the measured gain of a predictive controller over four loops to tuning or to its interaction terms.

Prerequisites Sections 1.4.2, 1.4.6 (endpoint detection), 1.6.2 (velocity form and loop action), 1.6.3 and 1.6.4; Chapter 6 for the etcher inside a run-to-run loop.

The app https://apmonitor.com/apps/plasma-etch/

See also Volume 6, Chapter 2 for pairing on a triangular gain matrix and Volume 6, Chapter 8 for a predictive controller above a decentralized structure. Volume 8, Chapter 7 for a phase sequencer above two loops, and Volume 8, Chapter 8 for a batch step ended on an estimate of the product.

7.1 Introduction

An inductively coupled plasma etch chamber has four measurements a control system can close on and four actuators to close them with. Chamber pressure goes to a throttle valve, plasma density to the source generator, self-bias voltage to the bias generator, and wafer temperature to the backside helium pressure. All four are measured directly, all four have dedicated actuators, and none of them carries meaningful dead time.

Table 7.1 shows that the slowest loop on the chamber settles in about a third of the time a wafer spends in it, and the fastest settles in under two tenths of a second. From the wafer's point of view the chamber is at steady state throughout. The dynamic control problem, which occupies most of a process control course, is largely solved by the hardware.

What is not solved is which four setpoints to choose. Each of them moves several of the properties the wafer is graded on, the properties trade against each other, and none of the trades has an interior optimum in more than one variable at a time. The emphasis of this application is therefore the reverse of the usual one. The loops are a prerequisite and the recipe is the controller.

The second reversal concerns what counts as a control element. A cassette of wafers is a batch sequence rather than a regulator. Section 7.6.5 measures a case in which loops holding pressure, density and self-bias to four decimal places scrap the first three wafers after a chamber clean, and Section 7.6.6 shows that the remedy is a detector that decides when to stop.

✦ DID YOU KNOW? A Clean Every Thousand Wafers

In a production trench etch studied in 2011, each chamber was stopped for maintenance after about 1000 wafers: its inner walls were cleaned and its ceramic parts and electrodes replaced. Each time the tool was out of service for up to fourteen hours [1].

If the loops are fast and nearly decoupled, a controller that plans all four actuators at once should add little, and the application carries one so that the claim is tested rather than assumed. Section 7.6.7 sets a four-input, four-output model predictive controller against the shipped loops and against the same loops retuned once, and assigns each measured difference either to a term of the controller's model or to tuning.

The chamber teaches a specific set of process dynamics and control principles. Loop pairing on an interacting plant is decided from a measured steady-state gain matrix and its relative gain array rather than from intuition, and the pairing that results is defensible only because the time-scale separation between the loops and the batch is large. Time-scale separation itself is the second principle, because a regulatory layer that is two orders of magnitude faster than the product converts a dynamic control problem into a static optimization problem. Feedforward recipe execution is distinguished from feedback regulation by a single measured disturbance, an unmeasured drift in an unmeasured state motivates inferential and batch-to-batch correction, and a discrete sequencer with an event-driven transition sits above the continuous loops as a supervisory layer. The physical fundamentals are a material balance on the chamber under a pumping speed set by the throttle, an energy balance on the wafer whose only cooling path is the backside helium conductance, and a surface reaction rate that requires ion flux and neutral coverage together. Those are the same conservation laws and the same loop structures that govern the rest of a process dynamics and control course [2–4], applied to a plant whose product is graded once and cannot be reworked.

The behavior matters industrially because an etch chamber processes a wafer whose accumulated value at that step is large, the pattern is carried in a resist that is consumed by the process itself, and a scrapped wafer is scrapped outright. This chapter describes the simulator in enough detail to rebuild it and measures what each control element is worth on a graded cassette where several of them are worth nothing. Section 7.2 reviews the ion-assisted etch mechanism, endpoint detection and the two disturbances the chamber instruments cannot see. Section 7.3 justifies the loop pairing from the measured gain matrix. Section 7.4 gives the chamber model, and Section 7.5 the four loops, the wafer sequencer, the endpoint detector and the predictive controller. Section 7.6 measures the recipe trades, the stabilization step, open-loop and closed-loop operation, chamber seasoning, over-etch, the graded cassette and the predictive controller against the loops. Section 7.7 describes the real sensors, final elements and controller, Section 7.8 states what the model leaves out, and Sections 7.9 and 7.10 draw the lessons.

7.2 Background

Plasma etching became the dominant pattern-transfer step because it is anisotropic where a wet etch is not. How its ion-assisted mechanism was found, and how the end of an etch came to be detected, is what the History box below tells. The synergy of ions and neutrals is what makes a directional ion flux able to select where a chemical etch proceeds, and it is the term that carries the etch rate in the model below. The discharge physics behind the ion flux, the sheath and the self-bias that accelerates ions into the wafer is standard material [5].

Knowing when to stop is a separate problem and it was solved separately. The optical emission of the discharge changes as the film being etched is consumed and the etchant is no longer being used. Endpoint detection is unusual among industrial measurements because it observes the product rather than the equipment, which is what makes it robust to disturbances the chamber instruments cannot see. Section 1.4.6 sets out the timed recipe against the endpoint on the running example of this volume.

Table 7.2 Measured steady-state gain matrix at the design point, all loops out of service, from a central difference over a 0.5 % perturbation of each actuator. The source column moves all four rows and the throttle column three.

	Source (1/W)	Bias (1/W)	Valve (1/%)	Helium (1/Torr)
n_e (1)	8.94×10^{-4}	0	-2.68×10^{-2}	0
V_{dc} (V)	-5.37×10^{-2}	4.00×10^{-1}	1.61	0
P (mTorr)	3.48×10^{-3}	0	−1.53	0
T_w (°C)	1.82×10^{-2}	1.21×10^{-1}	0	−3.50

HISTORY Learning to Aim and to Stop (1979 and 1980)

The first etching processes used liquids, which etch at the same rate in every direction and so limit how small a feature can be [1]. In 1979 an experiment exposed silicon to xenon difluoride gas and to argon ions, separately and together, and found the etch rate with both greater than the sum of the two acting alone. That ion-assisted chemistry goes with the directional profiles of plasma etching [1, 6]. Knowing when to stop followed at once. In 1979 the impedance of an aluminum etch discharge was shown to change at the end of the etch [1], and in 1980 a study measured the lower limit of detection of an optical emission endpoint for polycrystalline silicon and silicon nitride films [7]. Actinometry, which compares an emission line with that of an inert gas added in known amount, was described the same year [1]. Detectors that read thousands of wavelengths at once came decades later [8].

Two of those disturbances appear in this application. The interior of the chamber accumulates a fluorocarbon film that changes the radical recombination rate at the wall, so the first wafers after a clean are etched at a different rate from the wafers that follow at identical actuator positions [9], and the across-wafer distribution of etch rate in a high-density reactor is set by a competition between ion focusing at the edge and radical depletion at the center [10]. Production tools answer both with layers above the regulatory loops, chamber condition models that estimate what the instruments do not measure [11] and run-to-run control that adjusts the recipe between wafers on measured results [12], the subject of Section 1.4.3.

7.3 Interaction Analysis

The chamber is heavily cross-coupled in the open loop, and the natural pairing has to be justified rather than assumed.

Table 7.2 contains three physical couplings worth naming. The source generator raises the pressure, because dissociation turns each admitted molecule into more particles than it started with. The source generator *lowers* the self-bias, because a denser plasma supplies more ions to carry the same bias current at a lower sheath voltage, so turning the source up softens the ion bombardment. And both generators heat the wafer, which the helium loop then has to remove.

Despite that, the relative gain array for the natural pairing is

$$\Lambda = K \odot \left(K^{-1}\right)^{T} = \begin{bmatrix} 1.073 & 0 & -0.073 & 0 \\ 0 & 1.000 & 0 & 0 \\ -0.073 & 0 & 1.073 & 0 \\ 0 & 0 & 0 & 1.000 \end{bmatrix} \tag{7.1}$$

which is within 0.073 of the identity. The pairing is correct, the only real interaction is a mild one between the pressure and density loops through the source generator's contribution to throughput, and the couplings in the V_{dc} and T_w rows are one-directional and therefore invisible to (7.1). Volume 6, Chapter 2 meets the same blind spot in a reverse osmosis train, whose relative gain reads one while

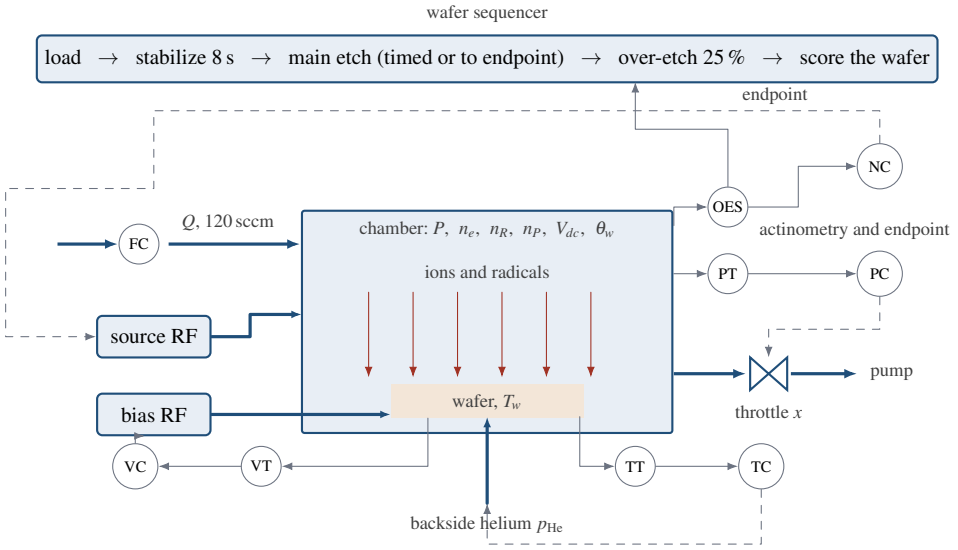


Figure 7.1 The simulated chamber. Four independent single-loop controllers hold pressure on the throttle valve (PC), plasma density on the source generator (NC), self-bias on the bias generator (VC) and wafer temperature on the backside helium (TC). Above them a discrete sequencer runs each wafer through load, stabilize, main etch and over-etch, taking the transition out of the main etch either from a timer or from the optical emission signal.

its loops interact through a triangular gain matrix. Combined with Table 7.1, this establishes that four independent single-loop controllers are the correct structure and that tuning them is a short task rather than a research problem. The remainder of this chapter is about what the setpoints those controllers hold are worth, except for Section 7.6.7, which measures what a controller that does see the one-directional terms gains from them.

7.4 Chamber Model

The chamber, its four loops and the sequencer above them are shown in Figure 7.1. The model carries pressure, plasma density, two neutral species, self-bias, wafer temperature, a chamber wall polymer coverage, three wafer depths and five actuator lags.

7.4.1 Working Units and the Design Point

The clock is in seconds, pressure in mTorr, flow in sccm and etch rate in nm/s. At design the source generator delivers 1200 W, the bias generator 150 W, the throttle valve sits at 28.71 % and the backside helium at 8 Torr. That gives 20.0 mTorr, a normalized plasma density of 1.000, 120 V of self-bias and 60 °C at the wafer, which etch the film at 5.831 nm/s with a mask selectivity of 12.00, an underlayer selectivity of 25.00, 2.200 % non-uniformity and an 88.500° sidewall.

7.4.2 Ion-Assisted Chemistry

The etch rate is the sum of an ion-assisted term and a spontaneous chemical term, inhibited by the passivating species and scaled by the pattern open area,

$$R = \left[k_{IA}\, n_e \sqrt{E_i - E_{\text{th}}}\; \theta_R + k_{CH}\, n_R\, \mathrm{e}^{-E_a/RT_w} \right] \times (1 - k_i \theta_P)\; \phi(A) \quad (7.2)$$

with θ_R and θ_P Langmuir coverages of the etchant and passivant. The ion-assisted term is the mechanism that makes anisotropic etching possible at all, and the synergy it represents was established by Coburn and Winters [6]: ions alone and neutrals alone each etch slowly and together they etch fast.

The mask and the underlayer are attacked by the same ions with different exponents,

$$R_{\text{mask}} = k_m\, n_e\, (E_i - E_m)^{0.85}\, \phi(A), \tag{7.3}$$

$$R_{\text{under}} = k_u\, n_e \sqrt{E_i - E_u}\, (1 - k_p \theta_P)\, \phi(A) \tag{7.4}$$

and the difference between those exponents and the square root in (7.2) is what makes selectivity fall as ion energy rises. The loading factor $\phi(A)$ is common to all three rates, so it cancels in both selectivities. Ion energy is set by the self-bias plus a plasma potential, and the self-bias follows (7.5).

$$V_{dc} = k_V \sqrt{P_{\text{bias}}/n_e} \tag{7.5}$$

A denser plasma supplies more ions to carry the same bias current, so the same bias power develops less voltage, which is where the negative source-to-V_{dc} entry in Table 7.2 comes from [5].

7.4.3 Uniformity and Sidewall Angle

Across-wafer uniformity is the difference of two competing terms,

$$U = U_0 + h_U \left| k_e \left(\frac{P_0}{P} \right)^{0.8} - k_c \left(\frac{P}{P_0} \right)^{0.7} \frac{R}{R_0} \frac{Q_0}{Q} \right| \tag{7.6}$$

with an edge-fast term that dominates at low pressure, where ions focus at the wafer edge, and a center-slow term that dominates at high pressure, where radicals are consumed before they reach the middle. The two cancel at one pressure, which is the interior optimum, and the competition it stands in for is documented in high-density reactors [10]. The reference rate R_0 in (7.6) is the design value of 5.831 nm/s, P_0 and Q_0 are the design pressure and flow, and h_U is a hardware uniformity factor equal to one except in the graded challenge.

Sidewall angle is not optimized at the same place,

$$\alpha = 90 - k_\alpha \left(\frac{P}{P_0} \right)^{0.6} \left(\frac{135}{E_i} \right)^{n_a} \frac{1}{0.5 + \theta_P} \tag{7.7}$$

because pressure scatters the ions and tapers the profile monotonically. Section 7.6.2 measures the resulting conflict.

7.4.4 Wall Polymer Coverage

A single state carries the polymer coverage on the chamber walls,

$$\dot{\theta}_w = \mathbb{1}_{\text{plasma}}\, k_w \left[n_P (1 - \theta_w) - n_R\, \theta_w \right] \tag{7.8}$$

deposited by the passivant and stripped by the etchant, and only while the plasma is lit. It feeds back into the radical balance by shielding the walls, so a chamber that has run wafers all morning has a different radical density from one that has just been cleaned, at identical actuator positions. That is the first-wafer effect [9], and it is the subject of Section 7.6.5.

7.4.5 The Clearing Transition

The film does not clear everywhere at once, because the etch rate is not the same everywhere. The application makes the cleared fraction a logistic in depth whose width is set by the non-uniformity,

$$c(t) = \left[1 + \exp\!\left(- \frac{z(t) - t_{\text{film}}}{\max(0.8,\ 0.60\, U\, t_{\text{film}}/100)} \right) \right]^{-1} \tag{7.9}$$

and the underlayer is only attacked in proportion to c. This one equation carries the whole over-etch argument. A poorly uniform wafer needs more over-etch to clear everywhere, and every second of over-etch is spent cutting into the underlayer wherever it has already cleared.

7.4.6 Initialization and Assumptions

The chamber steady state is algebraic once the pressure is known, and the pressure needs one short iteration because the plasma the source makes changes the throughput. Two inverse calculators return the throttle position for a chosen pressure and the helium pressure for a chosen wafer temperature.

The plasma is a global model with two lumped neutral species and no spatial structure, a normalized electron density rather than an absolute one, and no sheath model beyond the self-bias relation. Integration is fixed-step RK4 at 0.005 s with a 20 ms controller scan, on the engine of Section 1.6.1.

7.5 Controller Design

The control system has three layers, four continuous loops, a discrete wafer sequencer above them, and an endpoint detector that ends one phase of the sequence. A model predictive controller can take the place of the four loops in the regulatory layer, and Section 7.5.4 documents it.

7.5.1 Velocity Form Chamber Loops

Each loop computes an increment to its actuator every control interval, the velocity form of the PID law that Section 1.6.2 sets out [2, 13]. The derivative acts on the measurement rather than on the error, and the four chamber loops carry no derivative term, so the implemented controllers are proportional-integral. Each loop closes on a first-order filtered measurement, with time constants of 0.05 s on pressure and density, 0.02 s on self-bias and 0.30 s on wafer temperature. The error and the three increments are given by (7.10) and (7.11).

$$e_k = a\,(\mathrm{SP}_k - \mathrm{PV}_k) \tag{7.10}$$

$$\begin{aligned} \Delta u_k^P &= K_c\,(e_k - e_{k-1}) \\ \Delta u_k^I &= \frac{K_c}{\tau_I}\,e_k\,\Delta t_c \\ \Delta u_k^D &= -\,K_c\,\tau_D\,a\,\frac{\mathrm{PV}_k - 2\mathrm{PV}_{k-1} + \mathrm{PV}_{k-2}}{\Delta t_c} \end{aligned} \tag{7.11}$$

where $a = +1$ on the density and self-bias loops and $a = -1$ on the pressure and wafer-temperature loops, whose process gains are negative, $\tau_D = 0$, and $\Delta t_c = 20\,\mathrm{ms}$. In the convention of Section 1.6.2 the density and self-bias loops are therefore reverse acting and the pressure and wafer-temperature loops direct acting, which is also what the app's tuning notes call the last two. The increments are combined and the demand is clipped to the actuator range, as shown in (7.12).

$$\begin{aligned} s_k &= s_{k-1} + \Delta u_k^P + \Delta u_k^D \\ c_k &= s_k + I_{k-1} \\ u_k &= \mathrm{clamp}\,(c_k,\ u_{\min},\ u_{\max}) \\ I_k &= I_{k-1} + \Delta u_k^I + \frac{\Delta t_c}{T_t}\,(u_k - c_k) \end{aligned} \tag{7.12}$$

Here s_k, c_k and I_k are the running sum S_k, the output before limiting v_k and the integral I_k of Equations (1.13) to (1.15), with no output bias and a tracking time $T_t = \tau_I$. The actuator ranges are 4 to 100 % of throttle, 0 to 2600 W of source power, 0 to 700 W of bias and 0.5 to 25 Torr of helium. Anti-windup is back-calculation on the integral alone, applied only when the integral gain is nonzero, and the running sum s_k is never truncated. In manual operation and under the open-loop recipe I_k

is set each scan so that $s_k + I_k$ equals the actuator position in force, so the transfer to closed loop is bumpless with no additional logic.

During each wafer exchange the sequencer switches both generators off. With no plasma the measured density and self-bias collapse, so in closed loop the density and self-bias loops hold their last commands and set I_k each scan as in manual operation, and every strike begins from the source and bias power in force before the exchange. The pressure and temperature loops keep regulating, because the throttle and the helium remain in service.

The tuned set used in every closed-loop case has gains of 0.50 %/mTorr on pressure, 600 W per unit of normalized density, 1.20 W/V on self-bias and 0.45 Torr/°C on wafer temperature, with integral times of 0.40 s, 0.20 s, 0.15 s and 15 s. The pressure and temperature integral times are close to the open-loop time constants of Table 7.1, the choice the SIMC rules make for a lag-dominant loop [14], while the density and self-bias integral times are two to two and a half times their τ_{63}. The application also offers an open-loop mode in which the recipe commands generator power and valve position directly, which is how a large fraction of production recipes run, and Section 7.6.4 measures the difference.

7.5.2 The Sequencer

Each wafer runs through four phases. Load takes 6 s with the plasma off, and the new wafer arrives at 23 °C. Stabilize takes 8 s with the plasma lit and the wafer shielded, so the strike transient happens before the product is exposed. Main etch runs either for a fixed time or until the endpoint detector fires, subject to a 150 s backstop. Over-etch runs for a set percentage of the main etch time. The wafer is then scored against seven independent criteria with no partial credit, namely cleared to at least 0.995 of its area, mask remaining above 60 nm, underlayer loss below 12 nm, non-uniformity below 4 %, sidewall not tapered below 87°, sidewall not undercut above 90.2°, and resist never above 95 °C. Volume 8, Chapter 7 runs the same kind of phase sequencer above two loops inside every shot of a molding machine.

7.5.3 The Endpoint Detector

The endpoint detector watches an actinometric emission signal that falls as the film clears and the etch product stops being generated [7]. The modeled signal is $(1 - c) + \varepsilon$ with c the cleared fraction of (7.9) and $\varepsilon = 0.12$ the residual emission after the film is gone, passed through a first-order filter of 0.30 s and compared against a fixed threshold of 0.55. The threshold corresponds to 57 % of the wafer cleared, and the filter lag moves the firing on a design-point wafer to 64 % cleared. Thresholding a filtered emission trace is the classical implementation and remains the production standard even as spectral methods improve [8]. It is not a feedback loop. It has no setpoint, no actuator and no error signal, and Section 7.6.6 measures it as the most valuable control element on the tool.

🕮 **Pause.** Try Exercise 7.3 before reading on.

7.5.4 Model Predictive Control of the Four Actuators

The fourth chamber-control structure replaces the four loops by one receding-horizon plan for the throttle, the two generators and the backside helium, the multivariable form that dominates industrial practice [15, 16]. Section 1.6.3 gives the general form of the option, and this section gives what is particular to the chamber. Real-time feedback control of a reactive ion etcher on its measured plasma variables is documented in [17]. The sequencer, the endpoint detector, the recipe and every actuator limit are unchanged. While the plan acts, the four PI loops set I_k each scan so that their output equals the applied command, as in manual operation, which makes a hand-over in either direction bumpless.

The controller's model is identified rather than derived. With the plasma held, the sensor noise off and the actuators at the design point, each actuator is stepped by 5 % of its design value in each direction (1.4 percentage points of throttle, 60 W of source, 7.5 W of bias and 0.4 Torr of helium), the two responses are averaged, and every input-output pair is fitted with a second-order lag and a

Table 7.3 Identified elements of the predictive controller's model, from ±5 % steps of each actuator about the design point with the plasma held. Residual is the root-mean-square fit error as a share of the response.

Input	Output	K	τ_1 (s)	τ_2 (s)	θ (s)	Resid.
throttle	P	−1.542 mTorr/%	0.311	0.167	0.0435	0.27 %
throttle	n_e	−0.0269 1/%	0.291	0.197	0.0811	0.30 %
throttle	V_{dc}	1.61 V/%	0.299	0.180	0.0802	0.29 %
source	P	3.483 mTorr/kW	0.309	0.068	0.0294	0.30 %
source	n_e	0.893 1/kW	0.087	0.049	0.0237	0.71 %
source	V_{dc}	−53.6 V/kW	0.075	0.042	0.0319	0.82 %
source	T_w	18.19 °C/kW	15.00	–	0.0398	0.00 %
bias	V_{dc}	0.400 V/W	0.038	0.038	0.0102	0.48 %
bias	T_w	0.121 °C/W	15.00	–	0.0398	0.00 %
helium	T_w	−3.507 °C/Torr	15.04	0.391	–	0.00 %

Exactly zero: bias to P and n_e, helium to P, n_e and V_{dc}, throttle to T_w.

dead time,

$$G_{ij}(s) = \frac{K_{ij}\,\mathrm{e}^{-\theta_{ij}s}}{(\tau_{1,ij}s+1)(\tau_{2,ij}s+1)} \tag{7.13}$$

where K_{ij} is the steady gain of output i to input j, $\tau_{1,ij}$ and $\tau_{2,ij}$ are the two time constants and θ_{ij} is the dead time. Ten pairs respond and six measure exactly zero, as listed in Table 7.3. The pressure, density and self-bias are fitted on the instrument readings and the wafer on its true temperature, which omits the 0.30 s probe filter against a 15 s lag. The gains equal those of Table 7.2 to its three figures, each residual is at most 0.82 % of the response against 0.56 to 3.23 % for a first-order fit, and an independent set of step tests read at every 20 ms scan repeats the ten gains within 0.26 %.

Every dead time is shorter than one 0.1 s prediction step, so each element is realized as a chain of first-order lags with the dead time written as one more lag of its own length. The three wafer elements fit time constants of 15.00, 15.00 and 15.04 s, the single heat balance $C_w/h_{\mathrm{He}} = 60/4 = 15\,\mathrm{s}$ of the chuck, so they share one thermal state x_T with a heat-load disturbance w,

$$\tau_T\,\dot{x}_T = -x_T + w + \sum_j K_{Tj}\,v_j \tag{7.14}$$

where $\tau_T = 15\,\mathrm{s}$, the sum runs over the source, bias and helium, K_{Tj} are the three wafer gains of Table 7.3 and v_j is input j after its element's short lags. The result is a linear model of 26 states in deviation variables about the design point, discretized exactly under a zero-order hold at 0.1 s for the plan and at 20 ms for the estimator. Its numbers are frozen and it reads no chamber parameter or disturbance, so a change made to the simulated chamber reaches the controller only through the instruments, as model mismatch. Three models are therefore kept apart, the fifteen-state nonlinear plant of Section 7.4, the four filtered instruments, and this 26-state identification.

Offset-free tracking comes from a disturbance model in the estimator [18]. At every 20 ms scan, under every structure, the commands applied over the last scan drive the model, with zero source and bias power while the RF is off, and the four instruments correct it through steady-state Kalman gains. The pressure, density and self-bias each carry a constant output disturbance modeled as a random walk (0.045 mTorr, 0.0022 and 0.22 V per scan, against measurement noise of 0.2 mTorr, 0.01 and 1 V), which gives a gain of 0.200 per scan and an error time constant near 0.09 s. On the wafer the thermal state and the heat load of (7.14) are corrected together with gains of 0.618 and 0.309, so a reading is taken up within one scan while a heat load, which appears only through the 15 s lag, is estimated over about 30 s. While the RF is off the density and self-bias corrections are held, because there is no plasma to measure.

Every five scans (0.1 s, counted in whole scans) and at once whenever the RF strikes or goes off, the controller solves

$$\begin{aligned} \min \; & \sum_{k=1}^{N} \sum_{i=1}^{4} \left(\frac{\hat{y}_{i,k} - r_i}{b_i} \right)^2 + \sum_{j=0}^{N_b-1} \sum_{m=1}^{4} \left(\frac{\Delta u_{m,j}}{0.1\,\Delta_m} \right)^2 + \rho\, \varepsilon^2 \\ \text{s.t. } & u_{m,\min} \le u_{m,j} \le u_{m,\max}, \quad \left| \Delta u_{\mathrm{src},j} \right| \le 300\,\mathrm{W}, \\ & \hat{y}_{T,k} \le T_{\max} - 3\,^{\circ}\mathrm{C} + \varepsilon, \quad \varepsilon \ge 0 \end{aligned} \tag{7.15}$$

over $N = 50$ steps of 0.1 s, a 5 s horizon, with the inputs held constant across $N_b = 8$ blocks whose lengths in steps are $(1, 1, 2, 2, 4, 6, 10, 24)$. Here $\hat{y}_{i,k}$ is the predicted measurement including its disturbance estimate, r_i the setpoint, b_i the loop's tolerance band (0.6 mTorr, 0.04, 4 V and 3 °C), $\Delta u_{m,j}$ the move of actuator m at block j, measured for $j = 0$ from the command in force, Δ_m the actuator span (96 %, 2600 W, 700 W and 24.5 Torr), $T_{\max} = 95\,^{\circ}\mathrm{C}$ the resist limit, ε one slack over the whole horizon and $\rho = 10^4\ 1/\mathrm{K}^2$. A band's worth of error at one step therefore costs as much as a move of a tenth of the actuator's span. These weights were fixed by that rule before any run, and one pass at 0.3 and 3 times the move weight, made before the rate limit was added, changed the wafer result of the new scenario by less than 1 %. The actuator box is hard and equal to the clamps of the loops. The source rate limit is hard and applies at every block boundary, because a real RF generator is ramped at ignition and a match network tuning into a plasma that has just lit can reflect power and arc. The plant has no reflected-power or arc model, so the benefit of that row is visible only as the move it removes, which Section 7.6.7 measures. The wafer ceiling at 92 °C is soft. While the RF is off, the blocks before the strike hold the source and bias commands and the density and self-bias terms carry no weight. The decisions are scaled by the actuator spans before the shared dual active-set solver receives the condensed quadratic program, because watts against percent of throttle would otherwise spread its Hessian over eight decades.

The only future event the plan is given is the strike that ends each wafer exchange, which the sequencer schedules from its own load time. With this preview on, the plan knows that the source and bias power reach the chamber at that instant, and with it off the plan holds the present RF state over the horizon and learns of the strike when it happens. Recipe setpoint changes are previewed by neither structure. For the comparison the loops receive the same signal as a feedforward that adds the identified dissociation term to the throttle, $K_{P,\mathrm{src}} P_{\mathrm{src}} / |K_{P,x}|$ or 2.7 percentage points at 1200 W, from 0.3 s before each strike. That feedforward is a row of the comparison script, not an option of the application.

Before a plan reaches the actuators it is checked. The solver's status must be optimal, every number it returns (first move, later moves, predicted outputs, cost and slack) must be finite, each planned command must sit within its box, the source moves must respect the rate limit, and the generators must stay held while the RF is off. Any other outcome is rejected with its reason and marked in the run record, and the four loops, which have tracked every applied command, continue from the same commands. The next accepted plan takes the actuators back over a 0.3 s ramp. None of the roughly 9010 plans of any 900 s cassette reported here failed these checks. Each plan costs about 1.0 ms in Node.js, and in the browser, on the page's own frame loop, the solves take 1.6 to 3.0 % of one core at real time and 5.8 to 7.3 % at five times real time, over two measurements with different amounts of other work on the computer.

7.6 Results

7.6.1 Ion Energy Against Mask Selectivity

Table 7.4 is the central trade and it is not a two-way one. Raising the ion energy from 95 to 185 eV makes the etch 42 % faster, the uniformity 38 % better and the sidewall 0.93° closer to vertical, and it costs 35 % of the mask selectivity because (7.3) sputters the mask with a higher exponent than (7.2) etches the film.

Table 7.4 Closed-loop steady state against self-bias setpoint, all other setpoints at design. Every column but the mask selectivity improves with bias.

V_{dc} (V)	E_i (eV)	R (nm/s)	Mask sel.	U (%)	α (°)
80	95.0	4.828	15.34	2.716	87.942
120	135.0	5.831	12.00	2.200	88.500
170	185.0	6.852	9.99	1.675	88.870

Table 7.5 Closed-loop steady state against pressure setpoint, all other setpoints at design and the wall coverage at its design value of 0.5. Uniformity reaches its floor of 1.20 % at 22.8 mTorr and the sidewall has no interior optimum.

P (mTorr)	R (nm/s)	U (%)	α (°)	Mask sel.
12	3.613	5.919	88.601	7.44
16	4.867	3.840	88.543	10.02
20	5.831	2.200	88.500	12.00
23	6.404	**1.256**	88.468	13.18
26	6.881	2.211	88.435	14.16
34	7.864	4.450	**88.337**	16.18

Scenario 8 runs that trade through the wafer sequence. At 220 V the main etch finishes in 63.8 s instead of 85.9 and leaves 230.2 nm of mask instead of 247.4. At 75 V it takes 107.7 s, leaves 260.6 nm of mask, and pushes the sidewall to 87.84°, which is within a degree of the specification floor. All three pass. The choice among them is throughput against mask budget against profile margin, and it is a business decision rather than a control decision.

7.6.2 Pressure, Uniformity and Sidewall Angle

The uniformity column of Table 7.5 is a valley whose floor lies 2.8 mTorr above the design point, where the edge and center terms of (7.6) cancel, and the sidewall column is a monotone slope in the opposite direction, so no pressure optimizes both. The rate and the mask selectivity both improve monotonically with pressure, so the incentive to run at high pressure is direct and the reason not to is indirect.

Scenario 9 measures the cost of moving in the other direction. Moving from 20 to 13 mTorr for a better sidewall takes the non-uniformity to 5.34 %, above the 4 % specification, and scraps both wafers etched there. The main etch also lengthens from 85.9 to 127.3 s, so the decision costs throughput as well as yield.

> **TRY IT IN THE APP** **9 · Uniformity Against Sidewall**
>
> Press Run and let the cassette step its pressure setpoint. At 13 mTorr the sidewall is straighter, but the non-uniformity reaches 5.34 %, above the 4 % grade, and both wafers etched there are scrapped. Read the main etch time as well: it stretches from 85.9 to 127.3 s.

7.6.3 The Stabilization Step

Scenario 5 deletes the eight-second stabilization step, which is worth seven percent more wafers an hour (one every 113.6 s instead of 121.6 s). The transient is not removed by that change. It is relocated.

With the stabilization step in place, the pressure holds 20.00 mTorr and the self-bias 120.0 V through the whole main etch, and the etch rate varies only between 5.545 and 5.958 nm/s as the wafer warms. With it deleted, the strike lands in the main etch. The self-bias overshoots to 203.3 V while the density is still rising, then dips to 65.5 V, the pressure rises to 22.38 mTorr, and the rate climbs

Table 7.6 Eight wafers after a wet clean, with the wall coverage and the etch rate read at the end of each main etch. The loop setpoints never move.

Wafer	θ_w	R (nm/s)	Cleared	Result
1	0.078	5.393	0.9288	scrapped
2	0.152	5.464	0.9742	scrapped
3	0.213	5.526	0.9897	scrapped
4	0.264	5.577	0.9954	good
8	0.393	5.714	0.9995	good

from near zero. By the time the pressure is within 0.6 mTorr of setpoint, 1.3 s later, the top 5.6 nm of every feature has been etched at the wrong rate and ion energy, where the sidewall shows it.

The application still passes those wafers. Its scoring computes uniformity and sidewall from the final chamber state rather than integrating the profile over depth, so a defect confined to the top of the feature is not represented. The traces show the excursion clearly and the scorecard does not see it, which is a limitation of the model and is recorded as such in Section 7.8.

7.6.4 Open-Loop and Closed-Loop Chamber Control

Scenario 6 detunes the RF match at 130 s so that ten percent of the forward power reflects. The generator continues to report 1200 W.

Under the open-loop recipe the plasma density falls from 1.000 to 0.900, the etch rate at the end of each main etch from 5.834 to 5.209 nm/s and, less obviously, the self-bias *rises* from 120.0 to 126.5 V, because (7.5) has fewer ions carrying the same bias current. The recipe has changed both the rate and the ion energy while every reported actuator value is unchanged.

Putting the chamber loops in service at 400 s raises the source command to 1333 W, returns the density to 1.0000 and the self-bias to 120.0 V within 5 s. The pressure loop opens the throttle from 28.70 to 29.01 % to absorb the extra dissociation, which leaves the rate 0.6 % below that of the undisturbed closed-loop chamber. That is the argument for closed-loop chamber control in one measurement. Match drift, mass flow calibration, wall condition and gas purity all pass straight through an open-loop recipe and are absorbed by a closed-loop one, which is why production tools carry chamber-condition models beside their loops [11].

Pause. Try Exercise 7.2 before reading on.

7.6.5 Chamber Seasoning After a Wet Clean

Scenario 7 starts with a freshly cleaned chamber and runs a cassette with a timed recipe, an 86 s main etch and 12 % over-etch.

> **TRY IT IN THE APP** **7 · Chamber Clean, and What the Timer Does Not Know**
>
> Load the scenario and press Run. Watch the wall-polymer trace on the chamber strip climb while the pressure, density and self-bias tiles hold their setpoints. Then read each wafer's verdict note as it is graded: wafers 1, 2 and 3 are scrapped because the film has not cleared, and wafer 4 is good, with nothing on the tool changed.

Table 7.6 loses three wafers to a variable that is not measured, not controlled and not displayed anywhere on the tool. The wall polymer coverage of (7.8) climbs from 0.078 to 0.393 over eight wafers, the etch rate rises 6 % with it toward the design value of 5.831 nm/s that the model reaches at a coverage of 0.5, and the recipe's fixed 86 s is too short until the chamber has seasoned. The pressure, density and self-bias loops read 20.000 mTorr, 1.0000 and 120.00 V at the end of every main etch, and the temperature loop repeats the same warm-up from 42.6 to 66.4 °C on every wafer, which is the

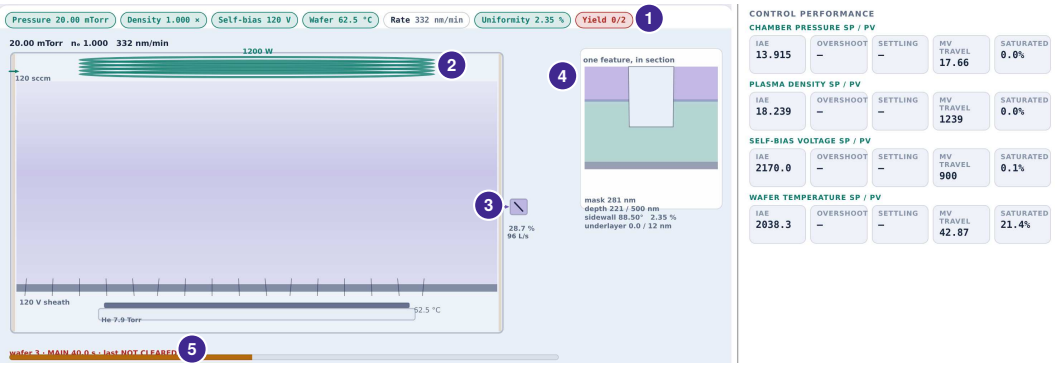


Figure 7.2 Scenario 7 at 274.8 s (4.58 min), 40.0 s into the main etch of the third wafer out of a clean chamber, cropped to the chamber view and, right of the rule, the loop rows of the Learn panel. The four loops read 20.00 mTorr, 1.0000 normalized density, 120 V of self-bias and 62.5 °C at the wafer (tiles along the top). (1) The yield tile. (2) The coil, at 1200 W of source power. (3) The throttle valve on the pump line. (4) The feature cross-section. (5) The wafer sequencer's status line. The integrated absolute errors are 13.9 mTorr s on pressure and 18.2 s on density, while the self-bias and temperature integrals (2170 V s and 2038 °C s) accumulate during each exchange, when there is no plasma to bias, and while each wafer warms from 23 °C. The yield is zero of two (1) and the last wafer is reported as not cleared (5). The feature cross-section (4) shows the etch 221 nm into a 500 nm film with 281 nm of the 300 nm mask left.

point: the loops are controlling the plasma and the plasma is not the whole plant. With the main etch ended by the emission detector instead, the same eight wafers all pass, and the main etch it chooses shortens from 93.6 s on the first wafer to 87.8 s on the eighth as the walls season.

🕮 **Pause.** Try Exercise 7.1 before reading on.

> ✦ **DID YOU KNOW? Close Rates, Opposite Verdicts**
>
> From the simulator: in Table 7.6 the first wafer after the wet clean etches at 5.393 nm/s, clears 0.9288 of its area and is scrapped. The fourth etches at 5.577 nm/s, clears 0.9954, just past the 0.995 grade, and is good. The pressure, density and self-bias read the same on both.

7.6.6 Over-Etch and the Graded Cassette

Over-etch has a two-sided optimum that scenario 10 walks across. At 5 % the wafers do not clear, at 0.987 of the wafer against a 0.995 threshold. At 30 % they clear and lose 6.2 nm into the underlayer. At 75 % they clear and lose 15.2 nm, above the 12 nm limit. Three good wafers out of six, failing at both ends for opposite reasons.

The graded challenge combines that with everything else: a cassette after a wet clean (wall coverage 0.05), hardware with $h_U = 1.15$ in (7.6), a chuck seal already leaking at 80 % of nominal, a match that drifts to 93 % efficiency at 200 s, a product with 1.45 times the open area entering the cassette at 450 s and the seal degrading to 62 % at 700 s. It loads with an open-loop recipe, a timed 86 s main etch, 10 % over-etch and loop gains about an eighth of the tuned set, and it scraps all eight wafers. The challenge is met when, after 900 s, no more than one wafer has been scrapped and at least five are good.

Table 7.7 is the chapter's main result and it is ordered by increasing value. Retuning the four loops is worth nothing, because in the loaded recipe they are not in service and the tuning constants are inert. Closing the four chamber loops is also worth nothing on its own, with the loaded gains or the tuned set, because a timed main etch in a fresh chamber is too short whatever the plasma does.

Table 7.7 The graded cassette. Each row adds one element to the recipe as loaded, not to the recommended one.

Change from the loaded recipe	Wafers good	Worst cleared
none	0 of 8	0.001
retune all four loops	0 of 8	0.001
close the four chamber loops	0 of 8	0.006
close the loops with the tuned set	0 of 8	0.029
raise over-etch to 32 %	5 of 7	0.965
switch to the OES endpoint	**7 of 7**	**0.996**
close the loops and use the endpoint	**7 of 7**	**0.998**

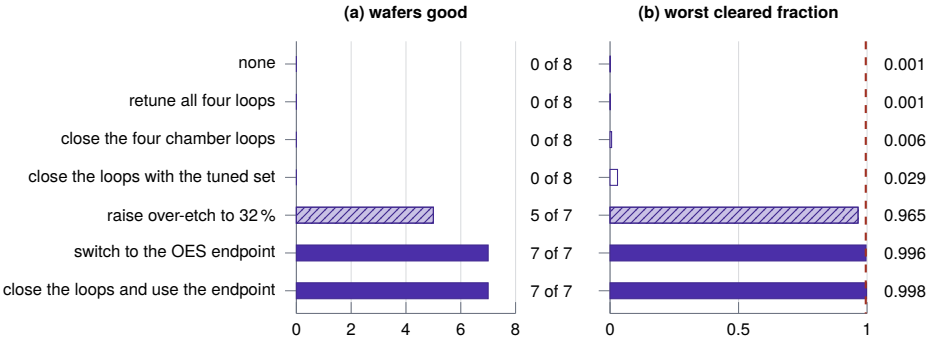


Figure 7.3 The graded cassette of Table 7.7: (a) good wafers of those graded in 900 s, and (b) the worst cleared fraction, against the 0.995 grade (dashed). Open bars change only the loops, the hatched bar the over-etch and the solid bars the end of the main etch. Notice that no change to the loops moves either panel, while the emission endpoint alone takes the worst wafer past the grade.

Enlarging the over-etch is worth five wafers and cannot fix the two that need the most, because a fixed percentage of a fixed main etch is still a fixed time. Figure 7.3 draws the seven rows as bars.

The single most valuable change on the tool is the endpoint detector, worth seven of seven on its own and enough by itself to meet the challenge, and the reason is visible in the times it produces. Under the recommended recipe (closed loops, the emission endpoint, 32 % over-etch and setpoints of 21 mTorr, 108 V and 56 °C), all six wafers graded in 900 s are good and the main etch runs for 95.0, 93.6, 92.5, 101.2, 103.5 and 102.2 s on successive wafers, a spread of 12 %, as the chamber seasons and the product open area changes. The timer in the loaded recipe holds 86 s in every case. The recommended settings give the same run whether they are chosen before Run or carried through a Reset.

Closing the loops as well gives the same seven good wafers at a worst clearance of 0.998 against 0.996. Neither element substitutes for the other's failure mode. The loops reject disturbances inside the chamber and the detector rejects everything the chamber cannot see, including the wafer itself.

7.6.7 Predictive Control Against the Four Loops

Scenario 11 is the cassette of scenario 4 run by the predictive controller of Section 7.5.4, with the strike preview on, a timed main etch and a recipe step of the self-bias setpoint from 120 to 150 V at 300 s. Bias power heats the wafer, which is a coupling that (7.1) cannot see because it runs in one direction only. The comparison of Table 7.8 runs scenario 11, scenario 8 (self-bias steps to 220 and then 75 V, emission endpoint) and scenario 5 (the strike inside the main etch) for 900 s each, at the same seed with the sensor noise off, the same instruments and the same sequencer, and grades

Table 7.8 Predictive control against the four loops. Seconds per wafer outside the loop bands from the true outputs, seven good wafers per 900 s run in every row. Scenario 11 steps the self-bias setpoint, scenario 8 (Sc. 8) runs the bias trade and scenario 5 strikes the plasma inside the etch.

Structure	Scenario 11				Sc. 8	Scenario 5		Source
	T_w out (s)	all out (s)	all out, lit (s)	T_w peak (°C)	T_w out (s)	peak $\lvert\Delta P\rvert$ (mTorr)	all out (s)	travel (W)
Loops, shipped tuning	37.08	37.18	48.31	67.5	42.86	2.38	43.98	482
Predictive controller	3.54	3.56	16.54	62.0	3.74	0.78	19.77	1071
Loops retuned once (SIMC)	4.14	4.16	13.50	61.9	4.89	2.26	15.11	1309
Predictive, diagonal model	4.18	4.20	13.74	62.8	12.13	2.13	15.28	1132
Predictive, strike preview off	3.54	3.56	16.44	62.0	3.74	0.55	19.70	1088
Predictive, no constraint rows	3.85	3.88	12.51	62.0	4.00	0.33	14.06	3553
Predictive, no source rate limit	3.49	3.52	16.47	62.0	3.75	0.25	19.69	3749
Loops with strike feedforward	37.05	37.16	47.53	67.5	42.84	1.46	43.18	432

The retuned loops have gains of 1.01 %/mTorr, 1294 W per unit density, 2.44 W/V and 11.1 Torr/°C with integral times of 0.39, 0.11, 0.057 and 1.56 s. Source travel is the total source command movement per wafer in scenario 11. The predictive rows solve about 9010 plans per run at a mean of 1.0 ms each in Node.js (0.43 ms with the diagonal model, measured with other work on the computer), and none is rejected. "No constraint rows" removes the actuator box, the source rate limit and the wafer ceiling and clips the first move afterward. Without the rate limit the plan of scenario 11 is the one shipped before that limit was added.

seven wafers per run. Every wafer of every row is good. The figure of merit is the time per wafer that each true output spends outside its loop's tolerance band while the wafer is exposed (main etch and over-etch), summed over the four outputs, and the same while the plasma is lit with a wafer in the chamber, which adds the stabilization step. The rows are the shipped loops, the predictive controller as scenario 11 ships it, the same loops retuned once by the SIMC rules [14] from the four diagonal elements of Table 7.3, and five counterfactuals, each with one mechanism of the controller removed or one of its signals given to the loops. No case is retuned for any row.

The predictive controller cuts the time the wafer spends outside its 3 °C band from 37.08 to 3.54 s per wafer on scenario 11 and from 42.86 to 3.74 s on scenario 8, and its hottest wafer is 62.0 °C against 67.5 °C, as listed in Table 7.8. Almost all of that gain is tuning. The shipped wafer loop is slow (0.45 Torr/°C with a 15 s integral time), and the same four loops retuned once by a standard rule reach 4.14 and 4.89 s, which is 98 and 97 % of the controller's improvement. What the bias-to-wafer interaction term adds is the remainder, 0.60 and 1.15 s per wafer, and it is obtained with a seventh and a quarter of the retuned loop's helium travel (24 against 170 and 39 against 155 Torr per wafer), because the plan raises the helium when the bias power rises instead of waiting for the wafer to warm. The diagonal-model row overstates the term on scenario 8 (12.13 s), and most of that loss is its own move weight, since the same diagonal plan reaches 8.41 and 6.01 s at 0.3 and 0.1 times the weight while the full plan stays at 3.74 s.

The strike pressure has a different source. With the stabilization step deleted, the shipped loops let the pressure move 2.38 mTorr at each strike and the predictive controller 0.78 mTorr. That gain belongs to the source-to-pressure term, because the same plan on a diagonal model leaves 2.13 mTorr, and the loops given the preview's signal as a throttle feedforward reach 1.46 mTorr. The re-plan at the moment of the strike tells the plan the generators are on, so the term acts with or without the preview. The preview itself now costs pressure, 0.78 against 0.55 mTorr with it off. Knowing the source will ramp, the plan closes the throttle from 25.8 to 22.9 % in the last half second before the strike, raising the pressure about 0.85 mTorr so that the higher pressure brings the density up sooner, and the strike then lands on a chamber already above setpoint. Before the source rate limit was added the preview helped instead (0.25 against 0.57 mTorr). On scenario 11 the preview changes no measured second, and the scenario keeps it on as it was fixed before these results.

The rate limit exists because of what the unconstrained plan did at ignition. Its linear model starts the density from zero power and knows nothing of ignition, so at every strike it commanded 2432 W

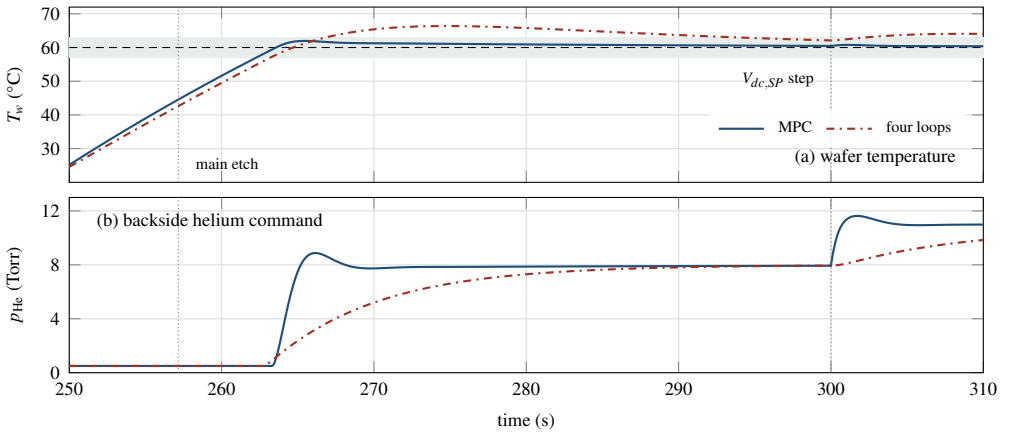


Figure 7.4 Scenario 11 in the 60 s before $t = 310$ s, the third wafer warming from 25 °C into its main etch and the self-bias setpoint stepping from 120 to 150 V at 300 s. (a) Wafer temperature, with its setpoint (thin dashed) and its band (shaded), and (b) backside helium command, under the predictive controller (solid) and the four loops (dash-dotted). Dotted verticals mark the start of the main etch and the setpoint step. At 310 s the wafer reads 60.37 °C with 10.99 Torr of helium under the plan, and 64.05 °C, outside its band, with 9.84 Torr under the loops.

of source for one decision and then 924, 1101 and 1328 W, which sent the true density to 1.42 (the actinometry read 1.19) and reached the 2600 W clamp at every strike once the match had drifted to 0.85. With the limit of 300 W per decision the first move is 1500 W, the density peaks at 1.15, and the source travels 2.2 times as far as under the loops instead of 7.8 times (1071 and 3773 W per wafer in scenario 5 against 482). The limit costs 0.05 s of wafer time on scenario 11 and the strike pressure discussed above.

The plan also decouples. On scenario 3 with the plasma held, a density step from 1.00 to 1.10 drags the self-bias 1.35 V under the predictive controller, 2.16 V under the loops and 5.98 V under the diagonal plan, and a 2 mTorr pressure step is inside a quarter of its band in 0.33 s against 0.90 s under the loops. The diagonal plan matches that speed (0.32 s), so speed comes from the model and the 0.1 s plan, while the smaller cross-excursions come from the interaction terms.

The instant that decides the wafer comparison is shown in Figure 7.4. Both structures warm the third wafer from 25 °C with the helium at its 0.5 Torr floor and reach the band within a second of each other (262.3 and 263.2 s), after which the loops overshoot to 66.4 °C at 274.7 s while the plan stays below 62.0 °C. When the self-bias setpoint steps, the plan raises the helium at once and the wafer peaks at 60.78 °C, while the slow loop waits for the temperature error and the wafer reaches 64.07 °C at 308.8 s.

The losses are as measured. Over the lit time, which includes the warm-up in the stabilization step, the controller spends 16.54 s per wafer outside the bands against 13.50 for the retuned loops and 12.51 for itself without its constraint rows. Knowing the helium is at its floor while a cold wafer warms, the plan heats the wafer with the only other actuators that can, the source and bias power, and holds the density and self-bias up to a band high for several seconds. With the stabilization step deleted that trade falls inside the etch, where the density spends 4.69 s per wafer outside its band against 0.77 under the loops and 0.31 without the rows, and scenario 5's total of 19.77 s loses to the retuned loops' 15.11 s. Sensor noise passes through at about twice the loops' amplitude, a peak pressure error during etching of 0.446 against 0.224 mTorr with the noise slider at one and 1.336 against 0.678 mTorr at three on scenario 4. On scenario 7, with a timed recipe after a chamber clean, the controller grades four good wafers of eight against five, because it holds the wafer at its 60 °C setpoint where the loops overshoot to 66 °C, so the chemical rate is lower and a timer calibrated

Table 7.9 Worst case over every scenario run for 900 s under each structure. Hottest wafer and good wafers of those graded.

Scenario	Four loops	Predictive
4, cassette	66.4 °C, 7 of 7	62.0 °C, 7 of 7
5, stabilization deleted	66.4 °C, 7 of 7	62.0 °C, 7 of 7
6, open-loop recipe[a]	66.9 °C, 7 of 7	66.6 °C, 7 of 7
7, chamber clean, timed	66.4 °C, 5 of 8	62.0 °C, 4 of 8
8, rate against selectivity	76.1 °C, 7 of 7	62.6 °C, 7 of 7
9, uniformity against sidewall	67.4 °C, 4 of 6	62.0 °C, 4 of 6
10, over-etch	66.4 °C, 3 of 6	62.0 °C, 3 of 6
11, predictive control	67.5 °C, 7 of 7	62.0 °C, 7 of 7
12, challenge as loaded	80.2 °C, 0 of 8	62.6 °C, 0 of 8
12, suggested setpoints	66.4 °C, 6 of 6	58.9 °C, 6 of 6
4, chuck seal at 0.40	78.0 °C, 7 of 7	65.4 °C, 7 of 7

[a]The structure named runs from $t = 0$, and the scenario's own event puts the four loops in service at 400 s.

under the loops falls short. The helium also sits at a limit longer, 112.5 against 108.3 s on scenario 4 and 773 against 435 s with the chuck seal at 0.40 of nominal.

No run of Table 7.9 brings any wafer near the 95 °C resist limit, the predictive controller's hottest wafer is the cooler of the two in every scenario that runs wafers and at every disturbance slider of scenario 4 set to its limit, and no plan is rejected anywhere. The match at 0.70, the flow calibration at 0.85 and 1.15, the wall loss at 1.6 and the hardware uniformity factor at 2.2 leave both structures with seven good wafers of seven. The two failures that remain at their limits, a doubled open area and a reduced etchant fraction, leave every wafer uncleared under both structures. The predictive controller is less favorable on scenario 7, by one wafer, and under sensor noise, as described above.

The cost leaves the browser at 60 frames per second throughout. Each plan takes 1.1 to 3.0 ms on average in Chrome, ten plans per simulated second take at most 7.3 % of a core at the top speed, and no frame reached the 25 ms stepping guard that would slow the run (the longest frame stepped for 16.8 ms). The four loops step at five times real time in under 1.0 % of a core.

7.6.8 The Running Application

The application as it runs is reproduced in Figure 7.2. It runs in a browser with no installation and computes everything on the client, and offers manual operation, an open-loop recipe, the four closed loops and the predictive controller. Each run exports as a comma-separated file carrying the four loops, the two neutral densities, the ion energy, the wall polymer coverage, the etch rate, both selectivities, the uniformity, the sidewall angle, the etched depth and the phase, so the recipe trades of Tables 7.4 and 7.5 can be reproduced offline. Manual operation offers two helpers that work on the tool's declared model of a seasoned chamber rather than on the chamber's true state. The first sets the four actuators for the recipe setpoints, through the two inverse calculations of Section 7.4, and so returns the chamber to the design point from which the steps of Table 7.2 can be repeated by hand. The second sets the main etch time from the rate that model predicts.

7.7 Technology in Practice

The loops of Section 7.5 read a pressure gauge, an optical emission spectrometer, a voltage sensor on the bias line and a wafer temperature probe, and they move a throttle valve, two radio-frequency generators and a helium valve. Figure 7.5 draws them with their tags.

What Measures It

Both the density loop and the endpoint detector read the light of the plasma, the first through actinometry and the second through the emission signal of Section 7.5.3.

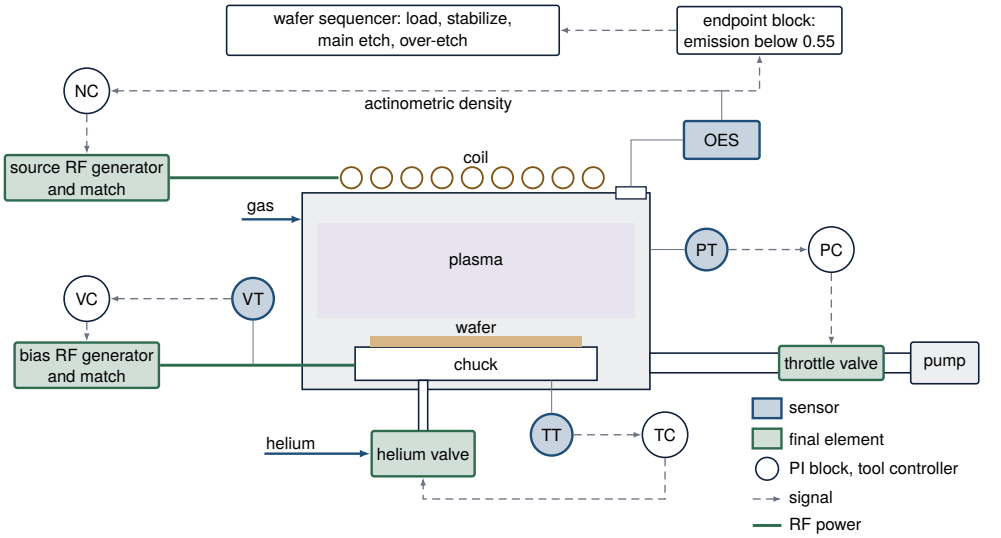


Figure 7.5 The four chamber loops of Figure 7.1 drawn with their hardware. The optical emission spectrometer (OES) at the viewing window serves twice: its actinometric density feeds the density loop NC, and its filtered emission ends the main etch in the wafer sequencer. All four PI loops, the sequencer and the endpoint block run in the etch tool's controller, scanned every 20 ms. Notice that the endpoint signal moves no actuator: it ends a step of the recipe.

TECHNOLOGY Sensor: An Optical Emission Spectrometer

Light from the plasma passes a window and a lens onto a detector, from a filtered photodiode that reads one wavelength to a spectrometer that records the whole spectrum at regular intervals. Each species emits at its own wavelengths, so the light follows the gas composition in real time without touching the plasma. The intensity also depends on the electrons that excite the gas, which is why actinometry compares a line with that of a reference gas. The integration time filters the signal: a longer one gives less noise and a slower reading. Deposits cloud the window and weaken the signal wafer after wafer, and a spectrometer that is knocked can lose its calibration [1].

The 0.30 s filter on the modeled emission stands for the integration time, and the model has no clouded window. The self-bias is read electrically on the bias line, where sensors between the matching network and the electrode need no optical access and do not suffer from clouded windows [1].

What Moves It

The throttle valve on the pump line is clipped to 4 to 100 %, the source and bias generators to 0 to 2600 W and 0 to 700 W, and the backside helium to 0.5 to 25 Torr.

TECHNOLOGY Final Element: A Radio-Frequency Generator and Its Match

The source and bias powers come from radio-frequency generators, which typically have an output resistance of 50 Ω. Connected straight to a plasma, a generator would deliver its power poorly, so a matching network between generator and electrode presents the generator with a purely resistive load equal to its own. Its variable elements, often two vacuum capacitors called load and tune beside a fixed inductor, are adjusted automatically as the plasma impedance changes with pressure, temperature and gas flow. The network also protects the generator from reflected power. Setting one up takes real expertise, because every wire has appreciable inductance and capacitance at these frequencies [1].

The match efficiency of scenario 6 is the model's form of a detuned network. A tenth of the forward power reflects while the generator still reports 1200 W, so on a trend the forward power stays flat while the density falls.

Where the Controller Runs

The four loops and the sequencer belong to the etch tool, which runs the recipe: setpoints of pressure, power and gas flow for each step, developed by experiment for each product and distributed to the fabs [1]. Supervisory and lower-level loops run with cycle times from milliseconds to minutes, and what counts as real time has to be stated for each application [19]. The app scans the four loops every 20 ms and the predictive controller plans every 0.1 s, while a wafer spends about two minutes in the chamber. The structure of this chapter is three layers in that controller: four PI blocks in velocity form, a step sequencer that executes the recipe, and an endpoint block that ends one step.

> **IN PRACTICE** **A Production Recipe Stops One Step on Light**
>
> A production trench etch studied in 2011 ran five steps. A timed step removed about three quarters of the exposed nitride layer, a gentler chemistry removed the rest and stopped on the endpoint signal of a monochromator, and an over-etch of fixed length followed. The main trench step was timed, and operators set its length off-line, adjusting it by hand for drift. Chambers drift as etch products coat their walls, and the first one or two wafers of a lot often etch differently. Maintenance cleans the walls and the viewing window, which restores the endpoint signal to full strength [1].

What Is Modernizing It

Three developments bear on this chamber. Multivariate endpoint methods read the whole emission spectrum instead of one line, which keeps the endpoint usable as the open area shrinks [1, 8]. Virtual metrology uses the in-situ measurements of every wafer, the emission spectrum and the electrical signals of the RF line among them, with models that estimate the etch result before any metrology tool sees the wafer [1]. And a 2024 review of the field, which faces three-dimensional devices and dimensions approaching the atomic scale, names digital twins of etch chambers, built from hybrid models and machine learning, among its opportunities [20]. Real-time multivariable feedback on measured plasma variables, the regulatory layer of this chapter, was demonstrated on a reactive ion etcher in 1995 [17].

7.8 Limitations

The most important limitation is the one Section 7.6.3 ran into. Wafer properties are scored from the chamber state at the end of the etch rather than integrated over the etched depth, so a defect confined to the top or bottom of a feature does not appear on the scorecard even when the traces show the excursion that caused it. Any conclusion about transients inside the product must be read from the traces rather than from the yield.

The plasma is a global model with two lumped neutral species, no spatial structure and a normalized rather than absolute electron density. Real reactors are designed around their spatial structure, and coil geometry, gas injection and wafer edge effects are what a process engineer adjusts to correct uniformity. Equation (7.6) is a fitted shape that reproduces the interior optimum and cannot tell anyone where it is on a particular tool.

Aspect-ratio dependent etching is absent. Real features etch more slowly as they deepen, and the resulting dependence on feature size is the reason a recipe that works on one product does not work on the next. The application carries an open-area load effect but no feature-scale transport, so its sidewall angle is uniform with depth where a real profile is not.

Chamber memory is one scalar. Real chambers accumulate several distinguishable conditions, respond differently to each cleaning chemistry, and drift over weeks as well as over wafers, which is why production tools carry run-to-run control layers that adjust the recipe between wafers on

measured results [12]. Nothing in this application represents that layer, and it carries most of the advanced control in a production fab. Chapter 6 adds it to an etcher whose measurement arrives two lots late. The endpoint is also only a threshold. A moving horizon estimate or a Kalman filter driven by the emission signal and the clearing model of (7.9) would follow the cleared fraction continuously instead of declaring one crossing, and would size the over-etch from the estimated spread instead of a fixed percentage.

The predictive controller's model is linear about one operating point and leaves out what the plant does elsewhere. It knows nothing of ignition, so a strike is to the plan only the two generators going from zero to their commands. Its wafer element is the 15 s lag at 8 Torr of helium, while at the 0.5 Torr floor the plant's wafer gain to bias power is 0.844 °C/W with a 104.5 s time constant against the model's 0.121 and 15 s, because the backside conductance scales with $p_{\mathrm{He}}^{0.7}$. The throttle's effect on the pressure departs from the model by 7.8 % for a 1.4-point step, and the wall coverage, the match, the seal, the flow calibration and the pattern loading enter only through the output and heat-load disturbances of the estimator. The plant also has no model of reflected RF power, arcing or generator slew, so the harm the source rate limit prevents is not scored.

Finally, there is one wafer state rather than a wafer map, so the application reports a single non-uniformity number and cannot show where on the wafer the problem is, which is the first question a process engineer would ask.

7.9 Discussion

Three conclusions generalize beyond this chamber.

The first is that fast loops on a slow product convert a dynamic control problem into a static optimization problem, and the engineering effort should follow. Table 7.1 and (7.1) together say that the regulatory layer here is a solved problem, and Tables 7.4 and 7.5 say that the interesting decisions are all about where to put the setpoints. A course that spent its time on tuning alone would not prepare a student for any of the decisions that matter on this plant. The predictive controller of Table 7.8 is the measured form of that claim, because the same four loops retuned once recover 97 to 98 % of its gain on the wafer and its interaction terms are worth about a second per wafer.

The second is that a control element need not be a loop. The endpoint detector in Table 7.7 has no setpoint, no actuator and no error signal, and it is worth more than the four PI controllers combined because it is the only element that measures the product rather than the process. The general form is that a measurement taken on the product usually outperforms a measurement taken on the equipment making it, and endpoint detection is universal in this industry and rare elsewhere because here the product measurement is available while the process is running. Where it is not, Volume 8, Chapter 8 can end each coating step on an estimated thickness instead.

> ⚠ **OPEN CHALLENGE** **Finding the End When Almost Nothing Is Open**
>
> Endpoint detection is easy when more than a tenth of the wafer is open to the etch, because the emission changes clearly. As devices shrink the open area falls, and the change in the signal sinks toward the noise. Molecular etch gases give a continuous spectrum with no clean line to watch, and the etched mask adds an emission of its own [1]. Multivariate methods read the full spectrum: one in 2023 raised the signal-to-noise ratio 2.78 times over a single wavelength on wafers with 1.0 % of the area open [8]. Progress would be an endpoint that fires reliably at any open area. The detector of this chapter thresholds one modeled signal that falls cleanly as the film clears.

The third is that regulatory performance can be perfect while the plant produces nothing of value. Figure 7.2 shows the pressure, density and self-bias loops on setpoint to the displayed precision in the middle of a main etch and a yield of zero, which is the same conclusion the other applications in this series reach from different directions [2, 4].

7.10 Conclusion

A plasma etch chamber settles far faster than the wafer it is processing, its fastest loop within 0.15 s and its slowest within 45 s against a wafer that occupies the chamber for about two minutes, and its four natural loops are nearly decoupled despite a gain matrix in which the source generator moves all four measured variables. That makes it the one plant in this series where tuning is genuinely a secondary concern, and the application is built to demonstrate what takes its place. Ion energy buys rate, uniformity and profile and sells mask selectivity, at 42 % against 35 % across the useful range. Pressure has a uniformity optimum near 23 mTorr and a sidewall that keeps improving below it. And in a graded cassette after a wet clean, a timed open-loop recipe scraps all eight wafers, retuning the idle loops is worth nothing, closing them recovers none, and an optical endpoint detector that has no setpoint at all recovers seven of seven, because the main etch a fresh chamber actually needs varies by twelve percent from wafer to wafer and a timer cannot know it. A model predictive controller over all four actuators cuts the time the wafer spends outside its band after a recipe step from 37.1 to 3.5 s per wafer, but the same loops retuned once reach 4.1 s. Beyond about a second of wafer time per wafer, the gain that belongs to its interaction terms is at the strike, where the pressure moves 0.78 instead of 2.38 mTorr.

KEY POINTS

- The loops settle far faster than a wafer is etched, so the setpoints, not the tuning, decide the wafer.
- Ion energy buys rate, uniformity and profile and sells mask selectivity, 42 % against 35 %.
- After a wet clean a timed recipe scraps three wafers while the plasma loops hold their setpoints, because the wall polymer that sets the rate is not measured.
- An optical endpoint with no setpoint recovers seven wafers of seven in the graded cassette, and retuning or closing the loops recovers none.
- A predictive controller cuts the wafer’s time outside its band from 37.1 to 3.5 s, and the loops retuned once reach 4.1 s.

Where else this applies. A phase sequencer above loops in Volume 8, Chapter 7, a batch step ended on an estimate of the product in Volume 8, Chapter 8, and a triangular gain matrix in Volume 6, Chapter 2.

References

[1] S. Lynn, “Virtual metrology for plasma etch processes,” Ph.D. dissertation, National University of Ireland Maynooth, 2011. [Online]. Available: https://mural.maynoothuniversity.ie/id/eprint/2657/

[2] D. E. Seborg, T. F. Edgar, D. A. Mellichamp, and F. J. Doyle, III, *Process Dynamics and Control*, 4th ed. Hoboken, NJ: John Wiley & Sons, 2016.

[3] B. W. Bequette, *Process Control: Modeling, Design, and Simulation*. Upper Saddle River, NJ: Prentice Hall, 2003.

[4] K. J. Åström and R. M. Murray, *Feedback Systems: An Introduction for Scientists and Engineers*, 2nd ed. Princeton, NJ: Princeton University Press, 2021.

[5] M. A. Lieberman and A. J. Lichtenberg, *Principles of Plasma Discharges and Materials Processing*, 2nd ed. Hoboken, NJ: John Wiley & Sons, 2005.

[6] J. W. Coburn and H. F. Winters, “Ion- and electron-assisted gas-surface chemistry: An important effect in plasma etching,” *Journal of Applied Physics*, vol. 50, no. 5, pp. 3189–3196, 1979.

[7] K. Hirobe and T. Tsuchimoto, “End point detection in plasma etching by optical emission spectroscopy,” *Journal of The Electrochemical Society*, vol. 127, no. 1, pp. 234–235, 1980.

[8] S. Lee, H. Choi, J. Kim, and H. Chae, “Spectral clustering algorithm for real-time endpoint detection of silicon nitride plasma etching,” *Plasma Processes and Polymers*, vol. 20, no. 6, p. e2200238, 2023.

[9] J. O. Clevenger, M. J. Buie, and N. L. Sandlin, “Effect of chamber seasoning on the chrome dry etch process,” in *Proceedings of SPIE*, vol. 5130, 2003, p. 92.

[10] M. Surendra, C. R. Guarnieri, G. S. Selwyn, and M. Dalvie, “Across wafer etch rate uniformity in a high density plasma reactor: Experiment and modeling,” *Applied Physics Letters*, vol. 66, no. 18, pp. 2415–2417, 1995.

[11] S. Gaddam and M. W. Braun, "Etch chamber condition-based process control model for shallow trench isolation trench depth control," in *IEEE/SEMI Conference and Workshop on Advanced Semiconductor Manufacturing*, 2005, pp. 17–20.

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

[13] K. J. Åström and T. Hägglund, *Advanced PID Control*. Research Triangle Park, NC: ISA – The Instrumentation, Systems, and Automation Society, 2006.

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

[15] S. J. Qin and T. A. Badgwell, "A survey of industrial model predictive control technology," *Control Engineering Practice*, vol. 11, no. 7, pp. 733–764, 2003.

[16] E. F. Camacho and C. Bordons, *Model Predictive Control*, ser. Advanced Textbooks in Control and Signal Processing. London: Springer, 2007.

[17] B. A. Rashap, S. Lafortune, J. R. Moyne, D. Teneketzis, F. L. Terry, M. E. Elta, H. Etemad, J. P. Fournier, J. S. Freudenberg, M. D. Giles, J. W. Grizzle, P. T. Kabamba, and P. P. Khargonekar, "Control of semiconductor manufacturing equipment: Real-time feedback control of a reactive ion etcher," *IEEE Transactions on Semiconductor Manufacturing*, vol. 8, no. 3, pp. 286–297, 1995.

[18] K. R. Muske and T. A. Badgwell, "Disturbance modeling for offset-free linear model predictive control," *Journal of Process Control*, vol. 12, no. 5, pp. 617–632, 2002.

[19] 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.

[20] G. S. Oehrlein, S. M. Brandstadter, R. L. Bruce, J. P. Chang, J. C. DeMott, V. M. Donnelly *et al.*, "Future of plasma etching for microelectronics: Challenges and opportunities," *Journal of Vacuum Science & Technology B*, vol. 42, no. 4, p. 041501, 2024.

7.11 Exercises

7.1 Concept check · 2 min

In scenario 7 the pressure, density and self-bias loops read 20.000 mTorr, 1.0000 and 120.00 V at the end of every main etch, yet the timed recipe scraps wafers 1 to 3 and passes wafer 4. What changed between wafer 1 and wafer 4?

(a) The wafer temperature loop settled, so the chemical term of the etch rate rose

(b) The density loop wound up its integral, so the plasma density rose during the etch

(c) The emission detector learned the clearing time and lengthened the main etch

(d) The wall polymer coverage rose, so fewer radicals were lost to the walls

7.2 Concept check · 1 min

Under the open-loop recipe of scenario 6 the RF match drifts so that a tenth of the forward source power reflects. The bias generator still delivers 150 W. What does the self-bias do?

(a) It rises, because a thinner plasma has fewer ions to carry the bias current

(b) It falls, because the plasma receives less power and develops less voltage

(c) It stays at 120 V, because the bias generator is a separate supply

(d) It falls, because less dissociation lowers the chamber pressure

7.3 Concept check · 2 min

In the graded challenge the emission endpoint alone is worth seven good wafers of seven. Which description of the endpoint detector is correct?

(a) A PI loop that trims the source power to hold the emission at its 0.55 setpoint

(b) A threshold on a filtered emission signal that ends the main etch, with no setpoint or actuator

(c) A model that predicts each wafer's clearing time from the chamber rate before the etch

(d) A derivative trigger on the emission signal, armed after a minimum main etch time

7.4 Theory and calculation · 7 min

Use Equation (7.5) and the gain matrix of Table 7.2. The design point is 120 V of self-bias at 150 W of bias power and a density of 1.000.

(a) Find $\partial V_{dc}/\partial P_{\text{bias}}$ and $\partial V_{dc}/\partial n_e$ at the design point. With the density gain of 8.94×10^{-4}/W from the table, predict the source-to-V_{dc} entry to three significant figures. *Checkpoint:* 0.400 V/W; –60 V per unit density; –0.0536 V/W.

(b) The density and pressure rows with the source and throttle columns form a 2×2 block. Compute its relative gain $\lambda = 1/(1 - K_{12}K_{21}/(K_{11}K_{22}))$ for density on the source and pressure on the throttle, to three decimals. *Checkpoint:* 1.073.

(c) Why are the self-bias and wafer-temperature diagonal entries of Equation (7.1) exactly 1, and what does the relative gain of density on the throttle say about that pairing? *Checkpoint:* The bias and helium columns reach no row above their own, so the matrix is block triangular; $1 - 1.073 = -0.073$ is negative, so that pairing is rejected.

7.5 Theory and calculation · 7 min

The wafer obeys $C_w\dot{T}_w = \eta\,(P_{\text{bias}} + s\,P_{\text{src}}) - h(p_{\text{He}})\,(T_w - T_{\text{chuck}})$ with $h = h_0\,(p_{\text{He}}/8\,\text{Torr})^{0.7}$, $C_w = 60\,\text{J/K}$, $h_0 = 4.00\,\text{W/K}$ and $T_{\text{chuck}} = 20\,^\circ\text{C}$. At the design point the wafer holds 60 °C at 8 Torr.

(a) From the wafer row of Table 7.2, find the fraction η of bias power that reaches the wafer and the source share s, to three decimals. *Checkpoint:* $\eta = 0.484$, $s = 0.150$.

(b) Predict the helium entry of that row from the balance. *Checkpoint:* –3.50 °C/Torr.

(c) A cold wafer warms with the helium at its 0.5 Torr floor. Find the conductance, the time constant C_w/h and the steady gain of the wafer to bias power there, and compare them with the predictive controller's 15 s and 0.121 °C/W. *Checkpoint:* 0.574 W/K, 104.5 s, 0.843 °C/W; about seven times both.

7.6 Theory and calculation · 7 min

Use Equation (7.9) for a design wafer: film 500 nm, non-uniformity 2.2 %, etch rate 5.831 nm/s and an 86 s main etch. The emission signal is $(1 - c) + 0.12$ and the detector fires when the filtered signal falls below 0.55, which the filter lag delays to 64 % cleared.

(a) Find the cleared fraction at the threshold, the width of the clearing transition, and the depth past the film thickness at which the detector fires. *Checkpoint:* 0.57; 6.6 nm; 3.80 nm.

(b) Find the further depth needed to clear 0.995 of the wafer, the time it takes, and that time as a percentage of the main etch. *Checkpoint:* 31.1 nm, 5.34 s, 6.2 %.

(c) Find the cleared fraction after a 5 % over-etch, and the over-etch, as a percentage of the same 86 s main etch, that a wafer of 4 % non-uniformity would need if the detector still fired at 64 % cleared. *Checkpoint:* 0.988; 11.3 %.

7.7 App challenge · 8 min

Load scenario 10, How much over-etch is enough (default seed), and run it as loaded at 5 s/s until the Yield tile reads 3 / 6 (wafer 6 is graded at about 12.5 min; the next wafer is graded after 15 min). Pause and export the CSV. Its time column is Time (min); the wafer number is `wafer_no`, and a wafer's rows run from its strike to the next.

(a) For each wafer, find the largest `cleared` and the largest `under_nm` in its rows, to ±0.001 and ±0.1 nm, and the over-etch setting `ctrl_over_pct` in its rows whose `phase` is `over`. Which wafers are good, and why do the others fail? *Checkpoint:* Wafers 1–2: 5 %, 0.987, 1.15 nm, not cleared. Wafers 3–5: 30 %, 1.000, 6.16 nm, good. Wafer 6: 75 %, 1.000, 15.18 nm, punch-through.

(b) The over-etch of the 30 % wafers lasted 25.8 s and that of wafer 6 64.5 s. Assume the loss grows in proportion to the over-etch time once the film has cleared. Find the loss per second, and the largest over-etch percentage of the 85.96 s main etch that keeps the loss under the 12 nm limit, to the nearest percent. *Checkpoint:* About 0.23 nm/s; about 59 %.

7.8 App challenge · 8 min

Load scenario 11, One plan for four loops: MPC against the closed loop (default seed), and run it as loaded at 5 s/s until the clock passes 10.0 min. Pause and export the CSV (its time column is Time (min)). Press Reset, set Chamber control to Closed loop, run again to 10.0 min and export a second file.

(a) In each file, find the largest `Tw_C`, to ±0.1 °C. *Checkpoint:* 62.0 °C under the predictive controller, 67.5 °C under the loops.

(b) In each file, read `Tw_C` and `he_cmd_Torr` in the row at 310 s (Time (min) 5.16666667), ten seconds after the self-bias setpoint steps to 150 V, to ±0.02. *Checkpoint:* 60.37 °C and 10.99 Torr; 64.05 °C and 9.84 Torr.

(c) Explain from the helium command why the plan's wafer stays inside its 3 °C band through the step while the loops' wafer leaves it. *Checkpoint:* The plan's model carries the bias-to-wafer gain, so it raises the helium as the bias power rises; the loop waits for the temperature error and is tuned slowly.

7.9 App challenge · 9 min

Load the graded challenge, Challenge · A cassette after a wet clean (default seed). Before Run, set Main etch ends on to Emission and change nothing else. Run at 5 s/s and Pause once the clock passes 15.0 min and before 16.5 min. Then reload the scenario with the reload button beside the scenario list, set only Over-etch (under Recipe timing) to 32 %, and repeat. The target is at most one scrapped wafer and at least five good ones after 900 s.

(a) Read the Yield tile after each run and say whether the target is met. *Checkpoint:* 7 / 7, met; 5 / 7, not met.

(b) Which wafers does the second run scrap, as the wafer notes say, and why can more over-etch not save them? *Checkpoint:* Wafers 5 and 6, not cleared. The more open product from 450 s etches slower, and a percentage of a fixed 86 s main etch is still a fixed time.

(c) Closing the four loops on their own leaves every wafer scrapped. Why does the endpoint, which has no setpoint and no actuator, meet the target alone? *Checkpoint:* It measures the product. The rate falls with the clean, the match and the product change, and the endpoint lengthens each main etch to match.

7.10 Application area · 3 min

Scenario 5 deletes the 8 s stabilization step to gain 7 % more wafers an hour, and every wafer is still graded good.

(a) What fraction of the 500 nm film is etched before the pressure is back within 0.6 mTorr of setpoint, and why does the Yield tile not see it? *Checkpoint:* About 1.1 % (5.6 nm); the wafer is scored from the chamber state at the end of the etch.

(b) What would a process engineer check before accepting the change? *Checkpoint:* The top of the feature profile on sampled wafers, and the strike transient in the tool's own trace data.

7.11 Application area · 4 min

After a wet clean, a fab running the timed recipe of scenario 7 burns seasoning wafers before product.

(a) From Table 7.6, how many seasoning wafers does it need, and how much tool time do they take at the scenario's 110.4 s per wafer? *Checkpoint:* Three; about 331 s, or 5.5 min.

(b) A run-to-run controller lengthens the main etch from the film left on each wafer, measured before the next one starts. Which wafers can it save, and what can the emission endpoint do that it cannot? *Checkpoint:* Wafers 2 onward at best; wafer 1 is lost. The endpoint saves wafer 1 too, because it measures the wafer being etched.

7.12 Review the agent · 5 min

An AI assistant reviews scenario 11 and writes: "The predictive controller cuts the time the wafer spends outside its band from 37.1 to 3.5 s per wafer and the hottest wafer from 67.5 to 62.0 °C. The relative gain array is within 0.073 of the identity, so the four loops cannot be the weak point, and the gain must come from the bias-to-wafer term the array cannot see. Replace the four loops with the predictive controller." Find the flaw, and say how you would check it in the app.