Encyclopedia of Systems and Control
DOI 10.1007/978-1-4471-5102-9_255-1

Run-to-Run Control in Semiconductor Manufacturing

James Moyne*
Mechanical Engineering Department, University of Michigan, Ann Arbor, MI, USA

Abstract

Run-to-run (R2R) control is a form of adaptive model-based process control that can be tailored to environments where the process is discrete, dynamic, and highly unobservable; this is characteristic of processes in the semiconductor manufacturing industry. It generally has, at its roots, a rather straightforward approach to adaptive model-based control such as a first-order linear plant model with moving average weighting applied to adapt the (zeroth-order) constant term in the model. Most of the complexity of R2R control science lies and will continue to lie in extensions to support practical application of R2R control in semiconductor manufacturing facilities of the future; these extensions include support for weighting and bounding of parameters, run-time modeling of a large number of disturbance types, and incorporating prediction information such as virtual metrology and yield prediction into the control solution.

Keywords and Phrases Run-to-run control • R2R control • Advanced Process Control (APC) • Wafer-to-wafer control • Model-based control • EWMA control • Adaptive control • Single-threaded control • Virtual metrology • Yield prediction • Feed-forward and feedback control

Introduction

The semiconductor manufacturing industry involves the processing of semiconductor "wafers" using a variety of physical and chemical processes to produce dies or "chips" that contain a number of nanometer size features organized in layers. As feature sizes shrink, the industry must innovate to maintain acceptable product yield and throughput. One effective dimension of innovation that has been utilized since the early 1990s is model-based process control. The use of this technology in semiconductor manufacturing has been largely industry specific due to unique industry requirements and been given the name "run-to-run (R2R) control."

R2R control is defined as "…a form of discrete process and machine control in which the product recipe with respect to a particular machine process is modified ex situ, i.e., between machine 'runs,' so as to minimize process drift, shift, and variability" (Moyne et al. 2000). (The "recipe" is the group of process settings for a process or process step, e.g., temperature, flow, and pressure.) The term "R2R control" was coined in the early 1990s in the semiconductor industry as the industry struggled to come up with mechanisms to keep critical semiconductor manufacturing processes such as chemical vapor deposition (CVD), chemical mechanical polishing (CMP), and reactive ion etching (RIE) under control. The processes are highly unobservable and are subjected

*E-mail: moyne@umich.edu

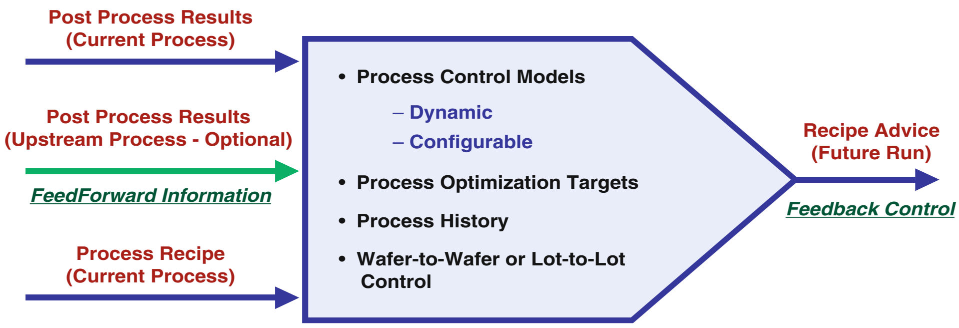

Fig. 1 Input/output structure of a typical R2R control solution

to a number of disturbances. However, many of these disturbances can be modeled or tracked as they create measurable shifts in the process (e.g., after a maintenance operation) or gradual drifts in the process (e.g., chamber wall "seasoning" of an etch process over time, resulting in polymer buildup on chamber walls, causes changes to the operational effectiveness of the tool). Process and product quality is generally assessed through metrology measurements made ex situ, i.e., after the process is complete; examples of post-process metrology parameters are wafer average deposited or removed film thickness and film uniformity. R2R control generally uses statistically developed models of tool process operation updated or "tuned" with process metrology feedback information on a "run-to-run" basis to keep the process under control and process quality high, in the face of these process drifts and shifts, as shown in Fig. 1. Note that the granularity of control could be wafer-to-wafer, or batch-to-batch ("lot-to-lot"), etc.

Run-to-Run Control Approach

Because the processes are highly unobservable and dynamic, rather simple model forms are usually employed with filtering techniques used to track process shift and drift. The most commonly utilized R2R controller in the industry is the exponentially weighted moving average (EWMA) controller. The algorithm uses a linear model with an additional constant term. (Equations will use the following notation: arrays of vectors will be capitals, vectors will be lower case, and indexing within a vector or matrix will be lower case with subscripts. In addition, the special subscript "t" will be reserved for time or run number information.)

$$Y = Ax + c \tag{1}$$

where:

y = System output,
x = Input (Recipe),
A = Slope coefficients for equation,
c = Constant term for linear model.

Each output represents a target of control (usually measured by pre- and post-process metrology tools), and each input represents an adjustable parameter in the recipe.

$$y_1 = a_{11}x_1 + a_{12}x_2 + \ldots a_{1m}x_m + c_1$$

$$\ldots$$
$$y_n = a_{n1}x_1 + a_{n2}x_2 + \ldots a_{nm}x_m + c_n \quad (2)$$

The models are generally developed by executing a design of experiments (DOE), where the process area is explored with respect to the allowed variation of the process inputs by processing wafers with various input settings (see, e.g., Box and Draper 1987). Statistical packages are then used to determine the base model of the form described in (1) at the normal process operating point. As the processes are dynamic, the base model is updated on a "run-to-run" basis to compensate for model error. The algorithm operates under the assumption that the underlying process is locally approximated by a first-order linear polynomial model in the form of equation (1) and that this polynomial model can be maintained near a local optimal point solely by updating the constant term "c."

The control process involves updating the model and then using that model to compute a recipe update. The model is updated by first comparing the actual process output, Y_t, to the model-predicted process output, AX_t. Using an EWMA filtering technique as an example, the constant term, c_t can be updated as follows:

$$c_t = \alpha(y_t - Ax_t) + (1 - \alpha)c_{t-1} \quad (3)$$

where α is a weighting factor between 0 and 1, often called a "forgetting factor." Note that because of the additive nature of the EWMA series, the C_t calculation only requires knowledge (and storage) of the previous run measurements; this, combined with its relative simplicity, led to the widespread adoption of EWMA as the R2R controller filter of choice in this industry during the 1990s and early 2000s.

Once the model is updated, the process recipe is calculated. Since there are generally more inputs that can be tuned than outputs measured, the process is underdetermined and there is an infinite solution space. Approaches such as Lagrange multipliers are used to determine the solution that is closest to the previous solution (Moyne et al. 2000).

Many extensions and alternatives to this basic approach have been developed and deployed over the past 10 years. These include (1) the replacement of EWMA filtering with other approaches such as the more general Kalman filtering, (2) explicitly modeling drift (termed "predictor corrector"), (3) modeling updates to first-order terms (in the "A" matrix), and (4) leveraging phenomenological models that capture process knowledge in equation forms, customized and tuned with statistical data. Perhaps the most important extensions to the basic approach involve addressing the practical issues associated with control systems application in this area. For example, providing capabilities for addressing bounding, weighting, and granularity (e.g., integer) of input and output settings often requires much more programming effort than supporting the core algorithm (Moyne et al. 2000).

Current Status and Future Extensions

Over the past 10 years, R2R control has evolved from a value-added capability applied to a few processes, to a required component to achieve cost and productivity competitiveness in most

DOI 10.1007/978-1-4471-5102-9_255-1

processes in the semiconductor manufacturing industries (ITRS 2014). As part of this evolution, a number of common trends in the R2R control space have emerged:

Support for fab-wide reusable and reconfigurable solutions for R2R control: As the benefits of R2R control were proven across multiple processes in semiconductor fabrication facilities, the focus turned to reusable and reconfigurable integrated "fab-wide" solutions for R2R control. The event-based capabilities described in Chapter 9 of Moyne et al. (2000) were leveraged to provide these solutions as they allow for integration and configuration of R2R control solutions to the particular application environment. This event-based approach has also been used to integrate R2R control with other capabilities such as fault detection and classification (FDC), work scheduling, and "virtual metrology" (see below), to provide another level of benefits towards improved product yield and throughput (Moyne 2004, 2009; Khan et al. 2007).

Movement to more granular control: The evermore stringent requirements on product quality are being addressed in large part by a movement from batch-level control (often called "lot-based control" in this domain), to wafer-level control (usually called "wafer-to-wafer" (W2W) control), to within-wafer (WIW) control. Although the granularity has changed, the basic approach to control has not. It is important to note that the improvement in quality associated with this trend results mostly from the use of pre-(process) metrology to reject incoming product disturbances, rather than post metrology to address the dynamics of the plant model (Moyne et al. 2000; ITRS 2014).

Support for control across multiple recipes using "single-threaded control": Semiconductor manufacturing process control systems are characterized by a number of disturbance types that usually can be modeled as independent from the base process model and from each other. Perhaps the most common type of disturbance that is addressed is recipe or product change. When there is a change in product and related product recipe, a single-process model must be adjusted to capture this disturbance while maintaining knowledge of process drift and/or shift. Oftentimes this process disturbance can be modeled as a shift to the overall process. Thus, the process model of equation (1) can be adjusted to the following:

$$\mathrm{Y} = \mathrm{Ax} + \mathrm{c}_1 + \mathrm{c}_2 + \mathrm{c}_3 + \ldots + \mathrm{c}_\mathrm{n} \tag{4}$$

where:

$c_1, c_2, \ldots c_{n-1}$ = constant terms associated with modeled disturbances such as product
c_n = constant term associated with process dynamics (drift and shift)
$c_1 + c_2 + \ldots c_n$ = c in Eq. (1)

Approaches have been devised for the assessment of c_i associated with a particular disturbance type (Edgar et al. 2004; Zou 2013); the result is that a single-control model can be used across multiple product recipes and other disturbance types.

Enhancing R2R control with "virtual metrology": Ex-situ metrology plays a crucial role in semiconductor manufacturing as it is often the only source of product quality data before and after a process. However, given its high capital equipment cost and cycle time impact on critical processes, optimizing metrology by minimizing wasteful use and optimizing measurement value is important. Virtual metrology (VM) is a new technology rapidly gaining acceptance in the marketplace as an efficient and cost-effective way to optimize and augment metrology value. VM is a modeling and metrology prediction solution whereby process and product data, such as in situ fault detection (FD) information and upstream metrology information, is correlated to post-

Encyclopedia of Systems and Control
DOI 10.1007/978-1-4471-5102-9_255-1

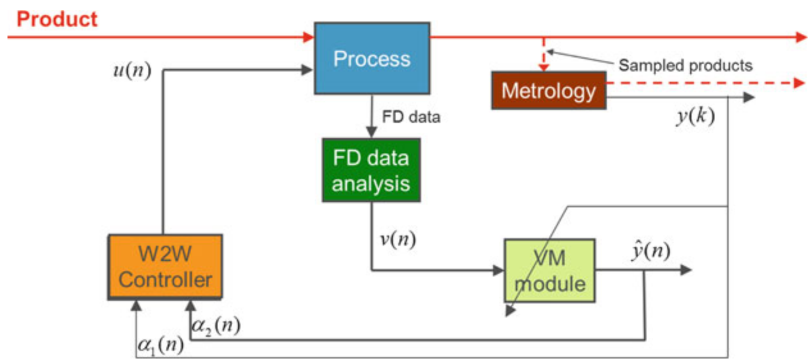

Fig. 2 Virtual metrology enhanced R2R control

process metrology data. This same data can then be used to predict metrology information when conventional metrology data is not available (Khan et al. 2007; Cheng et al. 2011).

One of the uses of VM that is expected to become prominent over the next decade is in support of enhanced R2R control. As shown in Fig. 2, fault detection (FD) summary information is used along with adaptive VM modeling to predict metrology information. The VM predictions are then used to fill in the measurement gaps in feed-forward and feedback control thus enabling wafer-to-wafer or even within-wafer control. One of the research challenges is to optimally tune the control to best utilize both the real and predicted metrology information. This requires that VM data contain information on predicted measurement data quality (Khan et al. 2007).

$u(n)$ Tunable process inputs
$v(n)$ FD summary information
$y(k)$ Metrology measurement data for measured wafers
$\hat{y}(n)$ Predicted metrology measurements for all wafers
$\alpha_1(n)$ Feedback filter coefficient for feedback of measured data
$\alpha_2(n)$ Feedback filter coefficient for feedback of predicted data

Movement towards interprocess and eventually fab-wide control: The generally accepted vision of the future of advanced process control (APC) in general is a fabrication-wide fully integrated solution that incorporates all of the APC capabilities (R2R control, FDC, fault prediction, and statistical process control) as well as predictive capabilities such as predictive scheduling, predictive maintenance, virtual metrology, and predictive yield (ITRS 2014). Opportunities for research and development exist with the integration of these technologies, especially as the powers of the predictive domain are tapped. For example, it is expected that R2R control will eventually incorporate predicted yield as a target with feedback to multiple coordinated process controllers (Moyne and Schulze 2010). Thus, the future of research in R2R control, while evolving, should remain strong in the coming years.

Summary and Future Directions

R2R control is a form of adaptive model-based process control that is tailored to environments where the process is discrete, dynamic, and highly unobservable; this is characteristic of processes in the semiconductor manufacturing industry. R2R control has evolved from a strictly research effort in the early 1990s to a required facility-wide capability in all of semiconductor manufacturing. It generally has, at its roots, a rather straightforward approach to adaptive model-based control. Most of the complexity of R2R control science lies and will continue to lie in extensions to support practical application of R2R control in semiconductor manufacturing facilities of the future.

The science of R2R control will continue to expand as the academic and industry communities look to incorporating capabilities that will allow R2R control to continue to be an integral part of the fabrication facility of the future. One key research direction over the next decade is the development of approaches for incorporating virtual metrology and yield prediction into control solutions. Other focus areas will likely include hybrids of R2R control and continuous process control, learning mechanisms for single-threaded control in "high-mix" environments where there are a large number of disturbances that should be modeled, phenomenological R2R control models, and model libraries that combine stochastic information with process physics and chemistry knowledge, control solutions that are more directly optimized to financial parameters such as yield and throughput, and R2R control solutions that incorporate other analysis capabilities, such as FDC, either algorithmically or via event-based control rule approaches. Each of these topics provides significant opportunity for research as well as benefit in application to semiconductor manufacturing facilities.

Cross-References

▸ Adaptive Control, Overview
▸ Controllability and Observability
▸ Event-Triggered and Self-Triggered Control in Hybrid Systems
▸ Experiment Design and Identification for Control
▸ Fault Detection and Diagnosis
▸ Kalman Filters
▸ Moving Horizon Estimation
▸ Nominal Model-Predictive Control
▸ Robust Model-Predictive Control
▸ Stochastic Model-Predictive Control

Bibliography

Advanced Process Control (APC) Conference Proceedings (2000–2014), titles available at http://www.apcconference.com

Box GEP, Draper NR (1987) Empirical model-building and response surfaces. Wiley, New York

Cheng F-T et al. (2011) Benefit model of virtual metrology and integrating AVM Into MES. IEEE Trans Semicond Manuf 24(2):261–272

Edgar TF, Firth SK, Bode C (2004) Multi-product run-to-run control for high-mix fabs. (session keynote) *AEC/APC Asia*, Hsinchu, (2004), (available at http://140.113.156.45/files/reference/2nd%20AEC-APC%20Symposium%20Asia/APCAEC/presentations/session_KeynoteInvited/Edgar_Thomas.pdf)

International Technology Roadmap for Semiconductors (ITRS), 2014 Edition, Semiconductor Industry Association. Available at http://www.itrs.net. (See especially "Factory Integration" chapter)

Encyclopedia of Systems and Control
DOI 10.1007/978-1-4471-5102-9_255-1

Khan A, Moyne J, Tilbury D (2007) Fab-wide control utilizing virtual metrology (invited). IEEE Trans Semicond Manuf-Spec Issue on Adv Process Control 20(7):364–375

Moyne J (2004) The evolution of APC: the move to total factory control (invited). Solid State Technol 47(9):47–52

Moyne J (2009) A blueprint for enterprise-wide deployment of advanced process control. Solid State Technol 52(7):35–37

Moyne J, Del Castillo E, Hurwitz A (2000) Run-to-run control in semiconductor manufacturing. CRC, Boca Raton

Moyne J, Schulze B (2010) Yield management enhanced advanced process control system (YMeAPC): part I, description and case study of feedback for optimized multi-process control. IEEE Trans Semicond Manuf Spec Issue Adv Process Control 23(2):221–235

Zou J (2013) Method and system for estimating context offsets for run-to-run control in a semiconductor fabrication facility. United States Patent, Patent Number US 8,355,810 B2 (Filed, Jan 2010; Issued, Jan 2013)