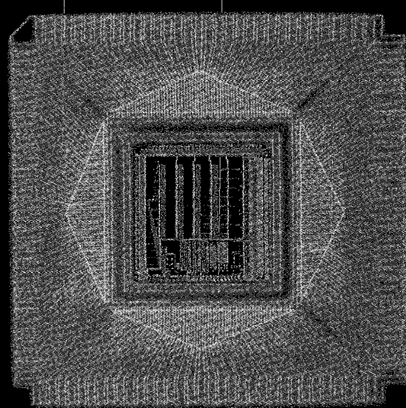
Monitoring and Control of Semiconductor Manufacturing Processes

Suttipan Limanond, Jennie Si, and Kostas Tsakalis

Semiconductor manufacturing is a complex and highly integrated industry involving several intermediate processes. The semiconductor market is currently driven toward low cost VLSI chips with higher circuit density. For example, in the year 2001, dynamic random-access memory (DRAM), which has historically employed a cutting-edge technology among various semiconductor products, is expected to reach a density of 1.7 Gbits per chip at a cost of only 0.03 cents per Kbit, whereas the microprocessor is predicted to have a density of 10 million logic transistors per squared-centimeter at a cost of 1 cent per 1000 transistors. By comparison, the corresponding numbers for 1997 were 267 Mbits per chip at the cost of 0.12 cents per Kbit for DRAM, and 3.7 million logic transistors per squared-centimeter at the cost of 3 cents per 1000 transistors for microprocessor.



The U.S. semiconductor community faces increasingly difficult challenges as it moves into production at feature sizes approaching 100nm. The grand challenges facing the research and development community are identified as the ability to continue affordable scaling, affordable lithography at and below 100nm, discovery of new materials and structures, GHz frequency operation on- and off- chip, as well as advanced metrology and test procedures [1]. The *National Technology Roadmap for Semiconductors* [1] has called for a concerted effort from various research communities and industries to advance technologies in several disciplines, ranging from device design to factory integration. Innovations are needed in design and testing, lithography, and interconnect, as well as assembly and packaging. To meet these challenges, an important component in the advancement of VLSI manufacturing is improved techniques in process monitoring and control as part of factory integration [1]. In the next two sub-sections, we first give a brief introduction of important processes in VLSI fabrication, and then elaborate on the role of monitoring and control technologies in advancing VLSI fabrication.

Introduction

Several chemical-based processes are involved in building VLSI chips. While their precise descriptions and sequences vary, depending on the desired device characteristics and the material used in fabrication, the typical VLSI fabrication steps include metal preparation, oxidation, photolithography, etch, diffusion, and deposition processes, as shown in Fig. 1. Most devices require multiple steps through the same process at different stages. Critical processes directly influencing the critical dimension (CD) and geometry of VLSI chips are photolithography, etch, chemical vapor deposition (CVD), and chemical mechanical polishing (CMP) process, which are briefly summarized as follows:

- *Photolithography* serves to transfer the pattern of the desired circuit ("mask") onto a wafer for the purpose of selective etching; the process includes photoresist coating, the transfer of mask onto the wafer ("direct wafer stepping"), and the UV light exposure step. The wafer is coated with photoresist material that softens when exposed to the UV light so that the wafer surface underneath it can be removed by an etching process. The mask is then applied to the photoresist surface to cover selected area from UV light exposure, which, in turn, determines the desired circuit pattern. Photolithography is the key technology driver for the semiconductor industry, responsible for the recent industry's rapid growth, and a significant economic factor representing over 35% of the chip manufacturing cost [1].
- *Etching* is a process for removing undesired material from the wafer. Performed immediately after photolithography, etching provides selective wafer surface removal from the area not covered by photoresist material. Etching techniques typically used in VLSI fabrications are plasma and Reactive Ion Etching (RIE). The main difference between

The authors are all with the Department of Electrical Engineering, SSERC, Arizona State University, Tempe, Arizona. Limanond's and Si's research is supported in part by NSF under grant ECS-9553202, by EPRI under grant RP8015-03, and by Motorola; Tsakalis's research is supported in part by SEMY Engineering Inc. under grant No. SEI-960815-002.

the two techniques is that plasma etching relies on a chemical reaction of the feed gas to remove unwanted material, while RIE uses both the chemical reaction of the feed gas and a physical reaction due to ion bombardment to remove material and operate at a lower pressure. Etching is an important process for providing reliable device isolation and interconnections within the VLSI chip.

- *Chemical Vapor Deposition* (CVD) is a process used to deposit a thin layer of material on the wafer. The reactant gases are introduced into the chamber and undergo chemical reactions with the heated wafer surface to form a thin film. Different CVD processes are currently used for VLSI fabrication, including Atmospheric Pressure CVD (APCVD), Low Pressure CVD (LPCVD), and Plasma Enhanced CVD (PECVD). The APCVD operates at a moderate wafer temperature that allows for fast deposition, but suffers from poor coverage and particle contamination. LPCVD, on the other hand, offers excellent purity and good coverage, at the expense of reduced material strength due to high operating temperatures. Finally, PECVD operates at the lowest wafer temperature by incorporating the RF-induced energy in the plasma, which allows for the fastest deposition rate. However, similar to APCVD, it suffers from contamination problems. CVD is another crucial technology for semiconductor industry for providing conformal-coating solutions for higher minimum feature density of the chip.
- *Chemical-Mechanical Polishing* (CMP) is a process of removing 'high' material on the wafer surface (or planarization) through mechanical and chemical reaction. This involves the use of a polishing pad and a slurry of abrasive material. CMP is typically used for dielectric and metal planarization so as to enable satisfactory imaging of polysilicon and metal layers.

Due to their significance for VLSI manufacturing, these processes have been under extensive investigation from different research viewpoints, including process monitoring and control.

Monitoring and Control for VLSI Fabrication

The recent surge of interest in the area of monitoring and control of semiconductor manufacturing process arises from the need to fabricate VLSI chips with higher feature densities and larger chip sizes. Until recently, however, monitoring and control of the various processes was fairly rudimentary relative to the state-of-the-art. It primarily relied on practical engineering experience and basic statistical process control (SPC) methods. Real-time feedback control was confined to simple single-loop controllers (e.g., PID). The lack of development of more sophisticated, high-performance control systems was due to the following:

A. Poor Process Understanding: VLSI fabrication processes are primarily based on chemical and physical reactions of small particles, which are generally complex and nonlinear. The analytical (first-principle) models currently available are often too complex and computationally intensive to be used for designing control systems. For example, in a $CF_4/O_2/H_2$ plasma system, there are 60 possible reactions that can occur [8]. Moreover, VLSI fabrication processes consist of several processing steps, such as PECVD (plasma generation and deposition step), etching (plasma generation and etching step), and photolithography (photoresist coating, photomask transfer, and develop steps). The internal interaction of various process variables leads to poor process understanding, hampering the application of systematic control design methods.

B. Lack of Appropriate Sensors and Measurement Techniques: During Integrated Circuit (IC) fabrication, wafers are processed inside a closed chamber on a cassette-to-cassette basis (e.g., one cassette may contain 24 product wafers; CVD/Diffusion furnaces can process four cassettes at a time). Furthermore, critical wafer parameters (or *process metrics*) such as the film thickness remaining after etching or line-width of the photomask, can only be determined using contact (as opposed to non-contact) measurements that can destroy the wafer surface and the associated electrical characteristics. For this reason, real-time control is often limited to variables (such as temperature, RF power, pressure, etc.) that affect the critical wafer parameters indirectly, but are easily measured.

These two factors led to a situation where monitoring and control of critical wafer parameters was performed outside the processing chamber in a non real-time fashion after a cassette of wafers had been processed; see Fig. 2 for a block diagram associated with this control strategy. More precisely, after the processing of a cassette, the critical wafer parameters are measured and used to update parameters of a (typically static) model, which relates a set of input variables (or *recipes*) to the critical wafer parameters. The updated model is then used to generate a suitable recipe correction, corresponding to the desired wafer parameters. This control design strategy is commonly referred to as Run-to-Run (RtR) Control, which is discussed in more detail subsequently.

These shortcomings have been actively investigated by both industry and research communities. Preliminary results are

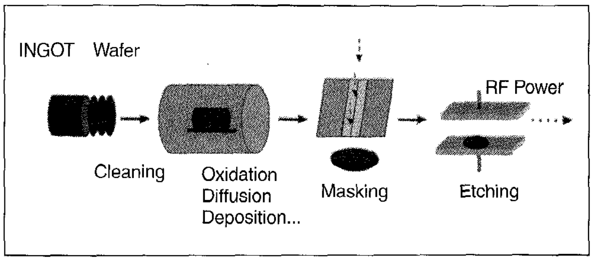


Fig. 1. Schematic of key semiconductor manufacturing processes relevant to system control.

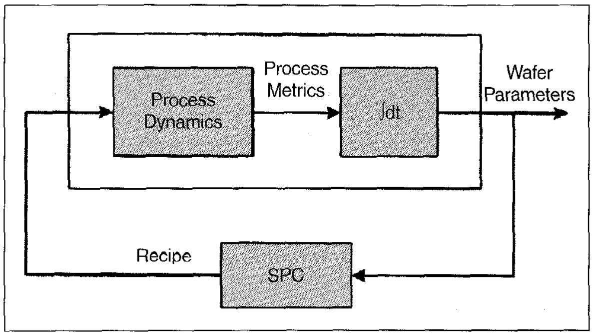


Fig. 2. Block diagram of conventional Run-to-Run control.

available for several VLSI fabrication processes in terms of modeling, control, and metrology (*i.e.*, instrumentation and measurement techniques). It should be mentioned that direct *in-situ* measurement of CD variables is still not practical. Nevertheless, optical-based sensors (e.g., optical emission spectroscopy (OES) and laser interferometry (LIF), see [15], [26], [27] and references therein) have been invented to measure various chamber-status process variables (or *in-situ* variables) such as temperature and pressure. This, together with the development of appropriate models that relate *in-situ* variables to the process metrics, leads to the integrated RtR and real-time control strategy [6], [8], [26], which can be designed in two steps. First, the RtR control uses these models to determine *in-situ* process variables corresponding to the desired critical wafer parameters. Second, a real-time control design is implemented to drive the chamber states to the set-point obtained from the RtR control. Notice that the latter step also requires a model relating manipulated process variables to the *in-situ* process variables. The associated block diagram is illustrated in Fig. 3. Table 1, on the other hand, lists the critical process variables, chamber state variables, and typical input variables for various VLSI manufacturing processes.

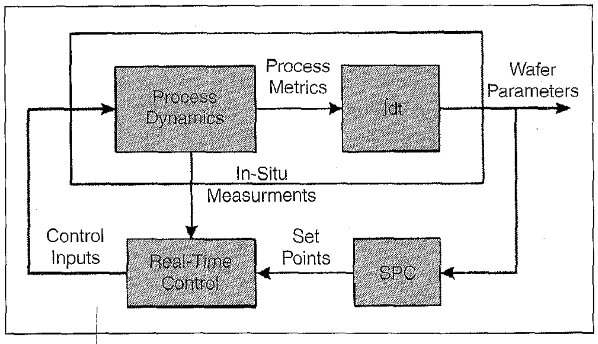


Fig. 3. Block diagram of the Integrated Run-to-Run and Real-Time control.

Organization

In this article, we report some representative results of emerging modeling and control techniques for VLSI fabrication. The underlying processes considered include photolithography, etching, CVD, and CMP. We would like to use these examples to demonstrate what could be done by the control systems community to advance the current practice of semiconductor manufacturing industry. In the mean time, the relevant issues of monitoring and control in semiconductor manufacturing are revealed to some extent through the examples.

In the next section, we discuss various modeling techniques that have been studied and, possibly, implemented for different VLSI processes. Following that, we consider monitoring and control issues for semiconductor manufacturing processes with emphasis on process monitoring, the Run-to-Run and real-time control.

In the fourth and fifth sections, we present two case studies based on research activities currently taking place at the Center for Systems Science and Engineering, Arizona State University. The first concerns a novel wafer-to-wafer end-point detection scheme for reactive ion etching. The second is an integrated design (from data to control) of a real-time multivariable temperature controller for CVD and Diffusion furnaces, a design that has been used successfully in about two hundred controller installations world-wide. Inner-outer loop control structures combining RtR and real-time controllers have special importance in semiconductor manufacturing due to the difficulty in measuring critical wafer parameters. In this way real-time control strives to maintain a repeatable process while run-to-run aims at adjusting the variable set-points so that product specifications are met. In particular, a practically desirable connection of these seemingly disparate controllers is that their designs are based on the following common objectives and requirements: 1) the ability to obtain models from data, tailored to a specific piece of equipment and

Table 1. Typical process metrics, input settings, and intermediate process variables for VLSI fabrication processes.

	Process Metrics	Intermediate Variables	Input Settings
Etching (Plasma and RIE)	etch rate etch anisotropy etch uniformity etch selectivity	dc bias gas species concentration pressure temperature	rf power throttle position gas flow
PECVD	deposition rate film stress film uniformity film purity	dc bias gas species concentration pressure temperature	rf power throttle position gas flow
Photolithography Spin coat and Bake step Exposure step Develop step	 photoresist thickness photoactive concentration (PAC) exposed PAC pattern line width		 spinning speed baking temperature exposure dose
CMP	removal rate uniformity		spinning speed pressure force

conforming to industrial needs of quick turn-around time; and 2) production-grade design and implementation.

In the final section, we summarize the current state of research in the area of monitoring and control of semiconductor manufacturing process and discuss some of the future research activities in the area.

Modeling of Semiconductor Manufacturing Processes

In this section we present a summary of various monitoring and control related models that have appeared in the literature, including regression, linear dynamical, physical, and neural network-based models. All have found successful applications to CMP, CVD, and etching processes. The role of these models in VLSI fabrication is two-fold. First, they serve to provide estimates of various *in-situ* process variables, for which the associated sensor and instrumentation technology is neither available nor suitable for implementation, for the purpose of process monitoring and diagnosis. Second, they serve as nominal models for traditional control system designs. Briefly summarized, process modeling from data consists of the following steps: selection of the model input, and possibly output; data collection; model structure selection; and, if necessary, model reduction. Widely used models that have appeared in VLSI fabrication literature are listed in Table 2.

Static Regression Models

Static models have found numerous applications in monitoring and control of semiconductor processes [2]–[5], [36], [40], [41]. Although, in principle, the regression model can be polynomial of any degree, a linear model often performs satisfactorily and is widely accepted as a standard model for statistics-based monitoring and control. For monitoring, a static model is typically implemented in combination with a conventional SPC chart (e.g., Shewhart and CUSUM chart) or a multivariate chart to generate alarms when the statistical properties of the monitored process variables deviate significantly from the set-point value [2], [3], [36]. Further, in the RtR control framework, the model can be used to determine an appropriate recipe adjustment.

Typically, a design-of-experiments is implemented to determine the appropriate process variables, those that have the greatest influence on the system outputs. For a modeling task involving large sets of data, principal component analysis or partial least-squares analysis can be employed to reduce the data correlation [3], [5], [7]. A linear regression model has the form:

$$y = Ax + d \tag{1}$$

where x and y are the input and output vectors, and A is a matrix containing the model parameters. Here, d is a constant term. This simple model is rather attractive for several reasons. First, the constant term d can be used to represent the slow process drift, common in semiconductor manufacturing processes. The slow drift is primarily due to chemical residue buildup inside the processing chamber and/or the loading effect of various electrical equipment. Exponentially weighted moving average (EWMA) adaptation is typically employed to adjust the term d [40]–[43]. Further, occasional shifts due to material variations and scheduled maintenance operations can be captured by the variation of the model parameters in A (see [2] for further details). Finally, the simplicity of the model allows for fast control computations. In fact, it has been successfully used in RtR control to generate suitable recipe adjustments for etching [2], [3], [5], photolithography [36], and CMP [40]–[42].

Linear Dynamical Models

Linear dynamical models are most commonly used in a generic control system design, mainly due to the fact that various systematic control design algorithms are readily available. In semiconductor manufacturing process, these models have been employed primarily for *real-time* control design [8]–[11], [26], [29] but they have also found a limited application in process monitoring. The selection of the model inputs and outputs is roughly similar to the one for static regression models.

A linear dynamical model is typically obtained by using step response and/or parameterization methods. In the former, the step response (or frequency response) plots are obtained so as to determine the associated gain and time constant for each single-input single-output channel of the process. This modeling method has been successfully used to determine a control-related model for RIE [6], [8], [9] and PECVD [26]. Another modeling technique relies on the parameterization of the system under consideration, and the minimization of a suitable error criterion for input-output data obtained during an identification experiment [44]. In particular, a simple parametrization arises from the so-called equation error; although this parameterization and estimation does not always lead to good prediction models, it is attractive for its computational simplicity as well as the control-oriented properties of the associated error system. It has provided a quick and successful mechanism for the temperature control problem of diffusion/CVD furnaces [29], presented in the second case study below.

Physical Models

Physical, or first-principle, models of VLSI manufacturing processes are complex and difficult to build. They are therefore

Table 2. Representative modeling publications for various VLSI fabrication processes.

	Etch	CVD	Photolithography	CMP
Static Model	[2], [3], [4]		[36]	[40], [41], [42]
Linear Model	[8], [9], [11], [12], [13]	[26], [28], [29]		
Physical Model	[12], [14], [15]	[30], [31], [32], [33]	[37], [38], [39]	
NN (MLP) Model	[15], [16], [17], [18], [20], [21], [23], [24], [25]	[34], [35]		[40]

Table 3. Representative monitoring and control publications for various VLSI fabrication processes.

	Etch	CVD	Photolithography	CMP
RtR Control	[2], [4], [6		[36]	[40], [41], [42]
Conventional Control Design	[8], [9], [11], [12], [14], [24]	[26], [29], [32], [33], [35]	[37], [38], [53]	
Monitoring	[3], [5], [13], [15], [17], [20], [21], [23], [45], [46], [47], [48]	[33], [34]	[54]	[42]

typically used for providing simulation studies under various operating conditions and settings or for the monitoring of process variables, see [15], [30], [39], [55] for examples. Nonetheless, with suitable assumptions and a model reduction step, simplified first-principle models can serve as nominal models for control design as well, as shown in [12] and [14]; [31]–[33]; and [37], [38], and [53], for, respectively, etching, CVD, and photolithography.

Neural Network-Based Models

Neural networks have recently found many applications in VLSI fabrication, especially for providing estimates of critical wafer parameters for real-time monitoring purposes. Briefly summarized, a neural network is an interconnected system of "neuron" units interacting with each other through their "weight" connections. These neurons are activated by the weighted sum of incoming signals (from other neurons) and fire outputs according to the associated "activation" function, which can be of sigmoid or Gaussian type. A special class of neural networks, known as multilayer perceptron (MLP), is commonly used as an approximating model for nonlinear dynamical systems, due to its good approximating property, which was rigorously established in [56]. An additional attractive feature lies in the fact that building an MLP model requires less training data than building a statistical model [16], [17].

MLP modeling requires similar building steps as statistical and linear dynamical modeling. The MLP structure is chosen based on the number of model inputs and outputs, typically with one or two internal layers. In order to determine a set of connecting weights that provide the least fitting error, a gradient descent training algorithm, such as backpropagation, can be implemented. However, better convergence properties can be achieved by using the Levenberg-Marquardt algorithm or other improved backpropagation training methods [18], [19]. For highly correlated sets of data, principal component analysis can be applied to reduce the correlation. In VLSI fabrications, MLP's have been used for process modeling for CMP [40], CVD [34], [35], and etching [16]–[18], [20]–[22], [24], [25].

Monitoring and Control of Semiconductor Manufacturing Processes

Semiconductor manufacturing processes suffer from a high level of system nonlinearity, internal state interaction and a lack of real-time measurement tools that hamper the effective application of the existing comprehensive diagnostic and control design procedures. Furthermore, these processes are often corrupted by slow drift and other process shifts due to, for example, maintenance operations. In this respect, several control schemes have recently been devised to address these issues. In this section we discuss monitoring schemes as well as control design schemes such as RtR and real-time control (see also Table 3).

Monitoring

Process monitoring serves to provide information concerning critical process variables in the form of real-time sensor readings or the associated statistics. Traditionally, process monitoring was done in an off-line and non real-time fashion, based on measurements available after several batches of wafers had been processed. Though sufficient for a simple systems, this type of process monitoring, is rather inadequate for complex processes such as etching and CVD. This leads to the development of different optical-based measurement techniques, such as OES and LIF for certain critical process variables (see [15], [26], [27], [46], [49] and references therein). In the absence of suitable sensors, two alternatives are available. First, state estimation algorithms such as Kalman filter [54], extended Kalman filter [13], [37], and, recently, a jump linear filter [48], have been implemented to estimate internal variables of the process from the available measurements. Second, sophisticated models can be built off-line and implemented on-line so as to provide real-time approximation or prediction of the critical variables. For this purpose, neural networks are instrumental and widely used in model-based monitoring of etch processes [5], [15], [18], [20], [21].

In addition to real-time control, another important application of process monitoring is in the detection of process end-point. In that case, the ability to correctly specify process end-points is of significant importance, since, for example, it can reduce production scrap from over or inadequate etching and increase throughput reproducibility. Briefly summarized, an end-point detection scheme monitors the so-called "end-point detection signals" obtained from optical-based sensors (i.e., OES and LIF) until they match a decision criteria for end-point [46], [49]. Another technique involves the use of MLP to compute the etch time based on various process measurements [45], [50], [52]. An example of the end-point detection schemes that involves the use of MLP and optimization algorithm [23] is discussed in the next section.

Run-To-Run Control

Run-to-Run (RtR) control is an integrated monitoring and control algorithm for semiconductor manufacturing processes which incorporates various modeling and (statistics-based) control design techniques in a systematic fashion. Traditional RtR control is based on static modeling and SPC charting techniques (RtR control is sometimes referred to as "integrated SPC") to provide monitoring and control functionality [2], [36], [41]. Recently, neural network-based modeling and control techniques have also been incorporated into the RtR control framework to

DOCKET ALARM

Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

Real-Time Litigation Alerts

Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

Advanced Docket Research

With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

Analytics At Your Fingertips

Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

LAW FIRMS

Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

FINANCIAL INSTITUTIONS

Litigation and bankruptcy checks for companies and debtors.

E-DISCOVERY AND LEGAL VENDORS

Sync your system to PACER to automate legal marketing.

WHAT WILL YOU BUILD? | sales@docketalarm.com | 1-866-77-FASTCASE

fastcase® Smarter legal research.