J Sched (2015) 18:195–205
DOI 10.1007/s10951-014-0381-1

Integration of scheduling and advanced process control in semiconductor manufacturing: review and outlook

Claude Yugma · Jakey Blue · Stéphane Dauzère-Pérès · Ali Obeid

Received: 21 May 2012 / Accepted: 25 April 2014 / Published online: 25 May 2014

Abstract Scheduling in semiconductor manufacturing is of vital importance due to the impact on production performance indicators such as equipment utilization, cycle time, and delivery times. With the increasing complexity of semiconductor manufacturing, ever-new products and demanding customers, scheduling plans for efficient production control become crucial. Scheduling and control are mutually dependent as control requires information from scheduling, for example, where jobs are processed, and scheduling requires control information, for example, on which equipment operations can be processed. Based on a survey of the literature, this article proposes a review and an outlook for the potential improvements by binding scheduling decisions and information coming from advanced process control systems in semiconductor manufacturing.

Keywords Scheduling · Dispatching · Advanced Process Control (APC) · Semiconductor manufacturing

C. Yugma (✉) · J. Blue · S. Dauzère-Pérès · A. Obeid
Department of Manufacturing Sciences and Logistics, Center of Microelectronics of Provence, École Nationale Supérieure des Mines de Saint-Étienne, Site Georges Charpak, 880 route de Mimet, 13541 Gardanne, France
e-mail: Yugma@emse.fr

J. Blue
e-mail: Lan@emse.fr

S. Dauzère-Pérès
e-mail: Dauzere-Peres@emse.fr

A. Obeid
e-mail: Obeid@emse.fr

1 Introduction

The semiconductor manufacturing industry is the fastest evolving and most highly competitive industry in the world. According to Moore's law, the number of transistors on integrated circuits doubles approximately every 2 years, and consequently new technologies appear (Moore 1965). The semiconductor industry is characterized by an increasing complexity of manufacturing processes in terms of elementary operations as well as of the number of constraints, finer geometries to realize on chips, increasing degrees of automation combined with equipment and infrastructure costs, high renewal rates of technologies that lead to a rapid obsolescence of products, an ever-increasing pressure on the cost of wafers due to worldwide competition, and high customer requirements in terms of quality. Furthermore, the current economic situation threatens the existence of companies where only the best survive. To succeed, companies must pay constant attention to manufacturing processes (flows), establish better and more intelligent controls at various steps of the fabrication process, and develop new scheduling techniques. For companies to remain competitive, innovations with acceptable cost limits, cycle time minimization, throughput maximization, and yield improvement are vital.

The overall semiconductor manufacturing process can be classified into four main stages: wafer fabrication, wafer probing, assembly or packing, and final test. Among these stages, wafer fabrication consists of the most complicated process flow and is the main focus of our research. In wafer manufacturing systems, a *lot* is a moving entity. Each lot contains a certain number of wafers, at most 25, in one cassette, or the Front Opening Unified Pod (FOUP, which is a container). Up to several thousand identical chips can be made on each wafer by building up the electronic circuits layer by layer in a wafer fabrication facility (wafer fab). Wafer fabrication

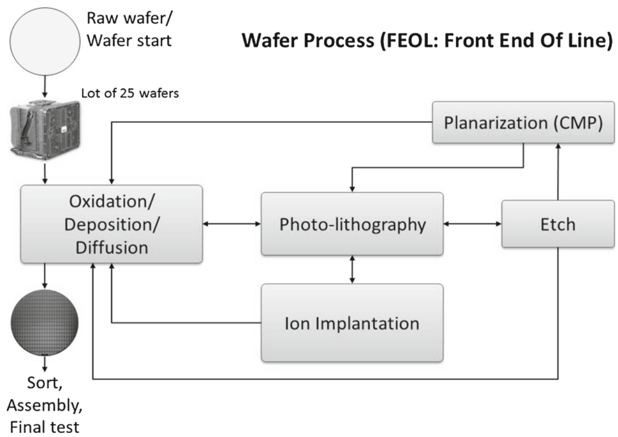

Fig. 1 Main work areas in a wafer fab

can be described as a multistage process with reentrant flows. Each chip layer requires several processing steps such as diffusion, chemical-mechanical polishing (CMP), film deposition, photolithography, doping, and etching. For each product type, and depending on the technology, a wafer may undergo more than 700 process steps over a period of several weeks. Figure 1 is adapted from Mönch et al. (2011) and gives the main process steps of wafer fabrication. In the figure, the arrows indicate the flow of the wafers between the different main work areas.

Advanced process control (APC) has become a key element in the semiconductor fabrication process, with aggressive competition for achieving better quality. The motivation for implementing APC is to improve device yield by controlling processes and equipment to collect information, reduce process variability, and increase equipment efficiency. To achieve this, a wide variety of mathematical/statistical and physical techniques have been developed and used, and standards and communication interfaces between equipment and IT systems have been defined. To maintain the processes at their specification levels and to survey equipment for existing possible faults, APC makes it possible to prevent process excursions, increase tool utilization, and reduce variability, non-product-wafer utilization, and maintenance and cycle time.

In the literature, only a few articles integrate specific APC information about how to make better scheduling decisions. In addition, few articles use scheduling decisions to improve information given by APC systems. We believe that the two domains (scheduling and APC), though separately studied, are interdependent because they both have as their objective improving productivity. For example, the heterogeneous equipment conditions of an APC system should be taken into consideration in scheduling decisions in order to prevent unexpected equipment excursions caused by assuming homogeneous equipment in scheduling plans. In this article, we discuss the achievable benefits by integrating scheduling decisions and information from APC. The remainder of the paper is organized as follows. Section 2 explains the features and role of APC in scheduling decisions. Section 3 presents a overview of the problem of integrating scheduling and APC information. Section 4 discusses the issues surrounding the integration of APC and scheduling. The paper concludes in Sect. 5.

2 Roles of advanced process control and scheduling in wafer fabrication

APC is the current practice to ensure a continuous process improvement in semiconductor manufacturing (Su et al. 2004). The controlling algorithms, system interfaces, and infrastructure have been widely developed and discussed in the literature over the last two decades. A comprehensive literature review is presented in Edgar et al. (2000) and Su et al. (2007), where the authors discuss the challenges and possibilities in automatic control in semiconductor manufacturing. APC can be regarded as a general heading for all kinds of equipment and process engineering systems in semiconductor manufacturing. It normally consists of several functional systems, for example, run-to-run (R2R) control, fault detection and classification (FDC), overall equipment efficiency (OEE), and e-Diagnostic Barna (1996); SEMATECH (2008). The fundamental aspects of APC are data collection and interface design, which enable the aforementioned engineering systems to function. The exact scope of APC is hard to clearly define because the techniques are being continuously developed and standards are constantly evolving. In addition, each integrated circuit (IC) maker has its own preferences in customizing APC modules, and thus the definition of APC differs slightly from fab to fab. For example, a common argument concerns the attribution of the FDC system. FDC is used to monitor in real time the process equipment of various process steps in order to allow for the detection of faults and their signatures associated with preventive actions. Based on an analysis of continuous data collected from equipment sensors, researchers and engineers would like to develop an equipment health index or equipment health factor (EHF) to better characterize the actual condition of equipment. Therefore, some manufacturers believe that FDC should be part of advanced equipment control while others place it within the scope of APC (Moyne and Patel 2007). In this paper, we will discuss APC in the most general framework such that APC is defined as those activities that collect and analyze data from process equipment for the purpose of improving productivity and yield. Consequently, APC addresses those systems mentioned at the beginning of this section.

Figure 2 presents a detailed diagram for an APC framework and an overview of this paper's focus, that is, the integration of APC and scheduling. In the left part of Fig. 2, one

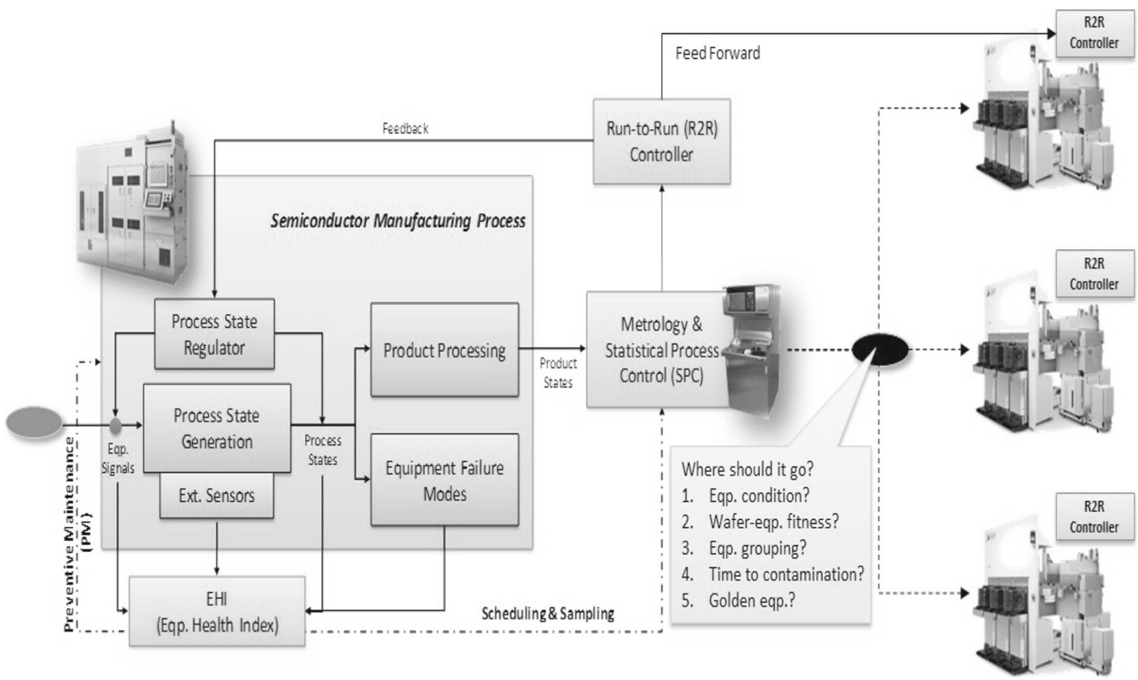

Fig. 2 Framework of APC and its integration with scheduling issues

sees APC functions inside a manufacturing process. When a wafer enters the equipment, the process starts according to its designated recipe in terms of its product type. A recipe is usually composed of a group of controllable parameters with set points that define the rise and fall in the process, for example, the pressure should arrive at 900 torr in the chamber and stabilize within 5 s at the beginning of the process. For a better understanding of the equipment and process, many IC makers implement the FDC system, which collects equipment signals, for example voltage and current, and process states, for example, pressure and temperature, from internal/external sensors. These real-time data are collected from beginning to end of the process and will be used to characterize the equipment condition, as mentioned earlier.

The commonest way to analyze FDC data is to apply statistical process control (SPC) methods for monitoring the equipment and process (May and Spanos 2006). Because FDC data are actually temporal series and nonstationary in relation to recipe settings, engineers usually summarize them in several indices using the statistics within temporal windows based on their engineering knowledge, for example, the slope of the temperature during the ramping-up stage or the average pressure during the steady stage. These summarized indices of each wafer are then monitored using SPC charts such as exponentially weighted moving average and cumulative sum control charts (Montgomery 2009). With the advancement of computing technology, engineers have tried to implement multivariate SPC methods, for example, Hotelling's T^2 distribution or principal component analysis, to handle multiple indices simultaneously.

When one wafer finishes its process in the equipment, it can be sampled and its physical characteristics measured using metrology equipment. Again, the metrology measurements collected here wafer by wafer are monitored through SPC methods to check not only the quality of the wafer but also the equipment condition. Because measuring a single wafer is very time consuming, a very limited number of wafers is sampled for measurement after certain key processes. As a result of the limited sample size and confounding interactions of preceding processes, analyzing metrology data is a very challenging task.

As can be seen in Fig. 2, the metrology measurements are sent to the R2R controller for fine-tuning of the recipes of the previous or next process, which is called feedback or feedforward control. In the feedback loop, the R2R controller compares the target with the real measurements and calculates the necessary adjustments in the recipes according to the algorithms designed with physical and mathematical models (May and Spanos 2006). For example, the polishing pad in CMP equipment is worn down over time, which impacts the polishing quality with each successive wafer Stamper et al. (2012); Sarfaty et al. (2002). The feedback loop then compensates the drift in the recipe settings of the next run. In the feedforward loop, the R2R controller receives the information from the previous process and calculates the necessary adjustments in the current process such that the variations caused by the previous process will be compensated.

For example, thickness measurements after a chemical vapor deposition (CVD) process represent critical information for the polishing time setting in the CMP process.

Other functions mentioned previously, such as OEE and e-Diagnostic, are more related to the design and implementation of monitoring/executing systems and are not discussed within the scope of this paper. Within the APC framework, research in integrating scheduling decisions for wafer sampling from process equipment have been active in recent times. This is mainly motivated by the long measuring time on metrology equipment, and the measurements are thus crucial for the followup analysis such as process qualification, the R2R controlling mechanism, and equipment condition evaluation. Moreover, APC information is very critical for the decision of either scheduling or dispatching from process to process. However, very few studies explicitly discuss utilizing APC information, such as the state of equipment and equipment grouping, or analyzing the wafer-equipment fitness and golden (the best performance) equipment/routes for comprehensive scheduling/dispatching decisions, as depicted in the right part of Fig. 2. This paper is motivated by a desire to survey and address the integration of APC information with scheduling decisions.

3 Literature survey on integration of scheduling decisions and APC

From the perspective of scheduling, APC could be seen a constraint that might delay or even jeopardize a scheduled production plan for the purpose of retaining an acceptable quality level. Once the product quality or equipment condition does not reach a satisfactory level in the APC loop, warning alarms will be triggered and the production flow delayed due to the extra actions of checking and evaluating the product quality. The better the APC is, the less deviations from the given production specifications are noticed and the less reworking is required for corrections (Engell and Harjunkoski 2012).

From the perspective of APC, the scheduling plan plays the role of material contributor that organizes the queue and feeds the lot into an APC loop. This is of very fundamental importance for the functioning of APC loops because all data related to process characteristics and product quality can be collected, analyzed, monitored, and controlled. In current practice, if an out-of-control action plan is triggered in an APC loop, then there is no choice but to reschedule the production plan. The unilateral effect of APC on scheduling plans has motivated us to address in this paper the integration of scheduling plans with APC information and the constraints on efficient production control.

Manufacturing technologies are getting more and more complex as the technology node advances. A scheduling plan without consideration of the requirements and limitations of APC would be infeasible. Focusing on the APC loop to improve product quality without complying with the planned schedule can interfere with factory-level plans. For example, a scheduling plan can take into account the setups and changeovers among heterogeneous products to alleviate the loadings in the APC. Scheduling can also make APC information, such as equipment condition and product measurement, available for updating R2R control models or enhancing product quality. On the other hand, advanced manufacturing information deduced from contemporary researches on APC loops, such as EHF, R2R control models, and advanced SPC charts, can be incorporated into scheduling decisions to maintain production efficiency without a concomitant loss in product quality. As a result, it is critical to integrate the input/output of APC loops when making scheduling plans because of the following properties and requirements in the modern semiconductor manufacturing environment:

- Products are becoming increasingly complex.
- Scheduling plans should avoid costly setups and changeovers and reduce cycle time and discarded products.
- Utilize the most comprehensive information in APC to assist with the precision of production control.
- Reduce nonproductive time by managing maintenance in an intelligent way.
- Avoid scheduling that leads to operational problems.

The literature on the integration of scheduling and APC functions is investigated and categorized in Table 1, which was constructed for the purpose of looking into the intersection between the characteristics of scheduling problems (single, parallel machines, and job-shop problems) and the functional modules in an APC framework (FDC, equipment qualification, EHF, preventive maintenance (PM), metrology, SPC, and R2R controller).

In general, few many works incorporate the characteristics of the two domains (APC and scheduling). Table 1 shows the following issues on the two domains:

- Single-machine scheduling problems.
  In the work of Purdy (2007) on metrology analysis, the author discusses the management of multiple and overlapping sampling rules on one piece of equipment. He proposed an integer linear programming approach to solving the problem. In Detienne et al. (2012), the problem of many different product types being manufactured with a given production schedule is studied. The goal is to schedule a limited number of lots on measurement equipment to minimize risks. Each lot to be measured has a set of due dates and associated risks, corresponding to the start times of the lots of the same product type in the production

Table 1 Articles on integration of scheduling and APC information

	Scheduling characteristics		
	Single machine	Parallel machines	Job shop
APC functions			
Fault detection and classification (FDC)			
Equipment qualification		Impact of qualification on scheduling decisions (Johnzèn et al. 2008)	
Equipment health factor (EHF)		Optimize scheduling plan with delay time constraints and equipment conditions (Obeid et al. 2012)	
Preventive maintenance (PM)			
Metrology	Management of multiple and overlapping sampling rules (Purdy 2007), study of different product types given production schedule (Detienne et al. 2012)	Penalty-based optimal rule selection for real-time measurement decision (Good and Purdy 2007); sampling decision system (SDS) with two-stage decision model (Kurz et al. 2012); scheduling with sampling constraints (Holfeld et al. 2007, Detienne et al. 2011)	Algorithm to sample, schedule, skip, and optimize metrology capacity (Dauzère-Pérès et al. 2010; Yugma et al. 2011)
Statistical process control (SPC)			
Run-to-run (R2R)	Scheduling with R2R constraints (Cai et al. 2009)	Scheduling with R2R constraints (Obeid et al. 2011, Li and Qiao 2008)	Information sharing between R2R and dispatching (Anderson and Hanish 2007); state error covariance matrix for optimizing R2R control (Pasadyn et al. 2008)

schedule. Lower and upper bounds, obtained from linear and Lagrangian relaxations of different integer linear programming formulations, are compared.

Cai et al. (2009) discuss the problem of qualification run on one piece of equipment. More precisely, the authors studied the interaction between scheduling and APC, specifically in R2R control. A single-machine multi-job-type makespan and R2R constraints are considered. An integer programming approach is used to solve the problem, and a heuristic is tested on generated instances.

– Parallel-machine scheduling problems.

Johnzen Johnzèn et al. (2008) proposed to model a semiconductor fabrication facility (in the case of high-mix production) by optimizing the capacity allocation of wafer fabs. The proposed approach supports effective qualification management in wafer fabs (qualifications of products on tools) and increases the flexibility of tools. Johnzen has studied the impact of flexibility on the scheduling decisions. The study was conducted by simulation on parallel machines using heuristics for scheduling.

Obeid et al. (2011) study the lot (job) scheduling of different product types (job families) on parallel machines. The time constraint associated with each job family should be satisfied for a piece of equipment to remain qualified to process a job family. A piece of equipment is qualified to process a certain job if and only if, at a given time instant, the time threshold corresponding to the job family is not violated. To solve this problem, a mixed-integer linear programming (MILP) model was proposed and solved. Purdy (2007) proposed a novel algorithm for selecting optimized wafers for measurement given a set of selection rules. The algorithm is based on assigning a penalty to each of the sampling rules and then using a mixed-integer linear program to pick wafers that minimize the sum of the penalties. In semiconductor manufacturing processes, metrology operations are so expensive and time consuming that only a certain number of wafers are measured. For that reason, virtual metrology (VM) methodologies that predict wafer metrology measurements in real time using equipment/process information are proposed to avoid real measuring operations. However, sampling designs in current practice do not take account of such information. Kurz et al. (2012) present a sampling decision system (SDS) that relies on virtual metrology data, suggesting an optimal strategy for measuring productive wafers. Detienne et al. Detienne et al. (2012) deal with the parallel-machine scheduling problem with the objective of minimizing a regular step total cost function. The authors are interested in solving an optimization problem in which the scheduling of inspection operations must be determined. Therefore, fixed production operations are processed at a higher level of quality. The inspection operation of a given lot is associated with the fact that lots of the same type must be produced. One lot can be scrapped if its production starts before the metrology information of other lots from the same product type is obtained. The objective is to minimize the overall risk incurred by production lots by optimally scheduling the set of inspection operations. In Li and Qiao (2008), the objective is to analyze the impact of the qual-run requirements of APC on the scheduling performance, such as the total weighted tardiness and makespan, in a number of experiments. Taking parallel machines, for example, the authors propose an ant colony optimization algorithm to arrive at a satisfactory solution in a reasonable amount of computation time under scenarios with or without consideration of the qual-run requirements of APC. A time constraint study was conducted by Obeid et al. (2011) in the area of photolithography. This constraint requires that there must be at most a given time interval between the processing of two jobs of the same job family on qualified equipment. A MILP model is proposed to solve the problem. A bicriteria objective function that includes scheduling and qualification criteria is considered, and dedicated heuristics are also proposed.

– Job-shop scheduling problem.
Dauzère-Pérès et al. (2010) and Yugma et al. (2011) address this problem, which consists in maximizing information, minimizing risk from measurement, and taking into account measurement capacities, and an algorithm is proposed to sample lots. An indicator, called a global sampling indicator, is developed to determine whether the lots are sampled and measured on metrology equipment. A simulator called S5 (Smart Sampling Scheduling and Skipping Simulator) was used and validated on actual data. It shows that the risk can be strongly reduced while retaining a limited number of measurements.
Anderson and Hanish (2007) discuss some benefits of integrating R2R and scheduling at the fab level. Pasadyn et al. (2008) study R2R control applications that require a constant stream of information about the state of the process in order to perform well. The trace of the state error covariance matrix from the Kalman filter is used as a metric for determining the information content of a particular data set in an R2R control algorithm. Processing decisions, such as batch scheduling, equipment allocation, and sampling plans, are shown to have an effect on estimator performance.

We examined the literature and could not find articles related to the blank field of the table. Because just a small body of work was found, we believe that it will be a critical subject for both academics and industry in the near future.

4 Issues concerning the integration of APC and scheduling

In our literature survey, we found that a very limited number of studies either utilized APC information or took into account APC constraints for scheduling. In this section, we will address the issues that could be raised with respect to scheduling and APC information or constraints.

4.1 APC information for scheduling

Information, whether generated from equipment/process or as the inputs within an APC framework, is usually used and analyzed specifically for process fine-tuning or equipment monitoring. In this survey, we discuss the possibility of taking information into account in the optimization of scheduling.

4.1.1 Equipment health index

Conventionally, equipment status is described as either up or down. With data circulated in an APC framework, we believe equipment status should be defined in a more progressive way. The data collected from FDC systems, metrology measurements, and even adjustments from R2R controllers should be used and incorporated into evaluations of equipment status. Scheduling decisions taking into account these indicators would then lead to a decrease in the rate of failure by promoting reliable processes/equipment. We can define exogenous or endogenous indicators depending on whether or not the processes involved in the decisions affect the equipment status. An exploration of this field might start with the development of health indicators that seem to be missing from the APC literature. This is a complicated matter because these indicators must be defined by equipment, process, and recipe and would evolve as a function of time. The difficulty will be not only in collecting a large amount of data for analysis but also in proposing a proper modeling technique for the equipment and recipe (Chen and Wu 2007). For example, a recipe-independent equipment health indicator and hierarchical monitoring scheme are proposed for the efficient evaluation of equipment condition (Chen and Blue 2009; Blue et al. 2013).

Most scheduling problems are tackled by considering not information derived from APC systems but largely from production execution systems. For example, equipment is generally considered to be capable of performing an operation or breaking down. In semiconductor manufacturing execution systems, the state of the equipment is always seen as a binary value: 1, meaning the equipment can process the operation, or 0, indicating the equipment cannot process the operation. However, a real equipment status should have any value between 0 and 1, for example, 0.2, 0.5, or 0.7. Depending on the criticality of an operation, the equipment with its associated status is selected to perform the operation. This is exactly the kind of information that can be provided by APC systems.

Recently, the health indicators used to monitor equipment status have come to be known as the EHF (Gleispach et al.

2012). EHF is defined as an indicator or factor associated with the equipment that is to describe the level of reliability (Chen and Wu 2007). On the other hand, a criticality indicator can be defined as associating with a lot, wafer, or product type (family) and indicates the states of the considered element. Based on these indicators, we can classify equipment/lots into quality categories, for example, good, average, and bad. If a scheduling plan is created without taking into account the EHF, a yield loss may occur. The cost of assigning a lot to a piece of equipment should be modeled in terms of the percentage/probability of yield loss. The challenging problem will be how to schedule while minimizing this kind of cost.

4.1.2 Equipment qualification

Equipment qualification is an important issue in semiconductor fabs. Qualification can help improve equipment efficiency, minimize work in process, and reduce cycle time, thereby affecting the overall fab performance (Johnzèn et al. 2008, 2011). Typically, lots cannot be sent to a piece of equipment unless that equipment is qualified. Qualification is usually maintained by sending test wafers to the equipment (Faruqi et al. 2008) and validating them by SPC monitoring or an R2R controller within the APC framework. Obviously, scheduling decisions are affected by equipment qualification (Johnzèn et al. 2011). The management of equipment qualification can be seen as a tactical problem. Long-term equipment qualification decisions are required since the initial qualification process often takes a large amount of time depending on the nature of the process in the manufacturing chain. Moreover, two types of qualification should be considered: equipment qualification and process condition qualification. To start processing a lot on a piece of equipment, both types of qualification are required. A scheduler that takes into account qualified and nonqualified equipment and processes will more effectively decide on the lots being dispatched in a fab.

4.1.3 Preventive maintenance

PM is a very fundamental activity in semiconductor manufacturing because a piece of equipment cannot be 100 % reliable for its entire life. A good PM schedule can increase the availability of equipment by trading off between planned unproductive downtime and the risk of unscheduled downtime due to equipment failures. Unscheduled downtimes will not only induce a loss in productivity but also disturb the manufacturing process. Thus, to lessen these negative effects, PM schedules must be planned carefully.

Despite the fact that there exists research on the integration of PM and scheduling decisions and their interdependency (Cassady and Kutanoglu 2005), classical PM schedules are mostly defined in a deterministic way such as by a fixed time span or fixed number of processed lots. It is very desirable to create a PM schedule that is predictably flexible, which is also referred to as the predictive maintenance (PdM) policy. Thus, the integration of both PM policy, especially the PdM, and scheduling will lead to a decrease in equipment downtime and an increase in utilization. This idea is not thoroughly investigated in the literature on semiconductor manufacturing. Cassady and Kutanoglu (2005) address this issue using a single machine scheduling problem. However, we believe that integrating scheduling and APC systems would require a fabwide point of view.

4.1.4 R2R control loop

As mentioned in Sect. 2, R2R controllers can perform essential calibration and regularly adjust the controllable parameters in recipe settings to compensate for process drifts. The potential problem lies in the fact that, over time, the parameters do not need to be kept fixed, in particular when the parameters of the last recipe become unsuitable for processing certain types of lots on the given equipment. Hence, lots should be sent regularly to the equipment in order to keep the R2R loop parameters up to date. More precisely, this problem lies not in the idleness of a piece of equipment but in the capacity to process different types of products using the same equipment. Therefore, the time between arrivals of two lots of the same type is considered idle time for the parameters.

In R2R control loops, nonproduct wafers (NPWs) are commonly used and processed separately from production wafers. They are used for multiple purposes and reused until some R2R criteria are satisfied, for example, the process conditions for certain recipes are stable. NPWs can also be used for equipment qualification, stabilization of process performance, process development, and process conditioning. For NPWs to be used multiple times, they are recycled or reclaimed. Because fabs consume a large number of NPWs, it is important to collect NPW information, including wafer location, specifications, and history of usage. According to SEMATECH (2008), an efficient management of NPWs is a major contributor to fab costs. The number of NPWs used in fabrication affects the cycle time, manufacturing costs, and fab capacity. Optimization of NPW use represents a potential approach to reducing manufacturing costs. The cost of NPWs varies considerably and should be improved by a NPW scheduling system that is capable of maximizing usage while at the same time minimizing the number of NWPs.

Patel (2000) presents a study on a dispatching rule that focuses on controllers to increase their accuracy by allocating lots on available equipment in order to minimize the risk of error under random perturbations. Cai et al. (2009) tackle a scheduling problem with APC constraints on a single machine. They study the interactions between scheduling and R2R control, with setup times between job families and a qualification run when R2R constraints are not respected. Li and Qiao (2008) analyze the impact of R2R constraints on parallel machines where qualification runs can be scheduled in relation to a threshold between two jobs of the same family.

The increasing complexity of semiconductor manufacturing processes directly impacts the management of NPWs. The quality requirements are more stringent, while the unit cost increases. The flow of these wafers, which are in turn prepared, used, recycled, reprocessed, sorted, and reused, is a reflection of the growing number of products being managed. In this context, numerous additional improvements and cost reductions can be undertaken.

4.1.5 Metrology and virtual metrology (VM)

To compensate for process drift and variation, R2R controllers demand a huge amount of data collected from wafers. This is not always easily done because metrology equipment is usually a limited and expensive resource in a fab. Taking measurements is very time consuming and definitely affects the production cycle time. Moreover, in extreme cases, inspection would not be 100 % reliable and could introduce errors of almost the same order as the fraction of defectives even if all the lots/wafers are measured (Pesotchinsky 1987). Under these circumstances, measurements for real-time adjustments of R2R loop parameters may not always be available. This problem is often referred to as time (or delay) metrology. The delay reflected on the overall cycle time of a product will reduce the overall responsiveness of the manufacturing unit and increase the risk of defects (Wein 1988). Moreover, Williams et al. (1999) discusses metrology operations needed for an APC system that incur significant costs that could be reduced by optimizing the measurement process itself.

VM was proposed in order to solve foreseeable metrology problems: it is time consuming and costly. The concept of VM is rather simple. With advancements in modern information technology, many process and equipment data become available within an APC framework. By utilizing these data to correlate with metrology measurements, a mathematical, statistical, or even physical model can be constructed. The validated model can be then used to generate a measurement prediction, which is so-called VM. Many studies have been devoted to the study of VM modeling techniques, which are then applied to different process operations. For example, Lynn et al. (2012) proposed a windowed VM scheme to model etch depth measurements immediately following an etch process based on Gaussian process regression. Besnard et al. (2012) used FDC data to predict CVD oxide thickness based on a partial least-squares regression (PLSR) and tree-ensemble method. However, a reliable model is very critical, and VM should be trustworthy so that the actual measuring activity can be skipped. Some researchers are working to develop a generic scheme to set up, monitor, update, and rebuild VM models so that VM methodologies can be practiced in a real production environment Cheng et al. (2007); Huang et al. (2007); Cheng et al. (2012). It remains a big challenge to put VM into real practice because of the issues, such as model stability, model updating criteria, prediction confidence, and the sampling strategy for maintaining a model's validity, waiting to be investigated in more detail.

Other than predicting measurements, the sampling strategy of selecting wafers to be measured is an alternative perspective to study. In the semiconductor industry, a sampling strategy determines the rate of measurements based on statistics. These rates can be obtained empirically based on knowledge of the product during its various manufacturing stages and more effectively by taking into account its life cycle (Boussetta and Cross 2005). This sampling decision depends on many factors, such as lot type, lot priority, and the capacity of the metrology equipment. Answering the questions of what, when and how to choose a lot for measuring forms a sampling strategy.

Alternatively, different pieces of equipment have differing processing times, and this results in an increase in cycle time variability. Hence, the idea of dynamic sampling was elaborated and consists in dynamically deciding which products should be measured to update certain control parameters of various APC systems (Dauzère-Pérès et al. 2010; Yugma et al. 2011. Williams et al. (1999) simulate the overall cost of production using static and dynamic strategies. Another alternative, called integrated metrology (Lensing and Stirton 2007), is to measure each product when it is in queue in front of the production equipment. However, its implementation on all equipment seems unrealistic given the cost of the metrology equipment needed.

Speaking of dynamic sampling strategies, Lee et al. (2003) outline a strategy to control queues and to skip certain measurements. Purdy (2007) presents a sequence modeling strategy for selecting lots for measurement. Holfed et al. (2007) describe a system applied to an entire manufacturing unit to balance the needs of metrology and minimize the risk under capacity and cycle time constraints. Most dynamic sampling techniques focus on statistical methods of determining the sampling rate. By considering the instantaneous information taken from an APC control system, dynamic sampling techniques may become "smarter" (one can speak of "smart sampling"). It is interesting to study how to make sampling not only dynamic but also intelligent by benefiting from the information related to the route that the wafer undergoes. Knowing the sequence followed by a lot (and this sequence is determined by the scheduler), it is possible to determine whether or not the lot should be selected depending on the level/degree of confidence in the equipment in the sequence.

This idea may lead to an additional scheduling criterion and help to reduce cycle times.

A more general point of view would be to simultaneously tackle the problem of scheduling lots on process and metrology equipment to minimize the production cost. The issue here is how to take the constraints of both processing and metrology equipment into account in the same scheduling algorithm with a view toward reducing the measuring costs and cycle times (Nduhura-Munga et al. 2013).

4.1.6 Wafer quality index

Given a wafer that is sampled for retrieval of metrology measurements, the SPC monitoring should deliver certain information, such as the quality of or confidence in these measurements. These data can not only feed back or feed forward through the R2R controller but they could also be correlated with the product yield. The FDC data of each wafer represent the processing condition of the wafer, so we could combine FDC signals, metrology data, and SPC results into a synthetic index to depict the wafer quality.

This index is associated with each wafer and can be used by the scheduler to know the states of all the equipment candidates and the state of the wafer itself. It may change the predefined route of the wafer depending on two factors: the equipment health and the wafer quality. Hence, fully dynamic scheduling plans will be available. The wafer quality index can be thought of in effect as complementing the EHF addressed earlier. The idea of updating and using dynamic data may be hard to apply in practice. This is why we add some restrictions and constraints in order to make this idea more applicable.

4.2 Scheduling with APC constraints

Constraints that are handled using the control/monitor mechanism in an APC framework, such as R2R adjustments or a backup plan on the equipment, can be considered as well in scheduling optimization, so that the dispatching of the wafer/lot will be reasonable and mitigate control/monitor loadings.

4.2.1 Wafer–equipment matching

By referring simply to the information on the state of the equipment, scheduling can avoid situations where critical products are processed by poorly maintained equipment and thus reduce the chances of turning out low-quality products. Due to the natural differences in equipment having the same capabilities, sometimes equipment in the best condition is not necessarily suitable for certain products. For example, in a set of equipment used to make depositions on wafer surfaces, a natural difference can be the acceptable deposition offsets among the pieces of equipment. In semiconductor manufacturing, this feature is commonly investigated in an APC framework from the perspective of R2R control. If we only take into account the equipment state and design a scheduling plan, a wafer that requires less thickness might be sent to equipment that is in the best condition but that usually puts on more materials in the deposition process. The idea behind imposing a wafer–equipment matching criterion is to make sure the right wafer goes to the *right* equipment instead of the *best* one in order to cancel out the accumulated variations resulting from previous processes. We believe that a scheduling plan that considers this criterion would ensure a high product quality and ultimately enhance the quality.

4.2.2 R2R feedforward adjustment

R2R control loop parameters need to be kept up to date constantly. This is usually achieved by sending the results of measuring tasks performed on metrology equipment to the R2R controller. If these results were not sent regularly, the parameters in R2R control loop would not provide the right recipe conditions, which might increase the risk of discarding the processed lot. Therefore, the first constraint corresponding to R2R parameters is to send at least one wafer measurement within a constrained time span, to be specified depending on several criteria, such as the type of processes, equipment, and the maturity of the control loop.

Another constraint would be the capacity on the total number of measurements (Pasadyn et al. 2008) since measuring all processed wafers is impossible. Scheduling under these constraints is a challenging problem and must maintain the R2R controller functioning to process wafers of high quality while avoiding unnecessary sampling for measurements to minimize costs. As an example, Sun et al. (2005) discussed scheduling under time-constrained processes. Furthermore, a semiconductor fab has a tremendous number of control loops that serve various types of processes and several types of products. From a local control viewpoint (equipment/process level), it is recommended to keep all these control loops updated. However, from a more strategic viewpoint, i.e., supervisory level, it is not necessary to have all these control loops updated all the time since we can assign less importance to noncritical loops. All these aspects could be integrated into a scheduling system at the supervisory level to improve overall factory efficiency. This might require a solution to complex decision problems and involve different operators and sometimes competing objectives.

5 Discussions and conclusion

In recent decades, scheduling techniques applied to the semiconductor industry are commonly addressed in the literature. Meanwhile, significant advancements have been made in an APC framework implemented by IC makers. However, we found that the idea of integrating APC information/constraints into scheduling has not been well investigated, which would surely lead to substantial improvements. These improvements are indispensable in the current extremely competitive environment of semiconductor manufacturing, where industry players are continuously searching for ways to make improvements on yields and costs.

Based on a literature survey, we discussed the problems that arise from the integration of scheduling decisions and APC information, as well as the associated benefits. Advanced ideas can be very theoretical, and thus fundamental research is still required before practical implementations are undertaken. This review started by investigating the intersections between scheduling problems and the functions of an APC framework. The very limited research in this field aroused our strong interest in defining the potential issues that arise in any attempts to make scheduling plans comply with information/constraints resulting from APC functions. We believe that these issues will produce helpful insights for researchers in this domain.

In the near future, the semiconductor industry will attain a higher precision level in fabrication processes. Each wafer will become a critical entity that in turn directly affects the overall factory efficiency. In particular, when the wafer size migrates toward 450 mm, lot losses will be extremely costly. Therefore, a fully integrated fabrication framework will be seriously considered.

Acknowledgments This work constitutes part of the IMPROVE (**I**mplementing **M**anufacturing science solutions to increase equi**P**ment p**RO**ducti**V**ity and fab p**E**rformance) ENIAC European project. In particular, we would like to thank Philippe Vialletelle and Jacques Pinaton from STMicroelectronics for their practical advice.

References

Anderson, M., & Hanish, C. K. (2007). An evaluation of the benefits of integrating run-to-run control with scheduling and dispatching systems. *IEEE Transactions on Semiconductor Manufacturing*, *20*(4), 386–392.

Barna, G. G. (1996). APC in the semiconductor industry, history and near term prognosis. In *Proceedings of advanced semiconductor manufacturing conference and workshop, Cambridge, (USA)* (pp. 364–369).

Besnard, J., Gleispach, D., Gris, H., Ferreira, A., Roussy, A., Kernaflen, C., et al. (2012). Virtual metrology modeling for cvd film thickness. *International Journal of Control Science and Engineering*, *2*(3), 26–33.

Blue, J., Gleispach, D., Roussy, A., & Scheibelhofer, P. (2013). Tool condition diagnosis with a recipe-independent hierarchical monitoring scheme. *IEEE Transactions on Semiconductor Manufacturing*, *26*(1), 82–91.

Boussetta, A., & Cross, A. (2005). Adaptive sampling methodology for in-line defect inspection. In *Proceedings of the IEEE/SEMI Advanced Semiconductor Manufacturing Conference, Munich (Germany)* (pp. 25–31).

Cai, Y., Kutanoglu, E., Hasenbein, J., & Qin, J. (2009). Single-machine scheduling with advanced process control constraints. *Journal of Scheduling*. doi:10.1007/s10951-010-0215-8.

Cassady, C. R., & Kutanoglu, E. (2005). Integrating preventive maintenance planning and production scheduling for a single machine. *IEEE Transactions on Reliability*, *54*(2), 304–309.

Chen, A., & Blue, J. (2009). Recipe-independent indicator for tool health diagnosis and predictive maintenance. *IEEE Transactions on Semiconductor Manufacturing*, *22*(4), 522–535.

Chen, A., & Wu, G. S. (2007). Real-time health prognosis and dynamic preventive maintenance policy for equipment under aging markovian deterioration. *International Journal of Production Research*, *45*(15), 3351–3379.

Cheng, F.-T., Huang, H.-C., & Kao, C.-A. (2012). Developing an automatic virtual metrology system. *IEEE Transactions on Automation Science and Engineering*, *9*(1), 181–188.

Cheng, F.-T., Huang, H.-C., & Kao, C.-A. (2007). Dual-phase virtual metrology scheme. *IEEE Transactions on Semiconductor Manufacturing*, *20*(4), 566–571.

Dauzère-Pérès, S., Rouveyrol, J.-L., Yugma, C., & Vialletelle, P. (2010). A smart sampling algorithm to minimize risk dynamically. In *Proceedings of the IEEE/SEMI Advanced Semiconductor Manufacturing Conference (ASMC)* (pp. 1078–8743). San Francisco: CA (USA).

Detienne, B., Dauzère-Pérès, S., & Yugma, C. (2012). An exact approach for scheduling jobs with regular step cost functions on a single machine. *Computers & Operations Research*, *39*(5), 1033–1043.

Detienne, B., Dauzère-Pérès, S., & Yugma, C. (2011). Scheduling jobs on parallel machines to minimize a regular step total cost function. *Journal of Scheduling*, *14*(6), 523–538.

Edgar, T. F., Butler, S. W., Campbell, W. J., Pfeiffer, C., Bode, C., Hwang, S. B., et al. (2000). Automatic control in microelectronics manufacturing: Practices, challenges, and possibilities. *Automatica*, *36*(11), 1567–1603.

Engell, S., & Harjunkoski, I. (2012). Optimal operation: scheduling, advanced control and their integration. *Computers & Chemical Engineering*, *47*, 121–133.

Faruqi, A., Goss, R., Adhikari, D., Kowtsch, T. (2008). Test wafer management and automated wafer sorting, In *Proceedings of the IEEE/SEMI advanced semiconductor manufacturing conference* (pp. 322–326).

Gleispach, D., Blue, J., Roussy, A., Haselmann, M. (2012). Hierarchical monitor scheme for recipe-independent tool condition evaluation. One approach of EHF (Equipment Health Factor). In *Proceedings of 12th European advanced process control and manufacturing conference (APCM), Grenoble, France*.

Good, P. R., & Purdy, M. A. (2007). An milp approach to wafer sampling and selection. *IEEE Transactions on Semiconductor Manufacturing*, *20*(4), 400–407.

Huang, H.-C., Su, Y.-C., Cheng, F.-T., Jian, J.-M. (2007). Development of a generic virtual metrology framework. In *Proceedings of IEEE International Conference on Automation Science and Engineering (CASE), Scottsdale, Arizona, (USA)* (pp. 282–287).

Holfeld, A., Barlovic, R., & Good, R. P. (2007). A fab-wide apc sampling application. *IEEE Transactions on Semiconductor Manufacturing*, *20*(4), 393–399.

Johnzèn, C., Vialletelle, P., Dauzère-Pérès, S., Yugma, C., & Derreumaux, A. (2008). Impact of qualification management on scheduling in semiconductor manufacturing. In *Proceedings of the winter simulation conference (WSC)* (pp. 2059–2066). Miami: Florida (USA).

Johnzèn, C., Dauzère-Pérès, S., & Vialletelle, P. (2011). Flexibility measures for qualification management in wafer fabs. *Production Planning and Control*, *22*(1), 81–90.

Kurz, D., Deluca, C., Pilz, J. (2012). Sampling decision system in semiconductor manufacturing using virtual metrology, In *Proceedings of the 8th IEEE international conference on automation science and engineering (case), Seoul (Korea)* (pp. 74–79).

Lee, S. B., Lee, T. Y., Liao, J., & Chang, Y. C. (2003). A capacity-dependence dynamic sampling strategy. In *Proceedings of the IEEE international symposium on semiconductor manufacturing* (pp. 312–314). San Jose: California (USA).

Lensing, K., & Stirton, B. (2007). Perspectives on integrated metrology and wafer-level control. In *Proceedings of the international symposium on semiconductor manufacturing* (pp. 1–4). California (USA): Santa Clara.

Li, L., & Qiao, F. (2008). The impact of the qual-run requirements of APC on the scheduling performance in semiconductor manufacturing. In *Proceedings of the IEEE international conference on automation science and engineering, (CASE)* (pp. 242–246). VA (USA): Arlington.

Lynn, S. A., Ringwood, J., & MacGearailt, N. (2012). Global and local virtual metrology models for a plasma etch process. *IEEE Transactions on Semiconductor Manufacturing*, *25*(1), 94–103.

May, G. S., Spanos, C. J. (2006). Fundamentals of semiconductor manufacturing and process control. Hoboken, New Jersey: Wiley (Chap. 6, Chap. 9, pp. 181–227, 333–378).

Mönch, L., Fowler, J. W., Dauzère-Pérès, S., Mason, S. J., & Rose, O. (2011). A survey of problems, solution techniques, and future challenges in scheduling semiconductor manufacturing operations. *Journal of Scheduling*, *14*(6), 583–599.

Moore, G. E. (1965). Cramming more components onto integrated circuits. *Electronics*, *38*(8), 114–117.

Montgomery, D. C. (2009). *Introduction to statistical quality control, Chap. 9*. Hoboken, New Jersey: Wiley.

Moyne, J. R., & Patel, N. S. (2007). Special section on advanced process control. *IEEE Transactions on Semiconductor Manufacturing*, *20*(4), 343–344.

Nduhura-Munga, J., Rodriguez-Verjan, G., Dauzère-Pérès, S., Yugma, C., Vialletelle, P., & Pinaton, J. (2013). A literature review on sampling techniques in semiconductor manufacturing. *IEEE Transactions on Semiconductor Manufacturing*, *26*(2), 188–195.

Obeid, A., Dauzère-Pérès, S., Yugma, C. (2011). Scheduling job families on parallel machines with a bi-criteria objective function under time constraints. In *Proceedings of the multidisciplinary international conference on scheduling: theory and applications (MISTA), Phoenix, Arizona, (USA)* (pp. 645–648).

Obeid, A., Dauzère-Pérès, S., Yugma, C. (2012). Scheduling on parallel machines with time constraints and equipment health factors. In *Proceedings of the* 8^{th} *IEEE international conference on automation science and engineering (case 2012), Seoul, (Korea)* (pp. 401–406).

Pasadyn, A. J., Lee, H., & Edgar, T. F. (2008). Scheduling semiconductor manufacturing processes to enhance system identification. *Journal of Process Control*, *18*(10), 946–953.

Patel, N. (2000). Control-oriented dispatching in semiconductor manufacturing. In *Proceedings of the modeling and analysis of semiconductor manufacturing conference* (pp. 61–66). Arizona (USA): Tempe.

Pesotchinsky, L. (1987). Problems associated with quality control sampling in modern ic manufacturing. *IEEE Transactions on Components, Hybrids, and Manufacturing Technology*, *10*(1), 107–110.

Purdy, M. (2007). Dynamic, weight-based sampling algorithm. In *Proceedings of the international symposium on semiconductor manufacturing* (pp. 1–4). California (USA): Santa Clara.

Sarfaty, M., Shanmugasundram, A., Schwarm, A., Paik, J., Zhang, J., Pan, R., Seamons, M. J., Li, H., Hung, R., Parikh, S. (2002). Advance process control solutions for semiconductor manufacturing, In *Proceedings of advanced semiconductor manufacturing conference and workshop, Boston, (USA)* (pp. 101–106).

SEMATECH. Non-product wafer (NPW) tracking guidelines. http://www.sematech.org/docubase/document/4955aeng.

Stamper, A., Sivaraman, G., Sankar, R. (2012). Advanced excursion control and diagnostics for CMP process monitoring. In *Proceedings of the 22nd annual SEMI advanced semiconductor manufacturing conference (ASMC), Saratoga Springs, (USA)* (pp. 1–3).

Su, A.-J., Jeng, J.-C., Huang, H.-P., Yu, C.-C., Hung, S.-Y., & Chao, C.-K. (2007). Control relevant issues in semiconductor manufacturing: Overview with some new results. *Control Engineering Practice*, *15*(10), 1268–1279.

Su, Y. C., Lin, T. H., Cheng, F. T., & Wu, W. M. (2004). Making the move to fab-wide apc. *Solide State Technology*, *47*(9), 426–434.

Sun, D.-S., Choung, Y. I., Lee, Y.-J., & Jang, Y.-C. (2005). Scheduling and control for time-constrained processes in semiconductor manufacturing. In *Proceedings of the IEEE international symposium on semiconductor manufacturing* (pp. 295–298). San Jose: California (USA).

Wein, L. M. (1988). Scheduling semiconductor wafer fabrication. *IEEE Transactions on Semiconductor Manufacturing*, *1*(3), 115–130.

Williams, R., Gudmundsson, D., Monahan, K., & Shanthikumar, J. G. (1999). Optimized sample planning for wafer detect inspection. In *Proceedings of the IEEE international symposium on semiconductor manufacturing conference* (pp. 43–46). California (USA): Santa Clara.

Yugma, C., Dauzère-Pérès, S., Rouveyrol, J.-L., Vialletelle, P., Pinaton, J., & Relliaud, C. (2011). A smart sampling scheduling and skipping simulator and its evaluation on real data sets. In *Proceedings of the winter simulation conference (WSC)* (pp. 1903–1912). Arizona (USA): Phoenix.