Proceedings of the 2017 Winter Simulation Conference
W. K. V. Chan, A. D'Ambrogio, G. Zacharewicz, N. Mustafee, G. Wainer, and E. Page, eds.

TOWARDS A NEW SIMULATION TESTBED FOR SEMICONDUCTOR MANUFACTURING

Michael Hassoun

Department of Industrial Engineering and Management
Ariel University
Ariel, ISRAEL

Adar Kalir

Department of Industrial Engineering and Management
Ben-Gurion University
Beer-Sheva, ISRAEL

ABSTRACT

We propose the creation of a new set of fab simulation testbeds. Extensions and additional features, not considered in the original MIMAC datasets, shall be incorporated in these new testbeds, thus allowing researchers to evaluate new methodologies with the same frame of reference. To do this, we surveyed the literature and mapped the pertinent research efforts of the past two decades. In this paper, we discuss in detail the various aspects of the new testbeds, in order to receive feedback from the simulation community on the importance of inclusion of some of the items in question; and the verification of the required inclusion of other items. Given the feedback, we aim to generate these testbeds within a year to serve as the new frame of reference for the benefit of the entire semiconductor manufacturing simulation community.

1 INTRODUCTION

Semiconductor manufacturing is known to be of extreme process complexity. This is largely a result of Moore's Law which states that the progress in scaling is such that approximately every two years, the number of transistors per given area (or die size) will double. For the past fifty years this holds true, and indicates that the rate of complexity and emerging challenges in this industry is exponential. What could be considered true or relevant a few years ago is likely outdated today (see Mönch et al, 2013 for an up to date description of the semiconductor manufacturing process.)

In this paper, we focus on the aspect of semiconductor process modeling and its complexities. In the spirit of enabling a satisfactory balance between the level of detail required for proper modeling and the need to not make things more complicated than necessary, we closely examine what has adversely changed over the past two decades since the creation of the MIMAC datasets (Fowler and Robinson, 1995), and should be captured in a new modeling framework and supporting datasets -- and at what level of detail.

We focus our attention to a side effect of this complexity, i.e. the difficulty of modeling. To paraphrase George Box's famous aphorism, let us say that when it comes to drawing conclusions directly usable by a fab manager, "all models are wrong. most are useless." The level of detail necessary to make a model credible to a practitioner is exceedingly high. The time span necessary to develop, validate, and study such a model represents a very high entrance fee for any research team willing to study phenomena related to semiconductor manufacturing. As a result, and also because of the need to study very specific scenarios, most fabs have their own "full fab" simulators, and conduct their research in-house. The stringent rules set to protect their intellectual property often mean that as interesting as these may be, they never find their way to the public domain. Researchers, therefore, are often lagging behind in studying emerging practical aspects, and are in somewhat of a 'danger zone' of studying irrelevant or outdated problems, and/or proposing inapplicable solutions.

There are very few cases of active cooperation where actual fab data is made available to the academic community, which in return creates outputs useful to the practitioner. To address these concerns, and create a common language both between practitioners and researchers and between different research teams, it has been customary to work on "testbeds," i.e. models of simplified fabs publicly available and widely accepted within the community as adequately representative to tackle the burning questions of the hour.

2 TWO DECADES OF SIMULATION BASED ON MIMAC DATASETS

The existing simulation models that have been shared and used to study topics related to semiconductor manufacturing to date are somewhat limited. Despite a significant body of work based on ad-hoc models aimed at highlighting phenomena and comparing techniques, for the most part -- these models are naturally a simplistic representation of reality, targeted to allow primarily proofs of concepts, and these are often dismissed by practitioners as non-representative of the actual fab dynamics or difficult to implement on existing manufacturing execution systems (MES).

In 1994, a five-machine six-step simulation model was devised and published by Intel, that included a great deal of intricacies, denoted as the 'mini-fab' (Kempf, 1994). In this model, two products are modeled as well as test wafers. Test wafers have further restrictions. They cannot be in the same batch on most tools and require setups. Several types of operators and technicians are described, as well as transportation times and stocker size limits. This model has been considered realistic in terms of its level of detail, but unrealistic in terms of scale (i.e. number of machines, process steps, routes, product types, etc.) – hence its original name: 'mini fab'.

A year later, led by SEMATECH, the Measurement and Improvement of Manufacturing Capacity (MIMAC) report was published (see Fowler and Robinson, 1995), proposing a set of methods to identify sources of loss in fabs. This was demonstrated via a dataset of seven detailed simulation testbeds. But the impact of these testbeds proved to go beyond what the MIMAC effort originally intended to accomplish, and Rose (2000) outlined why they have shown to be more relevant than simpler models. For the last 22 years since their publication, these datasets have filled an important void within the simulation community of semiconductor manufacturing, by providing simulation models of realistic scale. As an example, Dataset #1 of MIMAC, which has been extensively utilized over the years, is characterized by two high-volume products, 68 toolsets (i.e., group of identical tools), a total of 211 individual tools, and above 200 process steps per product. The level of detail for the process steps has enabled the modeling of complexities such as batching; cascading; and (sequence dependent) setups as well as rework and in-line scrap at both the wafer and the lot levels (some lots are fully reworked/scrapped, others are partially reworked/scrapped). Details were also provided with respect to the release of lots into the factory per each product (every 3 hours for product 1 and every 6 hours for product 2) resulting in total starts of 4,000 wafers per week. Since at the time, most semiconductor factories were manually operated, 83 human operators of 28 types were included in the dataset, supporting activities such as loading and unloading lots on machines, handling assists, and transporting lots from one step to the next. In their simulation dataset for Autosched AP ©, the embedded scheduling rule is First-Come First-Served (FCFS) with batching restrictions and setup avoidance rules at the relevant toolsets. The equipment cost and other data relevant for the economics of the fab was also provided.

As stated, a large variety of research work has utilized the MIMAC testbeds since their creation even until very recently. They have allowed, for example, Habenicht and Mönch (2002), Mittler and Schoemig (2000), or Rose and Zhugen (2012) to tackle scheduling and dispatching problems. Rose (1999) and Hassoun (2013) have used them to study ways to predict Cycle Time (CT).

It is important to note although there is also literature about transient simulation for non-steady-state periods (e.g. Klein and Kalir, 2006), neither the MIMAC nor the proposed testbeds are aimed to cover these particular cases.

These testbeds are now 22 years old, and it is about time to generate new datasets that are up to date, with features and complexities that have emerged over the past two decades in semiconductor manufacturing. In embarking upon this task, we first review what features should be included in a simulation testbed to make it flexible enough to be used by the research community, but also detailed enough to be accepted by practitioners as representative of actual fab dynamics.

We realize that some of the complexities used in the MIMAC project seem irrelevant today. Economic details or the distinction made between low and high capacity settings do not seem to be of prime interest when tackling operational questions. And with fully automated fabs, the need to model capacity limited machine operators has diminished. On the other hand, additional complexities that have emerged over the years and have become common practice in semiconductor manufacturing, such as dedication schemes or critical queue times (CQT), sometimes also referred to as Process Time Windows (PTW), are now necessary for the appropriate modeling. Features such as setups, hot lots or test wafers, were simply overlooked or greatly reduced within the MIMAC models. Others, like CQT or dedications are a result of technological advances and scaling. In the next section, we review the main features (with some detail) of a proposed future simulation testbed, to serve the semiconductor manufacturing community for the next decade.

3 DIRECTIONS FOR A NEW TESTBED

Much has changed since the publication of the MIMAC report, that requires treatment and representation within a revised and up to date simulation testbed. At the macro level, lot size has standardized at twenty five wafers per lot almost across the industry; and the material handling and transport systems have become automated. At the functional area level, some tools have seen their mode of operation change drastically. Wet tools have shifted to Single Wafer Processing (SWP). Many fabs reported using some form of Lot-to-Lens Dedication (LLD) in lithography tools to mitigate yield losses related to overlay issues between layers. Dynamic qualifications have increasingly become a source of concern for management, especially in areas such as dry etch. The appearance of more and more Critical Queue Times (CQT) throughout the process flow has significantly added to the complexity of WIP management. In the following sub-sections we describe our plans with respect to the future testbed and the way in which these features shall be implemented.

3.1 Types of fabs, products and configurations

We propose two fab configurations that, to the best of our understanding, should cover most of the industry needs. The first one, denoted as Low Volume/High Mix (LV/HM) will represent fabs running a relatively large number of products, each at low volume (with the total volume still being significant). This model shall run in a produce-to-order mode, which implies that each lot has a due date, and the priority of lots is based on the ratio of their remaining time to due date and their remaining raw process time (RPT) to completion, i.e. via the Critical Ratio (CR) rule. In this testbed, about ten different products shall be included. Due to the high cost of reticles, high mix fabs typically have a limited number of reticle sets. This consideration shall find its way into our testbed with a limitation to be placed on the number of available reticles per product and operation, allowing future research to study the pertinence of acquiring additional sets in order to add flexibility to the Litho area. For further reading on scheduling under due dates considerations, we refer the reader to Chung et al. (2014), Zhou and Rose (2011) or Rose et al. (2003).

The second fab will be a High Volume/Low Mix (HV/LM) model that will run only two product routes from the set of routes defined in the previous LV/HM model, each at high volume. With a produce to

stock mode of operation in mind, the scheduling rule will be First-Come First-Served (FCFS). Additionally, reticle sets will not be represented, since, typically in such cases, their cost is easily justified, and they are available in sufficient numbers.
In both models, we shall set the simulation horizon, and pre-determine the lots introduction to the line (see Kacar et al., 2013 for a discussion on lot release to the line). We plan to adopt a relatively stable starts plan for the HV/LM case, and a more chaotic one (reflecting dynamically changing customers' orders) in the LV/HM model.

3.2 Availability modeling

3.2.1 Unscheduled versus scheduled downtime

While the MIMAC testbed did model machine stoppages, it did so in a somehow crude fashion, with no distinction between Preventive Maintenance (PM) and breakage and with a common Mean Time Between Failures (MTBF) distribution. In the proposed model, a great deal of details shall be incorporated around downtime, providing two main benefits: 1) better representation of the stoppage distribution, and 2) enabling the study of PM policies which, certainly in recent years, has become a big factor of consideration (see for example Shin et al, 2016).

- Preventive maintenance incorporation: both time-based PM's (typical for lithography tools) and counter-based (or wafer-based) PM's (typical for etch, dielectric, etc.) shall be modeled. Time based PMs have fixed periods with a starting point set at each tool ('First One At' or FOA), set initially to space the PM's and mitigate the impact on toolset availability. Wafer-based PMs have a counter limit that is set for each tool. While running wafers on different operations, each operation contributes to counter progression at a different rate. When the counter limit is reached, a PM is triggered. In both cases, the duration of these PM's is uniform with a narrow range.
- Breakdowns: two types of unscheduled downtimes shall be incorporated. The Out Of Control (OOC) type of unscheduled downtime pertains to the case in which the downtime is caused by parametric evaluation of Statistical Process Control (SPC) and thus the root causes are likely understood and documented through a Response Flow Chart (RFC). Therefore, in this case, the unscheduled downtime typically behaves with a known duration and a uniform (or normal) distribution. The trouble-shooting type (TS) of unscheduled downtime pertains to the case in which the tool experienced a failure during operation (and it is not related to an SPC monitor) – and there is no known response for the failure, i.e. some sort of pathfinding activity is required. In that case, the response is likely to be unknown in duration with a broad exponential-type distribution.

3.2.2 Tools with sub-entities

Sub-entities in tools pose several questions related to their availability modeling. For the sake of simplicity, we allow for only one level of representation of availability. Therefore, when appropriate, i.e. when the tool behavior is mostly determined by its sub-entities, each of them is considered a separate tool (or entity). When the basic dynamics of the tool are set at the tool level itself, with interference from the sub-entities, the tool would be modelled as a unique entity (and we neglect the inner tool dynamics, because we do not think that such detail is needed to answer questions at the fab level). The following list of examples is provided as a baseline for feedback:

(a) In scope / each sub-entity modelled as a tool (or entity):
    - Reactors (tubes) in Diffusion tools.
    - Dry etch chambers. Tools with single or dual chamber processing (for higher RR) or independent process chambers. This includes tools that have sub-entities with the capability

to perform different process steps in the process flow, e.g. a dry etch tool that one of its process chambers supports a poly step and the other supports a spacer step.

- Lithography tools: two types of lithography tools shall be considered. First is the typical case in which the track (resist coating) is physically linked to the scanner. Their availability is therefore interconnected (or 'linked'). Typically, a good practice in fabs is to try and synchronize the track and scanner PMs to improve their availability. (We leave it as an opportunity for future research to study this topic. The second case is of a detached track, such that some tracks (and scanners) are separate individual entities.

(b) Out of scope are cases such as, for example, chambers in Thin Films (dielectric or metal) tools that may have a common operation in a pre-defined sequence or may work independently with a shared handling robot arm; PM's can be performed on the chambers separately but in such cases would require pre-setup of the entire tool to allow work on part of the chambers (PM partiality). Another type is Hybrid tools. These are tools with mixed chambers such that each chamber is capable of performing different operations (overlapping or not). Lastly, tools known as 'complex availability tools', for which the entity is comprised of several sub-entities (similar to chambers in a Thin Films tool) has an availability model that affects the run-rate of the tool in a non-linear (complex) manner. If some of the chambers are down, the tool can still operate but at a slower run-rate that is non-linear with the number of chambers.

3.3 Machine rate modeling

The basic machine rate (or run-rate) model is similar to MIMAC, with several enhancements. We adopt the structure of a loading time, followed by a processing time for a defined quantity of wafers (per batch/per lot/per wafer) followed by an unloading time. The case of a batch smaller than a lot is ignored (e.g. an implant tool able to run 7 wafers in parallel). In the MIMAC models, the load/unload time required the presence of a human operator, unnecessary in our case. Particularly, the following are captured:

- Batching tools: we adopt a level of detail similar to what was included in MIMAC, batching can be used on certain batch tools between a pre-defined minimum and maximum batch size. The maximum batch size is a technological limitation, while the minimum batch size may be a technological requirement (e.g. in diffusion) or an operational decision aimed at minimizing capacity loss on long process operations. For example of the use of simulation to tackle batching questions, the reader is referred to Habenicht et al. (2004), and to Mönch and Habenicht (2003).
- Cascading tools: In some tools, feeding the tool with a series (cascade) of lots from the same operation (or a subset of operations) allows to absorb the load/unload time between the cascading lots. This does not appear in MIMAC models, and shall be incorporated.
- Setup considerations: sequence dependent setup times and matrices are modelled in MIMAC models, but in limited use, while they are in fact crucial to explain WIP flow variability. We therefore intend to use them extensively. The special case when a PM is required between two different steps shall also be considered

3.4 Functional Area (FA) specific complexities

3.4.1 Photolithography

Scanners being distinctively the most expensive tools, the lithography area (or "Litho" in short) has traditionally been the exclusive fab constraint, at least by design (Kalir and Grosbard, 2014 As a result, it seems that one should be careful with approximations of its characteristics.

First, consider the fact that the number of Litho operations determines the number of layers on the wafer and the number of process steps in the process flow. This should be considered carefully as the re-entrant nature of the process has been shown to be an important factor of variability (see among others Rose, 1998 or Hassoun, 2008). Most of the MIMAC routings include between 10 to 15 Litho layers, up to 20 layers for the longer ones (Dataset #3); Given Moore's Law for technology scaling, this number went up significantly for recent processes and, subsequently, the operations in between any two Litho layers has changed over time, with some of the functional areas becoming more dominant (e.g. dry etch) for patterning and pitching.

With this exponential growth of Litho layers, the overall number of operations in the process has increased accordingly, and it is not uncommon today to find processes that are in the order of a thousand operations, split roughly equally between gate formation operations and metal (inter-connects) operations. Thus, the new testbeds would need to be modified accordingly to reflect the scaling and proportion shift.

The newer technology nodes are also more sensitive, that overlapping discrepancies between Litho layers have become a serious factor of yield loss. To remedy this issue, major semiconductor manufacturers deploy Litho dedication schemes, known as lot-to-lens dedication (LLD). What this means is that once a wafer/lot is processed on a specific Litho tool, it is from there onward committed to return to the same Litho tool in downstream layers, so that the same lens keep on performing the exposure. This mode of operation is a major operational constraint, especially when combined with product setups. In order to balance the future load on steppers, lots of the same type are distributed between different steppers. As a result, two lots waiting for the same Litho step, but with a different "lens history", would not be able to run on the same tool, and thus would trigger a setup necessary for the reticle to move from one stepper to the other (if reticle availability is limited.) In the new testbed, LLD shall be modeled. Although most simulation packages (Autosched AP is no exception) require customization to enable such modeling capability, we strongly believe that it is a must and shall be captured in our testbeds. For a recent example of research on dispatching under LLD constraints, we refer the reader to Chung et al. (2016).

3.4.2 Dry Etch

The main complexity of dry etch lies in its dynamic qualification matrix. Unlike most other tool types, dry etch tool qualifications may constantly changes based on the tool counter between PM's. Immediately after a PM is performed, a certain set of operations are allowed on the tool. After a certain amount of wafers go through the tool, some operations become restricted while other become open and so on, until the next PM takes place. This is particularly crucial since these operations are often part of a CQT segment (more on this in the next section) and shall be represented in the proposed model.

Another policy related to dry etch is soft-dedication with Litho tools. Based on end of line data, certain practitioners reach conclusions about certain combinations of Litho and Etch tools that are better than others in terms of yield. Consequently, soft dedications between these tools are imposed. This behavior is considered as out of scope in our testbeds.

3.4.3 Implant

Implanters present a particular complexity in their operations. Their main (and longer) PM is a replacement of the beam source which has a limited working life. It just so happens that certain sequences of operations are harder on the source than others and significantly reduce the source life duration until PM. We shall represent the dependence of the source PM on operation sequences but we plan not to apply any special policy aiming to reduce the source PM frequency, which leaves the door open for future research efforts in this domain.

3.4.4 Metrology

Quality control is performed in the fab by metrology tools which monitor various operations and tools. The main feature that separates metrology tools from production tools is the fact that they do not process all lots. The arrival rate to the operation is set by a skip rate usually determined based on the risk acceptable for the monitored toolset or segment of operations. When the MIMAC report was published, metrology tools were cheap and were not a concern in terms of capacity, and their representation in the MIMAC dataset is limited. This is not true anymore and over the last few years these became costly and are often actual fabs constraints or bottlenecks. We therefore plan to give metrology operations a representation consistent with this via realistic routes and levels of loading.

3.4.5 Overview of the planned model features

To conclude, we present a map of the intended future features of the testbeds, for the main functional areas in a fab. Unique operations (e.g. laser scribe) are not presented, but will be included in the model.

	Availability						Run Rate										
	Counter Based PM	Time based PM	Sequence dependent PM	Troubleshoot Unscheduled DT	Breakdowns	Sub-entities	Multiple tool configurations	Wafer based run rate	Lot based run rate	Batch based run rate	Cascading	Fixed duration setups	Sequence dependant setups	Lot to tool dedication	Dynamic qualification matrix	Skip rate	Scrap/rework
Litho		✓			✓		✓	✓			✓	✓		✓			✓
Etch	✓				✓	✓	✓	✓							✓		✓
Thin films	✓				✓	✓		✓									✓
Implant			✓		✓			✓			✓		✓				✓
Diffusion	✓					✓				✓							✓
Metrology				✓				✓								✓	
Wet	✓				✓			✓	✓		✓						✓
Planar	✓				✓			✓									✓

Table 1: Planned model features by functional area

3.5 Operational complexities (Non-FA)

3.5.1 Critical Queue Time (CQT) segments (loops)

This item has become a major challenge for the semiconductor industry since the early 2000's (see Scholl and Domaschke, 2000). A critical queue time segment refers to a sequence of operations in the process

flow that must be performed within a pre-specified elapsed time or otherwise the material is at high risk for yield loss and scrap.
The challenge that is posed by any CQT is as follows: when (and at what rate) should lots be releases into the CQT segment? On the one hand, it is desirable to release as much as possible to ensure proper utilization of the tools within, and maximize its output. On the other hand, the more aggressive the release policy, the higher is the risk of lots violating the CQT limit. Hence, there is a need to develop effective rules and logic for the release mechanism.
CQTs can be classified to types. They can be short or long in duration; contain few or many operations; and can be sequential, nested, or embedded. By 'sequential' we mean that they follow one another, i.e. at the operation that one finishes, another CQT begins. 'Nested' refers to the case that there is overlap between two or more queue times. The next CQT begins at an operation that is somewhere in the middle of the previous CQT in the process flow and ends after the ending operation of the previous CQT. An 'embedded' queue time is when a second queue time is entirely embedded within a primary queue time. Figure 1 illustrates all the three types of queue times that shall be captured in our dataset.

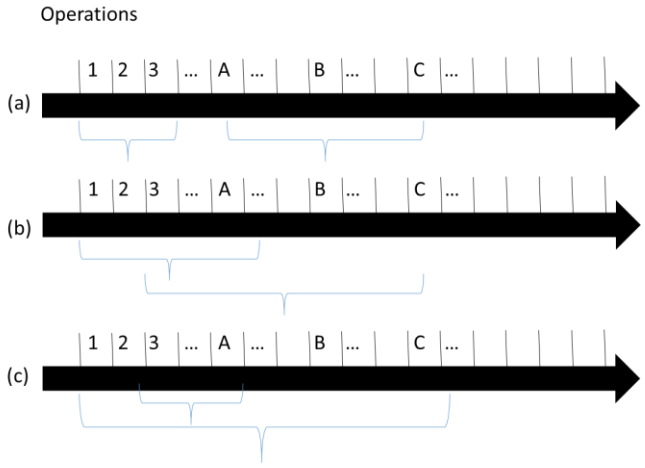

Figure 1: 3 types of queue times – (a) sequential, (b) nested, and (c) embedded.

3.5.2 Hot lots and Non-production (Engineering lots)

Another aspect that was not addressed within the MIMAC datasets is the indirect load on tools, emanating from non-production material. In particular, engineering lots – i.e. lots that serve as pilot lots for controlled engineering experiments and/or yield improvement. Additionally, there has been some published work on the negative impact of hot lots (sometimes also referred to as priority lots) but that was not included either in the original dataset.
Our intent it to include both non-production material and hot lots in the new dataset.
For hot lots, two categories shall be incorporated: 1) the extreme case of a hot lot that holds tools downstream open in advance (denoted as 'super hot lot'), and 2) the case of hot lots that are next-to-run at any operation (i.e. break batches and triggers setups). Engineering lots shall be modelled as short loop hot lots, i.e. lots whose route is a short segment of the regular production process, and that receive hot lot priority. Ziarnetzky et al. (2015) present an example of the complexity induced by special lots (in their case, engineering lots) to the fab operation.

3.5.3 Quality issues

Our representation of quality issues will not significantly differ from the MIMAC datasets. Some operations are characterized by a certain rate of wafer scrapping, or full lots scrapping. Other operations (particularly Litho) send a certain portion of their lots to a specific route for rework. As a result of scrap, arrival rate to downstream operations may be slightly lower. The operations and tools belonging to a rework loop obviously see an increased load by the amount of rework. This level of representation shall be incorporated in order to allow researchers to tackle questions related to rework, or to policies related to scrap, for instance by introducing hot lots to the line to compensate for the final product losses in a produce to order environment. It is important to note that, consistent with MIMAC, we do not plan to incorporate die yield considerations in our model.

3.5.4 Operational Controls (Scheduling rules)

Thus far, we described the physical and technological aspects of the planned testbeds. It is time to address the set of policies used to run it. Typically, these rules are extremely complex and are customized to the business needs and holistic philosophies of the specific companies that employ them -- which is why we see no incentive for devising very complex rules. Nevertheless, all of these can be relatively easily adjusted by practitioners, as needed, and serve as further opportunities embedded within the testbeds.

- **Generic dispatching rule**: we adopt a First-Come First-Served (FCFS) rule to prioritize between regular lots in the HV/LM model, and a Critical Ratio rule for the LV/HM model.
- **Setup avoidance**: setup tools are running under a setups avoidance rule aimed at limiting the time wasted on switching between operations. A tool continues to run the available WIP from its current operation to exhaustion, and only then would switch to a different operation. No priority is enforced between the types of setups. Incidentally, this policy is beneficial to the run rate of tools with cascading.
- **Batching**: by defining a minimum batch allowed (mainly in Diffusion), we set the bar for the losses induced by tools that are not fully loaded. The other side of the coin is obviously that the minimum batch size also impacts the inter-departure time from the operation. However, minimum batch size is not enforced on special lots (hot, super-hot).
- **Hot lots**: the presence of a hot lot in the queue of an operation triggers a stoppage in the first tool to finish a lot, and a setup to run the lot. Once the lot is finished, the tool resumes its regular mode of operation (per generic dispatch or setup avoidance run-to-exhaustion). **Super-hot lots**: as soon as a super-hot lot reaches an operation, the tool performing the next downstream operation is kept idle for it, and, if necessary, performs the required setup. As a result, these super-hot lots are supposed to run through the fab in a cycle time close to the theoretical process time.
- **CQT** policy: Lots will be released into CQT segments according to the following logic: a queue of lots waiting to enter the segment is ordered following the generic dispatch rule. A unique segment bottleneck operation is defined that sets the next release into the segment based on its own next end of operation, offset by the sum of the raw process between the segment first operation and the bottleneck and a predefined buffer aimed at protecting lots from availability variability. The larger the buffer, the less frequent excursions will happen, at the cost of a lower utilization in the segment, and vice-versa. The number of excursions will be collected as a performance measure of the model.

4 CONCLUSION AND NEXT STEPS

In this somewhat interactive paper, we presented the main characteristics of two planned fab simulation testbeds aimed at replacing the excellent but outdated MIMAC models. The extensive use of the MIMAC testbeds in semiconductor manufacturing research for the last two decades shows that there is a strong

need for a modified and up-to-date dataset. The complexities that emerged in the industry since then have prompted numerous additions, while other aspects (such as material handling) have been made simpler by automation and can therefore be dropped. The description provided in this paper, of the proposed testbeds is a call to the community for feedback, ideas, and hopefully, endorsement. MASM 2017 is the natural place for this to happen.

Once reviewed and confirmed by the community, our plan is to provide the testbeds along with a simulation build (Autosched AP and/or another platform) and a description of baseline performance to which future works can refer.

REFERENCES

Chung, Yong H., Kang H. Cho, Sang C. Park, and Byung H. Kim. "Dedication load based dispatching rule for photolithography machines with dedication constraint." In *Proceedings of the 2016 Winter Simulation Conference*, edited by T. M. K. Roeder, P. I. Frazier, R. Szechtman, E. Zhou, T. Huschka and S. E. Chick, Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc., pp. 2731-2739. IEEE Press.

Chung, Yong H., Sang C. Park, Byung H. Kim, and Jeong C. Seo. "Due date control in order-driven fab with high priority orders." In *Proceedings of the 2014 Winter Simulation Conference*, edited by S. J. Buckley, J. A. Miller, A. Tolk, L. Yilmaz, S. Y. Diallo, and I. O. Ryzhov, Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc., pp. 2544-2551. IEEE Press.

Fowler, John, and Jennifer Robinson. *Measurement and improvement of manufacturing capacities (MIMAC): Final report*. Technical Report 95062861A-TR, SEMATECH, Austin, TX, 1995.

Habenicht, Ilka, Lars Mönch, G. Linovszki, and István Molnár. "Evaluation of batching strategies in a multi-product waferfab by discrete-event simulation." In *Proceedings of the 2004 European Simulation Symposium*, pp. 23-28. 2004.

Habenicht, K., and Lars Monch. "A finite-capacity beam-search-algorithm for production scheduling in semiconductor manufacturing." In *Proceedings of the 2002 Winter Simulation Conference*, edited by, Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc., pp. 1406-1413. IEEE Press.

Hassoun, Michael, Gad Rabinowitz, and Shlomi Lachs. "Identification and cost estimation of WIP bubbles in a fab." *IEEE transactions on semiconductor manufacturing* 21, no. 2 (2008): 217-223.

Hassoun, Michael. "On improving the predictability of cycle time in an NVM fab by correct segmentation of the process." *IEEE Transactions on Semiconductor Manufacturing* 26, no. 4 (2013): 613-618.

Kacar, Necip Baris, Lars Monch, and Reha Uzsoy. "Planning wafer starts using nonlinear clearing functions: A large-scale experiment." *IEEE Transactions on Semiconductor Manufacturing* 26, no. 4 (2013): 602-612.

Kalir, A., and D. Grosbard. "On the use of simulation in support of capital utilization," In *Proceedings of the 2014 Winter Simulation Conference*, edited by S. J. Buckley, J. A. Miller, A. Tolk, L. Yilmaz, S. Y. Diallo, and I. O. Ryzhov, Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc., pp. 2617-2627. IEEE Press.

Kempf, K. "Intel five-machine six step mini-fab description." *Intel/ASU Report, Arizona State University* (1994).

Klein, M. and A. Kalir. "A Full Factory Transient Simulation Model for the Analysis of Expected Performance in a Transition Period." In *Proceedings of the 2006 Winter Simulation Conference*, edited by D. Nicol, R. Fujimoto, B. Lawson, J. Liu, F. Perrone, F. Wieland, Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc., pp. 1836-1839. IEEE Press.

Mittler, Manfred, and Alexander K. Schoemig. "Comparison of dispatching rules for reducing the mean and variability of cycle times in semiconductor manufacturing." In *Operations Research Proceedings 1999*, pp. 479-485.

Mönch, Lars, and Ilka Habenicht. "Factory scheduling and dispatching: simulation-based assessment of batching heuristics in semiconductor manufacturing." In *Proceedings of the 2003 Winter Simulation Conference*, edited by D. Ferrin, D. J. Morrice, P. J. Sanchez and Stephen Chick, Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc., pp. 1338-1345. IEEE Press.

Mönch, Lars, John W. Fowler, and Scott J. Mason. "Semiconductor manufacturing process description." *Production Planning and Control for Semiconductor Wafer Fabrication Facilities* (2013): 11-28.

Rose, Oliver. "WIP evolution of a semiconductor factory after a bottleneck workcenter breakdown." In *Proceedings of the 1998 conference on Winter simulation*, edited by J. S. Carson, M. S. Manivannan, D. J. Medeiros and E. F. Watson, Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc., pp. 997-1004. IEEE Press.

Rose, Oliver, and Zhugen Zhou. "WIP-Oriented Dispatching in Complex Manufacturing Facilities." *Decision Policies for Production Networks* (2012): 71-101.

Rose, Oliver, S. Chick, P. J. Sánchez, D. Ferrin, and D. J. Morrice. "Accelerating products under due-date oriented dispatching rules in semiconductor manufacturing." In *Proceedings of the 2003 conference on Winter simulation*, edited by D. Ferrin, D. J. Morrice, P. J. Sanchez, and S. Chick, Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc., pp. 1346-1350. IEEE Press.

Rose, Oliver. "Estimation of the cycle time distribution of a wafer fab by a simple simulation model." In *Proceedings of the international conference on semiconductor manufacturing operational modeling and simulation*, pp. 133-138. San Francisco, CA, 1999.

Rose, Oliver. "Why do simple wafer fab models fail in certain scenarios?." In *Proceedings of the 2000 conference on Winter simulation*, edited by P. A. Fishwick, K. Kang, J. A. Joines, R. R. Barton, Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc., pp. 1481-1490. IEEE Press.

Scholl, Wolfgang, and Joerg Domaschke. "Implementation of modeling and simulation in semiconductor wafer fabrication with time constraints between wet etch and furnace operations." *IEEE Transactions on Semiconductor Manufacturing* 13, no. 3 (2000): 273-277.

Shin, J., Morrison, J.R. and A. Kalir, "Optimization of preventive maintenance plans in G/G/m queueing networks and numerical study with models based on semiconductor wafer fabs," *International Journal of Industrial Engineering: Theory, Applications and Practice*. (2016): V-23, Issue 5, pp. 302-317.

Zhou, Zhugen, and Oliver Rose. "A composite rule combining due date control and WIP balance in a wafer fab." In *Proceedings of the 2011 conference on Winter simulation*, edited by P. White, M. Fu, S. Jain, R. Creasey, J. Himmelspach, Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc., pp. 2080-2087. IEEE Press.

Ziarnetzky, Timm, Lars Mönch, Thomas Ponsignon, and Hans Ehm. "A Reduced Simulation Approach for Wafer Fabs with Engineering Activities." *International Symposium on Semiconductor Manufacturing Intelligence (ISMI)*, 2015.

AUTHOR BIOGRAPHIES

MICHAEL HASSOUN is a lecturer at the Industrial Engineering Department of the Ariel University, Israel. His research interests focus on modeling and management of production systems, with a special interest in Semiconductor manufacturing. He earned his PhD and MSc in Industrial Engineering from Ben-Gurion University of the Negev, Israel, and his BSc in Mechanical Engineering from the Technion, Israel. He was a postdoctoral fellow at the University of Michigan in 2009. His email address is michaelh@ariel.ac.il.

ADAR KALIR received his B.S. and M.S. degrees in industrial engineering and management from Tel-Aviv University, Israel, and his Ph.D. degree in industrial and systems engineering from Virginia Tech. He is a Sr. Principal Engineer at the Fab/Sort Manufacturing network of Intel Corp., and an Adjunct Professor at Ben-Gurion University, Israel. He also serves as a co-chair of the IEEE Technical Committee on SMA (Semiconductor Manufacturing Automation.) His email is kalira@post.bgu.ac.il