Proceedings of the 2016 Winter Simulation Conference
T. M. K. Roeder, P. I. Frazier, R. Szechtman, E. Zhou, T. Huschka, and S. E. Chick, eds.

USING SIMULATION TO IMPROVE SEMICONDUCTOR FACTORY CYCLE TIME BY SEGREGATION OF PREVENTIVE MAINTENANCE ACTIVITIES

Kosta Rozen

Intel Corporation
Fab/Sort Manufacturing Division, Fab28
2 HaZoran ST
Qiriat-Gat, 82109, ISRAEL

Néill M. Byrne

Intel Corporation
Fab/Sort Manufacturing Division, Fab24
Collinstown Industrial Estate
Co. Kildare, W23CX68, IRELAND

ABSTRACT

Semiconductor manufacturing is a very costly and time-consuming process and any reduction in cycle time (CT) can result in a significant cost saving due to shorter time-to-market, reduced inventory and improved yield. This paper examines the topic of PM (Preventative Maintenance) segregation where the goal is to determine the optimum PM frequency that results in reduced fab CT. Much of the literature on this topic limits the scope of the study to CT performance of an individual toolset, without examining the wider impact on overall fab CT. The goal of this paper is to examine the fab-wide impact of PM segregation projects and reveal critical factors that will allow to identify toolsets with the most potential to improve overall fab CT by splitting PMs. Each toolset identified using these factors may improve fab CT by 1% or more, resulting in a very significant gain, if applied on multiple toolsets.

1 INTRODUCTION

Semiconductor factories, commonly known as fabs, are unique manufacturing environments, characterized by an extremely long manufacturing process, consisting of hundreds of sequential processing steps, performed by shared toolsets subject to restrictions such as step dedications, dynamic lock-outs of tools to steps, process time window limitations (manufacturing steps that must be completed within a predetermined amount of time) and rapidly changing product mixes. Operationally, the process has many sources of variability, mainly induced by a combination of machine down-time and highly re-entrant process flow.

One of the main goals of a typical high volume semiconductor fab are reducing cycle time (CT), given a defined fab capacity and loading plan. Reduced CT results in lower inventory costs, improved quality, better response to demand changes and shorter time-to-market for new products.

There are numerous methods for cycle time reduction, which can be roughly grouped in 3 main categories:

- **Operational improvements**: identification and efficient utilization of constraints, effective scheduling methods, elimination of non-value-added activities using lean manufacturing methods, variance reduction, standardization and efficient management of process time windows etc.
- **Process improvements**: removing operations, removing restrictions, removing process time windows, merging toolsets, reducing process time, etc.
- **Equipment improvements**: improving toolset availability, reducing availability variance by reducing frequency of unscheduled (unexpected) events or duration of scheduled (planned) events, scheduling planned events at desired times in order to reduce impact of unavailable tools.

Direct experimentation on the factory line can be expensive and infeasible, therefore models of the system are typically employed to assess the gain from a particular improvement and identify which improvements deliver the most economic gain. There are several modelling strategies, such as discrete event simulation, data mining and analytical approximations, discussed in depth by Hopp and Spearman (2011) .

A large portion of machine down-time is spent in Preventive Maintenance (PM) activities, which reduce the likelihood of unscheduled down-time events and ensures that the machine runs within defined control limits. Based on SEMI E10-0701, a typical fab PM comprises three sequential parts: pre-PM preparations, intrusive maintenance, and tool qualification. Pre-PM preparations and tool qualification are considered as Non-Value-Added (NVA) activities, while intrusive maintenance is a Value-Added (VA) activity. NVA activities are sometimes also referred to as PM Setups (PMSU).

Each PM has a duration and a PM frequency which is generally either defined in time between PM or wafers between PM. These times can be changed by modifying PM content or frequency, in order to maximize a given objective function. Most common PM optimization policies strive to maximize toolset availability (Barlow and Hunter 1960), but recent studies have shown that reduced variability may be a better objective function (Tirkel 2013).

2 PROBLEM STATEMENT

The topic of PM segregation (also called PM splitting) has been covered by Kalir (2013), Regev and Benson-Karhi (2015) amongst others. Longer PM activities increase the variance of WIP flow in the manufacturing line, resulting in increased cycle time. Reducing the PM duration usually leads to reduced variability and reduced cycle time. Hence, there is an incentive to reduce PM duration, perhaps even at expense of increased PM frequency and duration (Kalir 2013; Tirkel 2013).

When segregating a PM into two or more parts, we split the value added activities of the PM, but each new PM has to perform the PM Setups (PMSU). For example, if the value added PM time is 50 hours and non-value-added PMSU time is 10 hours, overall duration of PM is 60 hours. Splitting the PM VA time into 2 equal parts results in 2 PMs, lasting 35 hours each (25 VA hours + 10 NVA hours). Performing both halves of the split PM takes 70 hours instead of 60 hours of the original PM.

Therefore the problem of assessing the benefit from segregation of a specific type of PM is an optimization problem, examining the trade-off between reducing variance (due to shorter PM durations) and increasing tool down-time (due to more frequent PMs). In practice this can be a time-consuming task as there are hundreds of PM types in a typical fab. This paper proposes a generic guideline that will quickly identify a manageable amount of PMs that can be further examined in more detail and optimized. The initial filter involves selecting PMs with relatively long durations and relatively short PMSUs. This is a reasonable approach, and it will usually reduce the amount of "candidate PMs" from hundreds to several dozen.

The goal of this study is to examine the impact of additional toolset properties (factors) and identity the relative importance of each factor on expected gain from PM segregation. Following that examination, we will provide a set of heuristic guidelines for quick selection of toolsets with the greatest potential gain. The following 4 toolset factors will be examined: toolset loading, number of machines, number of operations and position in the manufacturing line. The objective of the current study is to identify which of these factors are more important than others, and as such should be used as main filters for selecting "candidate PM" for a more detailed segregation study.

Each of these factors can be examined in greater detail, but for the purpose of this study, only two values were selected for each factor: one typical value and one extreme value. By comparing the gain from PM segregation between scenarios with these two values, it is possible to estimate the relative importance of each factor. Future studies may then be able to focus on factors that carry the most impact and perform a more detailed study and sensitivity analysis of each parameter on the potential gain from PM segregation.

Other relevant parameters not examined in this study are PM and PMSU duration. Both of them have a major impact on the gain from PM segregation, but their impact have already been examined in other studies, like Kalir (2013). Therefore, these parameters were set to have constant representative values, as it is safe to assume that toolsets with short PMs or with very long PMSU durations are not good candidates and will be excluded from examination.

When analyzing the potential gain from PM segregation or PM scheduling, most studies (Kalir 2013; Morrison et al. 2014), focus on optimization of cycle time of the toolset under examination, without accounting for the cumulative effect of variance propagation to subsequent toolsets in the line. This is an incomplete analysis due to re-entrant nature of the fab production flow and complex scheduling rules on each toolset. Therefore, if we are to assess whether a project is beneficial for the entire fab (global optima vs local optima), it is necessary to analyze the CT impact on a full fab and not just the toolset itself. Using a full fab simulation model is required in order to assess the impact of local change on the entire fab. Analyses that limit results to the toolset under examination skew results in favor of non-segregation, as they capture the full loss from reduced availability due to PM segregation in the toolset's cycle time but do not comprehend potential gain in downstream toolsets CT due to variance reduction resulting from PM segregation.

3 SOLUTION METHODOLOGY

3.1 Simulation Model

In order to analyze the impact of proposed experiments, a full factory simulation model was used, based on AutoSched AP simulation software provided by Applied Materials. The software was further enhanced by multiple extensions developed by Intel's Decision Support Technology group, described in (DeJong and Fischbein 2000). Following the classification method in Jimenez et al. (2008), the details level of the simulation model may be classified under the third quadrant with a high level of capacity detail (level 5-6 per their classification) and a low level of AMHS detail (level B).

This simulation model has been used extensively and produces highly accurate and well-validated results. It has consistently captured fab CT within several percent of actual observed fab CT, and thus operates as a good test-bed for our experimentation. The model is highly complex and factors in process times, scheduled and unscheduled downtimes, setups, batching and dispatching rules, layer lock outs (inability of certain tools to run specific operations that tools of this toolset should usually run), lithography reticles, tool dedications, rework, test wafer production, time-constrained production sequences and many other typical fab complexities.

Typically, the model is used to examine transient simulations without a warm-up period, but for the purpose of this study, it was adjusted to a steady-state simulation by removing time-related changes and adding a warm-up period that lets the simulation stabilize regardless of initial starting conditions. The proposed simulation has not yet been validated as it requires a fab to be in relatively steady-state. Therefore, this constitutes a direction for further work.

3.2 Design of Experiments

The following 4 toolset factors will be examined in order to determine the gain from PM segregation on a toolset with different values of these factors: toolset loading, number of machines, number of operations and toolset position in the manufacturing line.

For each examined factor, two values were selected: one typical value (based on average expected values in a semiconductor fab) and one extreme value (in order to examine the impact of this factor). By comparing gain from PM segregation between scenarios with these two values, it is possible to estimate the relative importance of each factor. Selected values for each of the examined factors are listed below:

Table 1: List of selected values of each examined factor.

Factor	Typical Value	Extreme Value
Toolset loading	80% (Non-Constraint)	95% (Constraint)
Number of machines	6	2 (Few)
Number of operations	3	20 (Many)
Toolset position in the line	Spread evenly	Close to start of the line

To examine the impact of each of these toolset factors, the duration of base PM was set to the same value: 40 hours – based on an average PM duration for all PMs exceeding 24 hours in a real fab (shorter PMs are usually not good candidates for segregation due to major impact of PM setup time on segregated PMs). The duration of PMSU time in each scenario was set to be constant as well: 4 hours – based on an average NVA portion of a long PM in an average toolset, mostly consisting of repeated runs of calibration / cleaning / monitor wafers on the tool and completing the PM only when these wafers show expected results. Intuitively, the PM and PMSU durations have critical impact on gain from PM splitting, but the purpose of this article is to define a guideline for identifying a group of toolsets that are likely to have this gain, assuming reasonably long PM duration and reasonably short setup duration. All other toolset parameters, like mean time between PMs, unscheduled down-time, etc. were held constant in all scenarios to rule them out of interfering with experiment results. The only change between the scenarios is the relevant parameter (factor) of the examined toolset and its PM policy.

Two PM options were simulated for each scenario: original PM (without segregation) and PM split to two PMs with equal durations. Further PM splitting was ruled out of the examination as if there is no gain in splitting a PM into two parts, there is unlikely to be any in further splitting it into three or more parts. This notion is backed up by Kalir (2013) who showed that for reasonable values of PMSU (4 to 6 hours), the optimal gain is in splitting PM into 2 parts even when the duration is as long as 120 hours. In addition, further PM splitting is not always practical from a resources perspective.

Table 2: List of simulated PM duration scenarios.

PM	Single PM Value Added Time [hrs]	Single PM Duration [hrs]	Overall PMs Duration [hrs]	% of PMSU Time in PMs
Original	36	36 + 4 = 40	40	10%
Split (to 2 parts)	36 / 2 = 18	18 + 4 = 22	22 X 2 = 44	18.2%

In order to examine the impact of each factor, a total of 10 simulation experiments were run, with each experiment isolating one parameter only and keeping the values of the remaining parameters constant. The experiments are outlines in table 3.

For each scenario shown in table 3, 8 replications of a steady state simulation were run, resulting in a total of 80 simulation runs. Length of the warm-up period was defined as 20 weeks, based on the longest stabilization time observed in all simulated replications. Length of the single simulation (including warm-up period) was set at 100 weeks, resulting in 80 weeks of usable data, after excluding 20 warm-up weeks. The selected period of 80 weeks is significantly longer than the CT of a typical fab product, hence this length is sufficient for observing the long-term effect that each factor has on overall line performance. Running several replications instead of one long continuous simulation allows to separate typical fab variability perturbations (which take a long time to dissipate) from the variability resulting from experimental controls (Dummler 2000).

Table 3: List of simulated experiments.

EXP#	Examined Factor	PM	Loading	Machines	Operations	Toolset Position
1	Typical Toolset	Original	80%	6	3	Spread
2		Split	80%	6	3	Spread
3	Constraint	Original	**95%**	6	3	Spread
4		Split	**95%**	6	3	Spread
5	Few Machines	Original	80%	**2**	3	Spread
6		Split	80%	**2**	3	Spread
7	Many Operations	Original	80%	6	**20**	Spread
8		Split	80%	6	**20**	Spread
9	Line Start	Original	80%	6	3	**Line Start**
10		Split	80%	6	3	**Line Start**

4 SIMULATION RESULTS AND ANALYSIS

Averages were calculated on the following statistics (for steady-state period only, after discarding warm-up period data):

- **Normalized Fab Cycle Time –** measures the average velocity of the entire line. Fab CT of all experiments was normalized based on baseline fab CT of scenario “Typical Toolset with Original PM schedule”. In order to avoid local optima (optimizing PM duration for lowering CT of the toolset, without regard for downstream impact of variance generated by that toolset) we will define the true impact of each toolset factor based on this statistic.
- **Normalized Toolset Cycle Time** – measures the performance of the toolset being examined. Toolset CT of all experiments was normalized based on baseline Toolset CT of scenario “Typical Toolset with Original PM schedule”.
- **Toolset-Fab CT Gain Ratio –** measures what percentage of fab CT gain is attributed directly to CT gain on the examined toolset, as opposed to the gain derived from reduced on the manufacturing line. This statistic is calculated as toolset CT Gain in hours divided by fab CT Gain in hours. For example, if examined toolset has gained 1 hour of CT due to PM splitting, while the fab gained 10 hours, the value of this statistic will be 10%, meaning that major part of fab CT gain is due to reduced down-stream variance and not from direct CT reduction of the toolset.
- **Statistical significance of difference –** result of statistical test, that assesses the probability that two samples (CT in Original PM and Split PM scenarios) belong to the same population. The higher the percentage, the higher the confidence that relevant factor has statistically significant impact on Toolset or Fab CT.

The summary of all collected statistics is listed in Tables 4 and 5 below, with Gain% column representing percentage of CT gain resulting from PM segregation in this scenario (calculated as % of improvement in CT from same scenario with Base PM policy). Normalized Fab CT is presented in Figure 1, while Normalized Toolset CT is depicted in Figure 2.

Table 4: Simulation results - Normalized toolset CT.

Experiment	Base PM	Split PM	Gain%	Statistical significance of difference
Typical Toolset	100.0%	92.1%	7.9%	83%
Constraint	709.5%	881.4%	-24.2%	100%
Few Machines	261.7%	234.0%	10.6%	89%
Many Operations	249.7%	225.0%	9.9%	93%
Line Start	84.4%	84.6%	-0.3%	8%

Table 5: Simulation results - Normalized fab CT.

Experiment	Base PM	Split PM	Gain%	Statistical significance of difference	Toolset-Fab CT Gain Ratio
Typical Toolset	100.0%	99.9%	0.1%	5%	99.2%
Constraint	104.7%	106.5%	-1.7%	90%	65.5%
Few Machines	105.8%	104.7%	1.1%	63%	16.7%
Many Operations	103.7%	102.4%	1.3%	80%	12.8%
Line Start	96.5%	96.6%	-0.2%	18%	0.9%

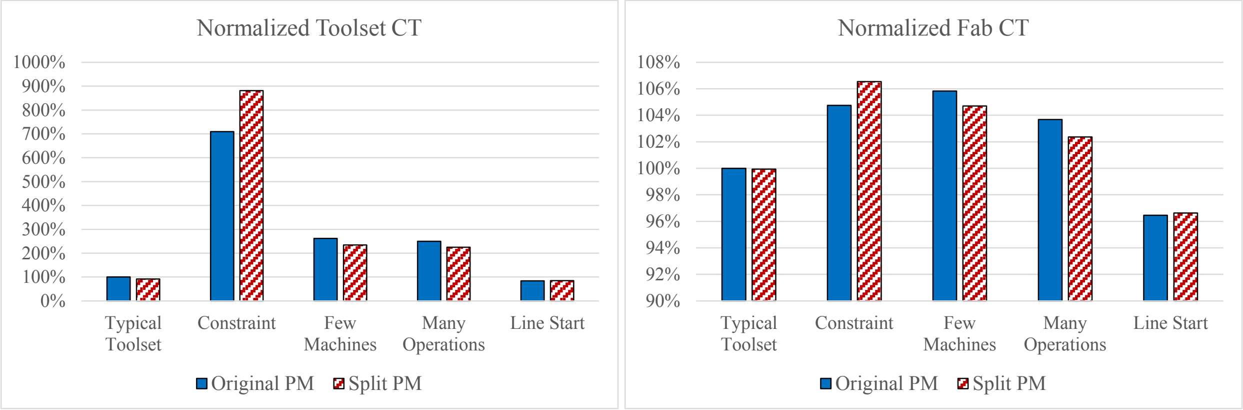

Figure 1 and 2: Charts of normalized toolset and fab CT for all examined scenarios.

When applying PM segregation on a **typical toolset**, there is a slight gain on toolset level, but not on fab level. Fab gain of 0.1% is not statistically significant, and if it exists it's fully derived from direct CT gain of the toolset and not from variance reduction (as the typical toolset is not a major variance generator, not being a constraint and having 6 machines). This is a good example of toolset that has a relatively large CT gain due to PM segregation, but has no real impact on overall fab CT.

When applying PM segregation to a **constraint toolset**, there is a negative impact both on the toolset itself (24.2% loss of CT compared to an original toolset) and on fab-wide CT (1.7% loss). 65.5% of fab CT loss is derived from increased toolset CT, and remaining 34.5% - due to increased variance generated by the constrained toolset, that becomes even more constrained in the PM segregation scenario because of increased down-time resulting from additional PMSU. This is an important finding and carries a practical impact when deciding the cases in which PM segregation is expected to be beneficial.

When applying PM segregation on a **toolset with few machines**, the gain is positive both in toolset and fab level: 1.1% gain in fab CT, most of it (83.3%) deriving from variance reduction and not from CT reduction on the examined toolset. As expected, the gain to toolset CT is larger than in a typical toolset, demonstrating that PM segregation has larger gain in a toolset with few tools, but the fab-wide gain is much more significant than toolset-level gain. Still, statistical significance of fab gain is only 63%, and more replications are required in order to increase the confidence levels (direction for future research).

When applying PM segregation on a toolset with many operations, the gain is positive both in toolset and fab level: 1.3% gain in Fab CT, most of it (87.2%) deriving from variance reduction and not from CT reduction on the examined toolset. Examined from a manufacturing physics perspective, the results are reasonable and expected: having long PMs on a toolset having many operations creates significant variance in many points in the manufacturing line. PM segregation reduces that variance, resulting in significant gain to overall fab CT. Toolsets with few machines or many operations constitute good examples of toolsets where PM segregation gain on fab level is much higher than the CT gain on the relevant toolset, due to cumulative effect of reduced variance on downstream toolsets. In this experiment as well, it is recommended to run additional replications to increase the significance of the results (80%).

When applying PM segregation on a toolset located in start of the line, there is no statistically significant gain neither on toolset nor on fab level, meaning that location of the toolset makes no major difference on gain from PM segregation.

5 CONCLUSIONS AND RECOMMENDATIONS

This paper has identified the most opportune toolset types to target as candidates for PM segregation based on the characteristics of the toolset, whether it is a constraint, has a low tool count (low parallel processing capacity), performs many reentrant steps or is positioned at the start of the line.

The results indicate that one of the most favorable characteristics for PM segregation in terms of both local toolset improvement and production line performance is number of machines. Our results indicate that the gain from splitting PM's with low parallel processing capacity (few machines) is significant and that the majority of the gain comes from reducing the downstream variability in the line. Here, the cost of decreasing the toolset availability due to splitting its PM and thereby introducing additional non-value added time can be absorbed by the toolset. The subsequent reduction in output variability from the toolset eliminates the 'boom-bust' phenomenon generally associated with tools with large burst capacity but with few tools and hence higher probability of being "lines-down".

Similarly toolsets that have many operations across the line and are therefore subject to high reentrancy can also benefit significantly from PM segregation. In this case as well, the gain was seen on both the local toolset cycle time and on the overall line cycle time, with the vast majority of the gain being attributed to line cycle time. The conclusion here is that tools with many operations (even if they are not deemed constraints) are good variability modulators for the factory and can control the variability propagation throughout the line by reducing the variability of their availability.

Splitting or segregating PMs on constraint tools has an adverse effect on both toolset and factory cycle time. Here, the availability of the toolset is a high modulator for factory output performance and hence any reduction in availability due to PM splitting removes 'precious' capacity from the constraint. Furthermore any potential downstream improvement in variability is outweighed by the reduced capacity and restricted flow.

It was also shown that tools positioned at the start of the line with all other parameters fixed do not see any statistically significant change by splitting PMs and therefore this attribute should not be considered as an individual positive characteristic for PM splitting decisions. The result suggests that toolset with operations spread throughout the line still generates sufficient variance on downstream toolsets, so that there is no statistically significant difference between it and toolset located in the start of the line. Still, it is expected that toolsets located only at the end of the line would have less impact on overall fab CT than toolsets located in the start of the line or spread throughout the line. This experiment was not inside a scope of the current article and may be examined in a follow-up study.

Another learning from this study is that measuring the true gain from PM segregation or other variance-impacting policies, requires measurement of overall fab CT, which is the real objective of the improvement. CT gain on a particular toolset is not a reliable estimator for gain in overall fab CT, as demonstrated by the typical toolset experiment that has CT gain on a toolset level but barely features on a fab level, and by experiments with few machines and many operations, that have fab CT gain significantly exceeding toolset CT gain.

In conclusion, it is generally considered that reducing variability of available capacity can lead to a more predictable factory output and reduce cycle time. However, improving variability of availability by reducing the impact of PM through PM splitting is not a panacea and comes with the additional cost of introducing non-value added maintenance related activities. It has been shown in this paper that there are only certain types of candidate tools that will improve factory velocity by segregating PMs. Most notably, non-constraint toolsets with many operations, few machines, long PMs and where possible short PMSUs are the best candidates for selection. Our results have shown that splitting PMs at toolsets with these individual characteristics showed an improved line performance and by deduction it is expected that tools combinations of these 'positive' characteristics will see an improvement as well. At best case, the improvement will be compounded due to combinations of these 'positive' characteristics, and at worst case, it will be close to an improvement based on their best individual characteristic. Furthermore, it is notable that much of the gain comes from downstream variability reduction, which is often not examined in these types of studies. As semiconductor manufacturing becomes more complex and the manufacturing cycle time increases, any gains to the manufacturing line will become more significant and a more holistic approach to PM segregation studies may be required.

ACKNOWLEDGMENTS

We express our appreciation to Dr. Adar Kalir, for his invaluable direction and advice.

REFERENCES

Barlow, R., and L. Hunter. 1960. "Optimum Preventive Maintenance Policies." *Operations Research* 8(1):90-100.

DeJong, D.C., and S.A. Fischbein. 2000. "Integrating Dynamic Fab Capacity and Automation Models for 300mm Semiconductor Manufacturing." In *Proceedings of the 2000 Winter Simulation Conference*., edited by J.A. Joines, R.R. Barton, K. Kang, and P.A. Fishwick, 1505-1509. Society for Computer Simulation International San Diego, CA, USA.

Dümmler, M.A. 2000. "Analysis Of The Instationary Behavior Of A Wafer Fab During Product Mix Changes." In *Proceedings of the 2000 Winter Simulation Conference*., edited by J. A. Joines, R. R. Barton, K. Kang, and P. A. Fishwick, 1436-1442. Society for Computer Simulation International, San Diego, CA, USA.

Hopp, W. J., and M. L. Spearman. 2011. *Factory Physics*. Waveland Press, 2011.

Jimenez, J.A., G.T. Mackulak, and J. W. Fowler. 2008. "Levels of Capacity and Material Handling System Modeling for Factory Integration Decision Making in Semiconductor Wafer Fabs." In *IEEE Transactions on Semiconductor Manufacturing* 21(4):600–613.

Kalir, A.A. 2013 "Segregating Preventive Maintenance Work for Cycle Time Optimization." In *IEEE Transactions on Semiconductor Manufacturing* 26:125-131.

Morrison, J.R., H. Kim and A.A. Kalir. 2014. "Mean Cycle Time Optimization in Semiconductor Tool Sets via PM Planning With Different Cycles: a G/G/m Queueing and Nonlinear Programing Approach." In *Proceedings of the 2014 Winter Simulation Conference* edited by A. Tolk, S. D. Diallo, I. O. Ryzhov, L. Yilmaz, S. Buckley, and J. A. Miller. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Regev, I. and Benson-Karhi, D. 2016. “Segmentation of the Expected Duration of Maintenance Activities in Semiconductor Fabs.” *IEEE Transactions on Semiconductor Manufacturing*; accepted for publication (June 2016).

SEMI E10-0701, Specification for Definition and Measurement of Equipment Reliability, Availability and Maintainability (RAM). Accessed March 3, 2016.

Tirkel, I. 2013. “The Effectiveness of Variability Reduction in Decreasing Wafer Fabrication Cycle Time.“ In *Proceedings of the 2013 Winter Simulation Conference: Simulation: Making Decisions in a Complex World*. 3796-3805. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

AUTHOR BIOGRAPHIES

KOSTA ROZEN received his M.Sc in Industrial Engineering and Management from Ben-Gurion University, Israel. He is currently a Senior Simulation Engineer at Intel Corporation in Qiriat-Gat, Israel. He is also an Adjunct Lecturer with Ben-Gurion University, Israel. His email address is kosta.rozen@intel.com.

NEILL BYRNE received his M.Sc and PhD in Industrial Engineering from Dublin City University, Ireland. He is currently a Simulation Engineer at Intel Corporation in Leixlip, Ireland. His email address is neillbyrne@gmail.com.