A literature review on variability in semiconductor manufacturing: the next forward leap to Industry 4.0

Kean Dequeant, Philippe Vialletelle, Pierre Lemaire, Marie-Laure Espinouse

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.

A LITERATURE REVIEW ON VARIABILITY IN SEMICONDUCTOR MANUFACTURING: THE NEXT FORWARD LEAP TO INDUSTRY 4.0.

Kean Dequeant[1,2]
Philippe Vialletelle[1]

[1]STMicroelectronics,
F-38926, Crolles Cedex, FRANCE

Pierre Lemaire[2]
Marie-Laure Espinouse[2]

[2]Univ. Grenoble Alpes, CNRS, G-SCOP, 38 000
Grenoble, FRANCE

ABSTRACT

Semiconductor fabrication plants are subject to high levels of variability because of a variety of factors including re-entrant flows, multiple products, machine breakdowns, heterogeneous toolsets or batching processes. This variability decreases productivity, increases cycle times and severely impacts the systems tractability. Many authors have proposed approaches to better model the impact of variability, often focusing on specific aspects. We present a review of the sources of variability discussed in the literature and the methods proposed to manage them. We discuss their relative importance as seen by the authors as well as the limits current theories face. Finally, we emphasize the lack of research on some critical aspects related to High Mix Low Volume fabs. In this setting, the ability of practitioners to predict and anticipate the effects of changing product mix and client orders remains challenging, delaying the transition of semiconductor manufacturers towards Industry 4.0.

1 INTRODUCTION

Industry 4.0 is said to be the next industrial revolution. The proper use of real-time information in complex manufacturing systems is expected to allow more customization of products in highly flexible production factories. Semiconductor High Mix Low Volume (HMLV) manufacturing facilities (called fabs) are one example of candidates for this transition towards “smart industries”. However, because of the high levels of variability, the environment of a HMLV fab is highly stochastic and difficult to manage. The uncontrolled variability limits the predictability of the system and thus the ability to meet delivery requirements in terms of volumes, cycle times and due dates.

Typically, the HMLV STMicroelectronics Crolles 300 fab regularly experiences significant mix changes that result in unanticipated bottlenecks, leading to firefighting to meet commitment to customers. The overarching goal of our strategy is to improve the forecasting of future occurrences of bottlenecks and cycle time issues in order to anticipate them through allocation of the correct attention and resources. Our current finite capacity projection engine can effectively forecast bottlenecks, but it does not include reliable cycle time estimates. In order to enhance our projections, better forecast cycle time losses (queuing times), improve the tractability of our system and reduce our cycle times, we now need accurate dynamic cycle time predictions.

As these types of projections are made considering finite horizons (e.g., see Mhiri et al. 2014), the dynamic cycle time models to develop are closer to clearing functions (see Kacar et al. 2012) than queuing theory. Using historical data to derive these models is not appropriate, since the behavior of the toolsets changes with the mix of products and lots priorities. Simulation is a promising tool since it allows to create controlled and repeatable scenarios. However, creating accurate simulations for this purpose requires a deep understanding of the main factors responsible for the generation of queues, referred to as

"sources of variability" in this paper. It also requires a deep understanding of how each source of variability contributes to high cycle times in practice.

The literature of variability in semiconductor manufacturing focuses on efficiency: In this context, we interpret variability as everything responsible for efficiency losses by means of non-productive waiting times. What we therefore call the intrinsic variability of a manufacturing system (which we will simply refer to as variability) is its corrupting tendency to generate uncontrolled unevenness both in its flows and in the capacities of its manufacturing resources, resulting in local efficiency losses, i.e. non-productive waiting times. In the effort towards flexible and agile manufacturing, this article seeks to give an exhaustive overview of the various facets and specificities of variability arising in the complex environment of semiconductor manufacturing. This article is organized as follows: Section 2 illustrates the main characteristics of semiconductor manufacturing as found in the literature and confirmed by our own operational experience; the impact of variability is then explained and the main sources of variability in a semiconductor environment are listed. Section 3 provides an extensive study of the aforementioned sources of variability based on the literature and our field experience. The final part (Section 4) discusses the challenges faced when considering the sources of variability in mathematical or simulation models: Using real fab data, we show that some complex behaviors may often be overlooked when trying to capture the sources of variability.

2 THE COMPLEXITY OF SEMICONDUCTOR MANUFACTURING

2.1 Overview Of Semiconductor Manufacturing

Traditional production or assembly lines are usually organized to build products starting from raw materials or components, progressively transforming or assembling them in order to deliver "end products" or "finished goods". Such factories manage a linear flow of products going all in one direction from the beginning of the line to the end of the line. Regardless of whether process steps are performed by operators or machines, processing times are subject to inconsistencies, machines are subject to breakdowns, and operators and secondary resources are sometimes unavailable.

Semiconductor fabs are, like other industries, subject to the above mentioned "disturbances". Their main characteristic however, justifying the label of "complex environment", is the reentrancy of their process flows. Microchips are produced by stacking tens of different layers of metals, insulators and semiconductor materials on silicon substrates (called wafers). The similarities of layers, added to the extreme machine costs, lead the same groups of machines (toolsets) to process different layers. Hence products are processed several times by the same machines over the course of their hundreds of process steps. Dedicating machines to steps is only (partially) possible in very large facilities (such as memory fabs) where thousands of tools are generally producing one or two different products; these factories operate in what is generally referred to as HVLM: High Volume Low Mix manufacturing.

The second characteristic of semiconductor manufacturing is that it involves different types of physical processes that require different types of machines (see Mönch et al. 2011). Batching machines, such as deposition or diffusion furnaces for example, will process products in batches of different minimum and maximum sizes according to the specific recipe and process step. Other machines, such as cluster tools, consist of process chambers or modules sharing the same robotics. Depending on the products being processed at the same time, resource conflicts may appear within the tool, rendering processing times even more inconsistent. Many other characteristics may also be cited here as semiconductor manufacturing is above all a matter of technology and then only a matter of efficiency. A very good example is that most semiconductor vendors still propose "lots of 25 wafers" as manufacturing units, whereas the move to larger wafer sizes should already have challenged this paradigm.

A third characteristic of semiconductor fabs is the business model they use. While High Volume Low Mix units usually produce in the range of 3 to 4 different products at the same time over one or two

different technology generations, High Mix fabs propose a wide variety of products to their customers over several technology nodes. A difficulty in this case is that specific process steps may only be required by very few products. Therefore only 1 or 2 tools may be used to process a given group of steps. Moreover, qualifying a "technology level" on a tool requires time (as functional tests can only be performed on finished products) and money (as "testing wafers" cannot be sold). Therefore, not all tools of a given toolset will be qualified on all the operations associated to this toolset. High Mix fabs, as they address a wide range of customers and products, generally follow a make-to-order policy. Given the range of products, demands cannot be satisfied by stocks and lots are therefore started on demand with different requirements in terms of delivery times. Production control techniques and rules are then introduced to manage different levels of priorities necessary to achieve the individual required speeds.

2.2 Impact Of Variability

As utilization, as well as variability, also creates higher cycle times, one way to visualize the intrinsic variability of a manufacturing system is through "operational curves" or "cycle time throughput curves". These curves are found in 20 of the papers that we reviewed, and used in 12 of them to explain the impact of variability (Brown et al. 2010, Delp et al. 2006, Etman et al. 2011, Ignizio 2011, Kim et al. 2014, Martin 1999, Robinson et al. 2003, Schoemig 1999, Shanthikumar et al. 2007, Tirkel 2013, Wu 2005, and Zisgen et al. 2008). These curves represent the Xfactor (the ratio between cycle time and raw processing time) as a function of the utilization of available capacity at different variability levels (as illustrated in Figure 1.A). Figure 1.A shows that, for any given utilization, a higher level of variability translates into longer cycle times.

Another less obvious impact of variability is the capacity loss it generates. Robinson et al. (2003) indeed titled their article "**Capacity Loss** Factors in Semiconductor Manufacturing." The link between capacity loss and variability arises from the desire of semiconductor manufacturing companies to minimize their production costs by maximizing the utilization of their assets (building, personal and equipment). To guarantee an acceptable cycle time, this generally translates into a maximum allowed Xfactor (the ratio between cycle time and raw processing time). Figure 1.B shows that a first part of capacity (around 25 percent of capacity) is directly lost by the inefficiencies of tools (such as breakdowns), and that a second part of capacity is lost because of variability: In order to keep the Xfactor below its target value, the capacity utilization must be kept under a given value.

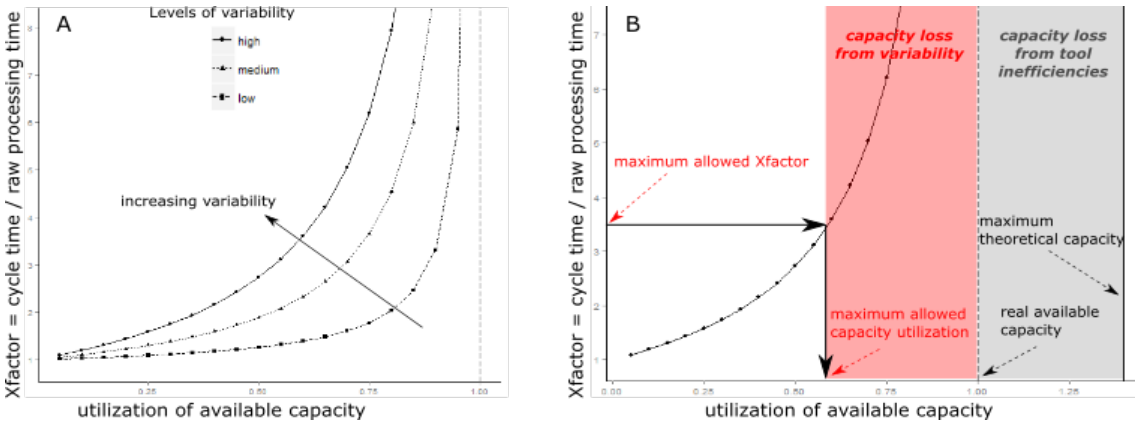

Figure 1: Operating curves showing different levels of variability (1.A) and the link to capacity loss (1.B).

Moreover, for a given maximum allowed Xfactor value, the maximum allowed capacity utilization decreases as the variability increases, which directly translates into capacity loss. This link is essential to understand the cost of variability in semiconductor fabs. As semiconductor manufacturing tools are

extremely expensive, a fab with high variability requires far more capital investment than one with low variability in order to achieve the same throughput under the same Xfactor constraint.

Another impact of variability is the difficulty of efficiently forecasting the capacities required and the cycle times of the products. Indeed, as the lack of information characterizing the system and its variability is high, changes in the production environment lead to unaccounted changes in variability and cycle times, making it ever so difficult to respect customer orders in terms of times and quantities.

2.3 Sources Of Variability

Table 1: Sources of variability identified by authors related to the variability of semiconductor manufacturing. The sources of variability are ordered by the number of times identified in the literature.

Sources of variability (equipment specific / product induced / operational / structural) / References	lack of tool redundancy	process time variability	downtime	product mix	batching	setup	rework / yield / scrap	reentrant flow	operator availability	tool dedication	dispatching policy	priorities	hold	WIP control strategy	heterogeneous toolsets	maintenance strategy	lot size	sequence specific processes	order release	time constraints between steps	operator cross-training	inspection	End of shift effects	factory shutdown	secondary resources
Martin (1999)			●	●	●				●																
Rose et al. (2000)	●	●	●																						
Huang et al. (2001)	●			●	●																				
Robinson et al. (2003)	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●		●		●	●	●	●	●	●	
Karabuk (2003)							●																		
Wu (2005)	●	●	●	●	●	●	●	●	●		●	●													
Delp et al. (2006)	●	●	●		●				●		●			●		●					●				
Morrison and Martin (2006)	●	●	●						●				●												
Hanschke (2006)	●	●			●												●								
Morrison and Martin (2007)	●	●	●		●	●			●	●			●					●							●
Shantikumar et al. (2007)	●	●	●	●	●	●	●	●		●	●	●	●		●			●	●						
Hopp et al. (2007)	●	●	●	●	●	●	●		●		●				●						●		●		
Zisgen et al. (2008)	●	●	●	●		●	●	●		●					●		●	●							
Wu and Hui (2008)	●	●	●	●					●							●				●					
Chang et al. (2008)	●	●		●								●													
Ignizio (2009a)	●		●	●				●		●	●	●		●	●	●									●
Brown et al. (2010)	●	●	●	●	●	●				●		●			●	●	●	●							
Etman et al. (2011)	●	●	●	●	●	●			●		●			●				●							
Ignizio (2011)	●	●	●	●			●	●		●															
Leachman (2012)	●	●	●		●	●	●				●	●	●									●			
Tirkel (2013)	●	●	●													●	●								
Kalir (2013)	●	●	●													●									
Godinho Filho and Uzsoy (2013)	●	●	●			●											●								
Schelasin (2013)	●	●		●				●																	
Rowshannahad et al. (2014)	●		●		●	●	●	●	●	●									●						
Kim et al. (2014)	●	●	●				●																		
Senderovich et al. (2015)	●											●													

The sources of variability of a manufacturing system can be defined as the primary factors that create, amplify, and propagate the local uncontrolled over-saturations of the system. We emphasize the fact that, given this definition, sources of variability do not *need* to be variable themselves. As there may be an unlimited number of sources of variability, our aim is to identify the main ones. To do so, we examined the recent literature related to the variability of semiconductor manufacturing systems. As authors do not necessarily use the same vocabulary and factors with minimal differences may exist, we grouped factors by what seemed to be the most relevant and distinctive names.

Moreover, since, as far as we know, there is no clear definition of the variability of a manufacturing system, we considered that authors identified a factor as a source of variability if they either mention it directly as a source of variability, mention that this factor increases the cycle time, mention that this factor increases the complexity of the system, mention it as a capacity loss factor, or include it in a formula in a way that this factor increases cycle time. Table 1 gives an overview of the sources of variability identified in our review.

3 LITERATURE REVIEW ON THE SOURCES OF VARIABILITY

3.1 Equipment-specific Factors

Natural process time, also referred to as raw processing time or theoretical processing time is, in a simplified manner, the time spent processing each lot on the machine. In the semiconductor environment, the definition becomes much more subtle (see Wu and Hui 2008 for more details). Natural process time variability is generally found to be a small source of variability. Both Kalir (2013) and Morrison and Martin (2007) computed small values of the coefficient of variation of natural process time and Tirkel (2013) also pointed that “the service time variability is not necessarily high since the processing time is usually automated and its variability is generated due to causes such as wafer lot size.”

Machine downtimes, if not cited the most, is unarguably the source of variability the most discussed by authors. Authors who speak about variability reduction always recommend doing so through reducing variability from downtimes: Increasing the frequency of preventive maintenances and therefore lowering the mean downtime is pointed by Tirkel (2013) and Delp et al. (2006) as a way to reduce variability. Brown et al. (2010) also discussed “how reduced variability in downtime durations […] could translate into shorter queues.” Godinho Filho, and Uzsoy (2013) show by means of simulation that many small improvements in downtime variability has a greater impact on cycle time reduction than few major improvements.

Batching is another equipment-specific characteristic. Most authors recognize the cycle time increase due to batching, and specific queuing equations for batch processes are actually proposed by Huang et al. (2001), Hanschke (2006), Brown et al. (2010), and Leachman (2012). Batches add variability to their toolset, but may add even more variability to other toolsets: As batch tools send batches of lots downstream, they greatly contribute to the unevenness of the downstream flows.

Setup, defined by Leachman (2012) as “[the time needed to] recondition, readjust or reconnect the machine when changing from performing one class of product steps to another,” is the last of our equipment-specific factors. Even though setup times may sometimes be as long as the processing times (in ion implantation for example), little research has focused on the time caused by setups. One reason may be because the consequences (additional waiting time, capacity loss and added variability) also depend on the setup *rules* and are thus fab specific. Shanthikumar et al. (2007) give extra references for approximations of cycle time under different setup rules.

3.2 Structural Factors

"Structural factors" refer to how resources are arranged together to process the flow of products. The way resources interact with each other through the flow greatly impacts the corrupting tendency of the manufacturing system to generate uncontrolled unevenness in the flows and in its capacities.

Tool redundancy is the simplest of examples, as it is found in many manufacturing systems and is measurable: the number of parallel tools that process a given process step. Increasing the redundancy smoothens the capacity of the toolset as tools are generally independent and the breakdowns happen more evenly. The redundancy therefore reduces the probability of high saturations and therefore decreases the variability of the system.

Tool dedication, defined by Shanthikumar et al. (2007) as "the specific relationship where certain tools in a toolset process only part of products or operation steps," disrupts the smoothing introduced by tool redundancy. In the same way, if characteristics or the performances of the tools in the toolset are uneven (heterogeneous toolset), the global performance of the toolset (i.e the global capacity) might vary greatly according to which tool in the toolset performs which process. Although such factors as tool dedication, heterogeneous toolsets or operator cross-training have been identified as sources of variability, little study has yet focused on their effects. Figure 2 illustrates the effects of poor redundancy and high tool dedication toolsets in HMLV fabs: It shows the lots processed by 3 tools of the same toolset, and more importantly which tools processed which products over the entire period. As one can see on Figure 2, some tools are qualified on many products (TOOL1 processes all products) whereas others are qualified on a limited quantity of products (TOOL3 only processes products *a*, *b*, and *c*). Some tools also see their qualification or dedication changing over time (TOOL2 only treats product *i* and *j* in the last third of the horizon).

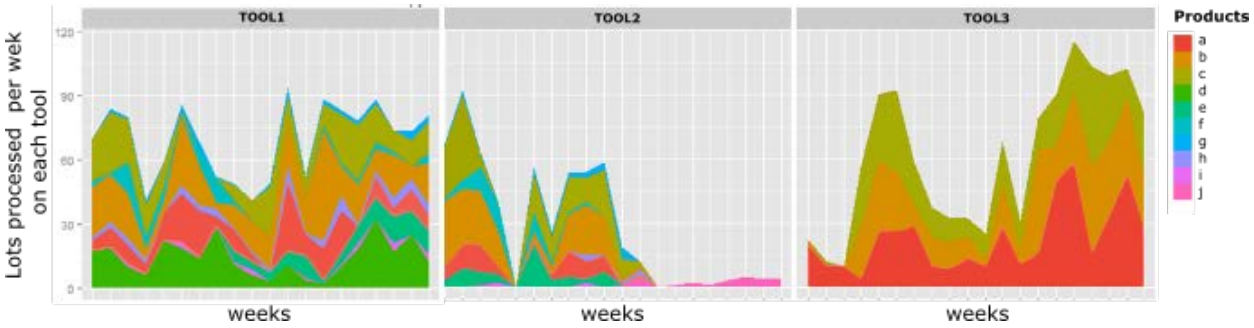

Figure 2: Product mix processed by different tools of a toolset over time.

Reentrancy is another, more global, structural factor. As resources interact locally through their redundancy, dedication and heterogeneity, they interact globally in the flow of products. As a part of the manufacturing system's variability can spread through the production line (through the unevenness of the flows), the reentrancy allows this spread to be much more extensive than in a linear production line: The effects of a single event can be seen by products completing their manufacturing process as well as by products just starting it. Moreover, in a reentrant system, so called "WIP bubbles" (Work In Progress) that originated independently can arrive at the same toolset at the same time, such as two waves joining to create a local splash. Hence reentrancy can create bottlenecks (and therefore queues) much larger than would be found in traditional linear lines. Reentrancy also forces machines to process different recipes (different product levels) creating high inefficiencies in the batches since the probability of forming a batch of identical recipes decreases. The same also goes for the setups, whose inefficiencies increase the more recipes are to be processed on the same tools, since this leads to more "switches" between recipes. Little study has been done on the added variability caused by the reentrancy of a system, even though Ignizio (2009b) introduced the notions of degree of reentrancy and nesting when considering reentrancy.

3.3 Product Induced Factors

The impact of product mix can be understood straightforwardly now that we have discussed batches, setups and reentrancy. More products mean more recipes on identical tools. Just as reentrancy adds more product *levels* to toolsets, higher product mix adds more *product* levels to toolsets. Therefore, an increased product mix results directly in reduced efficiency of batches and setups. Ignizio (2009a) performed simulations with low product mix and high product mix, and reports an increase in average total cycle time of 10 to 16 percent. Huang et al. (2001) included the number of recipes in their queuing approximations for batch servers and showed a significant increase in the queue length from 2 to 3 recipes. As seen in Figure 3, which represents the evolution of product mix over 20 weeks, a typical product mix in High Mix Low Volume fabs is high and variable.

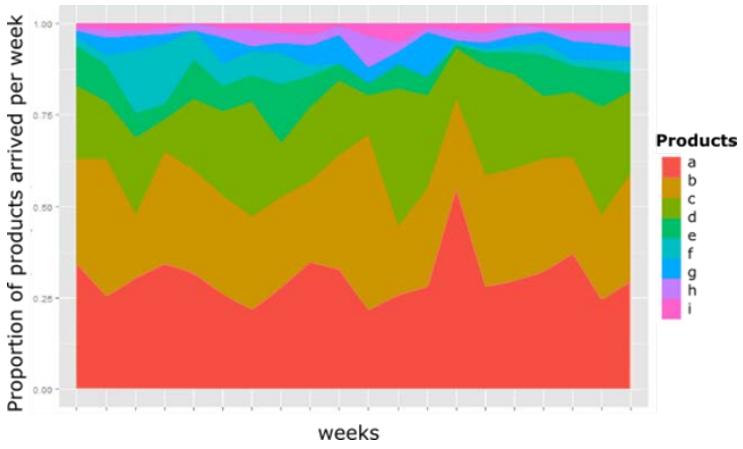

Figure 3: Evolution of product mix on a toolset of Crolles 300.

Priorities, induced by engineering lots, product prototypes lots or key customer orders, force inefficiencies in batches or setups by processing individual “hot lots” when batches or sequences could have been formed. Ignizio (2009a) shows with his simulation results that 5 percent of priority lots can increase the average cycle time of lots by approximately 15 percent. Chang et al. (2008) also show the impact of different priority mix on cycle times.

Product mix and priorities therefore both have a great impact on the cycle times observed at the different toolsets. As High Mix fabs operate in a *make-to-order* policy, the product mix is thus a very important source of variability as it is continually changing to follow the fluctuations of the demand.

3.4 Operational Factors

The other factors identified in the literature can then be classified as “operational”. These factors (which include dispatching policies, WIP control strategies, maintenance policy, end-of-shift effects, holds, factory shutdowns and inspection) describe the way the manufacturing system is operated. Hence there is no *number* or *parameter* that can translate their effects which makes them extremely hard to integrate. It is however clear that they have a significant impact: A dispatching policy based on an optimization algorithm can effectively impact cycle time in setup intensive areas. A maintenance policy that includes arrival forecasts can effectively prevent the combination of negative effects. Holding lots disrupts the flow. Factory shutdowns force lots to gather in “safe points” and therefore disrupt the flow and increase variability.

WIP control strategies are especially important. As some authors have reported, tool downtimes and product arrivals may not be independent because of specific WIP control strategies where production teams push back non-critical maintenance operations in case of high arrival rates.

4 THE CHALLENGES OF INTEGRATING VARIABILITY

4.1 Accounting For Variability

Our goal is to generate a taxonomy of the different toolsets versus their behavior from the cycle time point of view, using simulations to generate dynamic cycle time functions. The first step is then to fit the simulated toolsets to the reality. This means that the sources of variability need to be accounted for using parameters that adequately describe how these sources of variability generate queuing times in reality: This is where lies the real challenge.

Some sources of variability (such as lack of tool redundancy) can be added to the simulation as just an extra physical element. Other sources correspond to time related events (arrivals, breakdowns, processes). They require to be described by a set of parameters to be incorporated into simulations. The statistics generally used to describe these parameters (mean and standard deviation) should encompass the information on how much queuing time is added by the corresponding sources of variability. However, we emphasize the fact that the mean and standard deviation can only include the entire information of a given dataset under specific conditions: A known statistical distribution and the independency of the points in the dataset.

Figure 4 shows that this second assumption is not always true. Figure 4.A shows the number of arrivals per week at a toolset from STMicroelectronics Crolles300 fab, and Figure 4.B shows arrivals per week generated using the same inter-arrival rate as in Figure 4.A, but assuming independence of arrivals. Figure 4 shows that the order in which the inter-arrivals happen *also* contains information, and in this specific case, generates variability. Indeed, queues are more likely (in a non-linear way) to happen when saturation is higher. As Figure 4.A shows more high saturations than Figure 4.B, more queues will be created in the scenario of Figure 4.A than in the one of Figure 4.B. The mean and standard deviation may therefore be insufficient to measure the unevenness of the flows in this specific case.

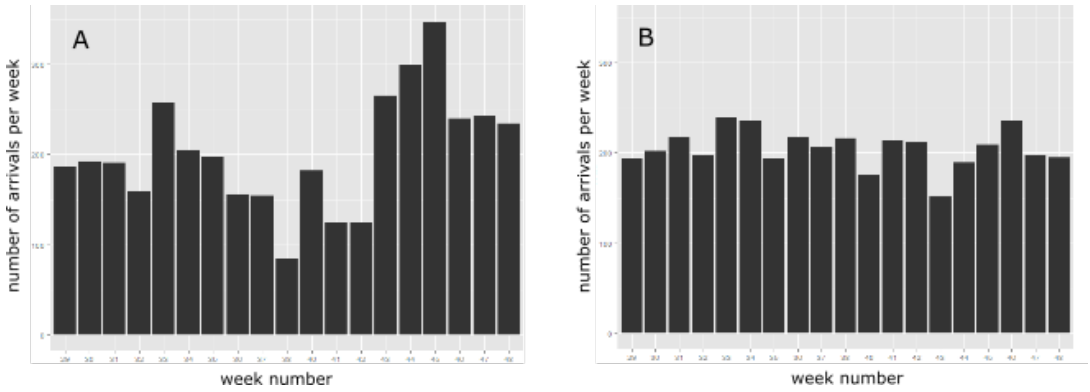

Figure 4: Number of arrivals per week from real data (A) and generated by removing dependencies (B).

Another dependency problem concerns natural process time variability. This next example was highlighted thanks to close collaboration with production teams, who pointed out that the process time variability was a major source of queuing time on a specific toolset of the Crolles300 fab. Contrary to the production team's intuition, our measurement of the coefficient of variation of processing times (the ratio between the standard deviation and the mean) indicated a low contribution to variability.

Figure 5 shows the histogram of processing times from this specific toolset. At first glance, process time variability seems low, with a standard deviation of 0.37h for a mean process time of 0.94h, resulting in a coefficient of variation of 0.39. However, several modes are visible on the histogram, corresponding to the number of available chambers when processing lots: When a chamber is down, lots are still processed, but by one less chamber, resulting in longer process times. The consequence is actually that

the processing times resulting follow a time-dependent (autocorrelated) sequence: Ten lots in a row might be processed with a degraded processing time. This temporarily reduces the toolset capacity, creating more over-saturations. This extra variability created by the dependency of the processing times cannot be captured by the standard deviation either: If we were to simulate this toolset with processing times described by a mean and a standard deviation, we would fail to incorporate this specific root cause of processing time variability.

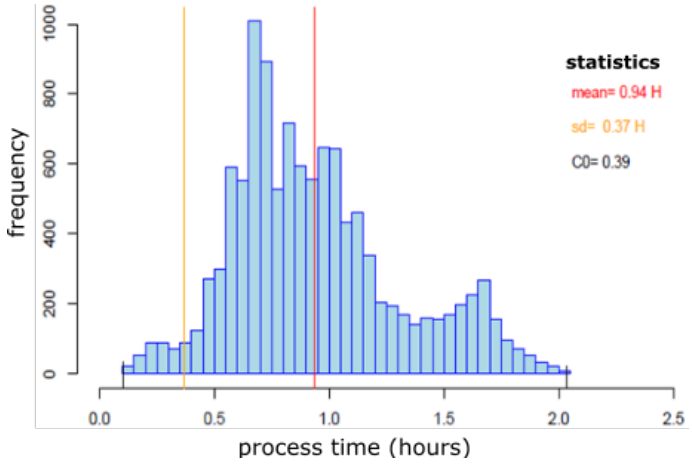

Figure 5: A histogram showing the effects of chambers on lot processing times.

Our two examples do not ignore the usefulness of the current statistics used to describe most parameters representing sources of variability. However, they point out that, in HMLV fabs, sources of variability can be far more complex than one might think.

4.2 Perspective: Variability Diagnostic Tree

As discussed above, identifying the sources of variability may not be enough: The root causes of the specific behaviors responsible for extra queuing times need to be understood in order to be sure that the statistics used to successfully model the sources incorporate these aspects. Therefore, in order to create valid simulation models, and in turn create dynamic cycle time models that fit the “to be” reality, it is essential to centralize the knowledge of experienced people working on the considered toolsets.

Following the work of Hopp et al. (2007), we intend to develop a Diagnostic Tree of the sources of variability. On top of referencing, for each toolset, the identified sources of variability, one objective of this Diagnostic Tree is to keep track of the root causes of each source as well as the statistics chosen to model these root causes after validation by means of simulation. This Diagnostic Tree will first help us to incrementally build our simulation model and therefore our dynamic cycle time model (which will in turn be the entry point of our projection engine). This Variability Diagnostic Tree will then help our Industrial Engineering team to find levers to control cycle times on the areas where high cycle time loss is forecast.

5 CONCLUSION

The proper use of real-time information is expected to lead the next industrial revolution. As high levels of variability lead to high levels of hidden information, it is essential to understand what variability is in order to both reduce it and integrate it in manufacturing management tools. To this extend, we first explained what the intrinsic variability of a manufacturing system is, and what its consequences are in terms of cycle time, capacity loss and system tractability. We then reviewed the sources of variability found in the literature (which, to our knowledge, had not been done prior to this article), giving for each

an explanation of its consequences as seen by the authors of our review (as well as personally experienced in a HMLV semiconductor fab). Using shop floor data, we finally pinpointed the complexity of some behaviors responsible for variability and the lack of usual statistical methods to capture some critical aspects of variability. In highly changing environments such as HMLV fabs, where projections play a major role for managing the line, this fundamental understanding is a milestone to create accurate production management tools.

Even though the component market is dominated by a few major players, thousands of production facilities exist through the world. And these factories will continue to produce components in the future due to the pervasion of electronics in every aspect of our lives, the “digitization” of everything. Besides traditional trends towards ever more computation power on ever smaller components (as reflected by the famous Moore’s law), the integration of heterogeneous functions on top of traditional computing or signal processing should allow smaller production units to develop their activities. As these companies will face the same tractability problems HMLV fabs face now, we call for more focus on the fundamental understanding of the manufacturing systems and a greater use of industrial data in research in order to find today the solutions for the “smart industries” of tomorrow.

ACKNOWLEDGMENTS

This work benefited from a funding of the French National Agency of Technical Research (ANRT) and of the Rhône-Alpes region.

REFERENCES

Brown, S. M., T. Hanschke, I. Meents, B. R. Wheeler, and H. Zisgen. 2010. "Queueing Model Improves IBM's Semiconductor Capacity and Lead-Time Management." *Interfaces* 40:397-407.

Chang, S. C., S. S. Su, and K. J. Chen. 2008. "Priority Mix Planning for Cycle Time-Differentiated Semiconductor Manufacturing Services." In *Proceedings of the 2008 Winter Simulation Conference*, edited by S. J. Mason, R. R. Hill, L. Mönch, O. Rose, T. Jefferson, and J. W. Fowler, 2251-2259. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Delp, D., J. Si, and J. W. Fowler. 2006. "The Development of the Complete X-Factor Contribution Measurement for Improving Cycle Time and Cycle Time Variability." *IEEE Transactions on Semiconductor Manufacturing* 19:352-362.

Etman, L. F., C. P. Veeger, E. Lefeber, I. J. Adan, and J. E. Rooda. 2011. "Aggregate Modeling of Semiconductor Equipment Using Effective Process Times." In *Proceedings of the 2011 Winter Simulation Conference*, edited by S. Jain, R. R. Ceasey, J. Himmelspach, K. P. White, and M. Fu, 1795-1807. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Godinho Filho, M., and R. Uzsoy. 2013. "The Impact of Simultaneous Continuous Improvement in Setup Time and Repair Time on Manufacturing Cycle Times Under Uncertain Conditions." *International Journal of Production Research* 51:447-464.

Hanschke, T. 2006. "Approximations for the Mean Queue Length of the Queue." *Operations Research Letters* 34:205-213.

Hopp, W. J., and M. L. Spearman. 2001. *Factory Physics: Foundations of Manufacturing Management*. Irwin/McGraw-Hill.

Hopp, W. J., S. M. Iravani, and B. Shou. 2007. "A Diagnostic Tree for Improving Production Line Performance." *Production and Operations Management* 16:77-92.

Huang, M. G., P. L. Chang, and Y. C. Chou. 2001. "Analytic Approximations for Multiserver." *IEEE Transactions on Semiconductor Manufacturing* 14:395-405.

Ignizio, J. P. 2009a. "Cycle Time Reduction Via Machine-To-Operation Qualification." *International Journal of Production Research* 47:6899-6906.

Ignizio, J. P. 2009b. *Optimizing factory performance*. New York: McGraw-Hill.

Ignizio, J. P. 2011. "Estimating the Sustainable Capacity of Semiconductor Fab Workstations." *International Journal of Production Research* 49:5121-5131.

Kacar, N. B., D. F. Irdem, and R. Uzsoy. 2012. "An Experimental Comparison of Production Planning Using Clearing Functions and Iterative Linear Programming-Simulation Algorithms." *IEEE Transactions on Semiconductor Manufacturing* 25:104-117.

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

Karabuk, S. 2003. "Coordinating Strategic Capacity Planning in the Semiconductor Industry." *Operations Research* 51:839-849.

Kim, D. J., L. Wang, and R. Havey. 2014. "Measuring Cycle Time Through the Use of the Queuing Theory Formula (G/G/M)." In *Proceedings of the 2014 Winter Simulation Conference*, edited by A. Tolk, S. Y. Diallo, I. O. Ryzhov, L. Yilmaz, S. Buckley, and J. A. Miller, 2414-2421. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Leachman, R. C. 2012. "The Engineering Management of Speed." In *Proceedings of the 2012 Industry Studies Association Annual Conference*.

Martin, D. P. 1999. "Capacity and Cycle Time-Throughput Understanding System (CAC-TUS) An Analysis Tool To Determine the Components of Capacity and Cycle Time in a Semiconductor Manufacturing Line." *Advanced Semiconductor Manufacturing Conference and Workshop*, 127-131.

Mhiri, E., M. Jacomino, F. Mangione, P. Vialletelle, and G. Lepelletier. 2014. "A Step Toward Capacity Planning at Finite Capacity in Semiconductor Manufacturing." In *Proceedings of the 2014 Winter Simulation Conference*, edited by A. Tolk, S. Y. Diallo, I. O. Ryzhov, L. Yilmaz, S. Buckley, and J. A. Miller, 2239-2250. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Mönch, L., J. Fowler, S. Dauzère-Pérès, S. J. Mason, and O. Rose. 2011. "A Survey of Problems, Solution Techniques, and Future Challenges." *Journal of Scheduling* 14:583-599.

Morrison, J. R., and D. P. Martin. 2006. "Cycle Time Approximations for the G/G/M Queue Subject To Server Failures and Cycle Time Offsets with Applications." In *The 17th Annual SEMI/IEEE Advanced Semiconductor Manufacturing Conference*, 322-326.

Morrison, J. R., and D. P. Martin. 2007. "Practical Extensions to Cycle Time Approximations for the G/G/m-Queue with Applications." *IEEE Transactions on Automation Science and Engineering* 4:523-532.

Robinson, J., J. Fowler, and E. Neacy. 2003. "Capacity Loss Factors in Semiconductor Manufacturing." Working paper. http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.582.1421 [Accessed July 12, 2016].

Rose, O., M. Dümmler, and A. Schoemig. 2000. "On the Validity of Approximation Formulae for Machine Downtimes." Research Report Series, Institute of Computer Science at the University of Würzburg, Germany, Report no. 250.

Rowshannahad, M., S. Dauzère-Pérès, and B. Cassini. 2014. "Qualification Management to Reduce Workload Variability in Semiconductor Manufacturing." In *Proceedings of the 2014 Winter Simulation Conference*, edited by A. Tolk, S. Y. Diallo, I. O. Ryzhov, L. Yilmaz, S. Buckley, and J. A. Miller, 2434-2443. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Schelasin, R. E. 2013. "Estimating Wafer Processing Cycle Time Using An Improved G/G/M Queue." In *Proceedings of the 2013 Winter Simulation Conference*, edited by R. Pasupathy, S.-H. Kim, A. Tolk, R. Hill, and M. E. Kuhl, 3789-3795. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Senderovich, A., M. Weidlich, A. Gal, and A. Mandelbaum. 2015. "Queue Mining for Delay Prediction In Multi-Class Service Processes." *Information Systems* 53:278-295.

Shanthikumar, J. G., S. Ding, and M. T. Zhang. 2007. "Queueing Theory for Semiconductor Manufacturing Systems: A Survey and Open Problems." *IEEE Transactions on Automation Science and Engineering* 4:513-522.

Tirkel, I. 2013. "The Effectiveness of Variability Reduction in Decreasing Wafer Fabrication Cycle Time." In *Proceedings of the 2013 Winter Simulation Conference*, edited by R. Pasupathy, S.-H. Kim, A. Tolk, R. Hill, and M. E. Kuhl, 3796-3805. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Whitt, W. 1993. "Approximations for the GI/G/M Queue." *Production and Operations Management* 2:114-161.

Wu, K. 2005. "An Examination of Variability and Its Basic Properties for a Factory." *IEEE Transactions on Semiconductor Manufacturing* 18:214-221.

Wu, K., and K. Hui. 2008. "The Determination and Indetermination of Service Times in Manufacturing Systems." *IEEE Transactions on Semiconductor Manufacturing* 21:72-82.

Zisgen, H., I. Meents, B. R. Wheeler, and T. Hanschke. 2008. "A Queueing Network Based System To Model Capacity and Cycle Time for Semiconductor Fabrication." In *Proceedings of the 2008 Winter Simulation Conference*, edited by S. J. Mason, R. R. Hill, L. Mönch, O. Rose, T. Jefferson, J. W. Fowler, 2067-2074. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

AUTHOR BIOGRAPHIES

KEAN DEQUEANT is an Industrial Engineer and Ph.D. student. He received an Industrial Engineering Degree specialized in Supply Chain Management from the Grenoble Institute of Technology (Grenoble INP). He currently works at STMicroelectronics Crolles, France, on the modeling of variability and products flow as part of an industrial Ph.D. with the Grenoble laboratory for Science, Conception, Optimization and Production (G-SCOP). His email address is kean.dequeant@grenoble-inp.fr.

PIERRE LEMAIRE is an assistant professor in the School of Industrial Engineering, Grenoble Institut of Technology, Grenoble-Alpes University, France. He received a M.S. degree in applied mathematics and computer science and a Ph.D. degree in computer science from the same university. His research focuses on optimization, algorithm design and analysis, and data mining, with applications to production and transportation systems, and medical diagnosis. His email address is pierre.lemaire@grenoble-inp.fr.

MARIE-LAURE ESPINOUSE is a full professor in the Department of electrical engineering and industrial computing at the Institute of Technology of the Grenoble-Alpes university, France. She received a Ph.D. degree and an habilitation in computer science from the university Joseph Fourier, Grenoble, France. She has a strong background in optimization and operations research. Currently, one of her research field focuses on taking into account uncertainty in optimization and one aspect is the analysis of variability in semiconductor facilities. Her email address is Marie-Laure.Espinouse@g-scop.grenoble-inp.fr.

PHILIPPE VIALLETELLE is Senior Member of Technical Staff at STMicroelectronics Crolles, France. After receiving an Engineering degree in Physics from the Institut National des Sciences Appliquees in 1989, he entered the semiconductor industry by working on ESD and physical characterization. His next experience was in Metrology before enlarging his scope of activities to Process Control. He finally integrated the Factory Integration world, through Industrial Engineering and is now responsible for the definition of advanced methodologies and tools for the Crolles 300mm production line. Expert in Manufacturing Sciences for Process and Production Control items, he is now in charge of driving collaborative projects at European level. His email address is philippe.vialletelle@st.com.