Proceedings of the 2024 Winter Simulation Conference
H. Lam, E. Azar, D. Batur, S. Gao, W. Xie, S. R. Hunter, and M. D. Rossetti, eds.

ANALYZING THE TRADE-OFF BETWEEN QUALITY AND SOJOURN TIME WHEN OPTIMIZING SAMPLING PLANS IN SEMICONDUCTOR MANUFACTURING

Stéphane Dauzère-Pérès[1] and Michael Hassoun[2]

[1]Mines Saint-Étienne, Univ. Clermont Auvergne, CNRS, UMR 6158 LIMOS, Gardanne, FRANCE
[2]Department of Industrial Engineering and Management, Ariel University, Ariel, Israel

ABSTRACT

Inspired by semiconductor manufacturing, this paper studies a system where the products processed on multiple production machines are sampled to be measured on a single metrology tool. In a previous research, we show that minimizing the expected number of defective products is not ensured by using the metrology tool at its maximum capacity, as it induces a congestion that impacts the expected product loss. However, the congestion of the metrology tool also impacts the expected sojourn time of products in metrology. Hence, in this paper, we analyze the trade-off between the expected product loss and the expected sojourn time when deciding how much of the metrology capacity should be used. Numerical results show that, depending on some parameters, the expected sojourn time can be reduced without increasing much the expected product loss.

1 INTRODUCTION

Let us not break away from the long established custom of beginning scientific papers on semiconductor manufacturing by stating that it is the most complex type of mass production in today's manufacturing. It is by all means the case. Among the countless accounts of this complexity, the design of efficient ways to control the production quality is the focus of numerous studies. Some emphasize the physical aspects of quality while others, like the current one, focus on operational aspects. Quality control along the process is conducted through a carefully designed set of metrology steps aimed either at monitoring parameters specific to the former production (also called process) operation, or at making sure that the level of particle contamination is kept at bay. At an operational level, these metrology steps share the particularity of not being mandatory to the product. Therefore, in order to keep products reasonably flowing, it is customary to define inspection sampling plans. The flexibility in inspection does not only pertain to the choice of which product is to be sent to inspection, but sometimes to the inspection recipe itself (wafers sampled, sites inspected on these wafers, control limits, etc.). The level of control required has therefore a direct impact on the Sojourn time (also known as Flow Time or Cycle Time in industrial contexts) of the products. Different aspects of the trade-off between quality and sojourn time have been the subject of numerous studies (Colledani and Tolio (2011); Bettayeb et al. (2012); Gilenson et al. (2015); Nduhura-Munga et al. (2012); Nduhura-Munga et al. (2013); Lee et al. (2003); Dauzère-Pérès et al. (2010); Rodriguez-Verjan et al. (2013); Shanoun et al. (2011); Tirkel et al. (2009); Tirkel and Rabinowitz (2012)). In a former publication Dauzère-Pérès and Hassoun (2020), we optimize the sampling rates of a set of production machines sending their products for inspection on a single metrology tool. The sampling rate of a production machine is defined through a sampling period, which is the number of products between two products sampled to be measured, and the sampling plan is the set of sampling periods for all production machines. We show in Dauzère-Pérès and Hassoun (2020) how maximizing the use of inspection capacity has the adverse effect of congesting the metrology tool, thus delaying inspection and allowing for more defective products to be produced. Moreover, the level of planned metrology utilization, associated to the sampling plan of the machines, can be efficiently determined depending on the level of variability in the production machines and the inspection tool. In Dauzère-Pérès and Hassoun (2020), the delay in the metrology queue of products sent to be inspected is considered solely through its impact on the expected flow of defective products over time (expected product loss). The sampling periods are therefore optimized to directly minimize this flow, with a resulting expected sojourn time of products in metrology.

In the present work, we broaden the scope of the problem by considering the trade-off between the expected product loss and the expected sojourn time of products in metrology. This allows practitioners to chose an inspection policy that might be sub-optimal in terms of quality but allows for the metrology sojourn time to be more acceptable. Let us consider a single metrology tool inspecting the products processed by several production machines, each following a specific sampling period, the number of products between two products sampled to be measured. First, the optimal sampling plan (i.e., the optimal set of sampling periods for all production machines) is determined by minimizing the expected product loss following the methodology presented in Dauzère-Pérès and Hassoun (2020)). Starting from this optimal sampling plan, the metrology utilization is gradually reduced, inducing a reduction of the expected metrology sojourn time at the expense of an increase of the expected product loss.

In Section 2, we recall the expression of the expected product loss (identical to the one derived in Dauzère-Pérès and Hassoun (2020) and derive the expected sojourn time of products in metrology. The associated optimization problem is also recalled in this section. Then, the trade-off between product loss and sojourn time is analyzed in Section 3 using illustrative instances. More extensive computational experiments are conducted in Section 4 to discuss the impact of different instance parameters on this trade-off. Section 5 concludes the paper by providing some perspectives.

2 PROBLEM MODELING AND RESOLUTION

After describing the problem in detail and providing the notations, the expected product loss introduced in Dauzère-Pérès and Hassoun (2020) is recalled in Section 2.1, and the expected sojourn time is formalized. Then, in Section 2.2, the optimization problem solved in Dauzère-Pérès and Hassoun (2020) is recalled, before presenting how the trade-off analysis was conducted.

2.1 Mathematical modeling

A single metrology tool is assigned to the quality control of a group of production machines subject to failures. The production rates are constant and machine-specific. A production cycle corresponds to the time spent by the machine to produce one item. At each production cycle, the machine has a specific (and stationary) probability to fail, following which all its production is defective and either reworked or scrapped, at a fixed cost. Being heterogeneous, production machines are controlled following specific sampling policies, characterized by a sampling period, i.e. the number of products between two consecutive inspections. The choice of a sampling plan, i,e. a set of sampling periods for all the production machines, results in a certain rate of product scrapping. The sampling plan also determines the level of congestion on the metrology tool, which in turns impacts the waiting time of the products at the metrology tool. Sending products to inspection more frequently is the natural way of shortening periods in which machines are out of control and produce defective products. In a former publication Dauzère-Pérès and Hassoun (2020), we show how requesting inspections too frequently is actually detrimental to the production quality due to the additional production occurring during the delayed response of the inspection. Sending products to inspection more often to better control the throughput quality of a production machine has the adverse effect of congesting the metrology tool. This induces delays in the inspection results during which more defective products may be produced, not only by the production machine but also by all other production machines controlled by the same metrology tool. However, in Dauzère-Pérès and Hassoun (2020), we do not explore the trade-off between the quality criterion and the sojourn time of sampled products in inspection. This trade-off is the subject of the current study. Note that we do not consider in this paper the case where a production machine is kept idle until the associated sampled product has been measured.

The following notations are used in the paper:

- R: Number of production machines,
- TP_r: Throughput rate of production machine r,
- TM_r: Throughput rate of the metrology tool when inspecting products from r,
- p_r: Failure probability (Bernoulli experiment) of production machine r at each production cycle, given that the machine is working properly before the production cycle,
- SP_r: Sampling period of production machine r, i.e. the number of production cycles between two consecutive inspections,
- $\lambda_r = \frac{TP_r}{SP_r}$: Inspection rate of machine r, i.e. the rate at which products are sent to the metrology tool,
- SP^{max}: Upper limit for SP_r over which the risk on production quality is deemed unacceptable by quality managers,
- $\mathscr{S} = \{SP_r; r = 1, \cdots, R\}$: Set of sampling periods.

When a product inspection returns a positive result (the machine is defective), the production machine is repaired after which normal production and inspection cycles resume. We assume that no products are sent by the production machine to metrology during repair. These events are rare enough to be considered insignificant in the metrology utilization.

By deciding on the sampling periods SP_r, $\forall\ r = 1, \ldots, R$, managers determine both the throughput of bad products from production machine r, and the portion of the metrology tool capacity consumed by machine r which is denoted by $g_r(SP_r)$ where $g_r(SP_r) = \frac{\lambda_r}{TM_r}$.

Following the mathematical development in Dauzère-Pérès and Hassoun (2020), the overall expected rate of bad products, called Product Loss, is recalled here. For simplicity, the full mathematical development is not presented and can be found in the original paper.

We assume that a queue is forming at metrology tool. Consequently, the response on a product sent to inspection is delayed. The overall rate of products sent to inspection is:

$$\lambda(\mathscr{S}) \triangleq \sum_{r=1}^{R} \lambda_r$$

The steady state composition of the queue is determined by expected proportion of the production machine, namely $\lambda_r/\lambda(\mathscr{S})$ for machine r, following which the expected metrology tool is the harmonic mean of rates TM_r weighted by the ratio of production machines in the queue:

$$\mu(\mathscr{S}) \triangleq \frac{\sum_{r=1}^{R} \lambda_r}{\sum_{r=1}^{R} \frac{\lambda_r}{TM_r}} \tag{1}$$

The congestion at the metrology tool, or expected traffic intensity, is:

$$\rho(\mathscr{S}) \triangleq \frac{\lambda(\mathscr{S})}{\mu(\mathscr{S})} = \sum_{r=1}^{R} \frac{\lambda_r}{TM_r},$$

We characterize the variability in the group of production machines by the coefficients of variation of the service time, and of the inter-arrival time to metrology, denoted by c_s and c_a, respectively. Kingman's G/G/1 queue approximation (see Kingman (1962)) is then used to estimate the expected sojourn time of products in metrology:

$$W(\mathscr{S}) \approx \frac{1}{\mu(\mathscr{S})} \left[\left(\frac{\rho(\mathscr{S})}{1-\rho(\mathscr{S})} \right) \left(\frac{c_a^2 + c_s^2}{2} \right) + 1 \right] \tag{2}$$

Let us recall that the analysis of the trade-off between the expected sojourn time of products in metrology and the quality, i.e. the expected product loss, is the main contribution of this research.

The overall expected rate of defective products, i.e. the product loss, of machine r under an inspection policy using the set of sampling periods $\mathscr{S} = \{SP_1, \ldots, SP_R\}$, is given by:

$$PL_r(\mathscr{S}) = \frac{TP_r \cdot \left[p_r \cdot \sum_{i=0}^{SP_r-1} (SP_r - i)(1-p_r)^i + (1-(1-p_r)^{SP_r}) \cdot \lceil W(\mathscr{S}) \cdot TP_r \rceil \right]}{SP_r + (1-(1-p_r)^{SP_r}) \cdot \lceil W(\mathscr{S}) \cdot TP_r \rceil} \tag{3}$$

Note that (3) includes the additional products that are scrapped during the delay $W(\mathscr{S})$ in the metrology queue (see Dauzère-Pérès and Hassoun (2020) for a detailed development of (3)).

Adopting a set of sampling periods $\mathscr{S}$ results in a total expected product loss given by:

$$PL(\mathscr{S}) = \sum_{r=1}^{R} PL_r(\mathscr{S}) \tag{4}$$

2.2 Optimization problem

The product loss minimization problem, denoted (P), is therefore:

$$\begin{aligned}
&\min \; PL(\mathscr{S}) \\
&\text{s.t.} \\
&\sum_{r=1}^{R} g_r(SP_r) < 1 \\
&SP_r \in \{1, \ldots, SP^{max}\}, \quad r = 1, \ldots, R
\end{aligned}$$

An important characteristic of this formulation is that $PL(\mathscr{S})$ is not separable by production machine as $PL_r(\mathscr{S})$ depends on the set of sampling periods for all machines $\mathscr{S}$.

The approach adopted to solve (P) is detailed in Section 3 of Dauzère-Pérès and Hassoun (2020). A heuristic is presented that solves the problem efficiently, resulting in both the optimal metrology capacity utilization, and the optimal sampling plan, i.e. set of sampling periods.

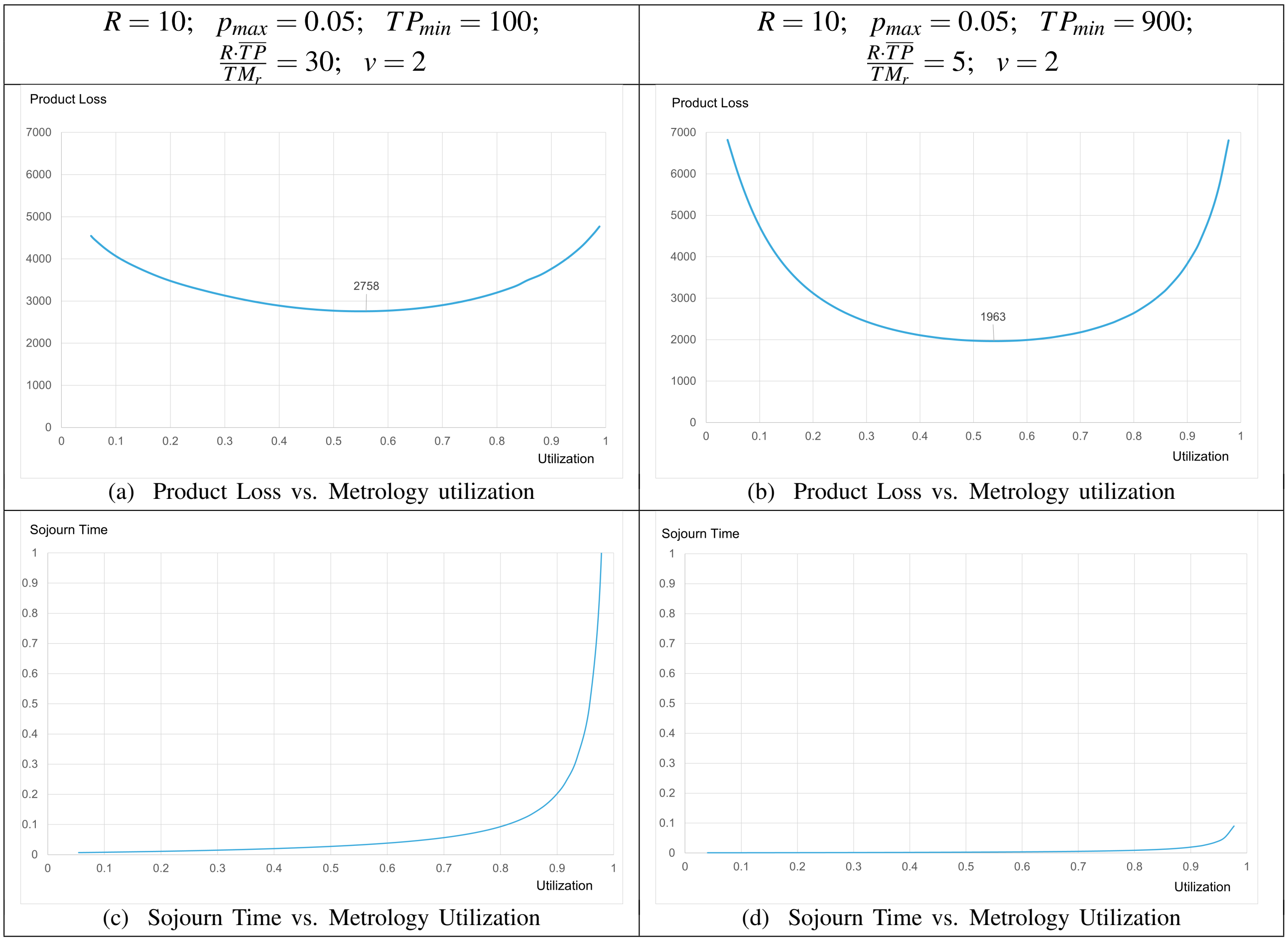
3 ANALYZING THE TRADE-OFF BETWEEN PRODUCT LOSS AND SOJOURN TIME

The solution structure is first recalled in Section 3.1. Then the trade-off between the product loss and the sojourn time, not explicitly considered as a criterion in Dauzère-Pérès and Hassoun (2020), is illustrated and discussed in Section 3.2. This trade-off is analyzed in more details on a large number of instances in Section 4.

3.1 Solution structure

Let us illustrate the qualitative relationship between Product Loss, Sojourn Time, and Metrology Utilization on two instances of the problem that differ in the relative speed of the metrology machine and the range of production rates. The characteristics of these two instances can be found in Section 4. Figures 1.a and 1.b show the impact of the metrology utilization on the expected product loss. Figures 1.c and 1.d show the increase of the expected sojourn time with the metrology utilization.

In accordance to any queuing model (assuming no balking), the sojourn time is increasing non-linearly with the metrology tool utilization, and goes to infinity when the utilization approaches 1 (Figures 1.c and 1.d). The product loss behavior in our problem is less obvious. At first, as the metrology tool utilization increases, i.e. inspection policies become more stringent (small values of SP_r), the product loss is reduced. However, sending more products to inspection also congests the metrology tool which, after a certain point (minima in Figures 1.a and 1.b), leads to a increase of the product waiting times in metrology that induces an increase in Product Loss. Dauzère-Pérès and Hassoun (2020) propose to determined the optimal point (that minimizes the product loss) by recursively reducing the metrology capacity (starting from a value close to 1) and solving problem (P) until $PL(\mathscr{S})$ reaches the inflection point and starts increasing. The minimal $PL(\mathscr{S})$ and the associated inspection plan $\mathscr{S}$ are recorded as the optimal solution.



(a) Product Loss vs. Metrology utilization

(b) Product Loss vs. Metrology utilization

(c) Sojourn Time vs. Metrology Utilization

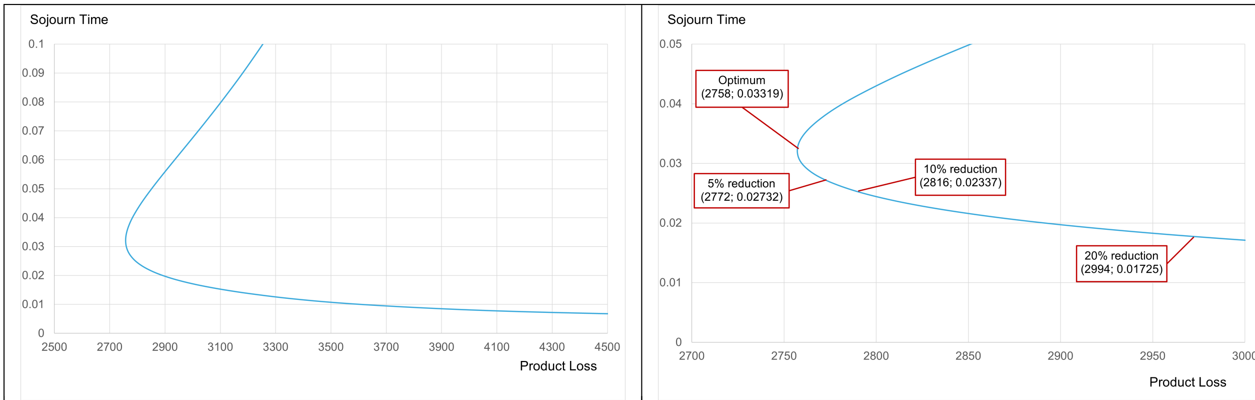
(d) Sojourn Time vs. Metrology Utilization

Figure 1: Product Loss vs. Metrology utilization and Sojourn Time vs. Metrology Utilization for two problem instances

3.2 Product loss vs. sojourn time trade-off

For the same two problem instances than in the previous section, Figures 2.a, 2.b, 3.a and 3.b illustrate the trade-off between expected sojourn time and expected product loss. In particular, Figures 2.b and 3.b zoom on the product loss and sojourn time obtained for different reductions (5%, 10%, 15% and 20%) from the optimal metrology capacity obtained by minimizing the product loss.

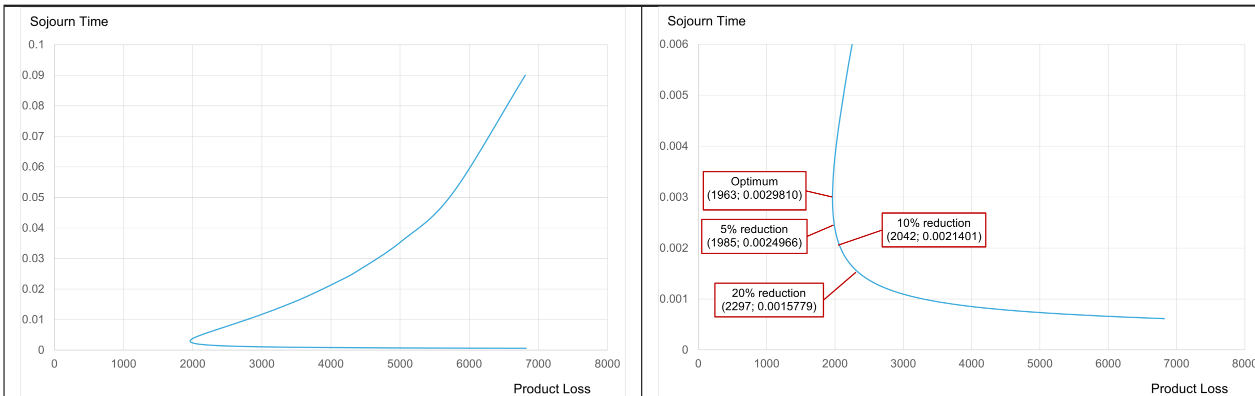
Figures 2.b and 3.b show that a significant reduction of the sojourn time can be obtained by reducing the optimal metrology capacity by up to 5% without significantly degrading the product loss in percentage. However, this is no longer the case when the optimal metrology capacity is reduced by more than 10%. The results do not differ so much when the case where the metrology machine is relatively slow ($\frac{R \cdot \overline{TP}}{TM_r} = 30$) is compared to the case where the metrology machine is relatively fast ($\frac{R \cdot \overline{TP}}{TM_r} = 5$).



(a) Sojourn Time vs. Product Loss

(b) Sojourn Time vs. Product Loss trade-off exploration

Figure 2: Sojourn Time vs. Product Loss for $R = 10$; $p_{max} = 0.05$; $TP_{min} = 100$; $\frac{R \cdot \overline{TP}}{TM_r} = 30$; $v = 2$



(a) Sojourn Time vs. Product Loss

(b) Sojourn Time vs. Product Loss trade-off exploration

Figure 3: Sojourn Time vs. Product Loss for $R = 10$; $p_{max} = 0.05$; $TP_{min} = 900$; $\frac{R \cdot \overline{TP}}{TM_r} = 5$; $v = 2$

4 COMPUTATIONAL EXPERIMENTS

4.1 Structure of the experiments

In order to demonstrate the idea of exploring the trade-off between metrology Sojourn Time and Product Loss, we conducted an experiment over a large number of scenarios, each characterized by the parameters described earlier. Most of these scenarios are similar to the ones used in Dauzère-Pérès and Hassoun (2020), and for the sake of clarity, we describe here how they were designed.

The number of production machines R is taken from the set $\{10, 20\}$. The failure probabilities p_r are generated from a uniform distribution $U[p_{min}, p_{max}]$, where p_{max} is chosen in the set $\{0.05, 0.2\}$ and p_{min} is kept constant ($p_{min} = 0.01$). The throughput rate TP_r is generated from a distribution $U[TP_{min}, TP_{max}]$, where $TP_{max} = 1,000$ and TP_{min} is chosen in the set $\{100, 900\}$. The measurement rate TM_r is determined using the ratio $\frac{R \cdot \overline{TP}}{TM_r}$ chosen from the set $\{5, 10, 30\}$, where $\overline{TP}$ is the average throughput rate for the considered scenario.

The variability factor is more complex to handle in the framework of this experimentation since it is affected by the choice of sampling periods, and by the problem parameters. Let us denote by v the variability factor in the Kingman's approximation (Kingman (1962)) $\left(v^2 = \frac{c_a^2+c_s^2}{2}\right)$. In Dauzère-Pérès and Hassoun (2020), we describe how the range of useful values for the variability factor is determined, and v is chosen from the set $\{0.6,\ 0.8,\ 1,\ 1.2,\ 1.4,\ 1.6,\ 2\}$. For each scenario, the optimal point was first determined, and its characteristics (metrology utilization, Product Loss, Sojourn Time) recorded. Then, the available metrology capacity was gradually reduced by steps of 5% up to 20%. For each of these reductions, the expected sojourn time and the expected product loss are compared to the optimal ones. Because Product Loss is impacted by the machine rates, presenting only the relative change may have been misleading. We therefore opted to also present the absolute variation of Product Loss.

Tables 4 and 5 present the averages of said variations for the four levels of capacity reduction, detailed by the metrology ratio $\frac{R \cdot \overline{TP}}{TM_r}$ and variability v. Both tables prompt several insights. First and foremost, across all cases, the impact on quality represented by the relative change in Product Loss seems limited when compared to the gain in Sojourn Time. For example, for 20 machines (Table 5), the decreases in CT are roughly an order of magnitude larger than the increases in product loss. By solving the metrology allocation problem for reduced metrology capacities of up to 20% we show that the time spent for inspection can be shortened by figures well above 40% with a relatively low impact on Product Loss (roughly between 5% and 12%). This alone justifies exploring the possibility of sacrificing a small portion of the production yield to gain a significant reduction of the time spent by lots in metrology.

Beyond this important point, the impact of both the stress on the metrology tool (metrology ratio) and of the variability seems overall quite limited. As an example, for 10 machines (Table 4) and a capacity reduction of 10%, the change in sojourn time ranges from -21.1% to -27.1% while the increase in product loss ranges from 1.3% to 3.3% over all values of the metrology ratio and the variability parameter.

Note that, while the present experiment shows the benefits of exploring the trade-off between quality and waiting times in metrology by allowing sampling policy to be set sub-optimally, in order for practitioners to actually choose a working point, considering the sojourn time of all lots, including those that are not sent to inspection, is necessary. This is beyond the scope of the present work.

		Capacity reduction - 5%			Capacity reduction - 10%			Capacity reduction - 15%			Capacity reduction - 20%		
Metrology Ratio	v	Relative change in Sojourn Time	Relative change in Product Loss	Absolute change in Product Loss	Relative change in Sojourn Time	Relative change in Product Loss	Absolute change in Product Loss	Relative change in Sojourn Time	Relative change in Product Loss	Absolute change in Product Loss	Relative change in Sojourn Time	Relative change in Product Loss	Absolute change in Product Loss
	0.6	-11.7%	0.9%	15.5	-21.1%	3.2%	52.4	-27.9%	6.5%	105.9	-33.1%	10.7%	174.5
	0.8	-12.2%	0.8%	15.3	-22.3%	2.9%	53.2	-30.0%	6.2%	110.9	-36.2%	10.5%	188.8
	1	-12.6%	0.8%	16.0	-23.1%	2.9%	55.6	-31.6%	6.2%	119.4	-38.2%	10.8%	205.7
5	1.2	-12.1%	0.8%	15.3	-23.3%	2.8%	56.6	-32.2%	6.3%	124.0	-39.5%	11.0%	215.8
	1.4	-12.9%	0.8%	16.6	-24.3%	2.9%	61.9	-33.5%	6.4%	134.8	-41.3%	11.4%	239.2
	1.6	-13.6%	0.8%	17.2	-25.1%	3.0%	63.9	-34.8%	6.6%	140.9	-42.8%	12.0%	252.5
	2	-13.8%	0.8%	18.9	-26.1%	3.3%	72.0	-36.5%	7.4%	161.7	-45.3%	13.5%	294.5
	0.6	-13.0%	0.8%	19.3	-23.0%	2.8%	67.8	-30.1%	5.7%	137.6	-35.4%	9.4%	225.5
	0.8	-13.5%	0.7%	18.8	-24.0%	2.5%	66.9	-31.9%	5.3%	138.3	-38.1%	8.9%	231.0
	1	-13.4%	0.7%	18.2	-24.3%	2.4%	67.3	-32.8%	5.2%	142.7	-39.6%	8.9%	243.5
10	1.2	-13.5%	0.7%	18.1	-24.7%	2.5%	67.1	-33.7%	5.3%	144.2	-40.9%	9.2%	248.1
	1.4	-13.6%	0.6%	18.3	-25.2%	2.4%	69.4	-34.6%	5.3%	150.9	-42.3%	9.3%	262.8
	1.6	-13.9%	0.6%	18.5	-25.8%	2.3%	71.5	-35.5%	5.2%	157.2	-43.6%	9.3%	277.1
	2	-14.2%	0.6%	19.6	-26.7%	2.5%	77.3	-37.1%	5.6%	173.7	-45.9%	10.1%	310.1
	0.6	-14.3%	0.5%	19.7	-24.4%	1.7%	69.2	-31.5%	3.5%	140.0	-36.9%	5.7%	228.6
	0.8	-14.3%	0.4%	18.2	-24.9%	1.6%	66.3	-32.8%	3.3%	137.7	-39.0%	5.6%	229.9
	1	-14.3%	0.4%	16.9	-25.2%	1.5%	62.2	-33.8%	3.1%	131.8	-40.5%	5.2%	222.0
30	1.2	-14.3%	0.4%	16.9	-25.6%	1.4%	64.2	-34.6%	3.0%	138.0	-41.9%	5.1%	235.9
	1.4	-14.2%	0.4%	15.9	-25.8%	1.4%	61.3	-35.1%	3.0%	133.3	-42.9%	5.2%	231.2
	1.6	-14.3%	0.3%	15.6	-26.2%	1.3%	60.7	-35.9%	2.9%	133.1	-44.0%	5.1%	232.2
	2	-14.6%	0.3%	16.1	-27.1%	1.3%	64.0	-37.5%	3.0%	143.4	-46.2%	5.3%	253.5

Figure 4: Exploration of the Pareto front for Sojourn Time and Product Loss - 10 machines

		Capacity reduction - 5%			Capacity reduction - 10%			Capacity reduction - 15%			Capacity reduction - 20%		
Metrology Ratio	v	Relative change in Sojourn Time	Relative change in Product Loss	Absolute change in Product Loss	Relative change in Sojourn Time	Relative change in Product Loss	Absolute change in Product Loss	Relative change in Sojourn Time	Relative change in Product Loss	Absolute change in Product Loss	Relative change in Sojourn Time	Relative change in Product Loss	Absolute change in Product Loss
5	0.6	-16.5%	1.1%	34.7	-28.1%	3.7%	113.9	-35.7%	7.3%	224.1	-41.5%	11.9%	364.3
	0.8	-16.4%	0.9%	30.3	-28.5%	3.3%	104.7	-37.1%	6.7%	214.8	-43.6%	11.2%	355.1
	1	-16.0%	0.9%	30.8	-28.3%	3.1%	109.2	-37.5%	6.6%	226.4	-44.6%	11.1%	381.9
	1.2	-15.6%	0.8%	33.0	-28.1%	3.1%	115.9	-37.6%	6.5%	241.7	-45.2%	11.1%	411.5
	1.4	-15.4%	0.9%	33.5	-27.9%	3.1%	116.5	-37.8%	6.7%	247.3	-45.7%	11.5%	424.2
	1.6	-15.4%	0.8%	30.3	-28.1%	3.0%	114.2	-38.2%	6.5%	248.1	-46.4%	11.5%	431.7
	2	-15.4%	0.8%	32.6	-28.3%	3.0%	121.7	-38.8%	6.8%	268.6	-47.5%	12.1%	476.3
10	0.6	-18.0%	1.0%	43.6	-29.9%	3.3%	146.6	-37.8%	6.5%	288.0	-43.5%	10.5%	462.5
	0.8	-17.5%	0.8%	41.4	-29.9%	3.0%	144.6	-38.7%	6.0%	292.0	-45.2%	9.9%	477.2
	1	-17.0%	0.8%	38.9	-29.6%	2.8%	138.3	-38.9%	5.8%	285.5	-46.0%	9.7%	474.9
	1.2	-16.5%	0.7%	38.1	-29.2%	2.7%	139.3	-38.9%	5.6%	291.7	-46.4%	9.6%	491.3
	1.4	-16.2%	0.7%	37.6	-29.1%	2.5%	139.7	-39.0%	5.5%	297.0	-46.9%	9.4%	504.7
	1.6	-15.9%	0.7%	36.3	-28.9%	2.5%	137.4	-39.1%	5.5%	295.7	-47.3%	9.5%	508.8
	2	-15.8%	0.6%	37.1	-28.9%	2.5%	143.0	-39.4%	5.5%	312.7	-48.1%	9.7%	545.4
30	0.6	-19.0%	0.6%	47.7	-31.1%	2.1%	161.0	-39.0%	4.2%	316.3	-44.7%	6.6%	503.5
	0.8	-18.3%	0.5%	41.0	-30.8%	1.8%	143.8	-39.6%	3.7%	290.2	-46.2%	6.0%	471.2
	1	-17.4%	0.5%	37.5	-30.1%	1.8%	135.7	-39.4%	3.7%	279.9	-46.5%	6.1%	463.1
	1.2	-16.8%	0.4%	35.7	-29.6%	1.6%	132.0	-39.3%	3.4%	277.6	-46.8%	5.8%	464.9
	1.4	-16.5%	0.4%	33.9	-29.4%	1.5%	127.1	-39.4%	3.3%	268.8	-47.3%	5.5%	453.8
	1.6	-16.3%	0.4%	33.2	-29.3%	1.5%	126.6	-39.4%	3.2%	270.7	-47.6%	5.5%	462.1
	2	-15.9%	0.4%	31.5	-29.0%	1.4%	122.7	-39.6%	3.1%	268.1	-48.3%	5.4%	465.0

Figure 5: Exploration of the Pareto front for Sojourn Time and Product Loss - 20 machines

5 CONCLUSIONS

This paper analyzed the trade-off between the expected product loss and the expected sojourn time when selecting the metrology capacity to use when optimizing the sampling periods of production machines that are controlled by a single metrology tool. Numerical experiments were conducted over a large number of scenarios. The results show that the metrology capacity that is optimal when minimizing the expected product loss can be safely reduced by up to 5% to decrease the expected sojourn time while not impacting much the expected product loss. Large decreases should be considered with scrutiny to determine if the expected cost in terms of quality issues is worth the gain in sojourn time. Over a relatively large set of experiments, it is also shown that, for a given number of production machines, reducing the metrology capacity has a relatively similar impact on the results regardless of the inspection rates. When the allocated metrology capacity is only slightly reduced (the general advocated change), the level of variability in the production cell has almost no impact on the results. This is no longer true if the allocated capacity is further reduced.

Our future research aims at considering the Effective Process Time (EPT) of all products instead of only the expected sojourn time in metrology of sampled products. Then, an optimization approach that minimizes both the expected product loss and the expected sojourn time will be developed. More extensive computational experiments will be conducted to analyze the trade-off between the two criteria and the impact of key parameters. Another interesting research avenue is to consider the scenario where production machines are waiting for sampled products to be measured before resuming production. In this case, product quality would not be impacted, only the sojourn times of products.

ACKNOWLEDGMENTS

This research was partly carried out while the second author was visiting the Department of Manufacturing Sciences and Logistics of the Center of Microelectronics in Provence of Mines Saint-Etienne, Gardanne, France, in September 2022 and September 2023.

REFERENCES

Bettayeb, B., S. Bassetto, P. Vialletelle, and M. Tollenaere. 2012. "Quality and exposure control in semiconductor manufacturing. Part I: Modelling". *International Journal of Production Research* 50(23):6835–6851.

Colledani, M. and T. Tolio. 2011. "Integrated analysis of quality and production logistics performance in manufacturing lines". *International Journal of Production Research* 49(2):485–518.

Dauzère-Pérès, S. and M. Hassoun. 2020. "On the importance of variability when managing metrology capacity". *European Journal of Operational Research* 282(1):267–276.

Dauzère-Pérès, S., J.-L. Rouveyrol, C. Yugma, and P. Vialletelle. 2010. "A smart sampling algorithm to minimize risk dynamically". In *2010 IEEE/SEMI Advanced semiconductor manufacturing conference (ASMC 2010)*, 307–310. IEEE.

Gilenson, M., M. Hassoun, and L. Yedidsion. 2015. "Setting defect charts control limits to balance cycle time and yield for a tandem production line". *Computers & Operations Research* 53:301–308.

Kingman, J. 1962. "On queues in heavy traffic". *Journal of the Royal Statistical Society. Series B (Methodological)*:383–392.

Lee, S., T. Lee, J. Liao, and Y. Chang. 2003. "A capacity-dependence dynamic sampling strategy". In *IEEE international symposium on semiconductor manufacturing*, 312–314.

Nduhura-Munga, J., S. Dauzère-Pérès, P. Vialletelle, and C. Yugma. 2012. "Industrial implementation of a dynamic sampling algorithm in semiconductor manufacturing: Approach and challenges". In *Winter Simulation Conference 2012 (WSC 2012)*, 1–9. IEEE.

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

Rodriguez-Verjan, G. L., S. Dauzère-Pérès, S. Housseman, and J. Pinaton. 2013. "Skipping algorithms for defect inspection using a dynamic control strategy in semiconductor manufacturing". In *Winter Simulation Conference 2013 (WSC 2013)*, 3684–3695. IEEE.

Shanoun, M., S. Bassetto, S. Bastoini, and P. Vialletelle. 2011. "Optimisation of the process control in a semiconductor company: model and case study of defectivity sampling". *International Journal of Production Research* 49(13):3873–3890.

Tirkel, I. and G. Rabinowitz. 2012. "The relationship between yield and flow time in a production system under inspection". *International Journal of Production Research* 50(14):3686–3697.

Tirkel, I., N. Reshef, and G. Rabinowitz. 2009. "In-line inspection impact on cycle time and yield". *IEEE Transactions on Semiconductor Manufacturing* 22(4):491–498.

AUTHOR BIOGRAPHIES

STÉPHANE DAUZÈRE-PÉRÈS is Professor at Mines Saint-Etienne, France, and Adjunct Professor at BI Norwegian Business School, Norway. He received the Ph.D. degree from Paul Sabatier University in Toulouse, France, in 1992 and the H.D.R. from Pierre and Marie Curie University, Paris, France, in 1998. He was a Postdoctoral Fellow at M.I.T., U.S.A., in 1992 and 1993, and Research Scientist at Erasmus University Rotterdam, The Netherlands, in 1994. He has been Associate Professor and Professor from 1994 to 2004 at the Ecole des Mines de Nantes, France. His research interests broadly include modeling and optimization of operations at various decision levels in manufacturing and logistics, with a special emphasis on production planning and scheduling, on semiconductor manufacturing and on railway operations. He has published more than 110 papers in international journals and contributed to more than 250 communications in national and international conferences. Stéphane Dauzère-Pérès has coordinated numerous academic and industrial research projects, including 5 European projects and more than 30 industrial (CIFRE) PhD theses, and also eight conferences. He was runner-up in 2006 of the Franz Edelman Award Competition, and won the Best Applied Paper of the Winter Simulation Conference in 2013 and the EURO award for the best theory and methodology EJOR paper in 2021. His email address is dauzere-peres@emse.fr.

MICHAEL HASSOUN is a senior 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.