Evaluating Cycle Time Performance of Integrated Metrology

Applied to 300 mm lithography and thermal treatment areas

R.A. van den Berg[1], E.J.J. van Campen[2], E. Lefeber[1], L.F.P. Etman[1] and J.E. Rooda[1]

6-February-2003

Abstract – Integrated metrology, one of the present developments in semiconductor fabs, influences the cycle time performance of a fab. A general hypothesis is that integrated metrology has a positive impact on cycle time performance due to reduction or even elimination of handling and waiting times. However, a negative impact on cycle time performance is also probable because of a reduced productivity of the combined tools (process tool with integrated metrology).

To investigate the overall effect of integrated metrology on cycle time performance, two 300 mm areas are considered: the lithography area and a mini-batch furnace in the thermal treatment area. For both areas the differences between integrated and stand-alone metrology are identified and for each difference the effect on cycle time is analyzed.

For the lithography area, integrated metrology results in about 20% cycle time improvement. For the mini-batch furnace this is about 3%. The difference between these results is explained and conclusions are drawn on the overall effect of integrated metrology on cycle time.

Index Terms – Integrated metrology, cycle time, 300mm lithography, thermal treatment

I. INTRODUCTION

Cycle time reduction is nowadays a key issue in semiconductor fabs. Integrated metrology is expected to have a positive impact on the cycle time by reducing handling and waiting times. On the other hand a negative impact on cycle time is probable because of a reduced productivity of the combined tools (process tool with integrated metrology). This paper aims to get a clear view of the overall effect of integrated metrology on the cycle time performance.

When reviewing the literature about the cycle time performance for integrated metrology ([1], [2], [3]), a common conclusion is that integrating metrology tools has a positive impact on the cycle time. However, this is not always the case. In this paper it will be made clear that under certain circumstances the cycle time performance can get worse for integrated metrology.

To investigate the cycle time performance of integrated metrology, it is compared to that of stand-alone metrology. Differences between the situation with integrated metrology and the situation with stand-alone metrology are described and the effect of these differences on the cycle time performance is discussed. Since the effect of integrated metrology on cycle time is not the same for different areas, it is interesting to consider more than one area. Therefore, two areas are investigated: the 300 mm lithography and thermal treatment areas.

Cycle time performances for the mentioned situations and areas are obtained using two analyses: a static analysis (using queuing theory) to determine the mean cycle time for an area, and a dynamic analysis (using discrete event simulation) to verify the results of the static analysis and to determine the variability of the cycle time for an area.

In this paper, first the two investigated areas are presented. Then the analyses that were used to determine the cycle time performance are described. Modeling assumptions are given as well as the assumed differences between the situations of integrated and stand-alone metrology. Next the queuing analysis and discrete event simulation results for the two investigated areas are presented. Finally the main conclusions on the cycle time performance of integrated metrology are drawn.

II. DESCRIPTION OF THE INVESTIGATED AREAS

The two areas under investigation concern the lithography and the thermal treatment area. The lithography area is considered as a whole, that is track, scan, inspections and measurements are all taken into account. For the thermal treatment area, one specific type of equipment is considered: a mini-batch furnace. This furnace is of special interest because of the possibility of advanced process control, which is enabled by integrated metrology.

The fab, in which the mentioned areas are present, is assumed to be in steady state, at a production level of about 2000 wafer starts per week (wspw). The considered process flow is a 120 nm node with DRAM [4]. In this flow a distinction is made between front-end (FE) and back-end (BE), because of contamination reasons: in the back-end a wafer gets in touch with copper, which may not contaminate wafers which are still in the front-end of the flow. Therefore, a distinction is also made between FE and BE operations.

First the lithography area is described, then the mini-batch furnace environment.

A. The lithography area

During the manufacturing process, a wafer is processed 44 times in the lithography area. After a lithography operation (193 or 248 nm), a lot can be measured or inspected using four types of metrology: macro inspection, micro inspection, overlay measurement and CD measurement. The number of lithography operations and their divisions in one complete process flow is indicated in Table 1. This table shows that some metrology tools are also used for lots that are not

[1] Systems Engineering, University of Technology, Eindhoven, The Netherlands.
[2] Philips Semiconductors Crolles2, Crolles, France

processed on lithography tools. The actual number of operations per wafer start at the metrology tools is follows from the data in Table 1 and a sampling rate, which will be discussed later.

FROM \ TO	248 nm FE	248 nm BE	193 nm FE	Macro FE	Macro BE	Micro FE	Micro BE	OVL FE	OVL BE	CD FE	CD BE	OUT
IN	24	17	3							3	12	
248 nm FE	X			10		11		2				1
248 nm BE		X			6		8					3
193 nm FE			X	1		1		1				
Macro FE				X		11						
Macro BE					X		6					
Micro FE						X		21		1		1
Micro BE							X		12			2
Overlay FE								X		22		2
Overlay BE									X		12	
CD FE										X		26
CD BE											X	24

Table 1: From-to matrix for lithography operations.

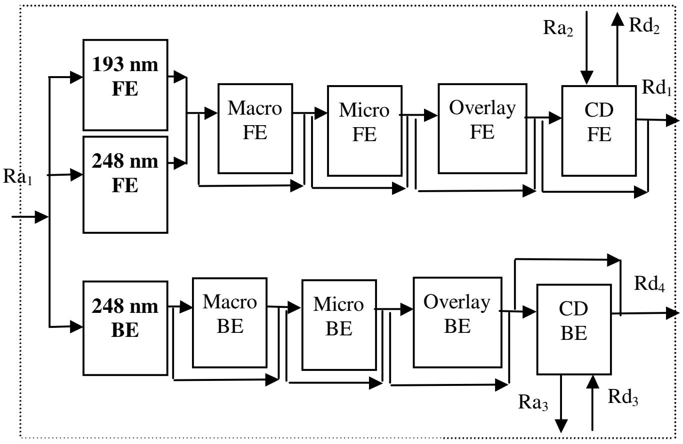

Figure 1: Structure of lithography area.

The structure of the lithography system is visualized in Figure 1. In this figure, each process block represents one buffer and a number of identical tools. The characteristics of these tools are presented in Tables 2a and 2b for the stand-alone situation. Front-end and back-end operations have the same characteristics.

Process operation	$t_{proc,0}$ [min]	$t_{is,0}$ [min]	Av	c_0^2
248 nm	40	16	70%	0.25
193 nm	40	16	70%	0.25

Table 2a: Lithography process data for one lot.

Metrology operation	$t_{proc,0}$ [min]	Av	c_0^2	Sample [w/l]	Sampling of lots
Macro	4	90%	0.25	5	50%
Micro	12	70%	0.25	3	30%
Overlay	13	90%	0.25	4	50%
CD	15	80%	0.25	2	50%

Table 2b: Lithography metrology data for one lot.

In Tables 2a-b, $t_{proc,0}$ is the mean raw process time, $t_{is,0}$ is the mean raw interstart time, *Av* is the availability of the tool on which the operation is performed and c_0 is the coefficient of variation of $t_{proc,0}$ (for a definition of $t_{is,0}$ and c_0, see section III). The sampling is indicated with two parameters: the number of wafers per lot (w/l) that is measured and the percentage of lots that is measured. For example: in one complete flow, a lot is inspected on average *11* (Table 1) *x 50%* (Table 2b) = 5.5 times with a FE Macro operation; each time only 5 wafers of this lot are inspected (Table 2b).

B. The mini-batch furnace environment

A wafer is processed 15 times at 5 different mini-batch furnaces during the manufacturing process. After a mini-batch furnace operation, three types of metrology are used to measure or inspect the wafers: particles measurement, micro inspection and thickness measurement. The number of operations and their divisions of one complete process flow are indicated in Table 3.

FROM \ TO	Nitride (FE)	Poly (FE)	TEOS (FE)	PMD (FE)	Cu&Final (BE)	Particles (FE)	Micro (FE)	Thickness (FE)	OUT
IN	3	1	3	1	7		7	17	
Nitride (FE)	X					3			
Poly (FE)		X				1			
TEOS (FE)			X			3			
PMD (FE)				X					1
CU&Final (BE)					X				7
Particles (FE)						X	7		
Micro (FE)							X	7	7
Thickness (FE)								X	24

Table 3: From-to matrix for operations in mini-batch furnace environment.

As can be seen in Table 3, the metrology steps follow only after a Nitride, Poly or TEOS operation.

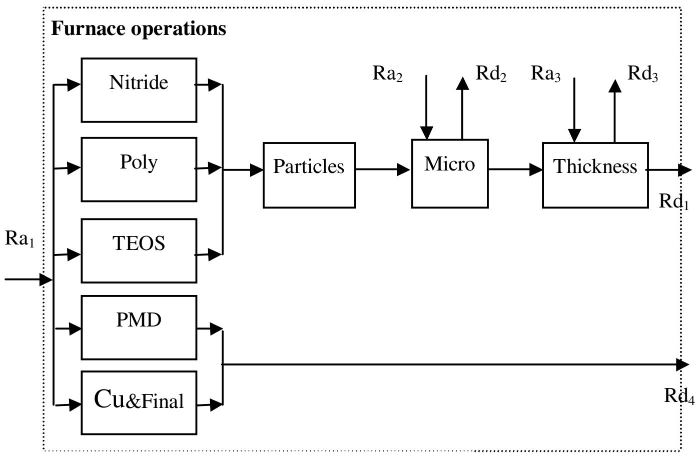

Figure 2: Structure of the mini-batch furnace environment.

In Figure 2 the structure of the mini-batch furnace environment is visualized. In this figure each process block represents one buffer and a number of identical process tools. The characteristics of these tools are presented in Tables 4a and 4b for the stand-alone situation.

Furnace operation	$t_{proc,0}$ [min]	$t_{is,0}$ [min]	Av	c_0^2
Nitride	111	108	90%	0.002
Poly	87	84	90%	0.000
TEOS	76	72	90%	0.003
PMD Densification	104	101	95%	0.000
Cu & Final Annealing	104	101	95%	0.000

Table 4a: Mini-batch furnace process data for one lot.

Metrology operation	$t_{proc,0}$ [min]	Av	c_0^2	Sample [w/l]	Sampling of lots
Particles	4	75%	0.25	4	100%
Micro	16	70%	0.25	4	30%
Thickness	5	80%	0.25	4	50%

Table 4b: Thermal treatment metrology data for one lot.

III. ANALYSES

To estimate the cycle time performance (mean cycle time and variance of the mean cycle time) of an area with integrated metrology, it is compared to the cycle time performance of an equivalent area with stand-alone metrology. These cycle time performances are determined using two analyses: a static analysis (using approximate relations for queuing theory) and a dynamic analysis (using discrete event simulation).

The static analysis is used to estimate the minimal required number of tools per equipment type as well as the mean cycle time for the area. The dynamic analysis is used for verification of the static analysis and to estimate the variance of the cycle time of the area.

Both analyses are performed for several production levels around the working point of 2000 wspw. Furthermore, a sensitivity analysis is performed for some assumptions of integrated metrology in order to determine their effect on the cycle time.

First the static analysis is described, then the dynamic analysis. In section IV the modeling assumptions are presented.

A. *Static analysis*

To estimate the mean cycle time for an area, the number of tools per equipment type should be known. However, since the choice of the number of tools also depends on the cycle time (this will be explained in subsection 2), the latter will be discussed first.

1) *Cycle time calculation for an area*

The mean cycle time for an area is a function of the mean cycle times per equipment type that is present in the area. The cycle time per equipment type is defined as the sum of the time to transport a lot to the equipment t_t, the time that a lot has to wait in a queue before being processed t_q, the time that a lot has to wait for an operator t_{op} and an effective process time $t_{proc,e}$:

$$CT = t_t + t_q + t_{op} + t_{proc,e}\,. \qquad (1)$$

Both t_t and t_{op} are assumed to be constant values, and estimation of these values is presented in section IV. The effective process time $t_{proc,e}$ is discussed later in this section. An approximation for the average queuing time t_q in a *G/G/m* system, with *m* identical machines, is presented by Whitt [5]:

$$t_q = \left(\frac{c_a^2 + c_e^2}{2}\right)\cdot\left(\frac{u^{\left(\sqrt{2(m+1)}-1\right)}}{m(1-u)}\right)\cdot t_{is,e}\,. \qquad (2)$$

Here, c_a and c_e are the coefficients of variation of respectively the interarrival rate r_a and the effective interstart time $t_{is,e}$; u is the utilization, which can be determined using Equation (3):

$$u = \frac{t_{is,e}}{m}\cdot r_a\,. \qquad (3)$$

The effective interstart time $t_{is,e}$ is defined as the inverse of the effective throughput of the tool, so the effective minimal time between the process starts of two subsequent lots. For the metrology operations, this time is equal to the effective process time. However, for a furnace operation and especially for a lithography operation, the effective interstart time is shorter than the effective process time, since there is some overlap of process times: more than one lot can be processed in one machine at the same time.

What is an effective process time and why use it instead of a raw process time? The effective process time [6] is introduced to simplify calculations with machine data. One could use the raw process time and take into account separately all ‘irregularities’ that influence this process time, such as setup times, machine availability, rework and operator availability. However, this would give rather complex computations. To simplify these computations, the behavior of the process time is described using only 2 parameters, in which all these ‘irregularities’ can be taken into account: the mean effective process time $t_{proc,e}$ and the coefficient of variation of this effective process time c_e. To determine the effective process time from the raw process time and the data of the ‘irregularities’, the following approximations can be used [6]. Since only the machine availability and rework have been taken into account (see section IV), these formulas are given here:

$$t_{proc,e} = \frac{t_{proc,0}}{A}\,, \qquad (4)$$

$$t_{proc,e} = \frac{t_{proc}}{(1-r)}\,. \qquad (5)$$

Equation (4) is used to take into account the availability A of a machine: the fraction of time that the machine can be used for production.
Equation (5) is used to take into account the effect of rework. Here t_{proc} is the effective process time without rework and r is the percentage of rework.
To determine the squared coefficient of variation of the effective process time formulas (6) and (8) can be used [6]:

$$c_e^2 = c_0^2 + \left(1 + c_r^2\right) \cdot A \cdot \left(1 - A\right) \cdot \frac{m_r}{t_{proc,0}}. \quad (6)$$

Equation (6) is used to take into account the effect of availability. Here, m_r is the mean time that a tool remains not available for production after it has been taken off-line; c_r^2 is the squared coefficient of variation of m_r. Further, c_0^2 is the squared coefficient of variation of the natural process time, which can be determined using Equation (7):

$$c_0^2 = \frac{\sigma_0^2}{t_{proc,0}^2}. \quad (7)$$

Here, σ_0 is the standard deviation of the mean raw process time $t_{proc,0}$. Equation (8) is used to take into account the effect of rework on c_e^2:

$$c_e^2 = \left(1 - r\right) \cdot c^2 + r. \quad (8)$$

Here c^2 is the squared coefficient of variation without the effect of rework taken into account.
Equations (4) – (8) can also be used to determine the effective interstart time from the raw interstart time. To do so, all process times should be replaced by interstart times in these equations.

Once the cycle time per equipment type is known, the cycle time for the area can be calculated using the information in section II. In one whole process flow, a lot visits on average p_i times equipment i in the area (sampling included) and enters the area altogether x times. The mean cycle time for the area is now defined as:

$$CT_{area} = \frac{\sum_i p_i \cdot CT_i}{x}. \quad (9)$$

It is noted that x is the number of times that the lot enters the area to be processed. If the lot enters the area only for a metrology operation, this is not included in x since in this case the metrology operation is performed after a process in another area. Since operations for other areas have nothing to do with the cycle time of this area, this entrance of the lot should not influence x. For the same reason, it is not included in p_i if the lot enters the area only for a metrology operation.

2 Number of tools per equipment type

The choice of the number of tools per equipment type is based on three factors:

- Utilization
- Cycle time factor
- Economics

The utilization of an equipment type may not exceed 100%, since this will cause an unstable system. However, since the utilization has a strongly non-linear effect on the cycle time (see Equation (2)) it is common to use an upper limit for the utilization that is less than 100% in order to prevent an excessive cycle time. This limit is given in section IV. Equation (3) shows that the utilization is dependent of the interstart time, the number of tools and the arrival rate. Since the interstart time is a fixed number, it is the number of tools that should be adjusted in order to get the utilization below the defined limit for a given arrival rate.
The cycle time factor CTF is defined in Equation (10):

$$CTF = \frac{CT}{t_{proc,e}}. \quad (10)$$

This cycle time factor is also bounded by an upper limit, which is given in section IV. The number of tools should be chosen in such a way that the cycle time (see Equation (2)) and thus the cycle time factor are small enough.
For economic reasons, the number of tools should be chosen as small as possible.

B. Dynamic analysis

For the dynamic analysis, discrete event simulations are used. First the simulation model is described and then the analysis of the output is discussed.

1) Description of the model

The discrete event model is written in the χ-language ([9], [10]). Within this language the behavior of system components is described by processes. In this model, processes are defined for each different action in the real system, such as transport, storage and processing of wafers. Furthermore, a few extra processes are defined to give the system a clear structure and to generate the output. The structure of the simulation model is visualized in Figures 3a and 3b.

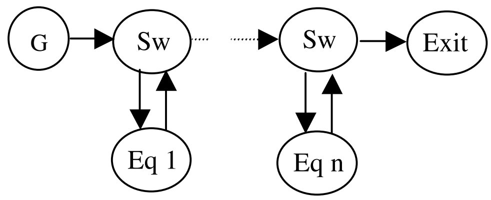

Figure 3a: Structure of the simulation model.

In Figure 3a, G represents a generator process, in which the lots are generated. Information about the equipment types that should be visited is also added to the lots. Sw is a switch process, in which it is decided whether a lot should visit the equipment type that is connected to the switch. This decision is based on the information tagged

to the lot and the sampling rate (for metrology tools). The sampling is performed with a Bernoulli distribution. In the *Exit* process all lots are received that leave the system, and output data is generated.
The equipment subsystems *Eq 1* to *Eq n* are subdivided into processes; this is clarified in Figure 3b. Each equipment subsystem describes one equipment type.

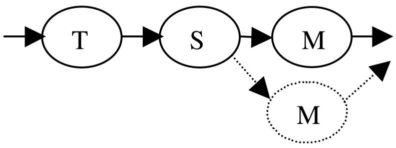

Figure 3b: The equipment subsystem.

As can be seen in Figure 3b, an equipment process consists of a transport process *T*, a stocker process *S* and one or more identical machine processes *M*. In the transport process, the transport time is determined per lot from an exponential distribution and the lot is delayed with this time before going to the next process. In the stocker process, lots are stored until they can enter one of the machine processes. In the machine process a lot waits a deterministic time for an operator to arrive and is processed afterwards for a time that is determined from a gamma distribution. After the lot has left the process, the next lot can enter. For the lithography equipment, this is slightly different, since this is a cascade machine: before one lot is finished the next lot can already start processing. Therefore, per lot that enters in a lithography process, two times are determined: the process time of the lot and the point of time that the next lot (if any) can enter the tool. Both times are determined using a gamma distribution.

2) Output analysis

The output analysis of the discrete event simulations is based on a steady state analysis. This implies that, due to an initialization period of the system, not all output data that are generated during the simulation can be used for this analysis; the data that are generated during the initialization period should be ignored. To determine the length of the initialization period a visual method is used. Mean cycle time and throughput of the system are plotted versus time. The initialization period is the time that is passed by until these values are stable.
To obtain a mean value with a small confidence interval [7] for the desired output data, several simulations should be run, and for each simulation the data generated during the initialization period should be ignored. This takes a lot of time. With the batch mean method [8], a simulation is only run once and the output data generated in the steady state part of the simulation are divided into batches. In this way the mean of a parameter can be calculated using the batch mean and the data of the initialization period has to be removed only once. However, because the batch means are all a result of one simulation, there is a positive correlation between the means, which causes that the mean and the confidence interval are not very reliable. To reduce this effect of correlation a large batch size is used and the simulation is performed twice, with 5 batches in each run.
The initialization period and batch mean method are visualized in Figure 4.
In each batch a mean value for all output parameters is calculated. With the results of all batches the overall mean and a confidence interval is calculated for all output parameters.

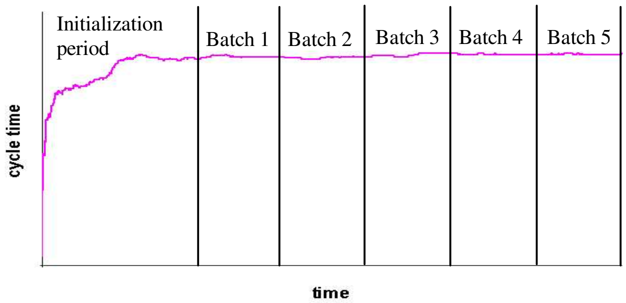

Figure 4: Initialization period and batches.

IV. MODELING ASSUMPTIONS

In order to analyze the cycle time performance of an area with integrated metrology, some assumptions have to be made. These assumptions are subdivided into general assumptions and assumptions that describe the difference between the situations of stand-alone and integrated metrology. Furthermore, in a sensitivity analysis it is determined which of these assumptions have the largest impact on the cycle time performance of integrated metrology.

A. General assumptions

- The arrivals for all equipment types are exponentially distributed.
- The control strategy is: First Come, First Serve. There are no ‘hot’ lots.
- The lithography module is regarded as one process, with one process time.
- The utilization of all machines may not exceed 90% and the cycle time factor should be lower than 2.5.
- The number of required tools is chosen in such a way that a process tool is the bottleneck.
- Each lot contains 25 production wafers.
- For all metrology tools and the lithography tool the c_0^2 is 0.25.
- Variability caused by availability and rework properties have been taken into account. Variability caused by setup times is negligible and operator variability has not been taken into account.
- There is 10% rework for the lithography operations. Rework for the mini-batch furnace is not possible.
- Mean transport time between two equipment types is 12 minutes. This is based on a tool-to-tool hand carried transport.
- Mean waiting time per lot for an operator is 10 minutes for a mini-batch furnace operation and 1 minute for a lithography operation.
- Mean waiting time per lot for an operator is 5 minutes for a metrology operation.
- m_r for a production tool is 4 hours.
- m_r for a metrology tools is 3 hours.
- m_r is exponentially distributed.

B. Assumptions for integrated metrology

- Wafers are still measured on the stand-alone metrology tools, but only at 20% of the original rate (verification or back up measurements). For the micro inspections, the original rate is maintained.
- The availability of the process tools is not influenced by integrated metrology.

1) Lithography

- Three metrology steps can be integrated in the process tool: macro inspection, overlay measurement and CD measurement.
- The measurement time for the three integrated metrology operations is 4.3 minutes per wafer (0.5 minutes for macro and CD; 3.3 minutes for overlay; handling times are ignored).
- Assuming that the last wafer of each lot is measured, the maximal increase of the raw process time per lot is 4.3 minutes. However, since not every lot is measured after each process (see Table 1 and the sampling in Table 2b), the mean increase of the raw process time is only 1.7 minutes.
- The throughput of the process tool is not influenced by the integrated metrology because of the parallel processing in the lithography cluster.

2) Mini-batch furnace

- Two metrology steps can be integrated in the process tool: particles and thickness measurements. The metrology is only integrated in process tools with the nitride, TEOS and poly operations. Here, the two metrology steps are combined in one integrated metrology tool.
- The measurement time for the two integrated metrology operations is 10.4 minutes per lot (4 wafers are measured; 1 minute per wafer for particles, 1 minute per wafer for thickness and 36 seconds per wafer for internal handling). However, since the thickness is only measured for 50% of all lots (see Table 4b) and 2 minutes of this time is masked by the discharge of other wafers of the same lot, the mean increase of the raw process time is 6.4 minutes.
- Since lots cannot be processed in parallel on the mini-batch furnace, the throughput is decreased by the integrated metrology; the raw interstart time is increased with 6.4 minutes.

C. Impact of integrated metrology assumptions

By integrating metrology, the production process is changed in several ways. The most important changes and their effect on the cycle time performance are summed up:

- Transport and waiting times for metrology tools are reduced or even eliminated. The longer these times are the more cycle time reduction is obtained.
- Measurement times of integrated metrology can be (partially) masked by production or discharge time of other wafers of the same lot. The more this can be masked the more cycle time reduction is obtained.
- Setup times for integrated metrology can be masked by production time. The longer these setup times are the more cycle time profit integrated metrology yields.
- By integrating metrology, the process tool becomes more complex, which can result in a lower availability. Reduction of availability results in a higher utilization and therefore longer waiting times (see Equation (4)) and thus a longer cycle time.
- The throughput of the process tool can be decreased, which results again in a higher utilization and thus a longer cycle time.
- There still can be a need for measurements on stand-alone metrology if the metrology step is integrated (verification or back up measurements). Since these are mostly extra measurements, this increases the cycle time.

Furthermore, it can be seen that if the cycle time for an area becomes shorter, the mean time to detect a failure is also decreased. This can result in less rework or scrap, which has again a positive effect on cycle time. This effect is not investigated further in this research.

V. CASE STUDY

The cycle time performance is analyzed for the areas described in section II, the results are presented in this section. First the results of the lithography area are presented, then those of the mini-batch furnace environment. Finally a comparison is made between the results of the two areas.

A. Lithography area

In the analyses, first the minimal required number of tools per equipment type is determined for a production level of about 2000 wafer starts per week. This number of tools is kept constant during the analyses.

Equipment	# Tools SA	Utilization SA Static analysis	Utilization SA Dynamic analysis	# Tools IM	Utilization IM Static analysis	Utilization IM Dynamic analysis
248 nm FE	6	0.81	0.86	6	0.81	0.85
248 nm BE	5	0.69	0.72	5	0.69	0.72
193 nm FE	2	0.30	0.31	2	0.30	0.31
Total ASML/TEL	*13*	-		*13*	-	
Macro FE	1	0.23	0.45	1	0.05	0.09
Macro BE	1	0.13	0.25	1	0.03	0.05
Micro FE	3	0.35	0.44	3	0.35	0.43
Micro BE	2	0.32	0.40	2	0.32	0.39
Overlay FE	3	0.52	0.68	2	0.16	0.20
Overlay BE	2	0.39	0.51	1	0.16	0.21
CD FE	4	0.54	0.66	2	0.31	0.40
CD BE	4	0.50	0.61	3	0.40	0.49

Table 5: Minimal required number of tools and utilization for 2000 wafer starts per week.

Regarding Table 5 it can be noticed that the utilization obtained from the dynamic analysis is higher for all

equipments. This is caused by the fact that in the simulation model the waiting time for an operator influences the interstart time of the tool, and thus the utilization (the tool is assumed to be busy, when it is waiting for an operator). For the static analysis, this is not the case. For equipment types with a high arrival rate (in terms of lots per unit of time) or a low number of tools this effect is larger, which is in agreement with Equation 5.
Furthermore it can be noted that some equipment types the number of tools is larger than one, while the utilization is very low. This is caused by the aim that the cycle time factor per equipment type may not exceed a certain level.

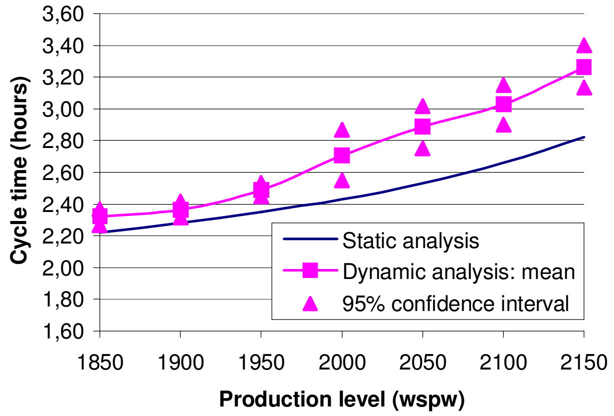

Figure 5a: Production level versus cycle time for stand-alone lithography area.

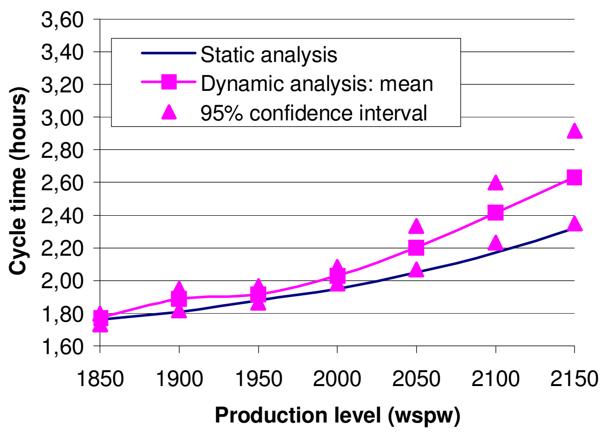

Figure 5b: Production level versus cycle time for lithography area with integrated metrology.

In Figures 5a and 5b the mean cycle time is visualized versus different production levels for the situations with respectively stand-alone and integrated metrology.
In these figures, the results of the dynamic analysis are structurally higher than those of the static analysis. This is caused by the differences between the two analyses. The most important difference is the utilization (see Table 5) and since this difference is larger for the stand-alone situation than for the integrated metrology situation, the cycle time improvement predicted by the dynamic analysis is a little higher than that of the static analysis.

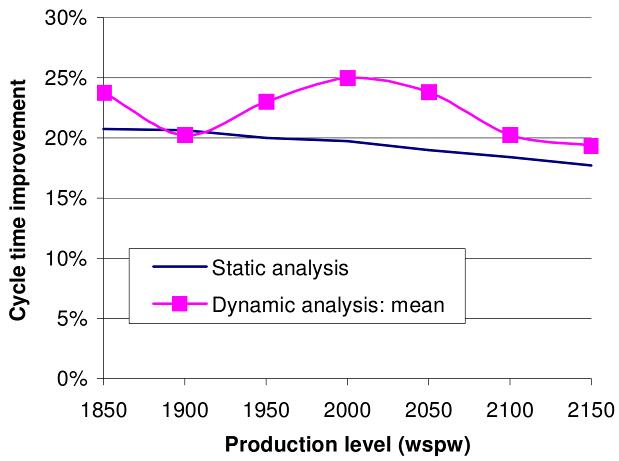

Figure 6: Cycle time improvement for integrated metrology.

In Figure 6 the cycle time improvement in terms of percentage is plotted. It can be seen that for a production level of 2000 wafer starts per week, about 20 % cycle time improvement can be gained. This corresponds to a mean cycle time improvement per lot of 29 minutes per processed layer or 21 hours in total.
However, these results are only valid if integrated metrology does not reduce the availability of the process tool. In the case that the availability is reduced the cycle time improvement will be significantly decreased. This is visualized in Figure 7 for a production level of 2000 wspw. Here it can be seen that for an availability decrease of only 4% the cycle time improvement is cancelled out entirely.

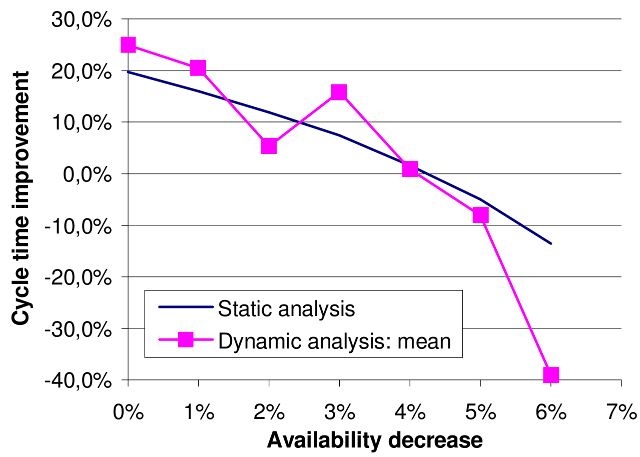

Figure 7: Effect of process tool availability decrease on cycle time improvement for 2000 wspw.

Another point of interest is the variability of the cycle time. From Figure 8 it follows that the standard deviation of the cycle time is smaller in case of integrated metrology. However, if the availability of the process tool decreases, the standard deviation will become larger. Furthermore it can be noticed that the reduction of the standard deviation of the cycle time is relatively smaller than the reduction of the mean cycle time. This implies that the coefficient of variation of the cycle time is larger in case of integrated metrology.

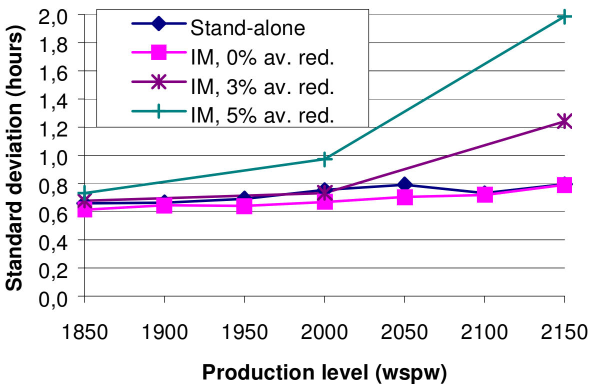

Figure 8: Effect of integrated metrology on standard deviation of cycle time.

As stated before, decrease of the process tool availability due to integrated metrology has a large impact on the cycle time improvement. But there are more parameters that have a large influence on the result. The most important parameters are identified:

- Throughput of the process tool: if the throughput drops with 3%, the cycle time improvement decreases from 20% to 14%.
- Setup time for stand-alone metrology: if there would be a mean setup time of 5 minutes for all stand-alone metrology in stead of no setup time, and this time is masked for the integrated metrology, the cycle time improvement increases from 20% to 27%.
- Need for stand-alone metrology: if there would be no need for measurements on stand-alone metrology in stead of 20% of the original rate, the cycle time improvement increases from 20% to 25%.

B. Mini-batch furnace environment

Since the results of the dynamic and static analyses show similar trends (see subsection A), only the static analysis is used to investigate the mini-batch furnace environment.

First the minimal required number of tools per equipment type is determined for a production level of about 2000 wspw and this number is kept constant during the analysis. These numbers are given in Table 6.

Furnace operation	**# tools SA**	***u* for SA**	**# tools IM**	***u* for IM**
Nitride	4	0.71	4	0.75
Poly	2	0.37	2	0.40
TEOS	3	0.64	3	0.69
PMD Densification	2	0.42	2	0.42
Cu&Final Anneal	7	0.84	7	0.84
Total mini-batch furnaces	*18*	-	*18*	-
Metrology operation				
Particles	2	0.15	1	0.06
Micro	2	0.38	2	0.38
Thickness	2	0.30	2	0.23

Table 6: Minimal required number of tools and utilization for 2000 wafer starts per week.

With these numbers of tools, the cycle time improvement for integrated metrology will be as indicated in Figure 9.

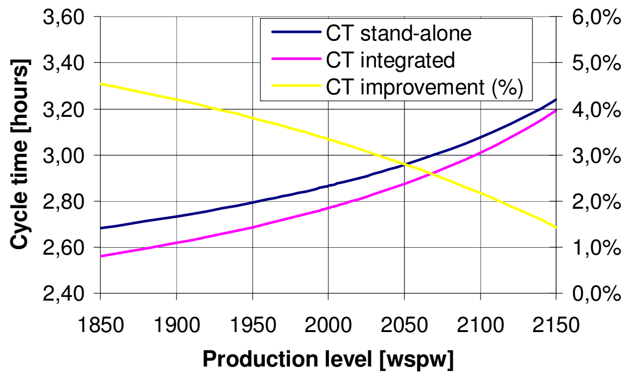

Figure 9: Cycle time improvement for integrated metrology.

It follows from Figure 9 that for a production level of 2000 wspw about 3% cycle time improvement is gained with integrated metrology. This corresponds to a mean cycle time improvement per lot of 6 minutes per layer or 86 minutes in total. However, these results are only valid if the process tool availability does not change as a result of integrated metrology. In Figure 10 can be seen that the effect of a decreased process tool availability on cycle time improvement is rather large: for a production level of 2000 wspw an availability decrease of about 2% already cancels out the cycle time improvement.

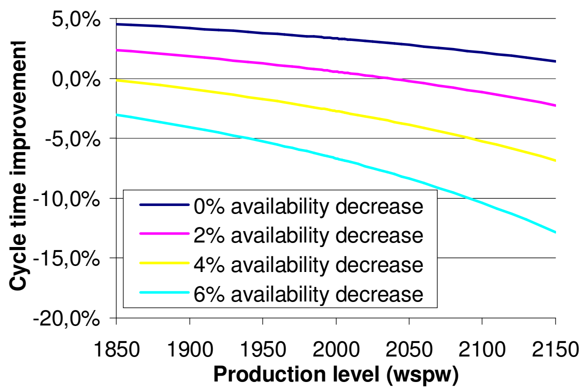

Figure 10: Effect of process tool availability decrease on cycle time improvement.

The process tool availability is not the only parameter that has a large impact on the cycle time improvement. Other parameters that have a large influence on the results are presented here:

- Throughput of the process tool: if the throughput would not be decreased by integrated metrology the cycle time improvement would be 8% in stead of 3%.
- Setup time for stand-alone metrology: if there would be a mean setup time of 5 minutes for all stand-alone metrology in stead of no setup time, and this time is masked for the integrated metrology, the cycle time improvement increases from 3% to 9%.
- Need for stand-alone metrology: if there would be no need for measurements on stand-alone metrology in stead of 20% of the original rate, the cycle time improvement increases from 3% to 5%.

C. Comparison of the two areas

Under the given circumstances the cycle time improvement is about 20% for the lithography area, where this is only 3% for the mini-batch furnace environment.

There are several reasons for this large difference, of which the most important will be described:

- For the lithography cluster the throughput is not decreased by integrated metrology, because of its cascade property: lots are processed in parallel. For the mini-batch furnace this is not the case: lots are processed sequentially and integrated metrology causes therefore a throughput decrease. In the previous subsections can be seen that the effect of throughput decrease on cycle time improvement is rather large.
- In the lithography area the measurement times of the integrated metrology are masked more than in the mini-batch furnace environment. This is again the result of the possibility in the lithography cluster to process wafers in parallel.
- Three metrology steps are integrated in the lithography cluster, where only two metrology steps (with shorter measurement times) are integrated in the mini-batch furnace.
- Since the cycle time for a lithography operation is shorter than that for a furnace operation, the cycle time improvement in terms of percentage is larger.

VI. Conclusions and Recommendations

A. Conclusions

Integrating metrology steps into process tools brings along changes in the production process. These changes can have a positive or a negative effect on the cycle time and it depends therefore on the conditions under which the metrology is integrated whether the overall effect is positive or negative.

Changes that have a positive effect on cycle time performance are:

- Reduction or even elimination of handling and waiting times for metrology tools
- Shorter time to detect failure: less rework/ scrap
- Masked setup times for integrated metrology tools
- (Partially) masked measurement times for integrated metrology tools

Changes that have a negative effect on cycle time performance are:

- Availability decrease for the process tools
- Throughput decrease for the process tools
- Need for both integrated and stand-alone metrology

The cycle time improvement for integrated metrology is different for different areas, since each area has its own characteristics. For the lithography area, a cycle time improvement of about 20% can be gained by integrating metrology. For the investigated mini-batch furnace in the thermal treatment area this is about 3%.

B. Recommendations

In this evaluation of integrated metrology only the aspect of cycle time has been investigated. Nevertheless, integrated metrology influences more aspects than just the cycle time, for example cost, quality and advanced process control. These aspects should also be considered when evaluating the overall effect of integrated metrology.

It is mentioned that integrated metrology causes a shorter time to detect a failure, and that this reduces rework or scrap. Although a reduction of rework or scrap probably has a large influence on the cycle time, this aspect is not investigated yet.

VII. References

[1] T. Claasen-Vujcic, A. Hasper and M. Abraham, *Analysis of a 200/300 mm Vertical Furnace with Integrated Metrology*, Solid State Technology, vol. 44, no. 4, pp S6-S12, April 2001.

[2] T. Stanley and J.C. Maia, *Evaluating Integrated Metrology Through Factory Simulation Models*, AEC/APC Symposium XIII, Banff, Alberta, Canada, October 2001.

[3] R. Noben, R. van Driel and T. Claasen-Vujcic, *Cycle Time Advantages of Mini Batch Manufacturing and Integrated Metrology in a 300 mm Vertical Furnace*, IEEE, ISSM 2001 Conference Proceedings Cat. No. 01CH37203, 2001.

[4] Process flow, HCMOS9: 120nm node with DRAM, version 16-09-2002, Crolles 2, Crolles, France.

[5] W. Whitt, "Approximating the GI/G/m Queue." *Production and Operations Management*, 1993.

[6] J.W. Hopp and M.L. Spearman, *Factory Physics: Foundations of Manufacturing Management*, Irwin/McGraw-Hill, New York, second edition, 2000.

[7] D.C. Montgomery and G.C. Runger, *Applied Statistics and Probability for Engineers*, John Wiley & Sons, Inc, New York, second edition, 2000.

[8] J. Resing, *Simulation*, Lecture notes 2S540, Eindhoven University of Technology, Department of Mathematics and Computing Science, 2001.

[9] J.E. Rooda, *Modelling of industrial systems*, Lecture notes 4C530, Eindhoven University of Technology, 2000, http://se.wtb.tue.nl.

[10] J.J.T. Klein and J.E. Rooda, χ *Manual*, Eindhoven University of Technology, Department of Mechanical Engineering, Systems Engineering Group, 2001, http://se.wtb.tue.nl.

APPENDICES

In these appendices, the models for the static and dynamic analyses are described in further detail. First the model for the static analysis is described, then the model for the dynamic analysis. Finally, the complete χ-code of the discrete event models is given.

APPENDIX 1: MODEL FOR STATIC ANALYSIS

The models for the static analysis of the lithography area and the mini-batch furnace environment are written in the Excel-files *Lithography.xls* and *Furnace.xls*. Since both models have the same structure, this structure is only described once.

The model is subdivided in 13 sheets. Calculations for the situation with stand-alone metrology and the situation with integrated metrology are both performed in the same model, but generally in different sheets. In the last sheets, the cycle time results of both situations are compared.

The first sheet is a general input sheet, in which most of the parameters that describe the system are inserted. This sheet also contains a field in which the final results are displayed, so that the effect of changing an input parameter can be investigated easily.

The second sheet is an input sheet for the machine parameters in the stand-alone situation. With these parameters and Equations (4) – (8), the mean effective process time and its squared coefficient of variation are determined for all equipment types. Only the effective process time data will be used in the rest of the model and not the other machine parameters.

The third sheet is a copy of the second sheet, but now with the data for the situation with integrated metrology. In fact this sheet is almost the same as the second sheet; only the machine parameters of the process tool are changed. The data for the stand-alone metrology tools (that are still present in the integrated metrology situation) does not change of course.

In the sheets *Flow* and *Flow IM*, an abstract of the process flow is described for respectively the situations with stand-alone and integrated metrology. In this abstract is defined for every time that a lot enters the area which equipment types should be visited. With this information and the sampling rate, the process- and transport times that are defined in previous sheets, the minimal cycle time (for a system without queuing) is calculated.

In the sheets *1st calculations* and *1st calculations IM*, the number of operations n_i in one whole process flow is defined per equipment type i and, using the information from previous sheets, the capacity of each equipment type is calculated, using Equation (A.1):

$$capacity_i = \frac{m_i}{t_{is,e} \cdot n_i}. \qquad \text{(A.1)}$$

With this information, the bottleneck capacity (in terms of wafer starts per unit of time) is determined. This is also the theoretical maximum throughput of the area. With the maximum throughput TP_{max} and the minimum cycle time CT_{min} that is determined in the previous sheets, the amount of work in process *WIP* is determined using Little's Law (see [6]), for the theoretical case of no queuing:

$$WIP = CT \cdot TP. \qquad \text{(A.2)}$$

To do so, first the critical WIP-level *WIP** is determined with Equation (A.3):

$$WIP^* = CT_{\min} \cdot TP_{\max}. \qquad \text{(A.3)}$$

Then, the cycle time and throughput of the area are determined for other WIP-levels using Equations (A.4) and (A.5):

$$CT = \begin{Bmatrix} WIP < WIP^* : CT_{\min} \\ WIP > WIP^* : \dfrac{WIP}{TP_{\max}} \end{Bmatrix}, \qquad \text{(A.4)}$$

$$TP = \begin{Bmatrix} WIP < WIP^* : \dfrac{WIP}{CT_{\min}} \\ WIP > WIP^* : TP_{\max} \end{Bmatrix}. \qquad \text{(A.5)}$$

This is done in the sheet *Results1* for both the situations with stand-alone and integrated metrology. With these results a first comparison can be made between the two situations. However, these results are based on a system without queuing, and therefore they are only an indication of the real results.

In the rest of the model, the system with queuing is considered.

In the sheets *2nd calculations* and *2nd calculations IM*, the cycle time per equipment type is calculated using Equations (1) – (3), for a production level of 2000 wspw. Then, using Equation (9) and the process flow information of the sheets *Flow* and *Flow IM*, the cycle time of the area is calculated for respectively the situation with stand-alone and integrated metrology.

In the sheet *Engine*, the calculations of the previous two sheets are repeated, but now for a series of different production levels between 0 and 2150 wspw. Furthermore, for each production level the difference in cycle time between the two areas is calculated.

The results from the sheet *Engine* are presented in the sheet *Results2*.

APPENDIX 2: MODEL FOR DYNAMIC ANALYSIS

A general description of this model is already presented in section III. Here, the source code of the model will be described. This source code is written in the language χ (chi), which is developed at the Eindhoven university of Technology ([9], [10]).

Two models have been written, one for the situation with stand-alone metrology *SA.chi* and one for the situation with integrated metrology *IM.chi*. Since these models have a lot in common, only the model for the situation with stand-alone metrology will be described and differences between the two models are indicated.

Some additional assumptions have been made in order to create the model. These assumptions will be presented first. Then a description of the χ-code is given.

A2.1. Additional assumptions for the dynamic model

- The system is in steady state: all 'layers' are processed in random order on the lithography tools.
- The effective process- and interstart times that are determined in the static analysis are used in this model as input parameters. Rework and availability data are included in these parameters.
- The buffers are assumed to be infinite.
- In reality, a lot cannot 'overtake' another lot during a lithography operation. However, in this model this can happen as a result of the used distributions.
- The transport times are exponentially distributed.
- The capacity of the transport system is infinite. This means that when there are more lots to be carried, more operators must be available, too.

A2.2. Description of the χ-code

The model is constructed with 2 different generator processes, a process that represents a lithography module, a process for the metrology operations, buffer processes, a transport process, collection processes, two switch processes and an exit process. These processes will be described one by one. Finally, the differences between the model for the situation with integrated metrology and the model for the situation without integrated metrology will be identified.

Generator processes

There is one generator process, *Glith*, for all the flows in the area that start with a lithography operation and one generator process, *Gmetr*, for the flows that visit only one metrology tool.

The first process has as input parameters: *b*, the production level (in lot start per hour), *c*, the number of times per process flow that a lot visits a lithography operation and *inp_SCV*, the squared coefficient of variation of the arrival rate. With a gamma distribution the time between two lot starts is simulated. The set of operations that should be undergone after being released in the area is chosen from 15 fixed combinations. This 'choice' is made using a uniform distribution, so that every combination appears as often as it should. The lot that is released carries along: the start time, a list with the names of the equipment types that should be visited, an id-number, a string in which comments can be placed (for verification) and a list in which the cycle time and process time per visited equipment type can be stored.

The process *Gmetr* is comparable to *Glith*, but simpler. This process has one extra input parameter: the name of the metrology operation that should be visited. The input parameter *c* now represents the number of times that this metrology operation is performed in one whole process flow. The lot that is released has the same structure as those from *Glith*, only the id-number is always 0. By doing so, these lots can be filtered out in the exit process, so that they are not taken along in the cycle time calculations.

By changing the input parameters of this process, it can be used for different metrology operations. In this model, the process *Gmetr* appears two times in the system: once for the CD_FE measurement and once for the CD_BE measurements. As can be seen in Table 1, the other metrology operations are only performed after an lithography operation.

Lithography process

The process *Mlitho* represents the lithography module. The input parameters of this process are the effective process- and interstart time and their squared coefficients of variation.

After a lot is received, a deterministic time is waited for an operator. This time is chosen deterministic to keep this model as close as possible to the static analysis. Then, two times are determined using gamma distributions: the time that the first wafer of the next lot can enter the lithography module and the time that the present lot can leave the module. The present lot is placed in a list that is sorted by the time that a lot can leave the module; the lot with the shortest remaining time is the head of the list. When the lot can leave the module, it is deleted from the list and placed in another list to be sent to the next process. Data about the process time are added to the lot. Also, as a proof that the lot has been processed, the name of the equipment type is added to the lot.

Process for other operations

For the operations on equipment types other then the lithography tool, only one process is written since all these operations are similar: a lot is received, some time is waited for an operator, the lot is processed (process data are added to the lot) and sent to the next process. Then, a new lot can be received.

The input parameters of this process are: the name of the equipment type, the effective process time and its squared coefficient of variation. By taking other values for the input parameters, another operation can be described with this process.

Buffer processes

There are 6 different buffers, for the situations that 1, 2, 3, 4, 5, or 6 parallel machines follow the buffer. The buffers are almost identical; only the channel through which the lot leaves the buffer is different. This is respectively a bundle of 1, 2, 3, 4, 5, or 6 channels. The buffer that is used is described in [9].

Transport process

The transport process is placed in the system between the switch process and a buffer process. In this way only one transport process has to be defined (with one incoming and one outgoing channel).

For each incoming lot a transport time is determined using an exponential distribution. Then the lot is placed in a sorted list, in which the lot with the shortest remaining transport time is the head of the list. When the transport time of the first lot in the list is passed, this lot is deleted from the list and placed in another list to be sent to the next process.

Switch processes

There are two different switch processes: one process *SwIN* that receives all the lots from the three generator processes and that functions as the switch for the first equipment type and one process *Sw* that receives a lot from the previous switch and that functions as a switch for one of the other equipment types.

In the process *SwIN*, a lot can be received from one of the generator processes. Once the lot is received it is determined, using the information in the lot, whether the lot should visit the equipment type that is connected to the switch. If so, the lot is placed in a list to be sent to the transport process. If the lot should not visit the equipment type, it is placed in a list to be sent on to the next switch process. Also a lot can be received from the 'connected' equipment type. Then, it is placed in the list to be sent on to the next switch process. If one of the mentioned lists contains one or more lots, these are sent immediately to their destination.

The other process, *Sw*, functions in the same way. The only difference is that this process receives lots from another switch process.

Further it is noted that for the switch processes that are connected to a metrology process, a sampling rate also is used to determine whether the lot should visit the equipment type or not. If the lot should visit the metrology process according to the information in the lot, then with a Bernoulli distribution the sampling is performed to determine whether the lot really should undergo the metrology operation.

Collection processes

The collection processes are placed in the system after the production processes and are used to receive lots from the parallel production processes and send those lots back to the switch using only one channel. The reason to include these processes instead of adapting the switch processes so that they can receive lots from multiple channels is that it simplifies the way to change the number of parallel metrology tools. Now, only other buffer- and collection processes have to be chosen (they already exist) and the number of parallel machines has to be adjusted, while the rest of the flow structure remains the same.

Exit process

In the exit process all lots are received that do not have to be processed or inspected anymore. After a lot is received, a selection takes place: only the lots that are processed on a lithography module are taken along in the cycle time calculations. This are the lots with an id-number > 0. With the data stored in the lot, the mean cycle time and throughput are updated and the new coefficient of variation of the cycle time is calculated. Then, with other data in the lot the mean effective process time and the cycle time per equipment type are updated for each visited equipment type. The data are matched with the right equipment type using an array.

Differences for integrated metrology

For the model with which the situation with integrated metrology is simulated, some processes are slightly different from the model just described. These differences are presented now.

- The effective process- and interstart time for the lithography operations are different. This is changed in the *Xper* of the model.
- An extra coefficient of sampling is added to the switch processes. The sampling coefficient that now is used as input parameter for the Bernoulli distribution is the product of the two sampling coefficients.
- The number of parallel tools for the (stand-alone) metrology operations is different. This is changed in the system *S* of the model.