

# Autocorrelation of Cycle Times in Semiconductor Manufacturing Systems

Alexander K. Schömig and Manfred Mittler

Bayerische Julius–Maximilians Universität Würzburg  
Institut für Informatik, Lehrstuhl für Informatik III  
Am Hubland, D-97074 Würzburg, Federal Republic of Germany  
Tel.: +49-931-8885511, Fax: +49-931-8884601  
e-mail: {schoemig|mittler}@informatik.uni-wuerzburg.de

**Abstract:** *Due to the high complexity of semiconductor manufacturing systems simulation has become a useful tool for the performance analysis of these systems. However, there are some serious problems in analyzing simulation output data. For example, the quality of confidence intervals for certain performance measures is affected by correlations among the data. Furthermore, analyzing the correlation of simulation output data might give valuable hints for understanding the behavior of complex systems such as semiconductor manufacturing systems.*

*In this paper we will present examples where we detected remarkable correlation phenomena in the simulation output of semiconductor fabrication facilities caused by exo- or endogenous factors. Based on real fab specifications we examine several semiconductor manufacturing systems in steady state. First we calculate the autocorrelation function of the sequence of successive cycle times and then compute the corresponding frequency spectrum. By this means, periodicities in the simulation output data can efficiently be determined.*

**Keywords:** *Analysis Methodology, Output Analysis, Semiconductor Manufacturing System, Cycle Times, Autocorrelation, Power Density Spectrum*

## 1 Introduction

Semiconductor manufacturing is among the most complex manufacturing processes according to Gise and Blanchard (1986) and Sze (1983). A *semiconductor chip* is a highly miniaturized, integrated electronic circuit consisting of thousands of components. Every semiconductor manufacturing process starts with raw *wafers*, a thin disc made of silicon or gallium arsenide. Up to a few dozens of identical chips can be made on each wafer, building up the electronic circuits layer by layer. Depending on the scale of integration, the type of chip, customer specs, the whole manufacturing process may require up to 500 single steps. Most of these operations involve cleaning, deposition, lithography, etching, ion implantation, and testing, and they are therefore basically of the same type. Note, that there are no assembly operations before reaching the backend stage of the production process.

Several performance measures are commonly used to describe and assess a semiconductor manufacturing facility. To highlight the most important of those we mention machine utilization, production yield, throughput, and cycle time. We define cycle time as the time a lot of

wafers needs to travel through the core semiconductor manufacturing process. The capability of meeting due dates has become a crucial factor in global manufacturing competitiveness. Consequently fab managers have to ensure short and predictable cycle times.

Simulation is gaining more and more importance in the planning process of a semiconductor fabrication facility, using sophisticated and well designed simulation tools. However, most studies published so far only consider the mean and variance of certain performance measures of the system. This is done although several authors, e.g. Law and Kelton (1991) and Fishman (1978), stress the fact that simulation output is always correlated and may lead to wrong variance estimates. Furthermore, the analysis of the correlation of the simulation output data might give valuable hints for the understanding of complex systems. Finally, one cannot deny the fact that correlated input to a queueing system may have a tremendous effect on its performance, cf. Livny, Melamed, and Tsiolis (1993) and Rose (1995). Consequently, if the output process of a system is heavily correlated it impairs the performance of downstream production stages, e.g. in terms of cycle times. This could be for example a facility where semiconductor products are packed for shipping or used as assembly parts.

The major methods that estimate the autocorrelation structure of time series in general and simulation output data in particular are the *autoregressive method*, developed by Fishman (1973) and the method of *spectrum analysis*, cf. Schlittgen (1987) and Oppenheim and Schaffer (1989). However, from our point of view, the latter method has not become a common tool among simulation practitioners, probably due to the fact, that this technique, according to Law (1983), requires “a fairly sophisticated background on the part of the analyst”.

In this paper we will present examples where we detected remarkable correlation phenomena in the simulation output of semiconductor fabrication facilities caused by exo- or endogenous factors, using the method of spectrum analysis.

## 2 Literature Review

Due to the growing competition among semiconductor manufacturers, research and development efforts in terms of the optimization of semiconductor wafer fabrication have increased. Dayhoff and Atherton (1984) and Lohrasbpour and Sathaye (1984) presented simulation models of semiconductor manufacturing facilities, and discussed the influence of system parameters on the overall performance of these facilities. Glassey and Resende (1988) introduced a closed loop job release mechanism for jobs in which randomness is primarily caused by machine downtimes. They adapted certain concepts of inventory control methods to the context of job shop scheduling and compare these concepts with other input control mechanisms. A more general approach is the *CONWIP principle*, cf. Hopp and Spearman (1991), Spearman (1992). Wein (1988) assessed the impact of several scheduling strategies on the performance of semiconductor wafer fabrication in terms of cycle time. He observed that input control has an even greater importance in reducing cycle times.

To the authors’ knowledge, there are very few papers concerning the application of the method of spectrum analysis. Law (1983) summarizes the techniques for statistical analysis of simulation output data. Duket and Pritsker (1978) discuss some issues of spectral methods. Heidelberger and Welch (1983) present a method for placing confidence intervals on the

steady state mean of a discrete event simulation output sequence. This method is based on obtaining estimates of the variance by estimating the spectral density at zero frequency. They show some typical examples for response time and waiting time spectra of queueing systems that do not exhibit long term correlations. Schruben and Cogliano (1987) present an experimental procedure for simulation response surface model identification. Since they run their experiments in the frequency domain, they use the frequency spectrum of the output process to determine significant factors. They focus on arbitrary chosen, exogenous factors that cause certain oscillations in the simulation output process. In a recent study, Rose (1995) deals with statistical properties of MPEG video traffic in ATM (Asynchronous Transfer Mode) networks. He shows that MPEG video traffic may have significant long term correlations that could cause several problems in ATM networks.

### 3 Power Density Spectrum

Correlations and periodicities in simulation output data can be determined efficiently by means of the *power density spectrum* which is also known as the *power spectrum*, see Oppenheim and Schafer (1989) and Fishman (1978). We briefly describe the idea behind this approach according to Oppenheim and Schafer (1989) and Schlittgen (1987).

Let  $X = (x_0, \dots, x_{N-1})$  a finite sequence obtained by sampling a strictly stationary continuous-time stochastic process  $X(t)$ . Assuming that this sequence can be represented by a number of harmonic oscillations with specific frequencies then the question is how does the oscillation with frequency  $\omega$  contribute to the entire time sequence as described by Schlittgen (1987). The intensity of this contribution is given by the power density spectrum.

There are two different approaches to estimate the power density spectrum, both make use of the *discrete Fourier transformation*. First, the discrete Fourier transformation can directly be applied to the finite sequence  $X$ . In the literature, this approach is referred to as *periodogram analysis*. Alternatively, one may first estimate the autocorrelation of the finite sequence  $X$  and then take the discrete Fourier transform of this estimate. The objective of both approaches is to obtain unbiased consistent estimators of the power spectrum. However, it is very difficult to achieve this goal such that we have to restrict ourselves to approximations, cf. Oppenheim and Schafer (1989).

Now, if  $X(\omega_k)$  denotes the discrete Fourier transform of  $X$  with  $\omega_k = 2\pi k/N$ ,  $k = 0, \dots, N-1$ , i.e.

$$X(\omega_k) = \sum_{n=0}^{N-1} x_n \cdot e^{i\omega_k n}, \quad (1)$$

then samples of the power spectrum  $I(\omega_k)$  are defined by

$$I(\omega_k) = N^{-1} \cdot |X(\omega_k)|^2. \quad (2)$$

We also refer to this estimate of the power spectrum as the *periodogram* since sequence  $X$  is obtained from a sequence of infinite length by simply cutting out  $N$  subsequent samples out of  $X$ . In the literature, this technique is called *windowing by a rectangular window*.

As mentioned above, the periodogram can alternatively be calculated by taking the discrete Fourier transform of the estimate of the autocorrelation of the sequence  $X$  denoted by  $c(l)$ . It is

$$c(l) = \sum_{n=0}^{N-1} x_n \cdot x_{n+l}, \quad (3)$$

where  $l$  denotes the lag between two samples of  $X$ ,  $|l| \leq N - 1$ . Then, the estimate of the power spectrum is

$$I(\omega_k) = N^{-1} \cdot \sum_{l=-(N-1)}^{N-1} c(l) \cdot e^{i\omega_k l}. \quad (4)$$

Note that since the periodogram is an odd function ( i.e.  $I(\omega) = I(-\omega)$  for any arbitrary  $\omega$ ) and since the periodogram has period 1 ( i.e.  $I(\omega + 1) = I(\omega)$  for any arbitrary  $\omega$ ) it is sufficient to show periodograms only for  $\omega \in [0; \pi]$ , cf. Schlittgen (1987).

The approximation of the estimate of the power spectrum can be improved by periodogram averaging following Oppenheim and Schaffer (1989): The original sequence is divided into  $K$  batches of length  $N/K$  each. Then, the average periodogram is obtained by averaging the  $K$  periodogram estimates  $I_r(\omega)$ ,  $r = 1, \dots, K$ , i.e:

$$\bar{I}(\omega) = K^{-1} \cdot \sum_{r=1}^K I_r(\omega). \quad (5)$$

Periodogram averaging yields an asymptotically unbiased consistent estimate of the power spectrum when the length of the sequence  $X$  tends to infinity so that  $N$  and  $K$  are allowed to increase.

We conclude this brief introduction by remarking that the discrete Fourier transform can efficiently be computed by using *Fast Fourier Transform* techniques (FFT), especially if the length of the sample sequence is a power of 2. Otherwise, we augment the sequence by zeros until the corresponding length is achieved. Finally, we refer the interested reader to the original literature for further details.

## 4 Simulation Models

We examine several production environments by means of simulation. These environments are introduced briefly.

### 4.1 $M/M/1 - \infty$ Queue with Bernoulli Feedback

We consider this simple model since one might expect that rework causes correlations in the sequence of successive cycle times. At this queue customers arrive according to a Poisson process with rate  $\lambda$  whereas service times of the single server are exponentially distributed with mean  $\mu^{-1}$ . In this model rework is incorporated via the feedback probability  $p$ . After

service time completion customers are fed back to the tail of the queue with probability  $p$ . Since for an arbitrary customer this probability is independent of its history, i.e. independent of the number of feedbacks that customer already experienced, this kind of feedback is called Bernoulli feedback, see for example van den Berg and Boxma (year). Taking into account the feedback probability  $p$  it can be shown that the load of this queueing system is given by  $\rho = \lambda/[\mu(1 - p)]$ .

We examined this model for a rework probability of 30% and adjusted the arrival and service rates in a way such that a load of  $\rho = 0.3$  was obtained. Further, we stopped the simulation after 25,000 customers have left the system and assumed that the initial transient period was finished after 5,000 customers.

## 4.2 Wein’s Semiconductor Fab

Here we use the model of a wafer manufacturing facility given by Wein (1988). We only sketch the major properties and parameters of this model and refer the reader to the original literature for further details. The facility consists of 24 work centers, where identical machines are grouped together. Lots are processed in 172 operation steps, building up a 13-layer structure. The machines at 21 work centers are subject to failures, such as unscheduled breakdowns or maintenance. Two different types of probability distributions are used to model processing times and times regarding failures. Processing times are sampled from a Erlang-2 distribution ( $CV = \frac{1}{\sqrt{2}} \approx 0.707$ ). Mean time between failures and mean downtimes are considered to be Gamma distributed with a shape parameter equal to 1.5 ( $CV = \sqrt{2} \approx 1.414$ ). We apply the closed loop input mechanism (also called CONWIP). This means, that the number of lots in the line is held constant to a certain value by dispatching a lot whenever a lot exits. We simulated 1,000,000 hours of operation and discarded statistical data collected during the initial 50,000 hours.

## 4.3 MIMAC Semiconductor Fabs

We simulate factory-level data sets of four different semiconductor fabs, including ASICs, microprocessors, and non-volatile memory. These datasets were developed by the MIMAC project (MIMAC: *Measurement and Improvement of MAnufacturing Capacity*, joint project between SEMATECH, Austin, TX, U.S.A., and the European JESSI project), can be obtained via anonymous FTP from SEMATECH. An overview of these data sets concerning the product to be manufactured, product mix, number of processing steps including rework, number of wafer starts per month (WSPM), average number of mask layers, etc is given in Tab. 1.

For evaluation purposes we collected the cycle time of 40,000 lots of product 1 for the original data set 3. Initial 10,000 lots of the same type were discarded. The load of this fab was adjusted in a way to obtain a load of 30% at the bottleneck tool.

We apply the methodology of the power density spectrum to sequences of successive cycle times obtained by simulating the aforementioned manufacturing facilities in statistical equilibrium. In each case we plot the autocorrelation function and the periodogram of the

sequence of successive cycle times.

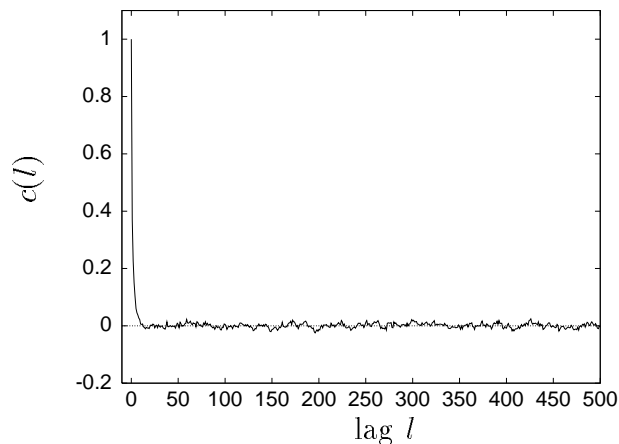
**Table 1:** SEMATECH Testbed Data Sets

Data Set	1 : Logic <sub>1</sub>	3 : Logic <sub>2</sub>	5a : ASIC <sub>1</sub>	6 : ASIC <sub>2</sub>
Type of Factory	Commodity	Commodity	ASIC	ASIC
Product Mix	2	11	21	9
approx. WSPM	16,000	21,400	10,000	5,500
Avg. #Mask Layers	15	35	30	30
#Processing Steps	210 / 245	298 – 533	117 – 259	234 – 356
#Tools	83	73	85	104

## 5 Numerical Results

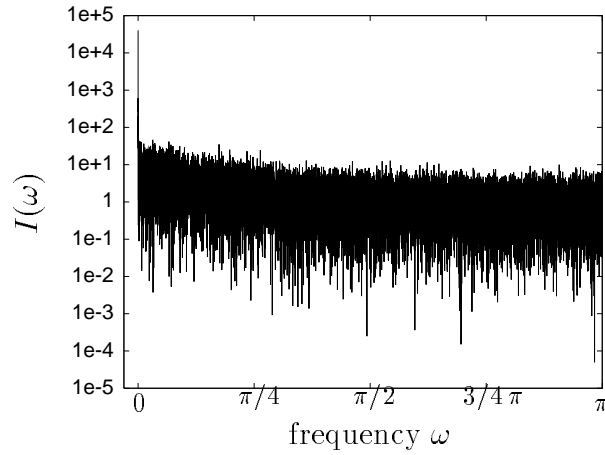
### 5.1 $M/M/1 - \infty$ Queue with Bernoulli Feedback

As far as the autocorrelation function is concerned the Bernoulli feedback has no impact with increasing lag, see Fig. 1. After a steep descent the autocorrelation function oscillates around the abscissa.



**Fig. 1:** Autocorrelation of Cycle Times in the  $M/M/1 - \infty$  Queue with Bernoulli Feedback

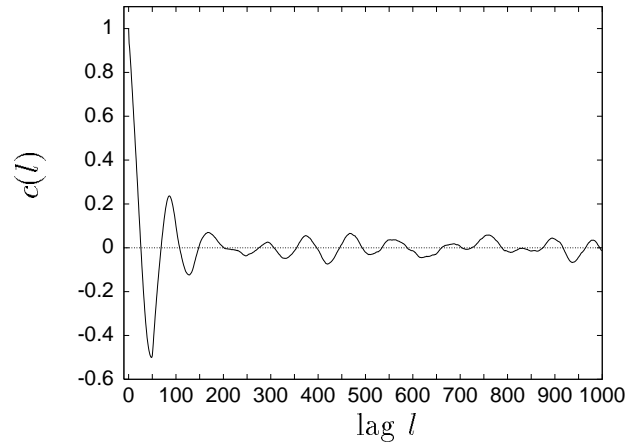
The conjecture that there are no periodicities in the sequence of successive cycle times is confirmed by the corresponding periodogram shown in Fig. 2. We also examined this system for varying loads and varying feedback probabilities. The results obtained are similar to these discussed here. This also holds for a  $D/D/1 - \infty$  system with Bernoulli feedback.



**Fig. 2:** Periodogram of the  $M/M/1 - \infty$  Queue with Bernoulli Feedback,  $K = 1$

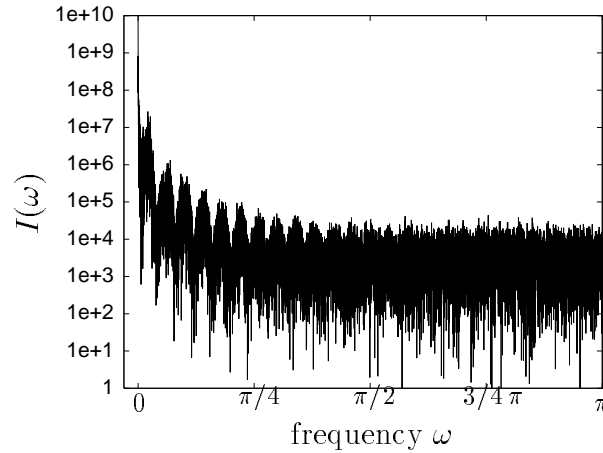
## 5.2 Wein's Semiconductor Fab

Apparently, Fig. 3 shows that the autocorrelation function of successive cycle times oscillates very strong around zero. Since we apply the closed loop rule with a WIP of 50 lots the autocorrelation is decreasing up to lag 50. Then it is increasing again.



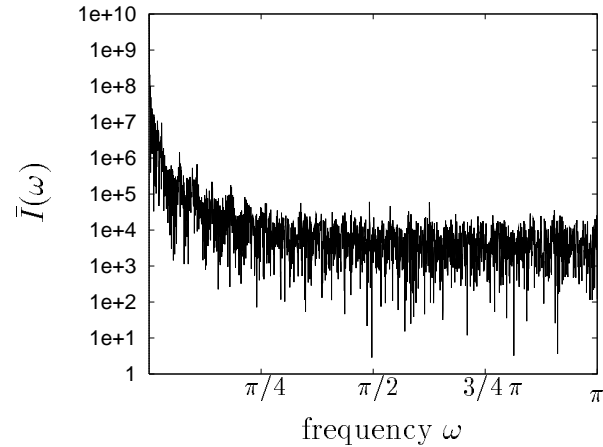
**Fig. 3:** Autocorrelation of Cycle Times in Wein's Semiconductor Fab

We conclude that due to the application of the closed loop rule with sequencing lots in order of their arrival (FCFS) the lot with number  $M$  is not allowed to enter the production system until the lot with number  $M - 50$  has left. However, this strong relationship becomes weak with increasing lag. Nevertheless, the periodogram shows some spectrally smeared peaks (see Fig. 4).



**Fig. 4:** Periodogram of Wein's Semiconductor Fab,  $K = 1$

Most probably, this smearing effect is due to the finite length of the sequence of successive cycle times. If we consider a sequence which is approximately *5 times* longer than the original one, we obtain the periodogram shown in Fig. 5. In this periodogram these peaks are narrower compared to those in Fig. 4.



**Fig. 5:** Periodogram of Wein's Semiconductor Fab for a longer sequence

Now let us discuss the peaks in the context of the corresponding sequence of successive cycle times: The distance  $\omega^*$  between the first 11 peaks of the periodogram is almost the same and equal to approximately  $0.04\pi$ . If we calculate the period  $P^*$  between two peaks according to Fishman (1978, page 261) by  $P^* = 2\pi/\omega^*$  we obtain  $P^* \approx 50$ . This implies a cyclic behaviour in the original sequence of cycle times with a period of approximately 50 lots. Note that the length of the period corresponds exactly to the maximum inventory of 50 lots which is due to the application of the closed loop rule.

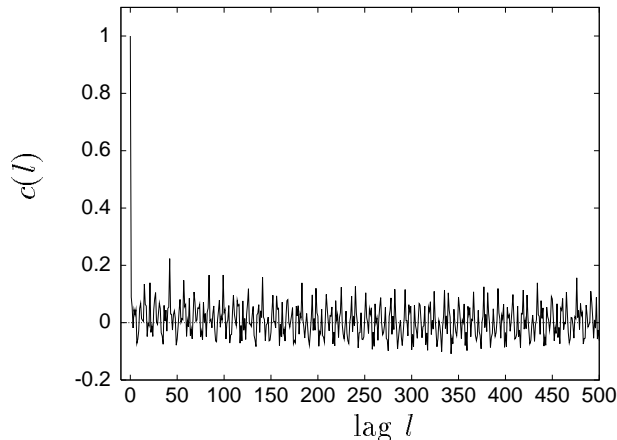
As Fishman (1978, page 262) states, the remaining peaks of the periodogram “are most likely harmonics of the fundamental frequency that contribute to the shape of the cyclic component” in cycle time. This periodicity is amazing and from a fab managers point of view not desirable.



Similar effects occurred for several side experiments using different seeds and different queuing disciplines, for instance shortest remaining processing time (SRPT). Thus, we conclude that most likely the periodicity in cycle times is due to the closed loop rule which is an exogenous factor.

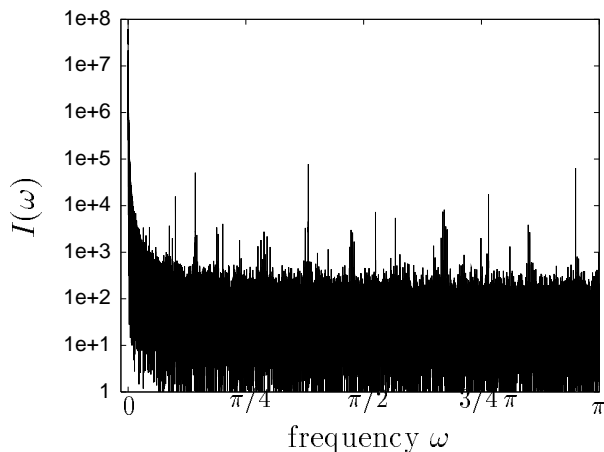
### 5.3 MIMAC Semiconductor Fabs

Unlike the autocorrelation function (see Fig. 1) of the Bernoulli feedback model we observe here (see Fig. 6) that the autocorrelation function for the MIMAC data set 3 does not settle near zero.



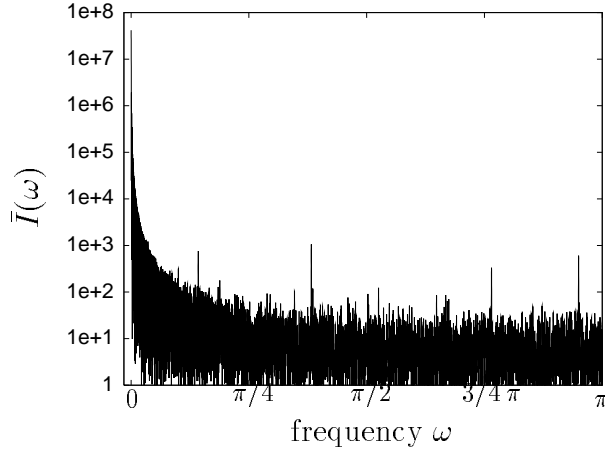
**Fig. 6:** Autocorrelation of Cycle Times, MIMAC Data Set 3

Like for Wein’s fab we find an oscillating autocorrelation function, here within an interval of  $[-0.1; 0.2]$  and with a much higher frequency. Again we use periodograms to obtain more information about periodic effects in the simulation output. Fig. 7 is the periodogram of the output of a single simulation run using MIMAC data set 3. Since *averaged* periodograms, according to Oppenheim and Schaffer (1989), show better the periodic effects by suppressing random influence, we turn to Figs. 8 and 9. These average periodograms are derived from independent simulation runs using *different* seeds.

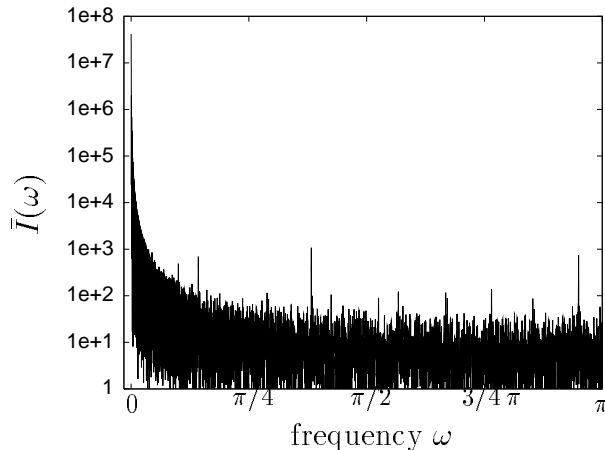


**Fig. 7:** Periodogram of MIMAC Data  
Set 3,  $K = 1$

In each periodogram one can easily identify several sharp, isolated peaks. There are dominant peaks at  $\omega = 0.10\pi, 0.14\pi, 0.38\pi, 0.52\pi, 0.67\pi, 0.74\pi,$  and  $0.95\pi$ . Note, that the peaks at  $\omega = 0.10\pi, 0.38\pi, 0.67\pi, 0.95\pi$  are equidistant, while the groups of peaks at  $\omega = 0\pi, 0.14\pi, 0.38\pi$  and at  $\omega = 0.38\pi, 0.52\pi, 0.74\pi$  form the same pattern. We did simulation runs using several different seeds. We present here the diagrams of two simulation runs for comparison purpose. In any case, we could identify these patterns in each diagram.



**Fig. 8:** Average Periodogram of MIMAC Data Set 3,  $K = 10$



**Fig. 9:** Average Periodogram of MIMAC Data Set 3 with Different Seed,  $K = 10$

Unfortunately, from our point of view there is no obvious explanation for these results, since we do not deal with an exogenous factor that itself shows some cyclic behavior. As Fishman (1978) points out, cyclic patterns in the simulation output data may also be caused by endogenous system features like batching as in his laundry shop example. In fact, in semiconductor manufacturing we have to deal with batching policies as well as rework, re-entrant

job flow, and machine breakdowns to name a few. All of these factors were incorporated in the MIMAC simulation model. Schruben and Cogliano (1987) showed that the interaction of factors results in a response at the sum and the difference of the frequencies of the single factors. In consideration of this result we believe that the observed patterns could be explained by the interaction of at least two endogenous factors. However, further research is needed to identify these factors and to explain the cause for the interaction.

## 6 Conclusion

We investigated the periodicity of cycle times in semiconductor fabrication facilities by means of the frequency spectrum. The result obtained exhibit periodicities which are due to either endogenous or exogenous factors. On the one hand, the limitation of the maximum inventory of the manufacturing system achieved by the application of the CONWIP rule represents an exogenous factor which leads to strong periodicities (as shown for Wein's semiconductor fab in Figs. 4 and 5).

On the other hand, there are also some hidden endogenous factors that contribute to the periodicity of cycle times (cf. Figs 7, 8, and 9). There are grounds for the assumption that rework, breakdowns, and batching effects are among those. However, further research has to be done to reveal the significance of each single factor.

**Acknowledgement** The authors would like to thank Prof. Dr. P. Tran-Gia for the support of this work and O. Rose for stimulating discussions. The programming efforts of M. Grasmann are greatly appreciated. This research was supported by the Deutsche Forschungsgemeinschaft (DFG), grant #Tr257/2-1.

## References

- Dayhoff, J. E. and R. W. Atherton (1984, December). Simulation of VLSI manufacturing areas. *VLSI Design*, 84-92.
- Duket, S. D. and A. A. B. Pritsker (1978). Examination of simulation output using spectral methods. *Math. Computer Simulation* 20, 53-60.
- Fishman, G. S. (1973). *Concepts and Methods in Discrete Event Digital Simulation*. New York, NY: John Wiley & Sons.
- Fishman, G. S. (1978). *Principles of Discrete Event Simulation*. New York, NY: John Wiley & Sons.
- Gise, P. and R. Blanchard (1986). *Modern semiconductor fabrication technology*. Englewood Cliffs, NJ: Prentice-Hall.
- Glassey, C. R. and M. G. C. Resende (1988, February). Closed-loop job release control for VLSI circuit manufacturing. *IEEE Transactions on Semiconductor Manufacturing* 1(1), 36-46.

- Heidelberger, P. and P. D. Welch (1983). Simulation run length control in the presence of an initial transient. *Operations Research* 31, 1109–1144.
- Hopp, W. J. and M. L. Spearman (1991). Throughput of a constant work in process manufacturing line subject to failures. *International Journal of Production Research* 29(3), 635–655.
- Law, A. M. (1983). Statistical analysis of simulation output data. *Operations Research* 31(6), 983–1029.
- Law, A. M. and W. D. Kelton (1991). *Simulation Modeling & Analysis* (2nd ed.). New York: McGraw–Hill.
- Livny, M., B. Melamed, and A. Tsiolis (1993, March). The impact of autocorrelation on queueing systems. *Management Science* 39(3), 322–339.
- Lohrasbpour, E. and S. Sathaye (1984). Simulation modeling of IC wafer fabrication lines. In *Semicon/West Technical Program*, pp. 93–99.
- Oppenheim, A. V. and W. Schafer, Roland (1989). *Discrete-Time Signal Processing*. Englewood Cliffs: Prentice Hall.
- Rose, O. (1995, Februar). Statistical properties of MPEG video traffic and their impact on traffic modeling in ATM systems. Forschungsbericht, Preprint-Reihe Nr. 101, Universität Würzburg, Institut für Informatik.
- Schlittgen, Rainer und Streitberg, B. (1987). *Zeitreihenanalyse* (2 ed.). München: Oldenbourg.
- Schruben, L. W. and V. J. Cogliano (1987, August). An experimental procedure for simulation response surface identification. *Communications of the ACM* 30(8).
- Spearman, M. L. (1992). Customer service in pull production systems. *Operations Research* 40(5), 879–894.
- Sze, S. M. (1983). *VLSI Technology*. New York, NY: McGraw-Hill.
- van den Berg, J. L. and O. J. Boxma (without year). Sojourn times in feedback queues. Technical Report OS-R8710, Centre for Mathematics and Computer Science, P.O. Box 4079, 1009 AB Amsterdam, The Netherlands.
- Wein, L. M. (1988). Scheduling semiconductor wafer fabrication. *IEEE Transactions on Semiconductor Manufacturing* 1(3), 115–130.