Efficient Scheduling Policies to Reduce Mean and Variance of Cycle–time in Semiconductor Manufacturing Plants[*][†]

Steve C. H. Lu, Deepa Ramaswamy and P. R. Kumar[‡]

Keywords: Control of Manufacturing Plants, Semiconductor Manufacturing, Cycle–time, Discrete–Event Systems

Extended Abstract

Recently, there has arisen a great deal of interest in the problem of controlling or scehduling semiconductor manufacturing plants. Such plants are characterized by a "re–entrant" flow, characteristic of both semiconductor manufacturing as well as thin film lines, that differentiates them in several important ways from both traditional flow shops as well as job shops. Hence machines are shared by all lots requiring the particular processing operation provided by the machine, even though they may be at different stages of their manufacturing life.

Typically, the sum of the processing times of the various manufacturing operations required is of the order of days. However, the manufacturing lead–time, which also includes the time spent by the lots in just waiting for service at machines, is of the order of weeks. This manufacturing lead–time, the time elapsed from the release of a lot into the plant till its emergence as a finished product, is called the "cycle–time" in semiconductor manufacturing parlance. The ratio $\frac{\text{Mean cycle–time}}{\text{Sum of processing times}}$ is called the "actual–to–theoretical ratio," and may range between 3 and 10, depending both on the type of the fabrication line as well as its loading. The typically large value of this ratio implies that lots spend most of their time in the plant simply waiting for service.

It is obviously of great economic importance to reduce the mean cycle–time. For device prototyping, typically involving several design changes, a shorter product development time allows a quicker response to rapidly changing market needs. For production lines, a smaller manufacturing lead–time provides an improved ability to satisfy customer requirements. Both enhance economic competitiveness. Also, for the same level of throughput, a shorter cycle–time results in a smaller work–in–process build up inside the plant. This not only reduces the capital tied up, but also leads to an uncluttered plant floor. Finally, the larger the cycle–time, the greater is the inventory buffer that needs to be maintained at the downstream end of the plant. When product designs become obsolescent, such inventory may lose value.

---

[*]Please address all correspondence to the third author.

[†]The research reported here has been supported in part by the National Science Foundation under Grants Nos. ECS-90-25007 and ECS-92-16487, and in part by the Joint Services Electronics Program under Contract No. N00014-90-J1270.

[‡]Department of Electrical and Computer Engineering, and the Coordinated Science Laboratory, University of Illinois, 1308 W. Main Street, Urbana, IL 61801, USA. Phone number: (217)-333-7476. Fax: (217)-244-1653. Email: prkumar@decision.csl.uiuc.edu.

In semiconductor manufacturing, there is also a technological reason for reducing the cycle-time. The shorter the exposure of wafers to aerial contaminants while waiting for processing, the smaller is the yield loss.

It is also important to reduce the *variance* of the cycle-time. Obviously, a smaller variance of the cycle-time allows a more accurate prediction of production completion time. This facilitates improved downstream coordination of further operations on completed wafers, such as assembly, etc.

To achieve these two objectives, it is necessary to properly control the manufacturing system. There are primarily two ways in which such control is exercised over the plant. First, one can regulate the entry of new lots into the plant. This is done by the *release policy*, which specifies when lots are released into the plant. Clearly, the release policy must satisfy some constraints such as maintaining an average throughput of lots. Second, for lots already in the plant, one has to decide which lot is processed next at each machine as it becomes available. This is done by the machine or lot scheduling policy, which we shall simply call the *scheduling policy.*

Due to the re-entrant nature of the line, there is much competition between lots at different stages of their life for the available machines. The manner in which this contention is resolved has a clear bearing on plant performance measures such as mean and variance of the cycle-time. That is the thesis of this paper. In the plant models studied in this paper, we show that, in comparison to our policies, the best previous policies are worse by 12.5% with respect to the mean queueing time, and worse by about 45% with respect to the standard deviation of the cycle-time, on average.

We now provide a brief account of some previous approaches. The basic difficulty is that there is an insufficient theoretical basis to guide us definitively in the choice of of release or scheduling policies for semiconductor manufacturing plants; see Kumar [1]. In some previous works, one concentrates on dealing with the "bottlenecks," see Harrison [2]. Thus, one attempts to release lots into the plant, and also schedule the lots already in the plant, to prevent starvation of the bottleneck machines, while maintaining the required average rate for lot releases. Such "starvation avoidance" policies have been advanced in Glassey and Resende [3] and Lozinski and Glassey [4]. We refer the reader to Uzsoy, Lee and Martin-Vega [5,6] for an excellent survey of these and other existing approaches to the problems of scheduling semiconductor manufacturing plants.

In this paper we pursue a new approach of attempting to smooth fluctuations in the flows in the network. Also, our method does not restrict attention to just the bottleneck machines in the plant, but attempts to control the behavior of the entire plant. The policies we develop, called *Fluctuation Smoothing Policies* (FS), can dramatically reduce the standard deviation of queueing time. We can also employ such Fluctuation Smoothing policies to reduce the *mean* of the cycle-time. As noted earlier, the net result is that in the plant models studied in this paper, the best previously studied policies are worse by 12.5% with respect to the mean cycle-time, and by 45% with respect to the standard deviation of the cycle-time, on average.

The conclusions regarding the effectiveness of our proposed scheduling policy are based on more than six thousand simulations conducted on two models of plants. The first is a model of a Hewlett-Packard Research and Development Fabrication Line. This line has earlier been modeled and studied by Wein [7], who performed a detailed comparison of 12 scheduling policies and four release policies. Wein's net conclusion was that among the set of scheduling policies considered by him, the mean cycle-time was not very sensitive to the choice of a scheduling policy, but could be significantly reduced by the proper choice of a release policy. For implementation work following up on these recommendations, see Miller [8]. Here, by introducing a new approach, and a new class of scheduling policies, we show that one can obtain a significant additional reduction simultaneously in both the mean and variance of cycle–time.

The second plant considered is an aggregated model of a full scale production line. The data used was not drawn from any existing plant, and should therefore only be regarded as a gross approximation. Nevertheless, since production lines tend to be more heavily loaded and feature much greater throughput than R&D lines, we felt it was useful to see if our results continued to hold for such plants. As we show, we obtained similar simultaneous reductions in both the mean and variance of cycle–times.

Statistical tests have also been performed to validate our conclusions.

The full details may be found in [9].

Acknowledgments The authors are grateful to Sarah Hood of IBM for informative discussions on modeling a production line.

References

[1] P. R. Kumar. Re-entrant lines. To appear in *Queueing Systems: Theory and Applications*, 1993.

[2] J. M. Harrison. Brownian models of heterogeneous customer populations. In W. Fleming and P. L. Lions, editors, *Stochastic Differential Systems, Stochastic Control Theory and Applications, IMA Volumes in Mathematics and its Applications*, chapter 10, pages 147–186. Springer-Verlag, 1988.

[3] C. R. Glassey and M. G. C. Resende. Closed-loop job release control for VLSI circuit manufacturing. *IEEE Transactions on Semiconductor Manufacturing*, 1(1):36–46, February 1988.

[4] C. Lozinski and C. R. Glassey. Bottleneck starvation indicators for shop floor control. *IEEE Transactions on Semiconductor Manufacturing*, 1(4):147–153, November 1988.

[5] R. Uzsoy, C.-Y. Lee, and L. Martin-Vega. A review of production planning and scheduling models in the semiconductor industry – Part I: System characteristics, performance evaluation and production planning. Research Memorandum 92–13, Purdue University, West Lafayette, IN, June 1992. Industrial Engineering.

[6] R. Uzsoy, C.-Y. Lee, and L. Martin-Vega. A review of production planning and scheduling models in the semiconductor industry – Part II: Shop floor control. Research Memorandum 92–14, Purdue University, West Lafayette, IN, June 1992. Industrial Engineering.

[7] L. M. Wein. Scheduling semiconductor wafer fabrication. *IEEE Transactions on Semiconductor Manufacturing*, 1(3):115–130, August 1988.

[8] D. J. Miller. Simulation of a semiconductor manufacturing line. *Communications of the ACM*, 33(10):98–108, 1990.

[9] Steve C. H. Lu and Deepa Ramaswamy and P. R. Kumar. Efficient scheduling policies to reduce mean and variance of cycle–time in semiconductor manufacturing plants. Technical report, University of Illinois, Urbana, IL, 1992.