Model-Based Systems Engineering and the Intel MiniFab Case

Leon F. McGinnis

Professor Emeritus

School of Industrial and Systems Engineering

The Georgia Institute of Technology

Last revision: April 28, 2023

Acknowledgements: This case is based on research conducted in the W. M. Keck Virtual Factory Lab at Georgia Tech and supported by the National Institutes of Science and Technology (grant 70NANB15H234), the W. M. Keck Foundation and the Gwaltney Chair in Manufacturing Systems. It has benefited from the participation of many individual researchers, particularly Dr. Tim Sprock (currently at NIST) and Dr. George Thiers (founder of ModGeno).

Table of Contents

Table of Figures

Forward

Model-Based Systems Engineering (MBSE) is a relatively new phenomenon that is transforming the way complex technical systems are designed. NASA JPL has been a leader in advancing MBSE through its application to the design of deep space missions, and today MBSE is becoming standard practice within DoD, as well as many major corporations, such as Boeing and Lockheed. The deployment of MBSE reduces the ambiguity that plagues document-based systems engineering, enables the integration and automation of a broad range of system analyses, and makes possible much more frequent critical review of system design and development decisions. It is making systems design and development faster, better and cheaper.

With the advent of Industrie 4.0, “the smart factory”, and “the internet of things”, production systems are becoming much more technical and complex. There is every reason to believe that the benefits of applying MBSE to space missions and aircraft programs also can be realized in applying it to the design and development of production systems.

This case study is one of a series intended to illustrate and promote the application of MBSE to production systems. The presentation of the case uses the semantics of discrete event logistics systems (DELS) developed over the course of several sponsored research projects performed in the W. M. Keck Virtual Factory Lab, beginning in 2007 and continuing today.

The work reflected in this case owes a large debt of gratitude to MBSE thought leaders, particularly Sandy Friedenthal, formerly of Lockheed, and the author of a leading textbook on the OMG Systems Modeling Language™ or SysML, and Dr. Chris Paredis, currently BMW Chair in Systems Integration at Clemson University. Any errors and/or omissions in this document are completely the responsibility of the author. Comments or questions may be directed to leon.mcginnis@gmail.com.

Model-Based Systems Engineering and the Intel MiniFab Case

Introduction

Factories that produce integrated circuits (ICs), such as microprocessors or memory chips, are among the world's most expensive production facilities, due to the cost of the key manufacturing resources, and the value of the work in process. In these factories, called "wafer fabs", it is imperative to achieve the highest possible utilization of expensive resources in order to reduce the production costs, and thus, there has been and continues to be great interest in discovering the best design of wafer fabs, including the best ways to manage and control them. The Intel MiniFab (IMF) case describes a hypothetical wafer fab that is greatly simplified compared to real wafer fabs, but nevertheless incorporates some of the key attributes of real wafer fabs that present significant challenges to production control. It has been widely used by researchers since it was first developed by Karl Kempf at Intel in the mid-1990's. Documentation of Kempf's original description of the MiniFab can be downloaded from https://aar.faculty.asu.edu/research/intel/papers/fabspec.html.

The combination of setup, batching, and failure/repair lends significant elements of realism to the Intel MiniFab case. In addition, there are practical constraints on the production and maintenance operators, corresponding to breaks, meetings, and shift start/end. The resulting problems of modeling control and estimating performance are quite challenging despite the very small size of the problem.

In short, while it is a relatively simple and small case, the Intel MiniFab accurately reflects many of the elements of real IC factories that present major challenges for factory design and operation, especially the control of product flow through the IC factory.

Concept of Operation

The IMF transforms raw silicon wafers into product wafers populated with integrated circuits. The only documented input to the IMF is the raw wafers. In reality, a wafer fab also would consume other materials, primarily gases used in wafer processing and water for wafer cleaning between process steps. The concept of operation is for lots of wafers to be moved between process-specific cells or "bays". (An alternative would be to have product-specific "farms" where lots would see all processes required for a specific product type.) The IMF output is the product wafers, which go to downstream factories for further processing. The IMF operates according to a production plan, which calls for product wafers to be produced at a specified rate.

IMF Product

The IMF produces two kinds of product wafers (denoted Pa and Pb) and also a test wafer (denoted TW) that is used to monitor the manufacturing processes. As with real wafer fabs, the product wafers for different product types are indistinguishable to the naked eye—they still look like silicon wafers. What distinguishes them is the patterning created by the sequence of process steps. It is assumed that there is a fixed release rate for each type of lot, and the system never runs out of bare wafers or other necessary materials.

On a weekly basis, the IMF will start, on the average, 51 lots of Pa, 30 lots of Pb, and 3 lots of TW, for a total of 84 lots.

IMF Processes

These three types of products (Pa, Pb, and TW) proceed through the manufacturing system in lots (sets of wafers, all of the same product. In the IMF, there are six processes, called steps, referred to as S1, S2, … S6, and every product follows the same six step sequence.

Each process step involves three elements: load a wafer lot, process the wafer lot, and unload the wafer lot. Each element of each process step is well-understood and has a highly predictable time, in minutes, as shown in the table below.

Step	Load Time	Run Time	Unload Time
S1	20	225	40
S2	15	30	15
S3	10	55	10
S4	15	50	15
S5	20	255	40
S6	10	10	10

Note that the times associated with the steps do not depend on the product wafer type.

IMF Resources

The IMF has three types of machines capable of executing the manufacturing steps described above. For a given type of machine, there may be more than one machine, and the machines are organized by type into cells. For each cell, the numbers of machines of each type, their process capabilities and preventive maintenance (PM) schedules are summarized in the table below.

Cell	Machine Type	Number of Machines	Process Capability	PM Time (min)
C1	M1	2	S1, S5	75/day
C2	M3	1	S3, S6	30/shift
C3	M2	2	S2, S4	120/shift

As a consequence of the process capabilities of each machine type, each lot will visit each type of machine exactly twice. This last aspect is referred to as “reentrant flow” and has been a major source of difficulty in planning and scheduling IC factory operations.

The production resources have other important properties, related to batching, failure/repair and setup.

M1 machines are described as “diffusion-like” resource, where batches of lots are processed together. They require no setup, never fail, and can accept batches with up to three wafer lots. When an M1 machine is executing S1, any combination of lots of Pa, Pb and TW is allowed, provided that a batch may contain at most one TW lot. When an M1 machine is executing S5, a batch cannot contain both Pa and

Pb lots, although it may mix TW with either Pa or Pb. Mixing of lots waiting for S1 with lots waiting for S5 is strictly forbidden.

M2 machines perform implant, require no setup, have a batch size of one lot, but are subject to failures requiring emergency repairs. There is some historical data for this machine which allows the failure/repair process to be described as follows. A failure occurs randomly once each half week but not within two shifts of the most recently completed repair. There are 168 hours per week, a half week is 84 hours, and two shifts is 24 hours, so a failure occurs between 24 and 84 hours after the previously completed repair. The repair, once started, takes between 6 and 8 hours. In order to keep the EM window within half a week with a worst case repair, this means a failure will occur between 24 and 76 hours after the previously completed repair (50 + 26 + 8 = 84, 50 - 26 = 24). When a failure occurs on an M2 machine, the lot being processed is "lost", i.e., it becomes scrap.

The M3 machine is described as a lithography-like resource that has specific configuration requirements depending on the product and step. If changing between products at the same step, the setup time is 5 minutes. When changing between steps for the same product, the setup time is 10 minutes. When changing both the product and the step, the setup time is 12 minutes. This machine has a batch size of one lot, and never fails.

The IMF has several other resources required to support production. Each cell contains a *stocker*, a piece of equipment that stores production lots both before and after processing on a machine. The stocker for C1 has a capacity of 18 lots, while the stockers for C2 and C3 each have a capacity of 12 lots. In addition, there is a "starts" stocker from which lots are released for production and an "outs" stocker where lots are delivered after production. The capacity of both is considered unlimited.

There is an automated material transport system that has a single transporter moving lots between the stockers. The transport time between any two adjacent stockers is 4 minutes. It is assumed that the arrangement of stockers is linear in the order "starts", C1, C2, C3, "outs". The transfer time between a stocker and the transporter is 1 minute for either loading or unloading a stocker.

There are two types of employees in the IMF. A production operator, or PO, is required to perform the setup (if required), load and unload elements of the process in each cell. There are two POs; PO1 can service M1 and M2, while PO2 can service M2 and M3. For M2, it is not required that the same PO perform both the load and unload. The POs move between cells with a move time of 1 minute between adjacent cells. POs are available for 540 minutes per shift, with two 60 minute breaks and one 60 minute meeting or training session each shift (540+2*60+60 = 720). The two POs do not have to synchronize their breaks/meetings in any way.

A single maintenance technician, or MT, is available for both preventive maintenance and emergency repairs. There is a well-defined window for the performance of the PM. A shift-based PM can be performed any time during the shift, except that a new PM cannot start within 6 hours of the most recently completed PM, cannot interrupt a process step to perform a PM and should be completed before the shift ends. Similarly, a daily PM can be performed any time during the shift, except that a new PM cannot start within 12 hours of the most recently completed PM, cannot interrupt a process step, and should be completed before the end of the day. If a PM is not completed within its required window, then no new lots may be processed at that machine until a PM is completed. The MT moves between cells with a move time of 1 minute between adjacent cells. The MT is available 600 minutes per shift, with two 45 minute breaks and one 30 minute meeting/training session (600+2*45+30 = 720). The MT does not have to coordinate or synchronize activities with any other employee, except that load/unload, PM and emergency repair cannot overlap.

Facility

The organization of the IMF resources is illustrated in Figure 1 below. In a real wafer fab, there might be thousands of machines, but they still would be organized in a similar if not exactly the same manner, i.e., groupings of similar machines, with groups of machines served by stockers. What does not scale up from the IMF to a real wafer fab nearly so well is the transporter. In a real wafer fab, the transporter system might be comprised of multiple subsystems, each with a complex network of guidepaths and many individual transporter vehicles in each subsystem. Issues of traffic management become critical in real wafer fabs, because the transport vehicles can create traffic jams and cause significant delays. In the IMF, these concerns are assumed away because there is a single transport vehicle.

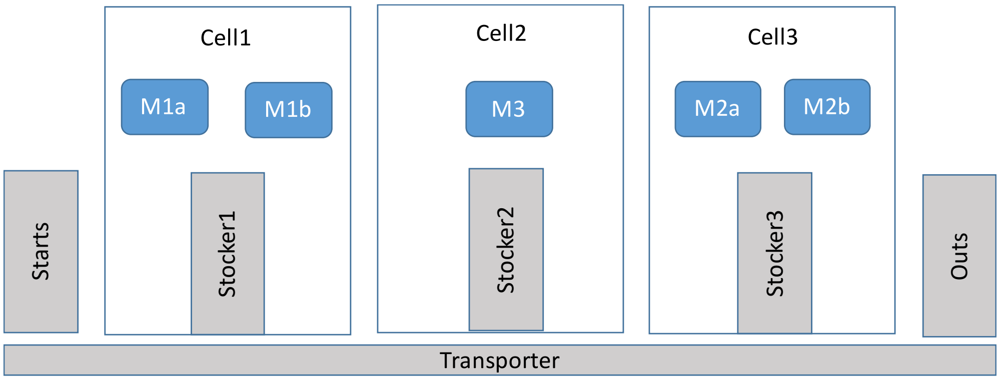

Figure 1 Facility Configuration for IMF

Operational Control

There are several distinct operational control decisions in the IMF. For the POs, the decisions are when to take breaks or meetings, and what setup/load/unload task should be done next. Note that these decisions can result in the PO moving from one cell to another, incurring a travel time. For the MT, the decisions are when to take breaks or meetings, and what maintenance task (PM or emergency repair) to do next. As with the PO, the selection of the “next” maintenance task can involve a travel time. For the transporter the decision is which lot movement to execute next; the sequence of lot moves also may incur “deadhead” or unloaded travel by the transporter.

In practice, at the beginning of a shift, a set of lots will be released to the starts stocker, i.e., they are available to be moved to the C1 stocker, provided there is space available. In the IMF, there is a weekly average release rate for each product. However, the exact composition of the set of lots released at the start of any shift is an operational control decision. Again, in practice the release of lots is not independent of the outs, and there may be considerations of customer order quantities, batching for transportation, or other issues that impact the release decisions. A major issue is the acceptable level of variability in starts/outs from the target averages.

There is uncertainty in the IMF, due to the occurrences of failures and the time required to effect their repairs. Uncertainty also results from the "on the fly" decisions about movement for the transporter, PO and MT. In reality, there also would be uncertainty associated with PM and EM repairs, and perhaps the load/unload processes. The technical processes of moving the transporter or executing the process steps likely would have much less uncertainty.

Some constraints are imposed on operational control. The batching constraints for M1 are the result of technical considerations and were described in the resources section. There are constraints that are not technical but imposed for performance reasons. One such constraint is that a TW lot may not ever go through the same machine twice, except for S3 and S6, where is only a single machine is available. This forced alternating between machines in C1 and C3 is to improve the coverage of the monitoring process.

An important consideration in operational control is the metrics used to assess the quality of the decisions rendered. In the case of the IMF, some metrics seem obvious. First of all, it is desirable for there not to be a large number of released but not yet moved lots in the starts buffer, because that would indicate that the IMF is not keeping pace with the target release rates. This concern could be viewed also from the perspective of the outs buffer, i.e., to maximize the rate of completed lots, while respecting the product mix implied by the release targets. The average throughput time is limited by the fact that work-in-process is limited by the capacity of the stockers, but as the economists say, *ceteris paribus,* less work-in-process, shorter cycle times and greater throughput all are desirable. Predictability is improved if the variance of outs rates and variance of cycle time are small. Some managers like to see high resource utilizations, although increasing utilization without increasing throughput doesn't really seem like a good idea.

System Summary

The IMF case is relatively short—less than five pages of text in total. What has been offered in this IMF presentation is information describing the IMF *plant*, i.e., the products, the process steps required by the products, the physical resources and their capabilities to execute processes, and the organization of the resources, along with a description of the kinds of operational decisions that must be made and the criteria that might drive those decisions. From this, we can understand the flows in the IMF, and the decisions about the flows. What is not described is *how those decisions are made*. In particular, there is no description of a decision architecture.

Thus, in order to develop formal IMF system models, or IMF system simulations, a design for the IMF control system must be developed. This design will require making assumptions about how the IMF should be controlled. As an example, consider the role of the two POs and how their work is managed. One possibility is that each PO is an independent actor ("agent" in the parlance of agent-based modeling), and makes decisions about what to do next based strictly on the list of available tasks, and the PO's current state with regard to location and mandatory breaks. Perhaps the PO carries a tablet that is updated with the status of the machines and maintains a list of load/unload tasks that are pending. This approach might require a mechanism to mediate between the two POs if both select the same "next task".

An alternative approach is to treat the two POs like machines whose tasks are assigned by a central controller. That controller would tell the POs when to move, and where to go, as well as what to do. Similar alternatives apply to the MT and to the transporter. It is worth noting that one function of the

POs is very similar to the function of the transporter, i.e., fundamentally they move lots between the stockers and the machines.

In order to apply MBSE principles and methods to the IMF, not only must the plant be described completely, but also the control architecture must be developed, and specific control decision methods defined.