Perspectives on Semiconductor Factories (And Some Simple Truths About Manufacturing)

GEORGE W HORN, Middlesex Industries

THE SUM OF INDIVIDUAL EVENTS DEFINES the fab. Innumerable number of parameters in each of scheduling, dispatch, and physical processes, controlled by APC (Advanced Process Control) systems. [1] Conflicts and interaction arise between the individual variables and then there is the conflict between cycle time and throughput. Management of manufacturing works with the acceptance of this complexity. And now, on the one hand there is big data handling, AI/ML (Artificial Intelligence and Machine Learning) [2, 3] which, on a fab wide scale may be like weather forecasting with supercomputers (deterministic dependence on practically infinite numbers of parameters) advancing towards dominance in managing the fab. And, on the other hand, there are the fundamentals in direct manufacturing: the physical capabilities of process equipment and material handling systems quietly making advances. Even though, it is evident that self-optimizing AI, or ML and APC have their limits set by equipment competence.

First, the manufacturing of transistors on silicon substrates became a multi-layered process on costly equipment sets. Then these costs demanded the reentrant use of that equipment set. Replacing the linear sequencing of value adds through a line of machines. Ideally, synchronizing the process steps would yield the maximum throughput and lowest cycle time, erasing the conflict between them. To achieve this in the linear case is relatively easy by balancing the line, while highlighting the very important aspect of a balanced line: the reduction of inter process idling of WIP. Doing the same balancing with reentrant manufacturing greatly multiplies the difficulties (makes it impossible). But we do want the advantages of reduced inter process WIP.

In short, equipment cost forced the reentrant manufacturing, which then prohibits balancing and thus synchronization. Still, we would like to emulate linear manufacturing. Accordingly, what we seek is some fundamental truth to help in achieving greater synchronization of process steps in reentrant manufacturing.

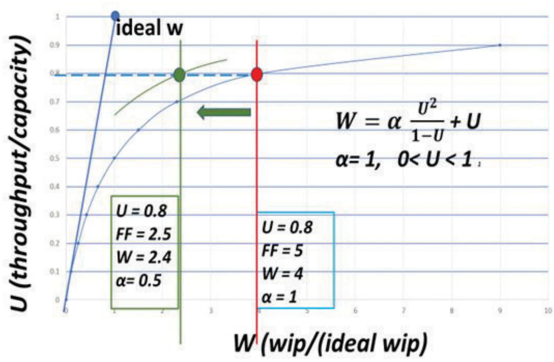

Figure 1. WIP vs. throughput of a fab at 80% capacity. Reducing WIP of a fab a new family of curves can be generated. Example: exiting the curve with $\alpha = 1$ will result in a new OC with α=0.5, requiring a WIP content of only 2.4 times the ideal WIP.

Synchronization is process coupling

In practice, we never sought synchronization from the start. Maximizing factory throughput was always equated to maximizing tool utilization. And with that we proceeded to fill the inter process gaps with WIP. And so reduced chances of a tool becoming idle.

Fundamentally, in a synchronized factory process tools are tightly linked (coupled). When the product exits a process, it immediately enters the next process. There is no product idling between tools. So, an indicator of the degree of synchronization is the amount of product idling between tools. And in today's fabs such idle products are 3-4 times more in quantity than products in process at the tools. But look again! This logic works on a curve of exponentially diminishing returns (fig 1). In other words, as we attempt to increase utilization (for the sake of factory throughput), for each incremental increase in throughput we need to add exponentially more idle products into the process gaps. Is this paradigm (cycle time vs. throughput) in manufacturing unbreakable?

Instead, the goal should be to aim for a minimal amount of WIP in manufacturing IC-s via higher synchronization of the processes. Eventually leading to sacrificing some individual tool utilizations (over capacity) and tool cycle times to achieve higher overall factory throughput and better factory cycle times. Switching the emphasis of the performance parameters from tool utilization to tool synchronization. Along with their control parameters. Today's many dispatch algorithms aim at optimizing downstream tool selection (balancing) to reduce overall cycle time by reducing idle WIP time in the gaps. Not advertised is the fact that the use of such tools achieves a higher degree of process synchronization. Therefore, the finding of control parameters for synchronization should be our goal. And the tool for such is offered by methodologies of better process coupling. How do we achieve better process coupling in the reentrant manufacturing, and so aim to emulate linear manufacturing even in the reentrant factory? For this, we need to focus on material handling equipment and inter process logistics.

Moulding the present by casting in the past

Process coupling is significantly affected in the domain of AMHS (Automated material Handling Systems). AMHS executes a logistics function on an infrastructure of inter process highway networks. It is fundamental in all manufacturing. The logistics executed on AMHS determines manufacturing efficiency (throughput & cycle time). Today's 300 mm fabs employ a highway structure of overhead tracks on which hoist vehicles move wafer lot carriers one by one. This choice limits the achievable with logistics due to its process variance multiplying. And variance is the throttle of fab throughput. It enhances process decoupling and makes the AMHS irrelevant to fab efficiency. This is a big loss. But how did we get here?

The 200 mm wafer format manufacturing was primarily developed by IBM. The investment was huge and not to be assumed again by a single IC maker. Thus, conversion to the 300 mm wafer format was charged to equipment makers, as guided by industry-government consortiums. Ultimately evolving into the non-government supported I300I and the Japanese Selete, formally independent. Global IC makers have then agreed to a unified and standardized approach to intra fab logistics. This agreement of uniformity cast the die for work in process logistics in favor of equipment concepts available at conglomerates of material handling companies. No one else could handle the global demand for volume forecasted for such prescribed equipment at that time. The winners were Muratec, Daifuku, and Shinko. These companies then applied their industrial overhead-hoist systems to semiconductor intra fab logistics. Furthermore, the development of these was extended outside of an overhead hoist's natural realm of competence (pick up and place within short distances), to a pickup and move the work over long distances, one by one. This is our 300 mm AMHS today.

The essential step is to improve process coupling

A tighter coupling between processes increases synchronization. And that in turn results in shorter cycle times. In tightly coupled processes cycle time and throughput become noncompeting performance metrics. What is the degree to which this could be achieved in today's fabs? Ideally, a giant cluster tool as a fab? Clearly that goal is too ambitious. But the logistics executed on the AMHS highway networks can significantly reduce the conflict between cycle time and throughput.

An example: in current fabs a finished wafer lot waits for a vehicle (or person) to be moved downstream to the next process step. Thus, a fab wide accumulation of wafer lots occurs at tool outputs. When the wafer lot finally is picked up and transported it may arrive to an open tool for immediate processing. That would be good. But if the fab operator does not want open tools waiting for wafer lots to be processed, then he will want to also to accumulate wafer lots waiting to be processed at tool inputs. As a result, a total accumulation of WIP in the process gaps occurs. The tools are decoupled due to this accumulation, and this decoupling is a fundamental character of the single vehicle move AMHS logistic. And it gets to be so sometimes that offline storage systems (stockers) are needed to hold the accumulated and idling wafer lots. If, on the other hand, we execute a logistic on the AMHS highways that eliminates the wait time of an emerging wafer lot, then at the pickup end of the move the accumulation is eliminated, and that in return also reduces the need for accumulation at the arrival end of the lot. The tools are now more tightly coupled. This significant achievement can be had if the running vehicle logistic on the AMHS highways are replaced by open flow conveyor logistics (input to transport on conveyors is always open). The perspective of this solution is the approximation of cluster tool principles by closer tying some bays of the fab.

In practice, implementation of this solution proves to be valuable between critical bays of existing fabs. Leaving existing vehicle systems in place but reprogramming them to primarily interface the conveyor and tool ports. This simple addition of conveyors to existing OHT systems achieves the shorter cycle times due to tighter coupling of tools. SD

GEORGE W HORN is engaged in developing factory logistics based on Digital Conveyors used in Hybrid AMHS, at Middlesex Industries. He was a participant in the standardization drive of SEMI during the 300 mm conversion years, and developer of the conveyor runner of FOUP-s and FOSB-s, within the Carriers and Physical Interfaces task force.

REFERENCES

1. 13th Annual IEEE/SEMI Advanced Semiconductor Manufacturing Conference. ASMC 2002 (Cat. No.02CH37259)
2. Autonomous order dispatching in the semiconductor industry using reinforcement learning. Andreas Kuhnle, Nicole Röhrig, Gisela Lanza
3. AI in Semiconductor Industry. River Publishers, eBook, Chapter 2. Cristina De Luca at all, Infineon Technologies AG