Proceedings of the 2020 Winter Simulation Conference
K.-H. Bae, B. Feng, S. Kim, S. Lazarova-Molnar, Z. Zheng, T. Roeder, and R. Thiesing, eds.

CHALLENGES ASSOCIATED WITH REALIZATION OF LOT LEVEL FAB OUT FORECAST IN A GIGA WAFER FABRICATION PLANT

Georg Seidel

Infineon Technologies Austria AG
Siemensstraße 2
Villach 9500, AUSTRIA

Ching Foong Lee
Aik Ying Tang

Infineon Technologies (Kulim) Sdn Bhd
Jalan Hi-Tech 7
Industrial Zone Phase II, Hi-Tech Park
Kulim, Kedah 09000, MALAYSIA

Soo Leen Low
Boon Ping Gan

D-SIMLAB Technologies Pte Ltd
8 Jurong Town Hall Road #23-05 JTC Summit
Singapore, 609434, SINGAPORE

Wolfgang Scholl

Infineon Technologies Dresden GmbH
Königsbrücker Strasse 180
Dresden, 01099, GERMANY

ABSTRACT

In the semiconductor industry a reliable delivery forecast is helpful to optimize demand planning. Very often cycle time estimations for frontend, backend production, testing and transits are used to predict delivery times on product level and to determine when products have to be started to fulfill customer demands on time. Frontend production usually consumes a big portion of the cycle time of a product. Therefore a reliable cycle time estimation for a frontend production is crucial for the accuracy of the overall cycle time prediction. We compare two different methods to predict cycle times and delivery forecasts on product and lot level for a frontend production: a Big Data approach, where historical data is analyzed to predict future behavior, and a fab simulation model.

1 INTRODUCTION

Demand planning is crucial for economic success. While this is true for any industry, maybe it is even more true for industries where the production process is complex and expensive. The semiconductor industry has a very complex production flow, consisting of frontend, backend, testing and storage facilities. Very often the facilities are not only logically but also physically distributed around the world. Apart from that increasing number of products, requiring different production steps, can be ordered by customers. A lot of challenging tasks can be derived from that. Accepting or rejecting new customer orders, decisions when and at which facility to start products, confirming delivery dates or communicating expected delays, creation of stock piles along the supply chain are only some examples of tasks within the demand planning process. A reliable delivery forecast of products within the supply chain is helpful to make some of these tasks at least easier (Nyhuis and Filho 2002; Geng and Jian 2009; Fowler et al. 2015; Seidel et al. 2017; Wang and Zhang 2016; Moyne and Iskandar 2017; Wang et al. 2018).

At Infineon Technologies AG, a German semiconductor company, a simplistic approach of using the estimated cycle times to calculate the expected delivery dates on product level is typically used. The estimations are derived by analyzing historical cycle times. Demand Planners need to make sure that confirmed customer orders can be completed in time. They use estimated cycle times to determine when to start products. Very often they add a time buffer to an estimated cycle time for a product and/or create stock piles to ensure that due dates can be met. However, even that cannot always ensure that demands are met in time because boundary conditions can change, e.g. a frontend production can be overloaded, therefore the cycle time increases and predicted cycle times can no longer be met. A cycle time estimation calculated by using historical data will show the increase only after some time. A cycle time prediction by using a simulation model of a frontend production could help to mitigate this problem.

We decided to use the existing simulation models from different Infineon frontend production sites to predict product cycle times and lot delivery dates. We compared the simulation results with a Big Data approach and later on with the reality.

In Section 2 we provide some more details why a forecast could generate a benefit. Section 3 describes the simulation and Big Data approach and explains how to compare results. In Section 4 we present the actual results and Section 5 gives a conclusion and an outlook of possible further studies.

2 BENEFIT OF A FORECAST

Demand planning at a frontend production site requires often manual effort, especially in situations where the demand is higher than the supply. Urgently needed products are sometimes tracked on lot level. Demand planners and line control experts are busy tracking down and prioritizing lots to ensure on time delivery or at least minimizing expected delays. In recent years this task became more and more complex and time consuming because the number of products increased and often there was a shortage for many products simultaneously. Meanwhile, time buffers and stock piles were reduced because production costs can be optimized by reducing WIP. This more often results in situations where products cannot meet the demand due dates. For high volume products, lot priority lists change daily because lots of the same product can overtake each other and therefore lot demand assignments change overnight. Low volume products have to be prioritized after line incidents, e.g. a lot scrap or line performance issues. But updating these lists and changing priorities daily creates disturbances in the line, not to mention the psychological effect on people chasing the lots. Additionally it is well known that a high share of priority lots can cause line performance losses.

A reliable lot and product delivery forecast can help to reduce the prioritized lot count, the daily changes and therefore also manual effort from planners and line experts. A decisive question is how good a forecast can predict fab out dates on a lot level or predict fab out on a product level. Only if we know the answer to these questions we can try to estimate possible benefits.

3 APPROACHES FOR LOT LEVEL FORECAST

3.1 Discrete Event Simulation

Discrete event simulation has widely been used in the semiconductor industry for operational use cases such as early warning on forecasted WIP (work-in-progress) piling at production areas, allocation of operator resources based on forecasted incoming WIPs, synchronization of preventive maintenance with material flow, dynamic dedication adjustment to minimize non-value adding setup switches, and many more (Bagchi et. al. 2008, Gan et. al. 2004, Scholl et. al. 2011, Scholl et. al. 2012, Seidel et. al. 2017, Scholl et. al. 2018, and Seidel et. al. 2018). The aim of the production control paradigm has been shifted from a purely reactive approach to a proactive approach, avoiding (or at least minimizing the impact of) issues instead of fixing them only when the problems occur. Ultimately, such a proactive operations management approach would help to improve the linearity of the production lines, which in turn reduces cycle time variability and thus enhances predictability of the fab performance.

The use cases so far has been focusing primarily on aggregated KPIs (key performance indicators), which could achieve high accuracy. Based on our experiences, the forecast accuracy (Seidel et. al. 2017 and Mosinski et. al. 2017) for fab level KPIs such as wafer out, WIPs, moves, and cycle time or dynamic flow factor, could go as high as 95%. The forecast accuracy for product level KPIs stay at approximately

95% for high volume products, and drops to around 70% for very low volume products (products with less than ten lots throughout the 8 weeks forecast period). The high forecast accuracy could be achieved with the precondition of good input data to the simulation model. Table 1 and Table 2 below give a summary overview of the modelling elements and the modelling fidelity required to achieve the forecast accuracy as described. The simulation model was built on a commercial simulation engine, D-SIMCON Forecaster (D-SIMLAB Technologies 2020) that provides all the essential wafer fab modelling elements.

Table 1: Modelling Elements and Considerations for High Accuracy Forecast Accuracy.

Modelling Element	**Description**
Work-In-Progress (WIP)	The model is initialized with WIP in the production line, with all the lots and their associated information being captured. The essential lot information are its current step, current state: in queue, in process (current equipment and remaining processing time is required), in rework, or on-hold (estimated hold release time is required), priority, start and due date.
Initial Equipment Down	All equipment that are in down or non-productive state are initialized as down before simulation starts. An estimation of when the equipment is coming back online is required. This information is obtained from either historical data (average duration for the corresponding down type) or provided by the maintenance department.
Wafer Start Plan	A wafer start plan up to the lot level is required. Typically wafer fabs do not have lot level wafer start plan beyond a week. To address this constraint, a product level weekly volume start is obtained from the planning department, and a lot level wafer start plan is created. An algorithm of batching lots of the same product to start to enhance batching efficiency at furnaces is used for realistic wafer start plan generation.
Process Flows	All process flows required by the WIP and wafer start production lots are considered in the model. We do not choose representative process flows as we need to ensure lots are following the exact path that they will run in the reality.
Rework	Rework is modelled as a random event, where rework rate is derived from historical data for all production steps that could trigger a rework process.
Hold	Hold is modelled as a random event, where hold rate and hold duration distribution are derived from historical data for all production steps that could trigger lot hold.
Split-Merge	Some equipment type such as Chemical-Mechanical-Polishing (CMP) and Lithography require pilot runs from time to time. This is modelled with the split-merge function, where split rate is calculated from historical data.

In this paper, we would like to evaluate the accuracy that we could achieve in a lot-level fab out forecast using the same simulation model that was built for the above-discussed operational use cases. We used the same simulation model that was used to support the use case of development lot level journey forecast (Scholl et. al. 2016), and managed to achieve high accuracy as development lot was moved through the fab with higher level of priority than normal lots. It is easier to forecast because these higher priority development lots always overtake normal production lots. Dispatching will thus have much less impact on the lot journey through the fab. Extending the lot level forecast to all production lots poses a very different challenge even though the focus of the use case is only at the fab out date and not the day at which the lot will be arriving at a particular production step. Choice of lot selection at each dispatching decision could change the fab out date, and random events of hold/rework could significantly influence the cycle time of the lot. Our end users (demand fulfillment planners), however, do not require a forecast accuracy of up to daily granularity. A lot level fab out forecast for weekly time buckets will already be sufficient for their purposes. So the question is whether the simulation model can enable such a use case, or whether a combination with a Big Data approach will be a better solution.

Table 2: Modelling Elements and Considerations for High Accuracy Forecast.

Modelling Element	**Description**
Equipment	All equipment in the production line are considered. Each equipment is mapped based on its specific behavior such as: single lot, single wafer, batch, or cluster.
Dedication	Dedication is modeled at recipe and product-recipe combinations, depending on equipment type. Long term inhibits are also considered in the model to ensure constraint in production line are portrayed accurately in the simulation model.
Process Time and Throughput	Data for recipe or recipe-product based process time and throughput for each equipment is gathered. Process time is defined as the time duration that lots/batches are spending in the equipment, while throughput is defined as the rate at which lots/batches are processed by the tools. Cascading of lots/batches are thus modelled when the throughput is higher than the process time. Limping effect (losing process speed) of chamber down is also modelled.
Setup Switching	Setup switching is modelled at some of the relevant equipment such as implantation. We consider the time required to switch from one recipe class to another. This overhead is important to be modelled as it reduces the tool capacity.
Equipment Down	Equipment down is modelled as a random event, where the mean time to failure and mean time to repair distribution are derived from historical data.
Reticle	Reticles are modelled as an additional resource required before lots can start processing at lithography equipment. This is essential because lithography equipment are typically the key bottleneck of the production line, and reticle availability could alter the selection of lots for processing. discussed above.
Dispatch Rules	Only global dispatch rules are considered in the model, such as lot priority, queue time priority, operation due date, maximum wait time and same setup. Some local dispatch rules such as prefer fast equipment were also considered.
Queue Time Constraint	Typically queue time constraints are controlled with KANBAN based dispatch rules. It is thus essential to construct a simulation model with such consideration as lots could be held back and not moving to the next step due to unavailability of KANBAN even though equipment capacity is available.

3.2 Big Data

Big Data analytics methods have been developed and applied in semiconductor manufacturing operations in recent years on use cases like fault detection (Chien et al. 2014; Chen et al. 2017), predictive maintenance (Lee at al. 2017) and forecasting cycle time (Wang and Zhang 2016). The evolution of Big Data has been developed with data peculiarities, in terms of volume, velocity, variety, veracity and value (Wang and Zhang 2016; Moyne and Iskandar 2017; Lee et al. 2017).With the massive amount of data in semiconductor manufacturing, data is stored in staging areas of data analytic tools before data analysis. Different types or combination of data analysis methods such as data mining, predictive analysis, machine learning or deep learning are provided according to the corresponding data analytic tool packages.

In this paper, a Hadoop cluster was chosen as the data analysis tool for our study. Basis for the Big Data approach on lot level forecast is a data table which holds all transactions lots do see while they are processed in frontend facilities. The number of rows in such table, where each row is a transaction, is more than 200.000.000 for one year of data from several production facilities. Such a data table is updated daily from local data warehouses. It is held by a Big Data data-lake and investigated by a Hadoop cluster. Data analytics on such table is done by Spark SQL.

Different lots of different products see different sequences of operations. Depending on the complexity of the product there are typically between 100 to more than 1000 operations for a lot in a frontend facility.

The operation identification number is unique in the sequence of the operations for one lot. The main transactions are the moves-out from the previous operation and the moves-in to the current operation.

Therefore for each lot, that has already left the facility the time it took from each operation to leaving the facility can be calculated. This is the remaining cycle time of the lot at the time when it was at that operation. To do so we determine the date-time when the lot left the facility of interest and subtract the date-time when the lot was at each single operation.

For a given product all the lots of such product can be used to make up a distribution of remaining cycle times for each operation. In *Spark SQL* it is straight forward to aggregate such distributions grouped by any operation on any product: we use *percentile_approx()* on a list of percentage values. This returns a vector holding the distribution of remaining cycle time. These distributions then can be used to forecast the time, when the facility will deliver a lot as a function of at which operation such lot currently is.

To some extent also the actual recent overall speed performance of the facility can be taken into account, to become more independent from the overall facility loading situation.

The accuracy of such predictions mainly depends on the level at which the aggregation of the distributions is calculated. We typically obtained a relative error of 8% for remaining lot cycle time over the complete life cycle of a lot.

3.3 Qualitative Comparison of Approaches

Both approaches have their strengths and weaknesses. The Big Data approach is easy to use and after the initial setup phase the effort to maintain it is small. However, there are limitations to overcome. The impact of changing boundary conditions on product cycle time, caused for example by fab ramping, a changing product mix, new incoming equipment or a change in tool dedications will not be reflected or only after some time, because historical data will be not related to this new environment. For new products no historical data will be available at all.

While there are some ideas to mitigate this risk, for example by using historical data from a similar product like the new one and then using speed factors to incorporate actual fab performance, this will be a tough challenge to solve. Furthermore, such mitigation strategies will increase the maintenance effort of the Big Data approach considerably.

In turn, maintaining a reliable simulation model typically requires a lot of effort. Data must be accurate, and continuous model validation is required. A bottleneck in the simulation caused by wrong input data that does not tally with reality can jeopardize all simulation results. On the other hand, a reliable simulation model can forecast product cycle time changes over time due to changing environment. New products' cycle times can be predicted even if no historical data is available.

Forecast accuracy comparison is done in section 4.3.

4 EXPERIMENTAL RESULTS

4.1 Forecast Quality Measurement for Simulation

To evaluate the usability of the simulation-based lot level fab out forecast, we need to measure the achievable forecast quality. This was done by choosing a time period where the fab was fully loaded in the past, running the simulation forecast to obtain each lot fab out week, and then comparing with the actual fab out week. Weekly buckets are used because demand planning at Infineon is done weekly. Therefore it is important to know how many wafer/lots of a particular product will be delivered within a certain week. The forecast quality was measured in two ways: (i) lot level, and (ii) product, as respectively illustrated in the equations below:

$$lot\ level\ forecast\ quality\ for\ evaluation\ period\ (n\ weeks) = \frac{number\ of\ lots\ with\ same\ fabout\ week\ (simu\ and\ reality)}{number\ of\ lots\ with\ fab\ out\ for\ evaluation\ period\ (reality)}$$

$$product\ forecast\ quality\ for\ evaluation\ period\ (n\ weeks) = \frac{\sum_{w=1}^{n} \frac{number\ of\ lots\ with\ fab\ out\ in\ simulation\ and\ reality\ for\ week\ w\ and\ product\ p}{number\ of\ lots\ with\ fab\ out\ in\ reality\ for\ week\ w\ and\ product\ p}}{n}$$

The lot level forecast quality is measuring the accuracy of forecasting whether a lot reaches fab out in both simulation and reality. A high accuracy at the lot level forecast is the most stringent measurement, and could provide us an insight into the usability/applicability of providing such forecast to the end users.

The product forecast quality is measuring the accuracy of forecasting the fab out volume for each product, and aggregated across weeks with a simple average. We did not carry out any weighted average calculation across weeks as the product mix was stable for the chosen time period of the study. The product forecast is an alternative forecast granularity that we were exploring to provide to the end user because in a typical business process our end users are matching the lots to a demand order upon fab out. The demand lot assignment is usually not fixed and can be changed over time. Some flexibility of reallocation of fab out lots to orders is still possible. This is an important point because it is possible that during a simulation run the random events such as rework and hold could extend some specific lots fab out time, or vice versa. This is also one of the main reasons why high accuracy at lot level forecast is not easy to achieve.

4.2 Forecast Quality Comparison: Actual Wafer Start vs Plan Wafer Start

The accuracy of a simulation forecast is highly influenced by the approach being taken to model various aspects of the wafer fab. Equally important, the simulation model needs to be fed with good quality input data. For a use case such as the fab out forecast, an accurate wafer start plan is thus crucial. In fact, obtaining a high accuracy wafer start plan is posing an additional challenge to achieve good forecast quality. Typically wafer fabs only have high accuracy wafer start plan for one to two weeks. Any weeks beyond that are still volatile and subject to further changes. Thus, besides comparing the achievable forecast quality at lot and product level, we extended the experimental study to compare scenarios with and without high accuracy wafer start plan. This is possible as we chose a historical time period for which we already knew which lots had been started, hence we also knew the wafer start plan available at the beginning of the evaluation period. The forecast quality data was collected for a time period of 8 weeks, as this is the required forecast time horizon by the end users.

Table 3 shows the lot level forecast quality achieved with simulation using the actual and plan wafer start. The actual wafer start forecast provided a 4% better forecast quality as compared to the plan wafer start. This forecast percentage is not very encouraging as it seems to indicate that we cannot use the simulation approach for lot level forecast use case. We thus added on another dimension to the evaluation of the applicability of the approach, i.e. to measure how much was the absolute deviation (in days) between the forecast and actual fab out day.

Table 4 shows the summary of this comparison. We observed that for 10.0% and 9.2% of the cases the fab out day in simulation and reality was exactly the same, using the actual and plan wafer start respectively. For 82.1% and 78.0% of the lots we saw a maximum gap of 7 days between simulation and reality. This means that the forecast could still be useful (but not ideal) for the end users because for approximately 80% of the cases they can be sure that lots will reach fab out within plus/minus 7 days of the forecasted value.

Table 3: Actual vs Plan Wafer Start Lot Level Forecast Quality.

Actual Wafer Start	**Plan Wafer Start**
50%	46%

Table 4: Actual vs Plan Wafer Start Forecasted Day Gap.

Forecasted Days Gap	Actual Wafer Start	Plan Wafer Start
=0	10.0%	9.2%
<=1	25.9%	25.1%
<=2	40.4%	38.1%
<=3	52.9%	49.3%
<=4	62.9%	59.1%
<=5	70.8%	66.9%
<=6	77.7%	73.1%
<=7	82.1%	78.0%

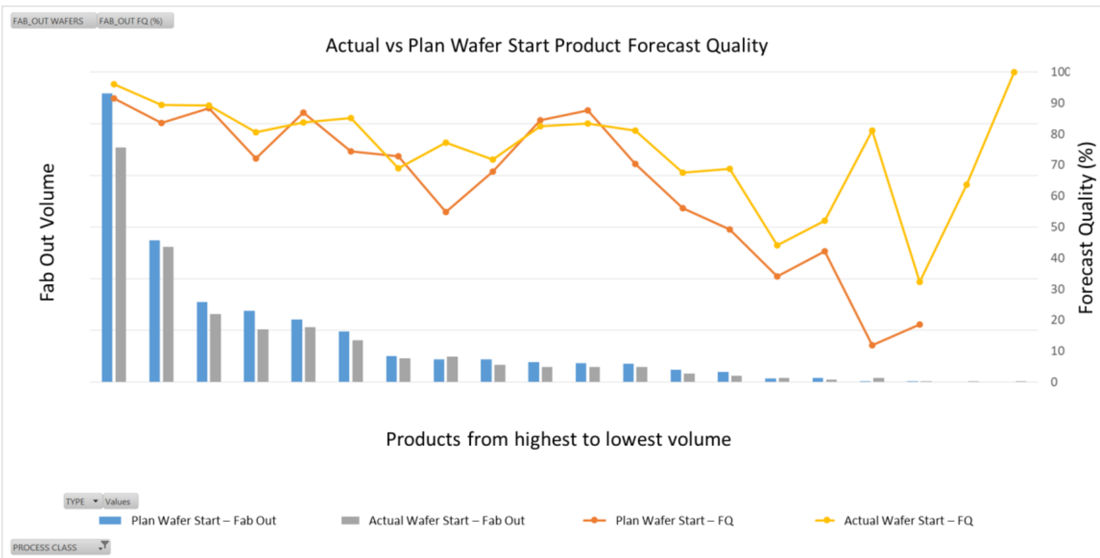

Figure 1: Actual vs Plan Wafer Start Forecast Quality Comparison.

Figure 1 shows the forecast quality achieved for product groups, sorted from highest to lowest volume. We presented the forecast quality at product group level instead of the product level because the number of products involved are more than 200 and it is not feasible to show them in a single chart. The product group forecast quality was calculated as a weighted average (by the product volume of the week) of the product forecast quality. As observed, high volume product groups (the first two product groups) achieved a forecast quality of above 90% for the runs with actual wafer start, and above 85% was achieved for the runs with plan wafer start. The forecast quality is steady between 78% and 75% for the next 11 product groups for runs with actual and plan wafer start respectively. Some of the extremely low volume products have shown very low forecast quality for plan wafer start. This is due to the fact that planned lot starts for low volume products are typically prone to high error. The observation for the product (group) level forecast quality is encouraging and shows that simulation can be used to forecast weekly product fab out with acceptable accuracy.

4.3 Challenges to Increase Forecast Accuracy

A detailed analysis on lot and product level revealed more reasons why the forecast accuracy was low, besides the wafer start forecast inaccuracy. Events that can cause a lot hold in production cannot be predicted for a particular lot. On the product level historical data is used to derive a distribution for lots on hold. The corresponding distribution is used in simulation to put lots on hold. Therefore it is possible to reach a good forecast accuracy for e.g. product moves. But it is very unlikely that exactly the same lots are put on hold in simulation as in reality. Therefore the prediction for fab out on lot level can never be that accurate.

Manual prioritization done by line control on lot or product level not known by any data system is also a challenge. Any simulation that does not have this information will maybe still have an acceptable fab move prediction but fab out predictions on a detailed level will be off. This problem was especially obvious in fabs with a low automation grade. The higher the degree of fab automation the less likely it is that crucial logistics information is not available in data systems.

Another challenge exists in fabs where the loading and unloading of tools is still done by operators. Very often operators do not follow the dispatch list for various reasons. Again, this is not necessarily a problem for a fab move prediction accuracy but it will impact any forecast accuracy on lot level.

4.4 Simulation vs Big Data Comparison

We compared the forecasts generated by two different fab simulation models with the respective Big Data results. One was a validated fab simulation model of an Infineon frontend facility in Europe, and the other one a validated simulation model of an Infineon frontend facility in Asia.

Figure 2 shows the gap between the forecasted remaining cycle time and the actual remaining cycle time for the European fab. Each data point represents the deviation for one lot. The lots are put into weekly buckets, dependent on the actual remaining cycle time of the lots. This means that data points in the first box plot of the figure represent lots where the remaining cycle time was less than 7 days from the respective simulation start time. The upper part shows the results obtained with Big Data, and the lower part the results obtained with simulation. Coloring indicates the deviation between forecasted value and reality. Red indicates a higher gap percentage as compared to green. The ideal result for a lot would be a dot with y value 0, indicating that the forecast met reality exactly.

The simulation showed better results within the first weeks of the forecasting period. Big Data overestimated the remaining cycle time at the beginning, simulation underestimated cycle time at the end of the forecast period slightly more than Big Data. The underestimation of cycle time with the simulation approach is caused by missing line disturbance modelling elements such as operator and production interference of changing lots priority. This could be mitigated through enhancement of the simulation model prior to delivery of the use case to the end users.

Figure 3 shows a similar box plots for the results from the Asian frontend facility, again with weekly buckets. The two leftmost boxplots show the results for all lots that finished processing within the first week from the simulation start time (remaining cycle time was less than 7 days). Again, each dot represents a lot. The blue boxplot represents results obtained with Big Data, the orange one results obtained with simulation. The results are similar for both frontend sites. Simulation prediction is more accurate at the beginning of the forecasting period. After some weeks simulation predicts shorter cycle times than reality, a similar effect can also be observed for Big Data although not as strongly.

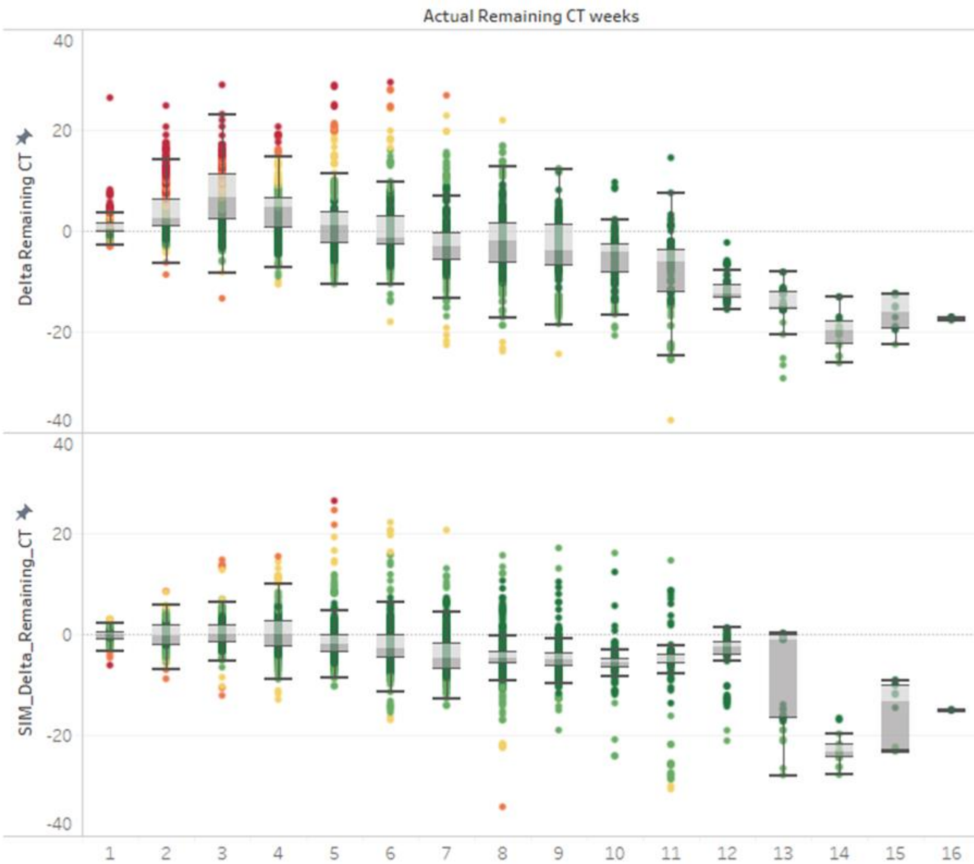

Figure 2: Cycle time forecast deviation for an European frontend site (big data upper part, simulation lower part).

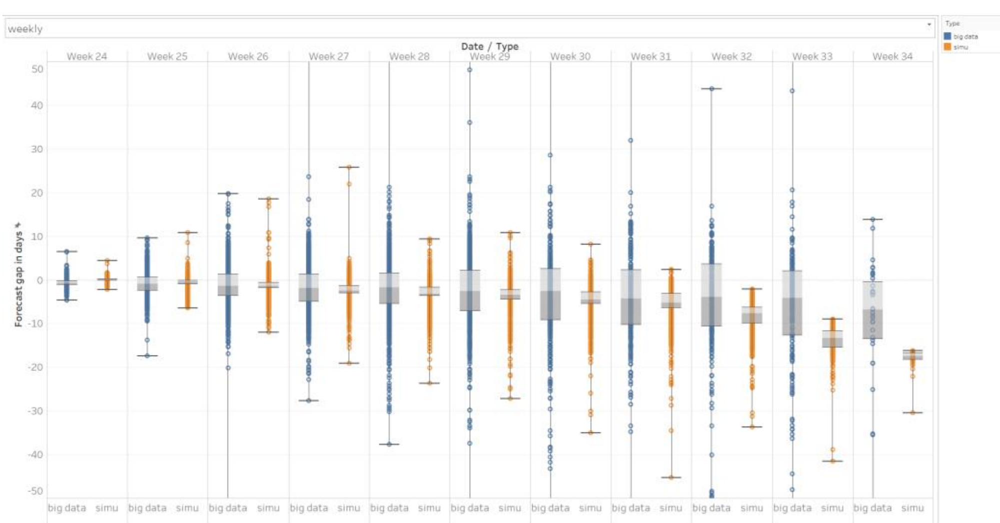

Figure 3: Cycle time forecast deviation for an Asian frontend site (big data blue boxplots, simulation orange boxplots).

Figure 4 shows the correlation of Big Data based and simulation-based forecasting. Each dot represents the data for a lot. If the forecast is 100% accurate for both simulation and big data the data point will end up right in the origin of the graphic. If the simulation predicts a higher/lower cycle time for the lot than observed in reality the data point will have a larger/smaller y-value and if big data predicts a higher/lower cycle time for the lot the x-value of the data point will be larger/smaller. Coloring of data points indicates different product classes.

Clustering of data points can be observed along the diagonal. This implies that for some lots cycle time predictions are tough to make. For example, lots going on hold or a change of lot priorities not known at the simulation start will be impossible to predict, irrespective of the method used.

Accuracy of cycle time predictions can also be dependent on process classes too as visible in Figure 4. For some product classes the spread of data points is higher as compared to others.

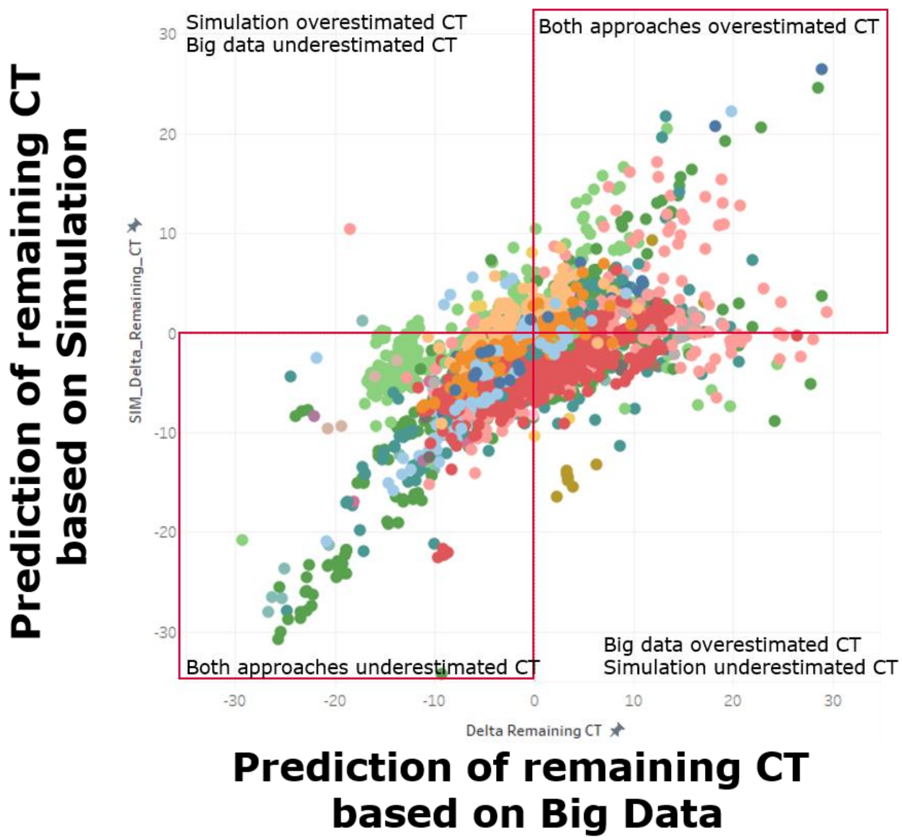

Figure 4: Correlation of big data and simulation forecast for an European frontend site.

5 CONCLUSIONS AND FUTURE WORKS

Based on our study, the simulation forecast results have been slightly better than big data forecast results. Big data has the advantage of less effort to generate predictions. We are working on improving both methods. Simulation tends to underestimate cycle times because the simulation model is too fast. Disturbances in the fab such as lots are waiting for operators to load equipment are not considered in our models yet.

Big Data can be improved by adding factors to incorporate actual fab performance. One idea is to combine both methods in the future. Prediction of future fab performance can be done once a week by a simulation run. These results can then be used as a factor for Big Data analysis.

This approach can reduce the effort for simulation because a weekly simulation run is sufficient. Big data can then be used for daily predictions. The question of whether this can improve forecast results and reduces the effort to generate the forecasts will be part of future studies.

ACKNOWLEDGMENTS

We would like to thank our colleagues in contributing towards the successful completion of this work. They are Dr Andreas Kyek, Mr Patrick Preuss, Mr Chew Wye Chan, and Ms Xiaoqing Yang.

REFERENCES

Bagchi, S., C.H. Chen, S.T. Shikalgar, and M. Toner. 2008. "A Full-Factory Simulator as a Daily Decision Support Tool for 300mm Wafer Fabrication Productivity". In *Proceedings of the 2008 Winter Simulation Conference*, edited by S. Mason, R. Hill, L. Mönch. 2021-2029. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Chien, C. F and Chuang, S. C. 2014. "A Framework for Root Cause Detection of Sub-Batch Processing System for Semiconductor Manufacturing Big Data Analytics". *IEEE Transactions on Semiconductor Manufacturing* 27(4): 475-488.

Chen, Y.J., Wang, B.C., Wu, J.Z., Wu, Y.C., and Chien, C.F. 2017. "Big Data Analytic for Multivariate Fault Detection and Classification in Semiconductor Manufacturing". In *13th IEEE Conference on Automation Science and Engineering*, Aug 20th-23rd, Xi'an, China, 731-736.

D-SIMLAB Technologies. 2020. Forecaster and Scenario Manager. http://www.d-simlab.com/category/d-simcon/products-d-simcon/forecaster-and-scenario-manager, accessed 20th April 2020.

Fowler, J.W., Mönch, L. and Ponsignon, T. 2015. "Discrete-Event Simulation for Semiconductor Wafer Fabrication Facilities: A Tutorial". *The International Journal of Industrial Engineering: Theory, Applications and Practice*. 22. 661-682.

Gan, B.P., P. Lendermann, Y.L. Loh, H.K. Tan, S.K. Lieu, L.F. McGinnis, and J.W. Fowler. 2004. "Analysis of a Borderless Fab Scenario in a Distributed Simulation Testbed". In *Proceedings of the 2004 Winter Simulation Conference*, edited by R.G. Ingalls, M.D. Rossetti, J.S. Smith, and B.A. Peters, 1896-1901. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Geng. N. and Z. Jiang. 2009. "A Review on Strategic Capacity Planning for the Semiconductor Manufacturing Industry". *International Journal of Production Research* 47(13): 3639-3655.

Lee, C., Cao, Y. and Ng, K.H. 2017. "*Big Data Analytics for Predictive Maintenance Strategies"*. In *Supply Chain Management in the Big Data Era*, edited by H.K. Chan, N. Subramanian, M.D. Abdulrahman, 50 – 74. IGI Global.

Mosinski, M., T. Weissgaerber, S.L. Low, P. Preuss. 2017. "An Easy Approach to Extending a Short Term Simulation Model for Long Term Forecast in Semiconductor Industry". In *Proceedings of the 2017 Winter Simulation Conference*, edited by W.K.V. Chan, A. D'Ambrogio, G. Zacharewicz, N. Mustafee, G. Wainer, and E. Page, 3636-3645. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Moyne, J. and Iskandar, J. 2017. "Big Data Analytics for Smart Manufacturing: Case Studies in Semiconductor Manufacturing". *Journal of the Processes* 5(39): 1-20.

Nyhuis, F. and Alves, P. 2002. "Methods and Tools for Dynamic Capacity Planning and Control". *Gestão & Produção* 9(3): 245-260.

Scholl, W., D. Noack, O. Rose, B.P. Gan, P. Lendermann, P. Preuss, and F.S. Pappert. 2011 "Implementation of a simulation-based short-term lot arrival forecast in a mature 200mm semiconductor fab". *In Proceedings of the 2011 Winter Simulation Conference*, edited by S. Jan, R. Creasey, J. Himmelspach, K.P. White, and M.C. Fu, 1932-1943. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Scholl, W., M. Mosinski, B.P. Gan, P. Lendermann, P. Preuss, D. Noack. 2012. "A Multi-Stage Discrete Event Simulation Approach for Scheduling of Maintenance Activities in a Semiconductor Manufacturing Line". In *Proceedings of the 2012 Winter Simulation Conference*, edited by C. Laroque, J. Himmelspach, R. Pasupathy, O. Rose, and A.M. Uhrmacher, 2263-2272. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Scholl, W., M. Förster, P. Preuss, A. Naumann, B.P. Gan, and P. Lendermann. 2016. "Simulation-Enabled Development Lot Journey Smoothening in a Fully Utilized Semiconductor Manufacturing Line". In *Proceedings of the 2016 Winter Simulation Conference*, edited by T.M.K. Roeder, P.I. Frazier, R. Szechtman, E. Zhou, T. Huschka, and S.E. Chick, 2559-2567. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Scholl, W., M. Förster, P. Preuss, A. Naumann, H.P. Ooi, and B.P. Gan. 2018. "Bridging the Gap Between Fab Simulation and Supply Chain Forecast". In *Proceedings of the 2018 Winter Simulation Conference*, edited by M. Rabe, A.A. Juan, N. Mustafee, A. Skoogh, S. Jain, and B. Johansson, 3461-3469. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Seidel, G., C.F. Lee, A.M. Kam, B.P. Gan, C.W. Chan, A. Naumann, and P. Preuss. 2017. "Harmonizing Operations Management of Key Stakeholders in Wafer Fab Using Discrete Event Simulation". *In Proceedings of the 2017 Winter Simulation*

Conference, edited by W.K.V. Chan, A. D'Ambrogio, G. Zacharewicz, N. Mustafee, G. Wainer, and E. Page, 3670-3678. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Seidel, G., P. Preuss, A. Naumann, S.L. Low, C.W. Chan, and B.P. Gan. 2018. "Operator Resource Modelling in a Multiple Wafer Sizes Wafer Fabrication Environment Using Discrete Event Simulation". In *Proceedings of the 2018 Winter Simulation Conference*, edited by M. Rabe, A.A. Juan, N. Mustafee, A. Skoogh, S. Jain, and B. Johansson, 3651-3659. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Wang, J. and Zhang, J. 2016. "Big Data Analytics for Forecasting Cycle Time in Semiconductor Wafer Fabrication System". *International Journal of Production Research* 54(23): 7231-7244.

Wang, L.C., A. Wang. and C.Y. Chueh. 2018. "Development of a capacity analysis and planning simulation model for semiconductor fabrication". *The International Journal of Advanced Manufacturing Technology.* 99(1-4): 37-52.

Zhang, M.T., Fu, J. and Zhu, E. 2008. "Dynamic capacity modeling in semiconductor assembly manufacturing". International Journal of Production Research. 46(3): 739-752

AUTHOR BIOGRAPHIES

GEORG SEIDEL is Senior Staff Engineer of Infineon Technologies Austria AG (Villach, Austria). He has been involved in simulation, WIP flow management and Industrial Engineering topics since 2000. He was responsible for WIP flow management, especially for Lot dispatching at Infineon's site in Kulim (Malaysia) from 2012 until 2015. He is now responsible to rollout Fab Simulation in Kulim and Villach. He holds a Master degree of Technical Mathematics. His email address is georg.seidel@infineon.com.

CHING FOONG LEE is Senior Specialist Engineer of Infineon Technologies (Kulim) Sdn. Bhd. She has been involved in Semiconductor System Development and Datamining since 2004. She joined Infineon Technologies Kulim in 2010 driving various projects in Production System Setup, Reporting and System Improvement under Factory Integration department. Currently she is responsible in Kulim for Simulation and WIP flow management topics under Operation Research and Engineering department. She holds Master of Business Administrative(MBA) and Bachelor Degree of Information Technology, majoring in Software Engineering. Her email address is chingfoong.lee@infineon.com.

AIK YING TANG is an engineer of Infineon Technologies (Kulim) Sdn. Bhd. She is currently involved in simulation topics under WIP Flow Management department. She holds a Doctor of Philosophy Degree specializing in Mathematics. Her email address is aikying.tang@infineon.com.

WOLFGANG SCHOLL works as a Senior Staff Expert for modeling and simulation for Infineon Technologies in Dresden (Germany). He studied physics at the Technical University of Chemnitz (Germany) and graduated in solid-state physics in 1984. From 1984 to 1995 he worked as a process engineer for ZMD in Dresden. In 1996 he joined Infineon Technologies (former SIMEC) and worked in the field of capacity planning. Since 2003 he is responsible for fab simulation. He supervises development and roll-out projects and is also a member of the Supply Chain Simulation community. His email address is wolfgang.scholl@infineon.com.

SOO LEEN LOW is a Project Manager at D-SIMLAB Technologies (Singapore). She is responsible for simulation modelling and analysis of Wafer Fabrication plants. She earned a Bachelor of Engineering in Computer Engineering from National University of Singapore (NUS) in 2014. Her email address is soo.leen@d-simlab.com.

BOON PING GAN is the CEO of D-SIMLAB Technologies (Singapore). He has been involved in simulation technology application and development since 1995, with primary focus on developing parallel and distributed simulation technology for complex systems such as semiconductor manufacturing and aviation spare inventory management. He led a team of researchers and developers in building a suite of products in solving wafer fabrication operational problems. He was also responsible for several operations improvement projects with wafer fabrication clients which concluded with multi-million dollar savings. He holds a Master of Applied Science degree, specializing in Computer Engineering. His email address is boonping@d-simlab.com.