

FACTORY CAPACITY OPTIMIZATION WITH INTEGER LINEAR AND NON-LINEAR PROGRAMMING APPROACHES

Fernando P. Elizaga Jr.
Jonathan S. Legarde
Snehalata More

Advanced Analytics Office, Head Backend Operations
Western Digital Corporation (Philippines), 109 Technology Ave, Laguna Technopark, Biñan, Laguna, PH

ABSTRACT

Semiconductors and high-tech industries are characterized with high volatility, dynamics, and complex manufacturing and supply chain operations. Driven by consumer, cloud and client end-markets, the demands might change highly over time, which creates significant challenges on factory and capacity planning domains. By their strategic nature, the related decisions consider a longer, involve major investment decisions and incorporate long equipment delivery lead times, particularly when special processes are considered.

In the paper, a novel capacity planning approach is presented that optimize the strategic, long-range capacity plans (LRP) in response to the following problems. First, the manual process takes 2 weeks to complete per cycle with only 3 scenarios explored in an iterative manner, considering that there are more than 15 cycles of assessment within 5 months causing catch-up issues with the many iterations. This limitation deprives optimization and sensitivity analysis leading to a plan that may not be robust against parameter changes and disruptions. Second, the manual nature of the existing process using MS excel is prone to errors either in the data input, excel formula, and product model assignment. Third, the static straight forward calculation with zero capacity flexibility buffer will lead to a wrong investment strategy with high capital expenditure (CAPEX) and may even lead to a failure to close the required build volumes.

The presented models leverage integer linear and nonlinear programming approaches, to identify the cost-optimal capacity investment plans, and furthermore, to identify the scenarios in engineering versus investment tradeoffs.

With the new system, the analysis time was improved from 14 days with only 3 scenarios to a single day but having multiple scenarios. It ensures the use of capital more efficiently, in a mathematically optimal way with scenario exploration and robust planning. Errors are also eliminated with the system calculation. Finally, the OR model-based system allows the strategic planning team to create a robust plan against price, volume, and capacity changes where available resources are maximized and capital is minimized.

1.0 INTRODUCTION

Semiconductors and high-tech industries are characterized by volatile demand, high variety of products and short product lifecycles coupled with complex manufacturing processes and special, high-cost equipment. These challenges require a combination of advanced forecasting and optimization techniques in long range factory capacity planning.

Typically, long range planning is based on a 5 year horizon, requiring a broad range of external and internal factors to be considered. Volumes can fluctuate dramatically due to changes in technology, demand, and macroeconomic factors. Predicting future demand accurately becomes difficult, especially when planning several years ahead. Therefore, the decision makers need to deal with highly uncertain parameters, and variance of the key driving factors.

The horizontal process segmentation in the semiconductor supply chain is intricate, involving multiple stages such as wafer fabrication, assembly, and testing. Coordinating capacity across these stages and ensuring alignment with market demand adds complexity. In a typical case, factories are dedicated to performing the major steps of the entire fabrication: as an example in the hard-disk drive (HDD) production, the overall process chain is broken down to media and substrate (disk), and head (reader/writer component) manufacturing, and dedicated sites perform the final product assembly. In head, both wafer manufacturing and backend assembly processes are highly complex, and therefore, are performed in separate and dedicated sites.

Building new head manufacturing facilities, or expanding existing ones requires significant lead times, often spanning several years. Accurately forecasting demand far in advance to justify these investments is challenging. Besides, the related technology evolves rapidly, new products with increasing complexity are introduced often. Long-range capacity planning (LRP) must account for technology advancements and ensure that future capacity aligns with emerging requirements. This means that the technology and equipment should be flexible and adaptive, to support the major shifts over the evolution of product generations.

In the backend operations, high-precision processes have a varying degree of automation, as some processes are still human labor intensive. Considering both the automation and equipment, manufacturing sites require capital-intensive investments, resulting in high costs to build and operate. Making long-term capacity decisions involves assessing financial risks and ensuring sufficient return on investment.

Another important factor is the stability of the process conditions that drives the unit capacity and process yield. Even though high-precision equipment is involved, and the conditions are under very strict control in cleanroom environments, tool and material modifications have a varying degree of product quality, especially in the ramping stage of the new product. Therefore, unit capacity and yields can vary due to process variations, equipment performance, and other factors. Ensuring adequate capacity while minimizing costs requires managing the variability of these parameters.

2.0 REVIEW OF RELATED WORK

Factory capacity planning is a strategic decision with a long-range, typically multi-years horizon, due to the lead times of related adjustments. Martinez-Costa et al. [1] revise various mathematical programming model to solve typical, long-range site capacity planning problems. They also specify the core problem as it follows. Single-site strategic capacity planning consists in determining the capacity expansion size of an existing plant in each period. Expansion alternatives include equipment acquisition from equipment vendors, rent and transfer by outsourcing. If the plant needs non-redundant, independent multiple resources, it is sometimes possible to identify one bottleneck resource. If this resource is expanded, all the capacity of the plan is too, provided that the others increase in the corresponding amount and the problem is equivalent to a single resource type. In the same way, if a plant produces multiple products but each product has its dedicated resources, they are independent because each type of resource serves the demand of a different product, and the capacity problem consists in determining the number of units of capacity expansion of each resource.

The state-of-the-art factory capacity planning techniques in the high-tech and semiconductors industry typically apply integer linear programming techniques, to address equipment conversion, investment timing and capacity ramping decision. As Rastogi et al. [2] highlight, the complexity of global capacity planning combined with the large capital expenditures to increase factory capacity makes it important to incorporate optimization methodologies for cost reduction and long-term planning. They present a two-stage stochastic integer-programming formulation to model a semiconductor supply network. The model makes strategic capacity decisions, (i.e., build factories or outsource) while accounting

for the uncertainties in demand for multiple products. Model was also used to analyze how variability in demand affects make/buy decisions and how correlation between demands of different products affects these strategic decisions.

Barahona et al. [3] investigates the impact of demand uncertainty on long-range planning decisions in semiconductors manufacturing. They present a stochastic programming approach to capacity planning, a mixed-integer program in which expected value of the unmet demand is minimized subject to capacity and budget constraints. This is a difficult two-stage stochastic mixed-integer program which cannot be solved to optimality in a reasonable amount of time. Therefore, they propose a heuristic that can produce near-optimal solutions and strengthens the linear programming relaxation of the formulation with cutting planes and performs limited enumeration.

Manufacturing equipment should be flexible and convertible to adopt to the changes in the product portfolio. Conversion kits provide the opportunity to qualify the resources for new products, without replacing the core manufacturing units. In the capacity planning, this involves investigating the tradeoff relation between new equipment investment and upgrading the existing ones. The former provides not only enhanced equipment capability (e.g. precision), but also volume scaling, however, the associated costs are significantly higher than kit-based upgrading. The related planning problem is analyzed by Zhang et al [4]. They proposed a two-level hierarchical planning methodology to generate a complete capacity planning solution using mixed-integer linear programming. It covers mid-range monthly planning and automated capacity allocation system covers short-range weekly planning. These systems are integrated to generate optimal capacity plans considering kit components.

A hybrid capacity planning method is proposed by Leu and Liu to solve combined capacity planning and volume allocation problems in semiconductors supply chains. They used linear programming to get a baseline solution, and then use discrete-event simulation to approximate a detailed solution for the production capacity planning of supply chain. Based on the proposed decomposition, the method is scalable, however, it is more suitable for mid-range capacity planning rather than strategic site capacity planning [5].

3.0 METHODOLOGY

In the coming section, the problem of a single-site strategic capacity planning is specified, considering alternative resources with different cost and precision attributes. Then the generic mathematical models is formulated as a non-linear integer programming model, with adjustable equipment capabilities that introduce decision tradeoffs.

1.1 Single-site Long Range Capacity Planning Problem with Alternative Resources

In the core investigated problem, multi-period capacity planning divided into quarters is considered with a single factory site. The volume forecast with multiple products is assumed to be known over the horizon, and demand uncertainty is disregarded. However, the experimental cases will assume alternative demand scenarios and therefore, sensitivity analysis will be proposed. Furthermore, the product roadmap is also specified, as well as manufacturing technology requirements, the type of the resources to be used, and the capacity required from each resource type. The routing of the products is not considered, which means that the capacity calculations disregard the sequence of operations, and only the aggregate capacity requirements are considered per equipment type and product.

The products are made by several different operations, each requiring specific equipment. The equipment capabilities are specified by the operations they can perform, and the cost equipment unit. Another important attribute is the process yield that specifies the quality rate of executed operations. These parameters are in a tradeoff setting, i.e., a higher spec equipment is associated with higher purchase cost and vice versa. Furthermore, capacity in the form of processing time is also an attribute of the equipment, and it determines the time associated with processing a unit of products.

A brownfield planning scenario is investigated with a manufacturing site that operates already in the beginning of the planning horizon. As initial parameters, the existing set of equipment is already known, including the products and capacity requirements. The planning decisions determine the equipment to be purchased over the horizon, specifying their type, the time of purchase, set-up leadtime, qualification and tool release to line. Additionally, equipment upgrades leveraging conversion kits is also possible, associated with upgrade costs. Tool end-of-life (EOL) and productivity improvement projects are also considered in the planning to sustain and improve the current plant capacity.

The principal planning rule is that the supply should always match the demand, i.e., the available capacity of the factory should be sufficient enough to meet (or exceed) the demand at any time. In addition, the equipment capabilities must also meet the technical requirements of the product portfolio. Important planning aspect is the availability of alternative equipment, i.e., the yield and unit capacity of equipment can be optimized to meet the demand, and therefore, they are considered to be variables in the planning model.

Most importantly, the overall objective is to minimize the total investment (capital) required by the site over time.

1.2 Mathematical Model of the Planning Problem

The base long range capacity planning problem is formalized as an integer linear programming model as it follows. In the coming sections, the nonlinear extension of the problem will also be proposed, in order to capture the adjustable equipment parameters, such as the unit capacity or yield.

In the model the set of product groups are denoted by i , including the products j . The process steps of making the products are denoted by k , and the time horizon is divided into equal length periods t . The manufacturing resources are denoted by l . The major parameters and decision variables are the volume X of products that are processed in a certain period by a set of equipment per operation. The tool capacities are quantified by Q and R , continuous and integer measures, respectively. The latter is used in case of special tool capacity requirements where rounding is necessary for the capacity calculations. Considering a brownfield planning scenario, there is a set of available tools A , while the decision variable N denotes the investments that are made over the horizon.

$$\text{minimize } \sum_{i,k,l} \left(\text{unitCost}_{i,k,l} \sum_t N_{i,k,t,l} \right) \quad (1)$$

$$\text{AvailableTools}_{i,k,t,l} \geq \text{ExistingTools}_{i,k,t,l} + \sum_{t=1}^{T-1} N_{i,k,t,l} \quad \forall i, k, l \quad (2)$$

$$\sum_{i,l} X_{i,j,k,t,l} \geq \text{Demand}_{j,k,t} \quad \forall j, k, t \quad (3)$$

$$R_{i,k,t,l} \leq Q_{i,k,t,l} + 1 \quad \forall i, k, t, l \quad (4)$$

$$R_{i,k,t,l} \geq Q_{i,k,t,l} \quad \forall i, k, t, l \quad (5)$$

If tool is existing tool, i.e.: $\forall i \in \text{AsIsTools}$

$$Q_{i,k,t,l} = \frac{\sum_j X_{i,j,k,t,l}}{UCap_i \cdot Yield_i \cdot Util_i \cdot NoDays} \quad \forall k, t, l \quad (6)$$

$$Q_{i,k,t,l} \leq \text{ExistingTools}_{i,k,t,l} \quad \forall i, k, t, l \quad (7)$$

$$\text{If is a special required tool, i.e.: } \forall i \in \text{SpecReqTools} \quad (10)$$

$$Q_{i,k,t,l} = \frac{\sum_j X_{i,j,k,t,l} - X_{i,j,k,t,l,AsIs}}{UCap_i \cdot Yield_i \cdot Util_i \cdot NoDays} \quad \forall k, t, l \quad (11)$$

$$Q_{i,k,t,l} \leq N_{i,k,t,l} + \text{ExistingTools}_{i,k,t,l} \quad \forall i, k, l \quad (12)$$

$$\quad (13)$$

If tool is a standard tool, i.e.: $\forall i \in \text{StdTools}$

$$Q_{i,k,t,l} = \frac{\sum_j X_{i,j,k,t,l}}{UCap_i \cdot Yield_i \cdot Util_i \cdot NoDays} \quad \forall k, t, l \quad (14)$$

$$Q_{i,k,t,l} \leq N_{i,k,t,l} + \text{ExistingTools}_{i,k,t,l} \quad \forall i, k, l \quad (15)$$

The objective function (1) minimizes the total investments, equivalent to the unit costs multiplied with the quantity of tools purchased. The first constraints state that the set of available tools at any period is the sum of available and additionally purchased equipment (2). The demand must fulfilled, i.e., undersupply is not allowed as provided in (3). The tool capacity rounding expressions are provided in (4) and (5). The tool capacity requirement for existing, standard and special tools are provided by (6)-(7), (8)-(9) and (10)-(11), respectively.

1.3 Solution of the LRP Problem

The base problem, i.e., the LRP model was formulated as a mixed-integer linear program (MILP). Depending on the problem size, this can be solved with general optimization techniques, e.g., the branch-and-bound algorithm. However, in several cases, the solution is somewhat straightforward, especially if volume reallocation across time periods is not allowed, i.e., it cannot be flexibly balanced with the capacities. This assumption holds in the present model, as stated in constraint (3). Therefore, in order to identify the base solution of the overall planning problem, a non-linear extension of the original model is proposed.

In this extended formulation, it is assumed that the unit capacity and yield parameters are variable, and hence they can be adjusted within ranges, and these adjustment are associated with engineering costs. As both parameters are in the denominator of e.g. constraint (6), the transformed model will be highly non-linear, and branch and bound techniques cannot be applied. The numerical complexity will significantly increase as the integrity constraints and binary variables remain unchanged.

optimization model for all parameter combinations within the ranges. A more sophisticated way of solving the problem is defining a custom heuristics for the scenario iteration, by implementing a search algorithm that explores the impact of parameter combinations on the linear relaxation of the original problem, which provides good guidance on setting the parameters of the original problem instance.

As an example for the latter, in each iteration, a fix set of key input variables (KIVs) is considered, however, these parameters are adjusted from scenario to scenario. For every scenario run, the model will always produce a POR (Plan-of-Record) case, a BEST case and a WORST case, where all cases meet the demand constraints.

The POR results is based on the original given KIV, and will provide the total capacity and capital required for investment. Alternatively, the BEST case results come after optimizing the yield, unit capacity and machine cost relative to the defined % improvement from the POR. The results will show the optimum capacity with a much lower investment required. This will help the engineering and procurement teams identify the critical processes and equipment as focus areas for improvement, and for purchase. Contrary, the WORST case will also be provided by the model which will show the opposite impact of the BEST case showing a higher investment requirement.

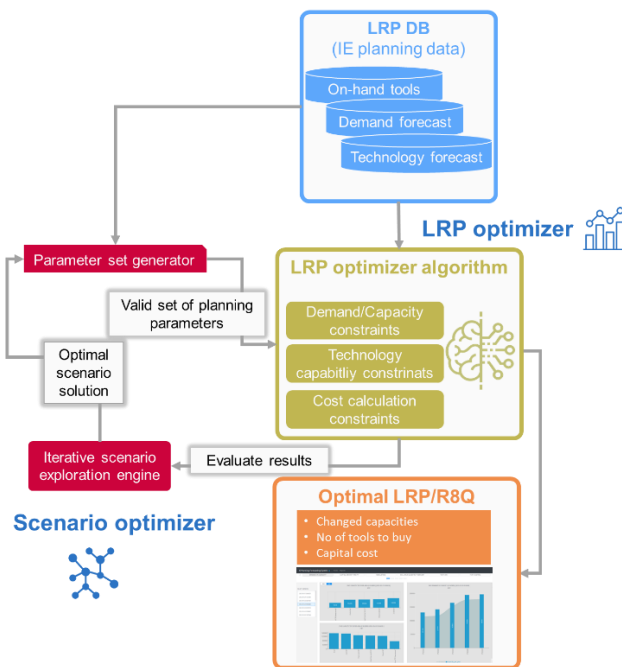
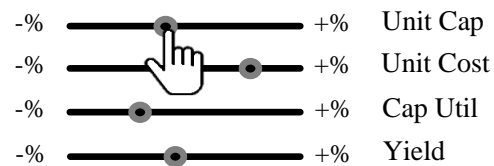


Fig. 1. Workflow of the overall planning problem, with the base parameters settings and the nonlinear extension of the problem, where scenario exploration (red) is applied to find the best input parameter (changed to decision variables) combinations.

Two alternative approaches are proposed to solve the extended model. If the adjustable parameters take discrete values within the specified engineering ranges, then a full factorial experiment plan can be defined to solve the

Inputs & Constraints Scenario



Having the two alternative approaches on top of the original POR case will allow the user and the model to understand the range at which the next set of key input variables (unit capacity, unit cost, yield, utilization, etc) can be adjusted.

Finally, the raw data produced by the model associated with the solutions provided by the optimizer is injected to a systems database (Microsoft SQL server). Stored procedures (referred as Scripts) are created in the database using SQL queries which generates the aggregated data format required by the users. These aggregated data are transferred to the web application and plotted into visualization (charts and summary tables). The web-based application was designed to visually explore and navigate the results.

4.0 RESULTS AND DISCUSSION

The overall planning workflow and related application has been implemented in a web-based application (see Figure 2), leveraging full-stack architecture. All planning parameters, scenario data etc. are stored in SQL databases, populated by users via a web application interface. The planning model has been implemented in FICO Xpress (MILP model implementation) and Python language was used for implementing the neighborhood search iterator.

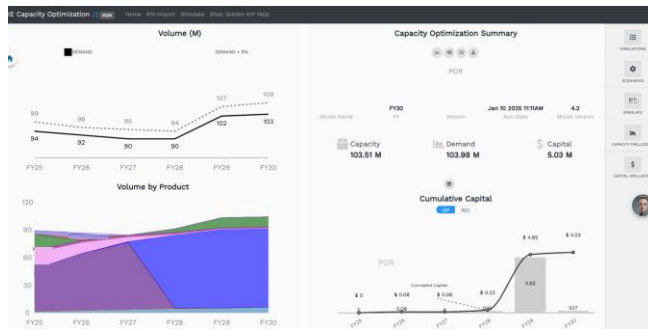


Fig. 2. The interactive Web-based application.

In a numerical experiment illustrated in Figure 3, a base volume forecast is taken, and the long-range capacity plan is calculated. Then, in a subsequent period a revision of the forecast is provided, anticipating significant volume changes in future quarters. Hence, the LRP experimentation was performed on a rolling horizon bases, and the numerical results between manual and optimization-based calculations were compared. During the experimentation, the non-linear extension of the original model was used, i.e., adjustable yield and unit capacity parameters for sensitivity analysis were also considered to identify the best equipment capabilities, in balance with the volume scenarios.



Fig. 3. Test planning scenarios with forecast revisions: Build volume plan significantly differs for the three (3) different forecasts showing the variability while growing, that needs to be address by the LRP model with proper capacity adjustment, avoiding excess capital expenditure.

In the numerical experiments, a full-site optimization scenario was considered, with a planning horizon of 5+ years, a large set of processes (150+) and portfolio of products (30+). The range of adjustable unit capacity and yield were provided as input, in the percentage of the original values. Taken all the input parameters, a series of numerical experiments were conducted, and the results were compared to those obtained with standard calculation methods.

Comparing the forecast-to-forecast revisions, the major business benefit of using the optimization model is aiding top management in making the right decisions to select the best strategy for the business which helps to avoid unnecessary investments by properly balancing the equipment parameters and capacity with the volume demand. Buying enough tools and achieve optimum capacity to support the volume with 10% fluctuation is strategic as compared to buying excess tools that later-on will require vendor negotiation for delivery push-out. With this, the optimization model helped to identify the best tradeoff settings among the key input parameters eventually contributing to several million annual capital expenditure avoidance over the planning horizon, comparing the solutions obtained by the manual planning process versus those provided by the optimization model.

Another benefit of the interactive web-based application is the ease of data exploration and navigation thru visualization. As a factory with hundreds of process steps in different production areas to make the finish product, it is important for planners to identify the bottleneck areas and processes that needs focus in case volume or demand suddenly increase.

The system capability in the web application includes the capacity drilldown function wherein users can see the overall plant-wide demand vs capacity as shown in Figure. 4, and can further be drilled down to the desired next level detail.



Fig. 4. Capacity drilldown showing total demand vs capacity.

The visualization will give the strategic planners a much better view and understanding on which specific production area is the gating capacity, and which specific processes are identified as the top detractors that needs focus and improvement as shown in Figure 5.

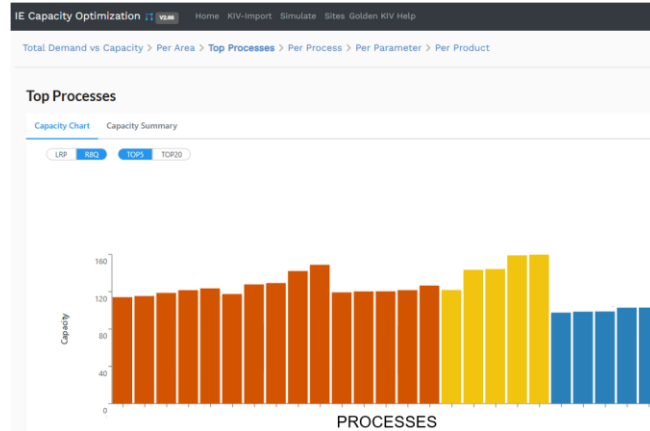


Fig. 5. Top Processes per Area to help planners identify bottleneck processes.

If there is a need to zoom in to a specific process, particularly those that require investments, the demand vs capacity per process step is also available as shown in Figure 6.

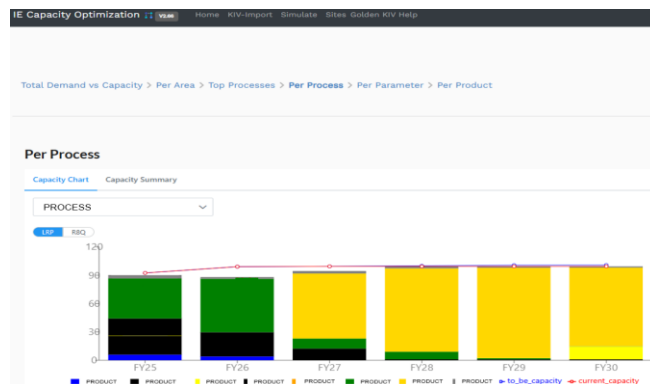


Fig. 6. Demand vs capacity for a specific process step

As product life cycle is fast in a semiconductor and high technology company, demand vs capacity analysis for new products is also presented in the visualization. To understand which specific tool and parameter is gating the capacity, the parameter level detail is also available per process and per specific tool.

To have the confidence in using the system, it is important to verify the results from the model. A validation was conducted for several months comparing the output of the manual calculation using excel spreadsheets and the system results from the model. As shown in Figure 7, both results show a gap delta of less than 1% per area confirming the viability of the model to replace the manual calculation process.

Area	Manual (Capital-M\$)	System (Capital-M\$)	Delta (M\$)	Manual (Capacity)	System (Capacity)	Delta (M Units)
AREA 1	3.45	3.45	0	108.9	109.3	0.4 (0.3%)
AREA 2	0.04	0.04	0	110.8	110.6	0.2 (0.2%)
AREA 3	0.66	0.66	0	94.8	94.8	0
TOTAL	4.15	4.15	0			

Fig. 7. Comparison of sample run using manual and system calculation.

5.0 CONCLUSION

In the paper, a long-range factory capacity planning approach was proposed with the base linear programming model, and its non-linear extension where equipment parameters are adjustable. Leveraging these parameter adjustments, the user and model can identify the optimal combination of capacity and equipment capability settings on a rolling horizon. To solve the nonlinear extension of the original problem, a neighborhood search iterator can be applied, which explores the impact of unit capacity and yield changes on the linear relaxation of the original problem and provides guidance to set parameters for the original MILP model.

A series of numerical experiments were conducted to assess the business benefits that can be achieved by the optimization. Hence, the optimizer was expected to detect the potential investment savings, to avoid capital spending in surplus capacities. The experimental results show potential in applying the proposed model in full scale real factory planning situations.

In addition, the web-based interactive user interface provides the ease of data exploration and navigation using the capacity drilldown function which helps planners to analyze demand and capacity results down to the parameter level detail. This allows planners and process owners to identify processes needing capacity improvements and can focus investments on new products and new technologies.

6.0 RECOMMENDATIONS

The key input variables (KIVs) are currently being uploaded to the system using an excel file and converted to the database. Thus parameter scenario adjustments are conducted manually based on the proposal of the model. It is recommended in the future to extract the KIVs directly online linking the system to data sources for volume, tool performance, and oracle (procurement cost), etc.

It is also recommended to complete ongoing enhancements on the existing beta model to not only cover the planning needs related to tool capacity and new technology, but also productivity and end-of-life tool sustenance, capital release and expenditure planning, depreciation, and capex forecast-to-forecast sensitivity gap analysis.

Currently, the authors are testing the model in other factories (China, Malaysia, Thailand) where the planning logic and rules are similar, and extended the model to multi-site planning scenarios. It is recommended to complete this project harmonization across the different sites to explore the potential in synergistic cross-site capacity balancing.

7.0 ACKNOWLEDGMENT

The authors would like to express their sincere gratitude to the following people who supported this project. It has been an honor for us to experience this cross-functional collaboration from the different functions and across different sites of western digital.

An immense thank you to David Gyulai, Phd (director), Divya Mutha, MSIE (data scientist), and Jhun Vitualla, SSBB (program manager) from the Advanced Analytics Office for their technical expertise and project management assistance especially in the backend development of the system.

A heartfelt thank you to Albert Zaldivar (senior technologist) and Marivie Vinuya (data analyst) from the Data Management Engineering for their continued support starting from the agile project sprints, frontend development of the system, until the project harmonization across different sites.

A deeper gratitude to Isaacs Chaipanich (director), Efren Levantino (principal engineer), Alvin Zeta (staff engineer), and Songtai Teera-Anuwat (technologist) from Global Industrial Engineering who had been our strong advocate being the subject matter experts in long range capacity and capital planning, for the several months of testing, validation, revision, and finalization of the system.

We are expressing our profound thanks to all of you for your continued support in helping our team and company to become a leader in the field of technology-driven industrial transformation.

8.0 REFERENCES

1. C. Martínez-Costa, M. Mas-Machuca, E. Benedito, A. Corominas, "A review of mathematical programming models for strategic capacity planning in manufacturing", *International Journal of Production Economics*, 153, 2014, pp.66-85.
2. A. P. Rastogi et al., "Supply network capacity planning for semiconductor manufacturing with uncertain demand and correlation in demand considerations" in *International Journal of Production Economics*, 134/2, 2011, pp. 322-332.
3. F. Barahona, S. Bermon, O. Günlük, S. Hood, "Robust capacity planning in semiconductor manufacturing" in *Naval Research Logistics (NRL)*, 52/5, 2005, pp.459-468.
4. M. T. Zhang, S. Niu, S. Deng, Z. Zhang, Q. Li, L. Zheng, "Hierarchical Capacity Planning with Reconfigurable Kits in Global Semiconductor Assembly and Test Manufacturing" in *IEEE Transactions on Automation Science and Engineering*, 4/4, 2007, pp. 543-557.
5. J. -D. Leu, F. -P. Liu, "Capacity Planning of the Semiconductors Manufacturing Supply Chain: A Decision Method and Application", *2022 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM)*, Kuala Lumpur, Malaysia, 2022, pp. 832-836.

9.0 ABOUT THE AUTHORS

Fernando P. Elizaga Jr. is currently a senior manager from the Global Industrial Engineering of Head Backend Operations in Western Digital. He received his Bachelor of Science in Electrical Engineering from Mindanao State University Iligan in 2000, and his master's degree in industrial engineering from Dela Salle University Manila in 2012. He is also a certified lean facilitator from the HGST Global Process Excellence.

Jonathan S. Legarde is a full-stack web developer from the Data Management Engineering at Western Digital Corporation. He received his Bachelor of Science in Information Technology major in Computer Programming from Quezon City University in 2014.

Snehalata More is a data scientist in the Advanced Analytics Office (AAO) at Western Digital. She holds a master's degree in industrial engineering from Texas A&M University. Her work focuses on capacity planning and factory optimization, leveraging operations research and data analytics to help businesses optimize performance and make data-informed strategic decisions.