Stochastic Optimization of Product-Machine Qualification in a Semiconductor Back-end Facility

Mengying Fu, Ronald Askin, John Fowler, Muhong Zhang

School of Computing, Informatics, and Systems Engineering,

Arizona State University, Tempe, AZ 85287, USA

Abstract

In order to process a product in a semiconductor back-end facility, a machine needs to be qualified first by having product-specific software installed and then running test wafers through it to verify that the machine is capable of performing the process correctly. In general, not all machines are qualified to process all products due to the high machine qualification cost and tool set availability. The machine qualification decision affects future capacity allocation in the facility and subsequently affects daily production schedules. To balance the tradeoff between current machine qualification costs and future potential backorder costs due to not enough machines qualified with uncertain demand, a stochastic product-machine qualification optimization model is proposed in this paper. The L-shaped method and acceleration techniques are proposed to solve the stochastic model. Computational results are provided to show the necessity of the stochastic model and the performance of different solution methods.

Key words: manufacturing; product-machine qualification; production planning and scheduling; stochastic programming

1 Introduction

The semiconductor manufacturing process consists of two main parts: the front-end process and the back-end process. The front-end process, also known as wafer fabrication, typically has a small number of products and very complex reentrant product flow. In contrast, the back-end process, also known as assembly and test, typically has hundreds or thousands of different products and relatively linear product flow. The research presented in this paper focuses on the back-end process. In a semiconductor back-end facility, each machine has to be configured for each of the products it will process in the future. This configuration (machine qualification) process includes installing and testing a software program for each product on the machine. Due to the wide product mix, if all machines were to be qualified for all products, the machine qualification process could take considerable time and engineering resources, thus incurring a high machine qualification cost. Meanwhile, not all machines are technologically capable of being qualified for all products. Because of short product life cycles and fast development of new products in the semiconductor industry, new machines may need to be procured frequently for new products. As a result, machines that perform the same operation could belong to different machine types/generations, with each type/generation only being able to be qualified for a subset of products. In addition, the product-machine qualification decision affects the capacity planning decision and subsequently the future daily production schedule. Poor product-machine qualification decisions could cause shortages by not qualifying enough machines for a given product, or machine utilization imbalance by qualifying too many products on a small subset of machines. Overqualification may also complicate scheduling decisions and lead to misallocation of capacity. In this paper, a mixed integer linear programming model (MILP) is first proposed to minimize product-machine qualification cost while considering future production scheduling. As the last part of the semiconductor manufacturing system, on time delivery of customer orders is generally the most important goal for the back-end process. Hence the objective of the MILP is set to minimize the weighted product-machine qualification costs and future backorder costs with a higher weight on the latter. Due to computational limitations and demand forecast data availability, the production scheduling horizon in the model is set to be a medium term (e.g. several weeks). In addition, the product demand is represented by a random distribution to reflect the uncertainty.

The remainder of the paper is organized as follows. Section 2 is a literature review about product-machine qualification. In Section 3, the problem is clearly defined and a mixed integer linear programming model (MILP) is proposed to optimize product-machine qualification in the semiconductor back-end facility. In Section 4, a stochastic MILP model is presented to account for the demand uncertainty in the production scheduling process. The L-shaped method and acceleration techniques are proposed to solve the stochastic model. This is followed by Section 5, in which computational results are presented to compare the deterministic and stochastic models as well as different solution methods of the stochastic model. Finally, conclusions and future research directions are provided in Section 6.

2 Literature Review

Product-machine or operation-machine qualification is a very common feature in the modern semiconductor manufacturing process. A few papers consider this feature in their scheduling models [9, 12, 5, 14, 17, 18], but none of them proposes to change or optimize the current machine qualification. There are also some other papers that utilize short-term machine dedication to schedule the production activities [6, 4]. An operation-machine qualification management system is proposed by [11] for a semiconductor front-end facility, in which four flexibility measures are developed to evaluate different operation-machine qualifications. The impacts of different operation-machine qualifications, with different scores according to the four flexibility measures, on production scheduling are shown through simulation. [1] present a mixed integer linear programming model (MILP) for the product-machine qualification optimization of parallel multi-purpose machines. The objective is to minimize machine configuration costs while obtaining a load-balanced capacity allocation. The MILP formulation is proved to be strongly NP-hard but could be relaxed to a transportation problem under certain assumptions. [15] presents a robustness measure for the multi-purpose machine configuration model developed by [1]. Maximal disturbance of the demand that changes the optimal configuration is used as the robustness measure. [10] proposes a binary optimization model for the operation-machine qualification of photolithography machines in a wafer fabrication factory. The objective is to obtain a load-balanced schedule at minimal machine qualification costs. The cycle time in the factory is shown to be decreased using the binary optimization model compared to machine qualifications developed by heuristic or "educated guess" means. In somewhat related work, [8] propose an integer programming model for long-term employee staffing based on qualification profiles. The objective is to accomplish all tasks with minimal total employment costs. Employee scheduling could be another application area of the methodologies developed for the machine qualification management in the factory.

None of these papers integrates the future production planning and scheduling of a multi-stage manufacturing system explicitly in their machine qualification optimization models. On the other hand, machine qualification decisions have a critical long-term impact on the future production planning and scheduling. Furthermore, the interaction between qualification decisions for different stages impacts delivery performance. In this paper, a stochastic mixed integer linear programming model is proposed to optimize product-machine qualification in a multi-stage manufacturing system while considering future production scheduling with demand uncertainty. In the following section, we define the problem first and then propose a deterministic model.

3 Problem Statement

The back-end facility has multiple stages and parallel machines at each stage. Products are processed in lots with a product-specific number of units in each lot. Setup times are typically sequence-dependent and not included in the lot processing time. However, for simplicity and computational purposes, in this paper, the setup times are not considered explicitly in the model. Instead, the setup times are modeled by decreasing the machine capacity by a certain percentage based on historical machine utilization data. Product-machine qualification is determined in the model, and thus only qualified machines can process a given product at a given stage. Initial product-machine qualification in the model could be empty or given by an existing configuration. In the semiconductor industry, once a machine is qualified for one product, it will not be de-qualified for extra cost. Therefore, in this model, no de-qualification is allowed. The objective of the model is to balance machine qualification costs and future backorder costs. The time horizon of future production scheduling in the model is limited to a medium term (i.e. a couple of weeks). The scheduling horizon is divided into small time buckets to model the movement of lots between stages. Meanwhile, the production quantity of each product on each machine will be scheduled for each time bucket and may be partial lots due to the assumption of continuous production. We assume that all the machine qualifications are finished at the beginning of production periods. A mixed integer linear programming (deterministic) model is proposed in this section.

The definition and notation of the elements for the deterministic machine qualification optimization (***D*-MQO**) model are listed below.

Notation:

P: number of products, with index p

N_p: number of stages for product p, with index n

$M[n]$: number of unrelated machines at stage n, with index m

T: number of time periods in the production scheduling horizon, with index t

C: capacity in minutes of a machine in each time period ($C_{n,m,t}$ if it is machine , stage, and time period dependent)

A: available percentage of machine capacity in each time period ($1-A$ percent of machine capacity is reserved for setup and downtime activities)

$B_{p,0}$: initial back order quantity of product family p

$I_{p,n,0}$: initial inventory of product p at (after) stage n

b_p: backorder cost per lot per time period for product p

$d_{p,t}$: demand quantity for product p at the end of time period t in lots

$t_{p,n,m}$: lot processing time of product p on machine m at stage n

$c_{p,n,m}$: cost of qualifying machine m at stage n for product p

S_Q : a set of (p,n,m)'s with machine m at stage n initially qualified for product p

$\overline{S_Q}$: the complement of set S_Q

Decision Variables:

$X_{p,n,m,t} \in \mathbb{R}^+$: production quantity for product p in time period t on machine m at stage n

$I_{p,n,t} \in \mathbb{R}^+$: inventory quantity of product p at the end of time period t after stage n

$B_{p,t} \in \mathbb{R}^+$: back order quantity of the product p at the end of time period t

$Q_{p,n,m} \in \mathbb{B}$: 1 if machine m at stage n is recommended to be qualified for product p, 0 otherwise

Deterministic Machine Qualification Optimization Model (*D*-MQO)

$$min \quad \sum_{(p,n,m)\in\overline{S_Q}} c_{p,n,m}Q_{p,n,m} + \sum_{p,t} b_p B_{p,t} \tag{1}$$

$$
\begin{aligned}
s.t.\quad & I_{p,n,t-1} + \sum_m X_{p,n,m,t} - \sum_m X_{p,n+1,m,t} = I_{p,n,t},\ \forall\ p, n < N_p, t && (2)\\
& I_{p,N_p,t-1} - B_{p,t-1} + \sum_m X_{p,N_p,m,t} - d_{p,t} = I_{p,N,t} - B_{p,t},\ \forall\ p, t && (3)\\
& \sum_m X_{p,n+1,m,t} \leq I_{p,n,t-1},\ \forall\ p, n < N_p, t && (4)\\
& \sum_p t_{p,n,m} X_{p,n,m,t} \leq C \cdot A,\ \forall\ n, 1 \leq m \leq M[n], t && (5)\\
& t_{p,n,m} X_{p,n,m,t} \leq C Q_{p,n,m},\ \forall\ p, n, m, t && (6)\\
& Q_{p,n,m} = 1,\ \forall\ (p, n, m) \in S_Q && (7)\\
& X_{p,n,m,t}, I_{p,n,t}, B_{p,t} \in \mathbb{R}^+,\ \forall\ p, n, m, t && (8)\\
& Q_{p,n,m} \in \mathbb{B},\ \forall\ p, n, m && (9)
\end{aligned}
$$

The objective (1) is to minimize the total machine qualification and backorder costs. Constraints (2) are the inventory balance constraints for every product at every stage, except for the last stage, in each time period. They indicate that the inventory quantity at the end of period t must be equal to the beginning inventory plus production at stage n in period t minus consumption at the next stage $n+1$ in period t. Constraints (3) are the inventory balance constraints for every product at the last stage in each time period. They are similar to constraints (2) except that the consumption at the next stage $n+1$ in period t is replaced by demand at the end of period t. Backorders are allowed but incur cumulative backorder costs as shown in the objective expression (1). Constraints (4) are the material availability constraints, which state that the production quantity at stage n in period t must be less than the inventory quantity at the previous stage $n-1$ at the end of period $t-1$. If a lot can flow through more than one stage in one time period, the right hand sides of constraints (4) can be expanded to include production at one or more prior stages. Constraints (5) are the capacity constraints for every machine in each time period, which state that the total production time over all products must be less than the available machine capacity after setup and downtime reservations. Constraints (6) are the machine qualification constraints, which state that production quantity $X_{p,n,m,t}$ is zero unless machine m at stage n is recommended to be qualified for product p. Constraints (7) define the initial qualification for machine m at stage n already qualified for product p. Constraints (8) and (9) are the positive and binary constraints for decision variables, respectively.

The model could be easily extended to include different process routes for different products and material handling time between stages by slightly modifying the subscripts. For example, instead of $X_{p,n+1,m,t}$, $X_{p,n+2,m,t}$ should be used in constraints (2) and (4) if product p skips stage n+1. If there is more than one operation performed at one stage, the stage subscript n can be substituted by operation subscript o in constraints (2), (3), and (4). Then in constraints (5) and (6), all the operations that could be performed on machine m at stage n should be considered in the left hand side. The material handling time for product p between stage n and stage n+1 is added on the subscript t of all $X_{p,n+1,m,t}$'s in constraints (2) and (4). If only bottleneck stages are modeled in the above formulation, which is possible when there are too many non-bottleneck stages in the manufacturing system, the material handling time can be further extended to include product-dependent delay times at non-bottleneck stages.

In the objective function (1), the total machine qualification cost is a one-time cost and the total backorder cost over the production scheduling period (e.g. a week) actually represents recurring costs. In addition, since our most important goal is to satisfy all demand, with minimizing machine qualification costs being the secondary objective, the machine qualification cost rates $c_{p,n,m}$'s are set to be very small compared to the backorder cost rates b_p's. In an alternative formulation, we may limit the total backorder cost $\sum_{p,t} b_p B_{p,t}$ to a constant in the constraints and minimize machine qualification cost. With the alternative formulation, we could generate the Pareto optimal frontier between the total backorder cost limit and the total machine qualification cost.

The medium-term production scheduling considered in the above formulation is a snapshot of future production scheduling. Therefore it should reflect a steady state of the production system. If we start with an empty system in the above formulation, the start-up effect could give us a non-optimal machine qualification for future steady state production scheduling. As a result, Little's law [13] is used to estimate initial inventory quantities in the above formulation in a steady state system:

$$I_{p,n,0} = \bar{t}_{p,n} \cdot \bar{d}_p, \ \forall p, n \tag{10}$$

where $I_{p,n,0}$ is the initial inventory of product p at (after) stage n, $\bar{t}_{p,n}$ is the average lot processing time of product p at stage n, and $\bar{d}_p$ is the average demand rate of product p. Average waiting time could be included in $\bar{t}_{p,n}$ if desired. To keep the production system in steady state, the ending inventory quantities at all stages should be greater than or equal to the corresponding starting inventory quantities or otherwise defined minimum. Therefore the following constraints should be added to the formulation during realization.

$$I_{p,n,T} \geq I_{p,n,0}, \ \forall p, n \tag{11}$$

In the above deterministic model, the demand quantities $d_{p,t}$'s used in the production scheduling are assumed to be certain at the time when the machine qualification decisions are made. However, the demand quantities are usually based on forecasts and thus uncertain in real world. Therefore, a stochastic model is proposed in the following section to consider demand uncertainty.

4 Stochastic Machine Qualification Optimization Model ($\boldsymbol{S}$-MQO)

Machine qualification is usually a long term factory configuration decision which incurs nonnegligible time and monetary costs. It affects capacity allocation and thus daily production schedules directly. In our model, the machine qualification decisions are integrated with medium term production scheduling. The objective is to minimize the total machine qualification costs and backorder costs. Since the demand data used in the production scheduling are uncertain, a stochastic machine qualification optimization model is proposed in this section with the objective of minimizing total machine qualification costs and expected backorder costs. The purpose of this stochastic model is to find a robust product-machine qualification matrix at minimal qualification cost. Cost parameters need to be assigned to machine qualification operations executed now and backorders that occur during the future planning horizon. Those parameters should be determined carefully considering that minimizing backorders is the primary objective and minimizing qualification costs is the secondary objective.

A two-stage stochastic machine qualification model is presented below. The demand is represented by a random vector $\xi = (d_{0,0}, ..., d_{P,T})^T$, with $d_{p,t}$ being the demand quantity of product p in period t. The objective (12) is to minimize the summation of total machine qualification costs $\sum_{(p,n,m)\in\overline{S_Q}} c_{p,n,m} Q_{p,n,m}$ and expected total backorder costs $\mathbb{E}[O(X, I, B, \xi)]$ over all possible demand scenarios.

$$\begin{aligned} min \quad & \sum_{(p,n,m)\in\overline{S_Q}} c_{p,n,m}Q_{p,n,m} + \mathbb{E}[O(X,I,B,\xi)] && (12)\\ s.t. \quad & Q_{p,n,m} = 1,\ \forall\ (p,n,m) \in S_Q && (13)\\ & Q_{p,n,m} \in \mathbb{B},\ \forall\ p,n,m && (14) \end{aligned}$$

$O(X,I,B,\xi)$ is the optimal value of the following production scheduling subproblem given a machine qualification matrix Q and a demand scenario ξ_s:

$$\begin{aligned} min \quad & \sum_{p,t} b_p B_{p,t} && (15)\\ s.t. \quad & I_{p,n,t-1} + \sum_m X_{p,n,m,t} - \sum_m X_{p,n+1,m,t} = I_{p,n,t},\ \forall\ p, n < N_p, t && (16)\\ & I_{p,N_p,t-1} - B_{p,t-1} + \sum_m X_{p,N_p,m,t} - d_{p,t}(\xi_s) = I_{p,N_p,t} - B_{p,t},\ \forall\ p,t && (17)\\ & \sum_m X_{p,n+1,m,t} \leq I_{p,n,t-1},\ \forall\ p, n < N_p, t && (18)\\ & I_{p,n,T} \geq I_{p,n,0},\ \forall p,n && (19)\\ & \sum_p t_{p,n,m} X_{p,n,m,t} \leq C \cdot A,\ \forall\ n,m,t && (20)\\ & t_{p,n,m} X_{p,n,m,t} \leq C Q_{p,n,m},\ \forall\ p,n,m,t && (21)\\ & X_{p,n,m,t}, I_{p,n,t}, B_{p,t} \in \mathbb{R}^+,\ \forall\ p,n,m,t && (22) \end{aligned}$$

The first-stage decision variables $Q_{p,n,m}$'s are determined before the realization of random demand vector ξ. The second-stage decision variables $X_{p,n,m,t}$'s, $I_{p,n,t}$'s, and $B_{p,t}$'s are determined based on the first-stage decision and the realized demand vector ξ.

For the ease of reading, we list the additional notation of the stochastic models as follows.

Additional Notation:

ξ_s: demand scenario, with index s; all the notations $\bullet$ defined in the deterministic model depending on the scenario are represented as $\bullet(\xi_s)$

$E^k_{p,n,m}$: cut coefficient of $Q_{p,n,m}$ generated in the L-shape method for iteration $k-1$

e^k: the constant term of the cut generated in the L-shape method for iteration $k-1$

Additional Decision Variables

θ: upper bound variable of backorder cost in the L-shape method

$\gamma(\xi_s)$, $\mu(\xi_s)$, $\sigma(\xi_s)$, $\varphi(\xi_s)$, $\pi(\xi_s)$, $\rho(\xi_s)$: dual variables of the subproblems of scenario ξ_s in the L-shape method

4.1 Deterministic Equivalent Formulation

If the random demand vector ξ can be represented or approximated by a discrete distribution with possible demand scenarios $(\xi_1, ..., \xi_S)$ and associated probabilities $(P(\xi_1), ..., P(\xi_S))$, the previous two-stage stochastic model could be rewritten as the following deterministic equivalent formulation. $X_{p,n,m,t}(\xi_s)$'s, $I_{p,n,t}(\xi_s)$'s, $B_{p,t}(\xi_s)$'s are the second-stage decision variables for demand scenario ξ_s.

$$min \quad \sum_{(p,n,m)\in\overline{S_Q}} c_{p,n,m}Q_{p,n,m} + \sum_{p,t,s} P(\xi_s)b_p B_{p,t}(\xi_s) \tag{23}$$

$$s.t. \quad I_{p,n,t-1}(\xi_s) + \sum_m X_{p,n,m,t}(\xi_s) - \sum_m X_{p,n+1,m,t}(\xi_s) = I_{p,n,t}(\xi_s), \ \forall \ p, n < N_p, t, s \tag{24}$$

$$I_{p,N_p,t-1}(\xi_s) - B_{p,t-1}(\xi_s) + \sum_m X_{p,N_p,m,t}(\xi_s) - d_{p,t}(\xi_s) = I_{p,N_p,t}(\xi_s) - B_{p,t}(\xi_s), \ \forall \ p, t, s \tag{25}$$

$$\sum_m X_{p,n+1,m,t}(\xi_s) \leq I_{p,n,t-1}(\xi_s), \ \forall \ p, n < N_p, t, s \tag{26}$$

$$I_{p,n,T}(\xi_s) \geq I_{p,n,0}(\xi_s), \ \forall p, n, s \tag{27}$$

$$\sum_p t_{p,n,m}X_{p,n,m,t}(\xi_s) \leq C \cdot A, \ \forall \ n, m, t, s \tag{28}$$

$$t_{p,n,m}X_{p,n,m,t}(\xi_s) \leq CQ_{p,n,m}, \ \forall \ p, n, m, t, s \tag{29}$$

$$Q_{p,n,m} = 1, \ \forall \ (p, n, m) \in S_Q \tag{30}$$

$$X_{p,n,m,t}(\xi_s), I_{p,n,t}(\xi_s), B_{p,t}(\xi_s) \in \mathbb{R}^+, \ \forall \ p, n, m, t, s \tag{31}$$

$$Q_{p,n,m} \in \mathbb{B}, \ \forall \ p, n, m \tag{32}$$

By solving this deterministic equivalent formulation, an optimal solution to the two-stage stochastic optimization problem (***S*-MQO**) can be obtained. The deterministic equivalent formulation is a mixed integer linear program. As a result, when there are a large number of demand scenarios, products, or machines, the deterministic equivalent formulation can be very difficult to solve. The L-shaped method and acceleration techniques are thus proposed to solve the ***S*-MQO** model for large problem instances.

4.2 L-Shaped Method

The extensive form of the deterministic equivalent formulation has a block structure. Taking the dual of the extensive form, we can obtain a dual block-angular structure. Therefore, it is natural to exploit Dantzig-Wolf decomposition [7] on the dual or Bender's decomposition [2] on the primal. [16] extend this method to take care of feasibility in stochastic programming, which is now called the L-shaped method. The classic L-shaped method was first developed only for stochastic linear programs. A valid set of feasibility cuts and optimality cuts is known to exist in the continuous case, based on duality theory in linear programming. This knowledge forms the basis of the classic L-shaped method. Those cuts can also be used in the case where only some first-stage variables are integers, e.g. the ***S*-MQO** model. The L-shaped method has been extended to stochastic integer programs. The integer L-shaped method is the integration of the classic L-shaped method and branch-and-bound, during which optimality and feasibility cuts are added to LP relaxations. Since the ***S*-MQO** has binary first-stage variables and continuous second-stage variables, the classic L-shaped decomposition algorithm is chosen instead of the integer L-shaped method. The L-shaped method is briefly described below as it applies to our problem.

Algorithm: L-Shaped Method

Step 0 Set lower bound $LB = -\infty$ and upper bound $UB = \infty$. Set the iteration count $i = 0$. Set δ.

Step 1 Solve the master problem for an optimal solution Q^i

$$\begin{aligned} LB = min \sum_{(p,n,m)\in\overline{S_Q}} & c_{p,n,m}Q_{p,n,m} + \theta \\ s.t.\ & Q_{p,n,m} = 1,\ \forall\ (p,n,m) \in S_Q \\ & Q_{p,n,m} \in \mathbb{B},\ \forall\ p,n,m \\ & \theta \geq \sum_{p,n,m} E^k_{p,n,m}Q_{p,n,m} + e^k, k = 1,2,...,i \end{aligned}$$

Step 2 For $s = 1, ..., S$, solve the following subproblem corresponding to Q^i and ξ_s

$$O(Q^i, \xi_s) = min \sum_{p,t} b_p B_{p,t}$$ Dual variables

$$\begin{aligned}
s.t.\ & I_{p,n,0} + \sum_m X_{p,n,m,1} - \sum_m X_{p,n+1,m,1} = I_{p,n,1}, \forall\ p, n < N_p && (\gamma_{p,n}(\xi_s)) \\
& I_{p,n,t-1} + \sum_m X_{p,n,m,t} - \sum_m X_{p,n+1,m,t} = I_{p,n,t}, \forall\ p, n < N_p, 1 < t \leq T && \\
& I_{p,N_p,t-1} - B_{p,t-1} + \sum_m X_{p,N_p,m,t} - d_{p,t}(\xi_s) = I_{p,N_p,t} - B_{p,t},\ \forall\ p, t && (\mu_{p,t}(\xi_s)) \\
& \sum_m X_{p,n+1,m,1} \leq I_{p,n,0},\ \forall\ p, n < N_p && (\sigma_{p,n}(\xi_s)) \\
& \sum_m X_{p,n+1,m,t} \leq I_{p,n,t-1},\ \forall\ p, n < N_p, 1 < t \leq T && \\
& I_{p,n,T} \geq I_{p,n,0},\ \forall p, n && (\varphi_{p,n}(\xi_s)) \\
& \sum_p t_{p,n,m} X_{p,n,m,t} \leq C \cdot A,\ \forall\ n, m, t && (\pi_{n,m,t}(\xi_s)) \\
& t_{p,n,m} X_{p,n,m,t} \leq C Q^i_{p,n,m},\ \forall\ p, n, m, t && (\rho_{p,n,m,t}(\xi_s)) \\
& X_{p,n,m,t}, I_{p,n,t}, B_{p,t} \in \mathbb{R}^+,\ \forall\ p, n, m, t &&
\end{aligned}$$

If $\sum_{(p,n,m)\in\overline{S_Q}} c_{p,n,m} Q^i_{p,n,m} + \sum_s P(\xi_s) O(Q^i, \xi_s) < UB$, update the upper bound.

Step 3 If $(UB - LB)/LB < \delta$, stop and return $Q = \{Q^i\}$ as the optimal solution and UB as the optimal objective value.

Step 4 For each $s = 1, 2, ..., S$, compute the cut coefficients

$$E^{i+1}_{p,n,m} = \sum_s P(\xi_s) (\sum_t \rho_{p,n,m,t}(\xi_s) \cdot C_{n,m,t})$$

and

$$\begin{aligned}
e^{i+1} = & \sum_s P(\xi_s)[- \sum_{p,n<N_p} I_{p,n,0} \cdot \gamma_{p,n}(\xi_s) + \sum_{p,n<N_p} I_{p,n,0} \cdot \sum_{p,t} \sigma_{p,n}(\xi_s) \\
& + \sum_{p,n} I_{p,n,0} \cdot \varphi_{p,n}(\xi_s) + \sum_p \mu_{p,1}(\xi_s) \cdot (d_{p,1}(\xi_s) - I_{p,N_p,0}) \\
& + \sum_{p,t>1} \mu_{p,t}(\xi_s) \cdot d_{p,t}(\xi_s) + \sum_{n,m,t} \pi_{n,m,t}(\xi_s) \cdot C \cdot A].
\end{aligned}$$

Update $i = i + 1$ and go to Step 1.

In the L-shaped method, the master problem solved in Step 1 provides a lower linear approximation for the function $\sum_s P(\xi_s) O(Q, \xi_s)$ through a continuous variable θ and optimality cuts $\theta \geq \sum_{p,n,m} E^k_{p,n,m} Q_{p,n,m} + e^k$, and therefore a lower bound LB for the objective function (23). The optimal solution Q^i obtained through the master program corresponds to a feasible solution for the stochastic program. It should be noted that in the first iteration $i = 0$, neither θ nor any optimality cut is included in the master problem. In Step 2, all S subproblems are solved using the optimal Q^i obtained from the master problem and corresponding demand scenario ξ_s. These S linear programs are solved independently, allowing for a computationally convenient decomposition or parallelization. If all S subproblems are feasible, which in our case is always true since backorders are allowed in all subproblems, these subproblem solutions together with the master problem solution yield a upper bound UB of the original problem. When the upper bound UB and the lower bound LB are sufficiently close within a preset relative error term δ, we conclude optimality. Otherwise the dual optimal solutions of the subproblems are used to construct an optimality cut added in the master program in the next iteration. Only dual variables corresponding to constraints with positive right-hand-side values or positive coefficients of first-stage variables ($Q_{p,n,m}$'s) will affect the cut coefficients. Those dual variables are represented as the $\gamma_{p,n}$'s, $\mu_{p,t}$'s, $\sigma_{p,n}$'s, $\pi_{n,m,t}$'s, $\varphi_{p,n}$'s, and $\rho_{p,n,m,t}$'s in the parentheses. It should be noted that the initial inventory quantities $I_{p,N_p,0}$'s at/after the last stage are assumed to be zero, because the demand quantities $B_{p,t}$'s can always be adjusted to make $I_{p,N_p,0}$'s zero. In Step 4, according to duality theory the optimality cut $\sum_s P(\xi_s)O(Q, \xi_s) = E^{i+1}_{p,n,m}Q_{p,n,m} + e^{i+1}$ is exact for Q^i and is a lower linear approximate for all other feasible Q's.

In the classic L-shaped method, two types of cuts are added to the master problem: feasibility cuts and optimality cuts. Optimality cuts are computed in the previous algorithm in Step 4. Feasibility cuts are added if and only if the master solution in Step 1 is infeasible for certain subproblems in Step 2. Since backorders are allowed in our model, all feasible master problem solutions are feasible for all the subproblems. As a result, no feasibility cut is added in this algorithm.

4.3 Acceleration of The L-Shaped Method

The number of iterations in the L-shaped method for real world problem instances can be very large. To improve the convergence behavior of the L-shaped method, the following acceleration techniques are proposed.

Cut Disaggregation

In the standard L-shaped method, one optimality cut is added at each iteration, which approximates the expectation of the second-stage objective functions given the current first-stage solution. Instead of one cut, S optimality cuts could be added at each iteration to approximate individual second-stage objective functions per scenario. The optimality cut corresponding to demand scenario ξ_s at iteration i is represented by

$$\theta^s \geq \sum_{p,n,m} E^{s,i}_{p,n,m} Q_{p,n,m} + e^{s,i},$$

in which

$$E^{s,i}_{p,n,m} = \sum_t \rho^i_{p,n,m,t}(\xi_s) \cdot C_{n,m,t}$$

and

$$\begin{aligned} e^{s,i} = & - \sum_{p,n<N_p} I_{p,n,0} \cdot \gamma^i_{p,n}(\xi_s) + \sum_{p,n<N_p} I_{p,n,0} \cdot \sum_{p,t} \sigma^i_{p,n}(\xi_s) \\ & + \sum_{p,n} I_{p,n,0} \cdot \varphi^i_{p,n}(\xi_s) + \sum_p \mu^i_{p,1}(\xi_s) \cdot (d_{p,1}(\xi_s) - I_{p,N_p,0}) \\ & + \sum_{p,t>1} \mu^i_{p,t}(\xi_s) \cdot d_{p,t}(\xi_s) + \sum_{n,m,t} \pi^i_{n,m,t}(\xi_s) \cdot C \cdot A. \end{aligned}$$

In the $(i+1)th$ iteration, the master problem takes the following form.

$$\begin{aligned} min & \sum_{(p,n,m)\in\overline{S_Q}} c_{p,n,m} Q_{p,n,m} + \sum_s P(\xi_s)\theta^s \\ s.t. \ & Q_{p,n,m} = 1, \ \forall \ (p,n,m) \in S_Q \\ & Q_{p,n,m} \in \mathbb{B}, \ \forall \ p,n,m \\ & \theta^s \geq \sum_{p,n,m} E^{s,i}_{p,n,m} Q_{p,n,m} + e^{s,i}, k = 1,2,...,i, s = 1,2,...,S \end{aligned}$$

This approach is referred to as multicut L-shaped algorithm [3]. In the multicut version, there is no information loss due to cut aggregation, thus providing a better approximation of the expectation of second-stage objective functions. Consequently, there are fewer iterations in the multicut L-shaped method. However, since more cuts are added at each iteration, the cost of the multicut algorithm is to solve larger master problems.

Qualification Cuts

In the early iterations of the standard L-shaped method there are very few cuts in the master problem. As a result, a minimal number of machines are qualified in the optimal solutions of the master problem, which results in large backorder quantities at the second-stage subproblems and a large number of iterations. To avoid such poor master problem solutions, information of the second-stage subproblems is integrated in the master problem by adding additional *qualification cuts.* Qualification cuts are added to impose a lower bound restriction on the number of machines to be qualified for each product at each stage.

The following formulation is defined as the single-scenario qualification subproblem for ξ_s $(1 \leq s \leq S)$.

$$\begin{aligned}
min \quad & \sum_{(p,n,m)\in\overline{S_Q}} c_{p,n,m} Q_{p,n,m}(\xi_s) + P(\xi_s) \sum_{p,t} b_p B_{p,t}(\xi_s) \\
s.t. \quad & I_{p,n,t-1}(\xi_s) + \sum_m X_{p,n,m,t}(\xi_s) - \sum_m X_{p,n+1,m,t}(\xi_s) = I_{p,n,t}(\xi_s), \ \forall \ p, n < N_p, t \\
& I_{p,N_p,t-1}(\xi_s) - B_{p,t-1}(\xi_s) + \sum_m X_{p,N_p,m,t}(\xi_s) - d_{p,t}(\xi_s) = I_{p,N_p,t}(\xi_s) - B_{p,t}(\xi_s), \ \forall \ p, t \\
& \sum_m X_{p,n+1,m,t}(\xi_s) \leq I_{p,n,t-1}(\xi_s), \ \forall \ p, n < N_p, t \\
& I_{p,n,T}(\xi_s) \geq I_{p,n,0}(\xi_s), \ \forall p, n \\
& \sum_p t_{p,n,m} X_{p,n,m,t}(\xi_s) \leq C \cdot A, \ \forall \ n, m, t \\
& t_{p,n,m} X_{p,n,m,t}(\xi_s) \leq C Q_{p,n,m}(\xi_s), \ \forall \ p, n, m, t \\
& Q_{p,n,m} = 1, \ \forall \ (p, n, m) \in S_Q \\
& X_{p,n,m,t}(\xi_s), I_{p,n,t}(\xi_s), B_{p,t}(\xi_s) \in \mathbb{R}^+, \ \forall \ p, n, m, t \\
& Q_{p,n,m} \in \mathbb{B}, \ \forall \ p, n, m
\end{aligned}$$

Let the $\bar{B}^s_{p,t}(\xi_s)$'s be the optimal backorder quantities obtained from the single-scenario qualification subproblem for ξ_s $(1 \leq s \leq S)$ and the $\bar{B}^o_{p,t}(\xi_s)$'s be the optimal backorder quantities obtained from the ***S*-MQO** model. When $c_{p,n,m} << P(\xi_s) b_p$ $(\forall p, n, m)$ holds, they must satisfy the following conditions:

$$\sum_{p,t} b_p \bar{B}^s_{p,t}(\xi_s) = \sum_{p,t} b_p \bar{B}^o_{p,t}(\xi_s), \ \forall \ s \tag{33}$$

Note that both $P(\xi_s) \sum_{p,t} b_p \bar{B}^s_{p,t}(\xi_s)$ and $P(\xi_s) \sum_{p,t} b_p \bar{B}^o_{p,t}(\xi_s)$ are equal to the minimal total backorder cost in demand scenario ξ_s given that every machine is qualified for every product. Therefore, if $\bar{Q}^s(\xi_s)$ is the unique optimal machine qualification matrix obtained from the single-scenario qualification subproblem for ξ_s $(1 \leq s \leq S)$ and $\bar{Q}^o$ is an optimal machine qualification matrix obtained from the ***S*-MQO** problem, they must satisfy the following conditions:

$$\sum_m \bar{Q}^o_{p,n,m} \geq \sum_m \bar{Q}^s_{p,n,m}(\xi_s), \ \forall\ p, n, s \tag{34}$$

Conditions (34) hold only when the following two assumptions are both valid: $c_{p,n,m} << P(\xi_s)b_p$ $(\forall p, n, m, s)$ and each single-scenario qualification subproblem has a unique optimal machine qualification matrix. The first assumption $c_{p,n,m} << P(\xi_s)b_p$ $(\forall p, n, m, s)$ holds if the cost parameters $c_{p,n,m}$'s and b_p's are carefully chosen. Because there are usually multiple optimal solutions for real world applications, the second assumption usually does not hold. As a result, adding inequalities (34) in the master problem leads to a sub-optimal solution for the original ***S*-MQO** problem. However, if the first assumption holds, the expected total backorder costs over all scenarios should still be the same with or without inequalities (34). Adding inequalities (34) will decrease the number of iterations in the L-shaped method. Thus the tradeoff here is between the total machine qualification cost and the solution time of L-shaped method. Inequalities (34) are referred to as qualification cuts in this paper.

Relaxed Qualification Cuts

When the problem size increases, even the single-scenario qualification subproblem can be difficult to solve since it is a mixed integer linear program. In such cases, we can solve the LP relaxation of the single-scenario qualification subproblem for an optimal continuous machine qualification matrix $\tilde{Q}^s$. Then a binary machine qualification $\bar{\bar{Q}}^s$ can be obtained using the following rule:

$$\begin{cases} \bar{\bar{Q}}^s = 1, & \tilde{Q}^s > \epsilon \\ \bar{\bar{Q}}^s = 0, & \tilde{Q}^s \leq \epsilon \end{cases}$$

where ϵ is a preset value between 0 and 1. A set of qualification cuts similar to inequalities (34) can be added using $\bar{\bar{Q}}^s$ instead of $\bar{Q}^s$. Those cuts are called relaxed qualification cuts. They require significantly less time for solving the (relaxed) single-scenario qualification subproblems. On the other hand, both optimal machine qualification cost and expected backorder cost with relaxed qualification cuts can be larger than those of the original ***S*-MQO** problem. Therefore, the tradeoff here is still between the solution quality and solution time.

5 Computational Experiments

In this section we will present a numerical experiment solving a 5-product problem instance with the proposed models and solution methods. First, the manufacturing system and demand information are introduced. Then the efficiencies of the two different stochastic solution methods for the ***S*-MQO** model will be discussed and compared using different numbers of scenarios. At the end, the solution quality of stochastic and deterministic models will be evaluated and thus compared through an optimization based scheduling system.

5.1 Data

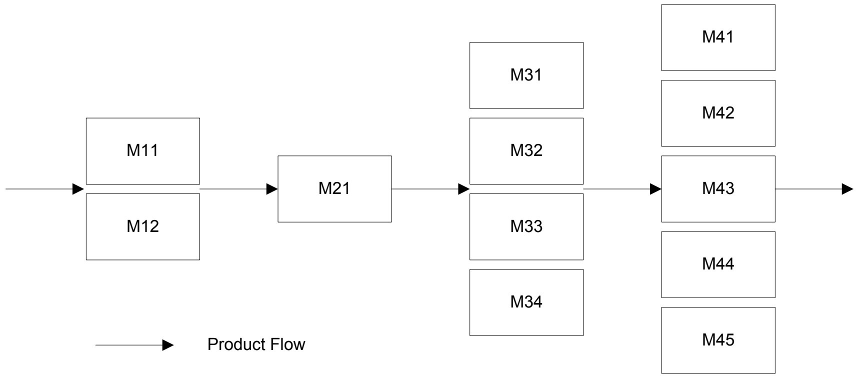


Figure 1: Manufacturing system description

The 5-product problem instance is based on a real semiconductor back-end facility with 4 bottleneck stages as shown in Figure 1. Usually there are 20 to 30 processing stages in a back-end facility. However, including all those stages in the mathematical model results in a significantly larger formulation size. Therefore all the non-bottleneck stages are modeled as constant delays between bottleneck stages, as stated in Section 3. The delay time on a non-bottleneck stage is estimated by the average throughput time at this stage. It is assumed there are multiple identical parallel machines at each stage, as shown in Table 1. Every machine can be qualified to process every product. The production scheduling horizon in the model is chosen to be 1 week, which is divided into 84 2-hr time buckets. All the times used in the experiments are in 2-hr units, e.g. processing time of 1.5 per lot in the experiment represents 3-hour per lot actual processing time. Two processing time distributions are used in the experiment to simulate production systems with approximately 60% and 90% machine utilizations. Processing times of all products at the same stage

Table 1: Manufacturing System Description.

Number of products	5
Num. of bottleneck stages	4
Num. of machines	(2,1,4,5)
Stage 1 Processing Time	$U(1.00, 2.00)$
Stage 2 Processing Time	$U(0.10, 0.20)$
Stage 3 Processing Time	$U(2.00, 4.00)$
Stage 4 Processing Time	$U(2.00, 4.50)$

Table 2: Weekly Demand.

	60% Utilization	90% Utilization
Weekly demand	$U(5, 25)$	$U(5, 35)$
Weekly demand average	15	20
Weekly demand maximal	25	35

are randomly generated based on the same distribution, as shown in Table 1. Although products are allowed to have different processing routes or skip certain stages in the proposed models, all products are assumed to go through all stages in the same linear sequence in the experiment. Customer orders or product types can be assigned with different priorities through their backorder cost rates (per lot per 2-hr time bucket), e.g. important orders or product types with higher backorder cost rates. However in the experiment, all product types and lots are assumed to have the same priority for simplicity, therefore the same backorder cost rate. The initial product-machine qualification matrix is assumed to be empty, with no machine qualified for any product. The weekly demand for future production scheduling is uncertain and randomly generated from a uniform distribution in the experiments as shown in Table 2. A small 5-product problem is design based on the real size 25-product problem to compare two solution methods of the stochastic ***S*-MQO** model. The production system description and weekly demand information for the 5-product problem are shown in Table 1 and Table 2 respectively. The available percentage of machine capacity in each time

Table 3: Size of the deterministic equivalent of the ***S*-MQO** problem.

S	Constraints		Variables	
	Equality	Inequality	Continuous	Binary
1	1,680	7,328	7,140	60
5	8,400	36,640	35,700	60
10	16,800	73,280	71,400	60
20	33,600	146,560	142,800	60

period A is set to be 80% in all cases. The WIP inventory in the system is estimated using Little's Law

$$I_{p,n,0} = \bar{t}_{p,n} \cdot \bar{d}_p, \ \forall p, n.$$

In the experiments, $\bar{t}_{p,n}$ is estimated by the expected processing time of product p at stage n from Table 1, and $\bar{d}_p$ is estimated by the expected demand of product p from Table 2 divided by the total number of periods (84) in a week.

The sizes of the deterministic equivalents of the $\boldsymbol{S}$**-MQO** problem for different S values are given in Table 3. There is a positive linear relationship between the number of constraints and continuous variables and the number of possible scenarios S. Even for a small problem instance with only 5 products and 12 machines, there are 60 binary variables in the formulation. For a typical test facility with 25 aggregated product families and 50 bottleneck-stage machines, there will be 1250 binary variables, thus making it very difficult to solve.

5.2 Performance of Different Solution Methods

In the experiment, two different solution methods for the $\boldsymbol{S}$**-MQO** model are tested. One is to solve the deterministic equivalent formulation (DE). The other is the L-shaped method (Bender). Proposed acceleration techniques of the L-shaped method are also tested, including cut disaggregation (CD), qualification cuts (QC), and relaxed qualification cuts (RQC). Solution times of all tested solution methods are listed in Table 4 and ploted in Figure 2 for different S values and machine utilizations. More details about the solution times of different methods are shown in Table 4, including solution/decomposition (BD) time, time for adding qualification cuts before the decomposition (QC time), number of iterations in the decomposition algorithm, and optimality gap at the end of runtime limit (36000 sec). The L-shaped method with cut disaggregation and relaxed qualification cuts ("Bender + CD + RQC") has the shortest solution times and fewest numbers of iterations. The L-shaped method with cut disaggregation and qualification cuts ("Bender + CD + QC") has relatively few iterations but unstable solution times. It is also noted that the time required for solving single-scenario qualification subproblems (QC time) increases significantly when S increases. As a result, adding qualification cuts is not suitable for real size problem instances. The L-shaped method with cut disaggregation ("Bender + CD") has relatively short solution times but relatively large numbers of iterations, which could make it unsuitable for real size problem instances. All other solution methods have both long solution times and larger number of iterations. The fewer number of iterations means that the cut disaggregation with qualification cuts or the relaxed qualification cuts work well by cutting off infeasible solutions. The use of relaxed qualification cuts sacrifice the solution quality to certain extend, while the computational time is much improved. As shown in Table 5, the total cost with relaxed qualification cuts are increased within 5%, while the computational times are improved largely.

The quality of solutions of different methods are listed in Table 5 for different S values and machine utilizations. Optimal solutions obtained with the first two methods are also optimal for the original $\boldsymbol{S}$**-MQO** model. However, optimal solutions obtained with the last four methods can be sub-optimal to the original $\boldsymbol{S}$**-MQO** model, due to QC/RQC cuts. Both the total qualification costs and the expected total backorder costs are shown in the "Q cost" and "B cost" columns respectively in Table 5. In the experiment, $c_{p,n,m} = 0.1$ $(\forall\ p, n, m)$ and $b_p = 1$ $(\forall\ p)$. From "Bender + CD" to "Bender + QC" or "Bender + CD + QC", the optimal "B cost" does not increase, and the optimal "Q cost" and "Total cost" increase slightly. For "Bender + RQC" and "Bender + CD + RQC", the optimal "B cost", "Q cost" and "Total cost" all increase. This is consistent with the previous analysis. The increase in "B cost" for "Bender + RQC" and "Bender + CD + RQC" is significant when S is 20. The reason is that $c_{p,n,m} < P(\xi_s)b_p$ does not hold anymore when S is 20. Therefore, $c_{p,n,m}$'s and b_p's should be chosen carefully to make sure that $c_{p,n,m} < P(\xi_s)b_p$ is valid if "Bender + CD + QC" is to be implemented. "Bender + CD + RQC" and "Bender + CD" are recommended for large size problem instances because of short solution times and small numbers of iterations. If "Bender + CD" does not find the optimal solution and "Bender + CD + RQC" finds one, thus providing an upper bound of the $\boldsymbol{S}$**-MQO** model, a lower bound can be estimated by the LP relaxation of the original $\boldsymbol{S}$**-MQO** problem.

At the end, the optimal qualification matrices obtained using the L-shaped method with cut disaggregation for different S values and machine utilizations are evaluated using a different set of 20 demand scenarios generated according to the distributions in Table 2. Each demand scenario is given an equal probability of 0.05. A production scheduling linear program is solved for each demand scenario and each optimal qualification matrix. The total qualification cost and expected

Table 4: Solution time comparison of different acceleration methods .

$S = 5$	**60% Utilization**				**90% Utilization**			
	BD time (sec)	QC time (sec)	Iterations	Gap (%)	BD time (sec)	QC time (sec)	Iterations	Gap (%)
Bender	36147		524	41	36143		605	11
Bender + CD	893		197	0	5919		452	0
Bender + QC	22750	3837	2494	0	30006	4960	2019	2
Bender + CD + QC	4315	3819	80	0	6534	4965	95	0
Bender + RQC	1064	3	200	0	6491	5	516	0
Bender + CD + RQC	177	3	28	0	244	4	12	0
$S = 10$	**60% Utilization**				**90% Utilization**			
	BD time (sec)	QC time (sec)	Iterations	Gap (%)	BD time (sec)	QC time (sec)	Iterations	Gap (%)
Bender	36122		349	61	36168		755	17
Bender + CD	2855		183	0	3326		161	0
Bender + QC	36012	13779	1907	31	36015	16290	803	2
Bender + CD + QC	14407	13764	55	0	17176	16286	38	0
Bender + RQC	7524	4	613	0	4235	5	228	0
Bender + CD + RQC	225	6	18	0	506	6	19	0
$S = 20$	**60% Utilization**				**90% Utilization**			
	BD time (sec)	QC time (sec)	Iterations	Gap (%)	BD time (sec)	QC time (sec)	Iterations	Gap (%)
Bender	36751		454	38	36067		314	67
Bender + CD	1924		126	0	4826		182	0
Bender + QC	36018	49315	1426	29	36031	26783	706	8
Bender + CD + QC	1246	50158	49	0	1719	26813	33	0
Bender + RQC	26998	4	946	0	4561	7	134	0
Bender + CD + RQC	450	4	17	0	387	6	11	0

total backorder cost for each optimal qualification matrix are listed in the ”Q cost” and ”B cost” columns of Table 6. Optimal qualification matrices from the deterministic model using the average or maximal demand are listed in the first and second row. For both the 60% and 90% machine utilization cases, the optimal qualification matrices obtained from the stochastic model outperform those obtained from the deterministic model. Not surprisingly, for the stochastic model, the optimal qualification matrix obtained with more demand scenarios also has better performance, because a larger number of demand scenarios provides a better approximation of the original continuous distribution. With the large number of scenarios, the Bender's approach with cut disaggregation and relaxed qualification cuts is preferred for the tradeoff of the computational efforts and solution quality.

6 Conclusion

In this paper, a stochastic mixed integer linear programming model (***S*-MQO**) is proposed to optimize product-machine qualifications for a semiconductor back-end facility. Future production

Table 5: Solution quality comparison of different acceleration methods.

$S = 5$	**60% Utilization**				**90% Utilization**			
	Q cost	B cost	Total cost	Gap(%)	Q cost	B cost	Total cost	Gap(%)
Bender	2	4.2	6.2	41	4.0	17.3	21.3	11
Bender + CD	2.1	2.6	4.7	0	2.6	17.3	19.9	0
Bender + QC	2.1	2.6	4.7	0	2.5	17.7	20.2	2
Bender + CD + QC	2.1	2.6	4.7	0	2.7	17.4	20.1	0
Bender + RQC	2.3	2.5	4.8	0	2.8	17.5	20.3	0
Bender + CD + RQC	2.3	2.6	4.9	0	2.8	18.2	21.0	0
$S = 10$	**60% Utilization**				**90% Utilization**			
	Q cost	B cost	Total cost	Gap(%)	Q cost	B cost	Total cost	Gap(%)
Bender	5.8	0.9	6.7	61	4.6	12.5	17.1	17
Bender + CD	2.3	0.9	3.2	0	2.8	12.7	15.5	0
Bender + QC	3.6	0.9	4.5	31	3.0	12.6	15.6	2
Bender + CD + QC	2.3	0.9	3.2	0	2.9	12.6	15.5	0
Bender + RQC	2.4	0.9	3.3	0	3	12.7	15.7	0
Bender + CD + RQC	2.4	0.9	3.3	0	3.3	12.8	16.1	0
$S = 20$	**60% Utilization**				**90% Utilization**			
	Q cost	B cost	Total cost	Gap(%)	Q cost	B cost	Total cost	Gap(%)
Bender	3.5	0.7	4.2	38	4.5	12.4	16.9	67
Bender + CD	2.3	0.8	3.1	0	2.7	11.7	14.4	0
Bender + QC	3.2	1.1	4.3	29	3.1	11.6	14.7	8
Bender + CD + QC	2.5	0.7	3.2	0	3	11.5	14.5	0
Bender + RQC	2.4	0.8	3.2	0	3.1	17.4	20.5	0
Bender + CD + RQC	2.4	0.9	3.3	0	3.3	17.4	20.7	0

Table 6: Evaluation of different qualification matrices.

S	60% Utilization		90% Utilization	
	Q cost	B cost	Q cost	B cost
1 (avg)	2	8.4	2.2	17.5
1 (max)	1.7	16.3	1.9	39.4
5	2.1	5.5	2.6	16.3
10	2.3	2.8	2.8	14.9
20	2.3	2.9	2.7	14.2

scheduling in a medium term horizon with demand uncertainty is considered. Setup times and downtime are modeled indirectly by using the machine utilization rate from historical data. The model proposed doesn't bias to the different setup sequences for the setup time. Therefore, for the general optimal solution, the setup time has the same distribution from the historical data. Depending on the conservatism, the decision maker may choose different confidence levels of the utilization rate to be used in the model. The L-shaped method and several acceleration techniques are proposed to solve the stochastic model. In the numerical experiments, a 5-product example is used to evaluate different solution methods and their solutions. Bender's Decomposition with Cut Disaggregation and possibly Relaxed Qualification Cuts applied to the stochastic demand formulation are recommended for determining a robust qualification schedule. This approach is shown to have advantaged over deterministic problem formulations.

In this paper, we assume product-machine qualification decisions are made and implemented now for a foreseeable future with stationary demand. The models described in this paper could be readily expanded to include time-phased qualification decisions. An interesting topic for future research will be a multi-stage stochastic model for time-phased qualification decisions.

7 Acknowledgment

Intel Corporation supported this work. We are thankful to Jeffrey Pettinato and Naiping keng at Intel for their helpful feedback.

References

[1] A. Aubry, A. Rossi, M.L. Espinouse, and M. Jacomino. Minimizing setup costs for parallel multi-purpose machines under load-balancing constraint. *European Journal of Operational Research*, 187(3):1115–1125, 2008.

[2] J.F. Benders. Partitioning procedures for solving mixed-variables programming problems. *Numerische Mathematik*, 4(1):238–252, 1962.

[3] J. R. Birge and F. Louveaux. A multicut algorithm for two-stage stochastic linear programs. *European Journal of Operational Research*, 34(3):384–392, 1988.

[4] K.E. Bourland and L.K. Carl. Parallel-machine scheduling with fractional operator requirements. *IIE Transactions*, 26(5):56–65, 1994.

[5] P. Brucker, B. Jurisch, and A. Krämer. Complexity of scheduling problems with multi-purpose machines. *Annals of Operations Research*, 70:57–73, 1997.

[6] G.M. Campbell. Using short-term dedication for scheduling multiple products on parallel machines. *Production and Operations Management*, 1(3):295–307, 1992.

[7] G.B. Dantzig and P. Wolfe. Decomposition principle for linear programs. *Operations research*, 8(1):101–111, 1960.

[8] A. Drexl and M. Mundschenk. Long-term staffing based on qualification profiles. *Mathematical Methods of Operations Research*, 68(1):21–47, 2008.

[9] J. Hurink, B. Jurisch, and M. Thole. Tabu search for the job-shop scheduling problem with multi-purpose machines. *OR Spectrum*, 15(4):205–215, 1994.

[10] J.P. Ignizio. Cycle time reduction via machine-to-operation qualification. *International Journal of Production Research*, 47(24):6899–6906, 2009.

[11] C. Johnzén, P. Vialletelle, S. Dauzère-Pérès, C. Yugma, and A. Derreumaux. Impact of qualification management on scheduling in semiconductor manufacturing. In S.J. Mason, R.R. Hill, L. Monch, T. Jefferson, and J. Fowler, editors, *Proceedings of the 40th Conference on Winter Simulation*, pages 2059–2066. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, 2008.

[12] B. Jurisch. Lower bounds for the job-shop scheduling problem on multi-purpose machines* 1. *Discrete Applied Mathematics*, 58(2):145–156, 1995.

[13] J.D.C. Little. A proof of the queuing formula L=λW. *Operations Research*, 9(3):383–387, 1961.

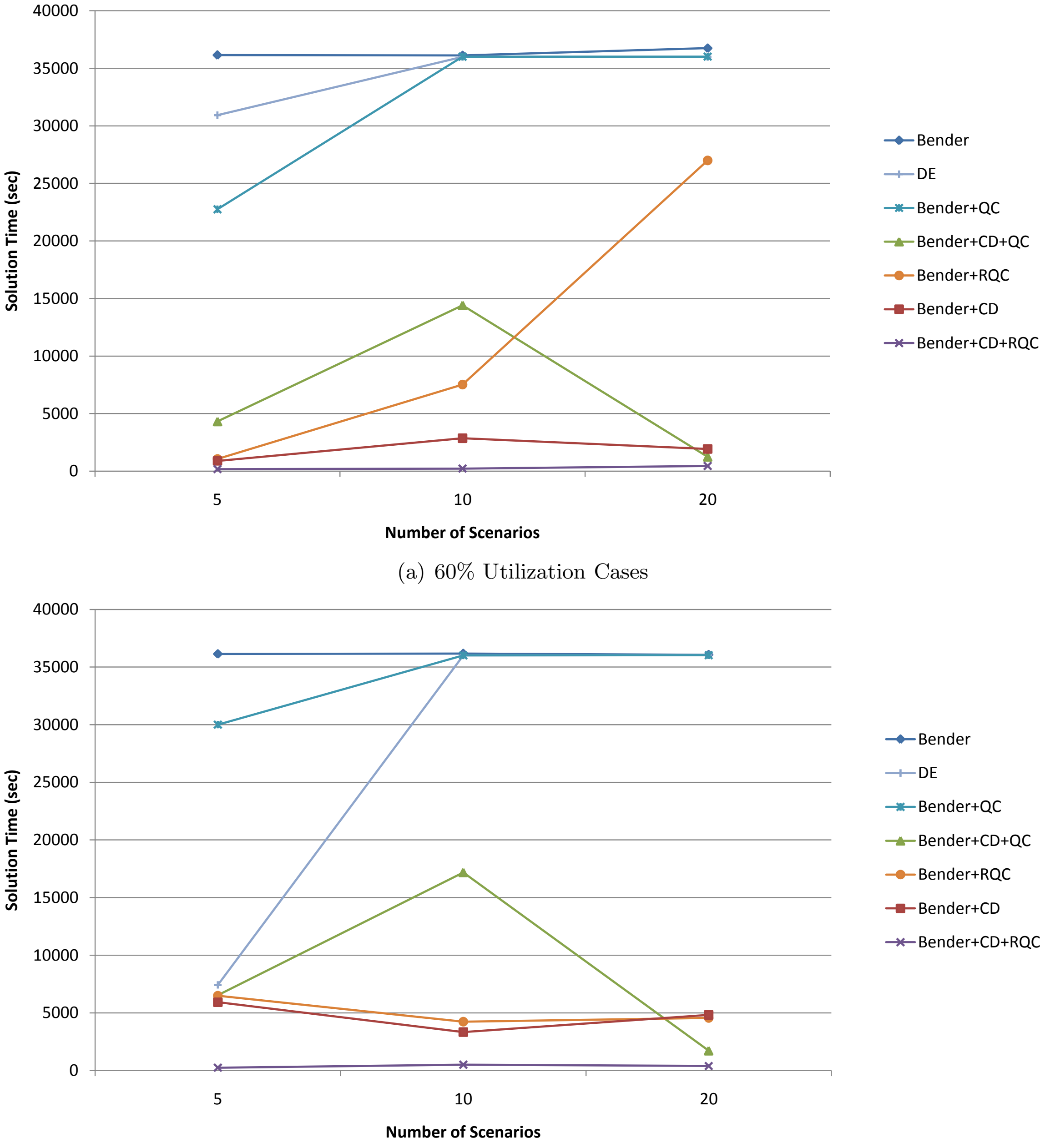
[14] Y. Mati and X. Xie. The complexity of two-job shop problems with multi-purpose unrelated machines. *European Journal of Operational Research*, 152(1):159–169, 2004.

[15] A. Rossi. A robustness measure of the configuration of multi-purpose machines. *International journal of production research*, 48(3-4):1013–1033, 2010.

[16] R.M. Van Slyke and R. Wets. L-shaped linear programs with applications to optimal control and stochastic programming. *SIAM Journal on Applied Mathematics*, pages 638–663, 1969.

[17] M.C. Wu, YL Huang, YC Chang, and KF Yang. Dispatching in semiconductor fabs with machine-dedication features. *The International Journal of Advanced Manufacturing Technology*, 28(9):978–984, 2006.

[18] M.C. Wu, H. Jiang Jr, and W.J. Chang. Scheduling a hybrid MTO/MTS semiconductor fab with machine-dedication features. *International Journal of Production Economics*, 112(1):416–426, 2008.



(a) 60% Utilization Cases

(b) 90% Utilization Cases

Figure 2: Solution times of different solution methods