Proceedings of the 2023 Winter Simulation Conference
C. G. Corlu, S. R. Hunter, H. Lam, B. S. Onggo, J. Shortle, and B. Biller, eds.

DECENTRALIZED DECISION-MAKING FRAMEWORK FOR MANAGING PRODUCT ROLLOVERS IN THE SEMICONDUCTOR MANUFACTURING

Carlos Leca
Reha Uzsoy

Edward P. Fitts Dept. of Industrial
and Systems Engineering
Campus Box 7906
North Carolina State University
Raleigh, NC 27695-7906, USA

Karl Kempf

Decision Engineering Group
Intel Corporation
5000 W. Chandler Blvd.,
Chandler, AZ 85226, USA

ABSTRACT

Competitiveness in the semiconductor industry requires continuous management of product rollovers, the process of introducing new products and retiring older ones to maintain market share. This paper presents a decentralized decision-making framework to coordinate product rollover decisions using Lagrangian decomposition of a centralized model using quadratic coordination errors in the subproblem objectives, and a decentralized heuristic that recovers the feasible solutions from the relaxed ones obtained from the Lagrangian procedure. Experimental results show that this decentralized framework delivers promising results, obtaining near-optimal solutions in modest CPU times.

1 INTRODUCTION

Semiconductor manufacturing firms are generally organized into Product Divisions (PD), each of which manages the product portfolio for a particular market segment such as mobile devices, servers, etc. Each PD is in charge of developing specifications for new products, forecasting their demand and estimating a target date for their introduction by which time all development activities must be completed. The forecast is presented to Corporate Management (CORP) to request financing. CORP represents the firm's overall interests, receiving the sales revenue and allocating operating budgets among the PDs to maximize corporate profit. The PD's demand forecast also serves to generate manufacturing orders for the firm's Manufacturing Group (MFG), which is responsible for producing the units to be sold in the market in return for payment from the PDs' operating budgets. A product can only be manufactured for sale if its development has been completed by the firm's Product Engineering Group (PEG), which converts the product specifications developed by the PDs into manufacturable circuit designs, for which the PDs pay PEG from their operating budget. Some development activities such as prototype and sample fabrication require manufacturing capacity, for which PEG compensates MFG.

The highly competitive nature and rapid technological innovation in the semiconductor industry require firms to continually introduce new products and retire old ones to maintain and grow market share. We define a *product rollover* as the process by which a firm introduces new products and retires old ones serving the same market segment over time (Billington et al. 1998). Several studies have shown that many product introductions fail, wasting resources that can place the firm in significant financial difficulties (Billington et al. 1998; Erhun et al. 2007). The successful management of product rollovers requires effective coordination of all the different business units involved. Centralized decision-making is neither practical nor desirable, as each unit has its own technological constraints knowledge which are not known outside

979-8-3503-6966-3/23/$31.00 ©2023 IEEE

the unit itself, as well as its own objectives. However, each agent needs resources allocated by another unit in addition to its own local resources. For example, the capacity available for PD A to manufacture its next generation product $A1$ will depend on both how much capacity is allocated to other PDS and PEG, as well as the capacity division A is using to manufacture its current product $A0$ and for the development of subsequent generations of its products. Furthermore, any agreed-upon plan between agents constantly evolves over time due to uncertainty in the production, development, and demand processes. However, this initial work focuses on deterministic problems to develop insight and solution techniques.

We first developed a centralized model (Leca et al. 2021) to capture the complex interactions between mutually dependent units of the firm pursuing different objectives with shared resources. A Lagrangian decomposition approach (Leca et al. 2022) suffered from subproblem degeneracy due to the linear nature of the constraints and linear prices update (Jennergren 1973; Jose et al. 1997; Kutanoglu and Wu 1999), rendering it computationally impractical. Motivated by the augmented tatonnement first proposed by Jennergren (1973) and the later application by Kutanoglu and Wu (1999), we applied an Augmented Lagrangian approach that yielded upper bounds but is no longer separable by agent, defeating our purpose of obtaining a decentralized solution approach. In this paper, we modify the Augmented Lagrangian approach with quadratic terms penalizing the deviation between solutions proposed by different agents. Since the Lagrangian solution is generally not primal feasible, it is necessary to restore the feasibility of the relaxed solution. Thus, we additionally develop a decentralized heuristic to recover a feasible (coordinated) solution from the relaxed solution which can also be used to obtain an initial feasible solution.

2 LITERATURE REVIEW

In the semiconductor industry, supply chain management is challenging due to the technical intricacy of the production process, which often possesses low overall yields compared to other industries (Mönch et al. 2018). Manufacturing complexity increases with the appearance of more advanced products, requiring new manufacturing technologies that bring more uncertainties to the process (Manda and Uzsoy 2020). The production process requires very costly, specialized equipment that is only available from a few suppliers, increasing the lead time to add capacity (Kempf et al. 2013). Kim and Uzsoy (2013) discuss problems arising from the competition between development and production activities for manufacturing capacity, noting the variability of manufacturing throughput and the cycle times due to congestion, equipment failure, and process engineering needs. In this environment, product rollovers can fail for a wide variety of reasons (Erhun et al. 2007), and significant financial investment in the development of an unsuccessful product can put the entire firm at risk (Billington et al. 1998).

Several studies examine capacity allocation in the context of a product rollover. Carrillo and Franza (2006) study the relation between time-to-market and ramp-up time decisions using a non-linear dynamic control model where time is treated as a continuous variable. Shen et al. (2014) consider how a capacity-constrained firm prices products during new product introductions, considering both pricing and capacity. Özer and Uncu (2015) present a dynamic programming model capturing the relationships between time-to-market, sales channels, pricing, and production decisions. Schwarz and Tan (2021) explores how limited production capacity affects the optimal unconstrained decision modifying the rollover strategy. Chung and Jang (2022) develop a stochastic model to analyze how the flow speed of development lots affects starvation of a bottleneck stage and the throughput of regular production lots in a semiconductor fabrication line.

In the papers discussed above, decision-making is centralized, and the firm acts as a single agent that can control all the aspects of the rollover strategy. However, the information and operational capabilities required for its execution lie within different units of the firm, such as PDs, MFG and PEG and are not easily available outside the unit. Hence, a centralized framework is not implementable. Kutanoglu and Wu (1999) were among the first researchers to develop a decentralized resource allocation model in the semiconductor industry in the form of a combinatorial auction for job scheduling. They analyze their mechanism from the perspectives of game theory and Lagrangian relaxation. An important element of this work is an augmented tatonnement to avoid subproblem degeneracy. The subproblem degeneracy problem is discussed by Jennergren (1973) and Jose et al. (1997) in the context of decentralized optimization.

Karabuk and Wu (2002) develop a coordination scheme for a semiconductor firm involving marketing and manufacturing agents, where the firm's objective is the sum of the agents' objectives. They propose a decentralized coordination scheme based on transfer prices that are updated iteratively until the decisions made by the different agents are consistent. They develop their mechanism using an augmented Lagrangian relaxation that avoids linear price updates and subproblem degeneracy but is not separable, preventing decentralization. To solve this problem they use auxiliary problem theory (Cohen 1980) to modify the agent's objective functions and recover separability. We follow the same approach in this paper.

3 CENTRALIZED MODEL

We first present the centralized model in order to describe the constraints and decision variables involved in the different agents' decisions and establish a mathematical point of departure from which alternative decentralized models can be derived. The formulation, whose notation is defined in Table 1, seeks to maximize corporate profit (1) subject to constraints describing the capabilities of each agent (CORP, PD, MFG or PE).

Table 1: Model parameters and decision variables.

Parameters	
ρ_{pit}	Unit selling price of product p of PD i in period t.
c^M_{pit}	Unit production cost of product p of PD i in period t.
c^D_{pit}	Cost of one development stage of product p from PD i at period t.
τ^{tr}_{pit}	Transistor production capacity required to process one unit of product p for PD i in period t.
τ^{mtl}_{pit}	Metal capacity required to process one unit of product p for PD i in period t.
ω^{tr}_{pit}	Transistor capacity required to process one unit of product (p,i,t) in the transistor's development stage.
ω^{mtl}_{pit}	Capacity required to process one unit of product (p,i,t) in the metal's development stage.
C^{tr}_t	Total capacity of transistor manufacturing process.
C^{mtl}_t	Total capacity of the metal manufacturing process.
ε^{tr}_{pit}	Total engineering capacity that product (p,i,t) needs in the transistor's development stage.
ε^{mtl}_{pit}	Total engineering capacity that product (p,i,t) needs in the metal's development stage.
ε^{dbg}_{pit}	Total engineering capacity that product (p,i,t) needs in the debugging development stage.
n^{tr}_{pi}	Number of units used in the transistor development stage.
n^{mtl}_{pi}	Number of units used in the metal development stage.
E_t	Total amount of engineering capacity to work on any of the development stages.
r_{pi}	Number of development cycles required by product p for PD i to complete development.
S	Initial budget available for the corporation at the beginning of the time horizon.
d_{pit}	Demand for product p that belongs to PD i market in period t.
h^{mtl}_{pit}	Inventory holding cost of one unit of final product.
h^{mtl}_{pit}	Inventory holding cost of one work in process unit.
Decision Variables	
OC_{it}	Operation budget allocated by the corporation to PD i in period t.
B^{corp}_t	Available budget at the corporation at period t.
B^{div}_{it}	Available budget at PD i in period t.

x_{pit}	Number of units of final product p from PD i sold to the market at period t.
m_{pit}	Number of units of final product p from PD i produced at period t.
w^{tr}_{pit}	Number of units of product p for PD i processed at the transistor process in period t.
I_{pit}	Inventory of final product (p,i,t) on inventory at the beginning of period t.
I^{mtl}_{pit}	Number of WIP units of product (p,i,t) between transistors and metal process.
z^{tr}_{pit}	Binary variable that takes the value of one if the transistor development stage of product (p,i) was performed during period t.
z^{mtl}_{pit}	Binary variable that takes the value of one if the metal development stage of product (p,i) was performed during period t.
z^{dbg}_{pit}	Binary variable that takes the value of one if the debugging development stage of product (p,i) was performed during period t.
R_{pit}	Binary variable that takes the value of one indicating that one entire cycle of the development process of product p from PD i has finished at the end of period t.

$$\textbf{maximize} \; : \; \left\{ \sum_{\forall p \in \mathbb{P}} \sum_{\forall i \in \mathbb{I}} \sum_{\forall t \in \mathbb{T}} \rho_{pit} x_{pit} \; - \; h^{mtl}_{pit} I_{pit} \; - \; h^{tr}_{pit} w_{pit} \; - \; c^{M}_{pit} m_{pit} \; - \; c^{D}_{pit} (z^{tr}_{pit} \; + \; z^{mtl}_{pit} \; + \; z^{dbg}_{pit}) \right\} \quad (1)$$

subject to:

Corporate Constraints:

$$B^{corp}_{t} = B^{corp}_{t-1} - \sum_{\forall i \in \mathbb{I}} OC_{it} + \sum_{\forall i \in \mathbb{I}} \sum_{\forall p \in \mathbb{P}} \rho_{pit} x_{pit} \qquad \forall t \in \mathbb{T} \qquad (2)$$

$$B^{corp}_{t} \geq 0 \quad ; \quad 0 \leq x_{pit} \leq d_{pit} \qquad \forall i \in \mathbb{I}; \forall p \in \mathbb{P}; \forall t \in \mathbb{T} \qquad (3)$$

CORP assigns each PD an operating budget OC_{it} at the start of each planning period, and receives the revenue $\rho_{pit} x_{pit}$ from that unit at its end. This allows CORP to subsidize temporarily unprofitable PDs to obtain increased revenues in the future. Constraints (2) ensure that the firm's expenditures do not exceed its revenues, while (3) ensure a balanced corporate budget and sales remain bounded by the demand forecast.

Product Division Constraints:

$$B^{div}_{it} = OC_{it} - [\sum_{\forall p \in \mathbb{P}} c^{M}_{pit} m_{pit} + c^{D}_{pit} (z^{tr}_{pit} + z^{mtl}_{pit} + z^{dbg}_{pit})] \qquad \forall i \in \mathbb{I}; \forall t \in \mathbb{T} \qquad (4)$$

$$I_{pit} = I_{p,i,t-1} + m_{pit} - x_{pit} \qquad \forall t \in \mathbb{T} \qquad (5)$$

$$B^{div}_{it} \geq 0 \quad ; \quad I_{pit} \geq 0 \qquad \forall i \in \mathbb{I}; \forall p \in \mathbb{P}; \forall t \in \mathbb{T} \qquad (6)$$

Each PD receives its operating budget (OC_{it}) from CORP at the beginning of each period. From this budget, it must pay MFG to produce its products for sale and PEG for any development work requested per the constraints (4). Each PD is responsible for its products' finished goods inventory holding costs since it provides demand forecasts. Constraints (5) define each PD's finished goods inventory, while (6) enforces budget constraints for all PDs.

Manufacturing Constraints:

$$\sum_{\forall p\in\mathbb{P}}\sum_{\forall i\in\mathbb{I}}\tau^{tr}_{pit}w_{pit}+\sum_{\forall p\in\mathbb{P}}\sum_{\forall i\in\mathbb{I}}n^{tr}_{pi}\omega^{tr}_{pit}z^{tr}_{pit}\le C^{tr}_t \qquad \forall t\in\mathbb{T} \tag{7}$$

$$\sum_{\forall p\in\mathbb{P}}\sum_{\forall i\in\mathbb{I}}\tau^{mtl}_{pit}m_{pit}+\sum_{\forall p\in\mathbb{P}}\sum_{\forall i\in\mathbb{I}}n^{mtl}_{pi}\omega^{mtl}_{pit}z^{mtl}_{pit}\le C^{mtl}_t \qquad \forall t\in\mathbb{T} \tag{8}$$

$$m_{pit}\le I^{mtl}_{pit} \qquad \forall i\in\mathbb{I};\forall p\in\mathbb{P};\forall t\in\mathbb{T} \tag{9}$$

$$I^{mtl}_{p,i,t-1}+w_{pit}-m_{pit}=I^{mtl}_{pit} \qquad \forall i\in\mathbb{I};\forall p\in\mathbb{P};\forall t\in\mathbb{T} \tag{10}$$

$$M[\sum_{t_p=0}^{t_p=t}R_{pit_p}]\ge w_{pit} \qquad \forall i\in\mathbb{I};\forall p\in\mathbb{P};\forall t\in\mathbb{T} \tag{11}$$

At the beginning of each period, MFG receives payment for production activities, which is used to pay for its own operations. The production process has two stages, transistor fabrication and metal fabrication, each requiring one planning period to complete. The primary constraints for MFG are capacity constraints for each processing stage (7) and (8), consistency of shipments with available material (9), and material balance constraints across planning periods (10). Constraints (11) ensure that MFG can only produce a product for sale if the product has completed development.

Product Engineering Constraints:

$$\sum_{\forall p\in\mathbb{P}}\sum_{\forall i\in\mathbb{I}}\varepsilon^{tr}_{pit}z^{tr}_{pit}+\sum_{\forall p\in\mathbb{P}}\sum_{\forall i\in\mathbb{I}}\varepsilon^{mtl}_{pit}z^{mtl}_{pit}+\sum_{\forall p\in\mathbb{P}}\sum_{\forall i\in\mathbb{I}}\varepsilon^{dbg}_{pit}z^{dbg}_{pit}\le E_t \qquad \forall t\in\mathbb{T} \tag{12}$$

$$\sum_{t_p=0}^{t_p=t}z^{dbg}_{pit_p}\ge r_{pi}R_{pit} \qquad \forall i\in\mathbb{I};\forall p\in\mathbb{P};\forall t\in\mathbb{T} \tag{13}$$

$$z^{tr}_{pit}+z^{mtl}_{pit}+z^{dbg}_{pit}\le 1 \qquad \forall i\in\mathbb{I};\forall p\in\mathbb{P};\forall t\in\mathbb{T} \tag{14}$$

$$z^{tr}_{pit}+z^{tr}_{p,i,t+1}+z^{mtl}_{p,i,t+1}+z^{mtl}_{p,i,t+2}+z^{dbg}_{p,i,t+2}+z^{dbg}_{p,i,t+3}=3 \qquad \forall i\in\mathbb{I};\forall p\in\mathbb{P};\forall t\in\mathbb{T} \tag{15}$$

$$z^{mtl}_{p,i,t+1}+z^{mtl}_{p,i,t+2}\ge z^{tr}_{pit} \qquad \forall i\in\mathbb{I};\forall p\in\mathbb{P};\forall t\in\mathbb{T} \tag{16}$$

$$z^{dbg}_{p,i,t+1}+z^{dbg}_{p,i,t+2}\ge z^{mtl}_{pit} \qquad \forall i\in\mathbb{I};\forall p\in\mathbb{P};\forall t\in\mathbb{T} \tag{17}$$

$$\sum_{t_p=0}^{t_p=t}z^{tr}_{pit_p}\ge z^{mtl}_{pit} \qquad \forall i\in\mathbb{I};\forall p\in\mathbb{P};\forall t\in\mathbb{T} \tag{18}$$

$$\sum_{t_p=0}^{t_p=t}z^{mtl}_{pit_p}\ge z^{dbg}_{pit} \qquad \forall i\in\mathbb{I};\forall p\in\mathbb{P};\forall t\in\mathbb{T} \tag{19}$$

$$\sum_{t_p=0}^{t_p=t}R_{p-1,i,t_p}\ge z^{tr}_{pit} \qquad \forall i\in\mathbb{I};\forall p\in\mathbb{P};\forall t\in\mathbb{T} \tag{20}$$

At the beginning of each period, PEG receives payment from each PD for its development work. A development cycle comprises three different stages: transistor fabrication, which requires capacity from both MFG and PEG for one period; metal fabrication, which also consumes MFG and PEG capacity for one period and a debugging stage requiring only PEG resources for one period. Constraints (13) ensure that a new product can only be produced for sale after r_{pi} cycles, while (12) enforce resource capacity. (14) ensure that PEG can only work on one development stage per product during a given period. Constraints (15) force the PEG to keep working on a product development project over consecutive cycles, allowing only one period break between consecutive stages. Constraints (16) - (19) define precedence between development stages of each product, while (20) allows PEG to develop a product only if its previous generation has been developed (although it may not have been introduced into the market).

4 DECENTRALIZED MODELS

The objective function of the centralized formulation is separable by agents, and each constraint is associated with one type of agent. However, some variables appear in constraints related to different agents. For example, OC_{it}s appear in (2), which are associated with CORP, and also in (4), which involve the PDs. The first step towards decentralization was to identify all variables shared by more than one agent and create a copy of each variable for each of the agents involved, adding constraints requiring their values to be consistent, an approach known as Lagrangian Decomposition (Guignard and Kim 1987). For example, the variable OC_{it} in the centralized model is replaced by OC_{it}^{Corp} in set (2), and by OC_{it}^{Div} in (4), together with a linking constraint $OC_{it}^{Div} \leq OC_{it}^{Corp} \forall i \in \mathbb{I}; \forall t \in \mathbb{T}$ ensuring that PD i cannot exceed the operating budget allotted to it by CORP. The duplicated variables and linking constraints are summarized in Table 2.

Table 2: Summary of Lagrangian decomposition.

Original Variable	**New Variables**	**Linking Constraints**	**Lagrange Multiplier**
OC_{it}	OC_{it}^{Corp} and OC_{it}^{Div}	$OC_{it}^{Div} \leq OC_{it}^{Corp}$	$\alpha_{p,i}$
x_{pit}	x_{pit}^{Corp} and x_{pit}^{Div}	$x_{pit}^{Corp} \leq x_{pit}^{Div}$	β_{pit}
m_{pit}	m_{pit}^{Mfg} and m_{pit}^{Div}	$m_{pit}^{Div} \leq m_{pit}^{Mfg}$	γ_{pit}
z_{pit}^{tr}	$z_{pit}^{tr}{}^{[Div]}$ and $z_{pit}^{tr}{}^{[Eng]}$	$z_{pit}^{tr}{}^{[Div]} \leq z_{pit}^{tr}{}^{[Eng]}$	θ_{pit}
z_{pit}^{tr}	$z_{pit}^{tr}{}^{[Mfg]}$ and $z_{pit}^{tr}{}^{[Eng]}$	$z_{pit}^{tr}{}^{[Eng]} \leq z_{pit}^{tr}{}^{[Mfg]}$	η_{pit}
z_{pit}^{mtl}	$z_{pit}^{mtl[Div]}$ and $z_{pit}^{mtl[Eng]}$	$z_{pit}^{mtl[Div]} \leq z_{pit}^{mtl[Eng]}$	λ_{pit}
z_{pit}^{mtl}	$z_{pit}^{mtl[Mfg]}$ and $z_{pit}^{mtl[Eng]}$	$z_{pit}^{mtl[Eng]} \leq z_{pit}^{mtl[Mfg]}$	σ_{pit}
z_{pit}^{dbg}	$z_{pit}^{dbg[Div]}$ and $z_{pit}^{dbg[Eng]}$	$z_{pit}^{dbg[Div]} \leq z_{pit}^{dbg[Eng]}$	δ_{pit}
R_{pit}	R_{pit}^{Eng} and R_{pit}^{Mfg}	$R_{pit}^{Mfg} \leq R_{pit}^{Eng}$	ψ_{pit}

Relaxing the linking constraints with associated Lagrange multipliers resulted in a fully separable model that suffered from subproblem degeneracy (Jennergren 1973; Jose et al. 1997) due to the linear update of the Lagrange multipliers. We attempted to address this issue with an Augmented Lagrangian approach (Nocedal and Wright 2006) that includes quadratic penalties for violation of the relaxed constraints. However, the quadratic terms destroy separability by agents. Hence, following Karabuk and Wu (2002), we modify our objective function by duplicating our quadratic terms and then replacing one of the variables with the average value of the variables of the agents related. For example, the term $(OC_{it}^{Div} - OC_{it}^{Corp})^2$ in the Augmented Lagrangian function is replaced by $(OC_{it}^{Div} - OC_{it}^{Param})^2$ where $OC_{it}^{Param} = \frac{1}{2}[OC_{it}^{Div}$ and $OC_{it}^{Corp}]$ in the previous iteration. Thus in the new formulation the objective function of the CORP subproblem becomes:

<u>Corporation Problem:</u>

$$maximize: \left\{ \sum_{\forall p \in \mathbb{P}} \sum_{\forall i \in \mathbb{I}} \sum_{\forall t \in \mathbb{T}} \rho_{pit} x_{pit}^{Corp} + \sum_{\forall p \in \mathbb{P}} \sum_{\forall i \in \mathbb{I}} \alpha_{it} OC_{it}^{Corp} - \sum_{\forall p \in \mathbb{P}} \sum_{\forall i \in \mathbb{I}} \sum_{\forall t \in \mathbb{T}} \beta_{pit} x_{pit}^{Corp} \right.$$

$$\left. - \sum_{\forall p \in \mathbb{P}} \sum_{\forall i \in \mathbb{I}} \mu (OC_{it}^{Param} - OC_{it}^{Corp})^2 - \sum_{\forall p \in \mathbb{P}} \sum_{\forall i \in \mathbb{I}} \sum_{\forall t \in \mathbb{T}} \mu (x_{pit}^{Corp} - x_{pit}^{Param})^2 \right\}$$

subject to constraints (2) and (3).

For additional illustration, the objective of the subproblem for PD_i is given by

<u>Product Division i Problem</u>

$$maximize: \left\{ \sum_{\forall p \in \mathbb{P}} \sum_{\forall t \in \mathbb{T}} -h_{pit}^{mtl} I_{pit} - \sum_{\forall p \in \mathbb{P}} \alpha_{it} OC_{it}^{Div} + \sum_{\forall p \in \mathbb{P}} \sum_{\forall t \in \mathbb{T}} \beta_{pit} x_{pit}^{Div} \right.$$

$$- \sum_{\forall p \in \mathbb{P}} \sum_{\forall t \in \mathbb{T}} \gamma_{pit} m_{pit}^{Div} - \sum_{\forall p \in \mathbb{P}} \sum_{\forall t \in \mathbb{T}} \theta_{pit} z_{pit}^{tr_{[Div]}} - \sum_{\forall p \in \mathbb{P}} \sum_{\forall t \in \mathbb{T}} \lambda_{pit} z_{pit}^{mtl_{[Div]}} - \sum_{\forall p \in \mathbb{P}} \sum_{\forall t \in \mathbb{T}} \delta_{pit} z_{pit}^{dbg_{[Div]}}$$

$$- \sum_{\forall p \in \mathbb{P}} \mu (OC_{it}^{Div} - OC_{it}^{[Param]})^2 - \sum_{\forall p \in \mathbb{P}} \sum_{\forall t \in \mathbb{T}} \mu (x_{pit}^{[Param]} - x_{pit}^{Div})^2$$

$$- \sum_{\forall p \in \mathbb{P}} \sum_{\forall t \in \mathbb{T}} \mu (m_{pit}^{Div} - m_{pit}^{[Param]})^2 - \sum_{\forall p \in \mathbb{P}} \sum_{\forall t \in \mathbb{T}} \mu (z_{pit}^{tr_{[Div]}} - z_{pit}^{tr_{[Param]}})^2$$

$$\left. - \sum_{\forall p \in \mathbb{P}} \sum_{\forall t \in \mathbb{T}} \mu (z_{pit}^{mtl_{[Div]}} - z_{pit}^{mtl_{[Param]}})^2 - \sum_{\forall p \in \mathbb{P}} \sum_{\forall t \in \mathbb{T}} \mu (z_{pit}^{dbg_{[Param]}} - z_{pit}^{dbg_{[Div]}})^2 \right\}$$

subject the constraint sets (4) to (6).

The subproblems for MFG and PEG were derived using the same approach. This new formulation is completely separable. The basic idea of the approach is to penalize the differences between the values of shared decision variables and the average value of those variables obtained at the previous iteration of the procedure, driving the algorithm towards a coordinated solution where shared variables take compatible values. This introduces a new set of parameters related to the linking constraints that must be initialized at the start of the solution procedure. This is accomplished with a constructive heuristic that solves a sequence of modified subproblems of the agents using a conservative estimate of the raw demand as the sales target. The same sequential feasibility routine explained in the next section is employed to recover the feasibility of the relaxed solution, which is then used as a lower bound.

5 CONSTRUCTION OF A FEASIBLE SOLUTION

Figure 1 shows a simplified flow diagram of the feasibility routine. In Step 0, the sales targets for CORP based on the PDs estimated sales X_{pit}^{Div} in the (infeasible) relaxed solution (or the raw demand if it is the initial feasible solution). In Step 1, CORP then solves a modified subproblem including the constraints 2 to 20 whose objective is to minimize the squared deviation from the sales target. In the experiments in this paper, we assume that CORP is able to parameterize these constraints with the same information available to the various agents, although in general this is clearly not the case. A more accurate perspective is that CORP takes input from the PDs as to what they believe they can sell, and uses its knowledge of the constraints the other agents must obey to develop estimates of their actions given the budget allocations computed in this subproblem. The solution to this subproblem produces a budget allocation from CORP to the PDs that minimizes deviations from the sales goals of the PDs and is consistent with all constraints the different agents must observe. It is not unreasonable in practice to assume that CORP will be acquainted with the basic forms of the agents' principal constraints, although it is likely that the values of some parameters in these constraints will be inaccurate to some degree. The impact of these inaccuracies on the performance of the procedure will be explored in future work. The sales targets of the PDs are revised to reflect the values consistent with the CORP budget allocations obtained from this subproblem. This subproblem also computes an estimate of the MFG capacity that must be allocated to each PDs in each period. When used to compute an initial feasible solution without running the fully separable approach, the initial sales targets are set equal to a conservative estimate of the demand forecast provided by the PDs.

Given the budget allocations and revised sales targets, and tentative MFG capacity allocation computed by CORP in Step 1, in Step 2 each PD solves its own subproblem with constraints 2 to 20 seeking to minimize the squared deviation of their sales from the new sales targets. This solution yields revised sales targets for the PDs, the orders that each PD will request from MFG, and the introduction dates for new products that will be requested from PEG. In Step 3 PEG solves its subproblem seeking to minimize deviations from the introduction dates requested by the PDs, assuming that the MFG capacity will be available to it for development activities. Finally, in Step 4, MFG solves its subproblem seeking to minimize the squared deviation between the quantities of product it can deliver for sale and those requested by the PDs. In addition to constraints 2 to 20, the MFG subproblem is forced to match in its development estimated logic the debugging development stages given by the PEG. In other words, the MFG capacity must be released to the PEG such that the metal stage (the last development stage requiring MFG capacity) is one or at most two periods before the PEG's scheduled debugging stage given by the already solved PEG problem. Since the PEG debugging schedule guarantees an introduction date that, due to the PEG objective function, deviates minimally from the PDs introduction estimates, the rollover strategy coordinates through this new constraint set. Note that each agent solves a subproblem that contains the constraints from the centralized formulation. The subproblems are solved sequentially, starting with CORP, proceeding to the PDs, followed by PEG and MFG. Each subproblem takes as input the decisions made at the previous step and seeks to minimize deviations imposed by its own constraints. After each step of the algorithm, the sales target can only be revised downwards, guaranteeing the feasibility of the solution relative to manufacturing and development capacity. However, consecutive reductions in sales targets may reduce revenue to the point that it cannot cover the costs incurred. In this situation, the PD subproblems are resolved with an extra restriction on the number of units that MFG can produce, yielding the true operating budget allocation *OC* required from CORP to produce this output. If this solution is not feasible due to reduced revenue, the CORP subproblem is reoptimized to minimize deviations from the sales target while meeting the PDs financial requests. If this CORP subproblem is infeasible, the firm does not have enough initial budget to develop its sales plan, so we delay the introduction date of the least profitable product that is going to be developed during the planning horizon. (We exclude generation zero from this delay).

6 EXPERIMENTAL RESULTS

For our computational experiments, we consider six problem configurations E1 through E6, whose characteristics are summarized in Figure 2, varying the number of divisions, number of generations of each product, number of periods in the planning horizon generations, and the timing of product introductions. Under synchronous demand, new generations of different divisions must enter the market within an interval of three periods, while under asynchronous demand, the new introductions are distributed over twelve periods. The need to introduce multiple product generations for different divisions within a short interval ought to result in more difficult problem instances. For each problem configuration, we consider eight capacity profiles, as shown in Figure 3, varying the manufacturing capacity, development capacity, and the number of development cycles required. For each combination of instance configuration and capacity profile, we solve three randomly generated problem instances using the centralized formulation, the constructive heuristic,

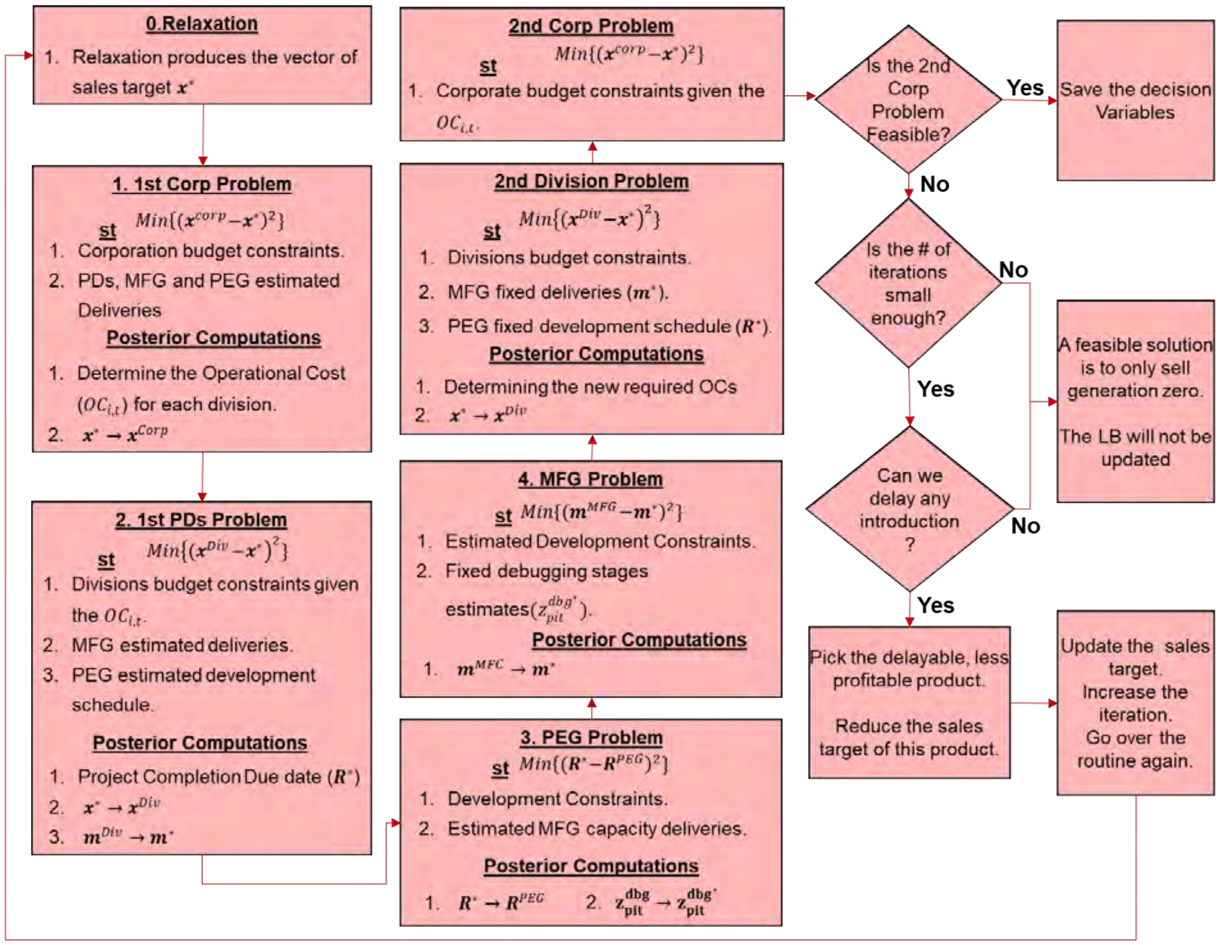

Figure 1: Flowchart of procedure for constructing a feasible solution.

and the fully separable model. We compare the solutions in terms of solution quality and computation time in Figure 4, where each cell in the table represents the average of the quantity of interest over the instances solved. Although we also solved most instances with the Augmented Lagrangian procedure we do not report detailed results here for brevity. However, we note that the fully separable model yields solution times two orders of magnitude faster than the Augmented Lagrangian (average CPU time of 1.06 minutes as opposed to 255.15 minutes). We thus focus our discussion on the constructive heuristic and the fully separable model.

In our experiments, we observe that the initial feasible solution does not deviate more than 25% from the centralized solution; in 45.8% of all instances solved, the fully separable Lagrangian procedure yielded no improvement over the initial solution. The fully separable model yields significant improvements in the gap over the initial solution, with the worst average deviation from the centralized solution being of the order of 14% for the difficult instances where MFG capacity is limited (80% of average demand), synchronous product introductions and four development cycles. A few of these instances required extremely long CPU times, as indicated by the difference between the average and maximum CPU times. In general, the solution procedure behaves as expected, with increasing capacity (MFG capacity and no. of development teams) yielding easier instances and increasing workload (number of development cycles and synchronous introductions) yielding harder ones. Overall, the fully separable approach yields decentralized solutions that are acceptably close to the centralized solution in quality; better choices of algorithm parameters, such as stopping conditions appear likely to yield further improvements and more consistent solution times.

Instance	Number of Divisions	Number of Generations	Number of Periods	Introduction Synchrony
Intance_E1	2	3	63	Synchronous
Intance_E2	2	3	63	Asynchronous
Intance_E3	2	2	55	Synchronous
Intance_E4	2	2	55	Asynchronous
Intance_E5	3	3	63	Synchronous
Intance_E6	3	3	63	Asynchronous

Figure 2: Problem configurations summary.

Sub-Instance	Manufacturing Capacity	Product Development Teams	Number of Development Cycles
#0	80%	2	3
#1	100%	2	3
#2	80%	6	3
#3	100%	6	3
#4	80%	2	4
#5	100%	2	4
#6	80%	6	4
#7	100%	6	4

Figure 3: Capacity profiles summary.

The main reason for the more efficient performance of the completely separable model is that it always begins with a good initial feasible solution obtained after a run of our feasibility finder routine in which the entire demand is used as the sales target. This initial solution not only provides a good initial lower bound, but because it is also in the parameters of the quadratic terms, it drives all agents towards this solution. It is important to notice that the initial feasible solution is obtained by solving sequentially special agent problems, in which they make two vital assumptions. First, they assume that the rest of the agents will act in the best interest of the company, truthfully reflecting their capabilities and limitations. Moreover, the agents assume that they completely understand the other agent's processes.

			3 Development Cycles				4 Development Cycles			
Introductions	Mfg. Capacity	Dev. Teams	Avg. IS Gap	Avg. FS Gap	Avg. CPU (min)	Max. CPU(min)	Avg. IS Gap	Avg. FS Gap	Avg. CPU (min)	Max. CPU(min)
Asynchronous	80%	2	0.13	0.06	1.53	3.53	0.20	0.10	135.97	605.62
		6	0.11	0.07	1.30	2.12	0.16	0.08	2.12	4.68
	100%	2	0.04	0.03	0.97	1.69	0.08	0.06	77.34	281.92
		6	0.05	0.03	0.97	1.77	0.05	0.04	0.91	1.33
Synchronous	80%	2	0.14	0.09	4.13	13.60	0.17	0.14	234.46	1241.43
		6	0.17	0.10	1.71	2.73	0.19	0.13	2.26	4.58
	100%	2	0.10	0.06	1.66	3.86	0.21	0.12	146.50	552.10
		6	0.13	0.08	1.48	2.35	0.22	0.09	1.63	2.47
Overall			**0.11**	**0.06**	**1.72**	**13.60**	**0.16**	**0.10**	**75.15**	**1241.43**

Figure 4: Summary of the experimental results.

From the computational time perspective, the instances that require four development cycles demand significantly more time, with one experiment replica requiring almost twenty-one hours to solve. This computational complexity is more intensive for instances E_5 and E_6, which are the ones that have three divisions and four generations, increasing the number of variables and constraints.

7 CONCLUSION

In this paper, we present a completely separable approach derived from previous centralized and semi-decentralized models presented in Leca et al. (2021) and Leca et al. (2022). This separable formulation achieves good solution quality relative to the centralized solution faster than the Augmented Lagrangian approach. We also present a sequential decentralized heuristic that recovers a feasible solution from the relaxed solution obtained by the fully separable approach and can also be used to construct an initial feasible solution. The success of the fully separable approach is driven in large part by the good initial solutions obtained since it incentivizes agents towards coordinated solutions that improve on the initial solution. The fully separable approach represents a tradeoff between improving the objective function and achieving coordinated solutions.

The decentralized, sequential heuristic proceeds by solving a sequence of agent subproblems in which each agent is aware of the constraints other agents must satisfy, such as those governing the stages of the product development process in Section 5. The main idea is that each agent solves a problem that possesses the centralized model constraints in which the variables on its domain are treated as decision variables, and the rest of the variables represent beliefs as to what other agents will do. Thus the constructive heuristic makes two essential assumptions: that each agent will seek to achieve the goals communicated to it by other agents preceding it in the solution sequence and that each agent has a basic knowledge of the constraints governing the decisions of the others. In practice, agents may have incomplete (not all constraints are known) or inaccurate (parameters such as the number of development cycles needed are incorrect) knowledge of the constraints governing the decisions of others. Future work will examine the behavior of the decentralized decision procedures in the presence of such information discrepancies and different strategies agents may use to hedge against uncertainties in external parameters such as market demand and the decisions of other agents.

ACKNOWLEDGMENTS

This research was supported by the National Science Foundation (NSF) under Grant No. CMMI-1824744. Any opinions stated are those of the authors, and do not necessarily reflect the position of NSF.

REFERENCES

Billington, C., H. L. Lee, and C. S. Tang. 1998. "Successful Strategies for Product Rollovers". *MIT Sloan Management Review* 39(3):23.

Carrillo, J. E., and R. M. Franza. 2006. "Investing in Product Development and Production Capabilities: The Crucial Linkage Between Time-to-Market and Ramp-up Time". *European Journal of Operational Research* 171(2):536–556.

Chung, J., and J. Jang. 2022. "Influence of an RD Lot on Productivity in Semiconductor Manufacturing". *Computers Industrial Engineering* 168:108030.

Cohen, G. 1980. "Auxiliary Problem Principle and Decomposition of Optimization Problems". *Journal of Optimization Theory and Applications* 32(3):277–305.

Erhun, F., P. Gonçalves, and J. Hopman. 2007. "The Art of Managing New Product Transitions". *MIT Sloan Management Review* 48(3):73.

Guignard, M., and S. Kim. 1987. "Lagrangian Decomposition: A Model Yielding Stronger Lagrangian Bounds". *Mathematical Programming* 39(6):215–228.

Jennergren, P. 1973. "A Price Schedules Decomposition Algorithm for Linear Programming Problems". *Econometrica* 41(5):965–980.

Jose, Rinaldo and Harker, Patrick and Ungar, Lyle 1997, 01. "Coordinating Locally Constrained Agents Using Augmented Pricing". Techinal Report. University of Pennsylvania, Philadelphia.

Karabuk, S., and S. D. Wu. 2002. "Decentralizing Semiconductor Capacity Planning Via Internal Market Coordination". *IIE Transactions* 34(9):743–759.

Kempf, K. G., F. Erhun, E. F. Hertzler, T. R. Rosenberg, and C. Peng. 2013. "Optimizing Capital Investment Decisions at Intel Corporation". *Interfaces* 43(1):62–78.

Kim, S., and R. Uzsoy. 2013. "Modeling and Analysis of Integrated Planning of Production and Engineering Process Improvement". *IEEE Transactions on Semiconductor Manufacturing* 26(3):414–422.

Kutanoglu, E., and S. D. Wu. 1999. "On Combinatorial Auction and Lagrangean Relaxation for Distributed Resource Scheduling". *IIE Transactions* 31(9):813–826.

Leca, C., K. Kempf, and R. Uzsoy. 2021. "An Optimization Framework for Managing Product Transitions in Semiconductor Manufacturing". In *Proceedings of the 2021 Winter Simulation Conference*, edited by S. Kim, B. Feng, K. Smith, S. Masoud, Z. Zheng, C. Szabo, and M. Loper, 1–12. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Leca, C., K. Kempf, and R. Uzsoy. 2022. "Towards Decentralized Decisions for Managing Product Transitions in Semiconductor Manufacturing". In *Proceedings of the 2022 Winter Simulation Conference*, edited by B. Feng, G. Pedrielli, Y. Peng, S. Shashaani, E. Song, C. Corlu, L. Lee, E. Chew, T. Roeder, and P. Lendermann, 3442–3452. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Manda, A. B., and R. Uzsoy. 2020. "A Simple Model of Capacity Contention During New Product Introductions". *IEEE Transactions on Semiconductor Manufacturing* 33(2):240–251.

Mönch, L., R. Uzsoy, and J. W. Fowler. 2018. "A Survey of Semiconductor Supply Chain Models Part I: Semiconductor Supply Chains, Strategic Network Design, and Supply Chain Simulation". *International Journal of Production Research* 56(13):4524–4545.

Nocedal, J., and S. J. Wright. 2006. *Numerical Optimization*. 2e ed. New York, NY, USA: Springer.

Schwarz, J. A., and B. Tan. 2021. "Optimal Sales and Production Rollover Strategies Under Capacity Constraints". *European Journal of Operational Research* 294(2):507–524.

Shen, W., I. Duenyas, and R. Kapuscinski. 2014. "Optimal Pricing, Production, and Inventory for New Product Diffusion Under Supply Constraints". *Manufacturing & Service Operations Management* 16(1):28–45.

Özer, O., and O. Uncu. 2015. "Integrating Dynamic Time-to-Market, Pricing, Production and Sales Channel Decisions". *European Journal of Operational Research* 242(2):487–500.

AUTHOR BIOGRAPHIES

CARLOS LECA is a PhD student in the Edward P. Fitts Dept. of Industrial and Systems Engineering at North Carolina State University from where he also earned a Master degree. He holds a BS degree in Industrial Engineering from the Universidad Catolica Andres Bello in Caracas, Venezuela. His research interests include production scheduling, supply chain management, decentralized decision-making and mechanism design. His email address is clecape@ncsu.edu.

DR. KARL KEMPF is a Senior Fellow and Director of Decision Engineering at Intel Corporation. Since joining Intel in 1987 he has lead a team of decision scientists charged with building decision-support processes and tools focused on faster better decision making across the corporation. Kempf has co-edited three books and published more than 175 contributions in decision science. He has been a research adjunct at Missouri State, Arizona State, North Carolina State and Stanford Universities. He is a member of the National Academy of Engineering (NAE), a Fellow of the IEEE, and an INFORMS Fellow. His team at Intel has won the INFORMS Prize, the Wagner Prize, and the Edelman Award. Prior to joining Intel he was involved in motor racing, movie special effects, and aerospace factory automation. His e-mail address is karl.g.kempf@intel.com.

DR. REHA UZSOY is Clifton A. Anderson Distinguished Professor in the Edward P. Fitts Department of Industrial and Systems Engineering at North Carolina State University. He holds BS degrees in Industrial Engineering and Mathematics and an MS in Industrial Engineering from Bogazici University, Istanbul, Turkey. He received his Ph.D in Industrial and Systems Engineering in 1990 from the University of Florida, and held faculty positions in Industrial Engineering at Purdue University prior to joining North Carolina State University in 2007. His teaching and research interests are in production planning and supply chain management. He was named Outstanding Young Industrial Engineer in Education in 1997 and a Fellow of the Institute of Industrial Engineers in 2005, and has received awards for both undergraduate and graduate teaching. His email address is ruzsoy@ncsu.edu.