Column Generation for Semiconductor Manufacturing Scheduling with Family Qualification and Secondary Resource Constraints

Arthur Kramer[1], Claude Yugma[2]

[1] Mines Saint-Etienne, Univ Clermont Auvergne, CNRS, UMR 6158 LIMOS, Institut Henri Fayol, 42023 Saint-Etienne, France

[2] Mines Saint-Etienne, Univ Clermont Auvergne, CNRS, UMR 6158 LIMOS, Centre of Microelectronics in Provence, 13541 Gardanne, France

{arthur.kramer, yugma}@emse.fr

Keywords : *scheduling, column generation, semiconductor manufacturing.*

1 Introduction

In this work, we address a scheduling problem arising in the photolithography stage of semiconductor wafer fabrication. Photolithography is a well-known bottleneck in wafer fabs because stepper machines are extremely expensive and significantly costlier than most other tools. As a result, efficient scheduling at this stage is central for improving overall fab performance. During photolithography process, groups of wafers (referred to as “lots” or “jobs”) undergo circuit-pattern transfer through stepper machines using masks. Masks act as secondary and scarce resources, and their timing and availability critically influence the utilization of these costly machines. Semiconductor fabs also rely on “send-ahead wafers” (SWs), processed before regular lots to calibrate the stepper machine and validate the mask. In practice, lots belonging to the same family are often processed consecutively to reduce SW usage and maintain machine qualification.

Although stepper scheduling has been widely studied, the combined integration of qualification constraints and send-ahead wafers has received far less attention. In particular, SWs are seldom discussed explicitly in semiconductor manufacturing research. Notable exceptions include the works of Mönch and Yugma [1] and Obeid et al. [2], which investigate related parallel-machine scheduling problems involving qualification requirements. The present study builds on these contributions by explicitly incorporating re-qualification mechanisms via send-ahead wafers and modeling secondary resource constraints associated with mask availability.

2 Problem definition and proposed column generation

We consider a scheduling problem in which a set $J = \{1, \ldots, n\}$ of jobs must be processed on a set $M = \{1, \ldots, m\}$ of identical parallel machines over a time horizon $H = \{0, 1, \ldots, h\}$. Each job $j \in J$ is characterized by a processing time p_j, a release date r_j, a due date d_j, a tardiness penalty weight w_j, and belongs to a family in $F = \{1, \ldots, f\}$.

When switching from one family k to another family l, a setup time s_l is required. A machine may process a job from a given family only if (i) it is qualified for that family, and (ii) the mask associated with that family is physically present on the machine. Because only one mask per family is available, jobs from the same family cannot be processed simultaneously on multiple machines. Masks may be moved between machines, but this operation requires a transportation time tm_k for family $k \in F$.

Qualification is governed by time windows. For each family k, a maximum allowed idle time Δ_k specifies how long a machine may remain unused before losing qualification for that family. Machines are initially qualified for a given time window $\tilde{\Delta}_k$. Re-qualification for family k can be obtained by processing send-ahead wafers (SWs), which require a processing time sw_k.

The objective is to determine a schedule that satisfies all the operational constraints (job processing, non-overlapping, setups between families, qualification conditions, and mask availability) while minimizing the total weighted tardiness and the weighted number of send-ahead wafers used. The problem can be expressed as

$$P \mid r_j, s_i, aux, sw \mid wT + w'SW,$$

where wT denotes the total weighted tardiness and $w'SW$ the total weighted count of SWs.

To tackle this problem, we propose the following extended formulation:

$$(MP) \quad \min \quad \sum_{s \in S} c_s x_s \tag{1}$$

subject to:

$$\sum_{s \in S} x_s \leq m \tag{2}$$

$$\sum_{s \in S} a_{js} x_s \geq 1 \qquad j \in J \tag{3}$$

$$\sum_{s \in S} b^t_{fs} x_s \leq 1 \qquad f \in F,\ t \in H \tag{4}$$

$$x_s \in \{0, 1\} \qquad s \in S \tag{5}$$

Here, $x_s = 1$ if one-machine schedule $s \in S$ is selected, and 0 otherwise. The set S contains all feasible one-machine schedules respecting non-overlapping constraints, family-dependent setups, qualification constraints, and mask availability. Value a_{js} indicates how many times job j is present in schedule s while $b^t_{fs} = 1$ if family f occupies time instant t in schedule s, and 0 otherwise. The cost c_s represents the total cost (tardiness and SW usage) of schedule s.

Constraints (2)–(4) enforce, respectively: (i) the machine-limit condition, (ii) full job coverage, and (iii) mask availability restrictions (only one mask per family at any time).

Model (MP) results from the reformulation of a Mixed-Integer Linear Program based on a Network Flow representation (MILP-NF). We solve the linear relaxation of (1)–(5) using Column Generation (CG). The pricing problem consists in identifying one-machine schedules with negative reduced cost. It is modeled as a MILP based on the MILP-NF and solved using a commercial solver.

3 Preliminary results and concluding remarks

We tested the CG approach on 24 instances with $n = 12$ jobs, $m \in \{2, 3\}$ machines, and $f \in \{3, 4\}$ families, and compared it with the MILP-NF formulation. All methods were coded in C++ and solved with Gurobi 10.0.3; MILP-NF was given a 2-hour time limit.

CG solved the LP relaxation for all instances and proved optimality for 21 out of 24 cases, compared to 11 for MILP-NF. The bounds from CG were consistently tighter, confirming the strength of the decomposition.

These results are encouraging. Future work includes improving the pricing problem, developing a branch-and-price scheme, and testing larger, more representative instances.

References

[1] L. Mönch and C. Yugma. Scheduling jobs on parallel machines with qualification constraints. In *2015 IEEE Int. Conf. on Autom. Sci. and Eng. (CASE)*, pages 657–658, 2015.

[2] A. Obeid, S. Dauzère-Pérès, and C. Yugma. Scheduling job families on non-identical parallel machines with time constraints. *Annals of Operations Research*, 213:221–234, 2014.