Proceedings of the 2025 Winter Simulation Conference
E. Azar, A. Djanatliev, A. Harper, C. Kogler, V. Ramamohan, A. Anagnostou, and S. J. E. Taylor, eds.

AVAILABLE-TO-PROMISE VS. ORDER-UP-TO: A DISCRETE-EVENT SIMULATION OF SEMICONDUCTOR INVENTORY TRADE-OFFS

Ayhan Gökcem Cakir[1,2], and Hans Ehm[2]

[1]TUM School of Management, Technical University of Munich, Munich, GERMANY
[2]Corporate Supply Chain Innovation, Infineon Technologies AG, Munich, GERMANY

ABSTRACT

Complex supply networks, volatile demand, and up to 26-week lead times pose severe trade-offs between inventory efficiency and service flexibility in semiconductor supply chains. We develop a discrete-event simulation to compare Order-up-to policies against an Available-to-promise (ATP) approach. Results show that ATP dynamically reallocates capacity, cuts reliance on finished-goods buffers, and boosts operational agility without eroding financial performance. Our findings suggest integrating ATP with real-time decision support can reconcile efficiency and flexibility under extreme lead-time uncertainty by enabling proactive order adjustments, decreasing backorders, and cost overruns.

1 INTRODUCTION

Semiconductor supply chains involve intricate, multi-stage processes from wafer fabrication through assembly and testing, that can exceed 26 weeks per product cycle, imposing significant operational and capital demands (Mönch et al. 2018). Demand volatility across global networks further amplifies variability through the bullwhip effect, leading companies to hold inventories that inflate holding costs and risk obsolescence of semiconductors (Lee et al. 1997). Conventional inventory models, such as Economic Order Quantity (EOQ) and Just-in-Time (JIT), falter under this lead-time uncertainty, prompting either service breakdowns or costly overstock. Available-to-Promise (ATP) systems dynamically allocate capacity and inventory commitments in real time, yet their empirical benefits in the semiconductor industry remain underexplored. To address this gap, we build a discrete-event simulation, calibrate with one year of real order and order forecast data of Infineon Technologies AG. We compare static and dynamic Order-Up-To policies against an ATP framework to quantify trade-offs in service, inventory efficiency, and cost.

2 METHODOLOGY OF DATA REGENERATION AND SIMULATION FRAMEWORK

This study applies a discrete-event simulation (DES) framework to evaluate three inventory policies under semiconductor-specific constraints: Static Order-Up-To (S-OUT), Dynamic Order-Up-To (D-OUT), and Available-to-Promise (ATP). Following the staged procedure recommended by Banks et al. (2013), we (i) formulate the problem and performance objectives, (ii) translate a conceptual supply-chain model into AnyLogic, and (iii) verify and validate the model through stepwise debugging, extreme-condition tests, and historical-data back-checks. The simulated network comprises fabrication, assembly, and test facilities as manufacturing fabs, estimated capacity ceilings, inventory hubs, and 26-week production lead times mirroring Infineon's global operations.

Historical records for a year supply the baseline order sizes, requested delivery dates, and 26-week rolling forecasts. To protect confidentiality while preserving statistical fidelity, we regenerate Qualified Synthetic Data (QSD) by fitting distributional parameters (mean, standard deviation (σ), autocorrelation) to the raw series and resampling with stochastic perturbations that embed bullwhip amplification and ±8-10% annual growth trends. Forecast accuracy is monitored with the company-standard SMAPE-3 metric, allowing realistic error bounds. The synthetic timeline is extended to ten simulated years so that long lead times and inventory cycles fully materialize.

S-OUT keeps a fixed target level equal to 26 weeks of average demand plus two σ safety stock; production halts once finished-goods inventory meets the target. D-OUT recalculates the target weekly using the latest 26-week forecast, adding adaptive safety stock to absorb forecast error. ATP dispenses with finished-goods targets: weekly forecasts drive a rolling production schedule, while customer orders consume virtual quotas in real time. Capacity is reallocated at each planning period, shifting inventory buffers upstream into raw materials.

All three scenarios share identical order streams, capacity thresholds, yield assumptions, and inbound raw-material cadence to isolate the effect of the replenishment rule. Each run spans 10 simulated years. Tracked outputs include order-fulfilment lead time, backorder count, finished-goods inventory (FGI), raw-material inventory (RMI), and total costs. This design provides a statistically robust basis for quantifying the operational and financial trade-offs inherent to ATP versus OUT in semiconductor supply chains.

3 RESULTS OF THE SIMULATION STUDY

Our ten-year simulation experiments reveal clear trade-offs among the three replenishment policies. S-OUT delivers the shortest average backlog duration, 9.2 days, at the cost of higher inventory holding. In contrast, D-OUT and ATP suffer markedly longer delays of 80.2 days and 74.6 days, respectively, because both rely on forecasts that leave periods of understocking. Consequently, backorder incidence rises from 20.3% under S-OUT to 78.8% under D-OUT and 82.4% under ATP. Note that we do not apply delivery window flexibility; any order exceeding the planned delivery date is marked as a backorder.

S-OUT's service advantage comes at the cost of very high FGI. Over the horizon, it incurs €92.9 k in FGI holding charges versus €74.8 k for D-OUT and only €1.0 k for ATP. In contrast, RMI shows the opposite pattern: the flexible policies tie up more capital upstream in RMI. These figures confirm that upstream flexibility is more favorable for answering volatile demand by later product configuration.

S-OUT minimizes total cost despite its large FGI because its penalty exposure is small. ATP lowers total holding cost by 18% relative to S-OUT, but backorder penalties offset much of that saving. D-OUT performs the worst on every cost dimension because forecast error drives high RMI and penalty charges. Although S-OUT appears most profitable in the model, its advantage stems from the assumption of immediate delivery penalties; in practice, short permissible delivery windows and dynamic production constraints would erode this lead. Among the adaptive strategies, ATP strikes the better balance: it cuts FGI almost entirely, keeps total cost within 22% of S-OUT, and generates nearly triple the EBIT of D-OUT. The results underscore a key principle for semiconductor supply chains: Shifting buffers upstream into raw materials can cut finished-goods risk without catastrophic cost escalation, provided backorder penalties are managed. Hybrid policies prioritizing ATP's real-time allocation with optimized safety-stock targets could preserve S-OUT-level service while capturing ATP's cost efficiency and flexibility.

REFERENCES

Mönch, L., Uzsoy, R., & Fowler, J. W. (2017). 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*, 4524–4545. https://doi.org/10.1080/00207543.2017.1401233

Lee, H. L., Padmanabhan, V., & Whang, S. (1997). Information distortion in a supply chain: The bullwhip effect. *Management Science, 43*, 546–558. https://doi.org/10.1287/mnsc.43.4.546