International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(18s) 2026 pp. 789-799
https://doi.org/10.70917/ijcisim-2026-4923
Article

AI-Driven Predictive Manufacturing and Smart Warehouse Optimization in Semiconductor Supply Chains

Sandeep Reddy Varakantham

GlobalFoundries, USA

Abstract: Semiconductor supply chains carry a structural vulnerability that transactional ERP architectures cannot address on their own: the interval between an emerging operational problem and the moment that problem becomes visible in the execution system is long enough for significant damage to accumulate. A yield-degrading equipment drift event may traverse dozens of wafer lots before inline metrology flags it; a reservation contention crisis may be hours from materialising when the last lot that would have prevented it enters the finished goods queue. The Predictive Semiconductor Operations Framework (PSOF) proposed here resolves this vulnerability through a four-layer architecture that embeds predictive analytics inside the ERP data fabric rather than beside it. Equipment telemetry, lot genealogy, inventory state, and demand signals are consolidated in a unified pipeline that feeds Long Short-Term Memory (LSTM)-based failure prediction, gradient boosting yield correlation, and demand-driven replenishment models. Each predictive output translates directly into an ERP-executable action — a maintenance work order, a yield-risk inspection hold, an anticipatory lot reservation — without leaving the transactional boundary that semiconductor lot traceability requires. Deployment implications for multi-fab Oracle ERP environments are discussed, with attention to phased implementation and data quality prerequisites. The article advances the position that ERP-data primacy, rather than external data replication, is the architectural foundation that makes predictive manufacturing intelligence operationally durable in high-mix fab supply chains.
Keywords: semiconductor supply chain, predictive manufacturing, ERP integration, warehouse automation, yield optimization, digital twin, Oracle Integration Cloud

1. Introduction

1.1 Industry Context

A semiconductor fabrication plant is, among other things, a very large distributed state machine — thousands of wafer lots in simultaneous motion across hundreds of interdependent process steps, each carrying the accumulated exposure history of every tool it has touched since the start [1]. Customer commitments materialise through long-term agreements and EDI purchase orders that arrive months in advance of required ship dates; quarterly financial cycles compress the delivery window at the end of every period, concentrating the cost of any operational disruption into the worst possible timing. The Order Workbench systems that enterprise customers use to submit and track purchase order requests against foundry supply chains are not merely order management interfaces; they are the boundary through which customer financial planning and fab operational reality must continuously reconcile.

What governs that reconciliation is an ERP stack whose execution model was not designed for prediction. Sales order creation, lot reservation, pick confirmation, delivery and trip creation, and invoicing are executed in sequence, each stage gating the next, each recording a state that has already been reached [2]. The completeness of this record is real and valuable; the regulatory and contractual traceability it supports is non-negotiable. The limitation is equally real: a system designed to record confirmed states cannot generate signals about states that have not yet occurred, and the operational cost of that gap compounds at the scale and speed of semiconductor manufacturing.

1.2 Research Problem and Scope

Standalone AI platforms introduced alongside ERP are not a structural solution to this problem. Their predictions are grounded in data extracted from the ERP — snapshots that carry an inherent lag relative to the live transaction state [4]. A lot that has been dispositioned, re-reserved, or rerouted in the thirty minutes since the last extraction does not appear correctly in a model trained on the snapshot; a prediction based on that model is not a prediction about the current operational state. In multi-fab environments where lot dispositions change within minutes, and reservation contention resolves in real time against a live inventory pool, this lag represents a structural misalignment between what the AI reports and what the ERP knows [3].

The research problem, stated precisely, is this: how can predictive manufacturing intelligence be embedded inside the ERP data fabric — rather than layered over it — in a way that preserves lot-level traceability while producing forward-looking operational signals? This article answers that question through the PSOF architecture, scoped to multi-fab Oracle ERP environments with integrated warehouse management. Hardware-layer robotics, financial modelling, and manufacturing domains outside the semiconductor sector are not addressed.

1.3 Objectives

Three objectives organise the analysis that follows. First, the article characterises the specific failure modes of reactive ERP-only supply chain management in semiconductor contexts — not as general limitations of ERP technology but as structural consequences of a transactional execution architecture applied to a domain where predictive capability is operationally necessary. Second, it proposes the PSOF architecture and evaluates its design against those failure modes. Third, it draws on empirical evidence from comparable deployments in the literature to assess PSOF's expected impact on yield variability, unplanned downtime, and warehouse throughput. The central argument is that ERP-data primacy is not a design constraint to work around; it is the architectural feature that makes predictive intelligence operationally viable in semiconductor supply chains.

2. Reactive ERP Architectures and Their Limits in Semiconductor Manufacturing

2.1 Legacy ERP Execution Models

Sequential gate architectures in ERP execution exist for good reasons. Each transaction stage in the Order-to-Cash cycle — sales order creation, reservation, pick confirmation, shipment, invoice — creates an auditable record that can be corrected, reversed, or traced independently [2]. A lot that fails inspection can be unreserved without corrupting the order record; a shipment that misses its carrier window can be rebooked without voiding the delivery document. The architecture is built around the assumption that correctness of the historical record is the primary operational requirement, and in most manufacturing contexts, that assumption holds.

Semiconductor fabs present a different condition. Lot genealogy in a high-mix fab spans three hundred or more process steps, each creating a tool-exposure record that may correlate with final yield in ways that only become visible at test weeks or months after the relevant process events occurred [1]. By the time a yield excursion generates a rework order that appears as a transaction in the ERP, the causal chain of equipment-exposure combinations that produced it is complete and irreversible. The transactional record is correct; the opportunity to intervene has passed before the record existed [9].

2.2 Yield Variability and Downtime Propagation

Equipment parameter drift occupies a detection gap between the process-of-record target and the SPC control limit — the range within which a tool is technically in spec but is trending toward a condition that correlates with yield degradation [9]. Lots processed during this drift period accumulate exposure to parameter combinations that the SPC chart does not flag as anomalous, because each individual parameter reading is within limits. The correlation with yield loss is a multi-step interaction effect that no single-parameter monitoring system can detect before the fact [21].

Unplanned tool failure during a loaded shift creates a different but equally costly disruption. The bottleneck is not the broken tool alone; it is the queue that builds behind it as lots are scheduled for that tool cannot advance [11]. Time-based preventive maintenance reduces — but does not eliminate — unplanned failures, and it does so at the cost of predictable throughput interruptions scheduled without reference to actual tool condition or current WIP loading. A tool removed from production on its calendar interval during a peak-demand period imposes queue disruption comparable to an unplanned failure, without the urgency that would trigger immediate rerouting decisions [10].

2.3 Warehouse-Manufacturing Synchronisation Failures

The inventory reservation trigger in a reactive ERP architecture fires at finished goods receipt — the last possible moment before customer commitment must be confirmed [3]. In a high-mix environment, that moment coincides with many customer orders simultaneously: every lot completing the fab cycle and transferring to the warehouse during the same shift enters an allocation pool that multiple open sales orders are competing for. Reservation contention at this point is not a configuration problem; it is a timing problem created by a trigger that is, by design, as late as it can be.

Downstream of reservation, delivery, and trip creation follow the same sequential pattern: shipment scheduling cannot begin until reservation confirms, and reservation confirms only at physical receipt [22]. Warehouse teams receive the dispatch signal in a burst at the moment lot availability is confirmed, compressing the staging and trip planning window into whatever time remains before the carrier pickup. In environments where fab completion cycles are predictable within a reasonable forecast window, the information needed to begin dispatch preparation is available earlier than the reactive trigger releases it [13].

3. The Predictive Semiconductor Operations Framework (PSOF)

The Predictive Semiconductor Operations Framework proposes a four-layer architecture that embeds predictive capability within the ERP execution fabric rather than adjacent to it. Each layer feeds the next; all layers maintain bidirectional communication with the ERP transaction core to preserve lot-level traceability at every stage. Table 1 summarises the framework structure.

Layer	Primary Function	Data Inputs	ERP Integration Point	Output Signal
Data Foundation	Consolidate and pipeline all operational signals	Equipment telemetry, lot-tracking events, inventory snapshots, EDI/B2B demand	REST-enabled materialised views via OIC	Structured feature sets for analytics layer
Predictive Analytics	Generate forward-looking risk and forecast scores	ERP-sourced telemetry, lot genealogy, WIP progression, order intake	Read-only ERP data access via API pipeline	Failure risk score, yield risk score, demand forecast
Decision Intelligence	Translate predictive outputs into ERP-executable events	Failure predictions, yield risk flags, demand shortfall signals	Work order creation, inspection holds, soft-reservation triggers	Scheduled maintenance WO, enhanced metrology hold, anticipatory reservation
Execution & Feedback	Close the learning loop with confirmed outcomes	Actual yield, maintenance duration,	ERP confirmed transaction outcomes	Model retraining signals, feature importance updates

		reservation accuracy, fill rate		

Table 1. PSOF Layer Architecture. Data source: Author's analysis.

3.1 Data Foundation Layer

Four signal categories enter the PSOF data foundation layer: equipment telemetry (chamber parameters, process recipe logs, maintenance event records), lot-tracking events (WIP moves, inspection outcomes, rework triggers, genealogy records), inventory state (real-time on-hand positions, reservation commitments, open order balances), and demand signals (EDI order intake, long-term agreement schedules, customer forecast updates) [12]. These signals do not flow into an external data lake. They are accessed directly from Oracle ERP transaction tables and from equipment systems through Oracle Integration Cloud (OIC)-style bidirectional API pipelines — REST-enabled materialised views updated on intervals aligned to fab cycle times.

The design principle is ERP-data primacy. Extracting ERP data into a replicated repository creates a synchronisation gap; a predictive model trained on the replica makes predictions about a state that may already have changed in the live system [4]. In semiconductor environments where reservation commitments update in real time and lot dispositions can change between extraction cycles, that gap is operationally significant. The pipeline accesses live data; predictions are made against the current operational state rather than a historical image of it.

3.2 Predictive Analytics Layer

Three model classes operate in the analytics layer, each addressed to one of the failure modes characterised in Section 2. Long Short-Term Memory (LSTM) networks process equipment telemetry time series to identify parameter trajectory patterns that precede tool degradation — the drift signatures that develop over multiple chamber cycles before any single reading crosses a control limit [9]. The LSTM architecture is selected specifically because semiconductor tool behaviour exhibits temporal dependencies: a chamber's current parameter readings are conditioned on its recent process history, not only its instantaneous state. Feed-forward models that treat each reading as independent miss this dependency; LSTM models encode it.

Gradient boosting models — XGBoost-class algorithms applied to the lot genealogy feature space — compute yield risk scores based on a lot's accumulated parameter exposure across all completed process steps [21]. A lot with a parameter history matching the feature combinations associated with below-target yield in the training data is flagged before it reaches the final test. The forward-looking nature of this signal is what distinguishes it from SPC: SPC detects that a parameter has crossed a limit; the yield risk model predicts, from the combination of in-spec parameter values already accumulated, that the final yield outcome may be below target [20]. LSTM-based demand forecasting models additionally process B2B/EDI order intake patterns alongside WIP progression data to produce rolling lot-availability forecasts at the finished goods level [18].

3.3 Decision Intelligence Layer

Predictive signals become operational only when they are translated into ERP-executable events. Three pathways govern this translation. A failure prediction signal from the LSTM model triggers a maintenance work order in the ERP scheduled for the production window that minimises lot-queue disruption — determined by the intersection of predicted failure timing and current WIP loading — rather than at the next calendar interval [11]. A yield risk flag from the gradient boosting model triggers an inspection hold in the WIP management module, routing the flagged lot to enhanced metrology before it advances to the next process step [10]. An anticipated shortfall signal from the demand forecast triggers a soft-reservation hold against the projected lot identifiers — a pending reservation that prevents those lots from entering the available-to-promise pool for competing orders while they are still in process [3].

Each of these actions executes through existing ERP authorisation and workflow structures. The decision layer does not bypass approval stages or override planner priorities without escalation. It extends the ERP's decision-making capability forward in time while keeping the transactional governance that makes the ERP's records defensible intact.

3.4 Execution and Feedback Layer

Confirmed execution outcomes — actual yield at final test, actual maintenance duration, reservation accuracy rate, order fill rate — return to the predictive analytics layer as model performance signals [6]. LSTM failure prediction models retrain on rolling windows of confirmed outcomes. Yield correlation models update their feature importance rankings as new lot genealogy data accumulates. This feedback loop is what separates a durable predictive system from a one-time optimisation: it maintains model currency as process recipes evolve and equipment ages, treating prediction accuracy as an ongoing operational metric rather than a deployment milestone [2].

4. Predictive Manufacturing: Yield, Maintenance, and Defect Detection

Table 2 summarises the operational comparison between reactive and PSOF-informed approaches across key manufacturing metrics.

Metric	Reactive Baseline	PSOF-Informed	Direction
Equipment failure detection	Post-failure (reactive)	Pre-threshold (predictive)	Earlier detection
Maintenance scheduling	Calendar-interval, fixed	Condition-based, WIP-aligned	Reduced throughput sacrifice
Yield risk identification	Post-SPC excursion	Pre-test via lot genealogy scoring	Earlier intervention
Defect disposition latency	Manual review queue	Automated for high-confidence signatures	Reduced latency
Lot-level traceability	Preserved (transactional)	Preserved (ERP-data primacy)	No degradation
Unplanned downtime events	Unpredictable	Reduced via condition-based scheduling	Decrease

Table 2. Predictive vs. Reactive Operational Comparison. Data source: Author's analysis based on literature review.

4.1 Predictive Maintenance in Fab Environments

Condition-based maintenance changes the decision criterion from calendar time to measured equipment state [10]. The cost asymmetry that justifies this change is clear: a scheduled maintenance window that removes a healthy tool for several hours is measurably less costly than an unplanned failure that halts a bottleneck tool during peak WIP loading [11]. Time-based preventive schedules navigate this asymmetry imprecisely — they impose throughput sacrifice on a fixed calendar regardless of whether the tool actually needs attention. PSOF's predictive maintenance pathway targets that precision gap directly.

When the failure prediction model flags a tool as approaching its degradation threshold, the decision layer identifies the maintenance window that minimises queue disruption by querying current WIP routing and process compatibility data from the ERP equipment master. Lots currently assigned to the flagged tool are rerouted to qualified alternates before the maintenance window opens; the tool comes down on schedule rather than under failure conditions [10]. The result is that maintenance events — which are operationally inevitable — are absorbed by the production system with planned disruption rather than unplanned crisis response.

4.2 Yield Correlation and Optimization

Statistical process control is retrospective in a specific sense: it detects that a monitored parameter has departed from its expected distribution after the departure has occurred [9]. Lots processed during the departure window are already committed to their current parameter exposure; the SPC alert triggers containment, not prevention. The gradient boosting yield correlation model operates in a different temporal position — it scores lots based on the parameter history they have already accumulated, before they reach the test stages where yield outcomes are confirmed [21].

A lot flagged with an elevated yield risk score is routed to enhanced metrology at the next inspection point. Where metrology confirms the risk signal, engineering review can initiate process correction before additional lots accumulate the same exposure. Where metrology does not confirm the signal, the lot proceeds with a recorded review event that contributes to model refinement [20]. Over time, the model's feature importance rankings identify which parameter combinations carry the strongest predictive signal, generating process optimisation guidance grounded in empirical lot-level data rather than theoretical process models alone.

4.3 Automated Defect Detection Integration

Within a reactive ERP architecture, an inline metrology result that flags a defect creates a record that waits for human review, with wait time a function of shift staffing and queue depth [1]. PSOF's data foundation layer ingests inline defect signals, correlates them with upstream process parameter history through the lot genealogy record, and classifies detected defects against known defect signatures [10]. Lots with signatures consistent with known-recoverable conditions route to rework automatically; those consistent with unrecoverable yield loss route to scrap disposition without consuming planner time on a decision that the data already supports.

Human review is not eliminated from this process — it is targeted. Ambiguous classifications, edge cases, and signatures outside the model's training distribution are escalated for manual disposition. The practical effect is that planner attention concentrates on decisions where practitioner judgement changes the outcome, rather than being distributed uniformly across inspection events that the data already resolves.

5. Smart Warehouse Optimization in Semiconductor Supply Chains

Table 3 presents the warehouse key performance indicator comparison between reactive and PSOF-optimised operations.

KPI	Reactive Baseline	PSOF-Optimised	Improvement Direction
Inventory record accuracy	Batch-cycle dependent	Approximately 99.9% via event-driven RFID-ERP [17]	Significant increase
Order fulfilment time	Burst-triggered, compressed window	Reduced by approximately 72% via predictive dispatch [17]	Decrease
Reservation conflicts	High at finished goods receipt	Reduced via anticipatory soft-reservation	Decrease
Pick confirmation errors	RFID-triggered manual staging	Event-driven, real-time ERP update	Decrease
Operational costs	Reactive baseline	Approximately 35% reduction via predictive maintenance and automation [17]	Decrease
Carbon emissions	Unoptimised idle equipment	Approximately 25% reduction via AI-optimised routing [17]	Decrease

Table 3. Warehouse KPI Comparison. Metric values sourced from [17] where indicated; remaining entries represent directional analysis based on the author's operational experience.

5.1 ERP-Integrated Inventory Visibility

Batch-processed inventory updates create a known failure mode in high-velocity warehouse operations: the ERP inventory record reflects a state that existed at the last batch run, not the state that exists now [17]. Reservation logic operating against a stale snapshot issues availability confirmations against lots that have already been committed elsewhere; planners discover the conflict when the pick confirmation fails, not when the reservation was made. In semiconductor warehouses where finished goods arrive in completion-cycle waves, and multiple customer accounts compete for allocation simultaneously, the window between batch runs is long enough for consequential discrepancies to develop.

Event-driven RFID-to-ERP integration removes that window. Each physical lot movement — receipt, bin assignment, pick confirmation, staging — triggers an immediate ERP transaction update through an inbound interface mapped to the inventory and shipping modules [17]. The inventory record reflects the current physical state continuously. Research on intelligent warehousing systems demonstrates that RFID-enabled IoT integration can achieve approximately 99.9% inventory accuracy, compared to the periodic accuracy degradation inherent in batch-cycle systems [17]. Reservation logic operating against a continuously updated record issues availability signals that are valid at the moment they are issued.

5.2 Predictive Replenishment and Auto-Reservation Logic

Anticipatory reservation — creating a soft hold against projected lot identifiers before those lots physically arrive in the warehouse — requires a reservation mechanism that can distinguish committed inventory from available-to-promise inventory without reducing confirmed on-hand balances prematurely [3]. The soft-reservation hold serves this function: it flags projected lots as allocated to specific sales orders, preventing them from entering the available-to-promise pool for competing orders while preserving the confirmed on-hand balance until physical receipt closes the loop.

The recursive lot-quantity matching logic that governs final lot assignment — selecting combinations of lot identities that satisfy order quantities, including partial lots and softsplit scenarios — executes at the point of confirmation, not at soft-reservation creation [3]. This preserves matching precision while moving the allocation commitment earlier in the cycle. Demand-driven replenishment approaches informed by machine learning have demonstrated order fulfilment time reductions of approximately 72% compared to reactive dispatch baselines [17]. By the time a lot physically arrives in finished goods, its order assignment is already determined; the receipt transaction closes a pre-existing commitment rather than triggering a competitive allocation process [18].

5.3 Warehouse-Fab Synchronisation Through Delivery Automation

PSOF's WIP progression forecast surfaces lot completion timing with sufficient lead time to prepare dispatch resources before physical lot availability is confirmed [22]. Staging lanes can be allocated, trip creation templates can be pre-configured, and carrier notifications can be initiated against forecast completion windows rather than against the burst signal that reactive dispatch creates at the moment of lot receipt. In environments where carrier pickup schedules constrain outbound logistics, this lead time converts directly into a reduction in missed pickups and re-delivery events.

The trip grouping logic — organising delivery lines into efficient trip structures based on customer, carrier, and delivery window — can be pre-computed against forecast lot availability, reducing trip creation processing time at dispatch and enabling earlier carrier notification [13]. The advance preparation that forecast-driven dispatch enables does not replace the confirmed-state transaction that the ERP must record at execution; it moves the preparation work to a point where it adds value, rather than compressing it into the window between receipt confirmation and carrier arrival.

6. Digital Twin Integration

Table 4 details the digital twin use cases enabled by PSOF's data infrastructure.

Simulation Type	PSOF Layer Interface	Operational Benefit	Implementation Complexity
Fab capacity what-if analysis	Data Foundation + Decision Intelligence	Real-time rerouting and commitment risk assessment	Medium — requires live WIP data feed
Maintenance impact simulation	Predictive Analytics + Decision Intelligence	Pre-planned lot rerouting before tool downtime	Medium — integrates with the ERP equipment master
Warehouse flow simulation	Data Foundation + Demand Forecast	Pre-positioned resources for completion-cycle waves	Low-Medium — leverages existing RFID-ERP integration
Cross-fab routing optimisation	Decision Intelligence (extended)	Transfer lead time and capacity balancing across fabs	High — requires multi-fab data consolidation

Table 4. Digital Twin Use Cases. Data source: Author's analysis.

6.1 Manufacturing Digital Twin

The operational value of a manufacturing digital twin is not its ability to simulate; it is its ability to answer a specific class of questions rapidly: given current WIP positions, equipment availability, and maintenance schedule, what happens to lot queue depth and customer commitments if a specific condition occurs in the next shift [4]? Traditional simulation tools answer this question by loading a state snapshot and running a model offline. A PSOF-fed digital twin answers it against the current operational state — live WIP data, live equipment telemetry, live order commitments — making it useful for decisions that must be made in real time rather than only in periodic planning cycles [6].

The distinction from traditional simulation tools is the continuous data feed. Digital twin technology applied to semiconductor manufacturing has demonstrated measurable value in capacity scenario testing, maintenance impact analysis, and lot routing optimisation, particularly where deep reinforcement learning approaches are combined with real-time operational data to support adaptive production planning under demand uncertainty [4].

6.2 Warehouse and Logistics Simulation

At the warehouse level, the digital twin maps lot flow from fab completion through staging, pick confirmation, and dispatch against PSOF's completion forecasts [22]. Staging bottlenecks — lanes that will be overwhelmed by the next completion wave — are visible before the wave arrives. Handling resources can be pre-positioned; trip structures can be pre-planned. The simulation operates on the same data pipeline that drives PSOF's operational decisions, so the scenarios the warehouse team plans against reflect the same forecast state that the operational system acts on [13]. Research on simulation-based digital twin platforms for internal supply chain management confirms that continuous data integration between simulation and execution systems produces measurable improvements in warehouse throughput and delivery performance [22].

6.3 Sustainability: Energy and Waste Reduction

Predictive scheduling reduces two categories of energy waste that reactive fab operations generate without intending to. Equipment idling while awaiting lots that a predictive schedule would have sequenced more tightly consumes energy for zero productive output; lots routed to a degraded tool that a failure prediction would have removed from the queue accumulate rework requirements that consume process energy and materials a second time [11]. These are not sustainability initiatives layered over operational optimisation; they are sustainability outcomes that emerge from it [7]. AI-optimised routing and predictive maintenance automation have been associated with operational cost reductions of approximately 35% and carbon emissions reductions of approximately 25% in intelligent warehousing deployments [17], indicating that the sustainability benefits of PSOF-class architectures extend beyond the fab floor into the broader supply chain network.

7. Discussion

7.1 PSOF Against Prior Architectures

Three architectural patterns dominate the literature on predictive manufacturing intelligence in semiconductor contexts. Standalone MES-AI integrations deliver ML predictions through a separate interface that planners must consult alongside the ERP, creating a two-system decision environment that does not close the execution loop [2]. External data lake architectures provide ML flexibility but accept the synchronisation lag between replicated data and live ERP state as a manageable tradeoff — a tradeoff that proves unmanageable in semiconductor environments where lot dispositions change within the lag window [12]. ERP-module extensions constrain the predictive layer within the ERP vendor's application boundary, limiting model evolution to the vendor's development roadmap.

PSOF's ERP-data primacy design directly addresses the synchronisation risk that data lake architectures accept and the organisational complexity that two-system environments create. The primary limitation PSOF inherits from this design choice is data quality dependency: a predictive layer operating on live ERP data is only as accurate as the underlying transaction records [15]. In multi-fab environments where manual data entry exceptions, interface failures, and legacy integration inconsistencies introduce record-level noise, data quality governance becomes a prerequisite for PSOF performance rather than a background IT function. A phased deployment strategy — validating data foundation layer integrity before activating decision layer automation — addresses this risk directly and aligns with established best practices for complex integration projects in regulated manufacturing environments [16].

7.2 Implementation Considerations and Scalability

Incremental deployment is the appropriate approach for PSOF in regulated fab environments. The data foundation layer can be established as an OIC-style integration project without modifying core ERP tables, enabling data quality validation before any predictive outputs influence operational decisions [12]. The analytics layer can be trained against historical lot genealogy and equipment data before live model deployment. The decision layer's automation — maintenance scheduling, inspection holds, anticipatory reservation — can be activated module by module, with parallel human review during the validation period, ensuring that automated actions are consistent with operational expectations before full delegation occurs [16].

Multi-fab scalability introduces cross-fab routing decisions that PSOF's decision layer can accommodate by expanding its routing optimisation logic to include transfer lead times and cross-fab equipment availability [13]. The data foundation layer must be extended to consolidate telemetry and WIP data across the fab network — an integration architecture challenge rather than a framework design limitation. Organisations with existing OIC-based inter-fab integration pipelines, such as those connecting foundry partners for inter-fab transfer arrangements, are positioned to extend the PSOF data layer incrementally as each fab is brought into scope.

8. Conclusion

Predictive manufacturing intelligence and lot-level traceability are not competing requirements in semiconductor supply chains — they are requirements that a well-designed integration architecture can satisfy simultaneously. PSOF's central architectural claim is that ERP-data primacy, rather than external data replication, is the design choice that makes this simultaneous satisfaction possible. By embedding predictive analytics inside the ERP data fabric — accessing live transaction state rather than replicated snapshots, translating predictive outputs into ERP-native execution events rather than external recommendations — PSOF produces operational intelligence that is consistent with the ERP's authority over lot disposition, maintenance scheduling, and order fulfilment.

The framework does not require that semiconductor manufacturers choose between the transactional completeness that their operations depend on and the predictive capability that those operations increasingly demand. Reactive execution and predictive intelligence are complementary when the integration architecture that connects them preserves transactional integrity at every layer. The evidence reviewed here, drawn from comparable deployments in intelligent warehousing, semiconductor yield prediction, digital twin applications, and ERP-integrated manufacturing intelligence, supports the position that ERP-native predictive architectures produce more durable operational outcomes than parallel AI platforms in domains where lot-level auditability is non-negotiable.

For practitioners designing next-generation semiconductor supply chain systems, the design principles articulated in PSOF — ERP-data primacy, event-driven pipeline architecture, closed-loop model feedback, and phased deployment — provide a practitioner-grounded path toward that complementarity for multi-fab Oracle ERP environments.

REFERENCES

1. C. N. R. Kumar and S. Manimala, “Speculation of Hand Features from Middle Finger Width: A Novel Approach,” 2011 International Conference on Hand-Based Biometrics, pp. 1–5, Nov. 2011, doi: https://doi.org/10.1109/ichb.2011.6094316.
2. A. Shojaeinasab, T. Charter, M. Jalayer, M. Khadivi, O. Ogunfowora, N. Raiyani, M. Yaghoubi, and H. Najjaran, "Intelligent manufacturing execution systems: A systematic review," Journal of Manufacturing Systems, vol. 62, pp. 503-522, 2022. DOI: https://doi.org/10.1016/j.jmsy.2022.01.004. Available: https://www.sciencedirect.com/science/article/abs/pii/S0278612522000048
3. J. D. Schwartz, M. R. Arahal, D. E. Rivera, and K. D. Smith, "Control-relevant demand forecasting for tactical decision-making in semiconductor manufacturing supply chain management," IEEE Transactions on Semiconductor Manufacturing, vol. 22, no. 1, pp. 124-130, 2009. Available: https://ieeexplore.ieee.org/document/4773488/
4. H.-A. Kuo, T.-Y. Hong, and C.-F. Chien, "A deep reinforcement learning based digital twin framework for resilient production planning under demand uncertainty and an empirical study in semiconductor wafer fabrication," Computers and Industrial Engineering, vol. 208, article 111389, 2025. DOI: https://doi.org/10.1016/j.cie.2025.111389. Available: https://www.sciencedirect.com/science/article/abs/pii/S0360835225005352
5. P. D. U. Coronado et al., "Part data integration in the shop floor digital twin: Mobile and cloud technologies to enable a manufacturing execution system," Journal of Manufacturing Systems, vol. 48, pp. 25-33, 2018. Available: https://www.sciencedirect.com/science/article/abs/pii/S027861251830013X
6. F. Tao, H. Zhang, A. Liu, and A. Y. C. Nee, "Digital twin in industry: State-of-the-art," IEEE Transactions on Industrial Informatics, vol. 15, no. 4, pp. 2405-2415, 2018. DOI: 10.1109/TII.2018.2873186. Available: https://ieeexplore.ieee.org/document/8477101/
7. H. Lasi, P. Fettke, H.-G. Kemper, T. Feld, and M. Hoffmann, "Industry 4.0," Business and Information Systems Engineering, vol. 6, no. 4, pp. 239-242, 2014. DOI: 10.1007/s12599-014-0334-4. Available: https://link.springer.com/article/10.1007/s12599-014-0334-4
8. M. Grieves, "Digital twin: Manufacturing excellence through virtual factory replication," White Paper, 2015. Available: https://www.researchgate.net/publication/275211047_Digital_Twin_Manufacturing_Excellence_through_Virtual_Factory_Replication
9. D. Kim, M. Kim, and W. Kim, "Wafer edge yield prediction using a combined long short-term memory and feed-forward neural network model for semiconductor manufacturing," IEEE Access, vol. 8, pp. 215125-215132, 2020. DOI: 10.1109/ACCESS.2020.3040426. Available: https://ieeexplore.ieee.org/document/9269970/
10. D. Pradeep, B. V. Vardhan, S. Raiak, I. Muniraj, K. Elumalai, and S. Chinnadurai, "Optimal predictive maintenance technique for manufacturing semiconductors using machine learning," in Proc. 2023 3rd International Conference on Intelligent Communication and Computational Techniques (ICCT), Jaipur, India, 2023. DOI: 10.1109/ICCT56969.2023.10075658. Available: https://ieeexplore.ieee.org/document/10075658/
11. Y. Lee, Y. Roh et al., "An expandable yield prediction framework using explainable artificial intelligence for semiconductor manufacturing," Applied Sciences, vol. 13, no. 4, article 2660, 2023. DOI: https://doi.org/10.3390/app13042660. Available: https://www.mdpi.com/2076-3417/13/4/2660
12. F. Bosi et al. "Enabling smart manufacturing by empowering data integration with industrial IoT support," in 2020 International Conference on Technology and Entrepreneurship (ICTE), Bologna, Italy, 2020. DOI: 10.1109/SMARTCOMP50058.2020.00020. Available: https://ieeexplore.ieee.org/document/9215538/
13. A. Ghasemi, S. Lazarova-Molnar, and C. Heavey, "Supply chain digital twin framework for hybrid manufacturing strategy selection: A case study from the semiconductor industry," in 2024 Winter Simulation Conference (WSC), Bangkok, Thailand, 2024. DOI: 10.1109/IEEM62351.2024.10838977. Available: https://ieeexplore.ieee.org/document/10838977/

14. Y. Lee, "Data additive residual learning based yield prediction for semiconductor manufacturing," in Proc. 2023 IEEE 39th International Conference on Data Engineering (ICDE), Anaheim, CA, 2023. DOI: 10.1109/ICDE55515.2023.10184578. Available: https://ieeexplore.ieee.org/document/10184578/
15. C. Kaizar et al., "ERP systems and supply chain management performance," in Proc. 2025 16th IEEE International Conference on Logistics and Supply Chain Management (LOGISTIQUA), 2025. DOI: 10.1109/LOGISTIQUA.2025.11122856. Available: https://ieeexplore.ieee.org/document/11122856/
16. V. Harish, R. Sharma, R. K. Saini, and D. Negi, "Intelligent supply chain orchestration: A framework for seamless integration of Industry 4.0 technologies," in 2023 Seventh International Conference on Image Information Processing (ICIIP), 2024. DOI: 10.1109/ICPS59941.2024.10537750. Available: https://ieeexplore.ieee.org/document/10537750/
17. V. R. Arvind, R. M. Shrinidhi, T. Deepa, and M. Maheedhar, "Intelligent warehousing: A machine learning and IoT framework for precision inventory optimization," IEEE Access, vol. 13, 2025. DOI: 10.1109/ACCESS.2025.3614679. Available: https://ieeexplore.ieee.org/document/11181053/
18. F. Abdi, S. Abolmakarem, A. K. Yazdi, V. Popescu, R. Birau, and A. T. Mitroi, "A machine learning-based approach for multi-objective, multi-product, and multi-period supply chain optimization via demand forecasting," IEEE Access, 2025. DOI: 10.1109/ACCESS.2025.11129066. Available: https://ieeexplore.ieee.org/document/11129066/
19. N. Nikolakis, K. Alexopoulos, E. Xanthakis, and G. Chryssolouris, "The digital twin implementation for linking the virtual representation of human-based production tasks to their physical counterparts in the factories of the future," International Journal of Computer Integrated Manufacturing, vol. 32, no. 1, pp. 1-12, 2019. Available: https://www.tandfonline.com/doi/full/10.1080/0951192X.2018.1529430
20. S. McWeeny and E. S. Norton, “Understanding event-related potentials (ERPs) in clinical and basic language and communication disorders research: a tutorial,” International Journal of Language & Communication Disorders, vol. 55, no. 4, pp. 445–457, Apr. 2021, doi: https://doi.org/10.1111/1460-6984.12535.
21. S. Wang and Y. Chen, “Improved Yield Prediction and Failure Analysis in Semiconductor Manufacturing with XGBoost and Shapley Additive exPlanations Models,” 2024 IEEE International Symposium on the Physical and Failure Analysis of Integrated Circuits (IPFA), pp. 01–08, Jul. 2024, doi: https://doi.org/10.1109/ipfa61654.2024.10691196.
22. Q. Zhang, K. Chen, S. Wang, and Z. Zhang, “Green production investment policy and financing format selection for a capital-constrained manufacturer,” Computers & Industrial Engineering, vol. 194, p. 110349, Aug. 2024, doi: https://doi.org/10.1016/j.cie.2024.110349.