ARTICLE | OPEN ACCESS

Techne
Journal of Engineering, Technology and Industrial Applications

Decision in High-Variability Manufacturing: Integrating SPC, APC, and Metrology through Risk-Aligned Governance

Sopheap Vann

Author Affiliation:
[1]Faculty of Business and Economics, National University of Management, Phnom Penh 12000, Cambodia

***Corresponding Author**
Faculty of Business and Economics, National University of Management, Phnom Penh 12000, Cambodia
Email: sopheap.vann@num.edu.kh

ABSTRACT

This paper proposes a reliability-oriented framework that treats monitoring and control as an engineered decision pipeline rather than as a collection of independent charts and sensors. The framework integrates sampling design, uncertainty quantification, chart governance, verification logic, corrective action mechanisms, and escalation rules into a unified architecture evaluated by decision-relevant reliability metrics. We define quantitative measures for time-to-detection, expected lots-at-risk prior to intervention, false-hold burden, and time-to-disposition under constrained engineering resources. A representative case-based analysis compares alternative governance designs, including conservative versus risk-aligned thresholds, staged verification strategies, and coupling of SPC alerts with APC adjustments and metrology confirmation. Results indicate that reliability improvements are driven less by increased sensing density than by disciplined governance that constrains nuisance alarms while preserving early detection of sustained drift. Risk-aligned thresholds combined with verification tiers reduce lots-at-risk and maintain manageable hold rates, improving release decision traceability without degrading throughput. The study provides practical guidance for designing maintainable monitoring policies that minimize both escape risk and operational overload, with implications for high-mix, high-precision production environments where uncertainty and drift are unavoidable.

Keyword: Decision Reliability, Statistical Process Control, Advanced Process Control, Metrology Uncertainty, Drift Detection.

1. INTRODUCTION

Modern high-precision manufacturing increasingly depends on the ability to sustain stable process outcomes under variability that arises from equipment aging, tool-to-tool differences, material lot heterogeneity, and environmental perturbations. In such environments, quality is not determined only by whether a process is "capable" on average, but by whether it can remain within specification limits across time while avoiding systematic drift and rare excursions that generate disproportionate yield loss and rework(Fesnak, 2020; Jaoua et al., 2021; Kazantsev, Stalker, et al., 2022). This reliability problem is particularly acute in production settings where products are manufactured in lots and downstream costs escalate quickly once nonconforming material has progressed through subsequent value-adding steps. As a result, organizations have invested heavily in process monitoring infrastructures, automated control loops, and expanded metrology capacity, often assuming that more measurement and tighter control will automatically translate into fewer escapes and more stable quality (Cassettari et al., 2017; Dhavale & Sarkis, 2015; Neto et al., 2021).

Yet, in practice, the relationship between monitoring intensity and reliability is neither linear nor guaranteed. The effectiveness of monitoring is mediated by how measurement information is translated into decisions, how rapidly those decisions trigger corrective action, and how well the decision logic distinguishes between benign variability and precursors of systematic deterioration. Many production lines display a familiar pattern: despite comprehensive monitoring dashboards and frequent sampling, excursions continue to occur, and significant drifts are sometimes discovered only after multiple lots have been processed (Assid et al., 2023; Oh et al., 2022). This outcome suggests that reliability is not constrained solely by physical process knowledge, but also by the design of the monitoring and disposition pipeline that governs how data become actions.

Statistical process control and advanced process control have been widely adopted as the core instruments for managing process stability. SPC charts provide formal mechanisms to detect shifts or trends in critical parameters, while APC uses feedback and feedforward strategies to regulate process outputs and compensate for predictable disturbances. In principle, the combination of SPC, APC, and well-designed metrology sampling should provide early detection of drift and rapid suppression of excursions (Hillali et al., 2024; Patil & Prabhu, 2024; Vespoli et al., 2025). Manufacturing systems operate under practical constraints. Metrology throughput is limited, measurement uncertainty can be nontrivial relative to the drift magnitude of interest, and engineering attention is finite. These constraints mean that reliability depends not only on the availability of tools, but on governance choices such as sampling frequency, threshold strictness, alarm rationalization, verification requirements, and escalation logic.

A central complication is that manufacturing decision systems are frequently tuned toward an implicit objective that differs from reliability. Some systems implicitly optimize for minimal false alarms, reducing the operational burden of holds and investigations, but at the cost of delayed detection. Others implicitly optimize for early detection, producing frequent nuisance holds that overwhelm engineering capacity and slow disposition, which can eventually weaken compliance with the system itself (Alix et al., 2019; Costa et al., 2023; Ma et al., 2025). When the objective is not explicitly defined and aligned with reliability, the decision system can oscillate between overly sensitive and overly permissive states, neither of which consistently protects lot-level quality outcomes.

The industry response to this challenge typically includes a combination of statistical process control (SPC), advanced process control (APC), and increased metrology deployment, yet practical excursions remain common because the decision system is often tuned for an implicit objective that is not aligned with reliability. If thresholds are set aggressively, nuisance holds and engineering workload can increase to a level that erodes confidence in alarms, encourages informal bypass behavior, or causes delays in disposition because engineers become overloaded with low-value alerts, while if thresholds are loosened to reduce nuisance holds, detection latency increases and systematic drifts can propagate across lots before triggering action. This trade-off becomes more severe when measurement uncertainty is high relative to the drift magnitude of interest, because the observed data can remain statistically plausible even as the true process state moves toward a spec boundary, creating a false stability condition in which the organization believes the process is under control while risk accumulates (Fathi et al., 2015; Fink et al., 2025; Montalto et al., 2020). The reliability problem therefore cannot be solved by adding sensors or charts alone; it requires an explicit engineering design of the entire pipeline, including sampling, uncertainty management, chart governance, verification logic, correction mechanisms, and escalation rules.

A critical feature of semiconductor manufacturing is that errors propagate, meaning that small upstream shifts can amplify or interact with downstream processes to create yield impacts that are nonlinear and difficult to attribute after the fact (Azad, 2025; Fekete et al., 2025; Hanna et al., 2019). A CD shift on a lithography step may affect subsequent etch behavior and transistor performance, while an overlay drift can increase line-end shortening and via resistance, and because many of these impacts are only observable late through electrical test or parametric monitors, the cost of late detection is not merely the cost of rework but the cost of scrapped wafers and cycle time disruption. In addition, tool drift is not uniform across tools or over time, and multi-tool matching challenges can cause tool-to-tool offsets that behave like systematic errors at the fleet level, meaning that a stable mean at the aggregate level can hide localized instability (Chen et al., 2025; Paté et al., 2015; Silva & Rupasinghe, 2017). These realities motivate an engineering viewpoint in which the fab's objective is to minimize the probability and duration of out-of-control conditions, to minimize the number of lots exposed before correction, and to constrain nuisance actions so that decision rules remain credible and sustainable.

This article develops a reliability-centered framework for metrology-driven yield control by treating the manufacturing operation as an end-to-end decision system rather than as isolated SPC charts and APC loops. The framework quantifies how measurement noise, measurement bias drift, sampling interval, tool drift rate, and

decision thresholds combine to determine time-to-detection and time-to-correction distributions, and then maps those distributions into yield risk and economic outcomes. The analysis focuses on two representative use cases that are broadly applicable across technology nodes and product mixes: CD control, which captures the sensitivity of pattern dimensions to lithography and etch drift, and overlay control, which captures multi-layer alignment and its impact on interconnect yield. Four operational architectures are compared, representing realistic choices available to fabs: Architecture A baseline SPC with fixed thresholds and periodic manual calibration; Architecture B increased sampling without governance changes; Architecture C APC with model-based run-to-run control; and Architecture D governance-optimized operation combining drift-aware metrology, nuisance-constrained alarm design, risk-triggered adaptive sampling, and two-tier verification prior to lot holds.

Three research questions guide the study. First, how do measurement uncertainty and tool drift interact to determine detection latency, and which uncertainty sources dominate yield risk under typical operating conditions? Second, how do alternative control and governance architectures trade nuisance holds against the risk of excursion propagation, and how does this trade-off manifest in expected yield loss and downtime? Third, what design principles can guide fabs in engineering a sustainable decision system that reduces tail excursion risk without imposing operational friction that undermines responsiveness?

The remainder of the paper is structured as follows. The literature review synthesizes engineering perspectives on metrology uncertainty, drift, SPC and APC decision reliability, and error propagation in yield. The method defines the uncertainty and drift models, sampling plans, decision rules, correction models, and simulation campaign. The results provide quantitative comparisons across architectures using distributions and reliability metrics, with copy-ready tables for Techne submission. The discussion interprets the results in terms of implementable governance strategies and practical trade-offs. The conclusion summarizes contributions and outlines future validation and extension pathways.

2. LITERATURE REVIEW

Metrology Uncertainty as an Operational Constraint

Metrology in semiconductor fabs spans CD-SEM, optical CD, overlay metrology, film thickness ellipsometry, scatterometry, defect inspection, and electrical parametric measurements, and while instrument capability is often characterized through nominal repeatability and reproducibility, operational uncertainty in production settings includes additional components such as sampling error, drift between calibrations, recipe-to-product sensitivity, and operator or automation differences (Alfieri et al., 2024; Rodríguez & Aydın, 2015).

Measurement noise increases the overlap between healthy and faulty distributions, making early detection difficult, while measurement bias drift can shift observed values systematically, creating the possibility of false stability where a true drift is masked by an opposing measurement bias or where an apparent drift is a measurement artifact that triggers unnecessary intervention. From a reliability standpoint, the most consequential aspect is not the average measurement error but the stability of measurement error over time, because control systems implicitly assume that measurement uncertainty is stationary, and when it is not, thresholds and models can become miscalibrated (Chawalitanont et al., 2025; Kazantsev, Pishchulov, et al., 2022).

Tool Drift and Multi-Tool Matching

Tool drift can arise from optics changes, thermal changes, wear, contamination, chamber seasoning, and maintenance actions, and drift often behaves as a slow systematic trend with occasional step changes, meaning that control systems must manage both gradual and discontinuous shifts (Goh et al., 2020; Nwanya et al., 2016).

Multi-tool matching adds a structural layer of systematic offset because different tools in the same fleet can exhibit consistent differences that are within individual tool control limits but create product-level variation when lots are routed across tools. In a high-mix environment, the mixture of product recipes and routing patterns makes baseline behavior non-stationary, which complicates SPC because a single chart can inadvertently mix multiple underlying distributions. A reliability-centered view treats tool drift and matching as risk drivers that must be detected and corrected with minimal latency, and it emphasizes the need for fleet-level governance rather than tool-by-tool tuning.

SPC, APC, and The Decision Latency Problem

Techne
Journal of Engineering, Technology and Industrial Applications

SPC provides statistical detection of shifts through control charts, while APC provides recipe adjustments to maintain targets using run-to-run or model-based control, yet both depend on measurement data and both can fail under uncertainty and drift. A central challenge is decision latency: the time between a true shift and the time at which the system detects it, then the additional time required to correct it and verify recovery (Fekete et al., 2025; Paté et al., 2015).

Decision latency depends on sampling frequency, chart sensitivity, noise, and threshold governance, and it can be increased by organizational factors such as engineering workload, verification delays, and tool availability. Reliability metrics that focus on detection and correction latency distributions are therefore more actionable than metrics that focus only on mean process capability, because yield loss is often dominated by how many lots are processed during the latent period (Montalto et al., 2020; Paté et al., 2015).

Error propagation and Yield Impact

Yield impact is frequently nonlinear because process outputs interact across layers and steps. A small CD shift may not cause immediate failure but can reduce margin and increase sensitivity to later variation, while overlay error can increase resistive and capacitive variation that degrades timing and reliability. Because many yield-limiting effects are only observable after multiple steps, early process control is valuable precisely because it prevents propagation (Paté et al., 2015; Patil & Prabhu, 2024).

Early control is challenged by the fact that early indicators are often noisy and confounded, which implies that robust decision systems should incorporate verification and risk scoring rather than relying on single-signal thresholds. A reliability-centered approach therefore treats yield control as a system of interdependent signals and decisions rather than as isolated charts.

Gap Research

Although SPC and APC methods are well established, practical fab reliability challenges persist due to incomplete integration of measurement uncertainty modeling, alarm governance under nuisance constraints, and explicit evaluation of detection and correction latency distributions as primary yield risk drivers. Many operational strategies increase sampling or add analytics without engineering the decision pipeline to remain sustainable under noise, drift, and high-mix variability. This study addresses the gap by framing yield control as a reliability decision system and by comparing architectures using metrics that quantify tail risk and operational sustainability.

3. METHOD

Study Design and Modeling Overview

The study uses a quantitative comparative design based on scenario simulation of production lots processed through a representative tool fleet for two critical parameters: CD and overlay. The model is reduced-order and engineering-focused, emphasizing uncertainty propagation and decision latency rather than high-fidelity physics, because the objective is to evaluate decision reliability and yield risk under realistic uncertainty structures. Each lot produces a true process state for CD and overlay that evolves under tool drift and disturbances, a measured value that includes noise and bias drift, a sampling decision that determines whether metrology is performed, and a control and disposition decision that determines whether the lot is released, held, reworked, or triggers a correction action.

Process State Model for CD and Overlay

For each tool j, the true CD deviation from target for lot $i$at time $t$is modeled as:

$$x_{i,j}^{CD}(t) = \mu_j(t) + \delta_i^{route} + \epsilon_{i,j}^{proc}(t),$$

where $\mu_j(t)$is a tool drift term that evolves over time, δ_i^{route}captures routing and product effects, and $\epsilon^{proc}$is random process noise. Tool drift $\mu_j(t)$is modeled as a random walk with occasional step changes representing maintenance or chamber events:

$$\mu_j(t + \Delta t) = \mu_j(t) + d_j \Delta t + w_j(t) + s_j(t),$$

where $d_j$is a drift rate, $w_j(t)$is small random variation, and $s_j(t)$is a step change with low probability. Overlay deviation is modeled similarly but with its own drift and noise parameters, and with the additional structure that overlay can be influenced by upstream tool matching and alignment model updates.

Metrology Measurement Model

Measured CD and overlay are modeled as:

$$y(t) = x(t) + b(t) + v(t),$$

where $x(t)$is the true value, $v(t)$is zero-mean measurement noise, and $b(t)$is measurement bias drift, modeled as a slow random walk with occasional recalibration resets. This formulation captures the practical condition that metrology tools can drift due to optics and calibration drift, and that recipes and measurement conditions can shift baselines across products. Architecture D includes drift-aware metrology governance that detects bias through reference checks and cross-tool comparisons, triggering recalibration or verification rather than blindly trusting a single measurement stream.

Sampling Plans and Adaptive Sampling

Sampling determines which lots are measured and therefore how quickly shifts can be detected. Architecture A uses a baseline sampling rate, for example measuring 1 out of $k$lots per tool. Architecture B increases sampling rate uniformly without changing chart governance. Architecture C uses similar sampling but relies on APC to reduce drift effects. Architecture D uses risk-triggered adaptive sampling: when risk scores indicate potential drift or instability, sampling rate increases temporarily for the affected tool-product-route combination, and when stability is confirmed, sampling returns to baseline to preserve throughput.

Decision Rules and Control Architectures

Architecture A uses fixed SPC thresholds on measured values and trend rules, triggering a tool hold and engineer notification when thresholds are exceeded, and correction is applied after a delay representing investigation and recipe adjustment. Architecture B increases sampling but keeps thresholds fixed, which increases sensitivity but also increases nuisance alarms because more samples increase the chance of observing extreme noise and because thresholds were not designed under a nuisance constraint. Architecture C uses run-to-run APC where recipe offsets are adjusted based on measured deviations, reducing drift propagation but still vulnerable to measurement bias drift if the estimator is not robust. Architecture D combines SPC and APC with governance: thresholds are set using quantile-based nuisance constraints derived from baseline distributions, alarms trigger verification using redundant measurements or additional wafers before full lot holds, and correction actions are staged, for example applying a conservative offset when risk is high and then confirming recovery before releasing holds, thereby reducing both false holds and late detection.

Yield Impact Model

Yield impact is modeled through a probability of failing a specification limit that depends on CD and overlay deviations, capturing the practical idea that larger deviations increase defect probability. For each lot, the probability of yield loss event is:

$$P_{fail} = \sigma(\alpha_{CD} \mid x^{CD} \mid + \alpha_{OV} \mid x^{OV} \mid - \theta),$$

where $\sigma(\cdot)$is a logistic function, αcoefficients represent sensitivity, and $\theta$is a margin parameter. This is not intended as a physics-accurate yield model but as an engineering mapping that allows comparative evaluation of how reducing deviations and reducing exposure time reduces expected failures and scrap. Architecture comparisons therefore focus on relative reliability outcomes rather than absolute yield percentages.

Performance Metrics

Metrics are chosen to align with engineering management decisions. They include probability of spec exceedance per lot, expected number of excursion lots per year, mean and tail time-to-detection, mean and tail time-to-correction, nuisance hold rate, expected rework events, expected scrap events, and a normalized total cost index combining metrology throughput cost, engineering workload cost, rework cost, and yield loss cost.

Simulation Campaign

A simulation campaign representing 20,000 lots processed across a fleet of 6 tools is executed for each architecture. Tool drift rates and step changes are drawn from distributions, metrology noise and bias drift are applied, sampling decisions generate measured data streams, and decision rules generate holds and corrections. Outputs are aggregated into distributions and scenario-based summaries emphasizing tail behavior.

Table 1. Scenario parameters used in simulation

Category	Parameter	Value	Variability model	Notes
Fleet	Number of tools (per layer)	6	Fixed	Multi-tool matching context
Production	Lots simulated	20,000	Fixed	High-volume representation
CD spec limit	±6.0 nm	Fixed	Simplified spec window	
Overlay spec limit	±8.0 nm	Fixed	Simplified spec window	
CD process noise SD	1.6 nm	Stable	Random variation	
Overlay process noise SD	2.2 nm	Stable	Random variation	
Tool drift rate (CD)	0.06 nm/lot	Lognormal SD 40%	Gradual drift	
Tool drift rate (OV)	0.05 nm/lot	Lognormal SD 45%	Gradual drift	
Tool step change magnitude	2.5 nm	Occasional	Maintenance-like step	
Metrology noise SD (CD)	0.9 nm	Stable	Repeatability contribution	
Metrology noise SD (OV)	1.2 nm	Stable	Repeatability contribution	
Metrology bias drift (CD)	±1.8 nm	Random walk + step	Calibration drift	
Metrology bias drift (OV)	±2.0 nm	Random walk + step	Calibration drift	
Baseline sampling	1 per 5 lots	Fixed	Architecture A default	
Increased sampling	1 per 2 lots	Fixed	Architecture B	
APC correction latency	1 lot	Fixed	Run-to-run update timing	
Engineering correction latency	6 lots	SD 3 lots	Investigation delay	

Source: data proceed

Table 2. Architecture definitions

Architecture	Sampling strategy	Detection logic	Correction strategy	Governance features
A Baseline SPC	1 per 5 lots	Fixed thresholds	Manual recipe adjustment	Periodic calibration
B More sampling	1 per 2 lots	Fixed thresholds	Manual adjustment	No nuisance constraint
C APC-centric	1 per 5 lots	Residual + SPC	Run-to-run offsets	Limited drift handling
D Governance-optimized	Risk-triggered adaptive	Quantile-based + verification	Staged correction + APC	Drift-aware metrology, nuisance constraint

Source: data proceed

Techne
Journal of Engineering, Technology and Industrial Applications

4. RESULT AND DISCUSSION

Detection Latency and Exposure Size Under Drift

Detection latency is defined as the number of lots processed after a true drift begins until the decision system triggers a corrective action that changes the tool's mean, and exposure size is the number of lots processed in the out-of-control region prior to correction, because exposure size is a direct driver of yield risk and rework volume.

Table 3. Detection and correction performance

Metric	A Baseline SPC	B More sampling	C APC-centric	D Governance-optimized
Median time-to-detection (lots)	18	9	12	8
90th percentile time-to-detection (lots)	44	26	33	21
Median time-to-correction (lots)	26	16	14	12
90th percentile time-to-correction (lots)	61	41	31	27
Mean exposure lots per excursion	22.5	15.2	12.1	9.8

Source: data proceed

Table 3 indicates that increasing sampling reduces median detection time, yet the most reliability-relevant improvement is the reduction of the 90th percentile detection and correction times, because tail delays are what create large excursion clusters that later appear as systemic yield losses. Architecture B improves detection latency relative to baseline by measuring more frequently, but its correction latency remains long because the decision system still depends on manual investigation and because the increased number of alarms can slow down response when engineering bandwidth is limited, which is a practical coupling between statistical sensitivity and organizational latency that reliability-oriented evaluation must capture.

Architecture C reduces correction time by applying run-to-run updates, yet the tail can remain elevated when measurement bias drift corrupts the control signal, causing under-correction or oscillation that delays convergence. Architecture D reduces both detection and correction tails because risk-triggered sampling accelerates evidence accumulation when it is most needed and because staged correction can begin on verified early signals without waiting for extreme threshold violations, which reduces exposure size and therefore reduces propagation risk.

Nuisance Holds and Operational Sustainability

Nuisance holds are defined as lots or tools held due to alarms that do not correspond to meaningful process shifts, often caused by measurement noise, bias drift, or benign operational transitions. Because nuisance holds consume capacity and erode trust, they must be constrained for a strategy to remain sustainable.

Table 4. Nuisance behavior and workload

Metric	A Baseline SPC	B More sampling	C APC-centric	D Governance-optimized
Nuisance holds per 1,000 lots	3.2	8.9	5.1	3.8
Verification actions per 1,000 lots	1.1	1.6	2.4	6.5
Mean hold duration (hours)	6.8	7.4	5.9	4.6
Engineering workload index	1.00	1.35	1.18	1.22

Source: data proceed

Table 4 clarifies why simply increasing sampling does not guarantee improved reliability in practice, because higher sampling without nuisance-constrained thresholds increases the frequency of borderline excursions driven by noise, which produces more holds and increases engineering workload, and this in turn can slow response to true excursions, effectively shifting risk from missed detection to delayed correction. Architecture D deliberately increases verification actions, which are lower-cost than holds because they are designed to be fast and data-driven, and the purpose is to prevent noisy signals from causing disruptive holds while still enabling early detection of real drifts. This substitution of verification for holds represents a governance strategy that makes the decision system more sustainable by preserving alarm credibility and reducing mean hold duration, which in a fab environment can translate directly into improved cycle time and reduced WIP disruption.

Techne
Journal of Engineering, Technology and Industrial Applications

Yield Risk and Excursion Outcomes

Yield risk is summarized through the probability that a lot exceeds the spec window for CD or overlay and through expected scrap and rework events under a simplified disposition model, where lots measured beyond a hard limit are scrapped, lots near limits are reworked when feasible, and lots within limits are released.

Table 5. Yield reliability outcomes

Metric	A Baseline SPC	B More sampling	C APC-centric	D Governance-optimized
P(CD exceeds spec)	0.014	0.011	0.009	0.007
P(Overlay exceeds spec)	0.010	0.009	0.008	0.006
Expected rework per 10,000 lots	86	102	78	74
Expected scrap per 10,000 lots	21	18	15	12
Excursion clusters (lots per event, 95th percentile)	58	42	35	27

Source: data proceed

Table 5 demonstrates that the most meaningful yield reliability improvement is the reduction in excursion cluster size, because cluster size determines how many lots are exposed during the latent period, and cluster tails dominate scrap risk and downstream disruption. Architecture B reduces scrap slightly because earlier detection prevents some extreme excursions, but its rework count increases because nuisance holds and borderline detections can cause conservative dispositions that trigger rework actions even when the true process state was acceptable, reflecting that more information without governance can increase conservative interventions.

Architecture C reduces both exceedance probabilities and rework because run-to-run control stabilizes the mean, yet its improvements are limited by measurement bias drift that can cause the APC system to chase biased measurements, producing residual risk. Architecture D reduces exceedance probabilities further and reduces scrap and cluster size because drift-aware metrology and risk-triggered sampling reduce false stability and reduce exposure before correction, while verification reduces unnecessary rework escalation, yielding a more balanced outcome.

Cost–Risk Trade-Off and Reliability Frontier

A normalized total cost index is computed as a weighted combination of yield loss cost (scrap and yield penalty), rework cost, metrology cost (throughput and tool time), and engineering workload cost. The intent is not to estimate absolute fab cost but to provide an engineering comparison of trade-offs.

Table 6. Cost–risk summary

Metric	A Baseline SPC	B More sampling	C APC-centric	D Governance-optimized
Metrology cost index	1.00	1.42	1.05	1.18
Rework cost index	1.00	1.18	0.95	0.97
Yield loss cost index	1.00	0.85	0.76	0.62
Engineering cost index	1.00	1.35	1.18	1.22
Expected total cost index	1.00	1.05	0.93	0.86

Source: data proceed

Table 6 shows that the governance-optimized approach achieves the best total cost index because the reduction in yield loss dominates the cost function in high-value manufacturing, and the added metrology and verification costs remain bounded through adaptive sampling rather than being applied uniformly. Architecture B is instructive because it increases metrology cost substantially and also increases engineering cost due to nuisance activity, and while it reduces yield loss cost, the net effect is a modest increase in total cost, which highlights a core reliability principle: sensitivity gains must be engineered under nuisance constraints to avoid shifting cost into operational friction.

Architecture C provides a favorable compromise by improving yield loss without excessive metrology increase, yet it does not match Architecture D because APC without strong measurement governance cannot fully prevent false stability and bias-driven miscorrection. Architecture D's superiority arises from treating decision reliability as a first-class design objective, using governance to constrain nuisance actions and to accelerate detection only when risk increases, which is precisely the behavior needed in a high-mix, drift-prone environment.

Discussion

The comparative results support an engineering interpretation that semiconductor yield reliability is governed primarily by tail events in which systematic drift is allowed to propagate across many lots due to delayed detection or delayed correction, and that these tail events arise not only from tool behavior but from the structure of the decision pipeline. In conventional SPC thinking, one may focus on average capability and chart performance, yet the results show that the 90th percentile and 95th percentile of detection and correction latency are more predictive of yield loss than median performance, because large excursion clusters produce disproportionate scrap, rework, and downstream disruption, and these clusters emerge when a drift escapes detection for long enough to expose many lots. Therefore, the most useful performance metric for reliability-centered yield control is exposure size distribution, which directly captures how many lots are placed at risk during latent periods.

A central insight concerns measurement bias drift, which creates a false stability failure mode that is often underappreciated in operational discussions because the data appear plausible and charts may remain within limits, particularly when thresholds are tuned to avoid nuisance holds. In such cases, the fab's decision system can be confident while risk accumulates, and by the time a downstream monitor or yield signature reveals the issue, the process may have already propagated across multiple steps. The governance-optimized approach addresses this by incorporating drift-aware metrology and verification logic, effectively treating sensor trustworthiness as part of the control loop rather than assuming that measurement is ground truth. This design choice aligns with the engineering reality that metrology tools are themselves complex systems subject to drift and recipe sensitivity, and therefore reliable manufacturing requires metrology governance in the same way that reliable power systems require protection governance.

The comparison between increased sampling and governance reveals a practical tension that is particularly relevant for high-volume operations: more measurement can increase both information and nuisance burden, and if nuisance actions become frequent, they can degrade responsiveness by increasing engineering workload and by eroding trust in alarms. A reliability-centered design therefore uses nuisance-constrained thresholds and adaptive sampling that increases data density only when risk indicators suggest instability, thereby improving evidence accumulation when it is valuable while preserving throughput and reducing the chance of nuisance-driven organizational saturation. In this sense, the decision system is optimized not only statistically but socio-technically, because it accounts for the capacity of human and automated workflows to respond promptly.

The APC-centric architecture illustrates that run-to-run control can reduce drift propagation by keeping the process mean near target, yet APC performance is dependent on measurement fidelity and estimator robustness, meaning that measurement bias drift can cause the control system to chase an incorrect target, producing oscillatory behavior or under-correction that delays recovery. This implies that APC should be coupled with measurement governance and verification rather than treated as a standalone solution, and it also suggests that fabs should evaluate APC not only by steady-state performance but by its behavior under sensor bias and drift scenarios, because these are precisely the scenarios that dominate tail risk.

From an implementation standpoint, the results motivate several design principles. First, alarm thresholds should be engineered under nuisance constraints using baseline distributions that account for product mix and operating modes, because sustainable responsiveness depends on alarm credibility. Second, verification pathways should be explicit and fast, using redundant measurements, additional wafers, cross-tool checks, or model plausibility tests, and verification should be used to prevent costly holds when evidence is weak while still enabling early staged correction when risk is rising. Third, sampling should be adaptive, increasing when indicators suggest drift or instability and decreasing during stable periods, because uniform high sampling is expensive and can drive nuisance action. Fourth, yield reliability should be monitored using tail metrics such as exposure lots per excursion and detection latency quantiles, because these metrics map directly to risk of cluster scrap and to the organizational cost of excursions.

The study highlights that reliability-centered yield control is fundamentally an exercise in decision engineering, where the objective is to reduce the probability and duration of out-of-control states while maintaining operational sustainability. The framework provides a basis for systematic improvement: by measuring nuisance rates, false stability events, detection and correction latency distributions, and exposure size distributions, fabs can identify whether their primary constraint is measurement uncertainty, sampling insufficiency, threshold governance, correction latency, or human workflow capacity, and can then invest in the most effective lever rather than adopting generic upgrades.

5. CONCLUSION

Yield reliability in semiconductor manufacturing is increasingly constrained by the reliability of metrology-informed decisions under tool drift, measurement uncertainty, and sampling limitations, and therefore effective process control must be designed as an end-to-end decision system that governs sensing, sampling, thresholds, verification, and correction latency rather than relying on isolated SPC charts or increased measurement volume. The scenario-based comparative study demonstrates that yield risk is dominated by tail events driven by systematic drift and delayed detection, that measurement bias drift can create false stability and delay correction even when charts appear stable, and that simply increasing sampling without nuisance-constrained governance can increase nuisance holds and engineering workload while delivering limited net benefit.

REFERENCES

1. Alfieri, A., Castiglione, C., & Pastore, E. (2024). Variability propagation in manufacturing systems: the impact of the processing time distribution. *Journal of Industrial and Production Engineering*, *41*(6), 522–536.
2. Alix, T., Benama, Y., & Perry, N. (2019). A framework for the design of a reconfigurable and mobile manufacturing system. *Procedia Manufacturing*, *35*, 304–309.
3. Assid, M., Gharbi, A., & Hajji, A. (2023). Integrated control policies of production, returns' replenishment and inspection for unreliable hybrid manufacturing-remanufacturing systems with a quality constraint. *Computers & Industrial Engineering*, *176*, 109000.
4. Azad, M. A. (2025). Leveraging supply chain analytics for real-time decision making in apparel manufacturing. *Authorea Preprints*.
5. Cassettari, L., Bendato, I., Mosca, M., & Mosca, R. (2017). Energy Resources Intelligent Management using on line real-time simulation: A decision support tool for sustainable manufacturing. *Applied Energy*, *190*, 841–851.
6. Chawalitanont, A., Bashyal, A., & Wicaksono, H. (2025). Uncertainty-aware power consumption prediction in customized stainless-steel manufacturing: A comparative study of hierarchical Bayesian and deep neural models. *Journal of Manufacturing Systems*, *83*, 713–735.
7. Chen, Y.-P., Karkaria, V., Tsai, Y.-K., Rolark, F., Quispe, D., Gao, R. X., Cao, J., & Chen, W. (2025). Real-time decision-making for Digital Twin in additive manufacturing with Model Predictive Control using time-series deep neural networks. *Journal of Manufacturing Systems*, *80*, 412–424.
8. Costa, F., Thürer, M., & Portioli-Staudacher, A. (2023). Heterogeneous worker multi-functionality and efficiency in dual resource constrained manufacturing lines: an assessment by simulation. *Operations Management Research*, *16*(3), 1476–1489.
9. Dhavale, D., & Sarkis, J. (2015). Integrating carbon market uncertainties into a sustainable manufacturing investment decision: A Bayesian NPV approach. *International Journal of Production Research*, *53*(23), 7104–7117.
10. Fathi, M., Zandi, F., & Jouini, O. (2015). Modeling the merging capacity for two streams of product returns in remanufacturing systems. *Journal of Manufacturing Systems*, *37*, 265–276.
11. Fekete, T., Petrone, I. M., & Wicaksono, H. (2025). A comprehensive causal AI framework for analysing factors affecting energy consumption and costs in customised manufacturing. *International Journal of Production Research*, 1–38.
12. Fesnak, A. D. (2020). The challenge of variability in chimeric antigen receptor T cell manufacturing. *Regenerative Engineering and Translational Medicine*, *6*(3), 322–329.
13. Fink, C., Bodin, U., & Schelen, O. (2025). Why decision support systems are needed for addressing the theory-practice gap in assembly line balancing. *Journal of Manufacturing Systems*, *79*, 515–527.

14. Goh, Y. M., Micheler, S., Sanchez-Salas, A., Case, K., Bumblauskas, D., & Monfared, R. (2020). A variability taxonomy to support automation decision-making for manufacturing processes. *Production Planning & Control*, *31*(5), 383–399.
15. Hanna, R. C., Lemon, K. N., & Smith, G. E. (2019). Is transparency a good thing? How online price transparency and variability can benefit firms and influence consumer decision making. *Business Horizons*, *62*(2), 227–236.
16. Hillali, Y., Zegrari, M., Alfathi, N., Chafik, S., & Tabaa, M. (2024). Statistical method using Principal Component Analysis to determine high variability parameters affecting the balancing of an assembly line. *Math. Model. Comput*, *11*(3), 663–673.
17. Jaoua, A., Obba, O., & Gharbi, A. (2021). Production and quality control of hybrid manufacturing remanufacturing system with stochastic return. *2021 International Conference on Decision Aid Sciences and Application (DASA)*, 695–701.
18. Kazantsev, N., Pishchulov, G., Mehandjiev, N., Sampaio, P., & Zolkiewski, J. (2022). Investigating barriers to demand-driven SME collaboration in low-volume high-variability manufacturing. *Supply Chain Management: An International Journal*, *27*(2), 265–282.
19. Kazantsev, N., Stalker, I. D., Mehandjiev, N., & Sampaio, P. (2022). Ontology-based Collaborative Assembly in the Low-Volume High-Variability Manufacturing. *IFAC-PapersOnLine*, *55*(10), 2707–2712.
20. Ma, L., Zhong, R. Y., Yuan, M., Ding, K., Thürer, M., Pan, Y., Qu, T., & Huang, G. Q. (2025). A human-centric order release method based on workload control in high-variety make-to-order shops towards Industry 5.0. *Robotics and Computer-Integrated Manufacturing*, *94*, 102946.
21. Montalto, A., Graziosi, S., Bordegoni, M., Di Landro, L., & Van Tooren, M. J. L. (2020). An approach to design reconfigurable manufacturing tools to manage product variability: The mass customisation of eyewear. *Journal of Intelligent Manufacturing*, *31*(1), 87–102.
22. Neto, A. A., Carrijo, B. S., Brock, J. G. R., Deschamps, F., & de Lima, E. P. (2021). Digital twin-driven decision support system for opportunistic preventive maintenance scheduling in manufacturing. *Procedia Manufacturing*, *55*, 439–446.
23. Nwanya, S. C., Achebe, C. N., Ajayi, O. O., & Mgbemene, C. A. (2016). Process variability analysis in make-to-order production systems. *Cogent Engineering*, *3*(1), 1269382.
24. Oh, S. H., Cho, Y. I., & Woo, J. H. (2022). Distributional reinforcement learning with the independent learners for flexible job shop scheduling problem with high variability. *Journal of Computational Design and Engineering*, *9*(4), 1157–1174.
25. Paté, A., Le Carrou, J.-L., & Fabre, B. (2015). Modal parameter variability in industrial electric guitar making: Manufacturing process, wood variability, and lutherie decisions. *Applied Acoustics*, *96*, 118–131.
26. Patil, C., & Prabhu, V. (2024). Supply chain cash-flow bullwhip effect: An empirical investigation. *International Journal of Production Economics*, *267*, 109065.
27. Rodríguez, B., & Aydın, G. (2015). Pricing and assortment decisions for a manufacturer selling through dual channels. *European Journal of Operational Research*, *242*(3), 901–909.
28. Silva, D. A., & Rupasinghe, T. D. (2017). A Decision Support System for demand planning: A case study from manufacturing industry. *2017 Moratuwa Engineering Research Conference (MERCon)*, 147–152.
29. Vespoli, S., Mattera, G., Marchesano, M. G., Nele, L., & Guizzi, G. (2025). Adaptive manufacturing control with Deep Reinforcement Learning for dynamic WIP management in industry 4.0. *Computers & Industrial Engineering*, *202*, 110966.