



A configuration-specific analytical evaluator for energy-aware capacity reconfiguration in two-class CONWIP systems

Maria De Martino, Silvestro Vespoli, Andrea Grassi & Guido Guizzi

To cite this article: Maria De Martino, Silvestro Vespoli, Andrea Grassi & Guido Guizzi (02 Sep 2026): A configuration-specific analytical evaluator for energy-aware capacity reconfiguration in two-class CONWIP systems, International Journal of Production Research, DOI: [10.1080/00207543.2026.2723544](https://doi.org/10.1080/00207543.2026.2723544)

To link to this article: <https://doi.org/10.1080/00207543.2026.2723544>



© 2026 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.



[View supplementary material](#)



Published online: 02 Sep 2026.



[Submit your article to this journal](#)



Article views: 215



[View related articles](#)



[View Crossmark data](#)

A configuration-specific analytical evaluator for energy-aware capacity reconfiguration in two-class CONWIP systems

Maria De Martino , Silvestro Vespoli , Andrea Grassi  and Guido Guizzi 

Department of Chemical, Materials and Production Engineering, Università degli Studi di Napoli Federico II, Napoli, Italy

ABSTRACT

This paper studies energy-aware capacity configuration in two-class CONWIP systems with changing product mix. The goal is to select the number of active parallel machines at each station to minimise power consumption while meeting a throughput target. Because each capacity vector alters class throughputs, visit intensities, and local arrival mixes, the analytical aggregation must be rebuilt for every candidate configuration. The proposed configuration-specific evaluator adapts EMDA-style aggregation by updating station arrival mixes through a multiclass fixed point, reconstructing an equivalent single-class network, and solving a single-class MDA recursion at the total CONWIP population. Embedded in a hybrid search procedure, it enables rapid capacity comparisons without simulation during optimisation. Tests on structured scenarios, random instances, and a manufacturing-inspired case yield throughput errors of about 1%, strong ranking preservation, and reliable feasibility classification in the tested two-class region. The reconstruction is also decision-relevant: at the operational throughput tolerance, it changes the selected capacity vector in about one in five realistic instances within the exhaustively enumerated $m \in \{3, 4\}$ divergence pool. Among 39 DES-adjudicated divergent instances, no frozen rule found a better configuration, while in 24 cases it selected one that DES classified as infeasible.

ARTICLE HISTORY

Received 31 March 2026
Accepted 17 August 2026

KEYWORDS

CONWIP; energy-aware manufacturing; capacity configuration; two-class queueing networks; extended marginal distribution analysis; analytical evaluator

SUBJECT

CLASSIFICATIONS

PRODUCTION PLANNING;
ENERGY EFFICIENT
MANUFACTURING;
QUEUEING NETWORKS

1. Introduction

Reducing energy use on the shop floor is no longer a purely reputational objective. For many factories it is an operating constraint that affects unit cost, exposure to energy-price volatility, and the sustainability indicators now requested by customers and supply chains. In this context, one of the most immediate levers is also one of the least systematically exploited: the amount of production capacity that is actually kept online. Lines are often sized for peak demand, yet many shifts run far below that peak. When all machines remain active regardless of the realised mix, plants absorb the baseline and auxiliary power draw of resources that contribute little to throughput.

This issue becomes more acute when the product mix changes under a controlled-WIP regime. In a CONWIP system, the total number of circulating jobs is fixed, but changes in the mix still redistribute congestion across stations because classes differ in routing and processing requirements. A station that is critical for one mix may become lightly loaded for another. The operational question is therefore specific and tactical: *given tomorrow's*

expected mix, how many identical parallel machines should remain active at each station in order to reduce power demand without compromising throughput?

That question is not well served by the existing modelling toolbox. Discrete-event simulation (DES) can represent the system faithfully, but repeated simulation over many candidate machine vectors is too slow for rapid tactical evaluation. Online control and data-driven methods can react to the system state, but they typically require historical training data, continuous sensing, and real-time actuation. Analytical queueing models are fast, but in the present setting speed alone is not enough. The machine vector is itself the decision variable, and when it changes the class throughputs, station visit intensities, and arrival-mix proportions change with it. What is needed is a configuration-specific analytical evaluator that remains coherent with the capacity vector under test and can support repeated tactical comparisons across candidate configurations.

This paper addresses that problem for a *two-class* CONWIP system modelled as a closed queueing network in which two product classes share the same line

CONTACT Maria De Martino  maria.demartino13@unina.it  Department of Chemical, Materials and Production Engineering, Università degli Studi di Napoli Federico II, P.le Tecchio, 80, Napoli, Italy

 Supplemental data for this article can be accessed online at <https://doi.org/10.1080/00207543.2026.2723544>.

© 2026 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

under fixed class populations. The two-class setting is both industrially relevant and analytically non-trivial. It captures common situations such as urgent versus standard orders, high-runner versus low-runner families, or two product platforms sharing machining resources (Herer and Masin 1997; Ryan, Baynat, and Choobineh 2000). It is also the smallest setting in which mix-dependent bottleneck shifts make the capacity-on/off decision structurally non-trivial.

The proposed approach is a configuration-specific analytical evaluator built on EMDA/MDA aggregation logic. For each candidate active-machine vector, the evaluator updates the station-level class arrival mixes from the current class-throughput estimates, reconstructs the corresponding equivalent single-class closed network, and solves it through a single-class MDA recursion at the total CONWIP population. The resulting fixed-point procedure provides class throughputs, cycle times, station utilisation, and congestion estimates that are consistent with the capacity vector under evaluation.

This configuration-specific reconstruction distinguishes the proposed approach from an aggregation constructed once under a reference configuration and then reused throughout the search. Alternative active-machine vectors may produce different class-throughput balances, station arrival intensities, local class mixes, and equivalent service demands. The equivalent network is therefore rebuilt for every candidate configuration before its performance is evaluated. The evaluator is then combined with a steady-state power model and embedded in a hybrid optimisation procedure that rapidly compares alternative capacity vectors without requiring DES during the search.

A frozen aggregation, by contrast, becomes internally inconsistent once the active-machine vector alters the class throughputs, because the station mixes it assumes no longer match the throughputs the network actually produces. The configuration-specific reconstruction restores this consistency for every candidate and is decision-relevant: as reported in Section 7.2, at the operational throughput tolerance it changes the deployed configuration in roughly one realistic random instance in five within the divergence pool, where each rule's minimum-power selection is identified by exact enumeration. The underlying inconsistency is structural rather than search-dependent and can arise whenever bottleneck identity depends on the local class mix. In the 39 DES-adjudicated divergent instances, no frozen surrogate selected a better configuration; in 24 cases, a frozen surrogate selected a configuration that DES classified as infeasible.

The contribution of the paper is threefold. First, it formulates multiclass aggregation as a configuration-specific operation for active-capacity comparison in two-class

CONWIP systems. Second, it links the resulting analytical evaluator to an energy-aware capacity-reconfiguration problem with a throughput requirement. Third, it evaluates the framework on structured scenarios, a broad pool of random instances, and a manufacturing-inspired illustrative case, considering throughput accuracy, ranking preservation, feasibility classification, robustness to service-time variability, and the DES-verified quality of the selected configurations, and isolates the decision-level role of the configuration-specific reconstruction, showing that frozen-mix surrogates change and, in a sizeable fraction of instances, degrade the deployment decision.

The journal paper extends the conference version in De Martino et al. (2025) by introducing the two-class configuration-specific evaluator, the affine steady-state power model, the hybrid optimisation layer, and an extensive DES-based decision validation.

The framework is intended for tactical, shift-level capacity planning in two-class closed CONWIP systems with equivalent parallel machines. Within this setting, it serves as a fast screening and ranking tool, while detailed simulation can be concentrated on selected and near-boundary configurations.

The remainder of the paper is organised as follows. Section 2 reviews the relevant literature and positions the contribution. Section 3 defines the system, decision problem, and modelling scope. Section 4 presents the configuration-specific analytical evaluator, while Section 5 introduces the steady-state power model and the energy-aware optimisation framework. Section 6 describes the experimental design and validation protocols. Section 7 reports the predictive, decision-oriented, robustness, and optimisation results. Section 8 discusses the reliability of the decision signals, practical deployment, limitations, and future research. Section 9 concludes the paper.

2. Literature review and research positioning

The present problem lies at the intersection of three research streams: CONWIP and multi-product production systems, analytical evaluation of closed multiclass queueing networks, and energy-aware machine activation in manufacturing.

2.1. CONWIP and multiclass closed queueing networks

CONWIP was introduced by Spearman, Woodruff, and Hopp (1990) as a pull-control alternative to Kanban in which the total number of jobs circulating in the line is bounded. Its closed-WIP logic stabilises flow and makes

the system naturally interpretable as a closed queueing network (Spearman, Hopp, and Woodruff 1992; Spearman, Woodruff, and Hopp 2022). This representation supports analytical models of throughput, cycle time, and congestion, as well as optimisation-oriented formulations of production decisions (Herer and Masin 1997; Hopp and Roof 1998).

The modelling challenge increases in multi-product systems because classes may differ in routing, processing requirements, and WIP allocation. Ryan, Baynat, and Choobineh (2000) showed that heterogeneous routings and processing times can shift the effective bottleneck even when total WIP remains fixed. The associated WIP-allocation problem was formulated as a multiclass closed queueing network by Ryan and Choobineh (2003) and extended to settings with lost sales by Ryan and Vorasayan (2005). Subsequent studies examined WIP setting and class-dependent WIP allocation in multi-product CONWIP environments (Framinan, González, and Ruiz-Usano 2003; Jaegler et al. 2018). These contributions establish the importance of mix-dependent congestion, but focus primarily on WIP allocation and production performance under a given capacity structure. The tactical activation of installed capacity under changing product mix remains less developed.

2.2. Multiclass aggregation and MDA/EMDA approximations

The classical analytical foundation is provided by product-form closed queueing networks (Baskett et al. 1975; Gordon and Newell 1967). For single-class networks, the convolution algorithm evaluates performance through the normalising constant (Buzen 1973). The cost of evaluating this normalising constant grows with both the circulating population and the network size, which motivates aggregation when the same closed network is evaluated repeatedly across candidate configurations. For multiclass networks, multichain Mean Value Analysis recursively computes class throughputs, cycle times, and mean queue lengths (Reiser and Lavenberg 1980). At the station level Little's law binds these quantities together, the mean queue length being the product of class throughput and residence time, which makes the relationship between congestion and utilisation explicit as the active capacity changes.

Direct solution of the underlying continuous-time Markov chain becomes impractical as the number of stations, circulating jobs, and customer classes increases. Population-recursive methods such as MVA and marginal-distribution analysis avoid explicit enumeration of the global state space and provide efficient performance estimates when product-form assumptions, or

suitable approximations of them, are acceptable. In manufacturing systems, exact product-form conditions are often weakened by FCFS multi-server stations, class-dependent service requirements, re-entrant routings, finite buffers, priority rules, and machine-eligibility constraints. Analytical evaluation is therefore frequently based on approximate MVA, aggregation, decomposition, simulation, or hybrid methods (Akyildiz and Bolch 1988; Amjath et al. 2024; Baynat and Dallery 1996; Baynat, Buzacott, and Dallery 2002; Buzacott and Shanthikumar 1993; Roy 2016).

In the two-class CONWIP setting considered here, the main approximation challenge arises from the combination of FCFS multi-server stations, class-dependent service requirements, and configuration-dependent active capacity. Akyildiz and Bolch (1988) proposed approximate MVA methods for multi-server stations, while Baynat and Dallery (1996) developed a practical product-form approximation based on station-wise aggregation of a multiclass network into an equivalent single-class representation. The same aggregation principle was subsequently extended to multi-product kanban-like systems, where the accuracy of the equivalent network depends on maintaining consistency between station mixes and throughput estimates (Baynat, Buzacott, and Dallery 2002).

That consistency condition is central to capacity reconfiguration. When the active-machine vector changes, class throughputs, aggregate station arrival rates, and local class proportions may change with it. An aggregation derived under one capacity configuration does not therefore remain automatically coherent under another. Extended Marginal Distribution Analysis provides a suitable computational backbone because it combines equivalent-network aggregation with efficient population-recursive evaluation of a closed network (Buzacott and Shanthikumar 1993).

The structural literature on server allocation in closed networks is also relevant. For the non-blocking closed-network setting considered in those studies, Shanthikumar and Yao (1988) and Dallery and Kathryn (1990) establish throughput monotonicity with respect to the number of available servers and show that efficient allocations tend to concentrate capacity at stations with high workload or congestion. These results provide the structural motivation for using utilisation and congestion information to guide the exploration of alternative active-capacity vectors.

Their applicability must, however, be restricted to the non-blocking setting. In finite-buffer systems subject to blocking after service (BAS) or blocking before service

(BBS), a local capacity change can interact with downstream blocking, upstream starvation, and the propagation of congestion across stations. Consequently, the system-level throughput response cannot be inferred directly from the monotonicity results established for non-blocking closed networks. The present study therefore uses those results only as motivation for the search heuristic and explicitly excludes finite-buffer blocking from the analytical model. Extending the framework to BAS or BBS systems would require an explicit representation of buffer states and a separate performance approximation.

More generally, Bitran and Dasu (1992) discussed the role of queueing approximations in production systems, while Inman (1999) examined the empirical adequacy of exponential assumptions in automotive manufacturing data.

2.3. Energy-aware capacity activation

Energy-aware production control has shown that switching off under-loaded resources can reduce energy consumption without necessarily causing a proportional loss of output. Mouzon, Yildirim, and Twomey (2007) established the practical relevance of switch-off policies for non-bottleneck resources. Related results were later obtained for closed-loop serial lines and pallet-constrained flow shops, where state-dependent control reduced energy use while maintaining production performance (Jia et al. 2016; Mashaei and Lennartson 2013). The review by Gahm et al. (2016), together with subsequent work on sustainable and energy-efficient reconfigurable manufacturing systems (Battaia et al. 2020), identifies capacity reconfiguration under realistic manufacturing constraints as an area in which analytical decision support remains limited.

The work closest in decision purpose is Frigerio, Tan, and Matta (2024), who optimise the simultaneous energy control of multiple machines using simulation-based methods and parameters measured on CNC machining centres, extending earlier research on buffer-threshold policies for individual stations (Frigerio and Matta 2015). These approaches address online or state-dependent control, whereas the present problem concerns the selection of a shift-level active-capacity configuration before operations begin.

Machine-energy studies provide the basis for representing the power implications of such decisions. Empirical evidence shows that machine tools exhibit a substantial baseline power draw together with an additional processing-related component (Avram and Xirouchakis 2011; Kara and Li 2011; Lv et al. 2016). Station-level identification of these components can be

supported by monitoring methods (Behrendt, Zein, and Min 2012; Vijayaraghavan and Dornfeld 2010) and by the ISO 14955 standard series (ISO 14955-1:2017 2017; ISO 14955-2:2018 2018). Product- and strategy-dependent variation in processing power has also been documented (Aramcharoen and Mativenga 2014).

Data-driven surrogates provide another route to rapid decision support. Reinforcement-learning methods have been applied to energy-efficient control of individual machines and multi-stage production systems (Loffredo et al. 2024, 2023; Panzer and Bender 2022), while hybrid learning methods can improve throughput prediction by incorporating queueing knowledge into the training process (Dieleman, Berkhout, and Heidergott 2023). These methods are promising for data-rich and dynamically controlled environments, but serve a different purpose from a transparent analytical evaluator intended for repeated tactical comparison of capacity configurations.

2.4. Research gap and positioning of this study

Taken together, the literature provides models of multi-class CONWIP systems, analytical and approximate solution methods for closed queueing networks, and energy-aware machine-control approaches. A gap remains at their intersection: the rapid analytical comparison of active-machine configurations in a two-class CONWIP system when the product mix and congestion pattern depend on the capacity vector under evaluation. In this setting, capacity affects both throughput feasibility and energy use, while the analytical aggregation must remain consistent with the candidate configuration.

The framework uses the established single-class MDA recursion as its computational core and contributes a configuration-specific aggregation and closure mechanism. Its distinctive element is the role assigned to the aggregation step. Because the active-machine vector is the decision variable, each candidate configuration may induce a different class-throughput balance, aggregate station arrival intensity, local arrival mix, and effective service demand. The equivalent single-class network is therefore reconstructed for each candidate vector rather than retained from a fixed reference configuration.

The configuration-specific evaluator couples three operations. First, station-level class proportions are updated from the current class-throughput estimates. Second, these proportions are used to rebuild the equivalent single-class visit ratios and service parameters. Third, the resulting network is solved through an MDA recursion within a fixed-point closure for the class throughputs. The evaluator is then used as the analytical engine of an energy-aware capacity-search procedure.

Table 1. Positioning of the proposed framework relative to the literature.

Literature stream	Established basis	Role in this study
CONWIP and multi-product closed networks (Ryan, Baynat, and Choobineh 2000; Ryan and Choobineh 2003; Spearman, Hopp, and Woodruff 1992; Spearman, Woodruff, and Hopp 1990)	Establish CONWIP as a pull-control mechanism and model multi-product production systems using closed networks and class-dependent WIP levels.	Defines the production-control setting in which active machine capacity is selected subject to a throughput requirement and a changing product mix.
Product-form closed queueing networks and MVA/MDA methods (Buzacott and Shanthikumar 1993; Buzen 1973; Gordon and Newell 1967; Reiser and Lavenberg 1980)	Provide exact and population-recursive methods for evaluating throughput, cycle time, and congestion in closed queueing networks.	Provides the analytical foundation and the single-class MDA routine used to evaluate each equivalent network.
Multiclass aggregation and approximate evaluation (Akyildiz and Bolch 1988; Baynat and Dallery 1996; Baynat, Buzacott, and Dallery 2002)	Develop approximate multi-server evaluation and multiclass-to-single-class aggregation based on station-level class proportions and equivalent service requirements.	Provides the aggregation logic extended through a configuration-specific station-mix fixed point, which reconstructs the equivalent network for each active-machine vector.
Server allocation in closed networks (Dallery and Kathryn 1990; Shanthikumar and Yao 1988)	Relates server allocation to throughput, workload, and congestion, motivating capacity additions at highly loaded stations.	Provides the structural motivation for using utilisation and congestion information to guide exploration of the active-capacity space.
Energy-aware machine switching and production control (Frigerio and Matta 2015; Frigerio, Tan, and Matta 2024; Jia et al. 2016; Mashaei and Lennartson 2013; Mouzon, Yildirim, and Twomey 2007)	Shows that machine deactivation and state-dependent control can reduce energy use while maintaining production performance.	Motivates a tactical, shift-level machine-activation decision supported by rapid analytical evaluation.
Machine-energy characterisation (Avram and Xirouchakis 2011; ISO 14955-1:2017 2017; ISO 14955-2:2018 2018; Kara and Li 2011; Lv et al. 2016)	Characterises idle, standby, and processing-related machine power and provides measurement principles for industrial calibration.	Provides the basis for the steady-state, station-level power model coupled with the queueing-based throughput evaluator.

Table 1 summarises this positioning relative to the most closely related streams of work.

The methodological contribution therefore consists in adapting established aggregation and MDA/EMDA logic to a capacity-reconfiguration problem in which the network representation changes with the decision vector. The station-mix fixed point maintains consistency between the estimated class throughputs and the equivalent network, while the optimisation layer converts this analytical evaluation into an energy-aware machine-activation decision.

3. Problem formulation and modelling scope

3.1. System definition and decision variables

We consider a pull-controlled production system operating under a CONWIP regime (Spearman, Woodruff, and Hopp 1990). A fixed number of jobs circulates in the system, and each departure triggers a new release so that the total work-in-process remains constant. The system is represented as a closed queueing network (Baynat and Dallery 1996; Reiser and Lavenberg 1980).

The network comprises m stations and two product classes $r \in \{0, 1\}$. The classes may differ in their routing and processing requirements. The visit ratio V_{ir} denotes the expected number of visits made by a class- r job to station i , while S_{ir} denotes its mean processing time per visit. The class populations are fixed at $\mathbf{K} = (K_0, K_1)$, and

the total CONWIP population is

$$N = K_0 + K_1.$$

Capacity at station i is provided by identical parallel machines operating under FCFS. The installed capacity is denoted by $c_{i,\max}$, whereas x_i denotes the number of machines kept active in the configuration under evaluation. The active-machine vector is therefore

$$\mathbf{x} = (x_1, \dots, x_m),$$

with

$$1 \leq x_i \leq c_{i,\max}$$

for stations visited by at least one class. Stations not included in either class routing are assigned $x_i = 0$.

Figure 1 summarises the system. Two classes circulate with fixed populations, follow class-dependent routings, and share station capacity. The decision consists in selecting the number of active machines at each station.

The model requires four groups of inputs: class populations, routing and processing-time data, installed station capacities, and station-level power parameters. The expected product mix is represented through the class-population vector $\mathbf{K}$, while the configuration decision is represented through $\mathbf{x}$.

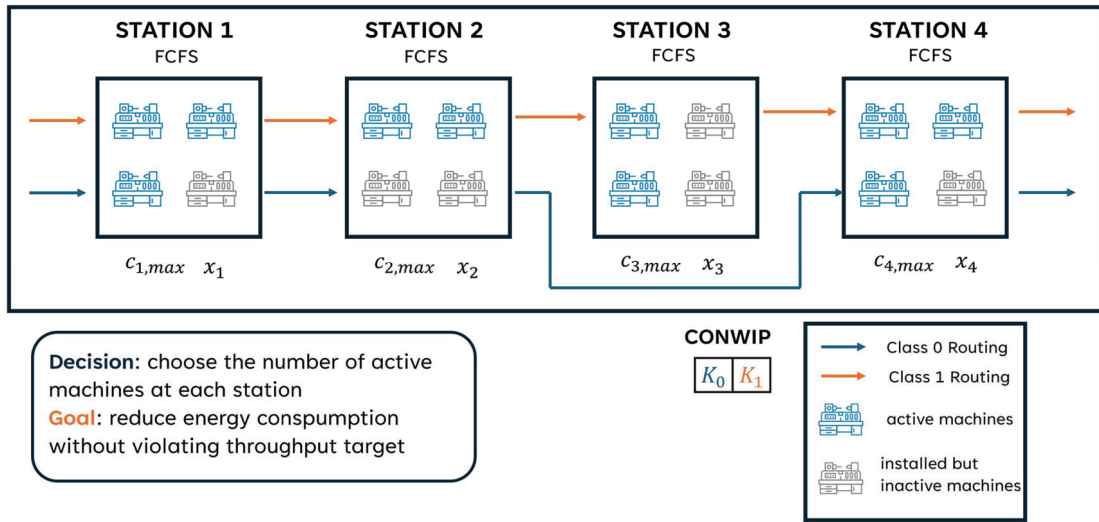


Figure 1. Schematic representation of the two-class CONWIP system. Two product classes circulate in a closed network with class-dependent visit ratios V_{ir} and mean processing times S_{ir} . Each station contains $c_{i,max}$ identical parallel machines, of which x_i are kept active in the candidate configuration.

3.2. Throughput requirement and capacity-reconfiguration problem

Let $TH(\mathbf{x})$ denote the steady-state system throughput associated with configuration $\mathbf{x}$, and let

$$\mathbf{c}_{max} = (c_{1,max}, \dots, c_{m,max})$$

denote the all-on configuration. The corresponding all-on throughput is

$$TH_{all} = TH(\mathbf{c}_{max}).$$

The production requirement is expressed relative to this reference. Given a maximum admissible throughput-loss fraction $\delta \in [0, 1)$, a candidate configuration must satisfy

$$TH(\mathbf{x}) \geq (1 - \delta)TH_{all}.$$

An optional limit $\bar{u}_i$ on the raw analytical utilisation may also be imposed at selected critical stations.

The capacity-reconfiguration problem is therefore to identify a feasible active-machine vector that reduces steady-state power while preserving the required throughput. Its analytical solution requires evaluating $TH(\mathbf{x})$, station utilisation, and congestion for many alternative capacity vectors. Since these quantities depend on the active capacity, the equivalent queueing representation used for evaluation is reconstructed for each candidate configuration, as described in Section 4.

3.3. Modelling assumptions and applicability

The model is designed for tactical capacity planning in closed CONWIP systems where the active-machine configuration is selected for a shift or comparable planning

horizon. Table 2 summarises the assumptions that define this operating setting and their implications for application.

In the analytical approximation, service times are assumed to be exponentially distributed with class- and station-specific means S_{ir} , consistent with the closed-network and EMDA/MDA approximation adopted in the analytical evaluator (Baynat and Dallery 1996; Baynat, Buzacott, and Dallery 2002; Buzacott and Shanthikumar 1993; Gordon and Newell 1967; Reiser and Lavenberg 1980). The service-time means determine the analytical evaluation, while the sensitivity of the resulting decision signals to alternative service-time distributions is examined through DES in the experimental study.

When the expected product mix is uncertain, the capacity vector can be evaluated across multiple class-population scenarios rather than against a single forecast. This scenario-based use is particularly relevant when the selected configuration lies close to the throughput requirement or when plausible mix changes move the effective bottleneck between stations.

The resulting scope is a two-class, non-blocking CONWIP system with interchangeable parallel machines, FCFS sharing, and shift-level capacity decisions. Within this setting, the evaluator supports rapid screening and ranking of candidate configurations. Applications involving many heterogeneous classes, machine-product eligibility, priority rules, sequence-dependent setups, finite-buffer blocking, or frequent within-shift switching require corresponding extensions of the system representation.

Table 2. Modelling assumptions and intended scope of application.

Assumption	Role in the model	Applicability implication
Two product classes	Represents the smallest multiclass setting in which changes in product mix can alter congestion and bottleneck patterns.	Appropriate for systems dominated by two product families, platforms, or operational categories. Extensions to additional classes require separate validation of aggregation accuracy and fixed-point convergence.
Fixed class populations	Preserves the closed-network representation of CONWIP through $\mathbf{K} = (K_0, K_1)$.	Suitable when class-specific WIP is controlled or planned at the beginning of the decision horizon. Forecast release proportions may be translated into expected class-population scenarios.
Identical parallel machines	Represents station capacity through the scalar decision variable x_i .	Appropriate when machines within a station are interchangeable. Product-specific eligibility and heterogeneous processing capabilities require a machine-level representation.
FCFS service	Provides a common queueing discipline for shared stations with class-dependent service requirements.	Suitable when priority rules, batching, expediting, and sequence-dependent setups do not dominate station behaviour.
Load-independent mean service times	Represents processing requirements through fixed mean service times S_{ir} .	Appropriate when capacity changes affect the number of available servers rather than their processing speed. Load-dependent service rates require an extended fixed-point formulation.
Non-blocking closed network	Models station interactions without finite-buffer blocking propagation.	Systems subject to blocking after service or blocking before service require explicit buffer-state information and a different analytical approximation.
Steady-state operation	Supports comparison of capacity configurations over a shift-level planning horizon.	Best suited to pre-shift or daily planning rather than real-time switching with transient start-up effects, warm-up periods, and minimum on/off times.

The main symbols are introduced where they first appear; a consolidated notation table is reported in the appendix.

4. Configuration-specific analytical evaluator

For a given active-machine configuration $\mathbf{x}$, the evaluator maps the original two-class CONWIP system into a candidate-dependent equivalent single-class closed queueing network. The fixed-point procedure returns the class throughputs, class cycle times, raw and clipped station utilisations, queueing estimates, and congestion indicators required by the optimisation layer. Its inputs are the class-population vector $\mathbf{K} = (K_0, K_1)$, the visit matrix $V \in \mathbb{R}^{m \times 2}$, the mean service-time matrix $S \in \mathbb{R}_{>0}^{m \times 2}$, and the active-machine vector $\mathbf{x}$.

4.1. Single-class MDA basis

The computational core of the proposed evaluator is the single-class Marginal Distribution Analysis (MDA) recursion for a closed queueing network (Buzacott and Shanthikumar 1993). This routine is inherited from the literature and is not modified in the present study. Its role is to evaluate an equivalent single-class network once the total circulating population, station visit ratios, service rates, and active server counts have been specified.

Consider a single-class closed network with total population N , station visit ratios v_i , effective service rates μ_i , and x_i identical active servers at station i . MDA evaluates the network recursively for population levels $l = 1, \dots, N$. At each level, the marginal probability $p_i^{(l)}(q)$

represents the probability that q jobs are present at station i when the total network population is l . The number of simultaneously busy servers when q jobs are present is represented by

$$r_i(q) = \min\{x_i, q\}. \quad (1)$$

Using the marginal distribution obtained at population level $(l-1)$, the mean station sojourn time at level l is calculated as

$$ET_i(l) = \sum_{q=1}^l \frac{q}{\mu_i r_i(q)} p_i^{(l-1)}(q-1). \quad (2)$$

The corresponding closed-network throughput is then obtained from the population-normalisation relation

$$TH(l) = \frac{l}{\sum_{i=1}^m v_i ET_i(l)}, \quad (3)$$

and the mean number of jobs at station i follows from Little's law:

$$EN_i(l) = v_i TH(l) ET_i(l). \quad (4)$$

The marginal station-population probabilities are subsequently updated and the recursion continues until $l = N$. At the final population level, the routine returns the network throughput $TH_{\text{mda}} = TH(N)$, the mean station sojourn times $ET_i = ET_i(N)$, and the mean station populations $EN_i = EN_i(N)$. The complete marginal-probability recursion and pseudocode are reported in Appendix 1.

In the proposed framework, this single-class MDA routine is used without methodological modification.

The contribution lies in how its input network is constructed. For every candidate active-machine vector $\mathbf{x}$, the multiclass routing and service information is aggregated into configuration-dependent visit ratios and effective service rates. These quantities are updated within a class-throughput fixed point before the resulting equivalent network is submitted to the inherited MDA routine.

4.2. Candidate-dependent equivalent network

The distinction between installed and active capacity is central to the evaluation. The parameter $c_{i,\max}$ denotes the number of identical parallel machines installed at station i , whereas x_i denotes the number kept active in the candidate configuration. The installed capacity defines the feasible range of the decision variable; the analytical evaluation itself represents station i as a multi-server node with x_i active servers. The all-on configuration is obtained by setting $x_i = c_{i,\max}$ at every visited station.

Figure 2 illustrates the transformation. The physical stations remain the nodes of the closed network, while class-dependent routing and processing information is aggregated into candidate-dependent visit ratios

and effective service rates. The total population of the equivalent network is $N = K_0 + K_1$.

Let $\mathbf{h} = (h_0, h_1)$ denote the current estimate of the class throughputs. The multiclass throughput implied by the fixed point is

$$TH_{\text{emda}} = \sum_{r=0}^1 h_r. \quad (5)$$

The aggregate arrival rate at station i is

$$\lambda_i = \sum_{r=0}^1 h_r V_{ir}, \quad (6)$$

and the corresponding aggregate visit ratio is

$$v_i^{\text{agg}} = \frac{\lambda_i}{TH_{\text{emda}}}. \quad (7)$$

The station-level arrival proportion of class r is

$$p_{ir} = \begin{cases} \frac{h_r V_{ir}}{\lambda_i}, & \lambda_i > 0, \\ \frac{1}{2}, & \lambda_i = 0. \end{cases} \quad (8)$$

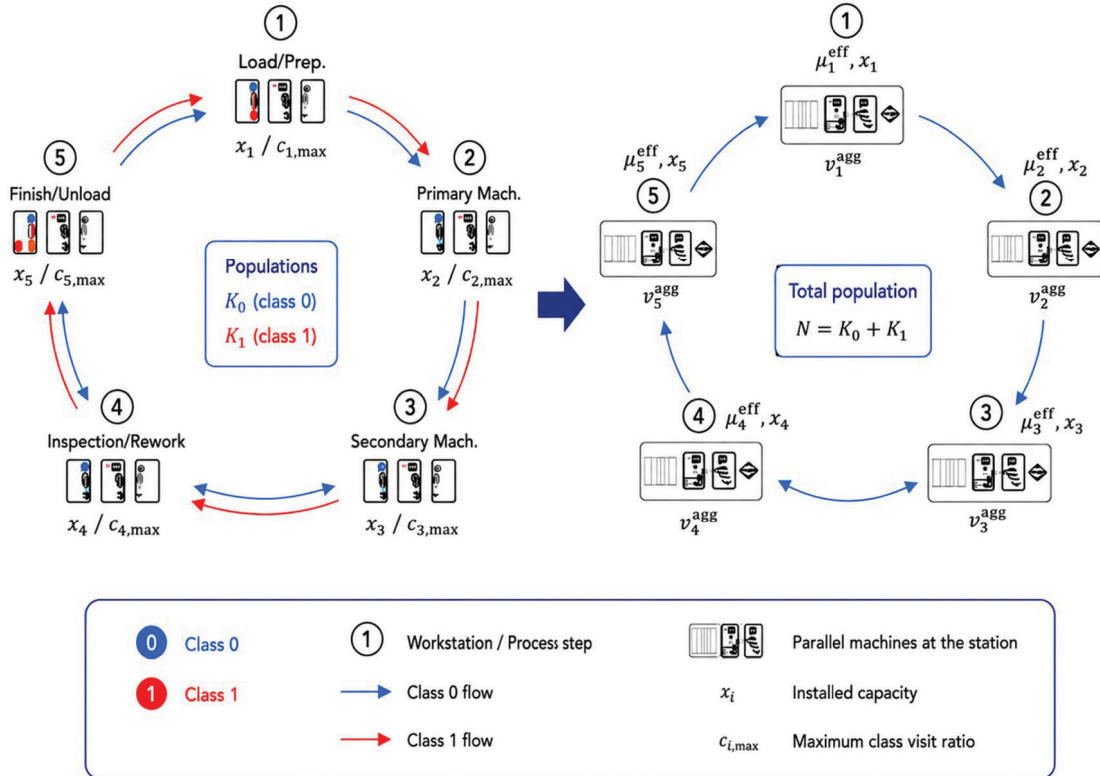


Figure 2. Two-class CONWIP system and corresponding equivalent single-class closed queueing network. The left-hand diagram represents two product classes, with fixed populations K_0 and K_1 , circulating through five processing stations. Blue and red arrows identify the routes of classes 0 and 1, respectively. At station i , x_i denotes the number of active parallel machines, while $c_{i,\max}$ denotes the maximum installed capacity. The right-hand diagram shows the equivalent single-class closed queueing network used for performance evaluation, in which the two class populations are aggregated into a total population $N = K_0 + K_1$. The equivalent network preserves the same sequence of stations while representing each station through aggregated service parameters.

The equivalent mean service time at station i is then

$$\bar{S}_i = \sum_{r=0}^1 p_{ir} S_{ir}, \quad (9)$$

with effective service rate

$$\mu_i^{\text{eff}} = \frac{1}{\bar{S}_i}. \quad (10)$$

Collapsing the two class-dependent FCFS service processes at station i into a single equivalent exponential server of rate μ_i^{eff} is a product-form (BCMP-type) aggregation. It is exact when the per-class service requirements coincide in mean, so that the station reduces to a single-rate FCFS node and the class mix is immaterial; under class-dependent service it is an approximation, because exact BCMP product form at an FCFS node requires a class-independent service rate rather than merely exponential service. Its accuracy on the secondary station-level quantities tapers as the per-class service-time disparity and the station utilisation grow. The operating regime validated against DES in this study lies within this scope.

The tuple $(N, \mathbf{v}^{\text{agg}}, \mu^{\text{eff}}, \mathbf{x})$ defines the equivalent single-class network associated with the current throughput estimates and the candidate configuration. This network is evaluated through the inherited population-recursive single-class MDA routine reported in Appendix 1. The routine returns the equivalent-network throughput TH_{mda} , the mean station sojourn times ET_i , and the mean station populations EN_i .

TH_{mda} and TH_{emda} have distinct roles. The former is the throughput returned by the internal single-class MDA solve; the latter is the total throughput implied by the class-throughput fixed point and is the quantity used for analytical feasibility assessment and optimisation. DES throughput, denoted by TH_{DES} , is introduced only in the validation stage.

4.3. Fixed-point station-mix closure

The configuration-specific reconstruction is embedded in a fixed-point loop. At each iteration, the current class-throughput estimates determine the station arrival rates and local class proportions. These quantities define a new equivalent network, whose queueing outputs are then used to update the class cycle times and throughputs. Iteration continues until the class-throughput vector and the equivalent-network parameters are mutually consistent.

The FCFS waiting component shared by the two classes is approximated as

$$W_i^{(q)} = \max\{ET_i - \bar{S}_i, 0\}. \quad (11)$$

Applying a single shared term $W_i^{(q)}$ to both classes is a shared-FCFS approximation, under which the two classes experience the same mean residual delay at station i . It is most faithful when the local class load mixes are comparable and becomes looser under strongly asymmetric per-class demand; retaining the class-specific service times S_{ir} in the cycle-time calculation preserves the dominant class differentiation, and the validated regime keeps the shared-delay component within its accurate range.

Class-dependent processing requirements are retained in the cycle-time calculation:

$$CT_r = \sum_{i=1}^m V_{ir} (S_{ir} + W_i^{(q)}), \quad r \in \{0, 1\}. \quad (12)$$

Little's law provides the undamped class-throughput update

$$\tilde{h}_r = \frac{K_r}{CT_r}. \quad (13)$$

To improve numerical stability, the fixed point uses the damped update

$$h_r^{(t+1)} = (1 - \beta)h_r^{(t)} + \beta\tilde{h}_r^{(t)}, \quad (14)$$

where $\beta \in (0, 1]$ is the damping factor. Convergence is reached when the maximum relative change in the class-throughput vector falls below the prescribed tolerance τ .

For boundary-mix scenarios, if $K_r = 0$ for one class, that class is assigned zero throughput and is excluded from the fixed-point iteration. The reconstruction is then performed over the positive-population class subset. Similarly, stations having zero aggregate visit ratio under that subset are excluded from the equivalent single-class MDA recursion and are assigned zero sojourn time, population, utilisation, and congestion outputs.

The fixed-point map is not claimed to be a contraction, and neither uniqueness nor global convergence is established analytically. The damping factor is used to improve numerical stability, while convergence is assessed through the stopping criterion and the post-convergence consistency check. Across the evaluated Campaign-B configurations, the iteration converged in a median of 22 iterations and at most 109 iterations; approximately 0.03% of evaluations failed the reliability screen and were treated as evaluator-unusable. These results provide empirical evidence of stable convergence within the tested operating region.

Algorithms 1 and 2 summarise the complete evaluator: the damped fixed-point iteration that produces the converged class-throughput vector, followed by the post-convergence consistency pass that returns the performance outputs. The initial throughput estimates

Algorithm 1: Configuration-specific two-class EMDA evaluator: damped fixed-point iteration

Input: $N, \mathbf{x}, \mathbf{K} = (K_0, K_1), \mathbf{V}, \mathbf{S}$, tolerance τ , maximum number of iterations T , damping factor β

Output: converged class throughputs $\mathbf{h}^{(t+1)} = (h_0^{(t+1)}, h_1^{(t+1)})$, or EVALUATOR-UNUSABLE

```

1 converged  $\leftarrow$  false
2 for  $r \in \{0, 1\}$  do
3    $h_r^{(0)} \leftarrow \frac{K_r}{\sum_i V_{ir} S_{ir}}$ 
4 end
5 for  $t \leftarrow 0$  to  $T - 1$  do
6    $TH_{\text{emda}}^{(t)} \leftarrow \sum_{r=0}^1 h_r^{(t)}$ 
7   foreach  $i$  do
8      $\lambda_i^{(t)} \leftarrow \sum_{r=0}^1 h_r^{(t)} V_{ir}$ 
9      $v_i^{\text{agg},(t)} \leftarrow \frac{\lambda_i^{(t)}}{TH_{\text{emda}}^{(t)}}$ 
10    if  $\lambda_i^{(t)} > 0$  then
11      for  $r \in \{0, 1\}$  do
12         $p_{ir}^{(t)} \leftarrow \frac{h_r^{(t)} V_{ir}}{\lambda_i^{(t)}}$ 
13      end
14    else
15      for  $r \in \{0, 1\}$  do
16         $p_{ir}^{(t)} \leftarrow \frac{1}{2}$ 
17      end
18    end
19     $\bar{S}_i^{(t)} \leftarrow \sum_{r=0}^1 p_{ir}^{(t)} S_{ir}$ 
20     $\mu_i^{\text{eff},(t)} \leftarrow \frac{1}{\bar{S}_i^{(t)}}$ 
21  end
22   $(TH_{\text{mda}}^{(t)}, \mathbf{ET}^{(t)}, \mathbf{EN}^{(t)}) \leftarrow \text{MDA\_Single}(N, \mathbf{v}^{\text{agg},(t)}, \boldsymbol{\mu}^{\text{eff},(t)}, \mathbf{x})$ 
23  foreach  $i$  do
24     $W_i^{(q),(t)} \leftarrow \max\{ET_i^{(t)} - \bar{S}_i^{(t)}, 0\}$ 
25  end
26  for  $r \in \{0, 1\}$  do
27     $CT_r^{(t)} \leftarrow \sum_i V_{ir}(S_{ir} + W_i^{(q),(t)})$ 
28     $\tilde{h}_r^{(t)} \leftarrow \frac{K_r}{CT_r^{(t)}}$ 
29     $h_r^{(t+1)} \leftarrow (1 - \beta)h_r^{(t)} + \beta\tilde{h}_r^{(t)}$ 
30  end
31   $\text{res}^{(t)} \leftarrow \max_{r \in \{0,1\}} \frac{|h_r^{(t+1)} - h_r^{(t)}|}{\max\{1, |h_r^{(t)}|\}}$ 
32  if  $\text{res}^{(t)} < \tau$  then
33    converged  $\leftarrow$  true
34    break
35  end
36 end
37 if  $\neg$ converged then
38   return EVALUATOR-UNUSABLE
39 end
40 return converged class throughputs  $\mathbf{h}^{(t+1)}$ 

```

are obtained from the class-specific no-queue service demands:

$$h_r^{(0)} = \frac{K_r}{\sum_i V_{ir} S_{ir}}.$$

Algorithm 2: Configuration-specific two-class EMDA evaluator: post-convergence consistency pass and performance outputs

Input: $N, \mathbf{x}, \mathbf{K} = (K_0, K_1), \mathbf{V}, \mathbf{S}$, and the converged class throughputs $\mathbf{h}^{(t+1)}$ from Algorithm 1

Output: $(\mathbf{h}^*, \mathbf{CT}^*, \mathbf{u}^{\text{raw},*}, \mathbf{u}^{\text{clip},*}, \mathbf{W}^{(q),*}, \mathbf{EN}^*, TH_{\text{mda}}^*)$

```

1  $\mathbf{h}^* \leftarrow \mathbf{h}^{(t+1)}$ 
2  $TH_{\text{emda}}^* \leftarrow \sum_{r=0}^1 h_r^*$ 
3 foreach  $i$  do
4    $\lambda_i^* \leftarrow \sum_{r=0}^1 h_r^* V_{ir}$ 
5    $v_i^{\text{agg},*} \leftarrow \frac{\lambda_i^*}{TH_{\text{emda}}^*}$ 
6   if  $\lambda_i^* > 0$  then
7     for  $r \in \{0, 1\}$  do
8        $p_{ir}^* \leftarrow \frac{h_r^* V_{ir}}{\lambda_i^*}$ 
9     end
10  else
11    for  $r \in \{0, 1\}$  do
12       $p_{ir}^* \leftarrow \frac{1}{2}$ 
13    end
14  end
15   $\bar{S}_i^* \leftarrow \sum_{r=0}^1 p_{ir}^* S_{ir}$ 
16   $\mu_i^{\text{eff},*} \leftarrow \frac{1}{\bar{S}_i^*}$ 
17 end
18  $(TH_{\text{mda}}^*, \mathbf{ET}^*, \mathbf{EN}^*) \leftarrow \text{MDA\_Single}(N, \mathbf{v}^{\text{agg},*}, \boldsymbol{\mu}^{\text{eff},*}, \mathbf{x})$ 
19 foreach  $i$  do
20    $W_i^{(q),*} \leftarrow \max\{ET_i^* - \bar{S}_i^*, 0\}$ 
21    $u_i^{\text{raw},*} \leftarrow \frac{\lambda_i^* \bar{S}_i^*}{x_i}$ 
22    $u_i^{\text{clip},*} \leftarrow \min\{\max\{u_i^{\text{raw},*}, 0\}, 1\}$ 
23 end
24 for  $r \in \{0, 1\}$  do
25    $CT_r^* \leftarrow \sum_i V_{ir}(S_{ir} + W_i^{(q),*})$ 
26 end
27 return  $(\mathbf{h}^*, \mathbf{CT}^*, \mathbf{u}^{\text{raw},*}, \mathbf{u}^{\text{clip},*}, \mathbf{W}^{(q),*}, \mathbf{EN}^*, TH_{\text{mda}}^*)$ 

```

The superscript $*$ identifies the outputs of the final post-convergence consistency pass. In the remainder of the paper, this superscript is omitted when no ambiguity arises.

4.4. Performance outputs and reliability screen

After the fixed-point convergence criterion is satisfied, the evaluator performs one final consistency pass using the converged class-throughput vector. Specifically, the station arrival rates, aggregate visit ratios, local class proportions, equivalent service times, and effective service rates are reconstructed from the final throughput vector, and the equivalent single-class network is solved once more through MDA. The returned station and class metrics therefore correspond to the same final fixed-point state.

If the convergence criterion is not satisfied within the maximum number of iterations, the configuration is classified as evaluator-unusable and is not admitted to feasibility assessment or optimisation.

$$u_i = \min \{ \max \{ u_i^{\text{raw}}, 0 \}, 1 \} \quad (15)$$

is used only for reporting and for the bounded steady-state power calculation.

Utilisation constraints are therefore imposed on u_i^{raw} , rather than on the clipped quantity, so that analytical overload or inconsistency is not hidden by truncation.

Utilisation and mean station population provide complementary information. The raw utilisation u_i^{raw} measures the average load intensity implied for the active servers, whereas EN_i measures the mean number of jobs accumulated at the station, including both jobs in service and jobs waiting. The two quantities are retained separately because stations with similar utilisation may exhibit different levels of local congestion. In the optimisation procedure, u_i^{raw} and EN_i are used as search-guidance indicators; throughput feasibility is always assessed through a complete re-evaluation of the candidate configuration.

$$g(\mathbf{x}) = \frac{|TH_{\text{mda}}(\mathbf{x}) - TH_{\text{emda}}(\mathbf{x})|}{TH_{\text{emda}}(\mathbf{x})} \quad (16)$$

A configuration is classified as evaluator-usable when the fixed point converges and

$$g(\mathbf{x}) \leq 0.05.$$

The 5% threshold acts as an operational screen for internally inconsistent equivalent-network reconstructions. It is not a measure of error relative to DES and is not used as a substitute for external validation.

Across the 6739 active-machine configurations evaluated across the Campaign-B solver runs (beam, greedy, and hill climbing), the damped iteration ($\beta = 0.5$, $\tau = 10^{-7}$) met the convergence criterion in a median of 22 iterations and in at most 109, and the converged states exhibited an internal consistency gap $g(\mathbf{x})$ with median $\approx 2 \times 10^{-6}$ and maximum 1.6×10^{-5} , far below the screening threshold. Only 2 of these evaluations (about 0.03%) were classified as evaluator-unusable, by failing to converge within the iteration budget or by exceeding $g(\mathbf{x}) > 0.05$, so the reliability screen removes a negligible fraction of internally inconsistent reconstructions and leaves the optimised configurations unaffected.

The evaluator therefore returns three groups of outputs:

- (i) class-level measures: h_r and CT_r ;
- (ii) station-level measures: u_i^{raw} , u_i^{clip} , $W_i^{(q)}$, and EN_i ;

- (iii) network-level diagnostics: TH_{emda} , TH_{mda} , and $g(\mathbf{x})$.

These outputs provide the throughput-feasibility, power-estimation, and search-guidance signals used by the energy-aware optimisation framework.

4.5. Computational complexity

Each fixed-point iteration consists of one single-class MDA evaluation, together with the station-level aggregation and class-throughput closure. The aggregation and closure operations require $O(mR)$ computations, with $R = 2$ in the present study.

In the implemented population-recursive MDA routine, the station residence-time and marginal-probability updates are evaluated for population levels $l = 1, \dots, N$, with an inner summation over up to l station-population states. The resulting computational cost is therefore

$$O(mN^2)$$

per single-class MDA solve, rather than $O(mN)$.

If I denotes the number of fixed-point iterations, one configuration-specific EMDA evaluation requires

$$O(ImN^2 + ImR),$$

which reduces to

$$O(ImN^2)$$

for the two-class setting. The final post-convergence consistency pass adds one further MDA solve without changing the asymptotic order. Memory usage is $O(mN)$ for the station marginal distributions.

Although quadratic in the total population, the routine remains sufficiently fast for the moderate CONWIP populations considered in the experiments and supports repeated evaluation of candidate capacity vectors.

5. Energy-aware optimisation framework

The analytical evaluator is embedded in a decision framework that combines throughput-feasibility assessment, steady-state power estimation, and search over the active-capacity space. Figure 3 summarises the resulting workflow. For each candidate machine vector, the configuration-specific evaluator returns throughput, utilisation, and congestion measures; the power model converts these outputs into an average energy-performance estimate; and the search layer compares alternative configurations under the production requirement.

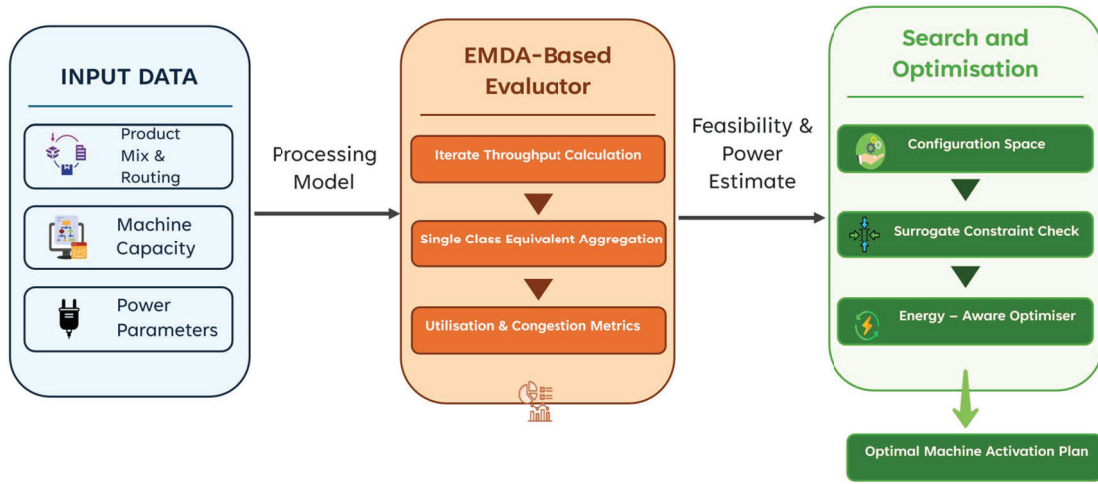


Figure 3. End-to-end workflow of the proposed decision procedure. Product-mix, routing, capacity, and power data are processed by the configuration-specific evaluator, which reconstructs and solves the equivalent network for each candidate active-machine vector. The resulting throughput, utilisation, and congestion measures are used by the power model and optimisation layer to identify an energy-aware capacity configuration.

5.1. Steady-state power model

Machine tools and related production resources typically exhibit a non-negligible baseline power demand together with an additional component associated with active processing (Kara and Li 2011; Lv et al. 2016). This behaviour is represented through an affine steady-state power model.

For one active machine at station i , the estimated average power is

$$P_i^{\text{mach}}(\mathbf{x}) = P_{\text{idle},i} + (P_{\text{proc},i} - P_{\text{idle},i}) u_i^{\text{clip}}(\mathbf{x}), \quad (17)$$

where $P_{\text{idle},i}$ is the power of one active but non-processing machine, $P_{\text{proc},i}$ is the power of one machine operating at full processing load, and $u_i^{\text{clip}}(\mathbf{x}) \in [0, 1]$ is the bounded average utilisation of the active parallel machines at station i .

Since x_i identical machines are kept active at station i , the corresponding station-level power is

$$P_i^{\text{station}}(\mathbf{x}) = x_i \left[P_{\text{idle},i} + (P_{\text{proc},i} - P_{\text{idle},i}) u_i^{\text{clip}}(\mathbf{x}) \right]. \quad (18)$$

The total steady-state power associated with configuration $\mathbf{x}$ is therefore

$$P_{\text{tot}}(\mathbf{x}) = \sum_{i=1}^m x_i \left[P_{\text{idle},i} + (P_{\text{proc},i} - P_{\text{idle},i}) u_i^{\text{clip}}(\mathbf{x}) \right] + P_{\text{aux}}, \quad (19)$$

where P_{aux} represents fixed auxiliary power that is assumed not to depend on the active-machine configuration.

Equation (19) separates two effects of the capacity decision. The first is the baseline contribution $x_i P_{\text{idle},i}$, which increases directly with the number of machines kept available. The second is the processing-related contribution, which depends on the configuration-specific station utilisation returned by the analytical evaluator. Capacity configurations may therefore differ both in the number of active machines and in the workload distribution generated by the resulting throughput and station mix.

For configurations satisfying $u_i^{\text{raw}} \leq 1$, so that $u_i^{\text{clip}} = u_i^{\text{raw}}$, the utilisation definition

$$u_i^{\text{raw}} = \frac{\lambda_i \bar{S}_i}{x_i}$$

gives the equivalent decomposition

$$P_{\text{tot}}(\mathbf{x}) = \sum_{i=1}^m \left[x_i P_{\text{idle},i} + (P_{\text{proc},i} - P_{\text{idle},i}) \lambda_i(\mathbf{x}) \bar{S}_i(\mathbf{x}) \right] + P_{\text{aux}}. \quad (20)$$

This decomposition is used only for interpretation. The numerical optimisation is evaluated directly through Equation (19), using the bounded utilisation returned by the evaluator.

For a fixed planning horizon H , the corresponding steady-state energy estimate is

$$E(\mathbf{x}; H) = H P_{\text{tot}}(\mathbf{x}). \quad (21)$$

Consequently, when the planning horizon is common to all candidate configurations, minimising average power is equivalent to minimising steady-state energy consumption over that horizon.

The first term is the capacity-dependent idle contribution associated with keeping x_i machines available. The second term is the processing-related contribution associated with the workload completed at the station. Consequently, differences among capacity configurations are driven directly by the number and location of active machines and indirectly by configuration-dependent changes in throughput and workload. This decomposition also helps explain why moderate utilisation errors do not necessarily change the power ranking when candidate configurations differ clearly in their idle-capacity contribution.

The clipped utilisation u_i^{clip} is used only to keep the reported power representation within its physical range.

The affine model is aligned with the decision granularity of the framework. It represents the average power associated with keeping a set of machines available over a shift-level planning horizon, rather than the sequence of transient machine states within the shift. The main activation decision therefore affects power through both the number of machines kept online and their estimated average utilisation.

The parameters $P_{\text{idle},i}$ and $P_{\text{proc},i}$ can be estimated from machine monitoring data, manufacturer information, or measurement procedures consistent with the ISO 14955 standard series (Behrendt, Zein, and Min 2012; ISO 14955-1:2017 2017; ISO 14955-2:2018 2018; Vijayaraghavan and Dornfeld 2010). The experimental values used in this study are stylised but compatible with machine-tool power ranges; plant deployment would require station-specific calibration.

Start-up energy, warm-up time, minimum on/off durations, and state-dependent switching dynamics are not included in the optimisation objective. Their relevance depends on the planning horizon and the frequency of reconfiguration. The numerical study therefore complements the steady-state model with break-even and energy-function sensitivity analyses.

5.2. Optimisation formulation

Let

$$\mathcal{I}_{\text{vis}} = \{i : \exists r, V_{ir} > 0\}$$

denote the set of stations visited by at least one product class, and let

$$\mathcal{I}_{\text{cap}} \subseteq \mathcal{I}_{\text{vis}}$$

denote the optional subset of stations subject to utilisation limits.

Given a maximum admissible throughput-loss fraction δ and a robustness margin $\varepsilon \geq 0$, the energy-aware

capacity-reconfiguration problem is formulated as

$$\min_{\mathbf{x} \in \mathbb{Z}_+^m} P_{\text{tot}}(\mathbf{x}) \quad (22)$$

subject to

$$TH_{\text{emda}}(\mathbf{x}) \geq (1 + \varepsilon)(1 - \delta)TH_{\text{all}}^{\text{emda}}, \quad (23)$$

$$u_i^{\text{raw}}(\mathbf{x}) \leq \bar{u}_i, \quad \forall i \in \mathcal{I}_{\text{cap}}, \quad (24)$$

$$1 \leq x_i \leq c_{i,\text{max}}, \quad \forall i \in \mathcal{I}_{\text{vis}}, \quad (25)$$

$$x_i = 0, \quad \forall i \notin \mathcal{I}_{\text{vis}}. \quad (26)$$

Because the all-on configuration is used as the throughput reference, the robustness margin and the admissible throughput-loss fraction must satisfy

$$(1 + \varepsilon)(1 - \delta) \leq 1. \quad (27)$$

Equivalently,

$$0 \leq \varepsilon \leq \frac{\delta}{1 - \delta}. \quad (28)$$

If this condition is violated, even the all-on configuration cannot satisfy the robust analytical target. In particular, $\delta = 0$ requires $\varepsilon = 0$.

Unless otherwise stated, the decision-oriented experiments use $\delta = 0.25$. The main Campaign-B solver benchmark, the 30-instance Campaign-B decision validation, and the 24-configuration end-to-end verification use $\varepsilon = 0$ to assess the unbuffered performance of the evaluator and search procedure. The manufacturing-inspired case and the supplementary Campaign-A solver-baseline comparison use $\varepsilon = 0.02$, while the dedicated sensitivity analyses vary ε explicitly. Pure prediction experiments, which do not apply a feasibility threshold, do not involve ε .

The value $\varepsilon = 0.02$ is used as an empirically motivated conservative setting for the tested two-class instances. It is not proposed as a universal calibration. Its suitability depends on the one-sided analytical overestimation observed near the application-specific throughput boundary and on the operational cost of a false-positive capacity decision.

Only evaluator-usable configurations, as defined by the convergence and internal consistency conditions in Section 4.4, are eligible for analytical feasibility and ranking.

5.3. Hybrid beam-GA search

The number of admissible machine vectors grows as

$$\prod_{i \in \mathcal{I}_{\text{vis}}} c_{i,\text{max}},$$

making exhaustive enumeration increasingly expensive as the number of stations and installed machines

increases. The optimisation layer therefore combines a congestion-guided beam search with a genetic-algorithm fallback.

The beam search begins from the structurally minimal configuration,

$$x_i = 1, \quad i \in \mathcal{I}_{\text{vis}},$$

and progressively adds capacity until the robust throughput requirement is satisfied. At each explored configuration, candidate expansions are prioritised through

$$\text{score}_i = \frac{EN_i u_i^{\text{raw}}}{P_{\text{idle},i}}. \quad (29)$$

for stations with $x_i < c_{i,\text{max}}$. The score favours stations that combine high local job accumulation, intensive use of the active capacity, and a relatively low baseline power cost for activating an additional machine.

The raw utilisation and congestion terms have complementary roles in this heuristic. The raw utilisation u_i^{raw} identifies stations whose active servers are heavily loaded or analytically overloaded, whereas EN_i identifies where jobs accumulate in the equivalent closed network. Their combination is used only to order neighbouring expansions. Every generated configuration is subsequently re-evaluated by the complete analytical evaluator, and final feasibility and ranking depend on $TH_{\text{emda}}(\mathbf{x})$ and $P_{\text{tot}}(\mathbf{x})$, not on the expansion score.

At each iteration, the beam retains a limited set of promising configurations ranked first by distance from the robust throughput target and then by estimated power. Once a feasible configuration is reached, a trimming phase attempts to remove machines one at a time while preserving analytical feasibility and reducing power. Memoisation avoids repeated evaluation of previously visited vectors.

A bottom-up beam is effective when successive capacity additions create an informative path from the minimal configuration towards the feasible region. Some capacity lattices, however, contain path-constrained structures in which early configurations are analytically unusable or provide weak guidance towards feasibility. In such cases, the restricted beam frontier may fail to reach a feasible region even when feasible configurations exist elsewhere in the search space.

The framework therefore invokes a genetic-algorithm fallback whenever the beam does not return an evaluator-feasible configuration. The GA uses the same analytical evaluator, power objective, throughput target, robustness margin, and capacity bounds; only the exploration mechanism changes. This design preserves a common decision model while providing a broader search over difficult capacity landscapes.

The detailed beam and GA pseudocode is reported in Appendix 4. Numerical search settings and comparison protocols are specified in the experimental-design section.

The hybrid procedure is intended to combine the computational efficiency of directed expansion with the broader coverage of population-based search. Exact enumeration remains the reference method when the capacity state space is sufficiently small, while DES is reserved for external validation of selected and boundary configurations.

6. Experimental design

6.1. Research questions and datasets

The experimental study addresses two broad objectives. The first is predictive: to assess how accurately the analytical evaluator reproduces DES throughput, ranking, and feasibility classification across different operating conditions. The second is decision-oriented: to determine whether the evaluator, when embedded in the optimisation framework, identifies lower-power configurations that remain feasible under DES.

These objectives are organised around five research questions:

- (RQ1) How accurately does the evaluator predict throughput and preserve the ranking of alternative capacity configurations?
- (RQ2) Does the configuration-specific station-mix update improve performance relative to frozen-mix aggregation rules?
- (RQ3) How reliably does the evaluator classify configurations as throughput-feasible or infeasible?
- (RQ4) How sensitive are the analytical decision signals to service-time variability and energy-model assumptions?
- (RQ5) How effectively does the hybrid search identify low-power configurations, and do the selected configurations remain feasible under DES?

The study uses three datasets with distinct purposes.

The first is a *structured validation battery* of 76 capacity configurations constructed from small three-station scenarios. These scenarios isolate routing symmetry, re-entrant flow, extreme product mix, variation in total WIP, and changes in service-time variability. Their purpose is to make approximation errors interpretable in relation to specific system mechanisms.

The second experimental component, denoted as Campaign A, supports a supplementary baseline comparison among beam search, greedy best-first search, and top-down hill climbing at the tighter robustness margin

$\varepsilon = 0.02$; greedy best-first search and hill climbing are additionally run on the Campaign-B instances under the identical evaluator in the main benchmark. Campaign A comprises 120 attempted random instances divided into small, medium, and large dimensional groups. One attempted instance produces a non-usable all-on analytical evaluation and is excluded from the reported comparison, leaving 119 usable instances. The Campaign-A dataset serves only this supplementary baseline comparison and is not pooled with the random instances used in the main optimisation and DES analyses.

The third experimental component, denoted as Campaign B, supports the main optimisation benchmark, the detailed random-instance DES validation, and the robustness-margin sensitivity analysis. Campaign B comprises 120 two-class random instances, with 40 instances in each dimensional group $m \in \{5, 8, 10\}$. The complete pool is used for the optimisation benchmark. Because replication-based DES is substantially more computationally demanding than analytical evaluation, a subset of 30 instances, with 10 instances per dimensional group, is selected from Campaign B for detailed simulation-based validation.

For each DES-validated Campaign-B instance, up to five decision-relevant configurations are considered: the all-on configuration, the evaluator-selected configuration, the analytically feasible configuration closest above the throughput threshold, the analytically infeasible configuration closest below the threshold, and one intermediate configuration. Duplicate vectors are removed, resulting in 146 distinct configuration evaluations. The 30-instance DES-validation set is therefore a subset of Campaign B and is not related to Campaign A.

A separate 15-instance subset of Campaign B, with five instances per dimensional group, is used for the robustness-margin sensitivity analysis. The optimisation is rerun independently for every value of ε .

The fourth experimental component is a manufacturing-inspired illustrative case. Its routing, processing-time, installed-capacity, WIP, and power parameters are selected to represent a plausible shared machining environment. The case illustrates model calibration and end-to-end use of the framework, but the parameters are not obtained from a specific operating

production line and the experiment does not constitute industrial field validation.

6.2. Structured scenarios

The structured scenarios are summarised in Table 3. All experiments use two product classes, consistently with the scope of the analytical evaluator.

The base scenarios are used to examine pointwise predictive accuracy under interpretable routing and mix conditions. The MIX and NSWEEP families assess how the approximation behaves as the class composition and total closed-network population change. The distributional stress tests retain the same mean processing times while changing the service-time distributions used in DES.

Full routing, processing-time, population, and installed-capacity parameters are reported in Appendix 2.

6.3. Random-instance campaigns, generation, and subset selection

Two purpose-specific random-instance campaigns are used. Their results are reported separately and their observations are not pooled.

Campaign A supports the supplementary comparison among beam search, greedy best-first search, and top-down hill climbing. It contains 120 attempted instances divided into small, medium, and large dimensional groups. The group-specific pseudo-random streams are initialised from values based on $42 + m$. Total population is sampled between 2 m and 5 m, processing-time parameters are sampled from uniform distributions, station capacities are generated and internally adjusted to avoid structurally unusable cases, idle powers are sampled from $U(1.5, 2.5)$ kW, and processing-power increments are sampled from $U(2, 4)$ kW. One attempted instance has a non-usable all-on analytical evaluation and is excluded, leaving 119 usable instances in the reported Beam-Greedy-Hill-Climbing comparison.

Campaign B supports the main optimisation benchmark, the random DES-validation experiment, and the robustness-margin sensitivity analysis. It contains 120

Table 3. Structured validation scenarios.

Family	Scenario	m	N	α	n_{cfg}
Base cases	BASE-A: symmetric routing	3	20	0.70	5
	BASE-B: re-entrant class-0 flow	3	30	0.70	2
	BASE-C: extreme 95/5 mix	3	20	0.95	5
Mix sweep	MIX: $\alpha \in \{0.0, 0.1, \dots, 1.0\}$	3	30	varied	4 per mix
WIP sweep	NSWEEP: $N \in \{10, 15, 20, 30, 45\}$	3	varied	0.60	4 per WIP
Distributional stress	DET and SCV sweep	3	30	0.50	4 per distribution

Note: Here, m is the number of stations, N the total population, α the fraction of class 0, and n_{cfg} the number of capacity configurations evaluated per scenario.

two-class instances, with 40 instances in each dimensional group. Small instances have ($m = 5$) and use seeds 501–540, medium instances have ($m = 8$) and use seeds 801–840, and large instances have ($m = 10$) and use seeds 1001–1040. The complete Campaign-B pool is used for the main optimisation benchmark.

Campaign A and Campaign B use the same analytical evaluator, power model, feasibility definitions, and decision variables, but they use different random-instance generators and serve different experimental purposes. Campaign A supports only the supplementary comparison among beam search, greedy best-first search, and top-down hill climbing. Campaign B supports the main exact/beam/GA benchmark, the 30-instance DES-validation subset, and the robustness-margin sensitivity analysis. Results from the two campaigns are reported separately and are never pooled.

For Campaign B, the class-probability vector is sampled from a symmetric Dirichlet distribution and the nominal population is allocated through a multinomial draw. A minimum class population of one is enforced. Visit ratios are generated independently as $V_{ir} = 1 + \text{Poisson}(1)$, so every class visits every station. A station-specific base processing time is sampled from $U(2, 12)$ min. Class heterogeneity is introduced through a lognormal multiplier clipped to $[0.4, 3]$, together with an independent station-class multiplier sampled from $U(0.8, 1.2)$. Service times are bounded below by 0.5 min. Idle power is sampled from $U(1, 3)$ kW, processing power is defined as $P_{\text{proc},i} = P_{\text{idle},i} + U(1.5, 5)$ kW, and $P_{\text{aux}} = 0.3$ kW.

For the additional DES assessment, 10 Campaign-B instances are selected from each dimensional group after ordering successful benchmark instances by identifier. Approximately equally spaced positions are used so that the 30-instance subset spans the full ordered group rather than concentrating on consecutive seeds.

A separate Campaign-B subset of 15 instances, with five instances per dimensional group, is used for the robustness-margin sensitivity study. Approximately equally spaced successful instances are selected within each ordered dimensional group, and the hybrid optimisation is rerun independently for every value of ε .

Complete Campaign-A baseline outputs, Campaign-B instance definitions, Campaign-B benchmark outputs, and the definitions of the Campaign-B DES and robustness-margin subsets are included in the supplementary reproducibility package.

A separate divergence-study pool supports the decision-level analysis of selection divergence in Section 7.2. It is generated with the same random protocol as Campaign B – the same analytical evaluator, power model,

feasibility definitions, Dirichlet class-probability sampling, visit-ratio construction, and service-time model – but at smaller station counts, $m \in \{3, 4\}$, and from an independent seed pool disjoint from the Campaign-B seeds, yielding approximately 300 usable instances per dimensional family. Smaller station counts are used deliberately, because they make mix-dependent bottleneck shifts both more frequent and easier to adjudicate by DES. Alongside this realistic-random pool, four mix-dependent stress families are constructed to isolate the regimes in which the bottleneck identity depends on the class mix. The *opposite per-class bottlenecks* family places the two classes' heaviest demands on different shared stations, in a balanced-population and a skewed-population variant, in which the nominal population K is split evenly or strongly unevenly across classes. The *wide capacity range* family samples installed capacities c_{max} over a broad interval, so that a single station alternates between binding and slack as the mix changes. The *extreme service asymmetry* family amplifies the per-class service-time multipliers, so that the dominant station differs sharply between classes. The *shared bottleneck (partial overlap)* family lets the two classes share one congested station while differing on the remaining stations. Each stress family retains the Campaign-B service-time, visit-ratio, and power conventions and contains approximately the same number of usable instances per family as the realistic-random pool. The complete divergence-study definitions and seed lists are included in the supplementary reproducibility package.

6.4. Validation metrics and feasibility definitions

Three notions of feasibility are distinguished.

A configuration is *evaluator-usable* if the fixed-point procedure converges and its internal throughput-consistency gap satisfies

$$\frac{|TH_{\text{mda}} - TH_{\text{emda}}|}{TH_{\text{emda}}} \leq 0.05.$$

A configuration is *evaluator-feasible* if it is evaluator-usable and

$$TH_{\text{emda}}(\mathbf{x}) \geq (1 + \varepsilon)(1 - \delta)TH_{\text{all}}^{\text{emda}}. \quad (30)$$

A configuration is *DES-feasible* if

$$TH_{\text{DES}}(\mathbf{x}) \geq (1 - \delta)TH_{\text{all}}^{\text{DES}}. \quad (31)$$

The analytical optimisation therefore uses the all-on EMDA baseline, whereas external verification uses the corresponding all-on DES baseline for the same instance.

Pointwise throughput accuracy is measured through mean absolute percentage error,

$$\text{MAPE} = \frac{100}{n} \sum_{j=1}^n \left| \frac{TH_{\text{DES},j} - TH_{\text{emda},j}}{TH_{\text{emda},j}} \right|, \quad (32)$$

and mean signed percentage bias,

$$\text{Bias} = \frac{100}{n} \sum_{j=1}^n \frac{TH_{\text{DES},j} - TH_{\text{emda},j}}{TH_{\text{emda},j}}. \quad (33)$$

Throughout the manuscript and the supplementary material, all throughput percentage errors, MAPE values, and signed bias values are normalised by the corresponding analytical estimate TH_{emda} .

A negative bias therefore indicates analytical overestimation of DES throughput. Ranking preservation is assessed through Spearman correlation between analytical and DES throughput values. Feasibility classification is evaluated through accuracy, precision, recall, false-positive rate, and false-negative rate:

$$\begin{aligned} \text{Precision} &= \frac{TP}{TP + FP}, & \text{Recall} &= \frac{TP}{TP + FN}, & \text{FPR} &= \frac{FP}{FP + TN}, \\ \text{FNR} &= \frac{FN}{FN + TP}. \end{aligned}$$

The validation evidence is interpreted at four levels. Pointwise metrics measure numerical throughput accuracy; Spearman correlation measures the preservation of configuration rankings; classification metrics assess the risk of accepting or rejecting configurations around the throughput boundary; and end-to-end verification assesses the DES performance of the configuration ultimately selected by the optimisation procedure. This distinction separates average prediction quality from decision reliability near the feasibility threshold.

6.5. DES protocol

The DES model mirrors the principal structural assumptions of the analytical evaluator: a closed two-class CONWIP population, FCFS service, identical parallel machines within each station, and fixed active-machine counts over the simulated horizon.

Unless a distributional stress test is specified, processing times are exponentially distributed with the same means used by the analytical model. The robustness experiments replace the exponential distribution with deterministic, Erlang, or lognormal distributions while preserving the service-time means.

The structured validation experiments use 30 independent replications per configuration, whereas the

random-instance DES validation uses 20 independent replications per configuration. Structured scenarios are simulated for 200,000 min after a warm-up period of 20,000 min. Re-entrant or highly congested scenarios use a 300,000 min horizon and a 30,000 min warm-up. The manufacturing-inspired case uses 20 independent replications, a 200,000 min horizon, and a 20,000 min warm-up. Even at the reduced count, the achieved sampling precision is high: across the random-instance DES configurations the 95% confidence-interval half-width is of the order of 0.3% of the mean throughput and at most about 0.6%. Feasibility near the boundary is decided on the conservative lower confidence bound rather than the replication mean, so the decision-relevant quantity is the throughput-to-target margin net of this half-width; at the achieved precision the half-width is at least about forty-five times below $\delta = 0.25$, and close to two orders of magnitude below it at the median, which keeps both the mean-based and the confidence-bound feasibility classification stable across the replication count. The half-widths achieved across the DES campaigns are reported in Appendix 2.

Throughput is calculated from post-warm-up departures. The relative percentage difference reported in the structured analyses is

$$100 \frac{TH_{\text{DES}} - TH_{\text{emda}}}{TH_{\text{emda}}}, \quad (34)$$

so a negative value indicates that the analytical evaluator overestimates DES throughput.

Replication uncertainty is represented through Student- t 95% confidence intervals. End-to-end feasibility is reported both with respect to the replication mean and, where decision risk is relevant, with respect to the lower confidence bound. Full confidence-interval calculations and seed policies are provided in Appendix 2.

6.6. Optimisation benchmark design

The optimisation experiments compare exact enumeration, beam search, the hybrid beam-GA procedure, and additional search baselines where appropriate. Exact enumeration is used as the reference whenever the capacity state space is sufficiently small. For larger instances, performance is assessed against the best solution found under the common evaluation protocol.

The main exact, beam, and GA benchmark described in this subsection uses Campaign B, and greedy best-first search and hill climbing are run on the same 120 Campaign-B instances under the identical evaluator within this subsection (Table 15, $\varepsilon = 0$). A separate supplementary baseline comparison on Campaign A, at

the tighter margin $\varepsilon = 0.02$ and with a different instance generator, is reported in Appendix A.15.

All search methods use the same EMDA evaluator, power model, throughput target, capacity bounds, evaluator-usability criterion, and robustness margin within a given experiment. Their performance is compared through:

- (i) success rate in finding an evaluator-feasible configuration;
- (ii) relative power gap from the exact or best-known solution;
- (iii) number of distinct EMDA evaluations;
- (iv) wall-clock computation time.

The main optimisation experiments use $\delta = 0.25$. The default analytical robustness margin is $\varepsilon = 0.02$, while selected benchmark tables also report $\varepsilon = 0$ to isolate search performance from the effect of conservative feasibility screening. The beam width and candidate-list size are adapted to instance scale. The baseline GA uses a population of 60 individuals, 150 generations, an elite size of 5, crossover probability 0.70, and mutation probability 0.40. GA runs use independent seeds, and separate sensitivity experiments examine lighter and heavier parameter settings.

The detailed search parameters, stopping rules, computational environment, and evaluation-accounting conventions are consolidated in Appendix 2. The beam and GA pseudocode is reported in Appendix 4.

6.7. Manufacturing-inspired illustrative case setup

The manufacturing-inspired case represents a five-station shared machining line with two product classes. Both classes visit the main preparation, machining, secondary machining, and finishing resources, while class 1 additionally requires an inspection or rework operation and makes an additional visit to the secondary machining station. Table 4 reports the routing, processing, capacity, and power parameters of the case.

Table 4. Manufacturing-inspired case: routing, processing, capacity, and power data.

Station	Resource	$c_{i,\max}$	V_{i0}	V_{i1}	S_{i0} [min]	S_{i1} [min]	$P_{\text{idle},i}$ [kW]	$P_{\text{proc},i}$ [kW]
1	Load/preparation	2	1	1	2.0	2.2	1.5	3.8
2	Primary machining	4	1	1	7.5	8.5	2.2	5.0
3	Secondary machining	3	1	2	5.0	6.0	2.0	4.6
4	Inspection/rework	2	0	1	–	4.0	1.4	3.2
5	Finishing/unload	3	1	1	3.0	3.5	1.6	3.7

Note: A dash indicates that the corresponding class does not visit the station.

The installed-capacity vector is

$$\mathbf{c}_{\max} = [2, 4, 3, 2, 3],$$

and the class-population vector is

$$\mathbf{K} = [20, 10],$$

giving a total CONWIP population of $N = 30$. The optimisation uses $\delta = 0.25$ and $\varepsilon = 0.02$. The case is evaluated through the same analytical and DES procedures used in the main experiments. The selected configuration and a lower-power boundary alternative are both verified by DES to illustrate the distinction between energy reduction and throughput feasibility.

7. Results

The results are organised around the five research questions introduced in Section 6.1. The analysis first considers throughput prediction and ranking, then isolates the contribution of the dynamic station-mix update, examines feasibility classification and end-to-end decision quality, evaluates robustness to service-time variability, and finally assesses energy savings, search performance, and the manufacturing-inspired case.

7.1. Predictive accuracy and ranking preservation

Table 5 reports pointwise accuracy for the three structured base scenarios. Throughput, which drives the capacity-reconfiguration decision, is predicted accurately across the tested configurations. Throughput MAPE ranges from 0.81% in BASE-B to 2.50% in BASE-A. Errors in cycle time, utilisation, and waiting time are larger, particularly in highly congested configurations, while aggregate throughput remains comparatively stable.

Table 5. Pointwise accuracy of the analytical evaluator on the structured base scenarios, averaged over the evaluated capacity configurations.

Scenario	Metric	MAPE (%)	Min err (%)	Max err (%)
BASE-A	TH	2.50	−10.83	0.39
	CT_r	9.88	−20.16	56.01
	u_j^{clip}	4.98	−24.75	17.62
	W_i	32.2	−40.79	117.0
BASE-B	TH	0.81	0.81	0.82
	CT_r	1.25	−2.32	0.27
	u_j^{clip}	0.46	−0.08	1.13
	W_i	4.79	−8.45	6.82
BASE-C	TH	0.95	−0.57	3.13
	CT_r	8.83	−35.92	1.20
	u_j^{clip}	8.01	−1.18	55.74
	W_i	35.7	−100.0	15.32

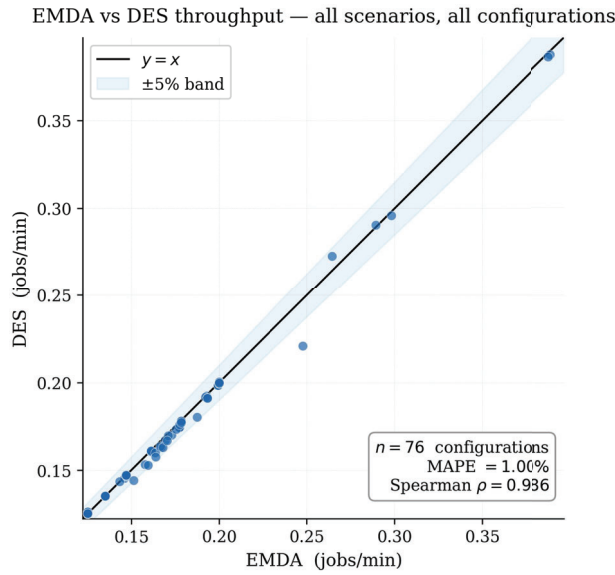


Figure 4. Analytical throughput versus DES throughput across the structured exponential scenarios ($n = 76$ configurations). The solid line is the $y = x$ reference and the shaded band marks the $\pm 5\%$ error region. MAPE = 0.96% and Spearman $\rho = 0.986$.

Across the full structured exponential battery of 76 configurations, throughput MAPE is 0.96% and Spearman rank correlation is $\rho = 0.986$. This coefficient pools all 76 configurations into a single ranking, whereas the per-scenario coefficients reported in Table 10 are computed within each scenario and are correspondingly close to unity. Figure 4 shows that most configurations lie within the $\pm 5\%$ error band. The largest deviation occurs in a low-capacity, highly congested BASE-A configuration.

The structured sweeps identify the operating conditions under which the approximation is most stable. Across the MIX scenarios, analytical and DES throughput remain within 0.25% for all eleven class-mix values, producing an overall MAPE of 0.39%. Station 2 remains the dominant bottleneck across the sweep, and the evaluator preserves the resulting weak sensitivity of total throughput to the class proportion. Lower rank correlations at extreme values of α arise because the throughput differences among configurations become very small.

Across the WIP sweep, the evaluator exhibits a modest positive throughput bias at low population levels. The relative difference decreases from -1.54% at $N = 10$ to -0.18% at $N = 45$, indicating closer agreement as the closed-network population increases.

The decision-oriented random-instance DES subset produces a similar pattern across the three dimensional groups. Table 6 reports the results for 146 configurations across 30 instances. Overall throughput MAPE is 0.913%, mean bias is -0.210% , and median Spearman correlation is 0.975. Across the 29 instances for which a within-instance throughput ranking is defined, the per-instance Spearman coefficient has a median of 0.97, an interquartile range of 0.87 to 1.00, and a minimum of 0.36, so ranking is essentially exact for most instances and weakens only on a few small instances whose candidate configurations differ in throughput by amounts comparable to simulation noise. Predictive accuracy improves with instance size, with MAPE decreasing from 1.363% in the small group to 0.574% in the large group.

In raw counts, the small-instance group of this set contributes 33 true positives, 2 false positives, 9 true negatives, and 2 false negatives, so its FPR = 0.182 corresponds to two of the eleven actually-infeasible small configurations being classified feasible; across the full set of 146 configuration evaluations the totals are 112 true positives, 2 false positives, 28 true negatives, and 4 false negatives. Reporting these counts makes the small-group false-accept rate auditable directly rather than only as a rate.

7.2. Contribution of the dynamic station-mix update

The configuration-specific evaluator updates the station arrival mix within the class-throughput fixed point. To isolate the contribution of this operation, the proposed dynamic rule is compared with three alternatives:

- STATIC-INIT, in which the station proportions are fixed at their initial values;
- STATIC-ALL-ON, in which the proportions are calculated under the all-on configuration and reused for every candidate vector;

Table 6. Predictive and decision-oriented validation on the 30-instance Campaign-B DES subset ($n = 146$ configurations).

Group	n	Regression quality			Decision quality				
		MAPE (%)	Bias (%)	Spearman	Acc	Prec	Rec	FPR	FNR
Small	46	1.363	-0.170	0.900	0.913	0.943	0.943	0.182	0.057
Medium	50	0.837	-0.247	0.975	0.960	1.000	0.949	0.000	0.051
Large	50	0.574	-0.208	0.950	1.000	1.000	1.000	0.000	0.000
Overall	146	0.913	-0.210	0.975	0.959	0.982	0.966	0.067	0.034

Note: FPR = $FP / (FP + TN)$; FNR = $FN / (FN + TP)$.

Table 7. Ablation of the station-mix update rule on the structured two-class exponential battery ($n = 76$ configurations, $\delta = 0.02$, $\varepsilon = 0$).

Variant	TH MAPE (%)	Bias (%)	Prec	Rec
Dynamic	0.96	−0.78	0.971	0.971
Static-init	1.16	−0.87	0.971	0.971
Static-all-on	1.11	−0.93	0.971	0.971
Global-K	1.06	−0.72	0.971	1.000

Note: MAPE and bias use the analytical throughput estimate Th_{emda} as the denominator, consistently with Equations (32) and (33).

- GLOBAL-K, in which class populations and visit ratios define a configuration-independent class-share approximation.

Table 7 shows that the dynamic update produces the lowest throughput MAPE, equal to 0.96%, compared with 1.06–1.16% for the configuration-independent alternatives. The dynamic update attains the lowest throughput MAPE against every frozen-mix rule, a consistent relative reduction of approximately 9% to 17%, and the decision-level analysis below shows this margin widens precisely where the bottleneck depends on the mix. Ranking and feasibility classification remain stable across the four variants, reflecting the mix-independent regime of this battery, in which the bottleneck identity does not depend on the aggregation rule.

Paired Wilcoxon signed-rank tests were also conducted on the absolute percentage errors of the 76 matched configurations. The unadjusted two-sided p -values for the comparisons of the dynamic update with Static-init, Static-all-on, and Global-K are 0.0438, 0.0199, and 0.0285, respectively. After Holm correction for the three related comparisons, the adjusted p -value is 0.0598 in each case.

All three comparisons rank the dynamic update ahead of the frozen rules, with Holm-adjusted p -values close to the 5% family-wise level (0.0598 in each case); because the configurations share structured scenarios, this ordering is read together with the decision-level evidence below, which carries the principal weight for the contribution.

On this structured battery, evaluated at the tight tolerance $\delta = 0.02$, the four aggregation rules are descriptively close. This is itself an informative robustness property: when the bottleneck identity is essentially mix-independent, the evaluator is robust to the aggregation rule. The decisive evidence for the configuration-specific update concerns instead the deployment decision at the operational tolerance, which we examine next.

7.2.1. When the configuration-specific update changes the decision.

The configuration-specific reconstruction operates at the deployment level, not only at the level of descriptive accuracy. At the operational tolerance $\delta = 0.25$ used throughout the energy experiments, every aggregation rule selects the minimum-power configuration it deems feasible, and a *selection divergence* occurs when the dynamic rule selects a different active-machine vector than a frozen surrogate; disagreements are then adjudicated by DES.

The divergence pool is drawn from the same random-instance protocol that generates the Campaign B benchmark, so the decision-relevance evidence is taken from the same instance family as the main experiments. The instances are sized at $m \in \{3, 4\}$ and drawn from an independent seed pool precisely so that the minimum-power configuration each rule deems feasible can be identified by exhaustive enumeration. The observed dynamic-versus-frozen divergences are therefore not artefacts of the search procedure (Section 6.3). Within this pool, selection divergence occurs in 18.1% of $m = 3$ instances and 21.4% of $m = 4$ instances, and the rate is stable across $\delta \in [0.15, 0.30]$ (Table 8). In the tested regimes with strong mix-dependent bottleneck shifts, opposite per-class bottlenecks on shared stations, wide capacity ranges, or skewed class populations, divergence rises to between 56% and 88%. The underlying inconsistency is structural rather than search-dependent: a frozen station mix cannot represent the configuration-induced re-weighting of the two classes at a shared station, so capacity changes at that station may be mis-evaluated. This mechanism can arise whenever bottleneck identity depends on the local class mix. However, the reported divergence rates pertain to the $m \in \{3, 4\}$ pool and do not establish size-independent prevalence or quantify the frequency of the effect for the $m \in \{5, 8, 10\}$ benchmark systems. Where the active-machine vector leaves the bottleneck identity unchanged, the four aggregation rules generally coincide

Table 8. Prevalence of selection divergence between the dynamic configuration-specific rule and the frozen-mix surrogates at the operational tolerance $\delta = 0.25$.

Instance family	m	Divergence at $\delta = 0.25$	Range over $\delta \in [0.15, 0.30]$
Realistic random, divergence study	3	18.1%	16–20%
Realistic random, divergence study	4	21.4%	21–28%
Opposite per-class bottlenecks, balanced K	3	78.9%	77–83%
Opposite per-class bottlenecks, skewed K	3	72.3%	65–72%
Wide capacity range	3	87.7%	82–88%
Extreme service asymmetry	4	78.9%	73–79%
Shared bottleneck, partial overlap	4	56.3%	44–56%

Note: Divergence occurs when the active-machine vector selected by the dynamic rule differs from that selected by at least one frozen surrogate. Each family contains approximately 300 usable instances.

Table 9. Mechanism of a selection divergence on a representative realistic random instance (divergence-study generator, same protocol as Campaign B, $m = 3$, seed 73002; $V = [[1, 2], [2, 3], [1, 3]]$, $K = [4, 5]$, $S = [[2.58, 5.10], [4.23, 9.89], [3.99, 5.94]]$, $c_{\max} = [1, 3, 3]$; DES target 0.123 jobs/min = $0.75 \text{ TH}_{\text{all-on}}^{\text{DES}}$ with $\text{TH}_{\text{all-on}}^{\text{DES}} = 0.164$).

Aggregation rule	Selected x	TH_{emda}	P_{tot} (kW)	DES TH ($\pm \text{CI95}$)	DES-feasible
Dynamic (proposed)	[1, 3, 2]	0.166	33.5	0.160 ± 0.0004	Yes
Static-init	[1, 2, 2]	0.131	26.9	0.118 ± 0.0004	No
Static-all-on	[1, 3, 2]	0.165	33.5	0.160 ± 0.0004	Yes
Global-K	–	–	–	–	–

Note: The frozen Static-init rule selects a cheaper vector it labels feasible, but DES places its throughput below target; Global-K finds no feasible candidate; only Static-all-on coincides with the dynamic selection on this instance.

Table 10. Decision-oriented validation on the structured set: throughput-rank correlation and feasibility classification for $\delta = 0.02$ and $\varepsilon = 0$.

Family	Scenario or range	Spearman ρ	Acc	Prec	Rec
Base cases	BASE-A	1.00	0.80	1.00	0.50
	BASE-B, re-entrant	n/a	1.00	1.00	1.00
	BASE-C, extreme mix	1.00	1.00	1.00	1.00
Mix sweep	MIX, α range	0.60–1.00	1.00	1.00	1.00
WIP sweep	NSWEEP, N range	1.00	0.75–1.00	0.50–1.00	1.00
Overall exponential set ($n = 76$)		≈ 1.00	0.974	0.971	0.971
Deterministic robustness	DET	–	0.25	1.00	0.25

in the tested cases; where bottleneck identity changes, the configuration-specific reconstruction may alter the deployed configuration.

When the rules disagree, DES provides an external comparison of the selected configurations. Across the 39 DES-adjudicated divergent instances (15 replications, 95% confidence intervals), no frozen rule selected a better configuration under the DES-based feasibility-and-power comparison. In 24 cases, a frozen rule selected a configuration that DES classified as infeasible – because the frozen mix over-estimated throughput at a bottleneck station whose load it could not re-weight – whereas the dynamic rule rejected that configuration and selected a DES-feasible alternative. In the remaining 15 cases, the frozen selection was feasible but no cheaper. Thus, within this DES-adjudicated sample, the dynamic rule was strictly better in 24 instances and tied with the frozen rule in 15. Table 9 illustrates a representative realistic random instance: the frozen Static-init rule selects a cheaper vector that it labels feasible, but DES places its throughput below the production target, whereas the dynamic rule's selection is DES-feasible.¹

Within the DES-adjudicated divergent sample, the configuration-specific reconstruction matched or improved upon the frozen surrogates and avoided the infeasible selections observed for the frozen rules. This is a sample-specific empirical result rather than a universal dominance guarantee. Near the feasibility boundary, both evaluators remain approximations, and marginal cases should therefore be resolved through DES. The result provides the operational counterpart of the consistency argument: a frozen aggregation can misclassify a lower-power configuration as feasible when bottleneck identity depends on the local class mix.

7.3. Feasibility classification and end-to-end validation

Feasibility classification on the same 30-instance Campaign-B subset (Table 6) gives overall accuracy 0.959, precision 0.982, and recall 0.966. The false-positive rate is 0.067 overall and falls to zero in the medium and large groups. The corresponding precision of one indicates that no evaluator-feasible medium or large configuration in the sample is rejected by DES as infeasible.

False positives represent the more direct production risk because they correspond to configurations for which the analytical model would deactivate excessive capacity. False negatives instead reduce the attainable energy saving by rejecting configurations that DES would accept. The observed error pattern is therefore predominantly conservative outside the small-instance boundary cases.

Table 10 reports the structured classification results at $\delta = 0.02$. Classification is perfect in most scenarios. The two exceptions are located close to the throughput threshold: BASE-A contains one false negative, while NSWEEP at $N = 20$ contains one marginal false positive. Across the full structured exponential set of 76 configurations, 34 are DES-feasible and 42 infeasible, and the analytical classifier yields 33 true positives, 41 true negatives, one false positive, and one false negative. The overall precision and recall of 0.971 therefore reduce to a single boundary misclassification in each direction, with a Wilson 95% interval on the precision of 0.85 to 0.99, while the accuracy of 0.974 corresponds to 74 correct decisions out of 76.

Across the structured battery, the ordering of configurations by estimated power is also preserved, with

Table 11. DES verification of evaluator-selected configurations. Configurations were selected using $\delta = 0.25$ and $\varepsilon = 0.0$.

Group	n	MAPE on TH (%)	DES-feasible by mean	DES-feasible by 95% CI	Median ΔP (%)
BASE structured	9	0.87	8/9	8/9	29.4
Random small	5	1.99	4/5	4/5	27.2
Random medium	5	0.55	5/5	5/5	38.3
Random large	5	0.30	5/5	5/5	32.0
Overall	24	0.92	22/24	22/24	31.0

Note: DES feasibility is assessed against $TH_{\text{target}} = (1 - \delta)TH_{\text{all}}^{\text{DES}}$.

Spearman correlation equal to 1.000 between analytical and DES-based power values.

The end-to-end test examines the configurations selected by the optimisation procedure rather than isolated candidate vectors. Table 11 reports DES verification of 24 configurations selected with $\delta = 0.25$ and $\varepsilon = 0$. The zero-margin setting is used to assess the unbuffered end-to-end decision quality of the analytical evaluator and search procedure, while the effect of positive robustness margins is examined separately in Appendix A.8.

The two unsuccessful cases lie close to the target boundary. In BASE-C-M, the analytical estimate exceeds the target by only 0.04 jobs/h before DES places the configuration slightly below it. The random small-instance failure is similarly located within the simulation confidence region of the threshold. All selected medium and large random-instance configurations remain feasible.

Feasibility by replication mean and feasibility by the lower confidence bound provide two levels of decision evidence. The first evaluates expected simulated performance, while the second incorporates replication uncertainty. Agreement between the two criteria for 22 configurations indicates that the corresponding selections are separated from the target by more than simulation noise. Treating the 24 verified selections as independent feasibility outcomes, the 22 successes give a success proportion of 0.92 with a Wilson 95% confidence interval of 0.74 to 0.98. Because these selections pool deterministic structured cases with random small, medium, and large instances whose underlying success probabilities differ, this pooled interval is reported as a descriptive summary rather than as an estimate of a single homogeneous success rate.

7.4. Robustness to service-time variability

The analytical evaluator uses service-time means within an exponential closed-network approximation. Its sensitivity to distributional shape is assessed by retaining the same means in DES while varying the service-time squared coefficient of variation.

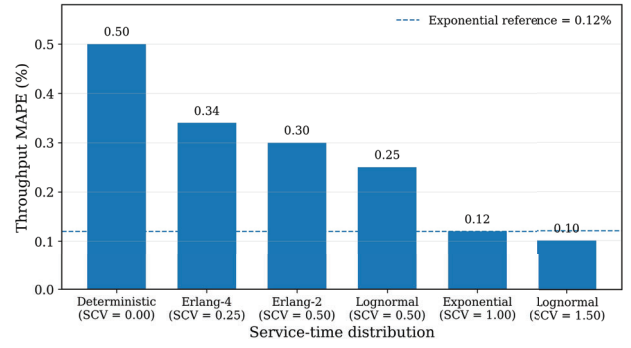


Figure 5. Throughput MAPE across alternative DES service-time distributions. The dashed line marks the exponential reference case. Approximation error remains at most 0.50% for all tested SCV values.

The MIX scenario with $\alpha = 0.5$, $N = 30$, and four capacity configurations is simulated under deterministic, Erlang-4, Erlang-2, lognormal, and exponential service times. The tested SCV values range from 0 to 1.5. Since the analytical inputs are unchanged, variation in the comparison results arises exclusively from the DES service-time distribution.

Figure 5 shows that throughput MAPE remains at most 0.50% for every distribution. Error is highest under deterministic service and decreases as variability approaches or exceeds the exponential reference.

The bias is positive when DES variability is below the exponential benchmark: deterministic and low-SCV systems produce slightly higher simulated throughput than the analytical estimate. Detailed results for each distribution are reported in Appendix A.7.

The separate deterministic boundary experiment highlights the distinction between numerical accuracy and feasibility classification. Although average throughput error remains small, three configurations cluster close to the threshold and are classified conservatively as infeasible by the evaluator. The resulting precision remains one, while recall falls to 0.25. Service-time variability therefore has limited impact on average throughput prediction in the tested scenarios but can affect the classification of near-threshold configurations.

7.5. Energy savings and solver performance

The structured optimisation experiments consider three capacity scales. Scale S uses the original installed capacities, scale M doubles them, and scale L triples them. The station power parameters are

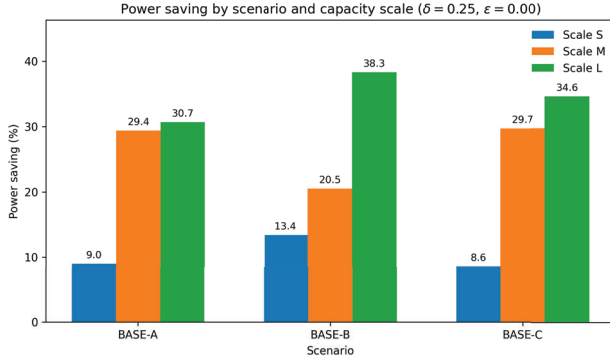
$$P_{\text{idle}} = [1.8, 2.2, 1.5] \text{ kW}, \quad P_{\text{proc}} = [4.5, 5.0, 3.8] \text{ kW},$$

with $P_{\text{aux}} = 0.5 \text{ kW}$.

Table 12. Power reduction as a function of the throughput tolerance δ at scale S.

Scenario	$\delta = 0.00$	$\delta = 0.02$	$\delta = 0.05$	$\delta = 0.10$	$\delta = 0.15$	$\delta = 0.25$
BASE-A	–	–	9.0	9.0	9.0	9.0
BASE-B	13.4	13.4	13.4	13.4	13.4	13.4
BASE-C	–	8.6	8.6	8.6	8.6	8.6

Note: Results use $\varepsilon = 0$. Dash: no evaluator-feasible reduced configuration.

**Figure 6.** Power reduction by structured scenario and capacity scale with $\delta = 0.25$ and $\varepsilon = 0$.**Table 13.** Energy-aware configurations at scale S with $\delta = 0.25$ and $\varepsilon = 0$.

Scenario	$\mathbf{x}_{\text{opt}}$	P_{all} (kW)	P_{opt} (kW)	ΔP (%)	TH_{emda} (jobs/h)
BASE-A	[2, 1, 2]	22.58	20.54	9.0	17.36
BASE-B	[1, 1, 1]	13.42	11.62	13.4	7.51
BASE-C	[2, 1, 2]	20.68	18.89	8.6	23.31

Table 12 reports the exact power saving at scale S as the admissible throughput loss δ changes. BASE-B supports a reduced-capacity configuration even at $\delta = 0$, whereas BASE-A and BASE-C require a small throughput tolerance before capacity can be removed.

At $\delta = 0.25$, the opportunity for energy reduction increases with installed capacity. Figure 6 shows that the medium and large scales contain more redundant capacity that can be deactivated while preserving the throughput requirement.

Table 13 reports the scale-S solutions. Power reductions range from 8.6% to 13.4%. At larger scales, savings reach 29.4% for BASE-A-M, 29.7% for BASE-C-M, and 38.3% for BASE-B-L.

The relevance of switching off capacity depends on the relation between the power reduction and any restart energy. If C_{start} is the restart energy per switched-off machine, the break-even horizon is

$$H^* = \frac{n_{\text{off}} C_{\text{start}}}{\Delta P}. \quad (35)$$

For the scale-S cases, published restart-energy values between 0.01 and 0.10 kWh imply break-even times of approximately 0.3–3 min (Frigerio, Tan, and Matta 2024; Mouzon, Yildirim, and Twomey 2007).

Table 14 reports the main search benchmark on the 120 Campaign-B instances with $\delta = 0.25$ and $\varepsilon = 0$. On successful runs, beam search reaches zero optimality gap while using fewer analytical evaluations and less computation time than the GA; the median runtime ratio relative to the GA is approximately $7.7\times$ on small instances and $2.8\text{--}3.1\times$ on medium and large instances. On this benchmark the hybrid reaches the reference solution on every instance, the exact optimum on the small and medium groups, where enumeration is feasible, and the best-found solution on the large group, so its defining property here is uniform reliability across the full instance set rather than raw search economy.

On the 120 Campaign-B instances, beam search identifies an evaluator-feasible solution in all three dimensional groups, corresponding to a 100% success rate. The 5% small-instance failure rate reported in the supplementary solver-baseline comparison belongs to Campaign A, which uses a separate random-instance generator and is reported independently. The structured BASE-A cases nevertheless show that bottom-up beam expansion can fail on path-constrained capacity landscapes, motivating the GA fallback.

Greedy best-first search and top-down hill climbing run on the same 120 Campaign-B instances under the identical EMDA evaluator, power objective, throughput target, and feasibility rule, placing the three search strategies on a common footing. Table 15 summarises the outcome. Both lighter baselines locate an evaluator-feasible solution on every instance, and on the typical instance both match the optimum, with a median optimality gap of zero. They diverge from the hybrid only in the tail. Hill climbing reaches the optimum on at least nine instances out of ten, with a ninth-decile gap of zero, and never loses more than 9.7% of the available power saving; greedy best-first stays within 0.7% of the optimum on ninety per cent of instances but degrades to 18.2% in the worst case. Across the full set, hill climbing reaches this reference solution on 113 of the 120 instances and greedy on 107, whereas the hybrid reaches it on all 120.

These figures place the value of the hybrid on the main benchmark in reliability rather than in raw evaluation economy. Greedy best-first is the lightest strategy, at a median of twelve evaluations, but it forgoes the reference solution on thirteen instances, by up to 18.2%; hill climbing is more reliable, at a median of thirty-seven evaluations, yet still settles on a plateau on seven instances, by up to 9.7%. The beam-GA hybrid reaches the reference on all 120 instances. On Campaign B the bottom-up beam already reaches it on every instance, so the GA fallback is not triggered here; the GA's corrective role is exercised on the structured BASE-A topology, where the minimal-capacity path stalls and

Table 14. Search benchmark on the 120 Campaign-B random instances with $\delta = 0.25$ and $\varepsilon = 0$.

Group	Beam search				GA			
	Success	Gap	Evals	Time (s)	Success	Gap	Evals	Time (s)
Small, $m = 5$	100%	0.00%	4	0.055	100%	0.00%	18	0.423
Medium, $m = 8$	100%	0.00%	45	3.171	100%	0.00%	127	8.770
Large, $m = 10$	100%	0.00%	67	14.72	100%	0.00%	183	46.10

Note: Gap is measured against exact enumeration when the state space does not exceed 500,000 configurations and against the best-found solution otherwise.

Table 15. Greedy best-first search and top-down hill climbing on the same 120 Campaign-B instances as Table 14, under the identical EMDA evaluator with $\delta = 0.25$ and $\varepsilon = 0$.

Method	Success	Reaches reference	Worst-case gap	Median evals
Hill climbing	100%	113/120	9.7%	37
Greedy best-first	100%	107/120	18.2%	12

Note: 'Reaches reference' counts instances at zero gap to the reference solution found by the beam-GA hybrid, namely the exact optimum on the small and medium groups (exhaustive enumeration) and the best-found solution on the large group; the worst-case gap is the largest power gap to that reference, and evaluation counts are medians over all 120 instances. The hybrid reaches the reference on all 120 instances (Table 14).

broader exploration restores the exact optimum. For a single energy-aware configuration decision that must be trusted without instance-by-instance tuning, this uniform reliability across the full instance set is the decisive property.

Comparisons under tighter robustness margins and alternative GA parameterisations are reported in the appendix. In the GA parameter-sensitivity experiment, all three tested settings recover the exact structured optima across all seeds, indicating that the structured conclusions are not sensitive to the tested GA parameterisation.

7.6. Manufacturing-inspired case and energy-model sensitivity

For the manufacturing-inspired case, the all-on configuration $\mathbf{c}_{\max} = [2, 4, 3, 2, 3]$ produces an analytical throughput of 27.22 jobs/h and draws 49.29 kW. With $\delta = 0.25$ and $\varepsilon = 0.02$, the nominal throughput target is 20.42 jobs/h and the robust analytical target is 20.82 jobs/h.

The optimisation selects $\mathbf{x} = [1, 3, 3, 1, 2]$. Its estimated power is 39.40 kW, corresponding to a 20.1% reduction relative to the all-on system. DES returns a mean throughput of 22.90 jobs/h with a 95% confidence interval of [22.87, 22.93], confirming feasibility above both the nominal and robust targets. Table 16 summarises this analytical selection and its DES verification.

To separate energy efficiency from energy curtailment, it is useful to report the specific energy $P_{\text{tot}}(\mathbf{x})/TH_{\text{DES}}$, the consumption per completed job. The all-on system draws 1.81 kWh/job and the selected configuration

1.72 kWh/job, a 4.9% reduction. The headline 20.1% power saving is curtailment-inclusive: it pairs the lower active capacity with a 16.0% reduction in throughput, from 27.25 to 22.90 jobs/h, and therefore measures the combined effect of running fewer machines at lower output rather than efficiency at fixed production. Specific energy isolates the part of the saving that persists at constant output, and for this manufacturing case the selected configuration is also the more efficient one per job. Across the Campaign-B instances the per-job efficiency gain is a distribution: the specific energy falls by a median of about 26% with an interquartile range of roughly 16–33%, and in the least favourable instance it is essentially unchanged (about -3.8%), where the saving is realised through curtailment rather than per-job efficiency. Since the model includes no exogenous demand process, the saving is reported at the throughput floor, the lowest output that still satisfies the production requirement.

The lower-power vector $[2, 4, 3, 1, 1]$ marks the point at which curtailment crosses into infeasibility. It would reduce power by 22.3%, but DES places its throughput at 19.06 jobs/h, below the nominal target of 20.42 jobs/h, so the analytical procedure selects the slightly higher-power configuration that preserves the production requirement. Read through specific energy, this lower-power boundary is in fact less efficient than the all-on system: at 38.32 kW over 19.06 jobs/h it consumes about 2.01 kWh/job, above the 1.81 kWh/job of the all-on baseline, because its power saving is bought entirely with lost output. The same vector also shows that aggregate power is governed by utilisation rather than by machine count: $[2, 4, 3, 1, 1]$ keeps more machines active than the selected $[1, 3, 3, 1, 2]$, eleven against ten, with extra units at the first two stations, yet draws less total power, since the lightly loaded units it adds contribute less to $P_{\text{tot}}(\mathbf{x})$ than the more heavily loaded station-five unit it switches off. Counting active machines is therefore not a proxy for power, and *more machines on* need not mean *more power*.

The selected vector remains unchanged under the three groups of energy-model sensitivity tests summarised in Table 17. Independent $\pm 20\%$ perturbations of station utilisation and power parameters leave the selected configuration unchanged in all 1000 Monte Carlo draws. Concave, affine, convex, and start-up-penalised energy functions also select the same vector,

Table 16. Manufacturing-inspired case: analytical selection and DES verification with $\delta = 0.25$ and $\varepsilon = 0.02$.

Quantity	All-on	Selected	Boundary infeasible
Active-machine vector	[2, 4, 3, 2, 3]	[1, 3, 3, 1, 2]	[2, 4, 3, 1, 1]
Power (kW)	49.29	39.40	38.32
Power reduction (%)	–	20.1	22.3
EMDA throughput (jobs/h)	27.22	22.89	19.07
DES throughput (jobs/h)	27.25	22.90	19.06
DES 95% CI (jobs/h)	[27.22, 27.28]	[22.87, 22.93]	[19.02, 19.10]
DES target-feasible	yes	yes	no
Specific energy (kWh/job)	1.81	1.72	2.01
Energy saving over 8 h (kWh)	–	79.1	87.8

Table 17. Summary of energy-model sensitivity in the manufacturing-inspired case.

Sensitivity family	Tested variation	Selected configuration	Power reduction
Station utilisation	Independent factors in [0.8, 1.2]	[1, 3, 3, 1, 2] in 100% of draws	18.28–22.17%
Idle and processing power	Independent factors in [0.8, 1.2]	[1, 3, 3, 1, 2] in 100% of draws	19.18–21.10%
Energy-function shape	Concave, affine, convex, and start-up variants	[1, 3, 3, 1, 2] in all variants	17.0–22.4%

although the numerical estimate of the saving varies with the objective shape.

The sensitivity analyses indicate that the manufacturing-inspired selection is driven primarily by the capacity–throughput structure rather than by a narrowly calibrated energy parameter or a specific affine load-to-power curve. The estimated magnitude of the saving remains model-dependent, while the selected active-capacity vector is stable across the tested perturbations. Detailed sensitivity tables are reported in Appendix 3.

8. Discussion and practical implications

8.1. Reliability of the analytical decision signals

The results distinguish clearly between the outputs that support the final capacity decision and those that primarily guide the search. Total throughput is the most accurately reproduced measure across the structured and random-instance experiments. Its low average error and strong rank correlation support the use of the evaluator for comparing candidate capacity vectors and identifying configurations that merit further consideration.

Feasibility classification is a more demanding task than pointwise throughput prediction because it depends on the position of each configuration relative to a fixed threshold. A throughput error that is small in percentage terms can still change the classification of a configuration located immediately above or below that boundary. This effect explains the few false-positive and false-negative cases observed in the experiments, as well as the weaker recall in the deterministic boundary scenario. The main source of decision uncertainty is therefore not broad prediction error across the configuration space, but local ambiguity around the throughput target.

The robustness margin ε provides a direct way to manage this boundary risk during analytical optimisation.

Increasing the analytical target reduces the likelihood of selecting a configuration that is only marginally feasible, at the cost of potentially rejecting some energy-saving alternatives. The appropriate margin depends on the accuracy observed for the application under study and on the operational consequences of a throughput shortfall. The default value used in the experiments reflects the error level observed in the validation set rather than a universal setting.

The configuration-specific reconstruction acts as a complementary safeguard at this boundary. Because it re-weights the station class mixes for every candidate vector, it keeps the feasibility estimate consistent with the configuration under evaluation and avoids the over-optimistic labels that a frozen aggregation can assign to marginally feasible configurations, as reported in Section 7.2. Boundary risk is therefore addressed from two complementary directions: the robustness margin raises the analytical target, while the configuration-specific reconstruction keeps the feasibility estimate aligned with the candidate being evaluated.

Station-level clipped utilisation, waiting-time, and congestion estimates serve a different role. Their errors are larger than those observed for aggregate throughput, particularly in highly congested configurations. The clipped utilisation u_i^{clip} enters the bounded power calculation, whereas the raw utilisation u_i^{raw} and the mean station population EN_i guide the beam-expansion heuristic and support overload diagnostics. None of these station-level quantities independently certifies throughput feasibility. Every candidate generated by the search is re-evaluated through the complete fixed-point procedure, and the final decision is based on network-level throughput and total power.

The manufacturing-inspired sensitivity analyses provide additional evidence on the practical effect of these secondary-metric errors. Perturbing station utilisation

and power parameters changes the estimated magnitude of the saving but does not change the selected capacity vector in the tested case. Similarly, alternative concave, affine, convex, and start-up-penalised energy functions preserve the selected configuration. This stability indicates that the decision is driven mainly by the capacity–throughput structure rather than by a finely tuned station-level estimate.

The evidence supports a layered interpretation of the evaluator. It is strongest as a rapid throughput-screening and ranking model, while local waiting and congestion outputs are most useful as directional information for search. Configurations with a clear analytical margin can be compared directly, whereas selected, near-tie, and near-threshold configurations benefit from DES verification before implementation.

8.2. Operational use and deployment guidance

The framework is intended for production planners, process engineers, and energy managers who must determine before a shift or production day which machines should remain active. Its decision horizon is tactical: the output is an active-capacity plan for a defined operating period, not a sequence of real-time switch-on and switch-off commands.

The required inputs are close to standard production-engineering data: class-dependent routings or visit ratios, mean processing times, installed machine counts, planned class populations or expected product mix, the required throughput level, and station-level idle and processing power. Optional utilisation limits and machine-availability constraints can be added when particular resources require additional protection.

The managerial deployment of the proposed model is structured as a sequence of planning, optimisation, validation, and control activities. As summarised in Table 18, the model converts production requirements and expected product-mix conditions into an implementable active-capacity decision, while preserving explicit throughput and robustness safeguards.

Table 18 also clarifies the respective roles of the different decision-support layers. The analytical evaluator

supports rapid tactical screening and capacity selection, DES or a detailed plant model is used to validate configurations close to the feasibility boundary, and shop-floor monitoring determines whether the selected plan can be maintained during execution. The resulting decision is therefore not an isolated optimisation output, but part of a closed-loop managerial process linking production planning, energy management, validation, and operational control.

Routine use does not require complete recalibration before every planning cycle. The quantities most likely to change from one decision period to another are the expected mix, the class populations, and the throughput target. Routing, processing-time, and power parameters can remain fixed while the production system is stable, and should be updated after process changes, machine replacement, maintenance interventions, new products, altered standby modes, or persistent differences between observed and predicted performance.

The energy parameters can be identified from short monitoring campaigns, machine documentation, or existing manufacturing information systems. Idle power should represent the average demand of an active but non-processing machine, while processing power should reflect representative loaded operation. The absolute saving estimate depends on this calibration, but the sensitivity analysis suggests that moderate parameter uncertainty may leave the selected capacity vector unchanged when the alternatives are well separated.

Product-mix uncertainty is the main operational source of model risk. A change in the realised mix may move congestion to another station and make a reduced-capacity vector less suitable than expected. When the mix forecast is uncertain, the same evaluator can be run over several plausible class-population scenarios. A robust operating choice can then be selected from the configurations that remain feasible across those scenarios, or a larger value of ε can be used to preserve additional throughput headroom.

During implementation, realised throughput and station utilisation should be monitored against the planned values. A safer configuration, including the all-on vector, can be retained as a recovery option when the realised

Table 18. Practical application of the proposed model.

Step	Managerial action	Decision output
1	Define the expected product mix, production volume, and machine availability for the planning period.	Planning scenario and class-population vector.
2	Evaluate the all-on configuration and determine the reference throughput.	Performance baseline and recovery configuration.
3	Set the admissible throughput loss δ and robustness margin ε .	Minimum acceptable throughput and safety margin.
4	Run the optimisation and select the lowest-power feasible machine-capacity vector.	Recommended number of active machines at each station.
5	Validate the selected configuration using DES and monitor throughput and utilisation during operation.	Approved operating plan or activation of a safer configuration.

product mix or operating conditions move outside the planned scenario. This creates a practical hierarchy in which the analytical model supports rapid planning, DES verifies the most consequential alternatives, and shop-floor monitoring provides operational feedback.

8.3. Limitations and future research

The framework is developed for two-class closed CONWIP systems with identical parallel machines, FCFS service, non-blocking flow, and shift-level capacity decisions. These assumptions define a coherent decision setting and also indicate the extensions required for broader application.

The first extension concerns the number and heterogeneity of product classes. The notation can be generalised to more than two classes, but the dimension of the throughput fixed point and the diversity of local station mixes increase with the number of classes. The resulting aggregation error and convergence behaviour should therefore be evaluated separately rather than inferred directly from the two-class results.

The second extension concerns machine heterogeneity. When machines at the same station differ in speed, eligibility, reliability, or power consumption, the scalar variable x_i no longer identifies the available capacity unambiguously. A machine-level decision model would need to specify which individual resources are activated and which product classes they can process. The same issue arises under strong product-machine compatibility constraints.

The third extension concerns operating rules and material-flow constraints. Priority dispatching, sequence-dependent setups, batching, rework policies, and finite-buffer blocking can alter the distribution of waiting time and the throughput response to added capacity. Blocking after service and blocking before service, in particular, introduce dependence on downstream buffer states and require an analytical representation beyond the non-blocking closed network used here.

The service model also assumes that mean processing requirements are independent of station load. In settings where interference, operator sharing, thermal conditions, batching, or coordination effects produce load-dependent service rates, the equivalent service parameters would need to become functions of utilisation and active capacity. This would add a further closure relation to the fixed-point procedure.

The power model represents average steady-state demand over the planning horizon. Applications involving frequent within-shift switching would require explicit treatment of start-up energy, warm-up time, minimum

on/off durations, machine-state transitions, and possibly time-varying electricity prices. These features would change the problem from tactical capacity availability to dynamic machine-state control.

Finally, the empirical evidence combines structured scenarios, random instances, and a manufacturing-inspired case, but does not constitute field validation on an operating production line. Industrial deployment would require calibration against measured throughput and power data and comparison with observed capacity decisions over repeated planning periods.

Future research can therefore proceed along four main directions: extending the aggregation to larger multiclass systems, introducing machine-level eligibility and heterogeneous resources, incorporating dynamic and measured energy behaviour, and validating the framework within industrial digital-twin or production-planning environments.

9. Conclusions

This paper develops a configuration-specific analytical evaluator for energy-aware capacity reconfiguration in two-class CONWIP systems. The evaluator combines a station-mix fixed point with an equivalent single-class MDA representation, reconstructing the analytical network for each candidate active-machine vector. This configuration-specific closure links class-throughput estimates, local arrival mixes, equivalent service parameters, and active station capacity within a common evaluation procedure. The resulting model is embedded in an energy-aware optimisation framework that compares alternative machine-activation plans under a throughput requirement. Within this construction the single-class MDA population recursion serves as an established computational engine and is employed unchanged, so that the methodological contribution resides in the configuration-specific reconstruction and fixed-point closure that render the equivalent network consistent with each candidate active-machine vector, and in their integration into the coupled evaluator, energy-aware optimisation, and discrete-event validation framework.

The empirical results support the use of the evaluator for rapid configuration comparison. Across the structured exponential battery, throughput MAPE is 0.96% and Spearman rank correlation is $\rho = 0.986$. Beyond the close descriptive agreement on the structured battery, the configuration-specific station-mix update is decision-relevant at the deployment level. At the operational tolerance $\delta = 0.25$, it changes the selected configuration in roughly one realistic random instance in five within the exhaustively enumerated $m \in \{3, 4\}$ divergence pool.

Among the 39 divergent instances adjudicated by DES, no frozen rule selected a better configuration; in 24 cases, a frozen rule selected a configuration that DES classified as infeasible. On the 30-instance Campaign-B DES-validation subset, overall throughput MAPE is 0.913%, precision is 0.982, and the false-positive rate is 0.067, with no false-positive configurations in the medium and large groups.

The manufacturing-inspired case illustrates the decision value of the framework. The selected configuration reduces estimated power by 20.1% while remaining feasible under DES, whereas a lower-power boundary configuration is correctly rejected because its simulated throughput lies below the production target. The sensitivity analyses also show that the selected vector remains stable under moderate perturbations of utilisation, power parameters, and energy-function shape. These results position the evaluator as a tactical screening and ranking tool: aggregate throughput provides the primary feasibility signal, while station-level utilisation and congestion measures support power estimation and search guidance. DES can then be concentrated on selected, near-tie, and near-threshold configurations.

The framework is intended for shift-level planning in two-class, non-blocking CONWIP systems with interchangeable parallel machines and a reasonably stable expected mix. Extensions to larger multi-class systems, heterogeneous machine eligibility, priority and setup-dependent operation, finite-buffer blocking, load-dependent processing rates, and dynamic switching would broaden its range of application. Further research should also combine the analytical evaluator with measured energy data and industrial digital-twin or production-planning environments to assess its performance under repeated field use.

Note

1. An extreme stress instance ($m = 3$, seed 12029, index 22) illustrates the same mechanism: the frozen-at-init mix conceals an approximately 42% throughput shortfall, selecting [1, 1, 1] with DES throughput 0.182 against a target of 0.314, whereas the dynamic and Static-all-on rules select [1, 2, 1] with DES throughput 0.369.

Author contributions

CRediT: **Maria De Martino**: Conceptualization, Investigation, Methodology, Validation, Writing – original draft; **Silvestro Vespoli**: Conceptualization, Methodology, Validation, Writing – review & editing; **Andrea Grassi**: Conceptualization, Methodology, Supervision, Writing – review & editing; **Guido Guizzi**: Conceptualization, Methodology, Supervision, Writing – review & editing

Disclosure statement

No potential conflict of interest was reported by the authors.

Notes on contributors



Maria De Martino is a PhD Student in Technology, Innovation and Management at the University of Bergamo and the University of Naples Federico II, Italy. Her research focuses on production planning and control, with particular interest in stochastic modelling, decentralised decision-making, and energy-aware manufacturing. Her work combines analytical models, simulation, and optimisation to support efficient and flexible production-system design and control.

Her e-mail addresses are maria.demartino13@unina.it and maria.demartino@unibg.it.



Silvestro Vespoli is an Assistant Professor in Industrial Mechanical Systems at the University of Naples Federico II, Italy. His research focuses on developing innovative methodologies and models for designing, managing, and controlling production systems within the digital factory (Industry 4.0) framework, to support sustainable processes and infrastructure design for both production and services in industrialised contexts. His research approach combines quantitative and interdisciplinary techniques, encompassing control models, data science, artificial intelligence, and forecasting methods. The results of his research activity have been published in top-ranked international journals and presented at major international conferences. His e-mail address is silvestro.vespoli@unina.it.



Andrea Grassi is a Full Professor at the Department of Chemical, Materials and Production Engineering at the University of Naples Federico II, Italy. His current research interests are in the areas of design and management of manufacturing systems and supply chains, including the development of stochastic analytical models, simulation modelling, performance analysis and prediction, Industry 4.0 distributed control architectures, and automation of manufacturing systems. Results of his research activity have been published in top-ranked international journals and in international conferences. He has been involved in several projects funded by companies and in Italian and European projects. His e-mail address is andrea.grassi@unina.it.



Guido Guizzi is an Associate Professor in the Department of Chemical, Materials and Production Engineering at the University of Naples Federico II. He has the title of Aggregate Professor in 2012. His current research interests include: modelling and simulation of stochastic systems, Supply Chain Management, and Operations Management, Logistics, Six Sigma. He obtained his PhD in Aerospace Engineering, Naval and Quality (University

of Naples Federico II – Italy) in 2006. He has successfully collaborated with industry and academia, authoring more than 60 peer-reviewed research publications. He has been involved in several projects funded by companies. His email address is g.guizzi@unina.it.

ORCID

Maria De Martino  <https://orcid.org/0009-0004-8729-4773>

Silvestro Vespoli  <http://orcid.org/0000-0003-2042-2668>

Andrea Grassi  <http://orcid.org/0000-0003-4638-9069>

Guido Guizzi  <https://orcid.org/0000-0002-3269-1307>

Use of Generative Artificial Intelligence (AI)

Generative AI tools (ChatGPT- 5.5, Claude Opus 4.5) were used exclusively for language polishing during the final proofreading stage. All technical content, methodology, code, and analysis were produced entirely by the authors, who reviewed and take full responsibility for every part of the manuscript.

Data availability statement

The data supporting the findings of this study are available in the supplementary reproducibility package accompanying the article. The package includes the input data, instance definitions, analytical and simulation outputs, optimisation results, sensitivity analyses, simulation seeds, methodological pseudocode, and documentation required to trace the reported tables and figures to the corresponding data files. The structured scenarios, the two random-instance campaigns, the DES-validation experiments, and the manufacturing-inspired case are reported separately according to their respective experimental purposes. The supplementary material additionally provides the executable source code required for end-to-end reproduction: the instance generators for both random-instance campaigns, the configuration-specific EMDA evaluator, the search procedures (beam search, the GA fallback, greedy best-first search, and top-down hill climbing), and the DES validation code. Together with the instance definitions, the simulation seeds, and the table-to-file map, these files let an independent researcher regenerate the instances, recompute the analytical evaluations, rerun every search procedure, and reproduce the DES validation underlying each reported table and figure.

References

- Akyildiz, Ian F., and Gilbert Bolch. 1988. "Mean Value Analysis Approximation for Multiple Server Queueing Networks." *Performance Evaluation* 8 (2): 77–91. [https://doi.org/10.1016/0166-5316\(88\)90015-6](https://doi.org/10.1016/0166-5316(88)90015-6).
- Amjath, Mohamed, Laoucine Kerbache, Adel Elomri, and James MacGregor Smith. 2024. "Queueing Network Models for the Analysis and Optimisation of Material Handling Systems: A Systematic Literature Review." *Flexible Services and Manufacturing Journal* 36 (2): 668–709. <https://doi.org/10.1007/s10696-023-09505-x>.
- Aramcharoen, Ampara, and Paul T. Mativenga. 2014. "Critical Factors in Energy Demand Modelling for CNC Milling and Impact of Toolpath Strategy." *Journal of Cleaner Production* 78:63–74. <https://doi.org/10.1016/j.jclepro.2014.04.065>.
- Avram, Ovidiu Iulian, and Paul Xirouchakis. 2011. "Evaluating the Use Phase Energy Requirements of a Machine Tool System." *Journal of Cleaner Production* 19 (6–7): 699–711. <https://doi.org/10.1016/j.jclepro.2010.10.010>.
- Baskett, Forest, K. Mani Chandy, Richard R. Muntz, and Fernando G. Palacios. 1975. "Open, Closed, and Mixed Networks of Queues with Different Classes of Customers." *Journal of the ACM* 22 (2): 248–260. <https://doi.org/10.1145/321879.321887>.
- Battaia, Olga, Lyes Benyoucef, Xavier Delorme, Alexandre Dolgui, and Simon Thevenin. 2020. "Sustainable and Energy Efficient Reconfigurable Manufacturing Systems." In *Reconfigurable Manufacturing Systems: From Design to Implementation*, edited by Benyoucef Lyes, 179–191. Cham: Springer. https://doi.org/10.1007/978-3-030-28782-5_9.
- Baynat, B., and Y. Dallery. 1996. "A Product-Form Approximation Method for General Closed Queueing Networks with Several Classes of Customers." *Performance Evaluation* 24 (3): 165–188. [https://doi.org/10.1016/0166-5316\(94\)00028-X](https://doi.org/10.1016/0166-5316(94)00028-X).
- Baynat, Bruno, John A. Buzacott, and Yves Dallery. 2002. "Multiproduct Kanban-Like Control Systems." *International Journal of Production Research* 40 (16): 4225–4255. <https://doi.org/10.1080/00207540210146198>.
- Behrendt, T., A. Zein, and S. Min. 2012. "Development of an Energy Consumption Monitoring Procedure for Machine Tools." *CIRP Annals – Manufacturing Technology* 61 (1): 43–46. <https://doi.org/10.1016/j.cirp.2012.03.103>.
- Bitran, Gabriel R., and Sriram Dasu. 1992. "A Review of Open Queueing Network Models of Manufacturing Systems." *Queueing Systems* 12 (1–2): 95–133. <https://doi.org/10.1007/BF01158637>.
- Buzacott, John A., and J. George Shanthikumar. 1993. *Stochastic Models of Manufacturing Systems*. NJ: Prentice Hall: Englewood Cliffs.
- Buzen, Jeffrey P. 1973. "Computational Algorithms for Closed Queueing Networks with Exponential Servers." *Communications of the ACM* 16 (9): 527–531. <https://doi.org/10.1145/362342.362345>.
- Dallery, Yves, and Stecke E. Kathryn. 1990. "On the Optimal Allocation of Servers and Workloads in Closed Queueing Networks." *Operations Research* 38 (4): 694–703. <https://doi.org/10.1287/opre.38.4.694>.
- De Martino, Maria, Andrea Grassi, Guido Guizzi, and Valentina Popolo. 2025. "Optimising Energy Consumption in Reconfigurable Manufacturing Systems: A Mathematical Approach." *IFAC-PapersOnLine* 59 (10): 1970–1974. <https://doi.org/10.1016/j.ifacol.2025.09.331>.
- Dieleman, Nanne A., Joost Berkhout, and Bernd F. Heidergott. 2023. "A Neural Network Approach to Performance Analysis of Tandem Lines: The Value of Analytical Knowledge." *Computers & Operations Research* 152:106124. <https://doi.org/10.1016/j.cor.2022.106124>.
- Framinan, José M., Pedro L. González, and Rafael Ruiz-Usano. 2003. "The CONWIP Production Control System: Review and Research Issues." *Production Planning & Control* 14 (3): 255–265. <https://doi.org/10.1080/0953728031000102595>.

- Frigerio, Nicla, and Andrea Matta. 2015. "Energy-Efficient Control Strategies for Machine Tools with Stochastic Arrivals." *IEEE Transactions on Automation Science and Engineering* 12 (1): 50–61. <https://doi.org/10.1109/TASE.2014.2344507>.
- Frigerio, Nicla, Barış Tan, and Andrea Matta. 2024. "Simultaneous Control of Multiple Machines for Energy Efficiency: A Simulation-Based Approach." *International Journal of Production Research* 62 (3): 933–948. <https://doi.org/10.1080/00207543.2023.2175175>.
- Gahm, Christian, Florian Denz, Markus Dirr, and Axel Tuma. 2016. "Energy-Efficient Scheduling in Manufacturing Companies: A Review and Research Framework." *European Journal of Operational Research* 248 (3): 744–757. <https://doi.org/10.1016/j.ejor.2015.07.017>.
- Gordon, William J., and Gordon F. Newell. 1967. "Closed Queuing Systems with Exponential Servers." *Operations Research* 15 (2): 254–265. <https://doi.org/10.1287/opre.15.2.254>.
- Herer, Yale T., and Michael Masin. 1997. "Mathematical Programming Formulation of CONWIP-based Production Lines and Relationships to MRP." *International Journal of Production Research* 35 (4): 1067–1076. <https://doi.org/10.1080/002075497195542>.
- Hopp, Wallace J., and Mark L. Roof. 1998. "Setting WIP Levels with Statistical Throughput Control (STC) in CONWIP Production Lines." *International Journal of Production Research* 36 (4): 867–882. <https://doi.org/10.1080/002075498193435>.
- Inman, R. R. 1999. "Empirical Evaluation of Exponential and Independence Assumptions in Queueing Models of Manufacturing Systems." *Production and Operations Management* 8 (4): 409–432. <https://doi.org/10.1111/j.1937-5956.1999.tb00316.x>.
- ISO 14955-1:2017. 2017. *Machine Tools – Environmental Evaluation of Machine Tools – Part 1: Design Methodology for Energy-Efficient Machine Tools*.
- ISO 14955-2:2018. 2018. *Machine Tools – Environmental Evaluation of Machine Tools – Part 2: Methods for Measuring Energy Supplied to Machine Tools and Machine Tool Components*.
- Jaegler, Yann, Anicia Jaegler, Patrick Burlat, Samir Lamouri, and Damien Trentesaux. 2018. "The ConWip Production Control System: A Systematic Review and Classification." *International Journal of Production Research* 56 (17): 5736–5757. <https://doi.org/10.1080/00207543.2017.1380325>.
- Jia, Zhiyang, Liang Zhang, Jorge Arinez, and Guoxian Xiao. 2016. "Performance Analysis for Serial Production Lines with Bernoulli Machines and Real-Time WIP-based Machine Switch-on/off Control." *International Journal of Production Research* 54 (21): 6285–6301. <https://doi.org/10.1080/00207543.2016.1197438>.
- Kara, Sami, and W. Li. 2011. "Unit Process Energy Consumption Models for Material Removal Processes." *CIRP Annals – Manufacturing Technology* 60 (1): 37–40. <https://doi.org/10.1016/j.cirp.2011.03.018>.
- Loffredo, Antonio, Nicla Frigerio, Ettore Lanzarone, and Andrea Matta. 2024. "Energy-Efficient Control in Multi-stage Production Lines with Parallel Machine Workstations and Production Constraints." *IIE Transactions* 56 (1): 69–83. <https://doi.org/10.1080/24725854.2023.2168321>.
- Loffredo, Antonio, Marvin Carl May, Lukas Schäfer, Andrea Matta, and Gisela Lanza. 2023. "Reinforcement Learning for Energy-Efficient Control of Parallel and Identical Machines." *CIRP Journal of Manufacturing Science and Technology* 44:91–103. <https://doi.org/10.1016/j.cirpj.2023.05.007>.
- Lv, Jingxiang, Renzhong Tang, Shun Jia, and Ying Liu. 2016. "Experimental Study on Energy Consumption of Computer Numerical Control Machine Tools." *Journal of Cleaner Production* 112:3864–3874. <https://doi.org/10.1016/j.jclepro.2015.07.040>.
- Mashaei, Masoumeh, and Bengt Lennartson. 2013. "Energy Reduction in a Pallet-Constrained Flow Shop through on/off Control of Idle Machines." *IEEE Transactions on Automation Science and Engineering* 10 (1): 45–56. <https://doi.org/10.1109/TASE.2012.2225426>.
- Mouzon, Gilles, Mehmet B. Yildirim, and James Twomey. 2007. "Operational Methods for Minimization of Energy Consumption of Manufacturing Equipment." *International Journal of Production Research* 45 (18–19): 4247–4271. <https://doi.org/10.1080/00207540701450013>.
- Panzer, Marcel, and Benedict Bender. 2022. "Deep Reinforcement Learning in Production Systems: A Systematic Literature Review." *International Journal of Production Research* 60 (13): 4316–4341. <https://doi.org/10.1080/00207543.2021.1973138>.
- Reiser, Martin, and Stephen S. Lavenberg. 1980. "Mean-Value Analysis of Closed Multichain Queueing Networks." *Journal of the ACM* 27 (2): 313–322. <https://doi.org/10.1145/322186.322195>.
- Roy, Debjit. 2016. "Semi-open Queueing Networks: A Review of Stochastic Models, Solution Methods and New Research Areas." *International Journal of Production Research* 54 (6): 1735–1752. <https://doi.org/10.1080/00207543.2015.1056316>.
- Ryan, M. Sarah, and Jumpol Vorasayan. 2005. "Allocating Work in Process in a Multiple-Product CONWIP System with Lost Sales." *International Journal of Production Research* 43 (2): 223–246. <https://doi.org/10.1080/0020754042000268875>.
- Ryan, Sarah M., Bruno Baynat, and Fred F. Choobineh. 2000. "Determining Inventory Levels in a CONWIP Controlled Job Shop." *IIE Transactions* 32 (2): 105–114.
- Ryan, Sarah M., and Fred Choobineh. 2003. "Total WIP and WIP Mix for a CONWIP Controlled Job Shop." *IIE Transactions* 35 (5): 405–418. <https://doi.org/10.1080/07408170304387>.
- Shanthikumar, J. George, and David D. Yao. 1988. "On Server Allocation in Multiple Center Manufacturing Systems." *Operations Research* 36 (2): 333–342. <https://doi.org/10.1287/opre.36.2.333>.
- Spearman, Mark L., Wallace J. Hopp, and David L. Woodruff. 1992. "Push and Pull Production Systems: Issues and Comparisons." *Operations Research* 40 (3): 521–532. <https://doi.org/10.1287/opre.40.3.521>.
- Spearman, Mark L., David L. Woodruff, and Wallace J. Hopp. 1990. "CONWIP: a Pull Alternative to Kanban." *International Journal of Production Research* 28 (5): 879–894. <https://doi.org/10.1080/00207549008942761>.
- Spearman, Mark L., David L. Woodruff, and Wallace J. Hopp. 2022. "CONWIP Redux: Reflections on 30 Years of Development and Implementation." *International Journal of Production Research* 60 (1): 381–387. <https://doi.org/10.1080/00207543.2021.1954713>.
- Vijayaraghavan, Athulan, and David Dornfeld. 2010. "Automated Energy Monitoring of Machine Tools." *CIRP Annals –*

Manufacturing Technology 59 (1): 21–24. <https://doi.org/10.1016/j.cirp.2010.03.042>.

Appendices

Appendix 1. Notation and single-class MDA details

A.1. Main notation

Table A1 consolidates the notation used in the system formulation, analytical evaluator, power model, and optimisation framework. Symbols introduced only in individual experiments are defined locally.

A.2. Single-class MDA routine

The configuration-specific evaluator constructs an equivalent single-class closed network defined by the total population N , aggregate visit ratios v_i^{agg} , effective service rates μ_i^{eff} , and active server counts x_i . This network is evaluated through the population-recursive marginal-distribution routine reported in Algorithm 3.

At population level l , the routine calculates the station residence times, system throughput, mean station populations, and marginal station-population probabilities. The active capacity

enters through

$$r_i(q) = \min\{x_i, q\},$$

which represents the number of busy servers when q jobs are present at station i . The installed capacity $c_{i,\text{max}}$ determines the admissible range of x_i but does not replace the active server count in the network evaluation.

The MDA recursion is used without modification as the computational core of the equivalent-network evaluation. The configuration-specific contribution developed in the main text lies in rebuilding the quantities v_i^{agg} and μ_i^{eff} within the two-class fixed point for each candidate active-machine vector.

Appendix 2. Reproducibility details

This appendix consolidates the structured-scenario parameters, random-instance generation rules, numerical settings, simulation protocol, and confidence-interval calculation used in the experimental study.

A.3. Structured-scenario parameters

Table A2 reports the class populations, visit ratios, mean processing times, and installed capacities used in the structured scenarios. A service-time entry associated with $V_{ir} = 0$ is not used in the analytical or simulation model. The deterministic and distributional stress experiments retain the same service-time means and modify only the distribution used by DES.

Table A1. Main notation used in the paper.

Symbol	Meaning
m	Number of stations.
i	Station index, $i = 1, \dots, m$.
r	Product-class index, $r \in \{0, 1\}$.
$\mathbf{K} = (K_0, K_1)$	Class-population vector in the closed CONWIP network.
$N = K_0 + K_1$	Total circulating CONWIP population.
V_{ir}	Visit ratio of class r at station i .
S_{ir}	Mean processing time per visit of class r at station i .
$c_{i,\text{max}}$	Number of identical parallel machines installed at station i .
x_i	Number of machines kept active at station i in the candidate configuration.
$\mathbf{x}$	Active-machine configuration vector.
h_r	Estimated throughput of class r .
TH_{emda}	Total throughput implied by the two-class fixed point; used for analytical feasibility and optimisation.
TH_{mda}	Throughput returned by the equivalent single-class MDA recursion; used as an internal consistency quantity.
TH_{DES}	Throughput estimated through discrete-event simulation.
$TH_{\text{all}}^{\text{emda}}$	Analytical throughput of the all-on configuration.
$TH_{\text{all}}^{\text{DES}}$	DES throughput of the all-on configuration.
λ_i	Aggregate arrival rate at station i implied by the current class-throughput estimates.
p_{ir}	Arrival proportion of class r at station i .
v_i^{agg}	Visit ratio of station i in the equivalent single-class network.
$\bar{S}_i$	Equivalent mean service time at station i .
μ_i^{eff}	Effective service rate at station i , equal to $1/\bar{S}_i$.
ET_i	Mean station sojourn time returned by the single-class MDA recursion.
EN_i	Mean number of jobs at station i .
$W_i^{(q)}$	Approximate queue-waiting component at station i .
CT_r	Estimated cycle time of class r .
u_i^{raw}	Unbounded utilisation implied for the active parallel capacity at station i ; used for utilisation constraints, overload diagnostics, and search guidance.
u_i^{clip}	Bounded utilisation $\min\{\max\{u_i^{\text{raw}}, 0\}, 1\}$; used for reporting and steady-state power estimation.
$P_{\text{idle},i}$	Idle or standby power of one active machine at station i .
$P_{\text{proc},i}$	Power of one active machine at station i under full processing load.
P_{aux}	Fixed auxiliary power.
$P_{\text{tot}}(\mathbf{x})$	Total steady-state power of configuration $\mathbf{x}$.
δ	Maximum admissible throughput-loss fraction relative to the all-on configuration.
ε	Robustness margin applied to the analytical throughput requirement.
τ	Fixed-point convergence tolerance.
β	Fixed-point damping factor.
$g(\mathbf{x})$	Internal throughput-consistency gap between TH_{mda} and TH_{emda} .

Algorithm 3: Single-class MDA routine for the equivalent closed network

Input: Total population N ; aggregate visit ratios v_i^{agg} ; effective service rates μ_i^{eff} ; active server counts x_i , $i = 1, \dots, m$

Output: Equivalent-network throughput TH_{mda} ; mean station sojourn times ET_i ; mean station populations EN_i

1 Initialise

$$p_i^{(0)}(0) = 1, \quad p_i^{(0)}(q) = 0 \quad \text{for } q > 0,$$

for all stations i ;

2 for $l = 1, \dots, N$ do

3 for $i = 1, \dots, m$ do

4 Compute the station residence time:

$$ET_i(l) = \sum_{q=1}^l \frac{q}{\mu_i^{\text{eff}} \min\{x_i, q\}} p_i^{(l-1)}(q-1).$$

5 end

6 Compute the closed-network throughput:

$$TH(l) = \frac{l}{\sum_{i=1}^m v_i^{\text{agg}} ET_i(l)}.$$

for $i = 1, \dots, m$ do

7 Update the mean station population:

$$EN_i(l) = v_i^{\text{agg}} TH(l) ET_i(l).$$

Update the marginal station-population probabilities:

$$p_i^{(l)}(q) = \frac{v_i^{\text{agg}} TH(l) p_i^{(l-1)}(q-1)}{\mu_i^{\text{eff}} \min\{x_i, q\}},$$

$$q = 1, \dots, l,$$

and

$$p_i^{(l)}(0) = 1 - \sum_{q=1}^l p_i^{(l)}(q).$$

8 end

9 end

10 return

$$TH_{\text{mda}} = TH(N), \quad ET_i = ET_i(N), \quad EN_i = EN_i(N).$$

Table A2. Parameters of the structured scenarios.

Scenario	Population rule	Station	V_{i0}	V_{i1}	S_{i0}	S_{i1}	$c_{i,\text{max}}$
BASE-A	$\mathbf{K} = [14, 6]$	1	1	0	8	1	3
		2	0	1	1	10	1
		3	1	1	6	9	2
BASE-B	$\mathbf{K} = [21, 9]$	1	1	0	4	1	2
		2	1	1	6	9	1
		3	2	1	5	5	1
BASE-C	$\mathbf{K} = [19, 1]$	1	1	0	4	1	3
		2	0	1	1	30	1
		3	1	1	5	10	2
MIX	$N = 30, \alpha \in \{0.0, 0.1, \dots, 1.0\}$	1	1	1	8	2	2
		2	1	1	2	8	2
		3	1	1	5	5	1
NSWEEP	$N \in \{10, 15, 20, 30, 45\}, \alpha = 0.60$	1	1	1	3	10	2
		2	1	1	8	2	1
		3	1	1	5	5	2
DET	$N = 30, \alpha = 0.50$, deterministic DES	1	1	1	8	2	2
		2	1	1	2	8	2
		3	1	1	5	5	1

For the MIX and NSWEEP scenarios, class populations are calculated as

$$K_0 = \text{round}(\alpha N), \quad K_1 = N - K_0.$$

The distributional SCV sweep uses the MIX structure with $N = 30$ and $\alpha = 0.50$. The DES service-time distributions are deterministic, Erlang-4, Erlang-2, lognormal with SCV 0.5, exponential, and lognormal with SCV 1.5. All distributions preserve the means reported in Table A2.

For the structured capacity-scale experiments, scale S uses the installed capacities in Table A2, scale M doubles every station capacity, and scale L triples every station capacity. The corresponding power parameters are

$$P_{\text{idle}} = [1.8, 2.2, 1.5] \text{ kW}, \quad P_{\text{proc}} = [4.5, 5.0, 3.8] \text{ kW},$$

with $P_{\text{aux}} = 0.5 \text{ kW}$.

A.4. Random-instance generation

The random experiments use a common pool of 120 two-class instances, comprising 40 small, 40 medium, and 40 large instances. Small instances use seeds 501–540, medium instances use seeds 801–840, and large instances use seeds 1001–1040. The full pool is used for the optimisation benchmark.

For each instance, the class-probability vector is sampled from a symmetric Dirichlet distribution and the nominal population is allocated through a multinomial draw. A minimum class population of one is enforced by replacing any zero class count with one. Consequently, the realised population can exceed the nominal population by one in cases where the multinomial allocation initially assigns zero jobs to one class.

Visit ratios are generated independently as $V_{ir} = 1 + \text{Poisson}(1)$. Therefore, every class visits every station and visit ratios may exceed two. For DES, a deterministic canonical route is reconstructed by ordering stations by increasing index and repeating station i round(V_{ir}) times for class r .

A station-specific base processing time is sampled from $U(2, 12)$ min. Class heterogeneity is introduced through a log-normal multiplier, clipped to $[0.4, 3]$, and an independent station-class multiplier sampled from $U(0.8, 1.2)$. The resulting service time is bounded below by 0.5 min. Idle power is sampled from $U(1, 3)$ kW, processing power is obtained by adding an

Table A3. Campaign-B random-instance generation protocol used for the main optimisation benchmark, DES validation, and robustness-margin sensitivity analysis.

Quantity	Generation rule
Number of classes	Two classes, $r \in \{0, 1\}$.
Small group	$m = 5, c_{i,\max} \in \{1, 2, 3\}$, nominal population $N_0 = 3m$, seeds 501–540.
Medium group	$m = 8, c_{i,\max} \in \{1, 2, 3, 4\}$, nominal population $N_0 = 4m$, seeds 801–840.
Large group	$m = 10, c_{i,\max} \in \{1, 2, 3, 4\}$, nominal population $N_0 = 5m$, seeds 1001–1040.
Class populations	Symmetric Dirichlet class probabilities followed by a multinomial allocation of N_0 ; zero class counts are replaced by one.
Visit ratios	$V_{ir} = 1 + \text{Poisson}(1)$.
DES route	Stations ordered by increasing index; station i repeated round(V_{ir}) times in the route of class r .
Station base processing time	Uniform on [2, 12] min.
Class multiplier	Lognormal multiplier clipped to [0.4, 3].
Station–class multiplier	Uniform on [0.8, 1.2].
Minimum service time	0.5 min.
Idle power	$P_{\text{idle},i} \sim U(1, 3)$ kW.
Processing power	$P_{\text{proc},i} = P_{\text{idle},i} + U(1.5, 5)$ kW.
Auxiliary power	$P_{\text{aux}} = 0.3$ kW.
Benchmark pool	120 instances, with 40 instances per dimensional group.
DES-validation subset	Ten approximately equally spaced successful instances per dimensional group, selected after ordering by instance identifier.

increment sampled from $U(1.5, 5)$ kW, and the fixed auxiliary load is 0.3 kW.

The 30-instance DES subset is obtained after identifying benchmark instances with at least one successful optimisation result. Within each dimensional group, instances are ordered by identifier and ten approximately equally spaced positions are selected. Complete definitions of the 120-instance pool and the 30-instance DES subset are supplied in the supplementary reproducibility package.

Table A3 refers exclusively to Campaign B. Campaign A follows the separate generation protocol described above and is used only for the Beam–Greedy–Hill–Climbing baseline comparison.

The implementation was executed in Python 3.12 on a laptop equipped with an Intel Core i7-1265U processor and 16 GB RAM. Wall-clock results are therefore hardware-dependent, while the number of distinct EMDA evaluations provides a more portable measure of search effort.

For small, medium, and large random instances, the beam uses adaptive beam widths of 20, 12, and 8 and candidate-list sizes of 4, 3, and 2, respectively. Memoisation stores the result of every evaluated active-machine vector, and all evaluation counts include distinct calls to the analytical evaluator. Where a trimming procedure is applied, its additional evaluator calls are included unless a table explicitly states a different accounting convention.

A.5. Numerical, simulation, and search settings

Table A4 consolidates the numerical, simulation, and search settings used throughout the experiments.

A.6. DES confidence intervals

For a configuration simulated over n independent replications, let TH_j denote the post-warm-up throughput of replication j .

Table A4. Consolidated numerical and experimental settings.

Component	Setting
EMDA fixed-point tolerance	$\tau = 10^{-7}$.
Maximum fixed-point iterations	300.
Fixed-point damping factor	$\beta = 0.5$.
Evaluator-usability criterion	Fixed-point convergence and internal throughput-consistency gap $g(\mathbf{x}) \leq 0.05$.
Structured-validation DES replications	30 independent replications per configuration.
Standard structured DES horizon	200,000 min.
Random-validation DES replications	20 independent replications per configuration.
Standard structured warm-up	20,000 min.
Re-entrant or highly congested horizon	300,000 min.
Re-entrant or highly congested warm-up	30,000 min.
Manufacturing-inspired DES replications	20 independent replications.
manufacturing-inspired DES horizon	200,000 min.
manufacturing-inspired DES warm-up	20,000 min.
DES seed policy	Independent replication seeds beginning from 1.
Main throughput tolerance	$\delta = 0.25$, unless otherwise stated.
Default robustness margin	$\varepsilon = 0.02$, unless otherwise stated.
Beam settings	Beam width and candidate-list size adapted to instance scale.
Baseline GA	Population 60; 150 generations; elite size 5.
GA probabilities	Crossover probability 0.70; mutation probability 0.40.
GA benchmark seeds	1, 2, and 3.
GA sensitivity seeds	1, 2, 3, 4, and 5.
Exact-enumeration benchmark	Used when the state space is sufficiently small; the random benchmark uses a 500,000-configuration threshold.
Confidence level	Student-t 95% confidence intervals across replications.

The replication mean and sample standard deviation are

$$\overline{TH}_{DES} = \frac{1}{n} \sum_{j=1}^n TH_j,$$

and

$$s_{TH} = \sqrt{\frac{\sum_{j=1}^n (TH_j - \overline{TH}_{DES})^2}{n-1}}.$$

The 95% confidence interval is calculated as

$$\overline{TH}_{DES} \pm t_{0.975, n-1} \frac{s_{TH}}{\sqrt{n}}, \quad (A1)$$

where $t_{0.975, n-1}$ is the corresponding Student- t critical value.

Mean-based DES feasibility compares $\overline{TH}_{DES}$ with the instance-specific DES target. Confidence-interval feasibility uses the lower bound of Equation (A1), providing a more conservative assessment for selected and near-boundary configurations.

The achieved sampling precision confirms that the replication counts are adequate for the feasibility decision. Over the $n = 146$ random-instance DES configurations evaluated with 20 replications, the half-width of Equation (A1) has a median of 0.29% and a worst case of 0.56% of the mean throughput; over a dedicated uniform-30-replication structured re-simulation of $n = 24$ groups—distinct from the mixed-replication 24-configuration end-to-end verification of Table 11, whose random part uses 20 replications – the corresponding values are 0.20% and 0.42%. In every case the half-width is at least about forty-five times smaller than the operational tolerance $\delta = 0.25$, and close to two orders of magnitude smaller at the median, so the reduced replication count preserves both the mean-based and the confidence-bound feasibility classification.

Appendix 3. Supplementary robustness and sensitivity analyses

This appendix reports the complete robustness analyses supporting the summary results presented in Sections 7.4 and 7.6. The analyses are organised around three sources of uncertainty: service-time distribution, analytical robustness margin, and energy-model specification.

A.7. Service-time distribution

Table A5 reports the complete distributional sweep for the MIX scenario with $\alpha = 0.5$, $N = 30$, and $\delta = 0.02$. The analytical evaluator uses the same mean service times in every case, while the DES service-time distribution varies from deterministic to lognormal with SCV 1.5.

The evaluator remains close to DES throughout the tested SCV range. The positive bias at low SCV indicates that deterministic and low-variability systems produce slightly higher DES throughput than the exponential analytical approximation.

Table A6 reports the separate deterministic boundary experiment. All four configurations satisfy the evaluator-usability condition, but three lie close enough to the throughput threshold for the distributional difference to change their feasibility classification.

Table A5. Throughput accuracy and feasibility classification under alternative service-time distributions.

Distribution	SCV	MAPE (%)	Bias (%)	Precision / Recall
Deterministic	0.00	0.50	+0.44	1.00 / 1.00
Erlang-4	0.25	0.34	+0.32	1.00 / 1.00
Erlang-2	0.50	0.30	+0.24	1.00 / 1.00
Lognormal	0.50	0.25	+0.17	1.00 / 1.00
Exponential	1.00	0.12	+0.10	1.00 / 1.00
Lognormal	1.50	0.10	−0.04	1.00 / 1.00

Note: MIX scenario with $\alpha = 0.5$, $N = 30$, and 30 DES replications per configuration.

Table A6. Feasibility classification under exponential and deterministic DES service times.

Configuration	Evaluator Throughput / feasibility	DES-EXP Throughput / feasibility	DES-DET Throughput / feasibility
[2, 2, 1]	12.00 ✓	12.01 ✓	12.00 ✓
[1, 2, 1]	11.60 ×	11.60 ×	12.00 ✓
[2, 1, 1]	11.60 ×	11.40 ×	11.77 ✓
[1, 1, 1]	11.25 ×	11.24 ×	12.00 ✓

Note: The throughput threshold is 11.76 jobs/h.

Table A7. Power reduction as a function of the robustness margin ε with $\delta = 0.25$.

Instance	$\varepsilon = 0.00$	$\varepsilon = 0.01$	$\varepsilon = 0.02$	$\varepsilon = 0.03$	$\varepsilon = 0.05$
BASE-A-S	9.0	9.0	9.0	9.0	9.0
BASE-A-M	29.4	29.4	25.1	18.4	18.4
BASE-A-L	30.7	30.7	30.7	28.3	28.3
BASE-B-S	13.4	13.4	13.4	13.4	13.4
BASE-B-M	20.5	20.5	20.5	20.5	20.5
BASE-B-L	38.3	38.3	33.2	23.1	23.1
BASE-C-S	8.6	8.6	8.6	8.6	8.6
BASE-C-M	29.7	22.3	22.3	22.3	22.3
BASE-C-L	34.6	34.6	34.6	34.6	34.6

Note: BASE-A instances use the GA fallback.

The deterministic experiment produces one true positive and three false negatives, with no false positives. The classification error is therefore conservative: the analytical approximation rejects configurations that deterministic DES accepts. This result confirms that distributional differences are most consequential when configurations cluster around the throughput boundary rather than when they are clearly feasible or infeasible.

A.8. Robustness-margin sensitivity

Table A7 reports the effect of the analytical robustness margin ε on the nine structured optimisation instances. Increasing ε raises the effective analytical throughput requirement and may therefore reduce the attainable power saving.

Six of the nine structured instances retain the same selected configuration across the full margin range. The remaining instances exhibit discrete changes in the selected capacity vector when the more conservative target excludes a lower-power boundary solution.

Table A8 reports the corresponding DES verification on a 15-instance random subset. Medium and large selected configurations remain DES-feasible across the tested values of ε . The persistent small-instance boundary case remains below the

Table A8. Sensitivity to ε on a dedicated subset of 15 random instances, with five instances per dimensional group.

ε	Group	Solver success	DES mean	DES 95% CI	Median saving (%)
0.00	Small	1.00	1.00	0.80	27.23
	Medium	1.00	1.00	1.00	38.35
	Large	1.00	1.00	1.00	32.01
0.02	Small	1.00	1.00	0.80	26.94
	Medium	1.00	1.00	1.00	38.35
	Large	1.00	1.00	1.00	32.01
0.05	Small	1.00	1.00	0.80	26.94
	Medium	1.00	1.00	1.00	38.35
	Large	1.00	1.00	1.00	32.01

Note: For each value of ε , the optimisation is rerun and the selected configuration is independently verified by DES. 'DES mean' denotes feasibility based on the estimated mean throughput, whereas 'DES 95% CI' denotes feasibility based on the corresponding confidence interval. This subset differs from the 24-configuration end-to-end verification reported in Table 11.

Table A9. Utilisation-to-power sensitivity in the manufacturing-inspired case.

Utilisation perturbation	Selected configuration	Power reduction (%)
Baseline	[1, 3, 3, 1, 2]	20.1
All utilisations reduced by 20%	[1, 3, 3, 1, 2]	20.7
All utilisations increased by 20%	[1, 3, 3, 1, 2]	20.1
Independent station factors sampled from [0.8, 1.2]	[1, 3, 3, 1, 2] in all 1000 draws	20.16 mean

lower-confidence-bound criterion at every tested margin. The robustness-margin sensitivity uses a dedicated five-instance subset for each dimensional group. The instances are selected by approximately equally spaced positions within the ordered benchmark groups, and the optimisation is rerun for every value of ε . Consequently, the feasibility counts in this sensitivity analysis are not directly comparable with the group counts reported in the main end-to-end verification table.

The margin therefore acts as a controllable trade-off between energy reduction and analytical headroom. It improves separation from the estimated feasibility boundary when an alternative higher-capacity vector is available, but cannot eliminate uncertainty arising from a configuration whose DES performance remains statistically close to the target.

A.9. Utilisation and power-parameter uncertainty

The energy objective depends on station utilisation and on the parameters $P_{idle,i}$ and $P_{proc,i}$. Their influence is examined first through deterministic perturbations and then through independent station-level Monte Carlo draws. Since these quantities do not alter the throughput-feasible set, the optimisation is repeated over the same analytically feasible configurations without rerunning DES. Table A9 reports the resulting utilisation-to-power sensitivity.

In the Monte Carlo utilisation experiment, the 5–95% interval of the power reduction is 18.28–22.17%. The selected vector is unchanged in every draw.

Table A10 reports the deterministic power-parameter sensitivity of the same case.

Table A10. Deterministic power-parameter sensitivity in the manufacturing-inspired case.

Power-parameter scenario	Selected configuration	Power reduction (%)
Baseline	[1, 3, 3, 1, 2]	20.1
$P_{idle} - 20\%$	[1, 3, 3, 1, 2]	18.9
$P_{idle} + 20\%$	[1, 3, 3, 1, 2]	21.2
$P_{proc} - 20\%$	[1, 3, 3, 1, 2]	21.4
$P_{proc} + 20\%$	[1, 3, 3, 1, 2]	19.1
High idle / low processing	[1, 3, 3, 1, 2]	22.6
Low idle / high processing	[1, 3, 3, 1, 2]	18.1

Table A11. Sensitivity of the power-reduction ratio to $\pm 20\%$ perturbations of power parameters in BASE-A-L.

Perturbation	P_{all} (kW)	$\Delta P / P_{all}$ (%)
$P_{idle} \times 0.8$	64.59	29.9
$P_{idle} \times 0.9$	65.44	30.3
$P_{idle} \times 1.0$	66.30	30.7
$P_{idle} \times 1.1$	67.15	31.1
$P_{idle} \times 1.2$	68.01	31.5
$P_{proc} \times 0.8$	54.85	31.6
$P_{proc} \times 0.9$	60.57	31.1
$P_{proc} \times 1.0$	66.30	30.7
$P_{proc} \times 1.1$	72.02	30.3
$P_{proc} \times 1.2$	77.75	30.0

Independent station-level perturbations in [0.8, 1.2] are also applied to both idle and processing power, with $P_{proc,i} > P_{idle,i}$ enforced after perturbation. The selected vector remains [1, 3, 3, 1, 2] in all 1000 draws. The mean power reduction is 20.08%, with a 5–95% interval of 19.18–21.10%.

Table A11 reports an additional parameter perturbation for the BASE-A-L configuration pair. In this experiment, station utilisations are retained at their nominal analytical values and the common power vectors are scaled uniformly.

The perturbation results separate two aspects of the energy decision. The numerical estimate of power saving moves with the assumed utilisation and power parameters, whereas the selected capacity vector remains unchanged in the manufacturing-inspired case. This stability reflects the discrete separation between the feasible alternatives rather than invariance of the absolute energy estimate.

A.10. Energy-function shape and start-up penalty

The following analysis is a static sensitivity study of the load-to-power mapping used in the shift-level objective. The functions u_i^γ are not dynamic machine-switching models. They do not represent the timing or frequency of idle windows, machine-state transitions, minimum on/off durations, warm-up trajectories, or state-dependent switch-off opportunities.

Instead, $\gamma = 0.5, 1$, and 2 provide alternative concave, affine, and convex steady-state mappings between average utilisation and average processing-related power. The additional start-up term is an amortised shift-level penalty used to test whether the selected capacity vector is sensitive to a conservative per-machine reconfiguration cost.

Applications concerned with online within-shift switching require an explicit dynamic machine-state model and are outside the scope of the present formulation.

Table A12. Sensitivity to energy-function shape in the manufacturing-inspired case.

Energy objective	Selected configuration	Power reduction (%)
Concave, $\gamma = 0.5$	[1, 3, 3, 1, 2]	22.4
Affine, $\gamma = 1$	[1, 3, 3, 1, 2]	20.1
Convex, $\gamma = 2$	[1, 3, 3, 1, 2]	18.1
Affine with start-up penalty	[1, 3, 3, 1, 2]	19.1
Concave with start-up penalty	[1, 3, 3, 1, 2]	21.5
Convex with start-up penalty	[1, 3, 3, 1, 2]	17.0

The alternative steady-state load-to-power mappings are defined as

$$P_\gamma(\mathbf{x}) = \sum_i x_i \left[P_{\text{idle},i} + (P_{\text{proc},i} - P_{\text{idle},i}) \left(u_i^{\text{clip}} \right)^\gamma \right] + P_{\text{aux}}. \quad (\text{A2})$$

The amortised start-up variant is then written as:

$$P_\gamma^{\text{start}}(\mathbf{x}) = P_\gamma(\mathbf{x}) + \frac{C_{\text{start}}}{H} \sum_i (c_{i,\text{max}} - x_i), \quad (\text{A3})$$

with an 8-hour horizon H and an illustrative penalty $C_{\text{start}} = 1$ kWh per deactivated machine. The term represents an average shift-level penalty rather than a dynamic switching process.

Table A12 reports the selected configuration and power reduction under each energy-function shape.

Every tested energy-function variant selects the same active-machine vector. The power-reduction estimate ranges from 17.0% to 22.4%, showing that the objective shape influences the estimated magnitude of the saving but not the configuration ranking in this case. For this manufacturing-inspired case the selected configuration is therefore stable across the tested monotone energy-function shapes and the amortised start-up penalty; production lines whose switch-off energetics are utilisation-dependent or stochastic call for the corresponding dynamic objective, which the framework accommodates by substituting the energy function while retaining the same search and feasibility procedure.

These sensitivity results support the use of the affine model for comparative shift-level screening in the tested setting. Applications with long warm-up periods, binding minimum on/off durations, product-specific energy states, or strongly non-linear consumption curves would require direct calibration of the corresponding objective rather than reliance on the generic variants examined here.

Appendix 4. Supplementary search algorithms and benchmarks

This appendix reports the detailed search pseudocode and supplementary solver comparisons supporting the optimisation results in Section 7.5. All methods use the same analytical evaluator, power model, capacity bounds, and throughput-feasibility definitions.

A.11. Beam-search procedure

Algorithm 4 reports the constructive beam search used as the primary optimisation method. The search begins from the minimum active-capacity vector and expands neighbouring configurations by adding one machine at a time. Candidate stations are prioritised by the utilisation-congestion score in Equation (29).

The search uses memoisation so that a capacity vector is evaluated only once. The beam retains multiple partial solutions rather than committing to a single greedy path, while the trimming phase removes redundant machines after the feasible region has been reached.

A.12. GA fallback

Algorithm 5 reports the genetic-algorithm fallback. The GA is activated when the beam does not identify an evaluator-feasible configuration. Candidate vectors are integer-valued and respect the installed-capacity bounds at every station.

The baseline setting uses $n_p = 60$, $G = 150$, $e = 5$, $p_c = 0.70$, and $p_m = 0.40$. These values are fixed across instances; their sensitivity is examined in Section A.16.

A.13. DES validation loop

Algorithm 6 summarises the external validation procedure applied to selected, boundary, and comparison configurations.

A.14. Structured optimisation benchmark

Table A13 compares exact enumeration, beam search, and the GA on the nine structured BASE instances with $\delta = 0.25$ and $\varepsilon = 0$. Exact enumeration provides the reference optimum.

When the beam reaches the feasible region, it matches the exact optimum while requiring fewer evaluations than enumeration or GA. Its failures occur in the BASE-A topology, where the minimum-capacity configuration and its immediate neighbours do not provide an informative path towards feasible solutions. The GA recovers the exact optimum in these instances through broader exploration of the capacity lattice.

A.15. Comparison with greedy and hill-climbing baselines

Tables A14 and A15 compare beam search with greedy best-first search and top-down hill climbing under $\delta = 0.25$ and $\varepsilon = 0.02$. These supplementary baseline comparisons use Campaign A and are separate from the Campaign-B benchmark reported in Tables 14 and 15.

Within Campaign A, greedy best-first search requires relatively few analytical evaluations but has the lowest success rate, particularly on larger instances, because it commits to a single local expansion path. Hill climbing is more reliable but can settle on local plateaus. These results provide a supplementary comparison of search behaviour and should not be combined numerically with the Campaign-B benchmark, which uses a different random-instance generator and experimental pool.

Evaluation counts in Table A15 are higher than those in the main $\varepsilon = 0$ benchmark because the robustness margin raises the effective throughput requirement and because all trimming calls are included in the count.

A.16. GA parameter sensitivity

Three GA settings are compared on the nine structured BASE instances: a light setting with population 40 and 100 generations, the baseline setting with population 60 and 150 generations, and a heavy setting with population 80 and 250 generations. Each setting is run with five independent seeds per

Algorithm 4: Congestion-guided beam search

Input: Installed capacities $c_{\max}$; class populations $\mathbf{K}$; routing matrix V ; service-time matrix S ; power parameters; robust throughput target $TH_{\text{target}}^{\text{robust}}$; beam width w ; candidate-list size k

Output: Best evaluator-feasible configuration $\mathbf{x}^*$ and power P^* , or INFEASIBLE

```

1  $\mathbf{x}_{\text{init}} \leftarrow \mathbf{1}_{\text{required}}$ ;
2  $\mathcal{B} \leftarrow \{\mathbf{x}_{\text{init}}\}$ ;
3  $\mathbf{x}^* \leftarrow \text{null}$ ;  $P^* \leftarrow +\infty$ ;
4 while  $\mathcal{B} \neq \emptyset$  do
5    $\mathcal{B}' \leftarrow \emptyset$ ;
6   foreach  $\mathbf{x} \in \mathcal{B}$  do
7     Evaluate  $\mathbf{x}$  using the configuration-specific EMDA evaluator;
8     if  $\mathbf{x}$  is evaluator-usable and  $TH_{\text{emda}}(\mathbf{x}) \geq TH_{\text{target}}^{\text{robust}}$  then
9        $\mathbf{x}' \leftarrow \text{TRIM}(\mathbf{x}, TH_{\text{target}}^{\text{robust}})$ ;
10      Evaluate  $P(\mathbf{x}')$ ;
11      if  $P(\mathbf{x}') < P^*$  then
12         $\mathbf{x}^* \leftarrow \mathbf{x}'$ ;  $P^* \leftarrow P(\mathbf{x}')$ ;
13      end
14      continue;
15    end
16    foreach station  $i$  with  $x_i < c_{i,\max}$  do
17       $\text{score}_i \leftarrow EN_i u_i / P_{\text{idle},i}$ ;
18    end
19    Select the top- $k$  expandable stations by score;
20    foreach selected station  $i$  do
21       $\mathbf{x}'' \leftarrow \mathbf{x}$ ;
22       $x''_i \leftarrow x''_i + 1$ ;
23      Add  $\mathbf{x}''$  to  $\mathcal{B}'$ ;
24    end
25  end
26  Rank  $\mathcal{B}'$  by throughput-gap ascending and power ascending;
27   $\mathcal{B} \leftarrow \text{top-}w \text{ configurations in } \mathcal{B}'$ ;
28 end
29 if  $\mathbf{x}^* = \text{null}$  then return INFEASIBLE
30 return  $(\mathbf{x}^*, P^*)$ ;

31 Function  $\text{TRIM}(\mathbf{x}, TH_{\text{target}}^{\text{robust}})$ :
32   repeat
33     foreach station  $i$  with  $x_i > 1$  do
34        $\mathbf{x}' \leftarrow \mathbf{x}$ ;
35        $x'_i \leftarrow x'_i - 1$ ;
36       Evaluate  $\mathbf{x}'$ ;
37       if  $\mathbf{x}'$  is evaluator-feasible and  $P_{\text{tot}}(\mathbf{x}') < P_{\text{tot}}(\mathbf{x})$  then
38          $\mathbf{x} \leftarrow \mathbf{x}'$ ;
39         break;
40       end
41     end
42   until no improving removal is found
43   return  $\mathbf{x}$ 

```

instance. Table A16 reports the success rate and optimality gap of each setting.

All three GA settings identify an evaluator-feasible solution and recover the exact optimum in all 45 runs. The structured

optimisation results are therefore insensitive to the tested population sizes and generation limits. Increasing the population and number of generations does not improve solution quality on these instances, although it increases computational effort.

Algorithm 5: Genetic-algorithm fallback

Input: $c_{\max}$, K , V , S ; power parameters; robust throughput target; population size n_p ; number of generations G ; elite size e ; crossover probability p_c ; mutation probability p_m

Output: Best evaluator-feasible configuration $\mathbf{x}^*$, or INFEASIBLE

- 1 Generate an initial population of integer vectors satisfying

$$1 \leq x_i \leq c_{i,\max}$$
 for all visited stations;
- 2 Evaluate every individual using the EMDA evaluator and power model;
- 3 Assign penalised fitness to evaluator-unusable or throughput-infeasible individuals;
- 4 **for** $g = 1, \dots, G$ **do**
 - 5 Retain the best e individuals as elites;
 - 6 Select parents from the current population;
 - 7 Apply crossover with probability p_c ;
 - 8 Apply mutation with probability p_m by changing one or more station capacities within their bounds;
 - 9 Evaluate all new individuals using the EMDA evaluator;
 - 10 Update the best evaluator-feasible individual found so far;
- 11 **end**
- 12 **if** an evaluator-feasible vector has been found **then**
return $(\mathbf{x}^*, P(\mathbf{x}^*))$
- 13 **return** INFEASIBLE;

The baseline setting is retained as an intermediate and conservative parameterisation for consistency across the remaining experiments; the sensitivity results are not intended to identify a uniquely optimal GA setting.

Algorithm 6: DES validation procedure

Input: Set of configurations $\mathcal{X}$; routing and processing-time data; class populations; simulation horizon T ; warm-up T_w ; number of replications n

Output: DES throughput mean, confidence interval, and feasibility label for each configuration

- 1 **foreach** $x \in \mathcal{X}$ **do**
 - 2 **for** $j = 1, \dots, n$ **do**
 - 3 Set independent replication seed j ;
 - 4 Simulate the closed two-class CONWIP system under configuration $\mathbf{x}$;
 - 5 Discard observations collected before T_w ;
 - 6 Estimate post-warm-up replication throughput TH_j ;
 - 7 **end**
 - 8 Compute $\overline{TH}_{\text{DES}}$ and s_{TH} ;
 - 9 Compute the Student- t 95% confidence interval;
 - 10 Compare the replication mean and lower confidence bound with the instance-specific DES throughput target;
 - 11 Store throughput and feasibility outputs;
- 12 **end**
- 13 **return** the DES validation table;

Table A13. Search benchmark on the structured BASE scenarios with $\delta = 0.25$ and $\varepsilon = 0$.

Instance	Method	State space	P_{sol} (kW)	Gap (%)	Evaluations	Time (s)
BASE-A-S	Exact	6	20.539	–	6	0.26
	Beam	6	–	–	2	0.19
	GA	6	20.539	0.00	6	0.45
BASE-A-M	Exact	48	31.488	–	48	0.28
	Beam	48	–	–	2	–
	GA	48	31.488	0.00	41	1.66
BASE-A-L	Exact	162	45.950	–	162	0.34
	Beam	162	–	–	–	–
	GA	162	45.950	0.00	88	2.84
BASE-B-S	Exact	2	11.616	–	2	0.13
	Beam	2	11.616	0.00	2	0.06
	GA	2	11.616	0.00	2	0.24
BASE-B-M	Exact	16	20.923	–	16	0.23
	Beam	16	20.923	0.00	9	0.25
	GA	16	20.923	0.00	16	0.62
BASE-B-L	Exact	54	24.191	–	54	0.38
	Beam	54	24.191	0.00	15	0.41
	GA	54	24.191	0.00	46	1.45
BASE-C-S	Exact	6	18.893	–	6	0.10
	Beam	6	18.893	0.00	6	0.10
	GA	6	18.893	0.00	6	0.27
BASE-C-M	Exact	48	28.549	–	48	0.30
	Beam	48	28.549	0.00	19	0.29
	GA	48	28.549	0.00	41	0.78
BASE-C-L	Exact	162	39.498	–	162	0.41
	Beam	162	39.498	0.00	29	0.40
	GA	162	39.498	0.00	84	1.38

Note: Gap is the relative power difference from the exact optimum. GA values are medians over three seeds.

Table A14. Baseline comparison on the structured BASE instances with $\delta = 0.25$ and $\varepsilon = 0.02$.

Instance	Hybrid beam-GA		Greedy		Hill climbing	
	Saving (%)	Evals	Saving (%)	Evals	Saving (%)	Evals
BASE-A-S	9.0	6	–	–	9.0	5
BASE-A-M	25.1	37	–	–	25.1	10
BASE-A-L	30.7	130	–	–	28.3	18
BASE-B-S	13.4	2	13.4	2	13.4	2
BASE-B-M	20.5	16	20.5	7	20.5	12
BASE-B-L	33.2	44	23.1	11	33.2	19
BASE-C-S	8.6	6	8.6	5	8.6	5
BASE-C-M	22.3	36	22.3	11	22.3	12
BASE-C-L	34.6	122	34.6	17	34.6	22

Note: A dash indicates that no evaluator-feasible solution was found.

Table A16. GA parameter sensitivity on the structured BASE instances.

GA setting	Population	Generations	Successful runs	Exact-match runs	Maximum gap (%)
Light	40	100	45/45	45/45	0.00
Baseline	60	150	45/45	45/45	0.00
Heavy	80	250	45/45	45/45	0.00

Table A15. Baseline comparison on Campaign A with $\delta = 0.25$ and $\varepsilon = 0.02$.

Group	Solver	Success	Saving (%)	Gap (%)	Evaluations
Small, $m = 5$	Beam	95%	33.6	0.00	143
	Greedy	90%	32.9	0.00	12
	Hill climbing	100%	33.6	0.00	29
Medium, $m = 8$	Beam	100%	36.8	0.00	246
	Greedy	98%	35.7	0.00	20
	Hill climbing	100%	35.8	0.00	72
Large, $m = 10$	Beam	100%	37.0	0.00	164
	Greedy	82%	36.1	0.00	23
	Hill climbing	100%	36.0	0.00	108

Note: Campaign A contains 119 evaluator-usable random instances from 120 attempts. Values are medians within each dimensional group.