ORIGINAL RESEARCH **Open access**

Asynchronous Multi-Fan Control for Energy-Efficient Server Cooling Using Evolution Strategy Optimization

Noah Williams[1*], Liam Johnson[1], Grace Collins[2], Ethan Brooks[1]

Abstract

A typical server within a data center houses multiple fans alongside numerous heat-producing components. The velocity of fans, spatial arrangement of parts, and workload profiles collectively shape the internal airflow patterns and heat distribution. Classifying server thermal management as a Multi-Input Multi-Output (MIMO) nonlinear control problem is therefore appropriate, given that straightforward linear analytical tools prove insufficient. By applying a weighting methodology—where each cooling fan receives a distinct coefficient for independent modulation—and integrating it with the Evolutionary Strategy (ES) algorithm guided by carefully designed fitness metrics, the present work brings to life a multi-fan thermal regulation system. This engineered system secures near-optimal weighting sets inside a finite search count, tailored to varying workload scenarios. Regulating fans independently according to these sets permits near-optimal energy conservation while still respecting server thermal boundaries, specifically the upper temperature thresholds of the CPU and PCIe. When benchmarked against uniform fan control, the independent modulation approach trims total fan power usage by an average of 43.1%, while scanning merely 14.2% of all conceivable weighting configurations across the eight experimental workload profiles. Additionally, the chance that an approximate optimal weighting corresponds to the global best is recorded at 47.5%. Empirical findings confirm that the introduced independent multi-fan regulation system simultaneously safeguards thermal limits and secures near-optimal power savings. In effect, the developed framework is both practical and high-performing in terms of significant energy reduction, all while circumventing the need to construct mathematical heat transfer representations or process voluminous data collections. With future adjustments to the system's software parameters, adaptation to a wider variety of server platforms becomes feasible.

Keywords Server thermal control, Energy-saving optimization, Weighting mechanism, Multi-fan control system with asynchronously fan weightings, Evolution strategy

*Correspondence:
Noah Williams
noah.williams@uq.edu.au
[1] Department of Artificial Intelligence for Energy Engineering, Faculty of Engineering, University of Queensland, Brisbane, Australia
[2] Department of Intelligent Energy Analytics, Faculty of Engineering, University of Sydney, Sydney, Australia

Introduction

The rapid maturation of 5G infrastructure, Artificial Intelligence (AI), and the Internet of Things (IoT), compounded by shifts caused by the COVID-19 crisis, has accelerated digitalization across sectors worldwide. Major cloud operators including Amazon, Microsoft, and Google stand out as prime gainers from this trend [1], with corresponding steady expansion in international revenues. Their sustained channeling of capital into infrastructure has fueled the erection of vast data center campuses, with projections placing the count of hyperscale facilities beyond the one-thousand mark by 2024 [2]. By 2016, the planetary yearly electrical draw of data centers had already surpassed 300 terawatt-hours (TWh)

This is an open access article distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License (CC BY-NC-SA 4.0), which permits non-commercial use, distribution, reproduction, and adaptation in any medium or format, provided that appropriate credit is given to the original author(s) and source, a link to the license is provided, and any changes made are indicated. Adapted material must be distributed under the same license.

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

[3], a quantity on par with the entire annual usage of an industrialized country. Forecasts for 2030 indicate this consumption could swell to 1,700 TWh [4, 5], edging toward 2.13% of worldwide electricity demand [6].

At present, cloud operators actively pursue green energy strategies for their facilities, aiming not just to honor decarbonization pledges aligned with the Paris Agreement [7] and the Glasgow Climate Pact [8], but also to curb operational electricity expenses and lift efficiency ratings. Power Usage Effectiveness (PUE) is the principal metric for gauging facility energy performance [9], expressed as the ratio of a site's total electrical consumption to that drawn by the IT hardware exclusively. A perfect PUE score of 1 would mean all inbound power is dedicated purely to computing equipment. Achieving this ideal remains unattainable in practice, since the overall consumption also covers electrical losses in transformers, uninterruptible power supplies, illumination, and, critically, cooling plant [10]. Cooling infrastructure alone is responsible for 37% of the aggregate electrical load, second only to the IT apparatus at 50%, while the specific contribution of cooling fans constitutes 12% of the facility's total draw. Data centers aggregate anywhere from a few dozen to tens of thousands of servers, with the individual server forming the foundational compute block. Economic considerations in capital build-out and ongoing upkeep keep air-cooling as the prevalent technique for heat rejection from servers, consequently driving considerable fan electricity expenditure. Thus, engineering a fan regulation scheme that achieves both component thermal compliance and energy-optimal operation has risen as a vital research avenue in the data center domain.

A typical air-cooled server is fitted with between 4 and 8 fans tasked with pulling heat away from vital elements like the Central Processing Unit (CPU), Graphics Processing Unit (GPU), and Peripheral Component Interconnect Express (PCIe) cards. Each machine incorporates several heat-dissipating parts. Their hardware makeup, relative positions on the board, and workload intensities all influence the resulting thermal picture. When heat sources are arranged with inadequate spacing, thermal interaction effects manifest. Beyond that, modulating individual fan rotational speeds reshapes the interior airflow trajectories. These interconnected variables complicate attempts at straightforward system characterization, defining server thermal regulation as a Multi-Input Multi-Output (MIMO) nonlinear challenge [11].

Shin *et al*. [12] detailed the fan affinity laws, establishing that a fan's rotational speed relates to the cube of its power consumption and linearly to its pulse-width modulation duty signal. Efficient fan control, as opposed to maintaining them at elevated duty settings for prolonged durations, can yield substantial reductions in wasted energy.

Also, Wang *et al*. [11] conducted experimental investigations on a rack populated with multiple blade-style servers and cooling fans. Drawing from per-server workload states, chip junction temperatures, intake air temperatures, and fan characteristic curves, they posited a thermal resistance relationship, extracted coefficients via data regression, and built a plant thermal transfer representation. Heat-map visualizations were employed to unpack the tangled thermal interdependencies linking separate servers and fans. Factoring in the cubic relation tying fan speed to electrical power, they engineered a MIMO control law for zonal fan orchestration aimed at optimizing the thermal regulation loop.

Lee and Chen [13] reported on a self-adjusting PID parameter methodology deployed for unified fan speed adjustments. Foregoing any prerequisite system identification step, they harnessed Neural Networks for real-time tuning of controller gains, guided by temperature limits and power minimization targets.

Acun [14] captured server telemetry covering individual CPU power draws, core utilization ratios, clock frequencies, silicon temperatures, fan revolutions, and ambient environmental readings spanning various workload scenarios. Following a data cleaning phase, a neural network model for temperature projection was trained using the Levenberg-Marquardt backpropagation technique. This predictive thermal model, operating alongside anticipatory control logic, trialed different fan speed combinations in simulation. A cost formulation then singled out the most favorable speed set, thereby implementing proactive fan management.

The exploration of data-driven regulation and digital twin methodologies has extended into server cooling applications. In 2019, Brännvall *et al*. [15] described a tuning procedure grounded in digital twins for server fan controllers, illustrating its

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

capacity to boost cooling performance. Separately, in 2020, Sarkinen *et al.* [16] performed an experimental assessment of autonomous server fan management schemes, highlighting real-world deployments capable of strengthening thermal handling within data centers. The contributions from those works yield pertinent knowledge that reinforces the direction taken in our study.

Chu *et al*. [17] put forward a set of reductive assumptions about server behavior and extracted the system's thermal representation via empirically derived relationships [16]. Using measurement data collected from thermal sources, they applied reinforcement learning based on the Deep Deterministic Policy Gradient (DDPG) framework and engineered reward formulations to deliver a fan management policy targeting energy reduction.

Chu *et al*. [18] presented a scheme that splits a bank of parallel fans residing within a server into a primary group and a secondary group. That effort trialed varied permutations of server load conditions and fan drive ratios to investigate patterns of thermal decoupling. They subsequently trained the MTSMFF (Multivariate Time Series Forecasting Framework) alongside the MsA (Multistage Attention) deep learning architectures to generate temperature projections.

Across the contributions [13, 14, 17], fans contained in servers receive uniform, synchronized actuation signals. Lee and Chen [13] and Acun [14] note that transplanting the concept of fan weighting [19] to enable non-uniform, independent fan operation holds promise for counteracting excessive cooling of parts—an artifact of synchronized actuation—and for paring down fan energy draw. Even so, the concrete process of formulating and determining those weightings is left undetailed. Furthermore, the learning-based models characterized in Acun [14], Chu *et al*. [17], and Chu *et al*. [18] lean on copious measurement records for both fitting and verification stages, a fact that translates to sizable computational burden. Questions surrounding the ongoing serviceability of such models once in production become paramount, particularly when hardware swaps may shift heat transfer characteristics or when substituting fan models redefines cooling effectiveness. On the ground, cooling fans inside servers fall under the command of the Baseboard Management Controller (BMC) silicon. Given the considerable processing footprint demanded by the models of Acun [14], Chu *et al*. [17], and Chu *et al*. [18], there is a pressing case for streamlining them; otherwise, embedding their logic directly onto BMC hardware risks becoming unmanageably intricate.

Grounded in the physical apparatus documented in Chu *et al*. [18], the present contribution introduces an original multi-fan management technique. Under this technique, a weighting scheme governs each fan individually. The Evolutionary Strategy (ES) algorithm is then coupled with the multi-fan management technique to assemble a server multi-fan regulation system that both respects thermal ceilings and secures energy-optimal behavior. The resulting framework sidesteps the need to derive a formal heat transfer representation, instead carrying out live parameter refinement without requiring extensive data accumulation. This feature translates to lean rollout overhead. In operational terms, one only needs to store the fan weightings linked to specific loading profiles onto the server BMC chip to realize multi-fan governance.

This research exclusively operates the server apparatus described by Chu *et al.*[18], inspired by the fan weighting notions found in Lee and Chen [13], Acun [14], and Byquist and Crocker [20], deploys the weighting mechanism to drive multiple fans independently. By wedging the Evolutionary Strategy algorithm to purpose-built fitness measures, this effort delivers the multi-fan regulation framework for the server—one that satisfies stipulated temperature requirements while capturing near-optimal reductions in energy expenditure. The construction of a thermal model is rendered superfluous. Because fan weightings are tuned online, the approach also forgoes amassing vast stores of data, ensuring that the expense of deploying the system stays low. Practically speaking, by loading the appropriate fan weightings for differing load scenarios onto the server BMC chip, the proposed framework can enact multi-fan control with heterogeneous weightings.

Experimental System

The experimental campaign underpinning this study utilized a 2U rack-mount server, as shown in **Figures 1-3**. With physical dimensions of 450 mm across the width, 760 mm along the depth, and 90 mm in height, the server hosts an

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

Inventec N880G3 DC motherboard. A pair of Intel Xeon CPUs, model E5-2699 v3, each capped at a thermal design power of 144 W, populate the board. Moreover, the unit houses a dozen TOSHIBA hard disk drives (type MG04ACA500E), sixteen SK Hynix dual in-line memory modules (type DDR4-2666 16GB 2Rx4 PC4-2133P), and two bespoke PCIe adapter cards (each peaking at 24 W). It further contains dual PSUs, with each one capable of furnishing up to 750 W. Thermal evacuation from the server is handled by a set of four Delta Electronics fans (model PFR0812UHE), each spanning 80 mm × 80 mm × 38 mm and topping out at 8000 RPM in rotational velocity. These four-conductor units supply a ground return path (black conductor), a 12 V power feed (red conductor), a tachometric speed signal (yellow conductor), and a PWM input pin (blue conductor), as depicted in **Figure 4.** In accordance with the fan affinity principles [12], standalone fan characterizations were run to track power draw across a sweep of drive duty values. **Figure 5**, together with **Eq. 1**, conveys the cubic mapping linking fan power to duty ratio, constructed through curve regression. Within the expression, P and R capture fan power and duty ratio respectively, while supplementary symbols correspond to fitted coefficients. This relation was later invoked to derive the aggregate power draw across the four fans by arithmetic summation of their isolated power figures.

$$P(R) = (8.418e-7)R^3 + (0.001761)R^2 + (0.08436)R + 3.249 \quad (1)$$



Figure 1. 2U server.

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press



Figure 2. The real location map of the internal components of the server.

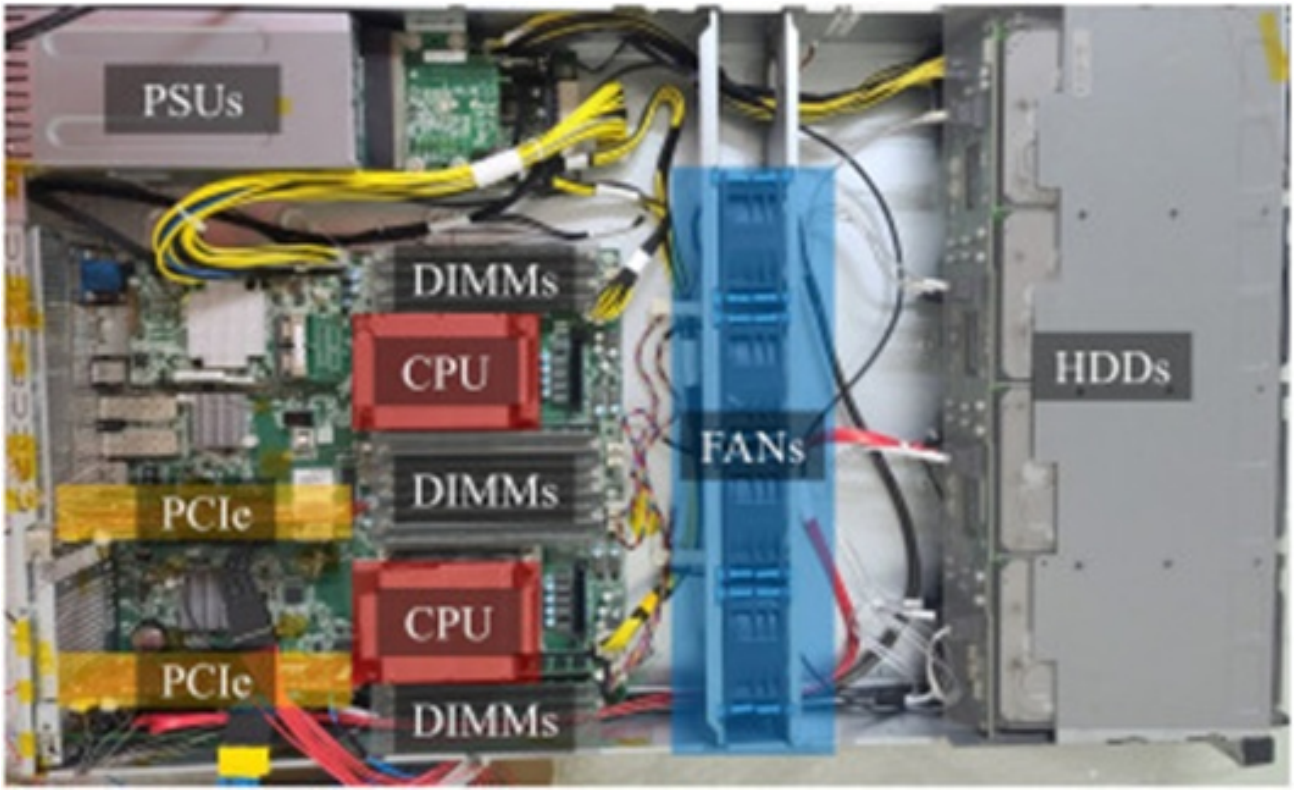


Figure 3. Schematic diagram of the internal components of the server.

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press



Figure 4. Delta Electronics PFR0812UHE server fan.

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

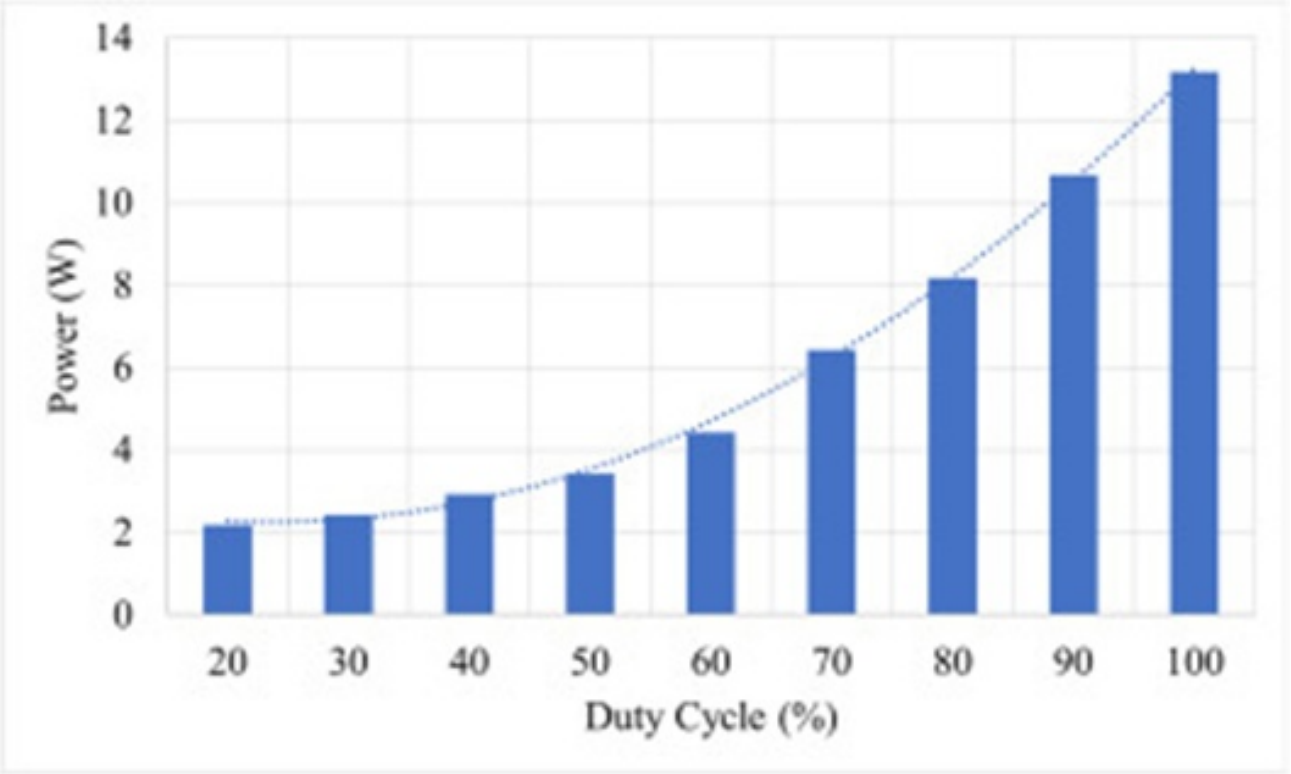


Figure 5. Power variation at different fan duty cycles.

Figure 3 renders the spatial allocation of the server's internal components. During the experimental window, an airflow guide scoop was secured directly atop the CPU to steer the airstream and amplify convective extraction, while enclosing the server with its outer chassis lid. The airflow scoop mounting is shown in **Figure 6.**

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press



Figure 6. Placement of wind scooper.

The server harnesses the Power Thermal Utility (PTU) toolset, which enables CPU power draw modulation and includes the Intelligent Platform Management Interface (IPMI) for telemetered oversight of server state. Throughout this investigation, the PTU environment was leveraged to vary CPU dissipation, yielding an assortment of loading regimes for testing. Owing to hardware constraints, however, component temperature telemetry and fan actuation did not route through the BMC chip. In its place, K-type thermocouple probes, an Agilent 34970A multiplexed acquisition instrument, and a PWM generation module were pressed into service. An Arduino Mega 2560 board managed a relay bank to switch power delivery to the PCIe cards. Because CPUs and PCIe cards are the primary heat-releasing components, thermocouple beads were affixed at their primary heat-concentrated locations to capture temperature readings. Those thermal readings were looped back to the regulation framework advanced in this paper. The regulation framework, in turn, emitted corresponding drive-level instructions to the PWM generation module, which relayed those commands to the PWM input terminals of the individual fans, thus carrying out server fan governance. This operational sequence is charted in **Figure 7.**

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

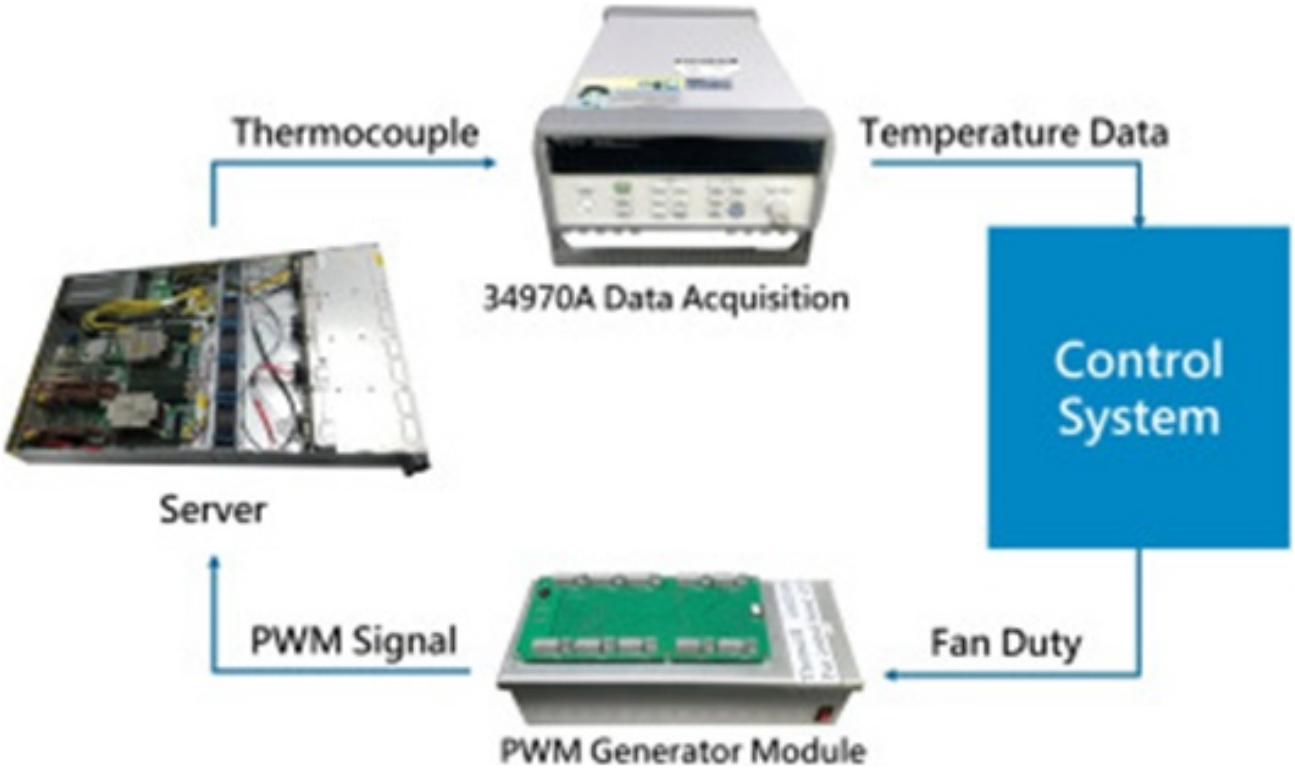


Figure 7. Schematic of experimental system.

Multi-Fan Control System

Weighting mechanism

The weighting scheme advanced in this paper is rooted in the weighting concepts documented in Lee and Chen [13], Acun [14], and Byquist and Crocker [20]. The feedback controllers are embedded with a weighting array to bring about independent, non-uniform fan actuation, as schematized in **Figure 8.**

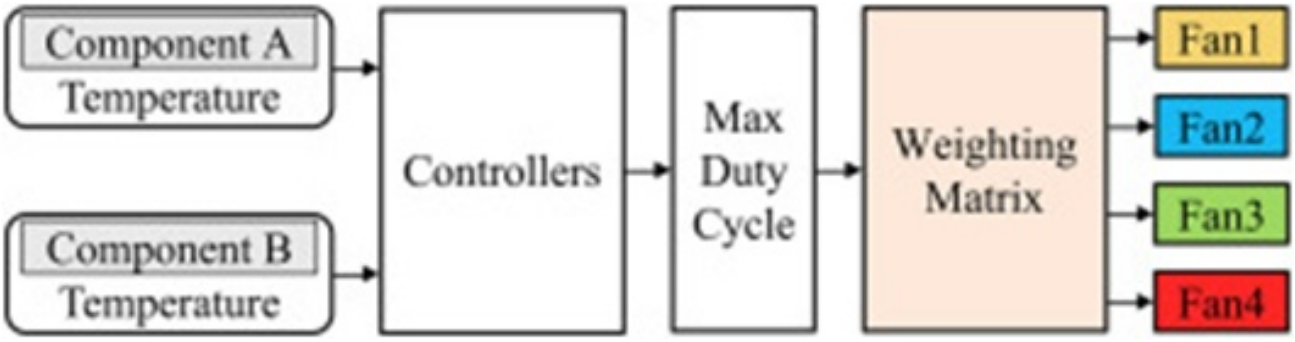


Figure 8. Asynchronous fans modulation.

The controllers, in continuous response to evolving temperature signals, perpetually recompute the ceiling drive output. The array of resulting values conveys the individual drive signals dispatched to the respective fans, as set out in **Eq. 2**. The term *Weighting fan* captures the scaling factor mapped to each fan, constrained within the interval 0.0 to 1.0. Should all fan weightings be uniformly fixed at 1.0, the fans revert to synchronized rather than independent modulation. *Max Duty Cycle* denotes the upper-bound duty output emanating from the controllers. *Fan* denotes the particular drive

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

command fed to the corresponding fan hardware. While the present method does not mandate any single controller topology, this work settles on the PID (Proportional-Integral-Derivative) architecture [18] given its transparent correspondence between physical quantities and tuning parameters, together with its straightforward realization. It endures as the prevalent control scheme across a broad industrial spectrum. The controller enforces closed-loop regulation via feedback signals originating from the managed thermal plant. Eqs capture the Laplace-domain transfer function Gc(s) of the PID structure. 3. At the same time, the temporal control effort u(t) follows the formulation in **Eq. 4**. Here, *Kp* signifies the proportional contribution, scaling corrective action directly against the instantaneous feedback error as dictated by the set gain. *Ki* embodies the integral contribution, accumulating past error over time per the set gain to drive out residual steady-state mismatch. *Kd* represents the derivative contribution, shaping corrective effort in proportion to the rate of change of the error as dictated by the set gain and furnishing anticipatory action aligned with the evolving control trajectory, thereby trimming settling duration and buttressing overall loop stability. The variable *s* stands for the complex-frequency operator, native to the Laplace transform domain. The signal *e(t)* denotes the instantaneous error quantity, derived by subtracting the plant's measured output from the desired setpoint. The symbol *t* indexes the present time coordinate. The practitioner is tasked simply with selecting the triad Kp, Ki, and Kd to conform to the intended closed-loop performance profile. Heeding the strengths articulated above, this inquiry adopts the PID construct. It iteratively revises the ceiling drive command informed by the fed-back temperature evolution of the critical heat-emitting parts. The operator need only assign appropriate Kp, Ki, and Kd values to bring the transient and steady-state behavior into line with the design envelope. In light of these merits, the present work implements the control layer using PID elements, focusing on feedback-based thermal management of essential components.

$$\begin{bmatrix} Weighting\ fan1 \\ Weighting\ fan2 \\ Weighting\ fan3 \\ Weighting\ fan4 \end{bmatrix} \times [\text{Max Duty Cycle}] = \begin{bmatrix} Fan1 \\ Fan2 \\ Fan3 \\ Fan4 \end{bmatrix} \quad (2)$$

$$G_c(s) = K_p + \frac{K_i}{s} + K_d s \quad (3)$$

Evolutionary strategy algorithm

Evolutionary Algorithms (EAs) [21] constitute a class of metaheuristic methods situated within the artificial intelligence landscape. The family encompasses several distinct paradigms, notably the Genetic Algorithm (GA), Evolution Strategy (ES), and Genetic Programming (GP). Rooted in the evolutionary theory advanced by Darwin [22], EAs replicate nature's selection and adaptation processes to refine successive populations toward optimal configurations progressively. Their utility spans diverse optimization contexts, delivering benefits that include curtailed computational analysis periods and diminished susceptibility to entrapment within local optima relative to conventional numerical techniques.

The Evolution Strategy (ES) encodes multidimensional real-valued problems straight into real-valued vectors destined for evolutionary processing. Its operation hinges on mutation operators whose defining parameters are automatically recalibrated as evolution proceeds, obviating any requirement for operator intervention. Beyond this, ES customarily attains convergence more rapidly than either GA or GP. This trait has cemented its status as the preferred option for parameter optimization challenges across engineering disciplines. **Figure 9** offers a schematic overview of the ES workflow. The procedure initiates with a setup phase, producing offspring until the designated population count λ is filled. Fitness functions are then computed against every member of the offspring pool. The top-scoring μ candidates, ordered by their fitness values, are designated as progenitors for the following round. Novel offspring subsequently emerge through the operations of recombination and mutation. This cyclical sequence (once initialization is complete) iteratively rejuvenates the population and drives fitness enhancement until one or more predetermined halting conditions become true. Acceptable halting conditions span reaching a preset generational cap, observing insufficient population variation, hitting a desired fitness benchmark, or detecting algorithmic convergence. **Table 1** collates typical parameter choices for ES, where each offspring is encoded as a real-valued vector. Discrete recombination selects parents stochastically and reproduces their vector contents without alteration. Mutation injects fresh variation into offspring by adding stochastic

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

disturbances—sampled from a Gaussian distribution [23]—to every component of the vector. For the Gaussian distribution N (ξ, σ), two defining quantities exist: ξ and σ. The first, ξ, captures the central tendency of the noise, normally anchored at zero, while the second, σ, quantifies the standard deviation that dictates how broadly the noise disperses. Recognized as the mutation step size, σ stands as a cornerstone parameter governing the algorithm's exploratory behavior.

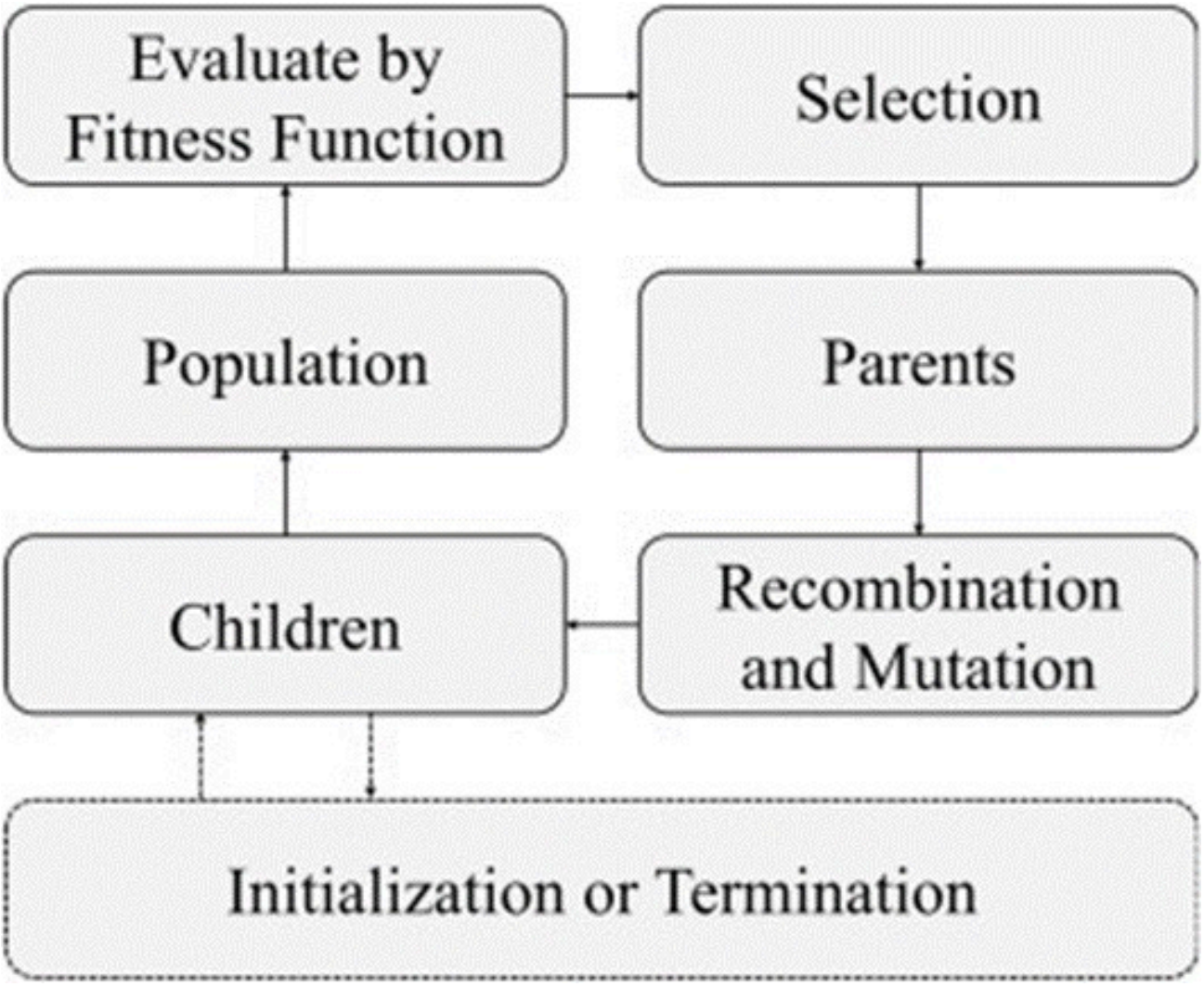


Figure 9. The flowchart of ES algorithm.

Table 1. Common parameter settings for Evolution Strategy.

Evolutionary Algorithm Component	Configuration
Solution Encoding	Real-valued vector representation
Crossover Strategy	Discrete recombination
Mutation Operator	Gaussian perturbation
Parent Sampling Method	Uniform random selection
Survivor Replacement Scheme	(μ, λ) selection strategy

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

Distinctive Feature	Self-adaptive adjustment of mutation step sizes

The treatment of mutation step size can be organized into two fundamental philosophies: parameter tuning, where the step size remains immutable throughout the evolutionary timeline, and parameter control, where the step size is permitted to fluctuate dynamically over the course of evolution. Parameter Control further branches into deterministic, adaptive, and self-adaptive subcategories. The self-adaptive paradigm is generally favored: under this scheme, the step size σ undergoes its own evolutionary update during each iteration. Whenever solution fitness degrades, σ inflates to boost diversity and broaden the territory being searched. In contrast, once high-fitness solutions dominate, σ shrinks to curb variability and intensify the search around the immediate neighborhood.

Complementing the self-adaptive parameter control framework, three principal formulations have been documented [21, 24]: Uncorrelated mutation with one step-size, under which a solitary step size value is uniformly applied across all components of the vector **(Figure 10)**; uncorrelated mutation with n step-sizes, in which each component of the vector is assigned a separate step size σ1, σ2, …, σn, and the step size for each dimension undergoes autonomous evolution **(Figure 11)**; and correlated mutation, wherein each vector component receives a dedicated step size σ1, σ2, …, σn, augmented by rotation angles α1, α2, …, αn(n-1)/2 defined between every pairing of dimensions, with both the step sizes and the rotation angles co-adapting as generations advance **(Figure 12).**

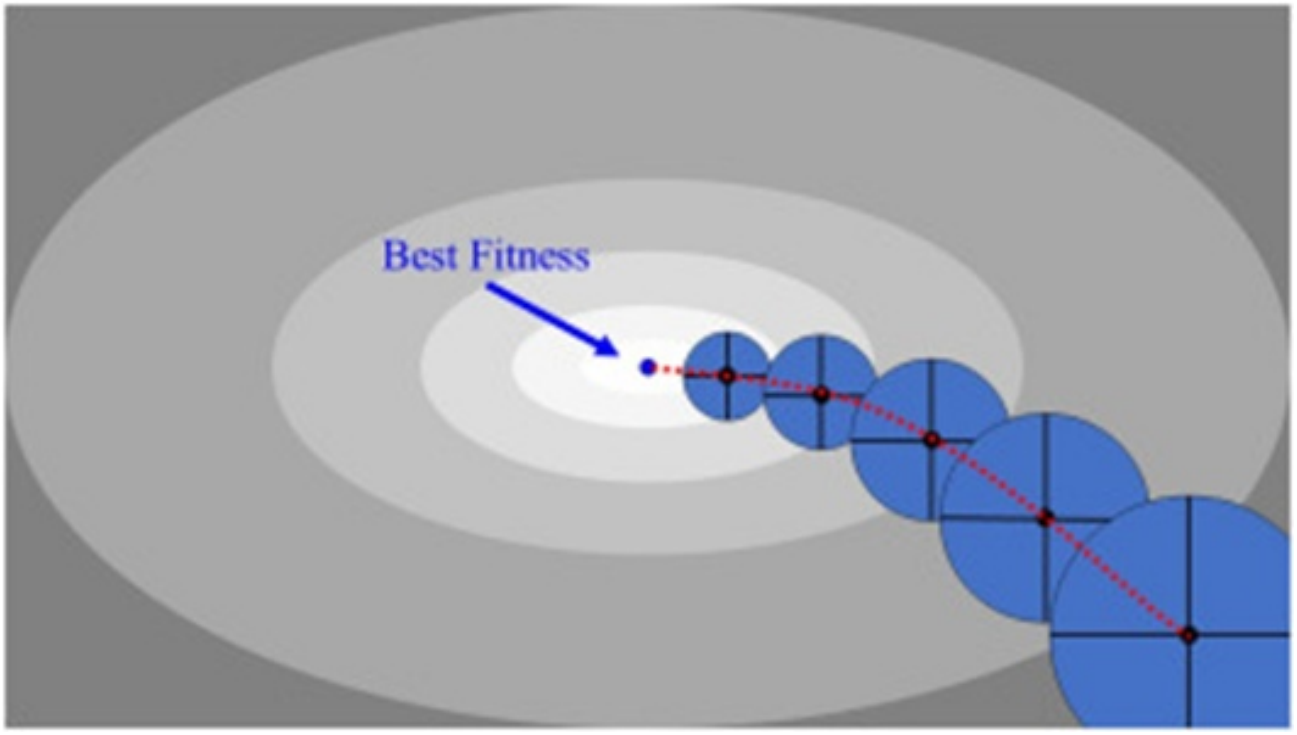


Figure 10. Uncorrelated mutation with one step-size.

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

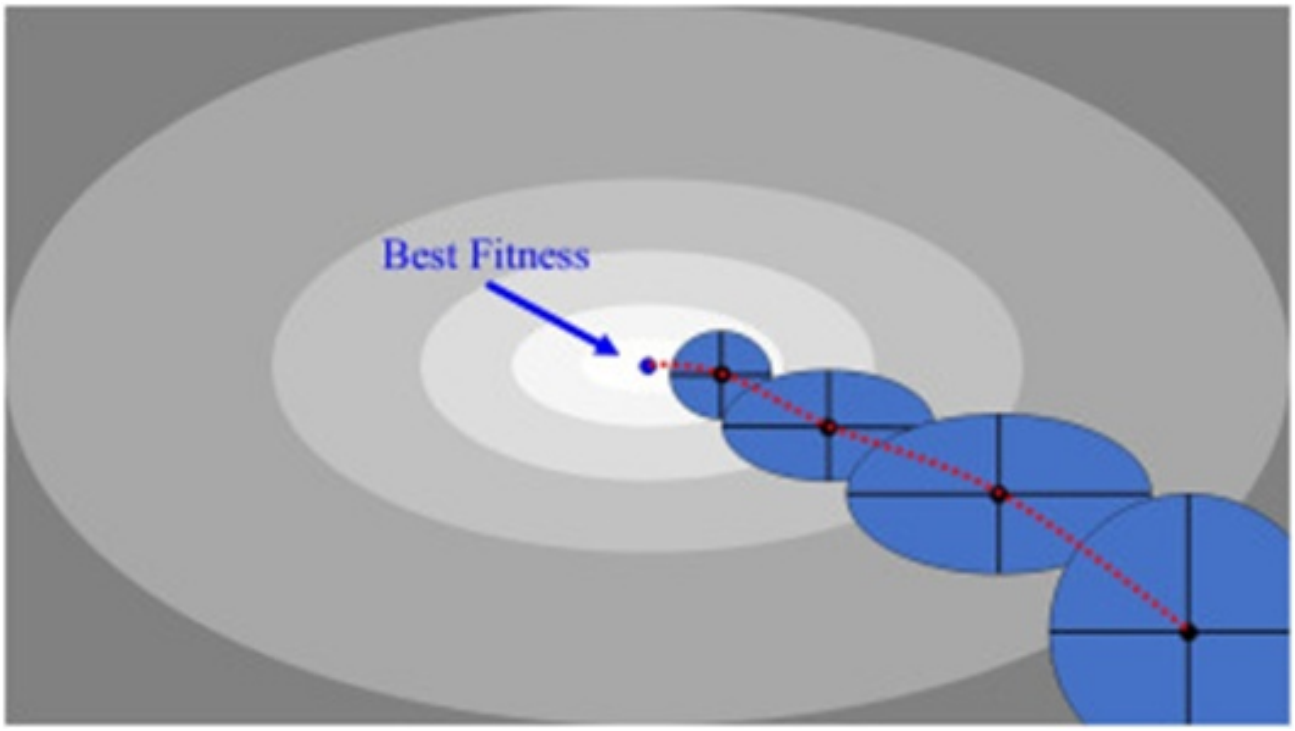


Figure 11. Uncorrelated mutation with n step-sizes.

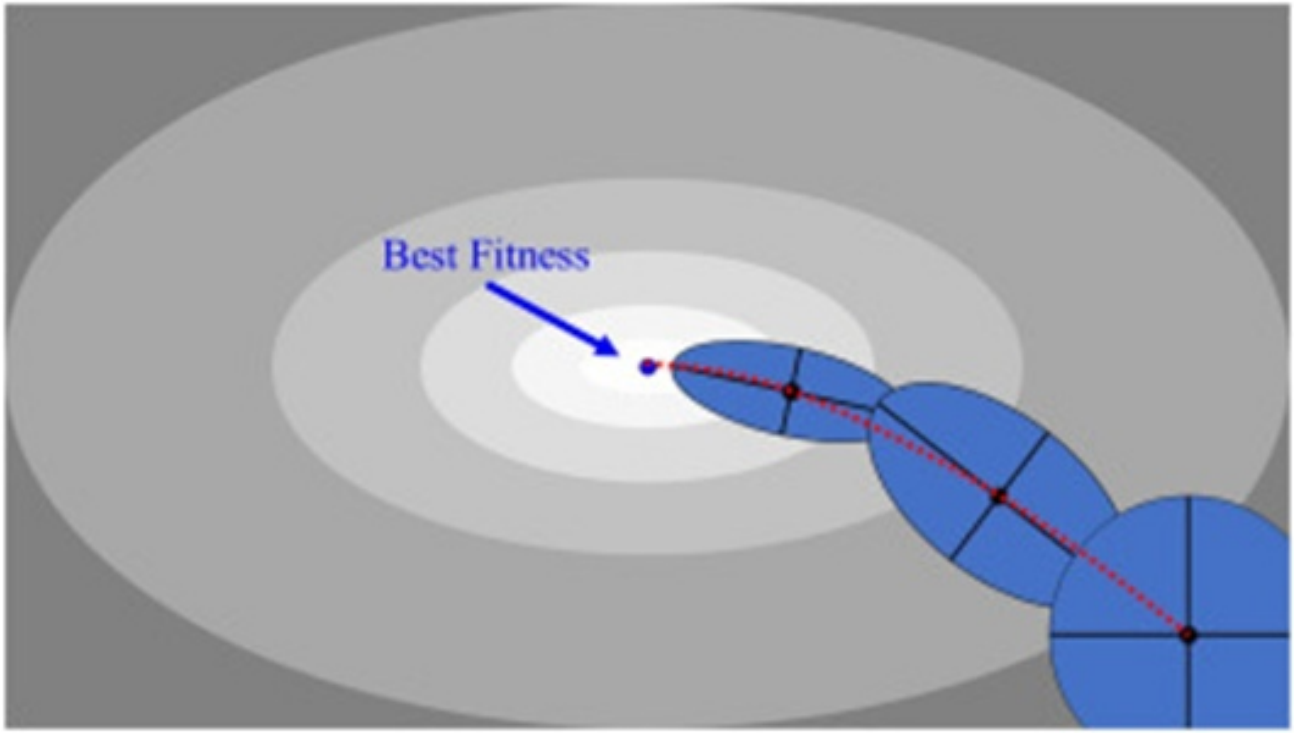


Figure 12. Correlated mutation.

Within the family of self-adaptive approaches, correlated mutation yields the most effective optimum-seeking capability. Its downside, however, lies in the profusion of adaptive parameters it must maintain (the real-valued vector itself, a distinct step size for every dimension, plus rotation angles bridging each dimensional pair), a burden that translates into ponderously slow convergence. Addressing this, Hansen and Ostermeier [25] presented in 2001 a reformulated Evolution

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

Strategy termed Covariance Matrix Adaptation Evolution Strategy (CMA-ES). CMA-ES substitutes the cumbersome parameter set of correlated mutation with a compact collection of adaptive quantities: the parental centroid, the covariance matrix, a global step size, and dedicated evolution paths. This innovation resolves the convergence bottleneck and unlocks an efficient search behavior comparable to correlated mutation. CMA-ES has subsequently found widespread application, with the algorithm undergoing successive rounds of enhancement; the most current edition was released in 2016 [26].

Additionally, optimization problems encountered in practical engineering tend to exhibit substantial complexity, as exemplified by multi-funnel function landscapes [27] that are riddled with local optima and lack any concave structural cues pointing reliably toward the global optimum. For this reason, most engineering applications content themselves with locating approximately optimal solutions rather than chasing the elusive global optimum.

Seeking to expeditiously uncover fan weightings that comply with temperature constraints and approximate energy-saving optimality, this investigation fields the multi-fan control framework. The framework leverages the weighting mechanism to drive fans asynchronously and harnesses CMA-ES to pinpoint near-optimal fan weightings, delivering simultaneous thermal compliance and energy-efficient operation.

A pair of fitness functions was purpose-built to judge fan weightings according to the results produced by the multi-fan control loop. When a candidate weighting set leads the control outcome to transgress the temperature bounds, its fitness is penalized through **Eq. 4**. The wider the gap separating the targeted and observed timing of the control response, the sooner that weighting set breaks the thermal constraints under the resulting control behavior. The trailing component in **Eq. 4,** namely “100% Fans Power,” communicates to the search algorithm that this weighting set is unfit for deployment in the multi-fan control scheme because it permits temperature violations. In the alternative case where the weighting set keeps temperatures within bounds, its fitness is quantified by **Eq. 5**. For a weighting set deemed compatible with the multi-fan control framework, its fitness amounts to the summed fan power draw observed under the resulting control behavior. The summation of fan power is obtained through **Eq. 1**. Furthermore, a reduced fitness score signals elevated quality for that fan weighting set.

$$f(\boldsymbol{x}_k) = (\ \text{IdealTime} - \text{ActualTime}\ ) + 100\%\ \text{FansPower} \quad (4)$$

$$f(\boldsymbol{x}_k) = \text{FansPower} \quad (5)$$

Within **Algorithm 1** and **Algorithm2**, Step 1 undertakes the initialization of all requisite parameters. Given that the fan weighting set is composed of *Weighting fan1, 2, 3, 4*—a formulation that maps onto a four-dimensional function optimization instance—n takes on the value 4. The count of offspring generated per cycle, λ, is assigned the value 8. The tally of candidates retained to form the parental pool, μ, is assigned the value 4, accompanied by the individual weighting coefficients (w1, w2, w3, w4) = (0.53, 0.29, 0.14, 0.04) as laid out in **Eqs. 6 and 7.** In this notation, wi denotes the individual weighting coefficient, which is keyed to the fitness-based rank of the i-th individual drawn from the μ parents. The symbol n captures the dimensionality of the problem. The symbol f represents the fitness function charged with scoring each offspring. The symbol xk designates the kth offspring, whose structure comprises the fan weight values. Various operational quantities—including the inaugural global step size σ0, the initial covariance matrix C(0), the initial parental centroid M(0), the initial evolution path tied to the covariance matrix P(0), the initial evolution path tied to the global step size P(0), and the ancillary constants μeff, c1, cc, cμ, cσ, and dσ—are populated with the default values endorsed by the developers of the algorithm. As remarked in Schwefel [24] and Hansen [26], these default settings are broadly general and well-suited to an extensive range of function optimization tasks. In **Algorithm 2**, Step 2 inspects whether the halting criterion has been satisfied. Step 3 draws on the parental centroid M(g), the prevailing global step size σ(g), and N(0, C(g)) to execute recombination followed by mutation, thereby generating λ fresh offspring, where the superscript g indexes the current generational cycle.

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

Within this description, N **(0**, $C^{(g)}$**)** signifies a Gaussian distribution characterized by a mean vector of zeros and covariance matrix $C^{(g)}$. Step 4 applies the fitness function to assign a fitness score to each freshly generated offspring. Steps 5 through 10 update assorted algorithmic state variables, including the parental centroid $M^{(g+1)}$, the covariance matrix evolution path $P^{(g+1)}$, the covariance matrix $C^{(g+1)}$, and other related quantities. It is worth emphasizing that $x^{(g+1)}$ corresponds to the ith highest-ranked offspring selected from among the λ offspring composing generation g+1, with the notation i: λ conveying the ith-ranked member within the set of λ offspring. The weighting coefficient wi mediates the selection pressure, thereby shaping the constitution of the parental pool. Steps 3 through 10 are repeated in a loop until the halting condition triggers, at which juncture the evolutionary search terminates in Step 11.

$$w_i = \frac{\ln(n+0.5) - \ln(i)}{\sum_{i=1}^{\mu} \ln(n+0.5) - \ln(i)} \text{ for } i = 1, \ldots, \mu \quad (6)$$

$$\sum_{i=1}^{\mu} w_i = 1 \quad (7)$$

Experimental Design

All test runs unfold under static server workload definitions. The temperature of the air surrounding the apparatus is regulated at a fixed 23 °C throughout. The subsystems that act as primary thermal sources—CPU0, CPU1, PCIe0, and PCIe1 as introduced—are commanded to dissipate at either 60% or 100% of their rated ceiling, producing a suite of eight discrete workload configurations, as enumerated in **Figure 13**. The thermal envelope imposes an identical 55 °C target on all primary thermal sources. During active thermal regulation, the momentary excursion of any sensed temperature above this target is capped at 5 °C, a safeguard intended to avert degradation stemming from thermal runaway. The settled temperature, once transients have decayed, is permitted to overshoot the target by no more than 2%; this bound defines the acceptable static offset. **Figure 14** lays out the spatial arrangement tying the primary thermal sources to the air movers housed within the enclosure. As regards the weighting apparatus, any given *Weighting fan* may adopt a value drawn from the discrete set {0.2, 0.4, 0.6, 0.8, 1.0}, subject to the restriction that unity must appear at least once across the four fans. This combinatorial constraint engenders a pool of 369 distinct fan weighting vectors. These are collated in **Figure 15** and constitute the entirety of fan weighting configurations available to the weighting apparatus during experimentation, permitting the fans to be driven in a non-uniform fashion. **Figure 16** captures the experimental protocol devised for non-uniform fan drive. The PID regulator is tuned with the triplet Kp = 15, Ki = 0.5, and Kd = 0.001. 400 seconds is earmarked for each fan regulation trial. From time zero to the 200-second mark, all fans receive an identical drive command. From 200 seconds onward until the trial concludes at 400 seconds, a candidate fan weighting vector is imposed, thereby switching the fans to non-uniform operation. The transient thermal response is assessed over the sub-interval spanning 200 to 300 seconds, whereas the static thermal condition is evaluated over the final 100 seconds, from 300 to 400 seconds. Whenever a candidate weighting vector breaches the thermal envelope, the trial is aborted, and that vector is branded as non-viable. Vectors that respect the thermal envelope throughout are conversely branded as viable. The experimental data facilitate a head-to-head comparison between uniform and non-uniform fan drive strategies, scrutinizing both the aggregate electrical burden of the fans and the temperature chronicles of the individual subsystems.

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

Initialize the CMA-ES Parameters

$n \leftarrow$ Number of dimensions

$\lambda \leftarrow$ Offspring Population size (4+floor(3ln(n)))

$\mu \leftarrow$ Parent Population for next generation (floor(λ/2))

$w_i \leftarrow$ Weight Value of i-th Individual

$f \leftarrow$ Function of Fitness

$\sigma^0 \leftarrow$ Initial Global Mutation Step Size

$\sigma^0 \leftarrow$ n dim. (uppest bound + lowest bound)/6

$\boldsymbol{C^0} \leftarrow$ Initial **Covariance Matrix** is $\mathbf{I}_{n \times n}$

$\boldsymbol{M^0} \leftarrow$ Initial **Mean** is **n dim. (uppest bound + lowest bound)/2**

$\boldsymbol{P_c^0} \leftarrow$ Initial **Cov. Matrix Evolution Path** is $[0,\ldots,0]^T_{1\times n}$

$\boldsymbol{P_\sigma^0} \leftarrow$ Initial **Mutation Step Size Evolution Path** is $[0,\ldots,0]^T_{1\times n}$

$\mu_{eff} \leftarrow$ Cov. Matrix & Mutation Step Size related parameter

$\mu_{eff} \leftarrow (\sum_i^\mu w_i^2)^{-1}$

$c_1, c_c, c_\mu \leftarrow$ Cov. Matrix related parameters

$c_1 \leftarrow 2((n+1.3)^2 + \mu_{eff})^{-1}$

$c_c \leftarrow (4+\mu_{eff}/n)(n+4+2\mu_{eff}/n)^{-1}$

$c_\mu \leftarrow \min(1-c_1, 2(\mu_{eff}-2+(\mu_{eff})^{-1})((n+2)^2+\mu_{eff})^{-1}$

$c_\sigma, d_\sigma \leftarrow$ Mutation Step Size related parameters

$c_\sigma \leftarrow (\mu_{eff}+2)(n+\mu_{eff}+5)^{-1}$

$d_\sigma \leftarrow 1+2\max(0, \sqrt{(\mu_{eff}-1)(n+1)^{-1}} - 1)(n+\mu_{eff}+5)^{-1}$

Algorithm 1. CMA-ES Pseudo Code - Part 1.

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

1 Initialize the CMA-ES Parameters (as Part1 show)

2 **while** stopping criterion is not met **do**

3 Generate **Offspring Population** for generation g+1

$$\boldsymbol{x}_k^{(g+1)} \sim \boldsymbol{M}^{(g)} + \sigma^g \boldsymbol{N}(\mathbf{0}, \boldsymbol{C}^{(g)}) \text{ for } k = 1, \ldots, \lambda$$

4 Evaluate Fitness

$$f(\boldsymbol{x}_k^{(g+1)})$$

5 Update **Mean** for generation g+1

$$\boldsymbol{M}^{(g+1)} \leftarrow \sum_{i=1}^{\mu} w_i \boldsymbol{x}_{i:\lambda}^{(g+1)}$$

6 Update **Cov. Matrix Evolution Path**

$$\boldsymbol{P}_C^{(g+1)} \leftarrow (1 - c_c)\boldsymbol{P}_C^{(g)} + c_c(2 - c_c\mu_{eff})^{0.5} (\boldsymbol{M}^{(g+1)} - \boldsymbol{M}^{(g)})(\sigma^g)^{-1}$$

7 Update **Cov. Matrix**

$$\boldsymbol{C}^{(g+1)} \leftarrow (1 - c_\mu)\boldsymbol{C}^{(g)} + c_1 \boldsymbol{P}_C^{(g+1)} (\boldsymbol{P}_C^{(g+1)})^T + c \sum_{i=1}^{\mu} w_i \boldsymbol{x}_{i:\lambda}^{(g+1)} w (\boldsymbol{x}^{(g+1)} - \boldsymbol{M}^{(g)})(\sigma^g)^{-1} [(\boldsymbol{x}_{i:\lambda}^{(g+1)} - \boldsymbol{M}^{(g)})(\sigma^g)^{-1}]^T$$

8 Update **Mutation Step Size Evolution Path**

$$\boldsymbol{P}_\sigma^{(g+1)} \leftarrow (1^\sigma - c_\sigma)\boldsymbol{P}_\sigma^{(g)} + c_\sigma(2 - c_\sigma)\mu_{eff} (\boldsymbol{C}^{(g)})^{-1/2} (\boldsymbol{M}^{(g+1)} - \boldsymbol{M}^{(g)})(\sigma^g)^{-1}$$

9 Update Mutation Step Size

$$\sigma^{g+1} \leftarrow \sigma^{g+1} \exp(c_\sigma d_\sigma^{-1} \| \boldsymbol{P}_\sigma^{(g+1)} \| (\mathrm{E}[\| \boldsymbol{N}(\mathbf{0}, \mathbf{I}) \|])^{-1} - 1)$$

10 Update best ever fitness solution

11 **end**

Algorithm 2. CMA-ES Pseudo Code - Part 2.

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

Case A		Case B		Case C		Case D	
PCIe1 100%	CPU1 100%	PCIe1 60%	CPU1 100%	PCIe1 100%	CPU1 100%	PCIe1 60%	CPU1 100%
PCIe0 100%	CPU0 100%	PCIe0 60%	CPU0 100%	PCIe0 60%	CPU0 60%	PCIe0 100%	CPU0 60%

Case E		Case F		Case G		Case H	
PCIe1 60%	CPU1 60%	PCIe1 100%	CPU1 60%	PCIe1 60%	CPU1 60%	PCIe1 100%	CPU1 60%
PCIe0 60%	CPU0 60%	PCIe0 100%	CPU0 60%	PCIe0 100%	CPU0 100%	PCIe0 60%	CPU0 100%

Figure 13. Eight cases of the server's loading scenarios.

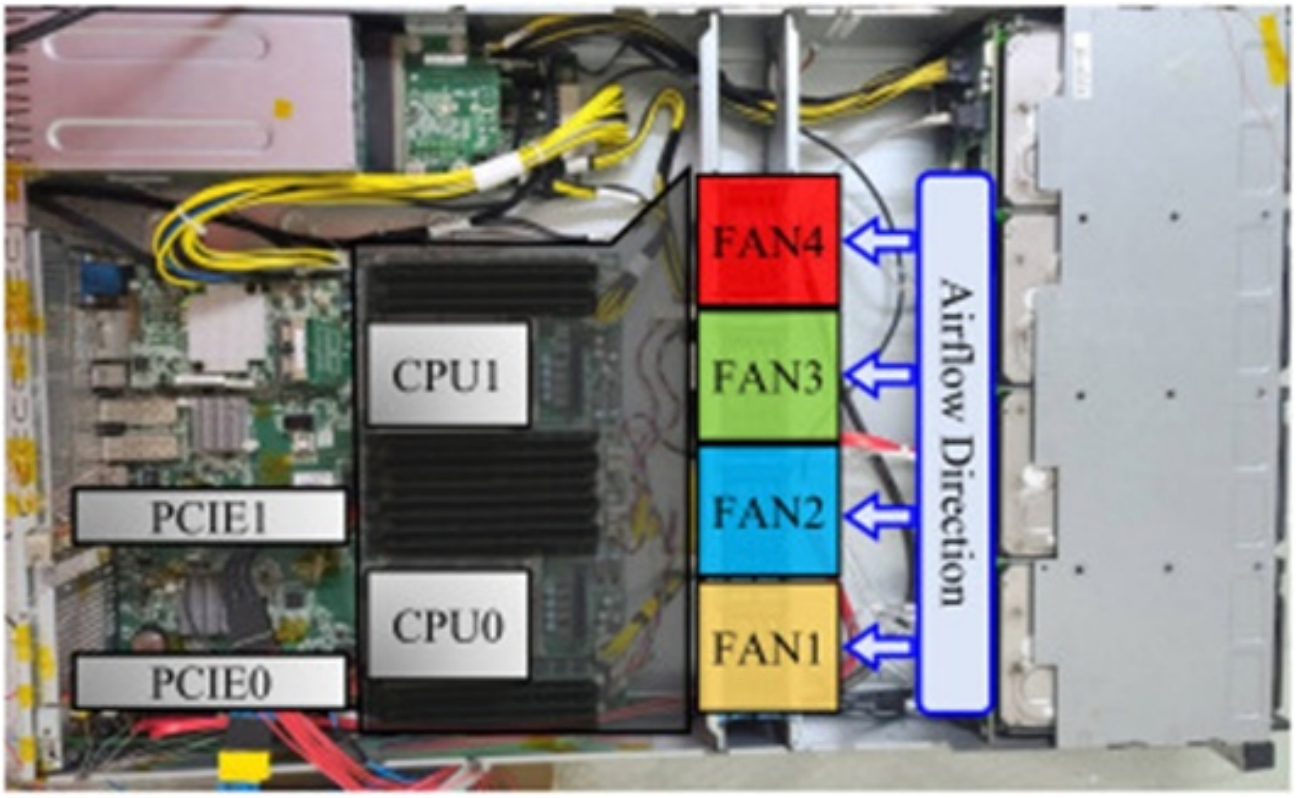
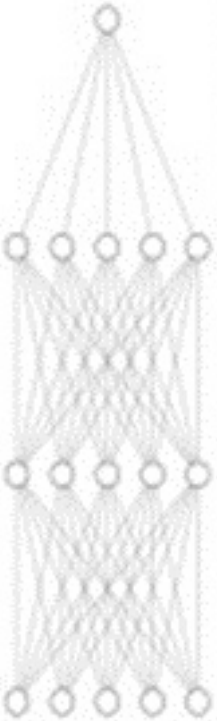


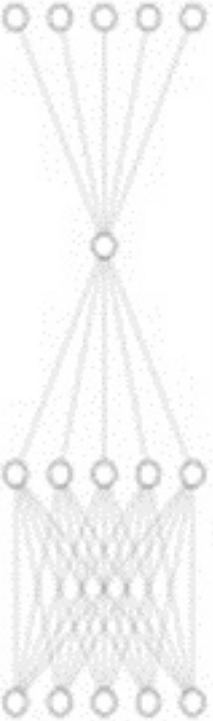
Figure 14. Schematic diagram of the server critical components and fans.

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

	Weighting
fan1	1
fan2	0.2, 0.4, 0.6, 0.8, 1
fan3	0.2, 0.4, 0.6, 0.8, 1
fan4	0.2, 0.4, 0.6, 0.8, 1



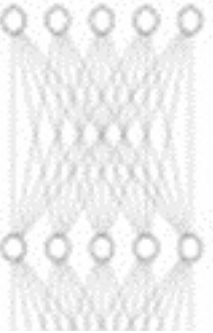
	Weighting
fan1	0.2, 0.4, 0.6, 0.8, 1
fan2	1
fan3	0.2, 0.4, 0.6, 0.8, 1
fan4	0.2, 0.4, 0.6, 0.8, 1



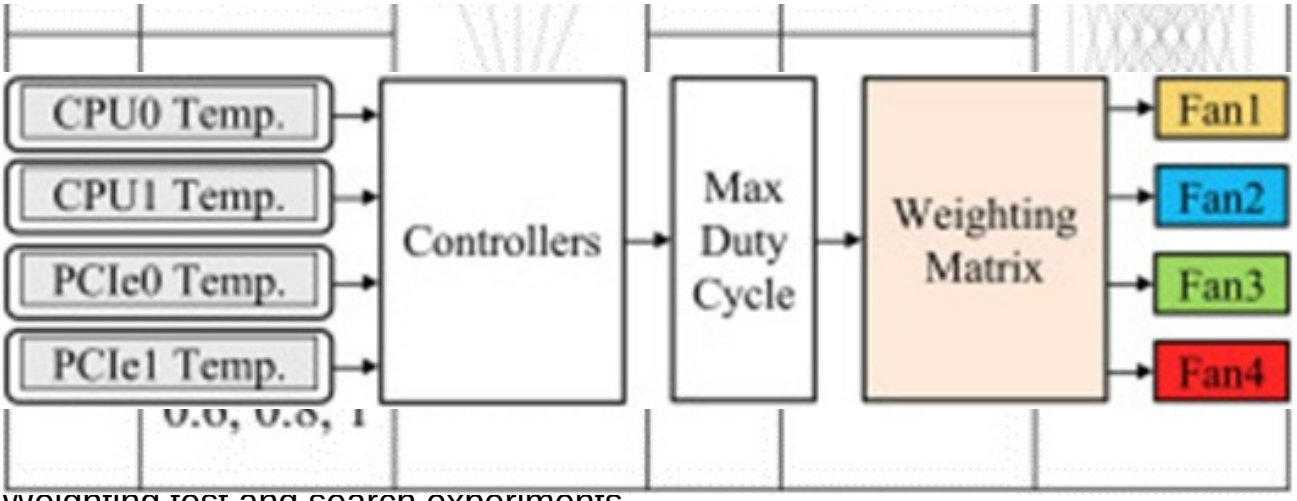
	Weighting
fan1	0.2, 0.4, 0.6, 0.8, 1
fan2	0.2, 0.4, 0.6, 0.8, 1



	Weighting
fan1	0.2, 0.4, 0.6, 0.8, 1
fan2	0.2, 0.4, 0.6, 0.8, 1



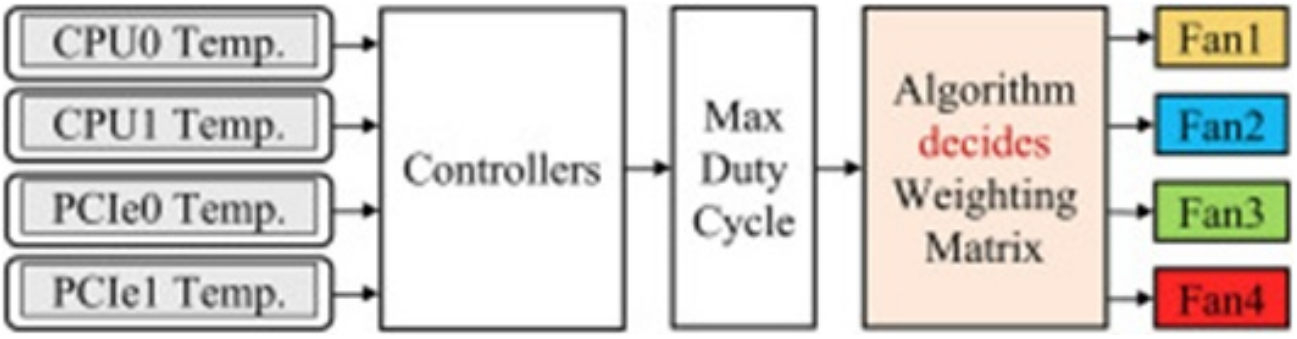
© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press



Weighting test and search experiments

The weighting test experiments subject the exhaustive catalogue of 369 fan weighting vectors to each of the eight workload configurations. The collected observations capture metrics such as the absolute number of viable vectors, the fraction they represent, the single vector yielding the lowest fan energy draw, and the percentage reduction in fan energy relative to the uniform drive baseline, thereby lending empirical weight to the case for non-uniform fan drive. Concurrently, this body of work posits a conjecture: that the primary thermal sources exhibit mutually consistent thermal behaviour. By placing side by side the temperature evolutions logged under uniform and non-uniform fan drive, the thermal interplay at each source is dissected. The motivation is twofold—first, to reinforce the plausibility of the conjecture, and second, to tease out the physical rationale that underpins the divergence in optimal fan weighting vectors observed across certain workload configurations.

The multi-fan regulation framework advocated in this study knits together three ingredients: the weighting apparatus, the evolutionary strategy optimiser, and the bespoke fitness measures. With these components integrated, the weighting search experiments are pursued. Under each of the eight workload configurations, the framework is presented with the full universe of 369 fan weighting vectors. The optimiser dictates which vectors actually undergo testing during the fan regulation procedure; after each test, the vector in question receives a fitness score. Vectors classed as non-viable are penalised as prescribed by **Eq. 4.** In contrast, those classed as viable are assessed through **Eq. 5**. Driven by the sequence of fitness evaluations, the optimiser continuously adjusts the heading of its exploration through the weighting space, homing in on near-optimal vectors after a restrained number of trials. Deploying these vectors for non-uniform fan drive simultaneously guarantees adherence to the thermal envelope and secures near-optimal parsimony in fan energy expenditure. The arrangement of the multi-fan regulation framework is schematised in **Figure 17.** Recognising the intrinsic randomness woven into the optimiser, each experimental condition is repeated across thirty independent runs to dampen statistical scatter when characterising true performance. A panel of statistical indicators—among them Euclidean distance, the best fitness value logged, the count of search steps consumed, the fraction of viable weightings encountered, the distribution of aggregate fan energy draws, and the percentage energy savings—is assembled, together building a case for the pronounced effectiveness of the multi-fan regulation framework herein proposed.



© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

Figure 17. The server multi-fan control system.

Results and Discussion

Weighting test experiments and analysis of thermal phenomenon

For the set of eight workload configurations depicted in **Figure 13**, weighting test experiments were conducted by exercising every one of the 369 fan weighting vectors catalogued in **Figure 15**. Working strictly within the thermal envelope prescribed, the tally and the proportion of viable vectors were catalogued. Among the viable vectors, the one that incurred the smallest aggregate fan energy draw was singled out as the foremost vector. **Tables 2 and 3** collate the percentage reductions in fan energy that non-uniform drive achieves relative to the uniform baseline. The regulation outcomes tied to the foremost vectors for each workload case, labelled A through H, are spread across **Figures 18-25;** within each figure, panel (a) traces the temperature evolution while panel (b) plots the fan drive signals. Furthermore, the collection of figures from **Figures 18-25** presents the static operational point data—subsystem temperatures, fan drive ratios, and aggregate fan energy draw—under both drive philosophies.

Table 2. Weight test experimental results — fan weightings' statistics.

Scenario	Evaluation of 369 Fan Coefficients		
	Optimal coefficient set [Fan1, Fan2, Fan3, Fan4]	Percentage of suitable coefficient sets	Count of suitable coefficient sets
A	[1.0, 0.8, 0.4, 1.0]	11.4%	42
B	[1.0, 0.8, 0.4, 1.0]	13.0%	48
C	[0.8, 0.6, 0.6, 1.0]	20.9%	77
D	[0.8, 0.6, 0.6, 1.0]	25.9%	95
E	[1.0, 0.6, 0.8, 1.0]	34.5%	127
F	[1.0, 0.6, 0.8, 1.0]	30.7%	113
G	[1.0, 1.0, 1.0, 0.6]	9.6%	35
H	[1.0, 1.0, 0.8, 0.6]	11.3%	42

Table 3. Weight test experimental results — steady-state data analysis.

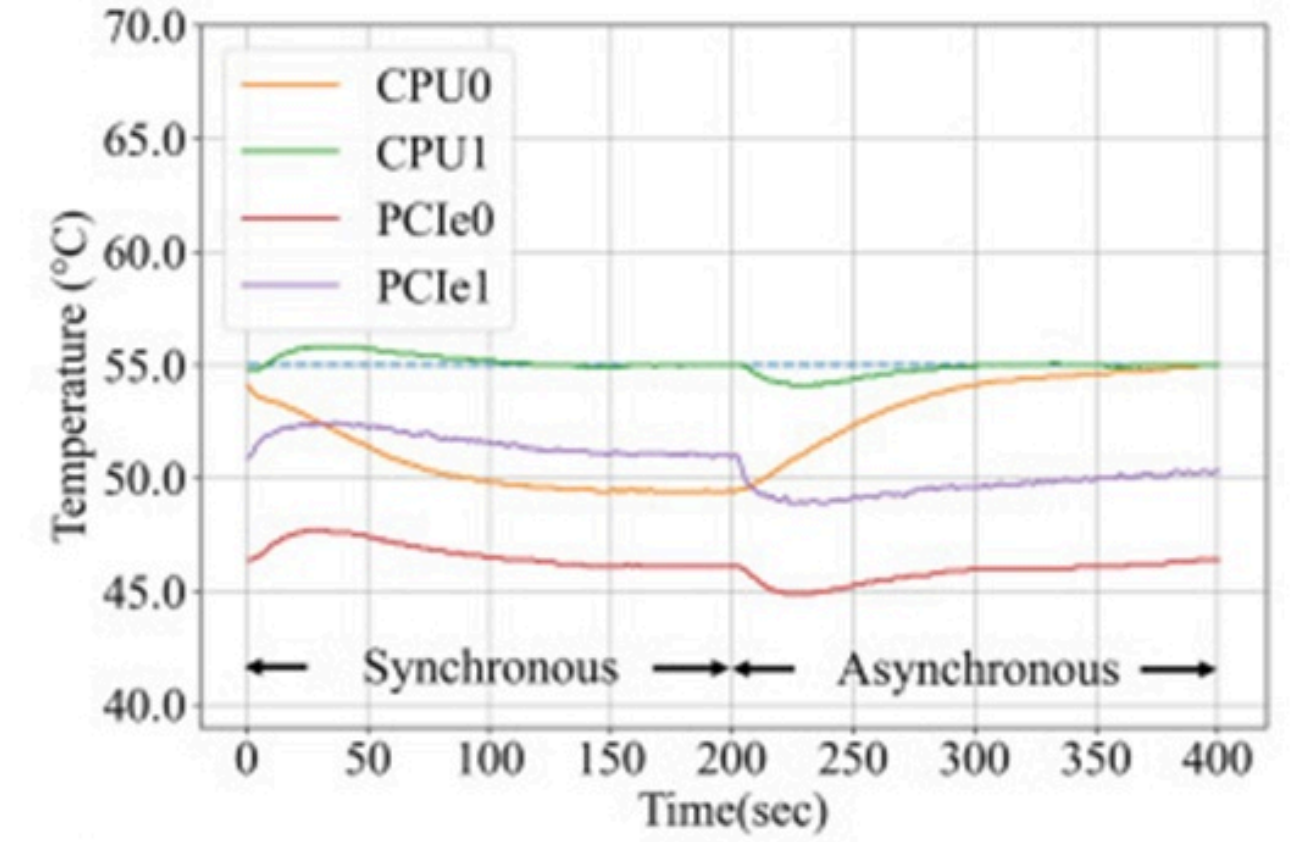
Scenario	Synchronized Fan Operation	Optimized Multi-Fan Operation	
	Power Consumption (W)	Power Savings (relative to synchronized)	Power Consumption (W)
A	41.6	45.5%	22.7
B	41.6	45.5%	22.7
C	41.6	59.4%	16.9
D	41.6	59.4%	16.9
E	21.7	42.5%	12.5

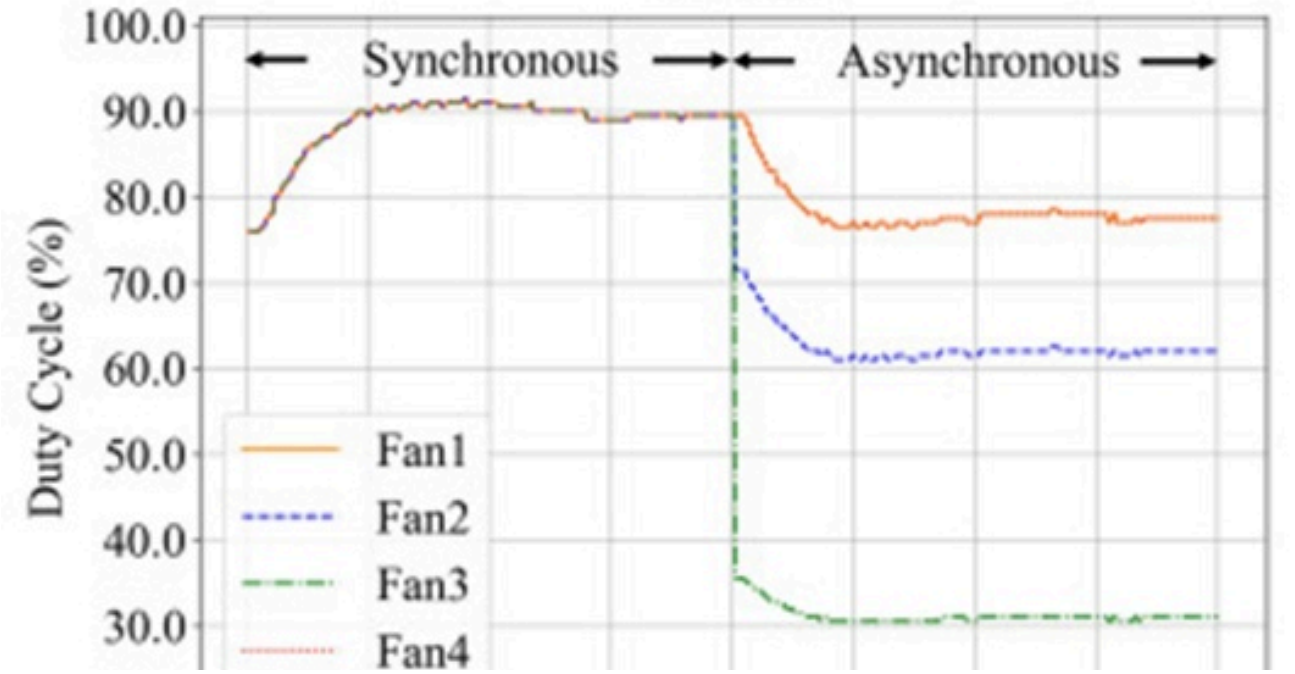
© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

F	24.7	46.6%	13.2
G	23.3	22.1%	18.3
H	28.0	45.3%	15.3

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

Modulation of multiple fans	Components (°C)				Fans duty cycle (%)				Fans power (W)
	CPU0	CPU1	PCIe0	PCIe1	Fan1	Fan2	Fan3	Fan4	
Synchronous	49.4	55.0	46.1	50.9	89.5	89.5	89.5	89.5	41.6
Asynchronous	55.0	55.0	46.4	50.3	77.5	62.0	31.0	77.5	22.7





© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

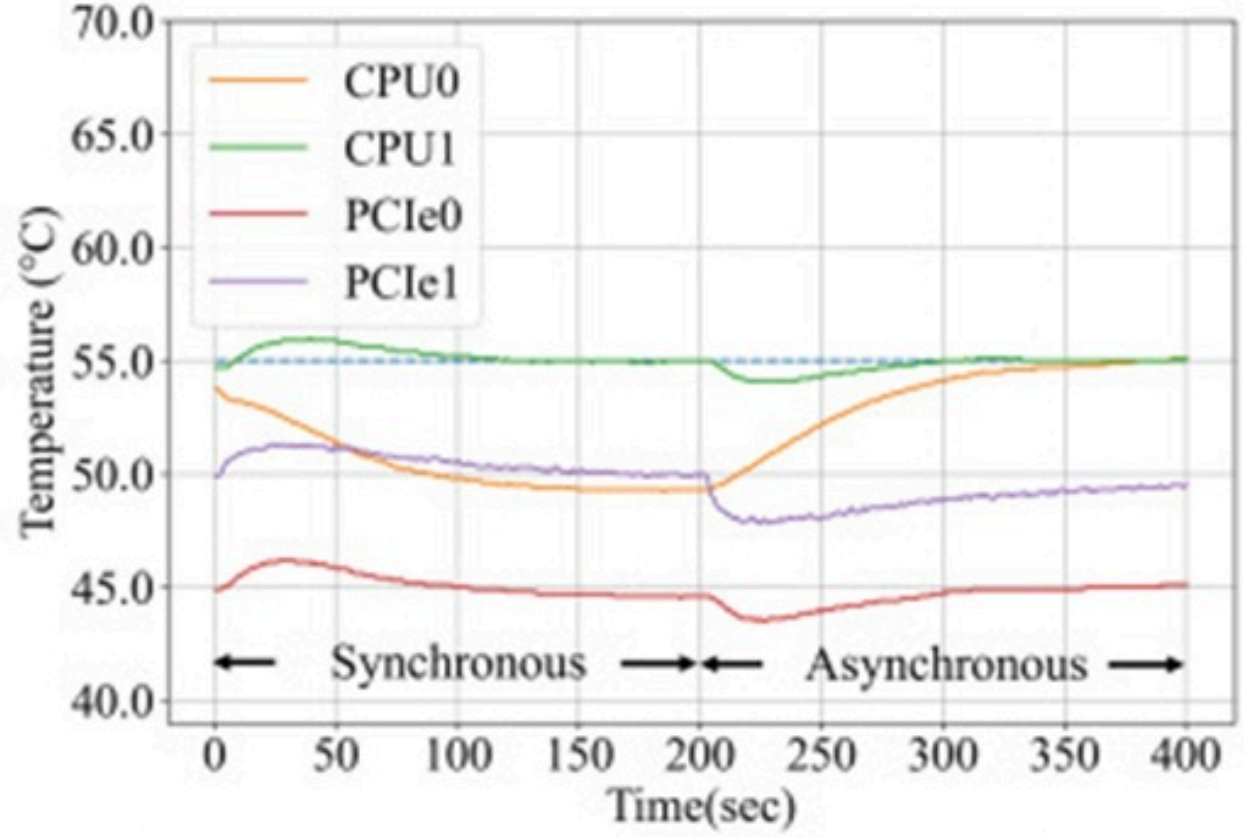


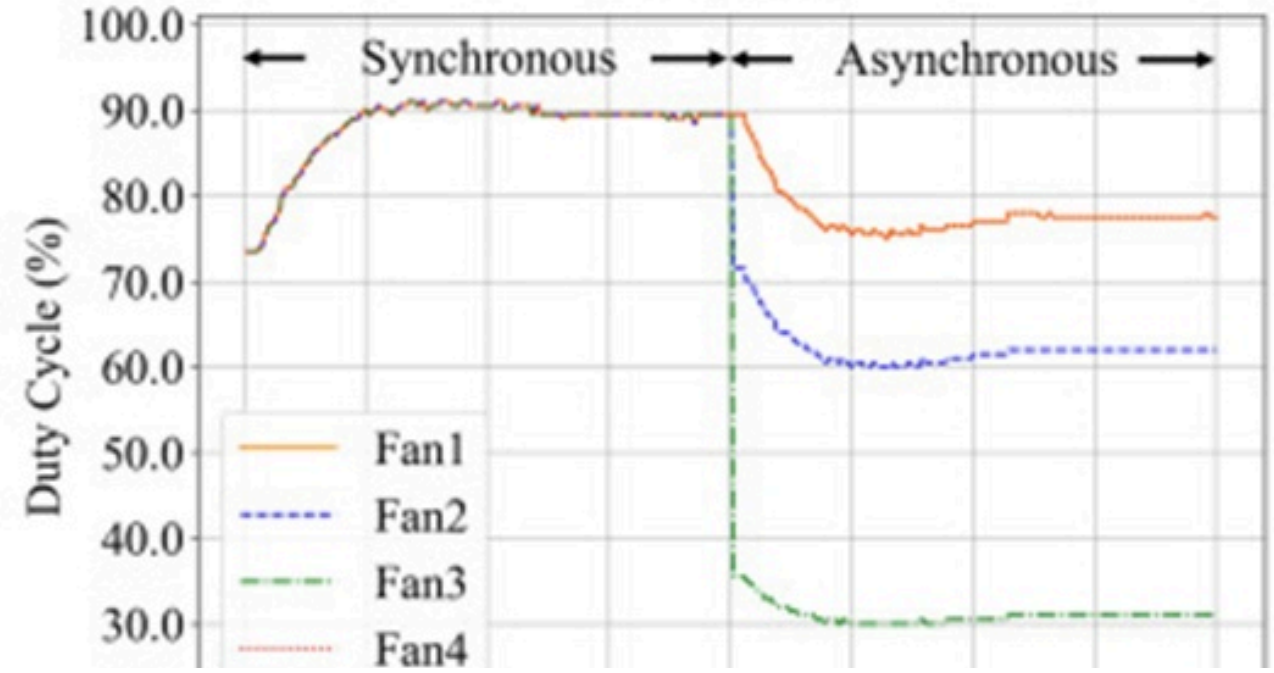
55.0, 46.1, and 50.9 °C, in that order. Upon switching to non-uniform drive (200–400 seconds) with the foremost vector [1.0, 0.8, 0.4, 1.0] enforced, the drive signals dispatched to Fan1, Fan2, Fan3, and Fan4 eventually settled at 77.5%, 62.0%, 31.0%, and 77.5%, respectively. The non-uniform drive regime produced static temperatures of 55.0, 55.0, 46.4, and 50.3 °C at CPU0, CPU1, PCIe0, and PCIe1. Placed alongside the uniform drive baseline, the non-uniform regime simultaneously alleviated the gratuitous sub-cooling of CPU0. It shrank the total fan electrical burden from 41.6 W to 22.7 W, corresponding to a 45.5% curtailment in fan energy.

Shifting attention to Case B as rendered in **Figure 19**, the uniform drive interval (0–200 seconds) again saw Fan1 through Fan4 converge to 89.5% duty. The static temperatures that materialised were 49.3, 55.0, 44.6, and 49.9 °C for CPU0, CPU1, PCIe0, and PCIe1. The subsequent non-uniform interval (200–400 seconds),

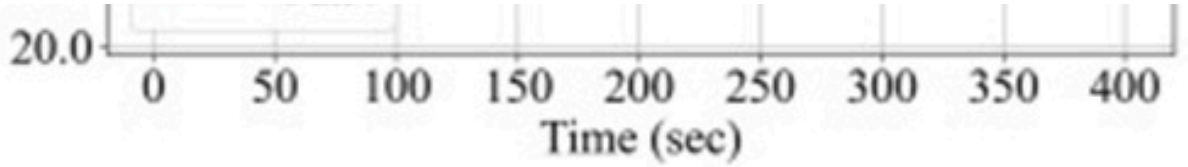
© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

Modulation of multiple fans	Components (°C)				Fans duty cycle (%)				Fans power (W)
	CPU0	CPU1	PCIe0	PCIe1	Fan1	Fan2	Fan3	Fan4	
Synchronous	49.3	55.0	44.6	49.9	89.5	89.5	89.5	89.5	41.6
Asynchronous	55.1	55.1	45.1	49.5	77.5	62.0	31.0	77.5	22.7





© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

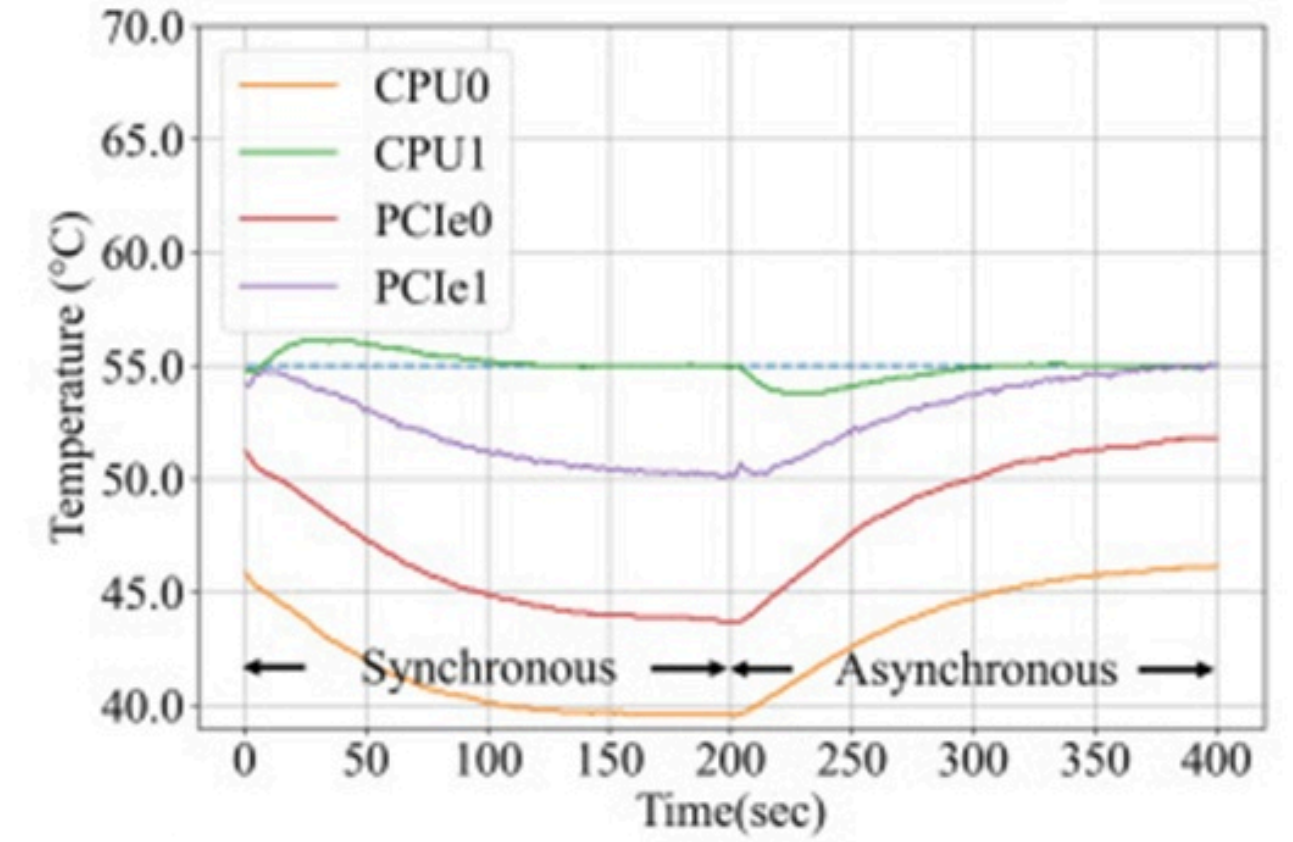


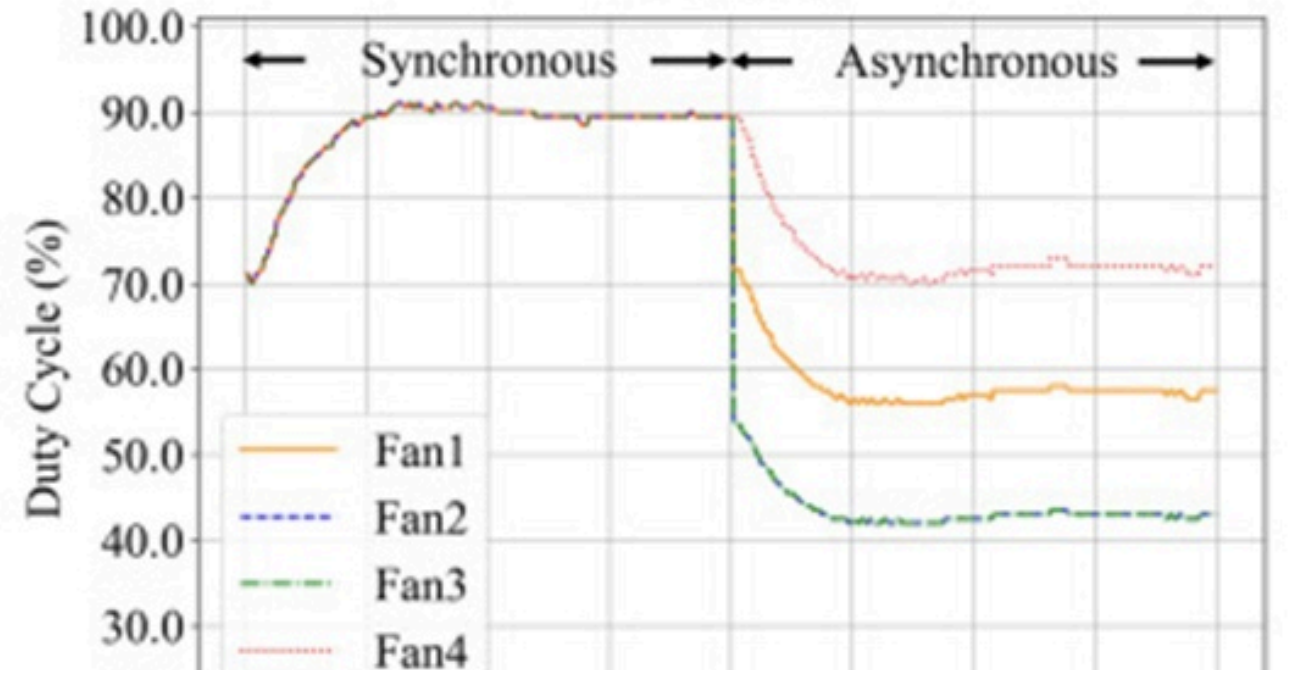
temperature. It cut the combined fan electrical demand identically from 41.6 W to 22.7 W, once more a 45.5% saving.

Case C, documented in **Figure 20**, unfolds with the uniform drive interval (0–200 seconds) stabilising all four fans at 89.5%. The static temperature picture that emerged was 39.6, 55.0, 43.7, and 50.2 °C for CPU0, CPU1, PCIe0, and PCIe1. The non-uniform interval (200–400 seconds), steered by the foremost vector [0.8, 0.6, 0.6, 1.0], settled Fan1, Fan2, Fan3, and Fan4 at 57.5%, 43.0%, 43.0%, and 72.0%. The ensuing static temperatures reached 46.2, 55.0, 51.8, and 55.1 °C at the four measurement points. Relative to uniform drive, the non-uniform strategy reduced needless cooling on CPU0, PCIe0, and PCIe1 and compressed the aggregate fan electrical draw from 41.6 W down to 16.9 W, a reduction of 59.4%.

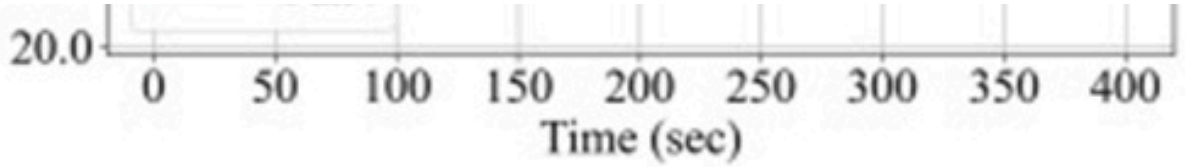
© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

Modulation of multiple fans	Components (°C)				Fans duty cycle (%)				Fans power (W)
	CPU0	CPU1	PCIe0	PCIe1	Fan1	Fan2	Fan3	Fan4	
Synchronous	39.6	55.0	43.7	50.2	89.5	89.5	89.5	89.5	41.6
Asynchronous	46.2	55.0	51.8	55.1	57.5	43.0	43.0	72.0	16.9





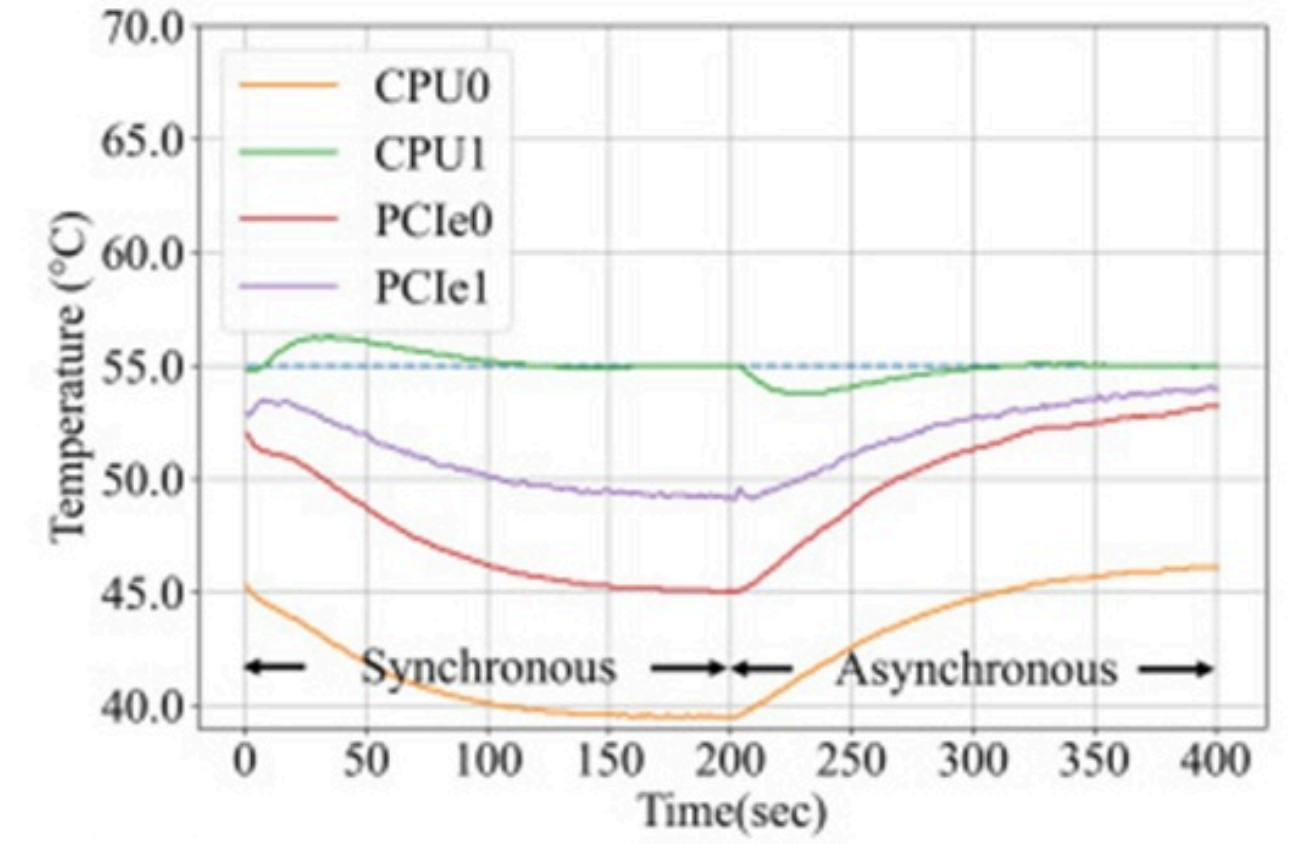
© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

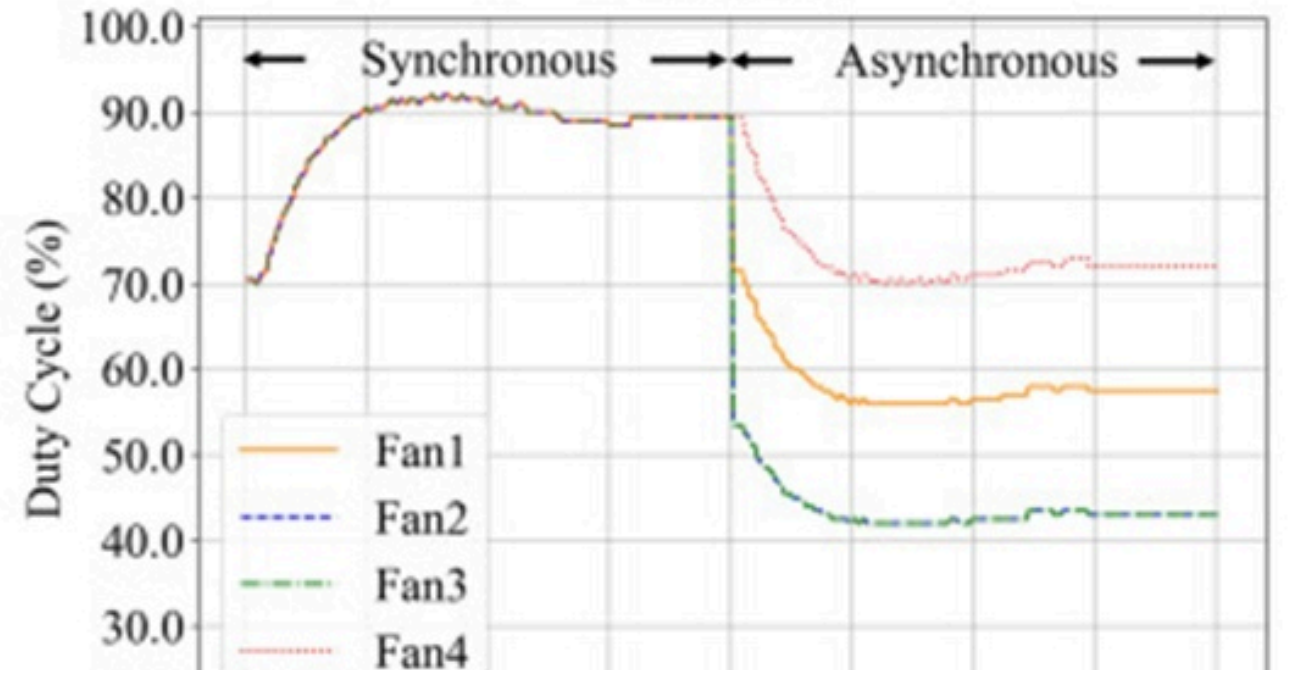


45.0%, 45.0%, and 72.0%. The static temperatures recorded under this regime were 40.1, 35.0, 33.2, and 34.0 °C. The non-uniform tactic once more lessened the unnecessary thermal suppression at CPU0, PCIe0, and PCIe1, and shrank the combined fan electrical draw from 41.6 W to 16.9 W, again a 59.4% improvement.

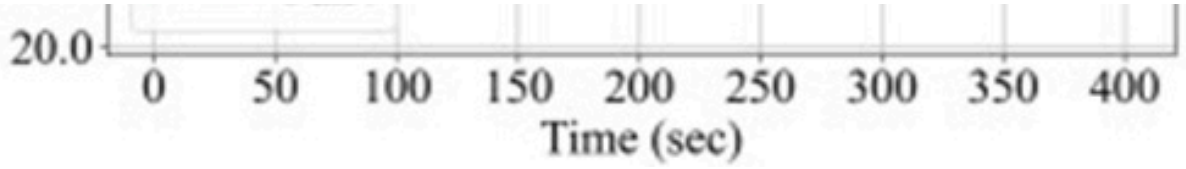
© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

Modulation of multiple fans	Components (°C)				Fans duty cycle (%)				Fans power (W)
	CPU0	CPU1	PCIe0	PCIe1	Fan1	Fan2	Fan3	Fan4	
Synchronous	39.5	55.0	45.1	49.2	89.5	89.5	89.5	89.5	41.6
Asynchronous	46.1	55.0	53.2	54.0	57.5	43.0	43.0	72.0	16.9



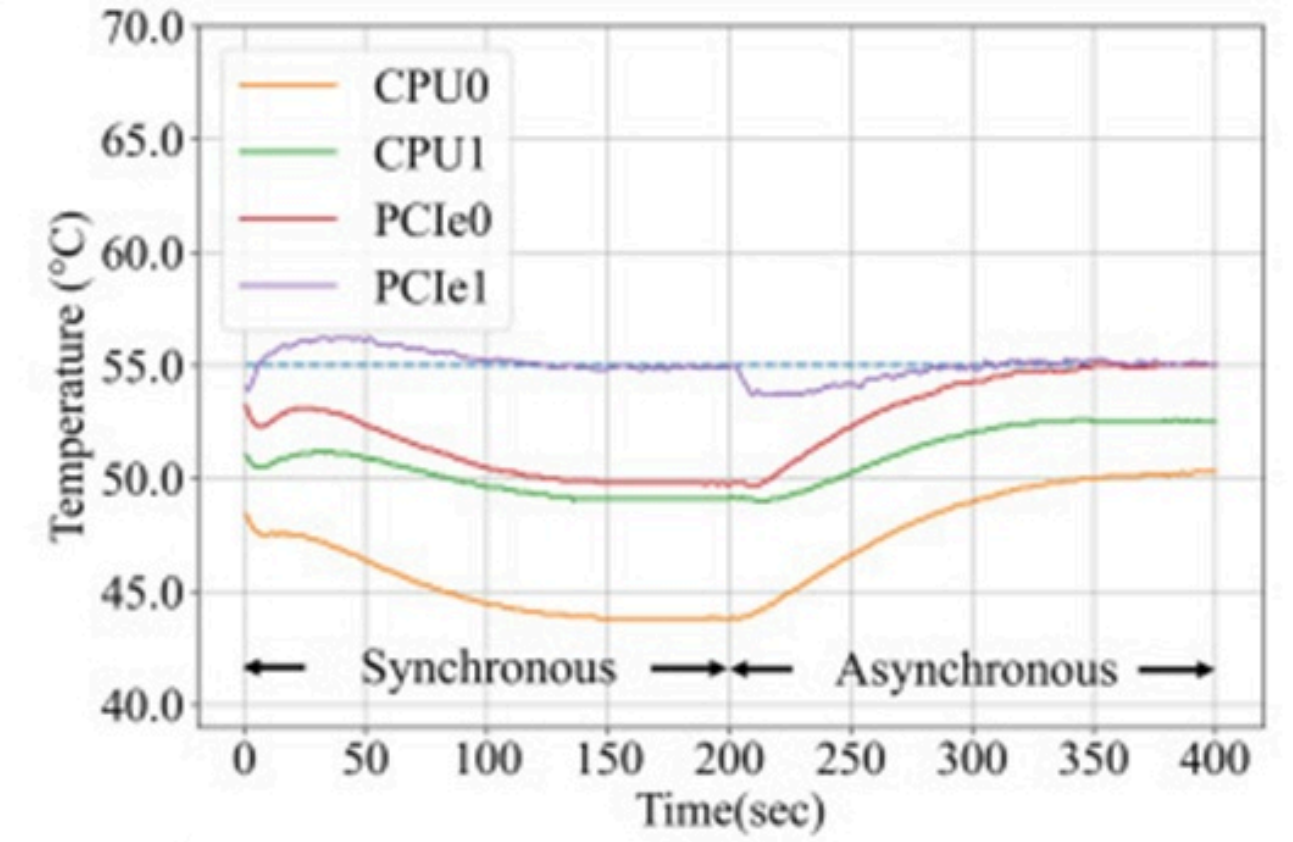


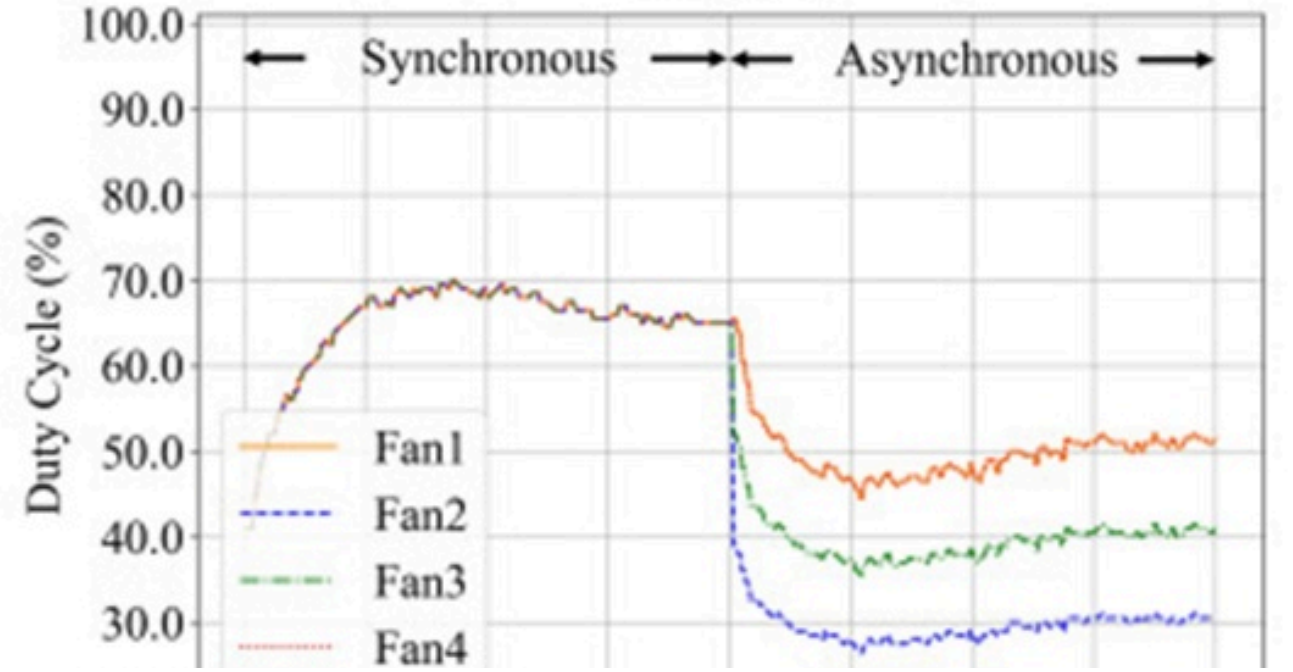
© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press



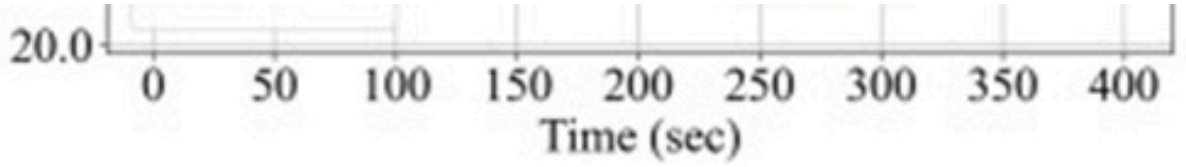
© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

Modulation of multiple fans	Components (°C)				Fans duty cycle (%)				Fans power (W)
	CPU0	CPU1	PCIe0	PCIe1	Fan1	Fan2	Fan3	Fan4	
Synchronous	43.8	49.1	49.7	54.9	65.0	65.0	65.0	65.0	21.7
Asynchronous	50.3	52.5	55.0	55.1	51.5	30.5	41.0	51.5	12.5





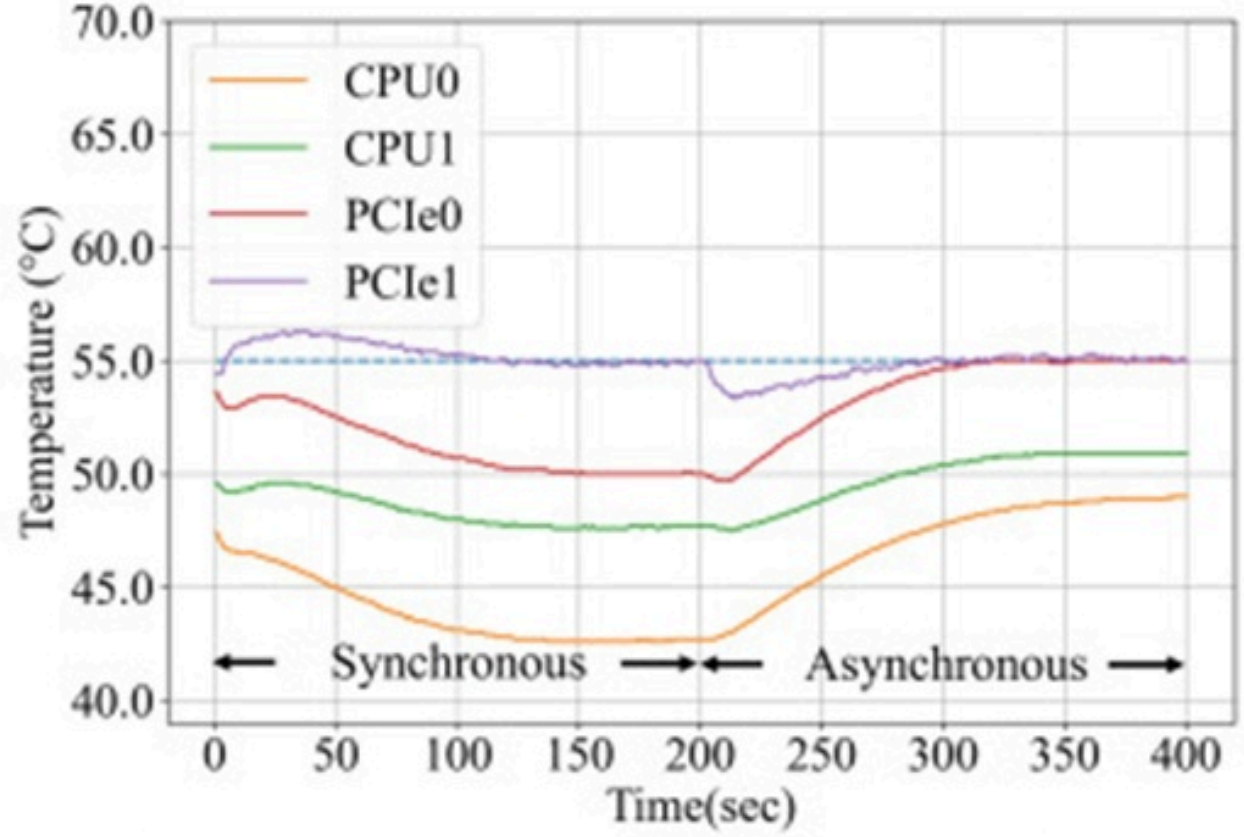
© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

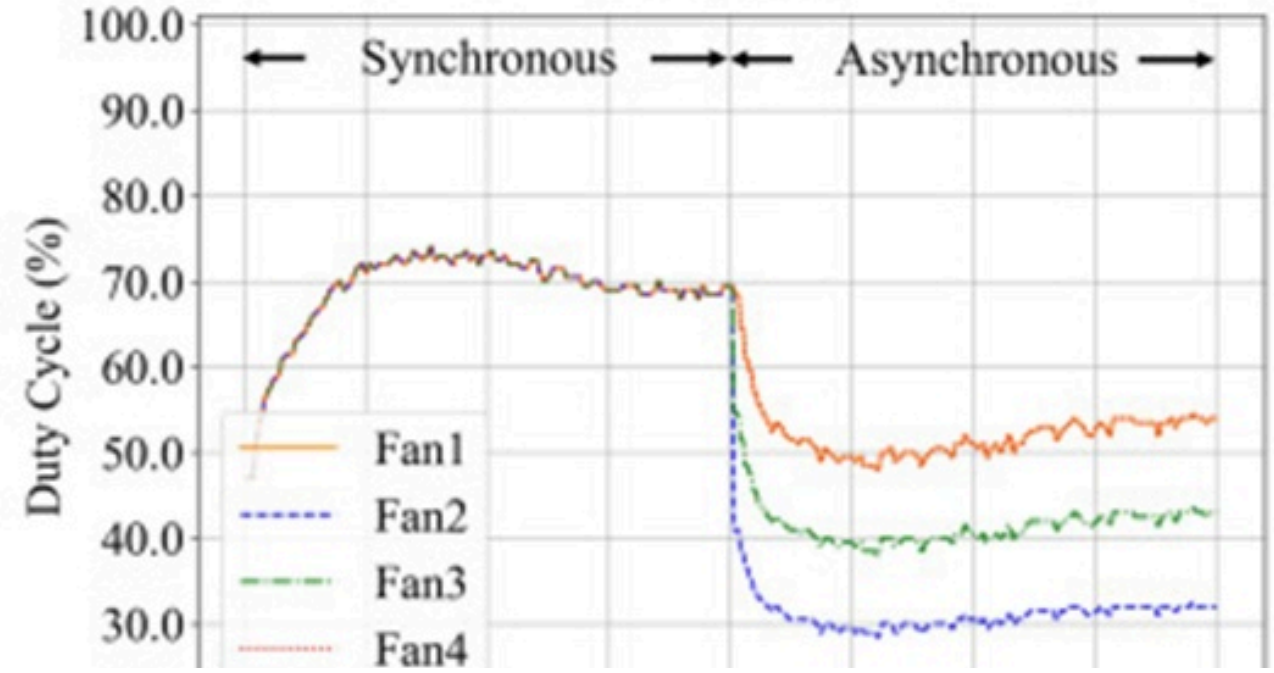


Finally, Case F, shown in **Figure 23**, began with the uniform drive interval (0–200 seconds) stabilising the fan duties at 69.5%. The static temperatures stood at 42.7, 47.7, 50.0, and 55.0 °C. The ensuing non-uniform interval (200–400 seconds), governed by [1.0, 0.6, 0.8, 1.0],

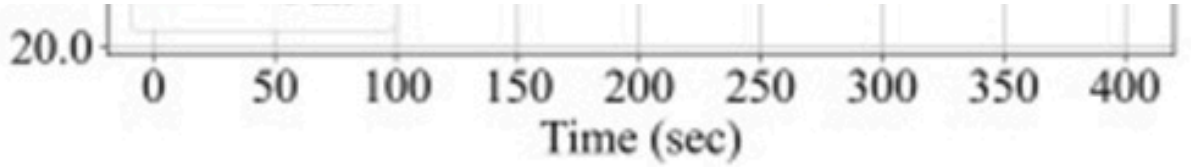
© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

Modulation of multiple fans	Components (°C)				Fans duty cycle (%)				Fans power (W)
	CPU0	CPU1	PCIe0	PCIe1	Fan1	Fan2	Fan3	Fan4	
Synchronous	42.7	47.7	50.0	55.0	69.5	69.5	69.5	69.5	24.7
Asynchronous	49.0	50.9	55.0	55.0	54.0	32.0	43.0	54.0	13.2





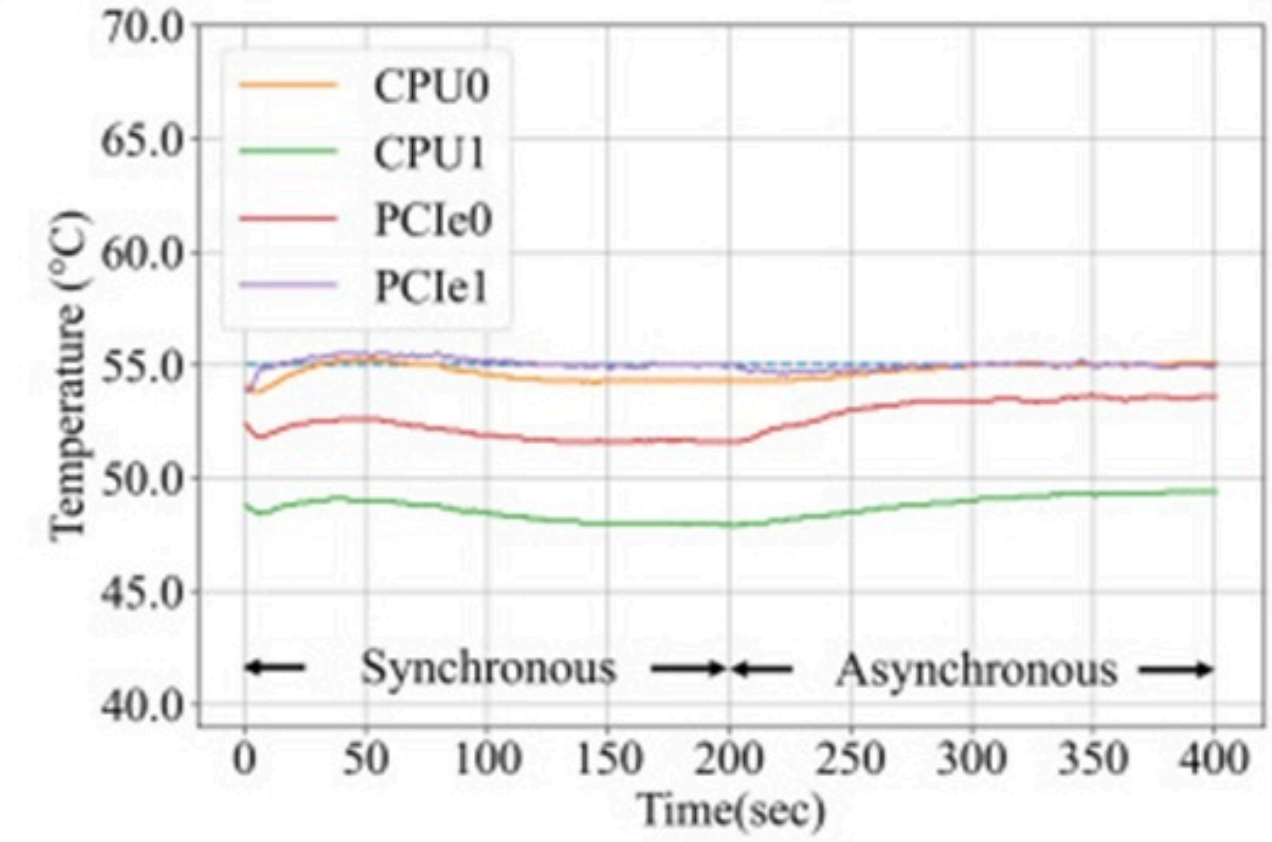
© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

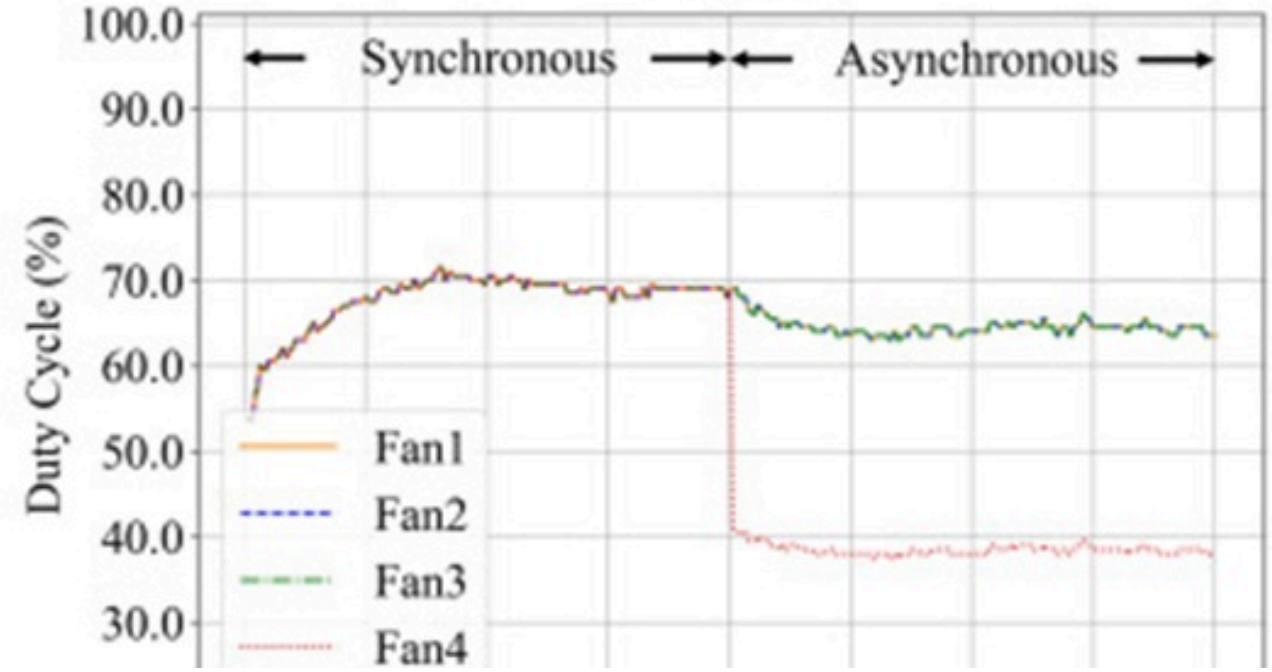


Turning attention to **Figure 24**, the data from Case G reveal that during the interval where all fans receive identical commands (0–200 seconds), the drive percentages for Fan1, Fan2, Fan3, and Fan4 eventually plateaued at 67.5%. The temperature readings once stabilised were 54.3 °C at CPU0, 47.8 °C at CPU1, 51.6 °C at PCIe0, and 54.9 °C at PCIe1. Shifting to the phase where fans operate under distinct commands (200–400 seconds) with the top-ranked vector [1.0, 1.0, 1.0, 0.6], the drive percentages settled at 63.5%, 63.5%, 63.5%, and 38.0% for Fan1 through Fan4, respectively. This independent operation mode brought the stabilised temperatures to 55.1, 49.2, 53.9, and 55.0 °C for the four components in order. Putting the two modes side by side, the independent scheme reduced the combined fan wattage from 23.3 W to 18.3 W, a percentage reduction of 22.1%.

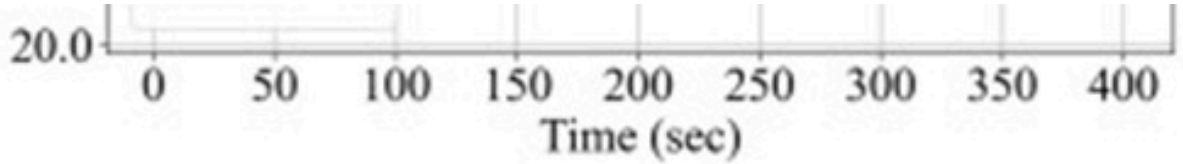
© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

Modulation of multiple fans	Components (°C)				Fans duty cycle (%)				Fans power (W)
	CPU0	CPU1	PCIe0	PCIe1	Fan1	Fan2	Fan3	Fan4	
Synchronous	54.3	47.8	51.6	54.9	67.5	67.5	67.5	67.5	23.3
Asynchronous	55.1	49.2	53.9	55.0	63.5	63.5	63.5	38.0	18.3





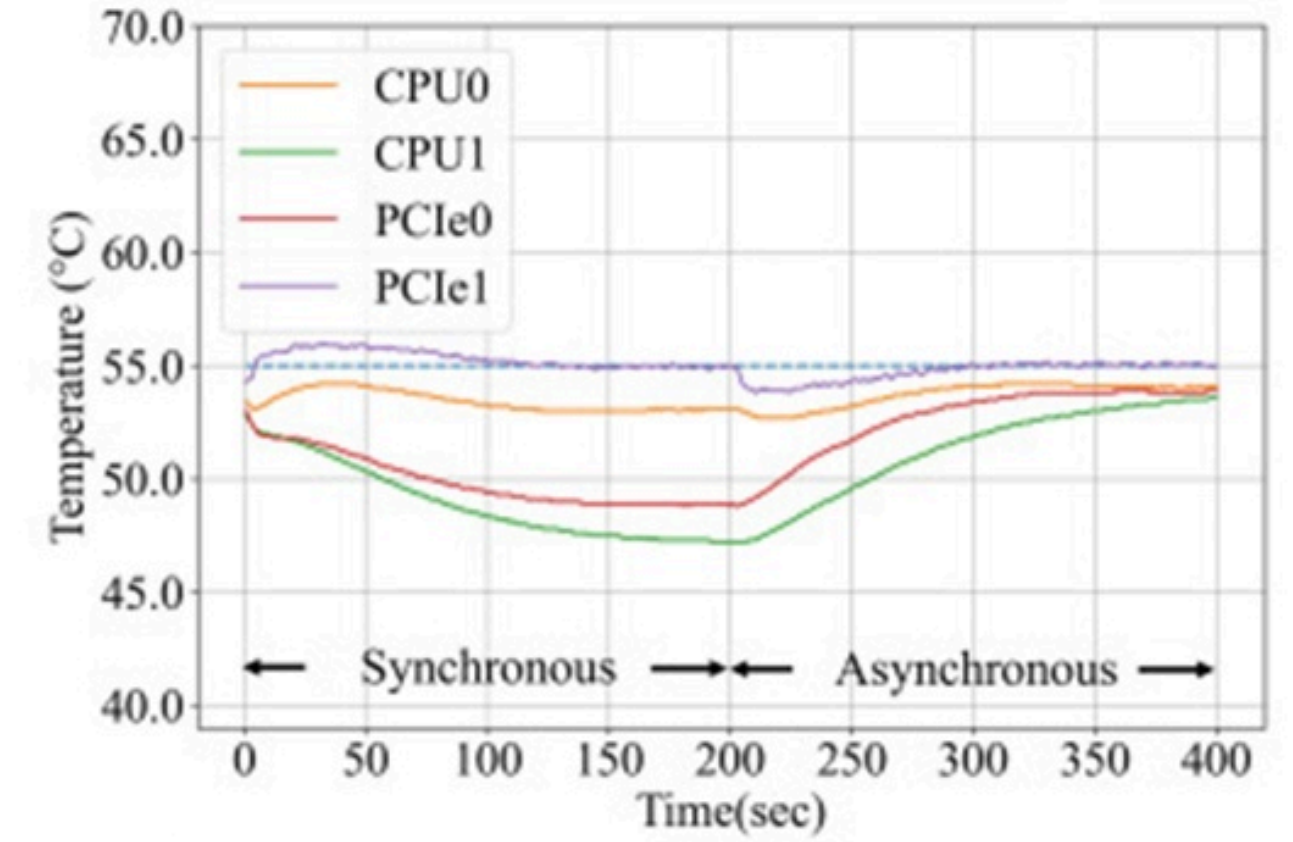
© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

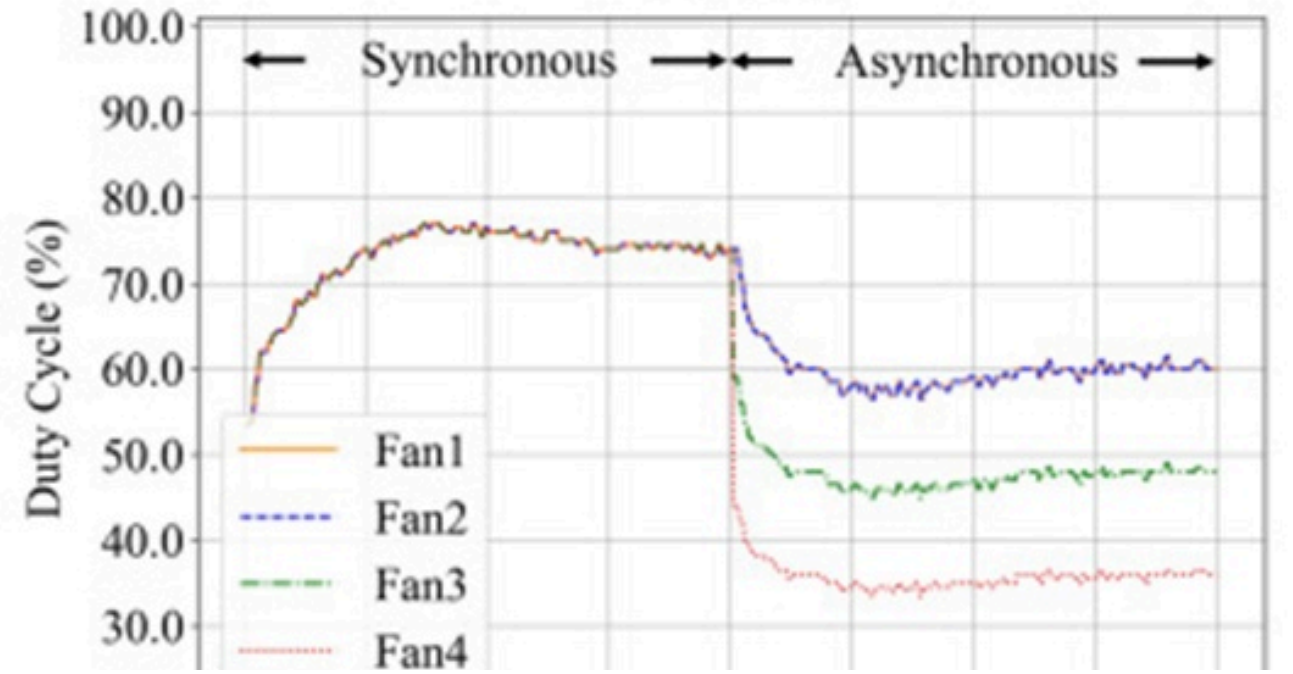


imposing the top-ranked vector [1.0, 1.0, 0.8, 0.6], Fan1 through Fan4 settled at 60.0%, 60.0%, 48.0%, and 36.0%. The resulting equilibrium temperatures were 54.1, 53.6, 53.9, and 55.0 °C in the same component order. Relative to synchronised commands, the independent strategy lowered the aggregate fan power draw from 28.0 W to 15.3 W, equivalent to a 45.3% saving.

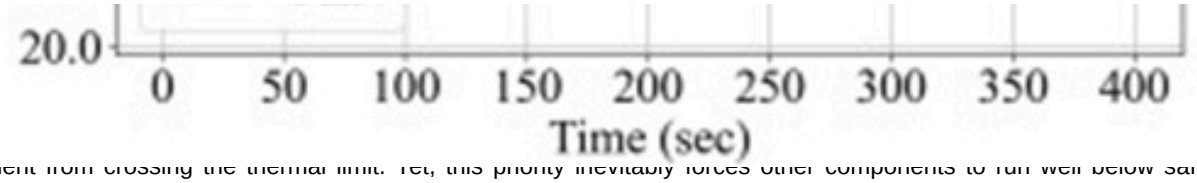
© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

Modulation of multiple fans	Components (°C)				Fans duty cycle (%)				Fans power (W)
	CPU0	CPU1	PCIe0	PCIe1	Fan1	Fan2	Fan3	Fan4	
Synchronous	53.1	47.2	48.8	55.0	74.0	74.0	74.0	74.0	28.0
Asynchronous	54.1	53.6	53.9	55.0	60.0	60.0	48.8	36.0	15.3





© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press



component from crossing the thermal limit. Yet, this priority inevitably forces other components to run well below safe thresholds, squandering fan energy in the process. For Cases A through D, the superfluously cooled components are CPU0, PCIe0, and PCIe1. For Cases E through H, they are CPU0, CPU1, and PCIe0. Adopting independent fan commands steered by the optimal vector allows the feedback loop to dial back its ceiling drive output. By operating the fans at non-uniform levels, meaningful curtailment of aggregate fan energy is realised, and the degree of needless thermal suppression imposed on certain components is softened. All temperature readings are held under the 55 °C cap. The eventual temperatures, fan drive levels, and aggregate fan power draw under both operating philosophies are governed by the dissipation intensities of the respective components. This causal link explains the observation that the set of permissible vectors, the percentage they represent, and the identity of the single best vector all shift when the workload configuration changes.

Within this portion of the study, having consulted the physical positioning of the principal heat sources relative to the air movers diagrammed in **Figure 14**, the authors formulate a supposition: that the principal heat sources display mutually mirroring thermal signatures.

Contrasting the outcomes of synchronised and independent fan operation across an array of workload configurations—while holding the weighting vector unchanged during the independent phase—reveals that the CPU temperature behaviours follow an identical template. The cooling airstream follows a defined path: the incoming ambient air first washes over the CPUs situated in the forward section, absorbing their dissipated energy, before continuing downstream to the PCIe devices occupying the aft section. As a logical consequence, the thermal state of the forward CPUs directly governs the thermal environment experienced by the aft PCIe devices. If the same weighting vector is utilised for independent fan operation under workload configurations that share an equivalent dissipation level for the forward CPUs, then the PCIe devices too will manifest thermal signatures that are replicas of one another.

Building outward from this supposition of mirrored thermal signatures and the control traces linked to the optimal vectors **(Figures 18-23)** gathered under the various workload configurations, the cases are arranged into matched pairs—A with B, C with D, E with F—and the temperature excursions between synchronised and independent operation are examined, as tabulated in **Table 4**. Within each matched pair, the optimal vector remains invariant, and the forward CPUs dissipate at the same level. The thermal signatures are then decoded through the framework of the supposition. Additionally, because the acceptable static regulation margin is pegged at 2% of the 55 °C ceiling, a temperature variation exceeding +1.1 °C is treated as an increase, a variation below −1.1 °C as a decrease, and any intermediate variation as flat. Extracting the control data from **Figures 18-23**, temperature transitions are traced between the static condition under synchronised commands and the static condition under independent commands using the optimal vectors. As collated in **Table 4**, these transitions for Cases A through F are logged, and the corresponding thermal interpretations are summarised in **Table 5**. The collected evidence supports the supposition advanced in this section. Provided identical weighting vectors are applied during independent fan operation, and the forward CPUs share a common dissipation intensity, both the CPU pair and the PCIe pair will reproduce the same thermal behaviour.

Table 4. Best weightings of Case A to H — temperature variations of critical components.

Case	Best weighting [Weighting fan1, 2, 3, 4]	Steady-state temperature variations (°C)			
		PCIe1	PCIe0	CPU1	CPU0
A	[1.0, 0.8, 0.4, 1.0]	-0.6	+0.3	0.0	+5.6

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

B		-0.4	+0.5	+0.1	+5.8
C	[0.8, 0.6, 0.6 ,1.0]	+4.9	+8.1	0.0	+6.6
D		+4.8	+8.1	0.0	+6.6
E	[1.0, 0.6, 0.8, 1.0]	+0.2	+5.3	+3.4	+6.5
F		0.0	+5.0	+3.2	+6.3
G	[1.0, 1.0, 1.0, 0.6]	+0.1	+2.3	+1.4	+0.8
H	[1.0, 1.0, 0.8, 0.6]	0.0	+5.1	+6.4	+1.0

Table 5. Best weightings of Case A to F — temperature variation trend.

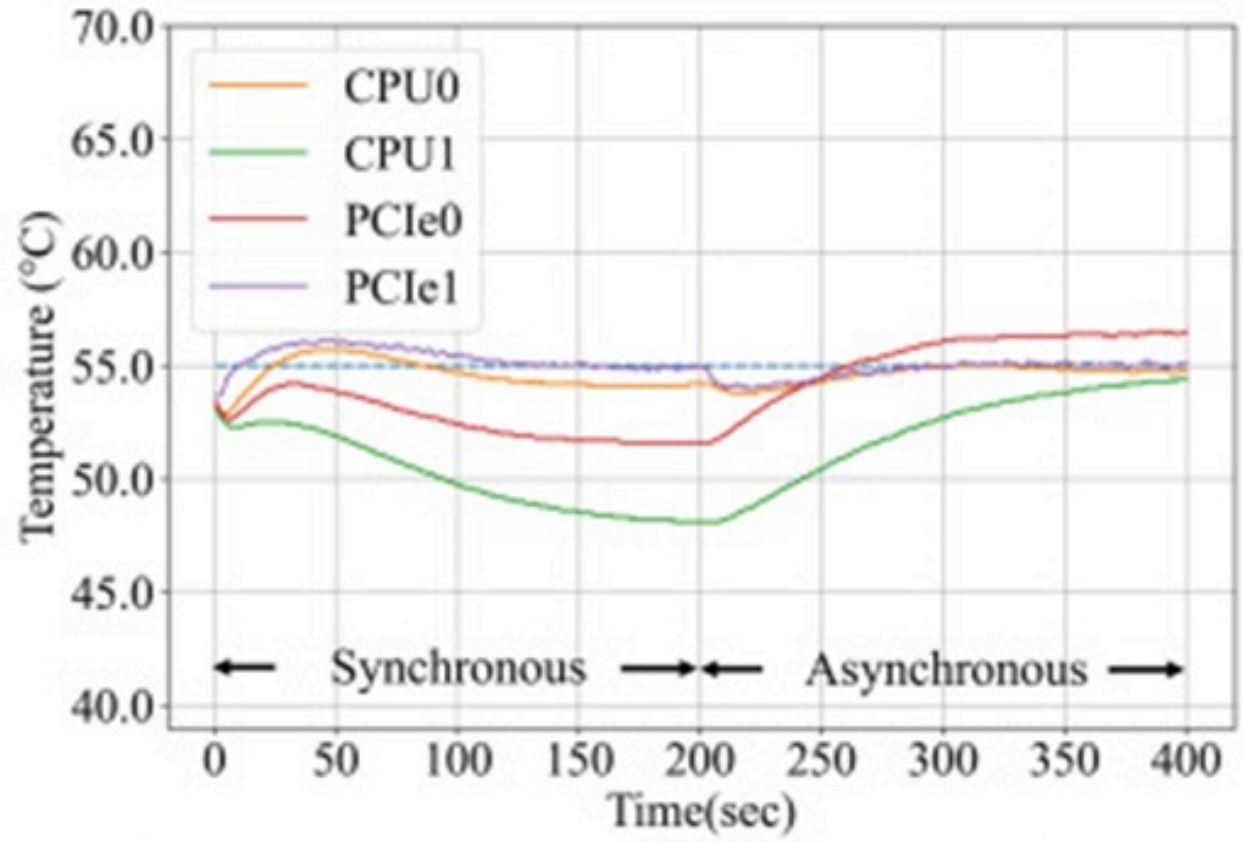
Scenario	Trend of Steady-State Temperature Variation				Optimal Coefficient Set [Fan1, Fan2, Fan3, Fan4]
	CPU0	CPU1	PCIe0	PCIe1	
A	rising	stable	stable	stable	[1.0, 0.8, 0.4, 1.0]
B	rising	stable	stable	stable	
C	rising	stable	rising	rising	[0.8, 0.6, 0.6, 1.0]
D	rising	stable	rising	rising	
E	rising	rising	rising	stable	[1.0, 0.6, 0.8, 1.0]
F	rising	rising	rising	stable	

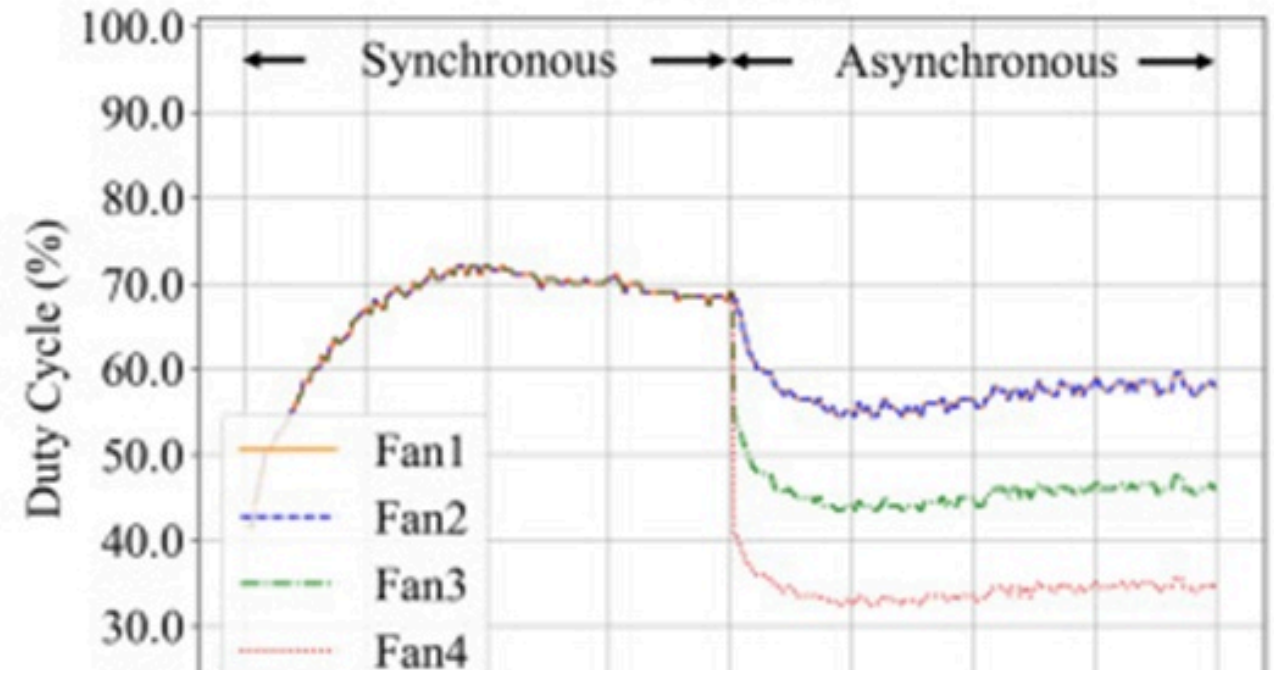
The stipulated trial window for the fan regulation exercise is fixed at 400 seconds in total, with the fans driven in lockstep for the opening 200 seconds.

Cases G and H resist interpretation through the prism of the mirrored thermal signature supposition because their optimal vectors differ from one another. To unearth the factors that drive this divergence, the static temperature constraints laid down are temporarily disregarded, and the fan regulation is allowed to proceed for its full duration, this time applying the optimal vector of one case to the other. The regulatory response of Case G when subjected to [1.0, 1.0, 0.8, 0.6] is displayed in **Figure 26**. During the independent command window (200–400 seconds), deploying [1.0, 1.0, 0.8, 0.6] causes the settled temperature at PCIe0 to overshoot the target by an amount exceeding the 2% allowance, thereby disqualifying [1.0, 1.0, 0.8, 0.6] as a workable choice for Case G. The regulatory response of Case H when subjected to [1.0, 1.0, 1.0, 0.6] is displayed in **Figure 27**. Even though [1.0, 1.0, 1.0, 0.6] passes the thermal check for Case H, the combined fan wattage under independent operation clocks in at 21.1 W—a figure substantially higher than the 15.3 W achieved when deploying its own optimal vector [1.0, 1.0, 0.8, 0.6]. This makes clear why [1.0, 1.0, 1.0, 0.6] fails to qualify as the leading choice for Case H.

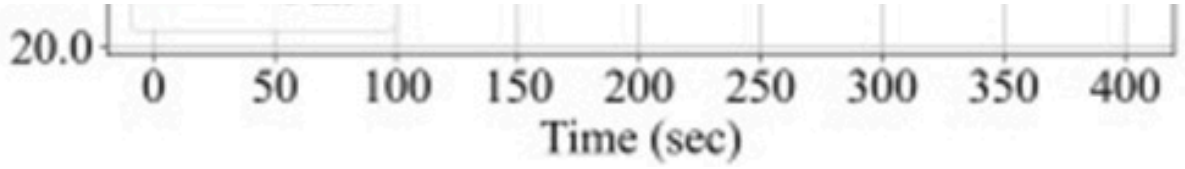
© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

Modulation of multiple fans	Components (°C)				Inappropriate weighting
	CPU0	CPU1	PCIe0	PCIe1	
Synchronous	54.2	48.1	51.6	55.0	PCIe0 steady-state temperature exceeds setpoint by more than 2%
Asynchronous	54.8	54.4	56.5	55.1	



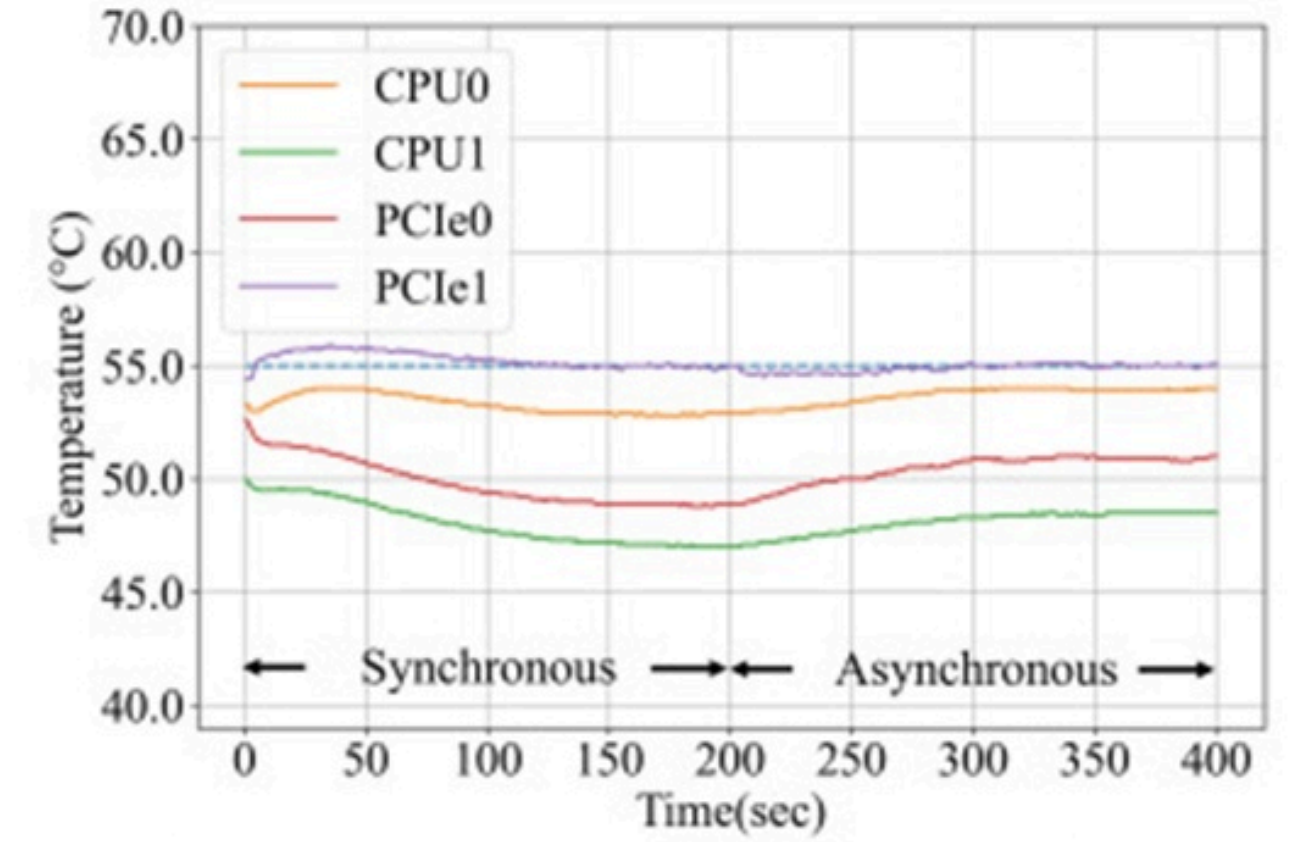


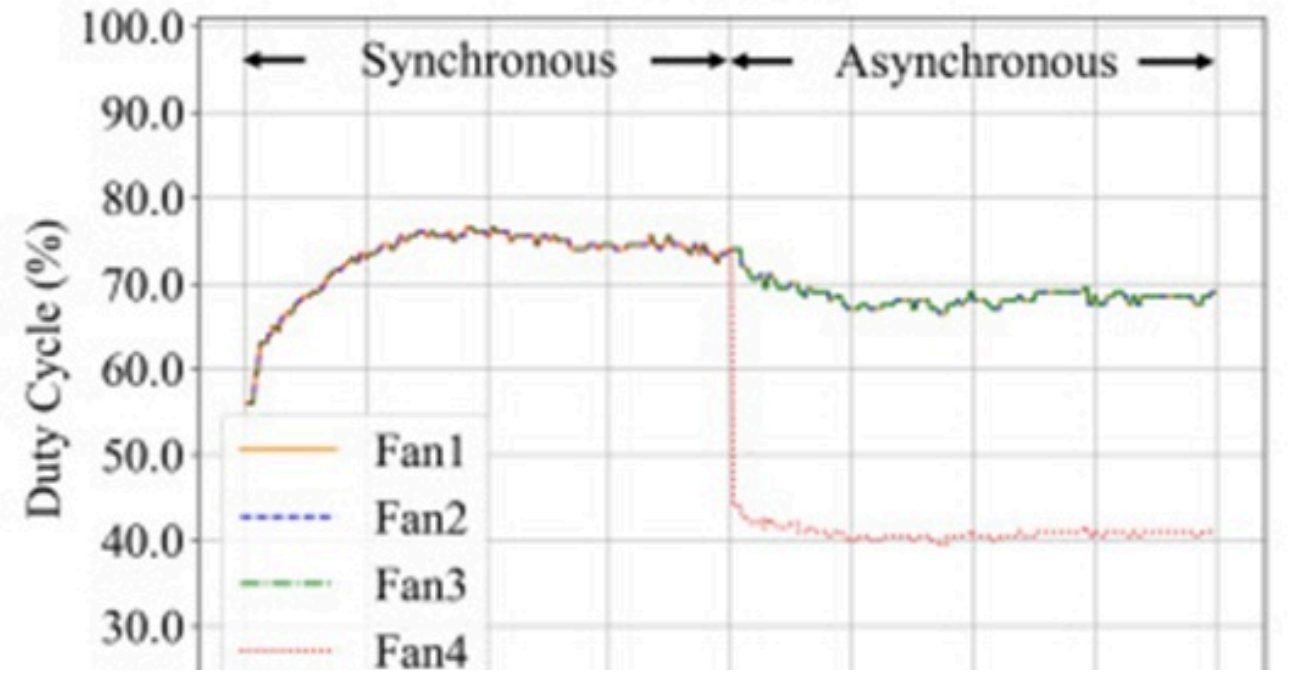
© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press



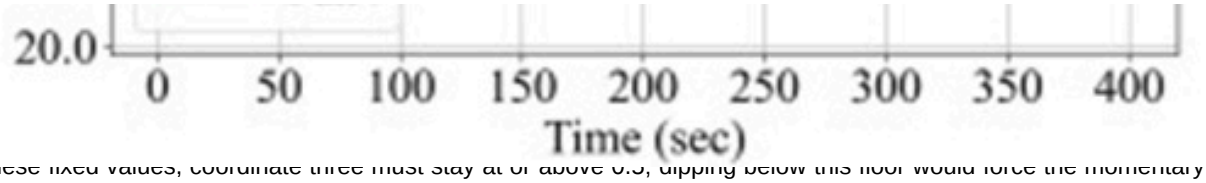
© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

Modulation of multiple fans	Components (°C)				Fans duty cycle (%)				Fans power (W)
	CPU0	CPU1	PCIe0	PCIe1	Fan1	Fan2	Fan3	Fan4	
Synchronous	52.9	47.0	48.9	55.0	74.0	74.0	74.0	74.0	28.0
Asynchronous	54.0	48.5	51.0	55.1	69.0	69.0	69.0	41.0	21.1





© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press



Given these fixed values, coordinate three must stay at or above 0.5, dipping below this floor would force the momentary temperature to overshoot the target by more than 5 °C. To protect against hardware stress from thermal excess, coordinate three was methodically swept from 0.5 up to 1.0. Through this sweep, temperature excursions were captured for both Cases G and H. Anchored in the supposition of a mirrored thermal signature, a thermal analysis was carried out, as summarised in **Table 7**. **Table 7** lends additional weight to the supposition of a mirrored thermal signature formulated in this section. The curves tracing temperature excursions for the principal heat sources in Cases G and H are rendered in **Figures 28 and 29.**

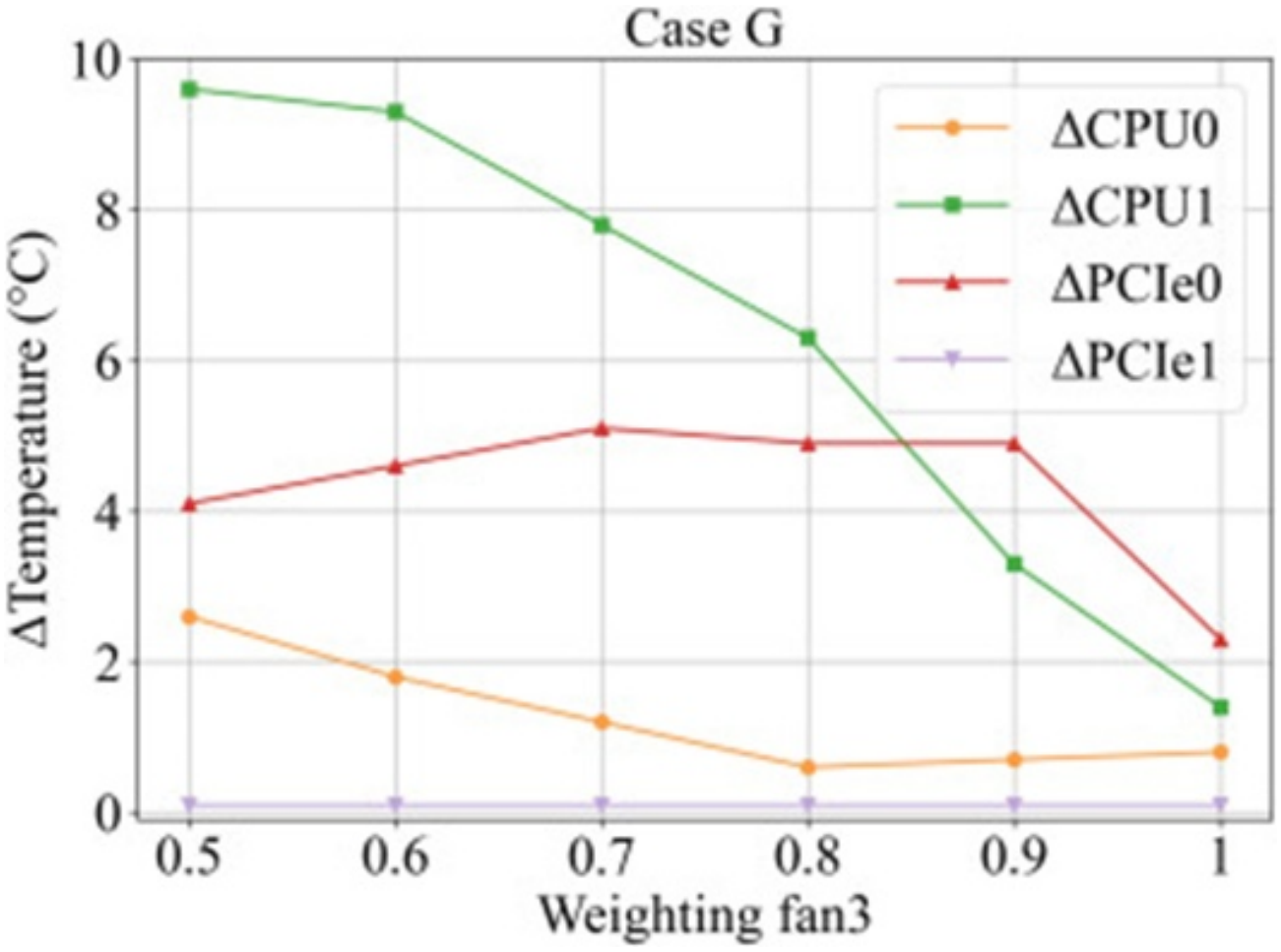


Figure 28. Weighting tuning of Case G — temperature variation curve.

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

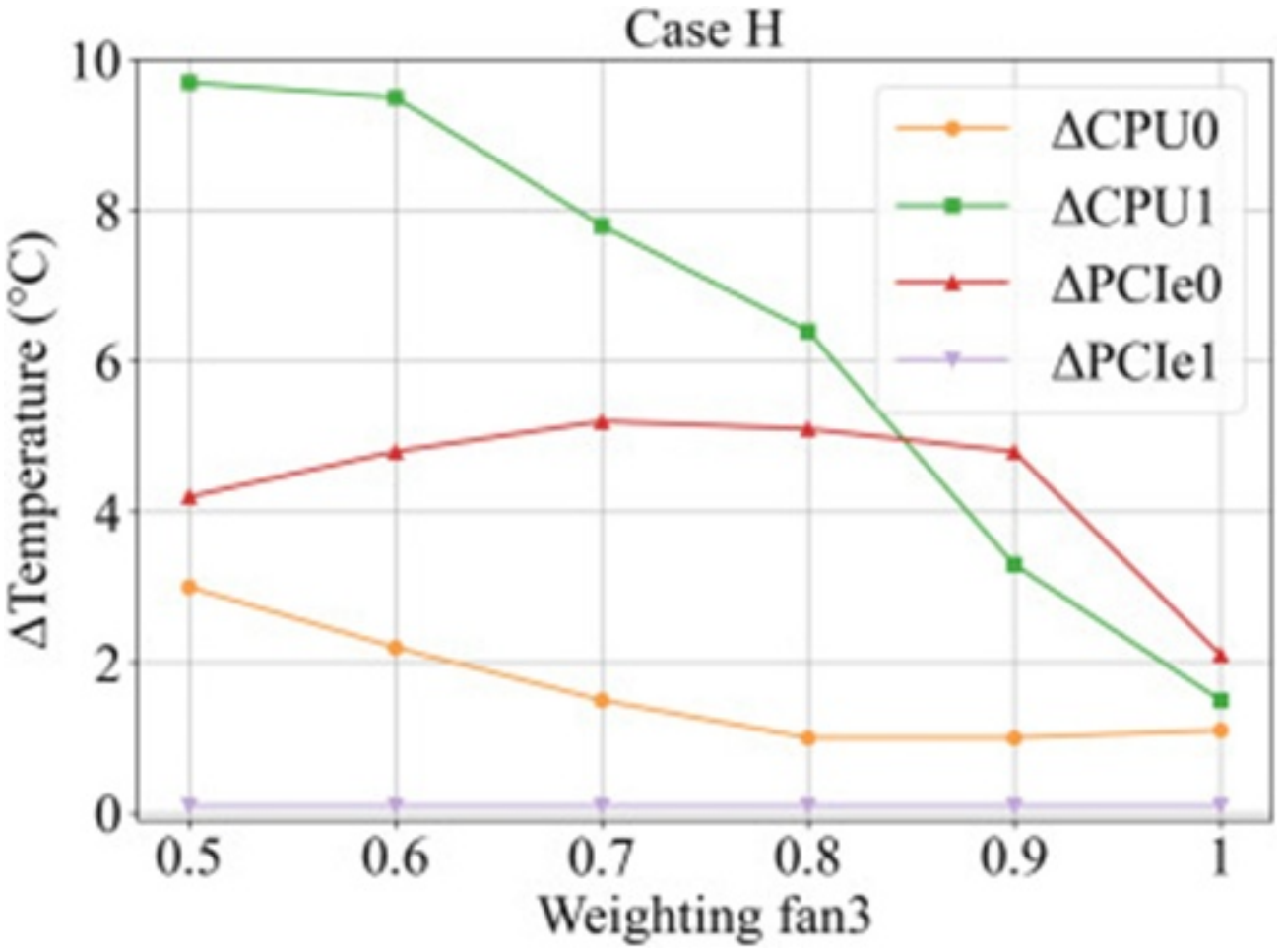


Figure 29. Weighting tuning of Case H — temperature variation curve.

What stands out is the striking degree of overlap between the two cases' profiles. Quantifying this similarity, the Mean Absolute Error (MAE) was determined to be 0.2, 0.0, 0.3, and 0.2 °C for CPU0, CPU1, PCIe0, and PCIe1, respectively. Complementarily, the Root Mean Square Error (RMSE) came out at 0.3, 0.1, 0.1, and 0.1 °C for the same four measurement points. Both error metrics stay well within the 5 °C neighbourhood for all components. A further observation drawn from the plotted curves **(Figures 28 and 29)** is that the thermal response of certain components—PCIe0 offering a case in point—does not scale linearly with adjustments to the weighting values, serving as additional confirmation that the thermal regulation of servers inhabits the domain of nonlinear problems [11].

Weighting search experiments

The weighting apparatus is interwoven with the evolutionary strategy optimiser and paired with the custom-designed fitness measures (5), (6) to instantiate the multi-fan regulation framework tailored for the server. Weighting search experiments were staged across the eight workload configurations enumerated in **Figure 13.** The framework's algorithmic engine draws from the catalogue of 369 candidate vectors **(Figure 15)**. During each generational cycle, a brood of eight offspring is created (λ = 8), which translates to selecting eight vectors (draws may repeat) from the 369-member catalogue for experimental fan regulation. For the initial 200-second span, synchronised fan commands prevail; for the subsequent 200-to-400-second span, each chosen vector is trialled in turn for independent fan operation. Every vector is adjudicated as either non-viable or viable according to the thermal envelope, after which its fitness is assessed via **Eqs. 4 or 5** as dictated by its classification. Informed by these fitness scores, the quartet of offspring posting the strongest scores out of the eight is elevated to act as progenitors for the upcoming generation, which is assembled through recombination and mutation. This generational loop cycles until the 50-generation mark is reached, with the ultimate aim of pinpointing a

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

near-optimal weighting vector after a modest number of search steps. That vector is subsequently employed for independent fan operation to conform to the thermal envelope while realising near-optimal energy thrift. The weighting search experiments are additionally repeated over 30 independent runs to temper the algorithm's stochastic fluctuations when estimating its true capability.

Table 6. Weighting tuning of Case G, H — temperature variations of critical components.

Case	Best weighting [Weighting fan1, 2, 3, 4]	Steady-state temperature variations (°C)			
		PCIe1	PCIe0	CPU1	CPU0
G	[1.0, 1.0, 0.5, 0.6]	+0.1	+4.1	+9.6	+2.6
H		0.0	+4.2	+9.7	+3.0
G	[1.0, 1.0, 0.6, 0.6]	+0.1	+4.6	+9.3	+1.8
H		+0.2	+4.8	+9.5	+2.2
G	[1.0, 1.0, 0.7, 0.6]	+0.1	+5.1	+7.8	+1.2
H		+0.2	+5.2	+7.8	+1.5
G	[1.0, 1.0, 0.8, 0.6]	+0.1	+4.9	+6.3	+0.6
H		0.0	+5.1	+6.4	+1.0
G	[1.0, 1.0, 0.9, 0.6]	+0.1	+4.9	+3.3	+0.7
H		+0.2	+4.8	+3.3	+1.0
G	[1.0, 1.0, 1.0, 0.6]	+0.1	+2.3	+1.4	+0.8
H		+0.1	+2.1	+1.5	+1.1

Table 7. Weighting tuning of Case G, H — temperature variation trend.

Scenario	Trend of Steady-State Temperature Variation				Optimal Coefficient Set [Fan1, Fan2, Fan3, Fan4]
	CPU0	CPU1	PCIe0	PCIe1	
G	rising	rising	rising	stable	[1.0, 1.0, 0.5, 0.6]
H	rising	rising	rising	stable	
G	rising	rising	rising	stable	[1.0, 1.0, 0.6, 0.6]
H	rising	rising	rising	stable	
G	rising	rising	rising	stable	[1.0, 1.0, 0.7, 0.6]
H	rising	rising	rising	stable	
G	stable	rising	rising	stable	[1.0, 1.0, 0.8, 0.6]
H	stable	rising	rising	stable	

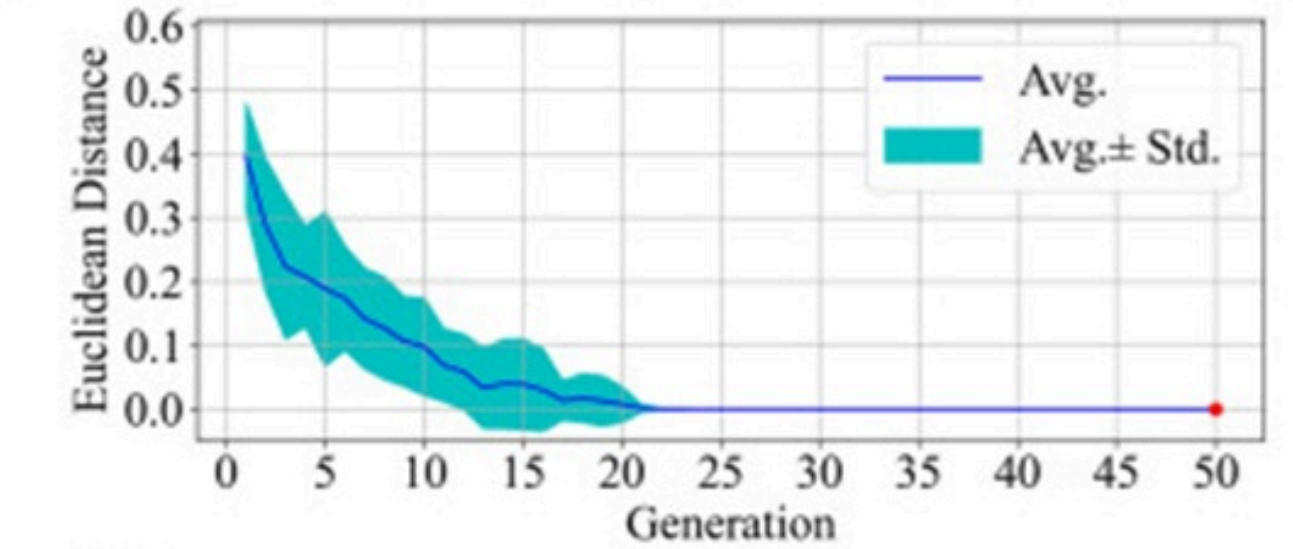
© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

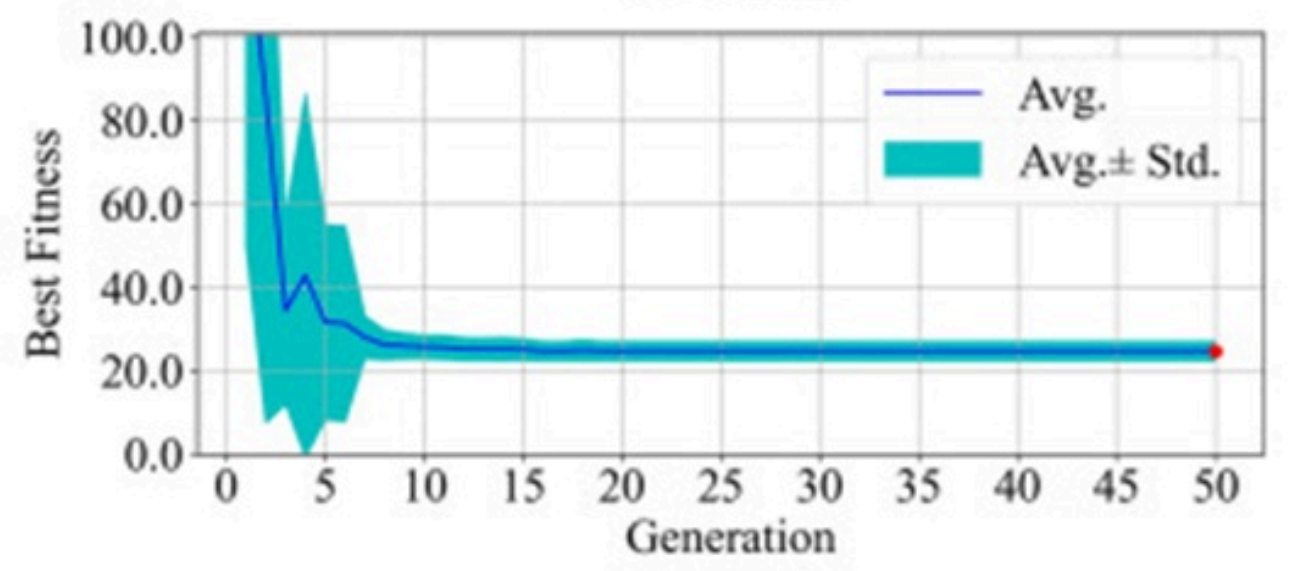
G	stable	rising	rising	stable	[1.0, 1.0, 0.9, 0.6]
H	stable	rising	rising	stable	
G	stable	rising	rising	stable	[1.0, 1.0, 1.0, 0.6]
H	stable	rising	rising	stable	

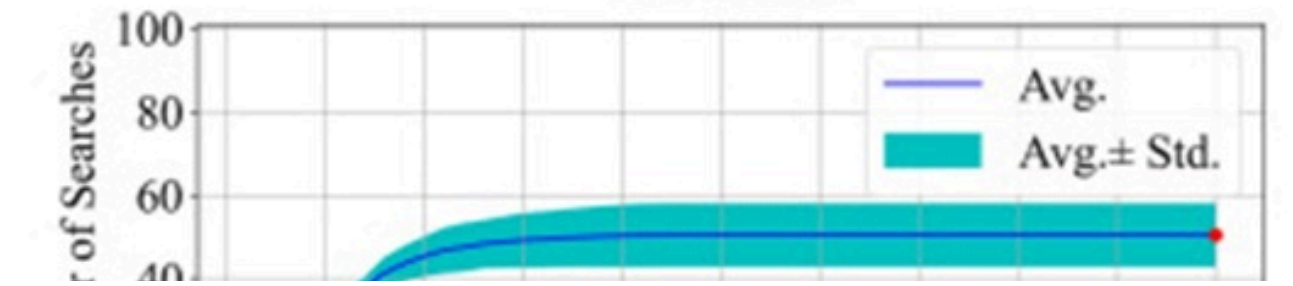
The findings from the weighting search experiments for Cases A through H are charted in **Figures 30-37**, which respectively map out the Euclidean distance metric, the top fitness value attained, and the cumulative tally of distinct search evaluations expended by the algorithm per generation. The Euclidean distance, formulated in **Eq. 8**, acts as a probe for the state of the algorithmic exploration. A Euclidean distance reading of zero communicates that the search population has coalesced, while the top fitness captures the strongest fitness reading recorded within each generation. In the interest of real-world deployability, the algorithm retains a memory of previously trialled vectors and avoids re-testing them under the same workload configuration. The tally of distinct search evaluations signifies the actual count of non-redundant vectors experimentally screened by the algorithm from the 369-candidate pool. **Figures 30-34** present the weighting search experiments for Cases A through H, and the accompanying statistical digest appears in **Table 8**. Across Cases A through H, the tallies of distinct search evaluations read 50, 52, 50, 49, 62, 52, 54, and 52, respectively. The Euclidean distance metric drops to zero in every instance, and the top fitness values are posted as 24.7, 25.2, 20.8, 18.1, 13.2, 13.9, 18.8, and 15.7, respectively.

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

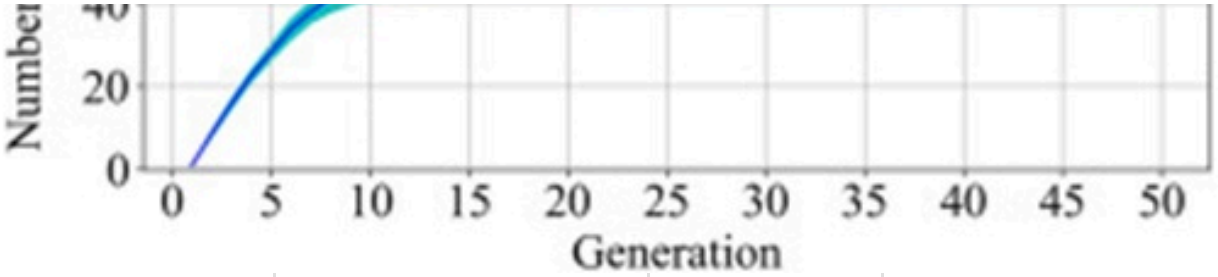
Weighting search results	Avg. ± Std.
Euclidean distance	0.0 ± 0.0
Best fitness	24.7 ± 2.2
Number of searches	50 ± 7







© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press



Case B	52 ± 8	25.2 ± 1.4	0.0 ± 0.0
Case C	50 ± 8	20.8 ± 2.9	0.0 ± 0.0
Case D	49 ± 12	18.1 ± 1.5	0.0 ± 0.0
Case E	62 ± 16	13.2 ± 1.1	0.0 ± 0.0
Case F	52 ± 11	13.9 ± 1.1	0.0 ± 0.0
Case G	54 ± 13	18.8 ± 0.5	0.0 ± 0.0
Case H	52 ± 9	15.7 ± 1.8	0.0 ± 0.0

Euclidean distance $(M(g+1), M(g)) = M(g+1) - M(g)$ (8)

$M^{(g+1)}$: Centroid vector of the parent set at generation g+1.

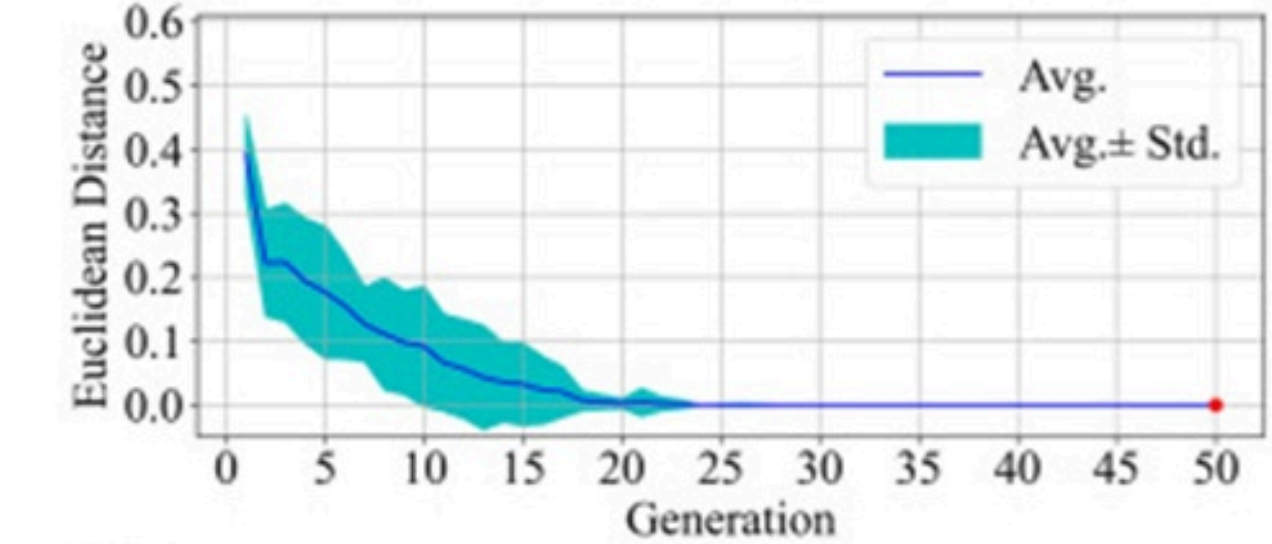
$M^{(g)}$: Centroid vector of the parent set at generation g.

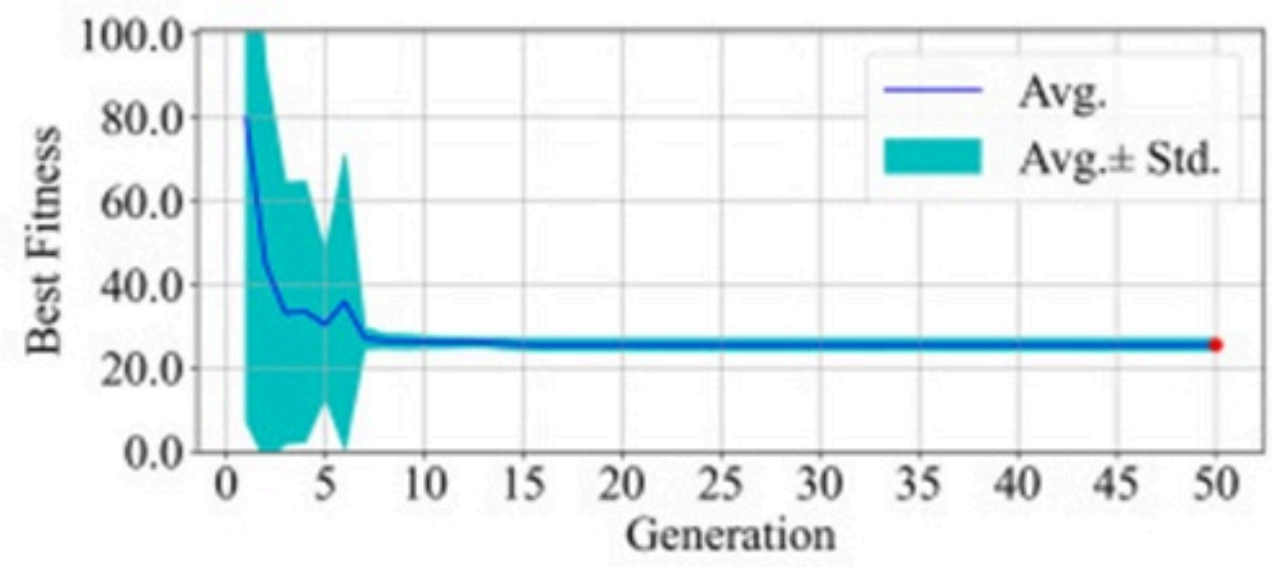
Figure 30 shows that, for Case A, the algorithm issues new generations and refines its search heading through the weighting space in response to the fitness signals received. Through this adaptive process, the top fitness logged at each generation progressively approaches the optimum. The search procedure is executed across 30 independent runs to extract the average ± one standard deviation for each metric. Once evolution has proceeded to the 22nd generation, the Euclidean distance collapses to zero, a sign that the search has fully played out. The tally of distinct search evaluations stands at 50 ± 7, signifying how many non-identical vectors from the 369-member pool required actual experimental verification. The top fitness registers at 24.7 ± 2.2, meaning that the near-optimal vector unearthed by the algorithm, when harnessed for independent fan operation, leads to an aggregate fan wattage corresponding to the quoted fitness value.

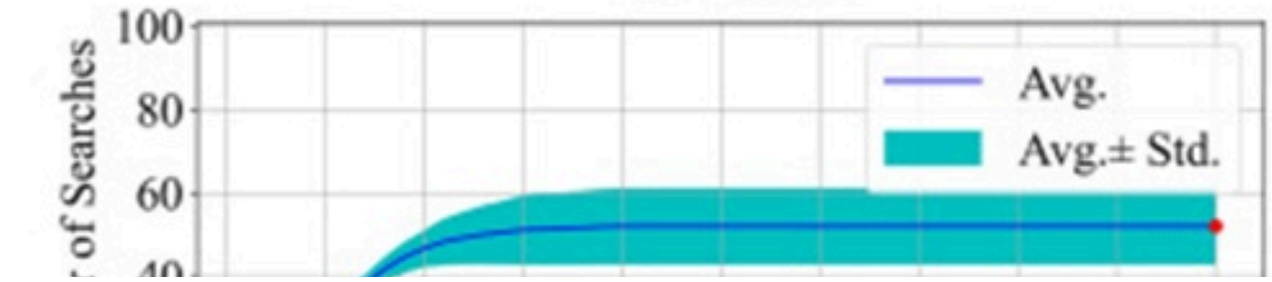
In the instance of Case B, graphed in **Figure 31**, the Euclidean distance falls to zero upon crossing the 24th generation, which denotes search completion. The tally of distinct search evaluations lands at 52 ± 8, encapsulating the number of non-overlapping vectors from the 369-entry pool that underwent experimental scrutiny. The top fitness is noted as 25.2 ± 1.4, implying the near-optimal vector produces an aggregate fan wattage aligned with this fitness number.

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

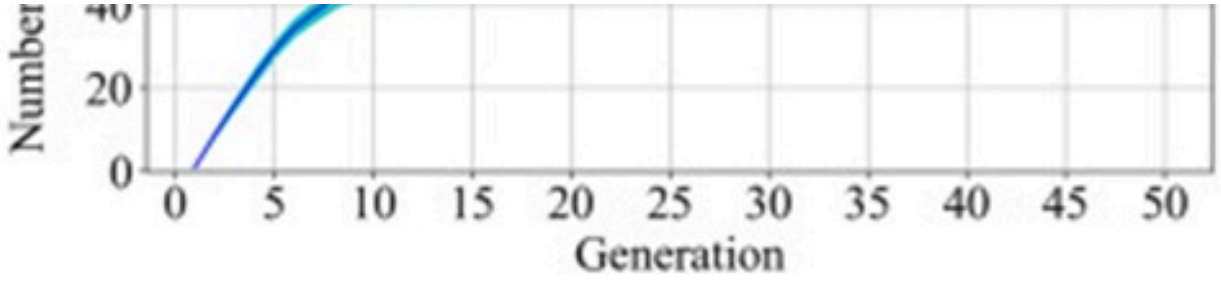
Weighting search results	Avg. ± Std.
Euclidean distance	0.0 ± 0.0
Best fitness	25.2 ± 1.4
Number of searches	52 ± 8





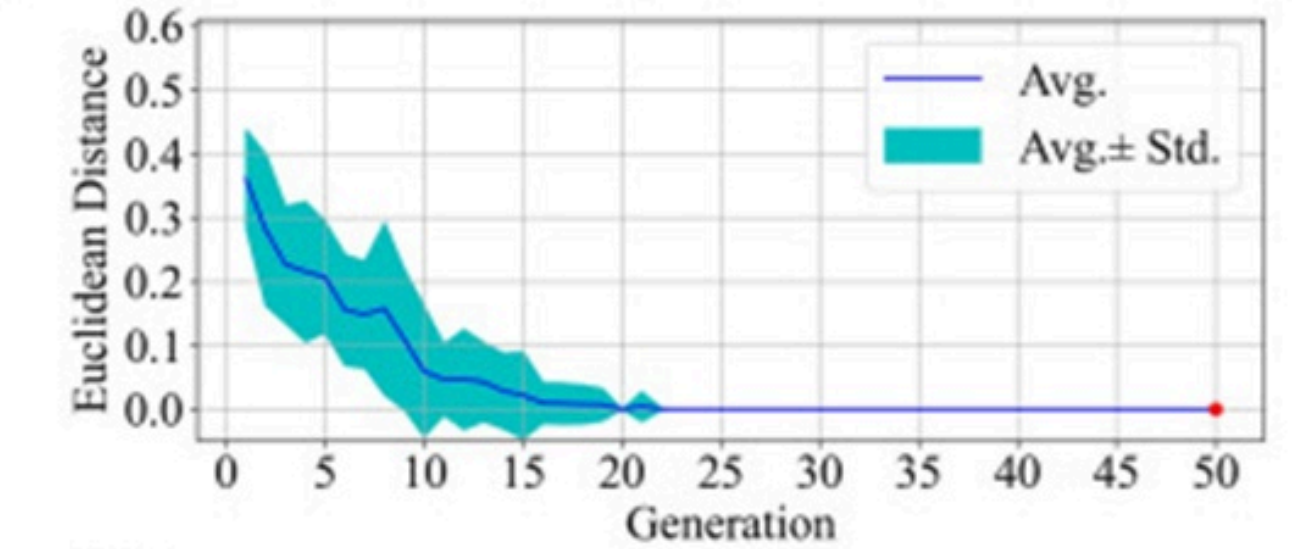


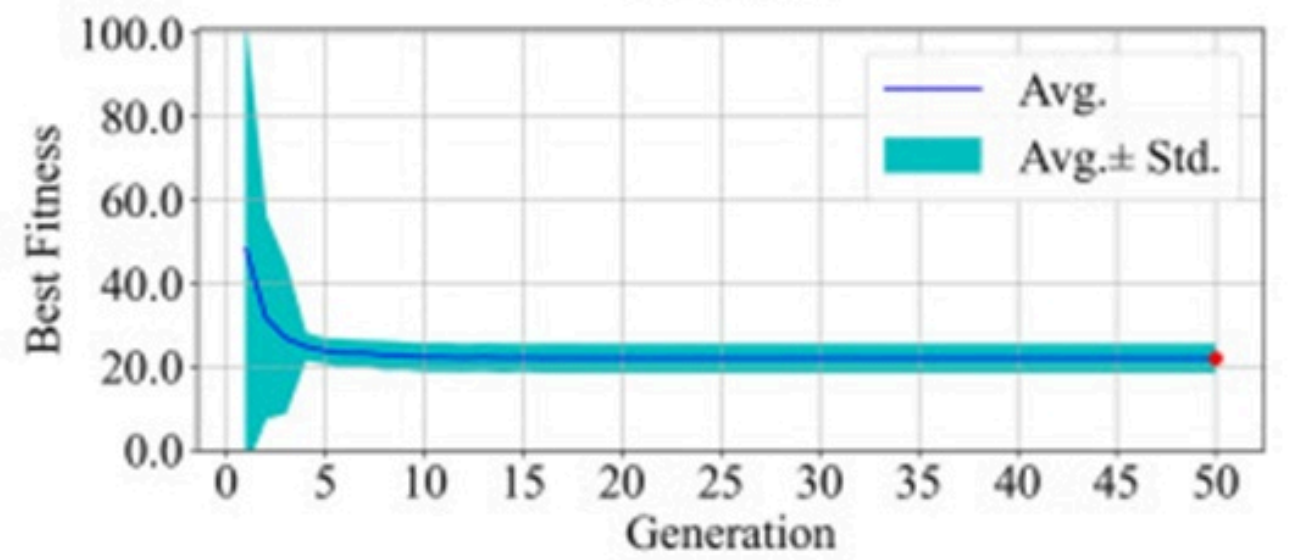
© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

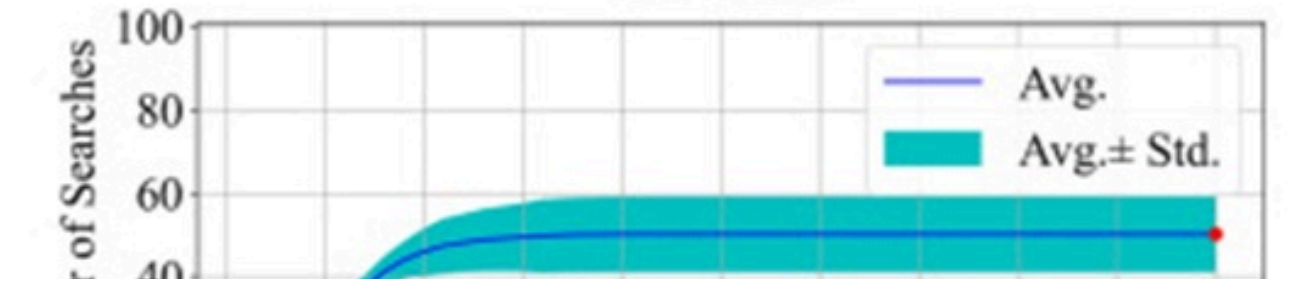


© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

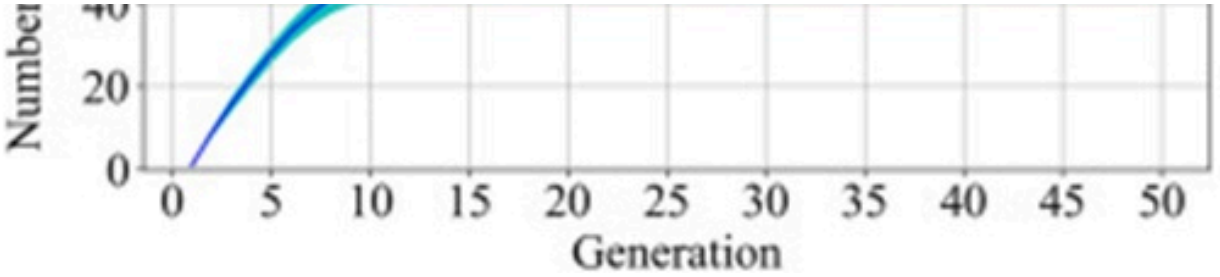
Weighting search results	Avg. ± Std.
Euclidean distance	0.0 ± 0.0
Best fitness	20.8 ± 2.9
Number of searches	50 ± 8





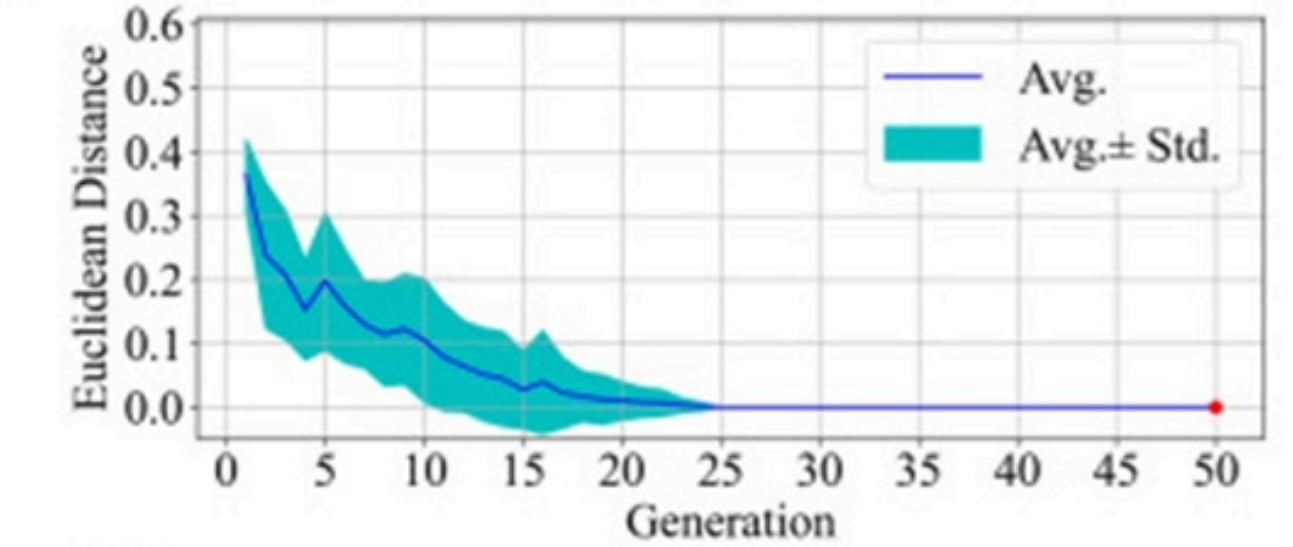


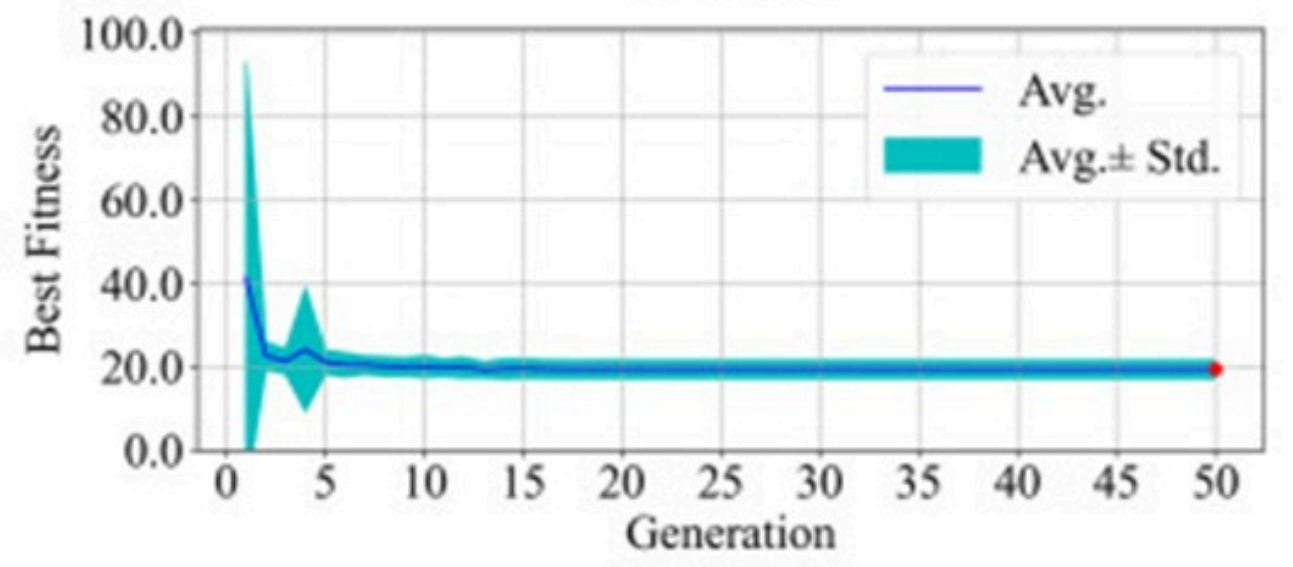
© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

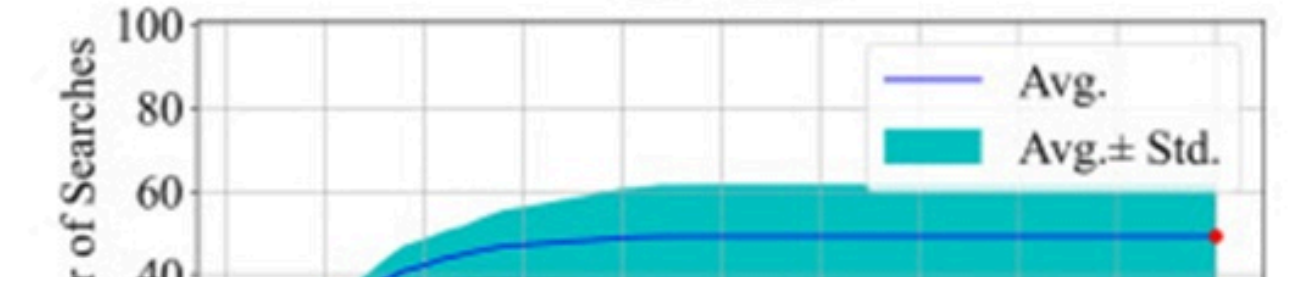


© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

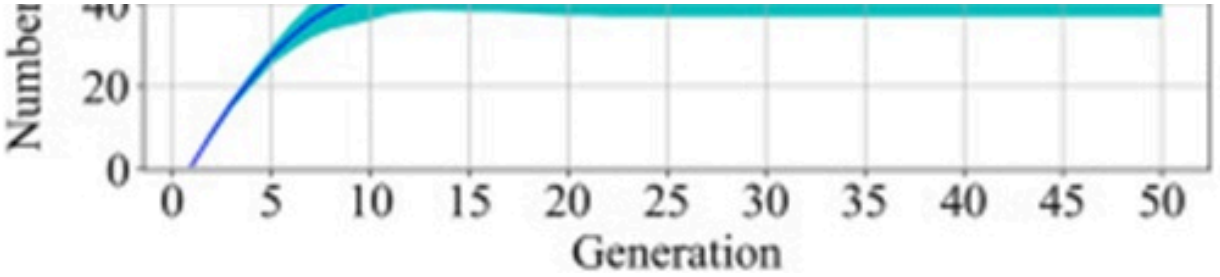
Weighting search results	Avg. ± Std.
Euclidean distance	0.0 ± 0.0
Best fitness	18.1 ± 1.5
Number of searches	49 ± 12





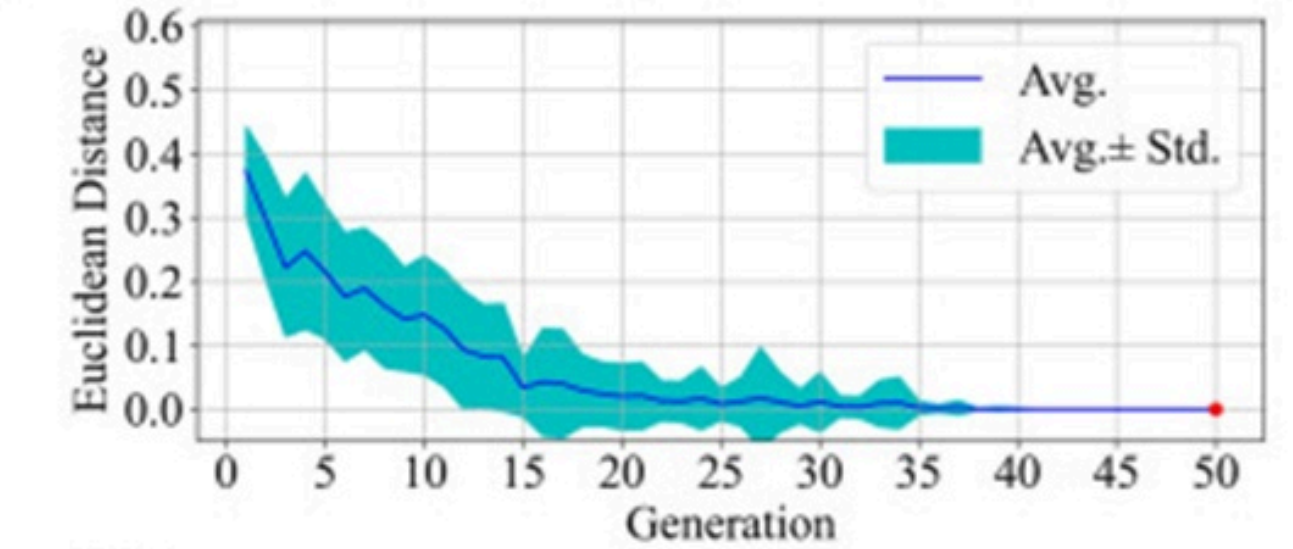


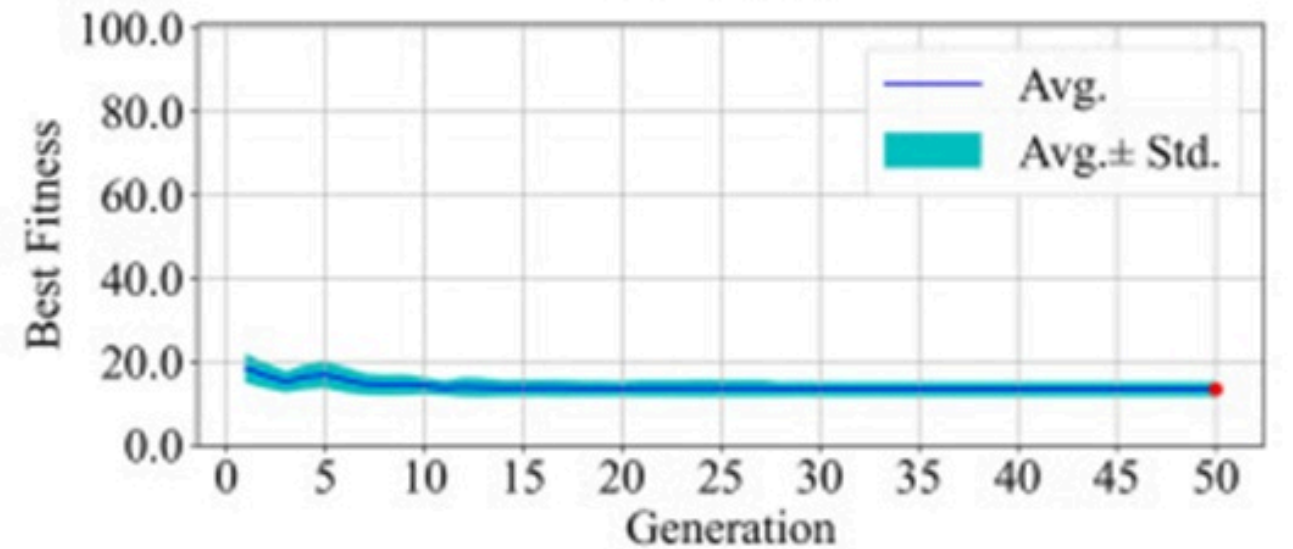
© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

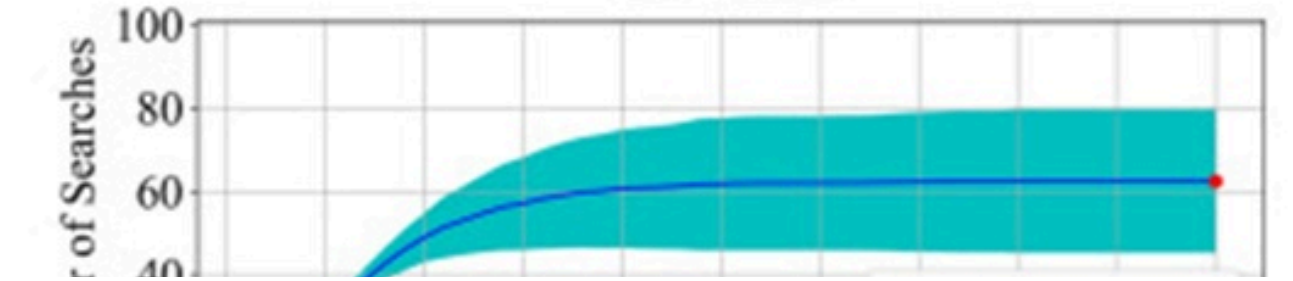


© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

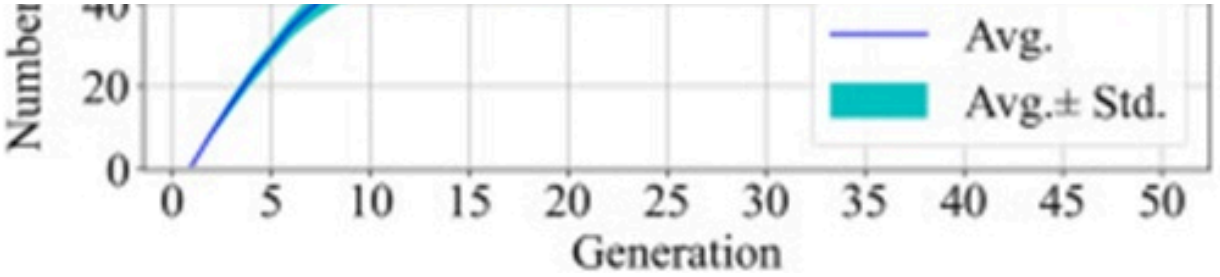
Weighting search results	Avg. ± Std.
Euclidean distance	0.0 ± 0.0
Best fitness	13.2 ± 1.1
Number of searches	62 ± 16





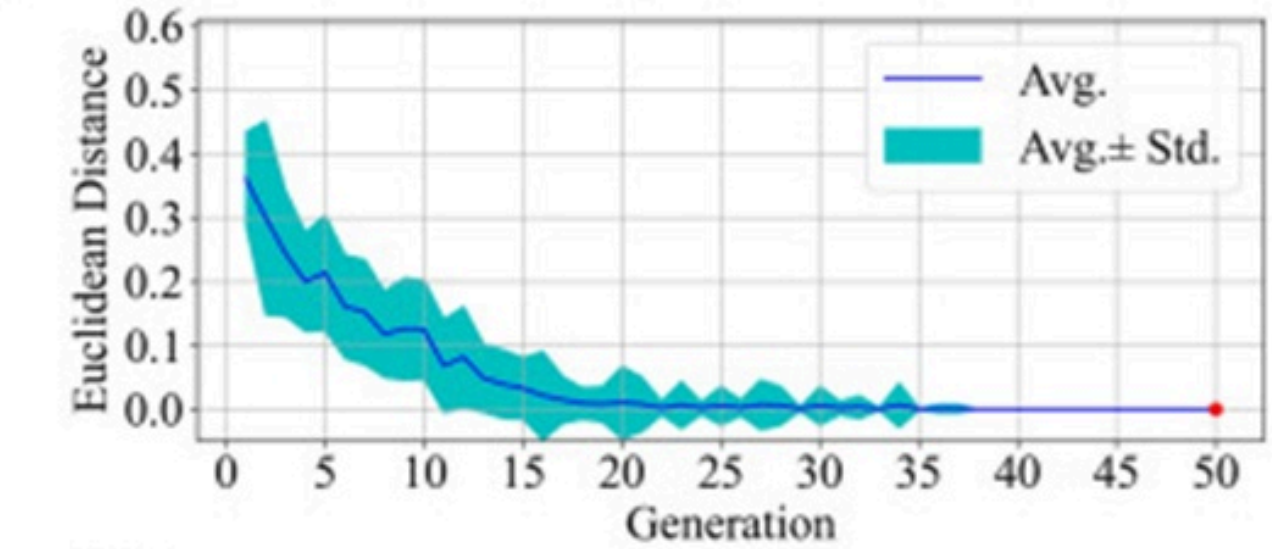


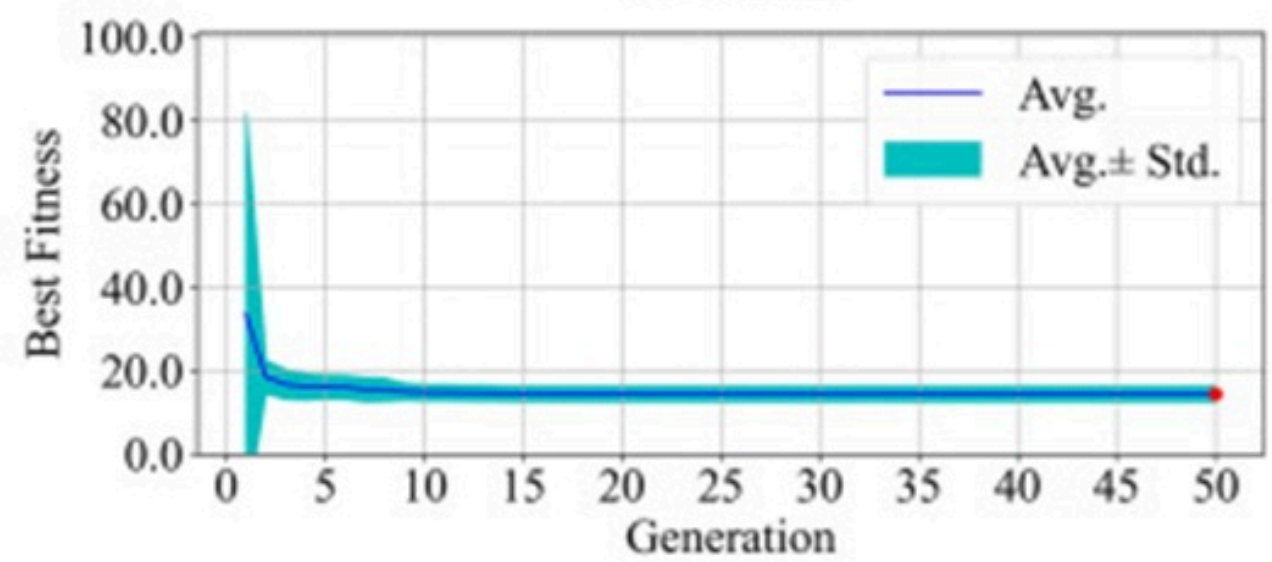
© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

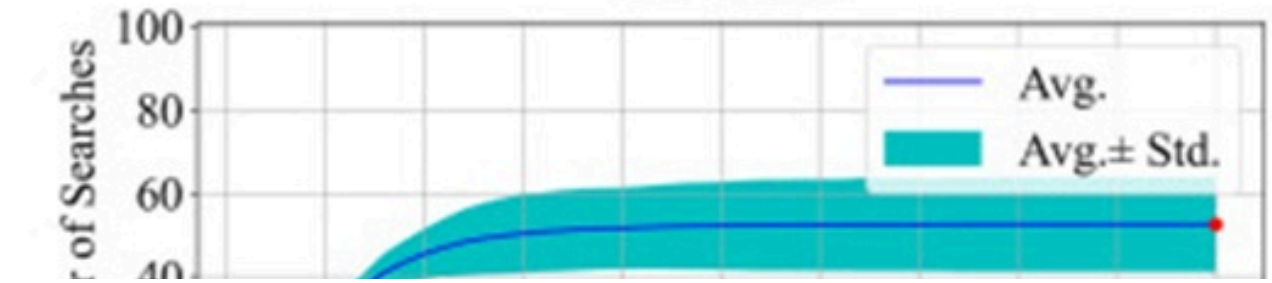


© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

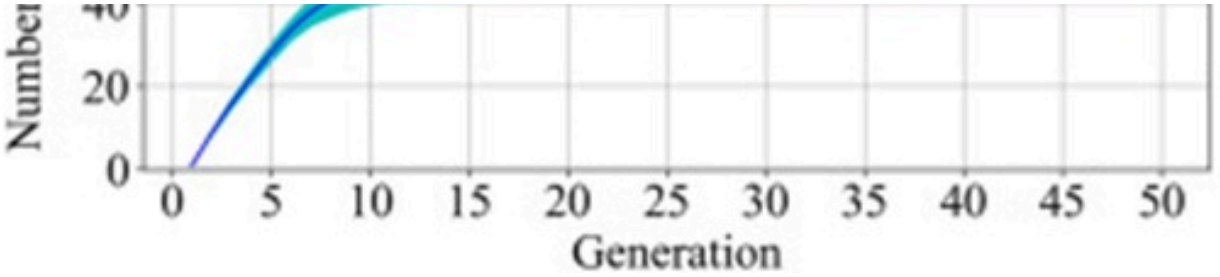
Weighting search results	Avg. ± Std.
Euclidean distance	0.0 ± 0.0
Best fitness	13.9 ± 1.1
Number of searches	52 ± 11





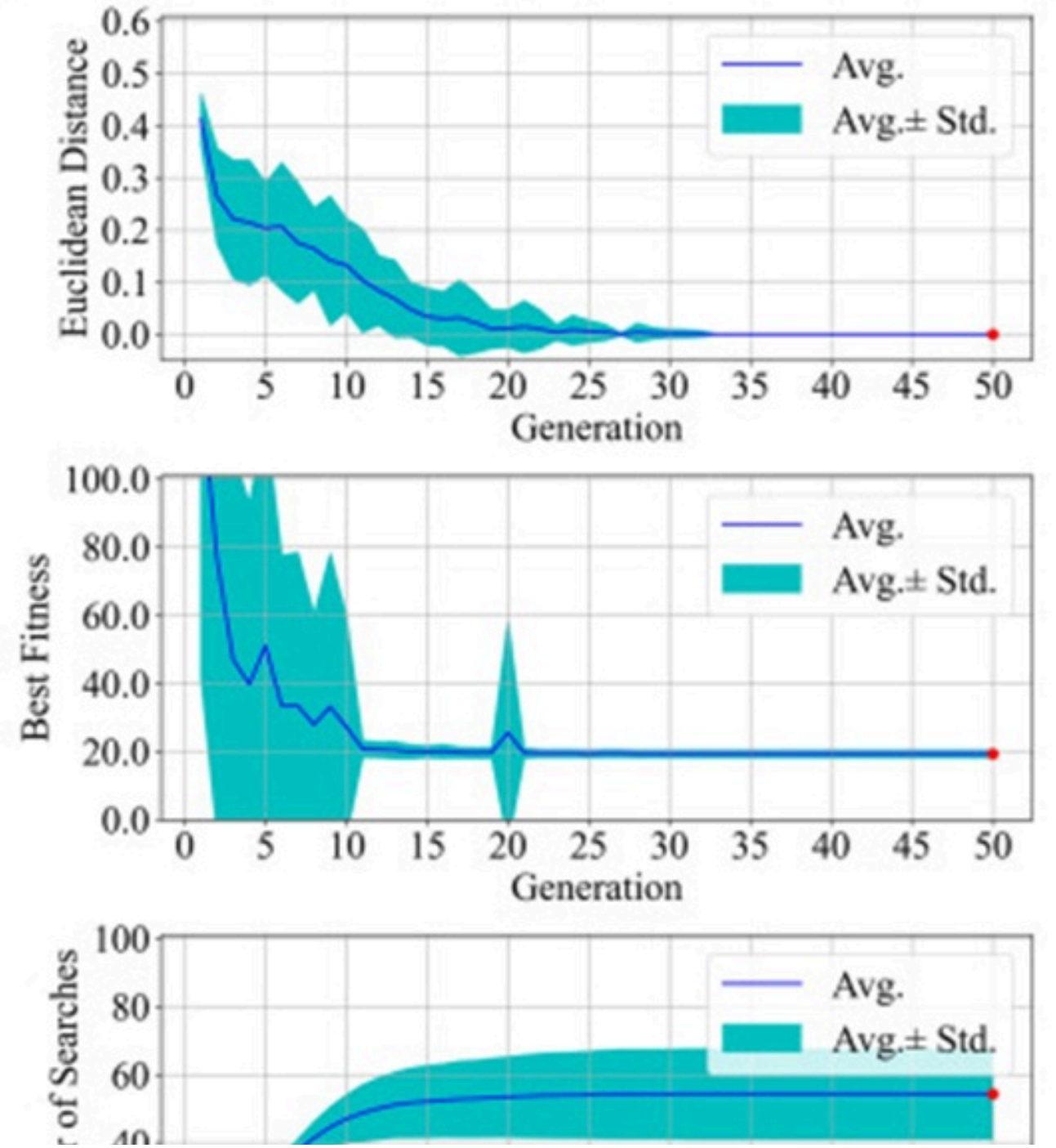


© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

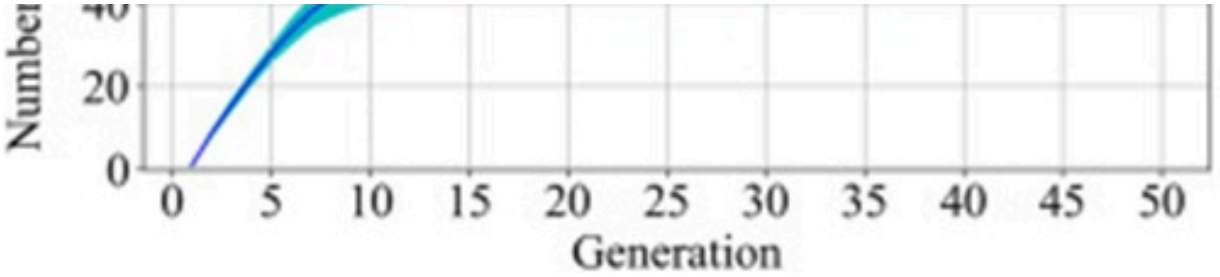


© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

Weighting search results	Avg. ± Std.
Euclidean distance	0.0 ± 0.0
Best fitness	18.8 ± 0.5
Number of searches	54 ± 13

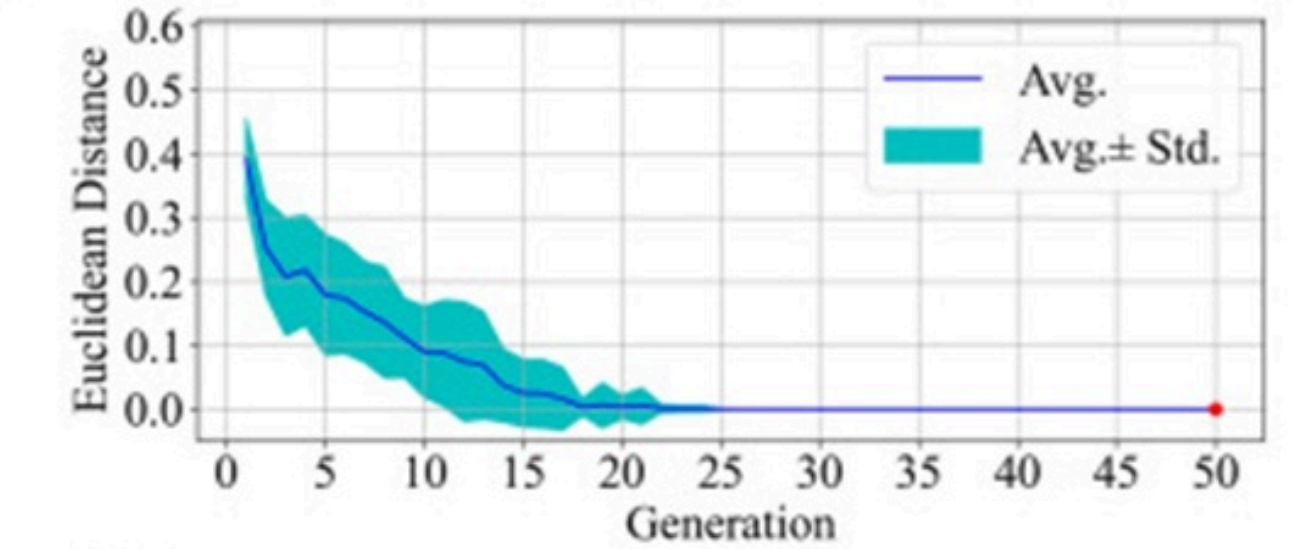


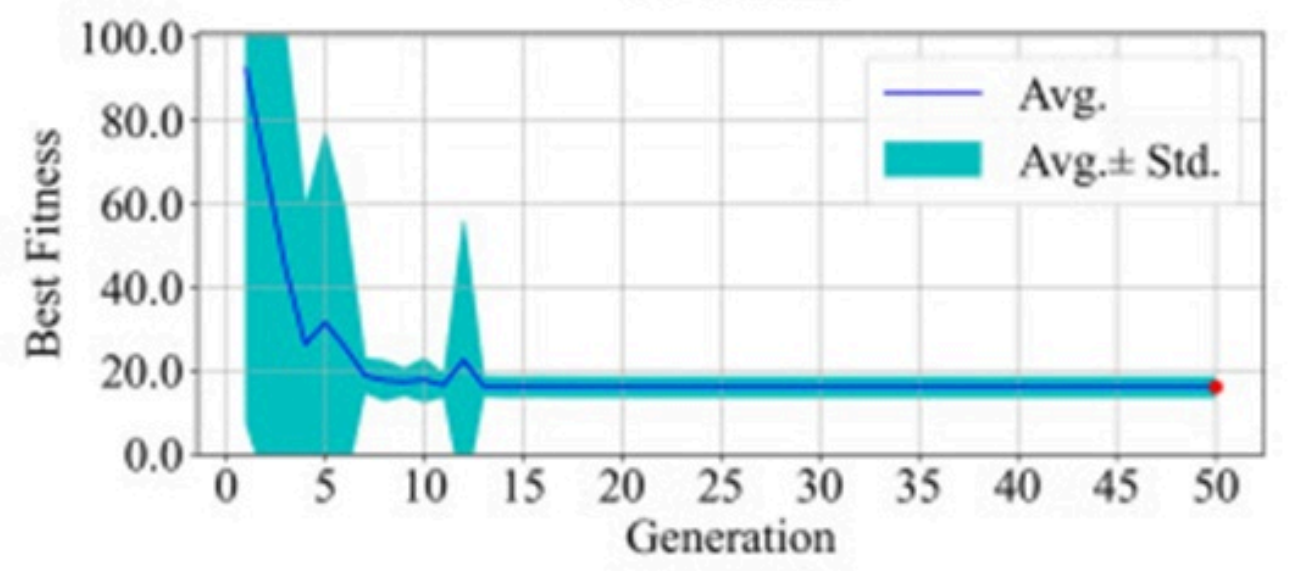
© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

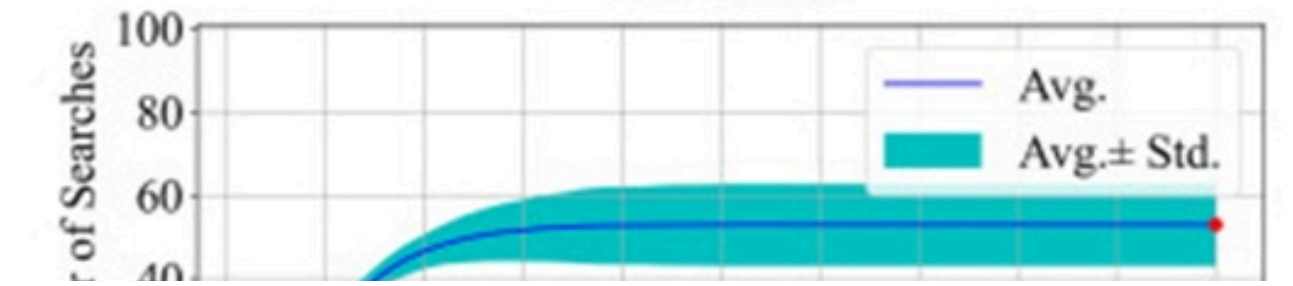


© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

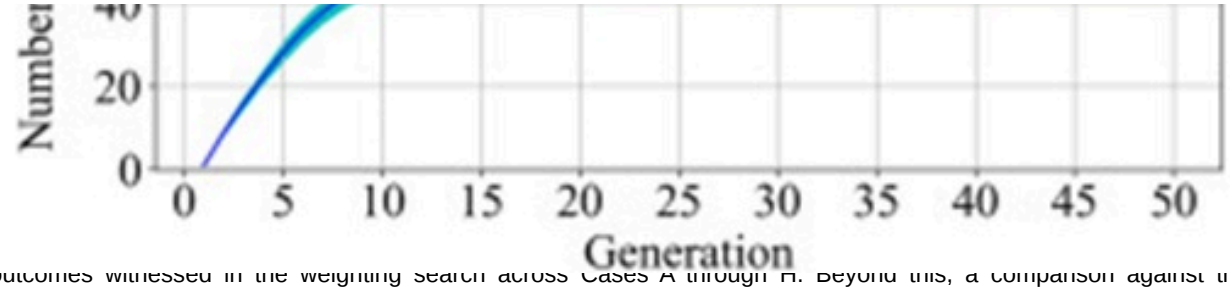
Weighting search results	Avg. ± Std.
Euclidean distance	0.0 ± 0.0
Best fitness	15.7 ± 1.8
Number of searches	52 ± 9







© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press



of outcomes witnessed in the weighting search across Cases A through H. Beyond this, a comparison against the weighting test experiments documented reveals that for Cases A through H, the weighting search experiments lifted the share of usable weightings by margins of 15.2%, 21.7%, 20.6%, 18.9%, 22.7%, 18.2%, 14.4%, and 18.3% respectively; this is visually captured in **Figure 38**. What **Figure 38** communicates is a dual achievement: the algorithm manages to converge after examining only a modest portion of the search space, and it does so while simultaneously expanding the fraction of weightings that prove viable.

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

The proportion of appropriate weightings								
Case	A	B	C	D	E	F	G	H
Weighting search	26.6%	34.7%	41.5%	44.8%	57.2%	48.9%	24.0%	29.6%
Weighting test	11.4%	13.0%	20.9%	25.9%	34.5%	30.7%	9.6%	11.3%

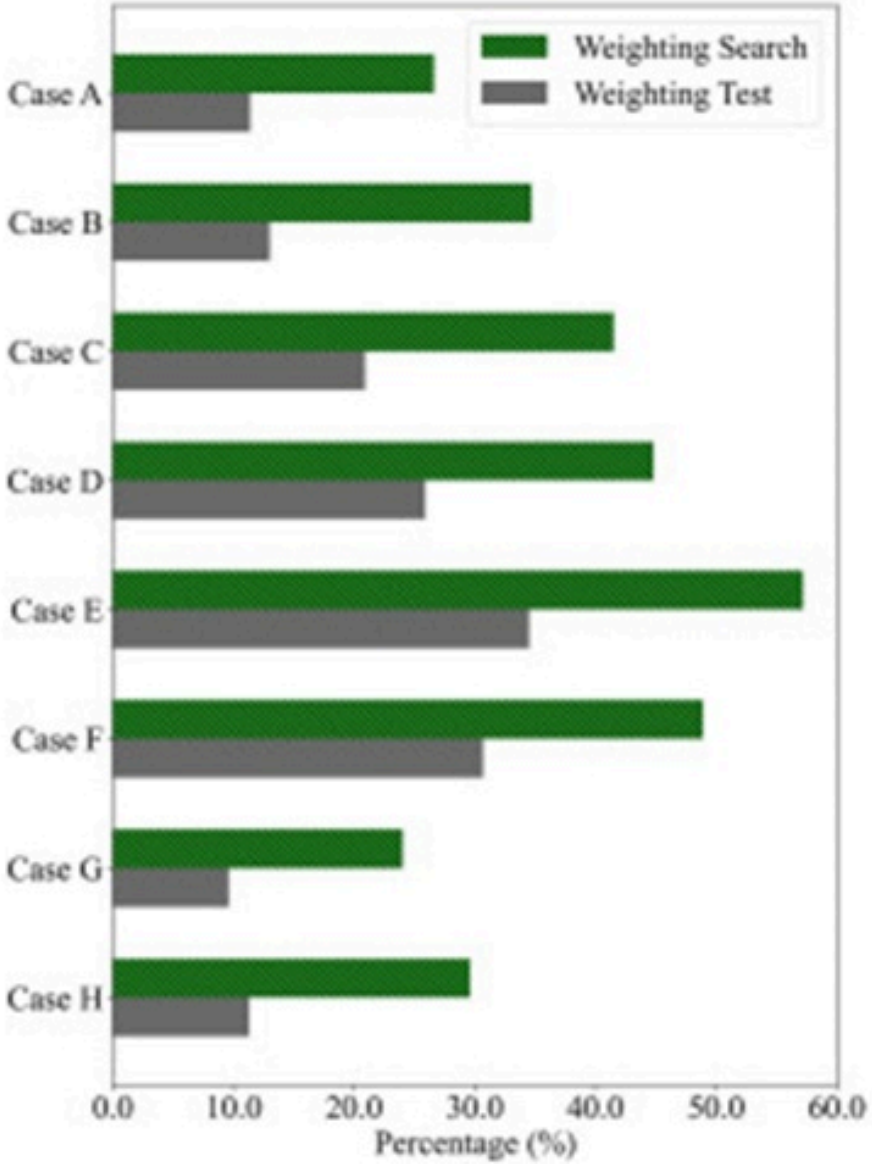


Figure 38. Weighting search and test — the proportion of appropriate weightings.

At the termination of the weighting search experiments, the peak fitness value directly maps onto the cumulative fan power consumption under independent fan operation within the multi-fan regulation framework. The distributional characteristics of this cumulative fan power—its minimum, median, maximum, and mean—are tabulated in **Table 9,** side by side with the mean percentage savings in energy when measured against the synchronized fan operation baseline. Employing Kernel Density Estimation (KDE) [27] to construct smoothed representations of the probability density, as portrayed in **Figure 39**, the horizontal coordinate represents the cumulative fan regulation power expressed in watts. The gradation of gray shading, moving from pale to dark, encodes the probability density distribution of cumulative fan power

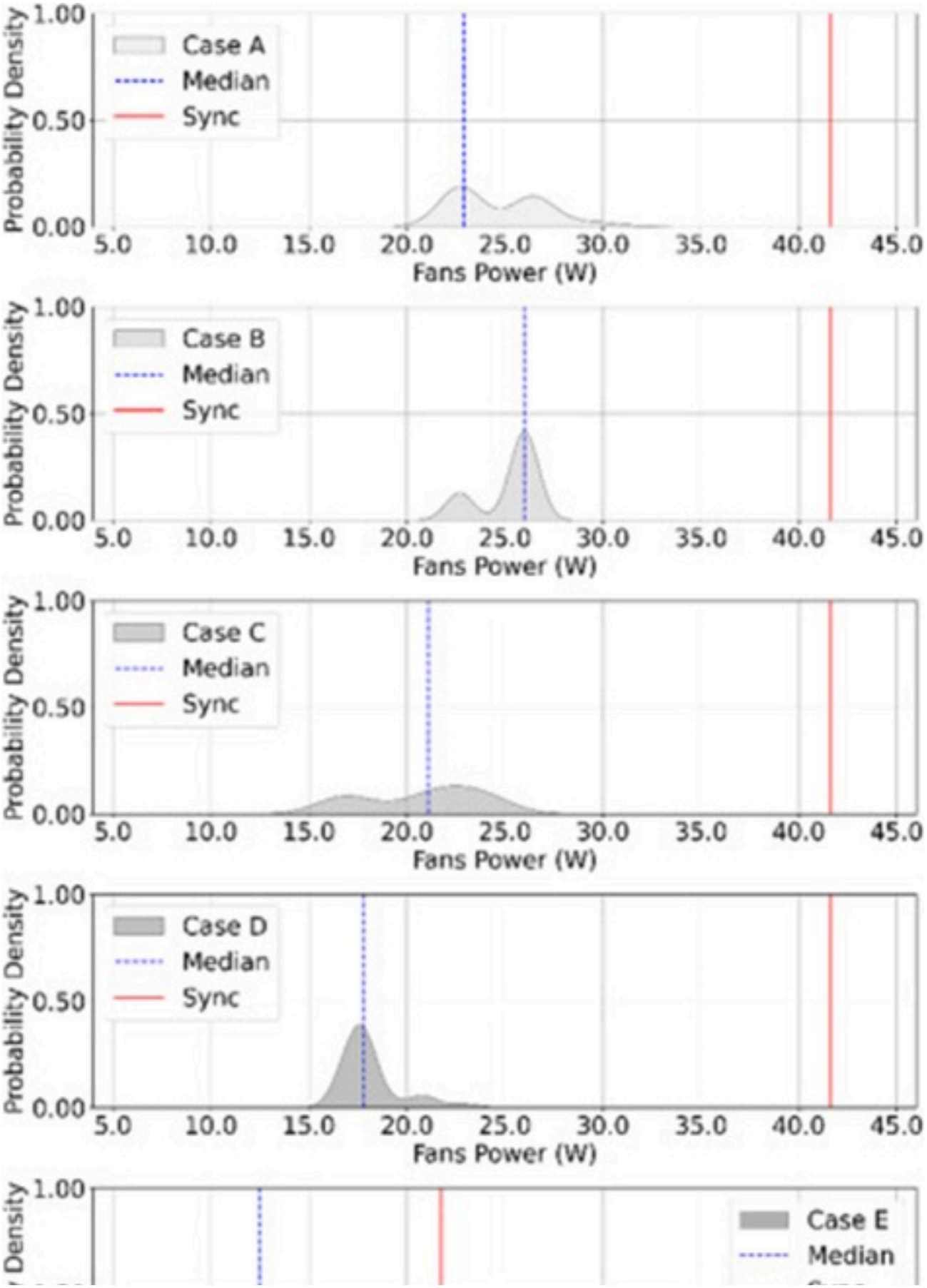
© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

under independent operation for each of the Cases A through H. The dashed blue vertical line pinpoints the median of these distributions across the diverse workload configurations. The unbroken red vertical line marks the cumulative fan power consumed when all fans operate in lockstep. As **Figure 39** makes abundantly clear, the framework yields energy-saving benefits in each of the eight workload configurations examined (Cases A through H).

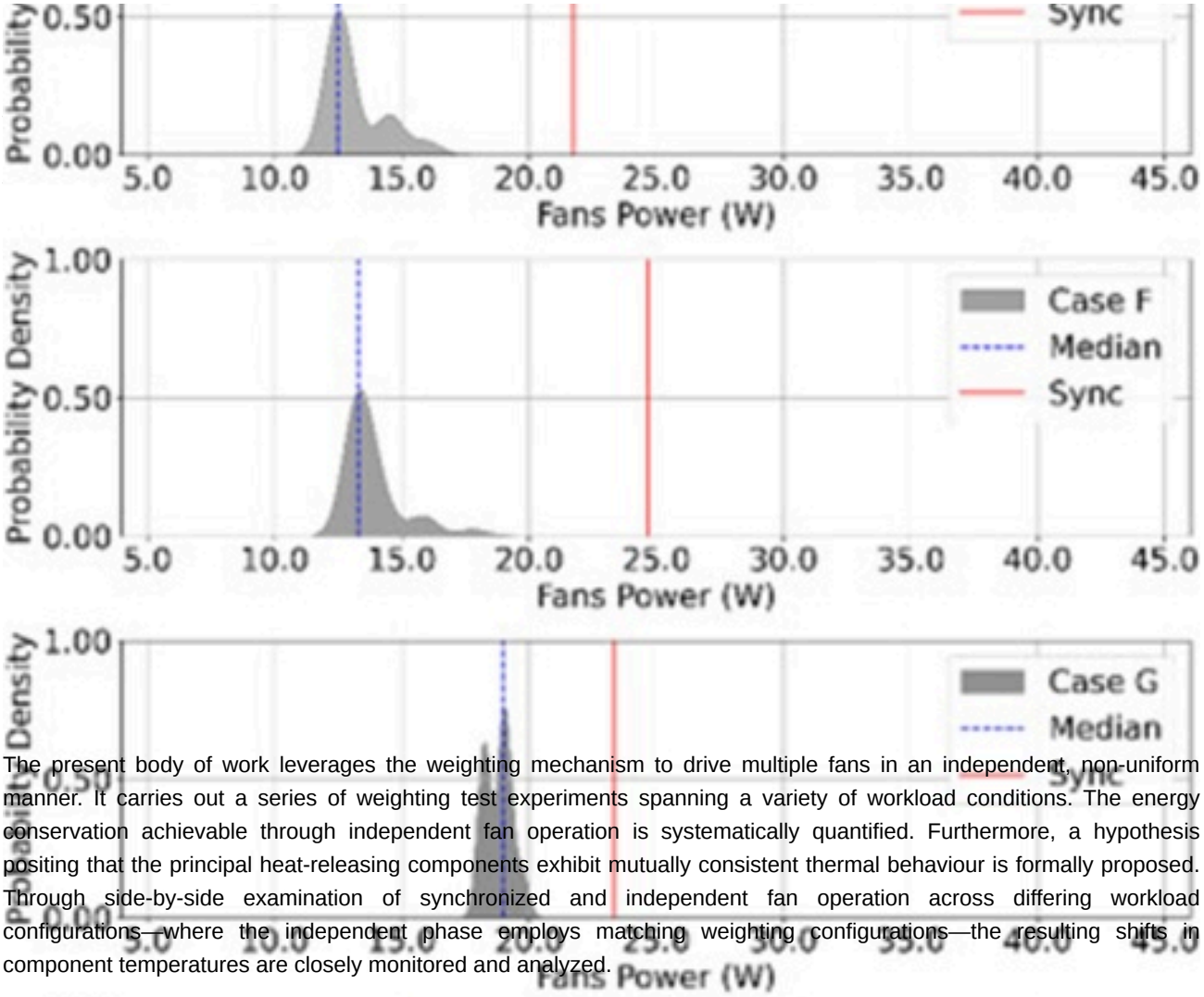
Table 9. Weighting search — the system's fans power statics.

cenario	Synchronized Fan Operation	Optimized Multi-Fan Operation				
	Power Consumption (W)	Average Power Savings (relative to synchronized)	Average Power (W)	Maximum Power (W)	Median Power (W)	Minimum Power (W)
A	41.6	40.7%	24.7	30.1	22.9	22.7
B	41.6	39.5%	25.2	26.2	26.0	22.7
C	41.6	50.0%	20.8	24.8	21.1	16.9
D	41.6	56.5%	18.1	23.0	17.8	16.9
E	21.7	39.3%	13.2	16.0	12.5	12.5
F	24.7	43.7%	13.9	17.8	13.3	13.2
G	23.3	19.5%	18.8	19.8	19.0	18.2
H	28.0	43.9%	15.7	25.1	15.3	15.3

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press



© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press



The present body of work leverages the weighting mechanism to drive multiple fans in an independent, non-uniform manner. It carries out a series of weighting test experiments spanning a variety of workload conditions. The energy conservation achievable through independent fan operation is systematically quantified. Furthermore, a hypothesis positing that the principal heat-releasing components exhibit mutually consistent thermal behaviour is formally proposed. Through side-by-side examination of synchronized and independent fan operation across differing workload configurations—where the independent phase employs matching weighting configurations—the resulting shifts in component temperatures are closely monitored and analyzed.

Realizing thermal compliance coupled with energy minimization demands that independent fan operation be paired with carefully chosen weighting configurations. To arrive at such configurations expeditiously, the present work integrates the weighting mechanism with the Evolutionary Strategy algorithm and devises two distinct families of fitness functions, thereby giving concrete form to the server multi-fan regulation framework. To substantiate the practical value of this framework, weighting search experiments are undertaken and replicated across thirty runs to neutralize the effects of algorithmic randomness. Guided by fitness evaluations, the algorithm iteratively refines the direction of its exploration through the weighting space, homing in on near-optimal configurations after a tightly bounded number of search iterations. The resulting independent fan operation can respect all thermal constraints while delivering near-optimal energy thrift. The inherently stochastic nature of the search process reduces the probability of premature convergence to a local optimum, making it well-suited to parameter optimization tasks commonly encountered in engineering practice. The experimental apparatus centres on a 2U server equipped with four air movers and four principal heat-dissipating components—CPU0, CPU1, PCIe0, and PCIe1. The experimental programme confirms three pillars: the practical feasibility of independent fan operation, the validity of the hypothesis that critical components exhibit congruent thermal behaviour, and the strong performance of the server multi-fan regulation framework advanced in this study.

1. Eight distinct server workload configurations reveal that when the entire catalogue of candidate weighting configurations is exercised for independent fan operation, the absolute number of admissible weightings, their proportional representation, and the identity of the most effective weighting all exhibit variation. The root cause of this variation lies in the differing power dissipation intensities of the components under each workload configuration, which in

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

turn modulate the thermal responses of the components, the drive ratios applied to the fans, and the cumulative fan energy consumption. Across Cases A through H, the most effective weightings coincide with the global optima for those respective workload conditions, satisfying thermal specifications while driving the cumulative fan energy draw to its absolute minimum from among the 369 candidate configurations. When the most effective weighting governs independent fan operation and is compared against the synchronized operation baseline, the ceiling drive output of the feedback controller is reduced. The independent scheme ameliorates the overcooling experienced by certain components without breaching thermal limits, slashing the cumulative fan energy draw substantially and yielding savings of 45.5%, 45.5%, 59.4%, 59.4%, 42.5%, 46.6%, 22.1%, and 45.3% for Cases A through H, respectively. These figures robustly validate both the energy-saving advantages and the operational feasibility of independent fan modulation.

2. Once the principal heat-dissipating components and their geometric relationship to the cooling fans have been established, a hypothesis of congruent thermal behaviour linking the CPUs and the PCIe devices across varying workload configurations is put forward. In comparing synchronized and independent fan operation across those configurations, matching weighting configurations are applied during the independent phase. Observations of the steady-state component temperatures indicate that the CPUs display thermal behaviours that mirror one another. The physical explanation stems from the trajectory of the cooling airstream: ambient air drawn into the chassis first traverses the CPUs positioned in the forward section and absorbs their dissipated thermal energy before proceeding to the PCIe devices situated further downstream. The thermal load imposed by the forward CPUs thus directly shapes the thermal environment encountered by the rearward PCIe devices. Under conditions where matching weighting configurations are deployed for independent fan operation and the forward CPUs share equivalent dissipation intensities, the PCIe devices likewise exhibit thermal behaviours that are replicas of one another. The experimental data furnish empirical support for the hypothesis of congruent thermal behaviour. Beyond this validation, the insights derived illuminate the reasons why the most effective weighting configurations differ between certain workload scenarios, and they provide further substantiation that server thermal regulation belongs to the class of nonlinear control problems.

In the weighting search experiments, the weighting mechanism was combined with Evolutionary Strategy (ES) algorithms and two families of fitness functions were designed to bring the server multi-fan regulation framework into operation. The experiments underwent thirty repetitions across eight different server workload configurations, a strategy adopted to mitigate the variability introduced by the algorithm’s random components. From the pool of 369 candidate weighting configurations, the algorithm physically tested an average of 53 distinct configurations across Cases A through H—a figure that represents a scant 14.2% of the complete set. In every case, the Euclidean distance metric ultimately registered zero. When placed alongside the weighting test experiments, the weighting search procedure boosted the proportion of admissible weightings by a mean of 18.8%. The experimental evidence points to a twofold conclusion: the algorithm not only attains convergence after a relatively modest number of search iterations, but it also expands the share of weightings that meet the thermal criteria.

The near-optimal weightings unearthed by the algorithm possessed, on average, a 47.5% likelihood of being the global optimum solution—where the global optimum is understood as the weighting configuration that minimizes cumulative fan energy consumption under independent fan operation. At the other end of the spectrum, the least favourable fitness scores registered among the near-optimal weightings—corresponding to the highest observed cumulative fan energy draws under independent operation—occurred with an average frequency of just 4.2%. Even these least favourable instances, however, still outperformed the synchronized fan operation baseline in terms of energy consumption. Summing up the evidence, the multi-fan regulation framework developed through this research, when benchmarked against synchronized fan operation, achieved an average reduction in cumulative fan energy draw of 43.1% across the eight workload configurations. These results strongly endorse the excellent performance delivered by the server multi-fan regulation framework.

The evolution strategy-based optimization procedure introduced here is conceived for offline execution: the ideal weighting configurations are determined in advance and then embedded directly into the firmware of the BMC (Baseboard Management Controller) before the server is deployed into production. This architectural choice guarantees

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

that the real-time computational footprint during live operation remains negligible, as the fan regulation system merely retrieves and applies the pre-optimized weights corresponding to the currently sensed workload state. That said, further research efforts are needed to investigate adaptive optimization techniques capable of dynamically reconfiguring weight values in real time, informed by streaming sensor data. Future inquiries may assess whether lightweight heuristic optimization methods or reinforcement learning frameworks can strengthen real-time decision-making while preserving the computational efficiency required for embedded deployment. While the independent fan regulation strategy advanced in this work has been demonstrated to optimize energy efficiency at the scale of an individual server, a comprehensive assessment of its behaviour within a full-scale data centre environment remains an open question. In particular, the influence of variable supply pressures characteristic of containment-based cooling architectures, and the dynamic coupling between server-level fans and rack-level airflow patterns, merit deeper investigation.

Furthermore, translating this independent fan regulation methodology onto commercial server platforms will necessitate implementation via industry-standard interfaces such as OpenBMC. Future development efforts will concentrate on migrating the experimental proof-of-concept into a production-ready system, ensuring seamless interoperability with established server management frameworks. These directions constitute important threads for ongoing and future research, extending the reach and applicability of the proposed regulation methodology well beyond the confines of the present experimental setup.

Acknowledgements

This study was supported by the Inventec Corporation of Taiwan. The financial aid and the equipments provided by the Inventec is gratefully acknowledged.

Conflict of interest

None

Financial support

None

Ethics statement

None

Received: 14 Mar 2022 Revised: 18 May 2022 Accepted: 03 Jun 2022
Published online: 15 June 2023

Rights and permissions

Open Access The author(s) retain copyright. This article is licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License. It may be shared and adapted for non-commercial purposes with appropriate attribution, an indication of changes, and distribution of adaptations under the same license. Third-party material may be subject to separate terms identified in its credit line. View the license at https://creativecommons.org/licenses/by-nc-sa/4.0/.

References

1. SRG. Q2 cloud market grows by 29% despite strong currency headwinds; Amazon increases its share [Internet]. 2022 [cited 2022 Oct 1]. Available from: https://www.srgresearch.com/articles/q2-cloud-market-grows-by-29-despite-strong-currency-headwinds-amazon-increases-its-share

2. SRG. Pipeline of over 300 new hyperscale data centers drives healthy growth forecasts [Internet]. 2022 [cited 2022 Oct 2]. Available from: https://www.srgresearch.com/articles/pipeline-of-over-300-new-hyperscale-data-centers-drives-healthy-growth-forecasts

3. Hintemann R, Hinterholzer S. Energy consumption of data centers worldwide - how will the internet become green? In: ICT4S. Lappeenranta, Finland; 2019.10-4.

4. Wei X, Xiao J, Wang Z, Wang Z, Tian Y. Integration and development of energy and information network in the pan-arctic region. Glob Energy Interconnect. 2019;2(6):504-11.

5. AKCP. The electricity consumption of data centers [Internet]. 2022 [cited 2022 Oct 1]. Available from: https://www.akcp.com/blog/the-real-amount-of-energy-a-data-center-use/

6. Koot M, Wijnhoven F. Usage impact on data center electricity needs: a system dynamic forecasting model. Appl Energy. 2021;291:116798.

7. Paris Agreement. Report of the conference of the parties to the United Nations framework convention on climate change (21st Session, 2015: Paris). In: Paris, France: HeinOnline; 2015. p. 2017.

8. Lennan M, Morgera E. The glasgow climate conference (cop26). Int J Mar Coast Law. 2022;37(1):137-51.

9. The Green Grid. Green grid metrics: describing datacenter power efficiency [Internet]. 2007 [cited 2022 Oct 1]. Available from: http://www.thegreengrid.org/gg_content/Green_Grid_Metrics_WP.pdf

10. Zhang Q, Meng Z, Hong X, Zhan Y, Liu J, Dong J, et al. A survey on data center cooling systems: technology, power consumption modeling and control strategy optimization. J Syst Architect. 2021;119:102253.

11. Wang Z, Bash C, Tolia N, Marwah M, Zhu X, Ranganathan P. Optimal fan speed control for thermal management of servers. In: International Electronic Packaging Technical Conference and Exhibition. San Francisco, California, USA; 2009 Jul 19-23. p. 709-19.

12. Shin D, Kim J, Chang N, Choi J, Chung SW, Chung EY. Energy-optimal dynamic thermal management for green computing. In: 2009 IEEE/ACM International Conference on Computer-Aided Design - Digest of Technical Papers. San Jose, CA, USA: IEEE; 2009 Nov 2-5. p. 652-7.

13. Lee C, Chen R. Optimal self-tuning pid controller based on low power consumption for a server fan cooling system. Sensors. 2015;15(5):11685-700.

14. Acun B. Mitigating variability in HPC systems and applications for performance and power efficiency [dissertation]. University of Illinois at Urbana-Champaign; 2017.

15. Brännvall R, Gabor H, Björkman M, Chen H, Xu W, Zheng W. Digital twin for tuning of server fan controllers. In: 2019 IEEE 17th International Conference on Industrial Informatics (INDIN). Vol. 1. IEEE; 2019.

16. Sarkinen J, Kunkel R, Williams T, Wang Z, Kumar A, Ellsworth M. Experimental analysis of server fan control strategies for improved data center air-based thermal management. In: 2020 19th IEEE Intersociety Conference on Thermal and Thermomechanical Phenomena in Electronic Systems (ITherm). IEEE; 2020.

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

17. Chu WX, Lien YH, Huang KR, Wang CC. Energy saving of fans in air-cooled server via deep reinforcement learning algorithm. Energy Rep. 2021;7:3437-48.

18. Chu HC, Lee CM, Tung KY, Chen R. Time series prediction for thermal management in server applied with deep learning. In: 2022 25th International Conference on Mechatronics Technology (ICMT). IEEE; 2022. p. 1-4.

19. Jonsson H, Moshfegh B. Modeling of the thermal and hydraulic performance of plate fin, strip fin, and pin fin heat sinks-influence of flow bypass. IEEE Trans Comp Packag Technol. 2001;24(2):142-9.

20. Byquist TA, Crocker MT. Method and system of advance fan speed control. United States patent US 7,490,479 B2. 2009.

21. Eiben AE, Smith JE. Introduction to evolutionary computing. Vol. 53. Springer; 2003.

22. Darwin C. On the origin of species. 1859. Routledge; 2004.

23. Kern S, Müller SD, Hansen N, Büche D, Ocenasek J, Koumoutsakos P. Learning probability distributions in continuous evolutionary algorithms–a comparative review. Nat Comput. 2004;3(1):77-112.

24. Schwefel HP. Numerical optimization of computer models. John Wiley & Sons, Inc.; 1981.

25. Hansen N, Ostermeier A. Completely derandomized self-adaptation in evolution strategies. Evol Comput. 2001;9(2):159-95.

26. Hansen N. The cma evolution strategy: a tutorial [Internet]. arXiv preprint arXiv:1604.00772. 2016 [cited 2022 Oct 1]. Available from: https://arxiv.org/abs/1604.00772

27. Lunacek M, Whitley D, Sutton A. The impact of global structure on search. In: Parallel Problem Solving from Nature – PPSN X. Dortmund, Germany: Springer; 2008 Sep 13-17. p. 498-507.

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press