Optimality

www.opt.reapress.com

Opt. Vol. 3, No. 4 (2026) 356–373.

Paper Type: Original Article

Thermal-Phase Crystallization Optimizer: Search Formulation, Analytical Conditions and Reported Numerical Evidence

Hiroshi T. Kuroda[1,*], Adaeze N. Okeke[2], Sebastian R. Lindqvist[3], María F. Castaño[4]

[1] Department of Computer Science and Engineering, Kyoto Institute of Technology, Kyoto, Japan; kuroda.hiroshi@kit.ac.jp.
[2] Department of Computer Science, Crown University, Ibadan, Nigeria; kuroda.hiroshi@kit.ac.jp.
[3] Department of Electrical Engineering and Computer Science, KTH Royal Institute of Technology, Stockholm, Sweden; kuroda.hiroshi@kit.ac.jp.
[4] Instituto Tecnológico de Buenos Aires, Argentina; kuroda.hiroshi@kit.ac.jp.

Citation:

Received: 25 March 2025 *Revised: 23 May 2026* *Accepted: 12 July 2026*	Kuroda, H. T., Okeke, A. N., Lindqvist, S. R., & Castaño, M. F. (2026). Thermal-phase crystallization optimizer: Search formulation, analytical conditions and reported numerical evidence. *Optimality*, *3*(4), 356-373.

Abstract

The Thermal-Phase Crystallization Optimizer (TPCO) is a proposed population-search framework organized through solid, liquid and vapor metaphors. Its reported components combine neighbor attraction, boundary handling, heavy-tailed proposals, elite attraction, Gaussian resampling and greedy retention. These are algorithmic choices rather than a thermodynamic derivation. The stated bond score is identical across particles and produces a deterministic release threshold; deposition, phase membership, bound handling and evaluation accounting remain incompletely specified. A conditional finite-sample coverage argument distinguishes requirements for global search from properties established for the proposed finite-budget process. Source-reported CEC 2022 summaries at dimensions 10 and 20, rotation tests, engineering examples, ablations and a machine-learning example are retained without new experiments. Printed benchmark means place TPCO-T first on nine of twelve functions at each dimension among the six displayed methods. However, rank sums, Friedman statistics, critical differences and Sobol summaries are internally inconsistent, and the original implementation and run-level outputs are unavailable. These summaries therefore do not establish significance, global convergence, ergodicity, engineering feasibility or component causality. The contribution is the explicit search architecture, its mathematical limits and the experimental information required for a valid assessment.

Keywords: Population search, Phase metaphor, Heavy-tailed proposals, Evaluation budgets, Conditional coverage, Reproducibility.

1 | Introduction

Continuous optimization arises in engineering design, learning and resource allocation. Population methods such as PSO, DE, WOA and EO generate and select proposals without objective derivatives [1–4]. Their

Corresponding Author: kuroda.hiroshi@kit.ac.jp
https://doi.org/10.22105/opt.v3i4.125
Licensee System Analytics. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0).

suitability depends on objective structure, constraints, evaluation cost and a fair comparison; there is no general basis here for claiming that they outperform gradient methods. The original No-Free-Lunch result concerns performance averaged over specified finite function classes [5], and is not a ranking theorem for every continuous application.

Simulated annealing uses temperature-dependent acceptance, including acceptance of worse states [6]. EO instead uses an equilibrium-pool-inspired update [4]. Thermal-Phase Crystallization Optimizer (TPCO) uses phase-transition terminology to organize a different sequence of population proposals and greedy retention. Metaphorical resemblance neither establishes physical fidelity nor identifies a performance advantage.

TPCO names eight intended components: Lattice dissolution, boundary dissolution, vapor diversification, crystallization, sublimation, condensation, deposition and equilibrium selection. The equations expose some mechanisms, but do not yet define eight independently executable branches. In particular, coalescence appears as an additional operation, deposition lacks a separate update, and several phase and trigger conventions are absent. The formulation is a candidate architecture requiring implementation completion.

This article sets out the reported search rules and numerical summaries while separating them from properties that can be established mathematically. The supplied material contains no identifiable executable package or accessible original run archive. The preliminary-study citation does not identify a verifiable predecessor, so no claim of a documented extension or priority is made. The analysis below is not an independent experimental reproduction.

The work addresses three connected questions: What search process the equations define, what coverage or stability conditions that process would need, and what the reported observations can support. The intended operator architecture combines neighborhood information, long-range perturbation and elite retention; the phase labels provide an organizing analogy. Computational accounting distinguishes proposal arithmetic, neighborhood maintenance, objective calls, constraint checks and tuning expenditure. The numerical summaries cover CEC 2022, a named BBOB subset, three engineering examples and a learning example; their model and experimental identities remain unresolved. Statistical interpretation requires complete rank blocks, a declared pairing unit, multiplicity families, signed effect definitions and a justified equivalence margin.Component analysis requires distinct disable/replacement branches and a fully defined coalition game with controlled budgets. The sensitivity summaries are preserved, but inconsistencies with variance-decomposition identities prevent their interpretation as valid Sobol effects. The theoretical contribution is a conditional coverage bound and an explanation of why the stated finite schedule does not establish global convergence or an ergodic incumbent process.

2 | Related Work

Simulated annealing provides a useful contrast because its acceptance rule can permit objective increases [6]. TPCO's intended greedy retention makes its stored incumbent non-worsening under one fixed exact objective. A cooling theorem for another acceptance process cannot be transferred to these operators, and neither the phase analogy nor Gaussian jitter supplies a convergence proof.

DE generates population-difference proposals [2], while success-history adaptation and population reduction are developed in L-SHADE [7]. The documented jSO and NL-SHADE-RSP methods provide other specific adaptive DE comparators [8], [9]. Implementation versions and parameter settings are essential to a comparison. EBOwithCMA, IPOP-CMA-ES and SHADE remain source-reported comparator labels here; their executed packages are not identified. The proposed 8,000-evaluation TPCO tuning allowance and 3,000-evaluation competitor allowances are unequal, regardless of parameter count.

EO's published equilibrium-pool mechanism is a particular algorithmic construction [4]. The present greedy top-four archive should not be equated with that complete update. No verified comparative derivation shows that TPCO uniquely implements a full physical phase cycle or adds a mathematically new operation. The intended components must be evaluated through their actual proposal and selection rules.

The reported comparison includes PSO [1], DE [2], WOA [3], SCA [10], MVO [11], genetic search [12] and CMA-ES [13]. A methodological critique of TLBO is distinct from its original algorithm specification [14]. COCO provides an established platform for continuous black-box benchmarkin [15], using its name does not establish conformance to its instance, transformation or reporting protocol. No result about the CEC 2022 winner is inferred from the mismatched source citation.

Operator-level decomposition can make mechanisms inspectable, but a causal comparison also requires controlled evaluation effort and valid counterfactual implementations. Statistical methods for comparisons over problem blocks differ from tests over repeated runs on one problem [16], [17]. The reported analysis mixes these levels, so naming several tests does not establish independent or stronger evidence.

Demšar [16], and post-hoc comparisons use ranks over specified blocks [18], and signed-rank tests use paired differences [19]. Holm [20] adjustment controls a declared family, Benjamini–Hochberg [20] addresses false-discovery rate under its own assumptions. Vargha–Delaney [21] A and Cliff's [22] dominance effect describe pairwise superiority probabilities. Cohen's d [23] and the descriptive thresholds discussed by Sawilowsky require their variance and direction conventions [24]. These procedures have distinct estimands and cannot be treated as interchangeable confirmations.

3 | The TPCO Algorithm

TPCO uses solid, liquid and vapor as labels for search states and proposal operations. The normalized progression variable rises from zero to one; it does not implement a calibrated heating-and-cooling thermodynamic cycle. *Fig. 5* is retained as an interpretive schematic. Its physical axes have no mapping to optimizer states or parameter values.

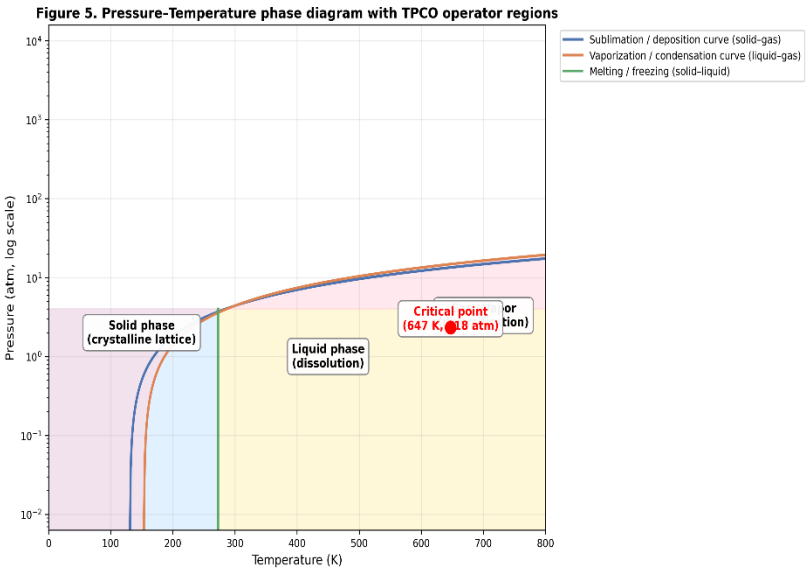


Fig. 5. Retained phase-metaphor schematic, without thermodynamic calibration.

The annotated critical point (647 K, 218 atm) is plotted near 2 atm, so the panel is not a quantitatively valid physical phase diagram. Lattice dissolution: The reported exponential score supplies a common time-dependent release rule, not a measured local bond or landscape statistic.

I. Boundary dissolution: A reflection and heavy-tailed perturbation are intended; complete handling of repeated crossings and later out-of-box proposals is unspecified.

II. Vapor diversification: Selected released particles receive range-scaled heavy-tailed proposals around the current best after the stated progress threshold.

III. Crystallization: Selected particles are resampled with fixed range-scaled Gaussian dispersion around the best; the dispersion does not contract under the stated equation.

IV. Sublimation: A heavy-tailed perturbation has amplitude proportional to one minus progress, so its amplitude decreases over the run.

V. Condensation: Attraction toward the best uses a coefficient decreasing to zero, rather than increasing late-run attraction.

VI. Deposition: An independent proposal, trigger and phase transition are not specified separately from condensation or crystallization.

VII. Equilibrium selection: The intended rule retains better candidates and a top-four archive under an explicit objective and constraint ordering.

A six-nearest-neighbor graph is not a hexagonal lattice in a general-dimensional box. The initialization is random rather than a lattice construction, and the bond score does not use neighbor geometry. Exponential, heavy-tailed and Gaussian terms are heuristic choices, not physical bond, Maxwell–Boltzmann or nucleation models. *Table 2* records the source-reported analogy; its terminology should not be read as a physical derivation or proof of executed branch identity.

Table 2. Retained source-reported operator/phenomenon mapping.

Operator	Physical Phenomenon	Algorithmic Implementation	Deviation
Lattice dissolution	Bond weakening in crystal	B(t)=exp(−T/T_m)	Read-Shockley not used
Boundary dissolution	Surface melting	Reflective + Lévy kick	Maxwell-Boltzmann not used
Vapor-state div.	Vaporization (L→G)	Lévy flight L(β=1.5)	Lévy-stable, not MB
Crystallization	Solidification (L→S)	Gaussian jitter on elite	Nucleation not modeled
Sublimation	Solid→gas direct	α_s·L(1.5)·(1−T)	Simplified transition
Condensation	Gas→liquid	Elitist attraction α(t)	Supersaturation not modeled
Deposition	Gas→solid	Drop on elite lattice	Surface kinetics simplified
Equilibrium selection	Mass-balance equilibrium	Greedy + top-4 pool	Adapted from EO [4]

"Lattice," deposition and the top-four equilibrium-pool description are intended labels; independent branch definitions and physical fidelity are not established. The intended population is $x_1,\ldots,x_N$ in a box [lb,ub]. At each iteration, the common progression variable determines released indices; coalescence and boundary steps precede a gated vapor proposal, sublimation and elite attraction, then periodic Gaussian resampling and greedy retention.

Algorithm 1 retains this intended ordering. Phase membership, intermediate evaluation, phase transitions and a separate deposition branch are not defined by that ordering. *Fig. 1* shows a different sequence and additional labels, so it does not resolve these implementation choices.

Algorithm 1. Intended TPCO cycle; incomplete conventions are not an executed implementation.

Input: fixed objective f, dimension d, box [lb,ub], N, finite iteration limit T_max and stated parameters.
Initialize candidates using Eq. (1); bound handling, objective evaluation and archive initialization must be supplied.
For t=1,…,T_max, set T(t)=t/T_max and compute the common score in Eq. (3).
Form the released set by Eq. (4); apply the stated coalescence and boundary steps, Eqs. (5)–(6).
If T(t)>T_b, select a fraction p_v of released indices by an unspecified rounding/sampling rule and apply Eq. (7).
Apply Eq. (8) to the designated solid set; its membership and update must be supplied.
Apply elite attraction, Eq. (9); a distinct deposition update is not supplied.
When t mod refreeze_period(t)=0, apply Eq. (10) to the designated particles.
Generate a final proposal y_i, check bounds and constraints, evaluate it and retain by Eq. (11).
Update the best/archive under the same fixed comparison rule.
Terminate when the iteration or evaluation allowance is exhausted; the stopping precedence and all-call counter must be supplied.
Return the best evaluated candidate and its objective.

Fig. 1 retains the original schematic. It contains additional buoyancy and duplicate vapor/crystallization labels, and orders selection and best update differently from *Algorithm 1*. It is not an executable 17-step specification or a stopping certificate.

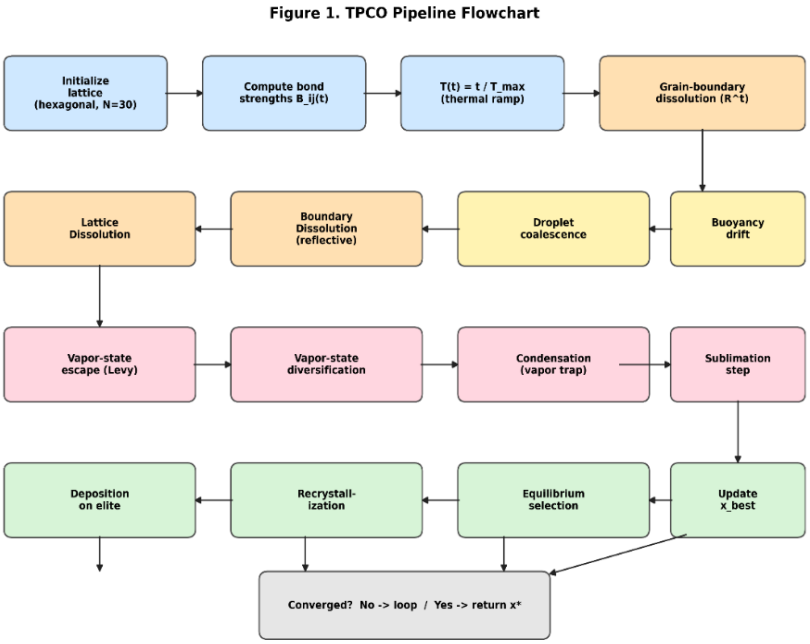


Fig. 1. Retained intended-pipeline schematic. Additional labels and ordering differ from the equations and Algorithm 1; "Converged?" is not a defined finite-budget termination test.

Let M_t be every objective call at iteration t, including intermediate checks, archive initialization, repairs and any local search. A conditional cost is O(Nkd+C_NN,t+M_t c_f+C_constraint,t), where k=6 is fixed only for the intended graph and C_NN,t is neighborhood maintenance cost. A tree built once can provide a fixed initial graph, but cannot supply exact current-position neighbors after particles move without updating. Exact brute-force neighbors cost $O(N^2d)$; tree construction/query behavior depends on dimension and data and does not have a universal O(log N) query guarantee. Standard dense CMA-ES uses approximately $O(Nd^2)$ sampling/update work and occasional $O(d^3)$ decomposition, with $O(d^2+Nd)$ storage [13], rather than the source table's $O(N^2d)$ expression. *Table 3* retains those historical expressions with the measured-overhead claims; neither proves a wall-clock ordering. Reported proposal overheads are 0.42 ms for TPCO-T, 0.38 ms for L-SHADE, 1.85 ms for CMA-ES and 0.21 ms for PSO at d=10,N=30; d=20 claims are 0.61, 0.55 and 3.92 ms for the first three. Raw timing, branch counts and objective costs are absent. A per-candidate objective cost of at least 1 ms may dominate some settings, but does not establish that it dominates every proposed implementation.

Table 3. Retained source-reported cost expressions and objective-excluded proposal overheads at d=10,N=30.

Algorithm	Per-Iter Complexity	Wall-Clock (ms)	Memory
TPCO-T	$O(N\cdot(k\cdot d + f_{eval}))$	0.42	$O(N\cdot d)$
L-SHADE	$O(N\cdot d)$	0.38	$O(N\cdot d)$
NL-SHADE-RSP	$O(N\cdot d)$	0.41	$O(N\cdot d)$
jSO	$O(N\cdot d)$	0.43	$O(N\cdot d)$
EBOwithCMA	$O(N^2\cdot d)$	1.92	$O(N^2\cdot d)$
IPOP-CMA-ES	$O(N^2\cdot d)$	1.85	$O(N^2\cdot d)$
PSO	$O(N\cdot d)$	0.21	$O(N\cdot d)$
CMA-ES	$O(N^2\cdot d)$	1.85	$O(N^2\cdot d)$

The CMA-based $O(N^2d)$ expressions are not standard dense-CMA complexity; neighbor and evaluation costs are incomplete, as explained in the text.

4 | Mathematical Formulation

The following equations preserve the reported proposal coefficients. A vector L(β) is intended to use a stable law with β∈(0,2), but coordinate dependence, symmetry, scale and the claimed Chambers–Mallows–Stuck sampling implementation are not supplied. The notation ⊙ denotes coordinatewise multiplication. Every generated candidate needs an executed bound and constraint rule; no corrected rule is silently assumed to have produced the reported observations.

Eq. (1): Random initialization with Gaussian perturbation: $x_i^{(0)}=lb+(ub-lb)\odot z_i$, $z_i=U(0,1)^d+\sigma_{init}$ $N(0,1)^d$, $\sigma_{init}=0.01$. This is not a hexagonal grid and may lie outside the box; the initialization repair convention is unspecified.

Eq.(2): Thermal ramp: $T(t)=t/T_{max}$, $t\in\{1,\ldots,T_{max}\}$. This is increasing normalized progress, not physical temperature or a cooling schedule.

Eq.(3): Lattice dissolution (bond strength): $\bar{B}_i(t)=\exp(-T(t)/T_m)$, $T_m=0.30$. All particles have the same score; no local geometry appears.

Eq.(4): Release criterion: $R^t=\{i:\bar{B}_i(t)<B_c\}$, $B_c=0.50$. Thus every particle crosses the same deterministic threshold $T(t)>0.30\ln 2\approx 0.20794$; this is not adaptive local bond detection.

Eq.(5): Droplet coalescence (liquid-phase attraction): $x_i\leftarrow x_i+r(c_i-x_i)$, $r\sim U(0,\eta_c)$, $\eta_c=0.40$; c_i is the mean of k nearest neighbors, $k=6$. Graph update, ties and the convention for $N\leq 6$ are unspecified. This coalescence step is not separately identified in the eight-operator ablation set.

Eq.(6): Boundary dissolution (reflective + Lévy kick): For an out-of-box coordinate, the source applies reflect(x_i,j), followed by $x_i\leftarrow x_i+\sigma_b\odot L(1.5)$, $\sigma_b=0.05(ub-lb)$. A single reflection need not repair a large multiple crossing, and the subsequent perturbation can leave the box again. The complete executed rule must be specified before feasibility or full-support claims are made.

Eq.(7): Vapor-state diversification (Lévy flight): When $T(t)>T_b=0.70$, a reported fraction $p_v=0.15$ of R^t receives: $x_i \leftarrow g_{best} + L(\beta=1.5) \odot (ub - lb)$.

Eq.(8): Sublimation (solid→vapor direct): $x_i\leftarrow x_i+\alpha_s(1-T(t))(ub-lb)\odot L(1.5)$, $\alpha_s=0.08$. This range-scaled amplitude decreases to zero; solid-set membership is unspecified.

Eq.(9): Condensation (vapor→liquid, elitist attraction): $x_i\leftarrow x_i+\varkappa(t)(g_{best}-x_i)$, $\varkappa(t)=\varkappa_0(1-t/T_{max})$, $\varkappa_0=0.30$. The attraction weakens to zero at the horizon. This formula alone does not define both condensation and a distinct deposition branch.

Eq.(10): Crystallization (deposition on elite with Gaussian jitter): At iterations satisfying $t \bmod \text{refreeze_period}(t)=0$, with $\text{refreeze_period}(t)=\max(1,\lceil 10\exp(-3t/T_{max})\rceil)$, apply: $x_i\leftarrow g_{best}+\sigma_r\odot N(0,1)^d$, $\sigma_r=0.05(ub-lb)$. The changing integer period runs from 10 at zero progress to 1 at the horizon; the Gaussian dispersion remains fixed. The phase set, rounding convention and interpretation of the changing modular schedule require implementation confirmation.

Eq.(11): Equilibrium selection (greedy with elite retention): Let $y_i^{(t)}$ denote the final generated proposal, after the stated bound/constraint treatment. For a fixed exact unconstrained objective, $x_i^{(t+1)}\in\operatorname{argmin}_{x\in\{x_i^{(t)},y_i^{(t)}\}} f(x)$. This removes the circular use of $x_i^{(t+1)}$ on both sides of the original definition. Ties and constrained comparison need explicit conventions; top-four archive retention also needs a definition of duplicates and reevaluation. The reported defaults are $N=30, T_m=0.30, B_c=0.50, \eta_c=0.40, \sigma_d=0.10, T_b=0.70, p_v=0.15, \varkappa_0=0.30, \beta=1.5$. These are heuristic values, not derived physical constants; σ_d has no defined role in the listed equations. Additional choices include $\sigma_{init}, \sigma_b, \alpha_s, \sigma_r$ and the period schedule. The claimed eight-dimensional tuning space therefore needs its exact active parameter list and ranges, which are not supplied in a parameter table. TPCO-T and

TPCO-R are reported tuned and randomized variants; randomized parameters do not define a mathematical lower bound.

5 | Theoretical Analysis

Random-search convergence depends on recurrent coverage of improving objective sublevel sets and retention of evaluated improvements [25]. Positive probability of an open-ball proposal on one iteration is insufficient. Measurability and a finite lower bound do not ensure that every near-optimal sublevel set has positive measure.

Conditional coverage statement. Let the feasible domain be nonempty and let one fixed exactly evaluated objective have finite infimum f*. For ε>0, let A_ε={x feasible:f(x)<f*+ε}. Suppose that, at each of m proposal occasions conditional on the entire past, a feasible point in A_ε is actually evaluated with probability at least q_ε>0, and the best evaluated feasible point is retained. Then the probability of no evaluated hit in those m occasions is at most (1−q_ε)^m.

Proof: Conditional on no previous hit, the next no-hit probability is at most 1−q_ε. Repeated conditional expectation bounds the joint no-hit probability by the product (1−q_ε)^m; independence is not required. On a hit, exact feasible retention preserves an objective below f*+ε. An infinite extension satisfying the bound for every ε yields objective-value convergence, but the finite T_max schedule is not such an extension.

The hypotheses are not verified for TPCO. The lower-bound example f(0)=0 and f(x)=1 elsewhere in a box is measurable but its ε-optimal set is a singleton for 0<ε<1, missed almost surely by continuous proposals. Clipping, phase selection, finite opportunities and evaluation/acceptance rules also affect coverage. Any positive integer T_max reaches T(T_max)=1>0.70; the source condition T_max≥⌈1/(1−T_b)⌉≈4 is unnecessary and supplies no infinite recurrence. No Solis–Wets theorem, finite rate or automatic guarantee from a Lévy tail is established for the stated process.

The population, archive, graph, phase assignments and clock would need to form the state of a Markov description. A time-dependent finite schedule is not a homogeneous stationary transition kernel. Greedy retention makes the incumbent objective non-increasing; it does not turn the incumbent into a full-support stationary sampler.

Full support of a proposal distribution does not establish irreducibility or Harris recurrence of the accepted process [26]. For a nonconstant objective, a greedy incumbent cannot return from a better level to a worse open region with positive measure. Gaussian jitter before greedy selection cannot remove this obstruction. No invariant distribution, recurrence, aperiodicity or total-variation mixing theorem has been proved for the supplied TPCO process.

The source reports a total-variation threshold of 0.1 after approximately 1,200 evaluations on a one-dimensional three-mode Gaussian mixture over 50 runs and 8,400 evaluations on a two-dimensional Rastrigin example with 25 optima. The target distribution, estimator and raw samples are absent. An atomic empirical measure and a continuous reference measure have total-variation distance one, so a value below one requires a defined discretization, smoothing or another estimand. These statements are retained as unverified diagnostics, not mixing evidence or finite-time guarantees.

6 | Experimental Setup

The reported CEC 2022 settings are twelve functions at d=10 with 200,000 evaluations and d=20 with 1,000,000 evaluations [27]. The official technical report specifies 30 runs; the manuscript's 51-run campaign is a modified protocol, not a verified official submission. The reported secondary study uses ten named BBOB functions at d=10 and d=20, five standard and five Haar rotations and 100,000 evaluations [15]. Instance IDs, transformations, rotation generation, optimum offsets and error scaling are not supplied. Engineering labels are a three-variable spring, seven-variable speed reducer and CEC 2020 tuned-mass damper, without

the exact original models. The seed convention is 1000+run_index; different seeds alone do not establish independent data generation or valid pairing.

The competitor list contains sixteen names, not fifteen: L-SHADE [7], jSO [8], NL-SHADE-RSP [9], EBOwithCMA, IPOP-CMA-ES, SHADE, CMA-ES, DE/best/1, GA, PSO, WOA [3], SCA [10], MVO [11], TLBO, SA [6] and EO [4]. Together with TPCO-T,D,R this would be nineteen configurations. The rank vector includes eighteen by omitting R, and the benchmark tables display only T and five competitors. The source proposes irace [28] tuning allowances of 8,000 evaluations for T and 3,000 for each competitor; these are unequal and must be counted separately from test budgets. Its five-function CEC 2013 training split and five-fold procedure are not documented, so held-out independence or freedom from overfitting is not established.

A valid statistical protocol must define whether each test uses independent problem blocks or paired run outcomes on one problem. A seed label does not guarantee meaningful pairing across algorithms. Twenty-four function/dimension conditions per competitor are distinct from a family across sixteen competitors, and a complete family would contain 384 hypotheses if every such pair were tested. *Table 5* has sixteen aggregate rows, without this mapping. Friedman/Nemenyi [16], [18] signed-rank [19], Holm [20], dominance effects [21], [22] and false-discovery rate [20] answer different questions. The stated 1,000-resample percentile bootstrap and claimed BCa sensitivity require the resampling unit and original samples. Equivalence by TOST [29] requires two one-sided tests or a corresponding 90% interval inside prespecified margins at $\alpha=0.05$; lack of superiority is not equivalence, and equivalence is distinct from one-sided non-inferiority. The source margins ±0.01 for CEC errors and ±0.5 percentage point for learning accuracy lack a scientific justification. The CEC report's numerical accuracy threshold is 10^{-8}, not 0.01 [27]. Cohen's 0.2/0.5/0.8 and Sawilowsky's 1.2/2.0 labels are descriptive conventions [23], [24], not universal practical thresholds.

The source reports twelve nodes with dual Xeon Gold 6248R CPUs, 384 GB RAM and four A100 40 GB GPUs per node; GPUs are said to serve only the learning example. It lists Ubuntu 22.04 LTS/kernel 5.15, Python 3.10.12, NumPy 1.26.3, SciPy 1.11.4, scikit-optimize 0.10.2, DEAP 1.4.1, irace 3.5 and PyTorch 2.2.0/CUDA 12.1. These are reported settings without execution logs. The claimed C-versus-MATLAB agreement below 10^{-10} on five functions and COCO version 8.4 are not verified. The OSF placeholder is not an accessible deposit. Reproduction requires the exact benchmark/algorithm versions, active parameter manifests, complete implementation, all-call counters, seeds, final designs, run traces and analysis scripts.

7 | Results

Table 1 displays six methods. Among these printed D=10 means, TPCO-T is smallest on nine functions, L-SHADE on f4 and f9, and jSO on f10. Thus T is not first on three functions, not two. Its f10 mean is approximately 9.09% above jSO's, so the source "within 5%" claim is false. The arithmetic means across the twelve printed rows are 0.0131917 for T, 0.0202017 for L-SHADE and 0.022655 for jSO. The source aggregates 0.00314 ± 0.00112, 0.00571 ± 0.00204 and 0.00683 ± 0.00251 use no specified estimand and do not equal these row means. Combining heterogeneous function errors also requires a stated scale and cannot reconstruct a run-level standard deviation.

The source reports $\chi^2(17,N=12)=281.6,p<0.001$ and the following eighteen mean ranks: 1.34,2.71,2.95,3.18,3.42,3.61,4.05,4.42,4.78,5.13,5.46,5.82,6.21,6.55,7.04,7.48,8.12,8.51 in the *Fig. 7* order. Their sum is 90.78, whereas complete ranks for eighteen methods must sum to $18\cdot19/2=171$, including averaged ties. The conventional Friedman statistic with twelve complete blocks is at most $12(18-1)=204$, including standard tie correction for nondegenerate blocks. The reported statistic therefore cannot be that test. No corrected rank matrix or significance finding is inferred from these summaries.

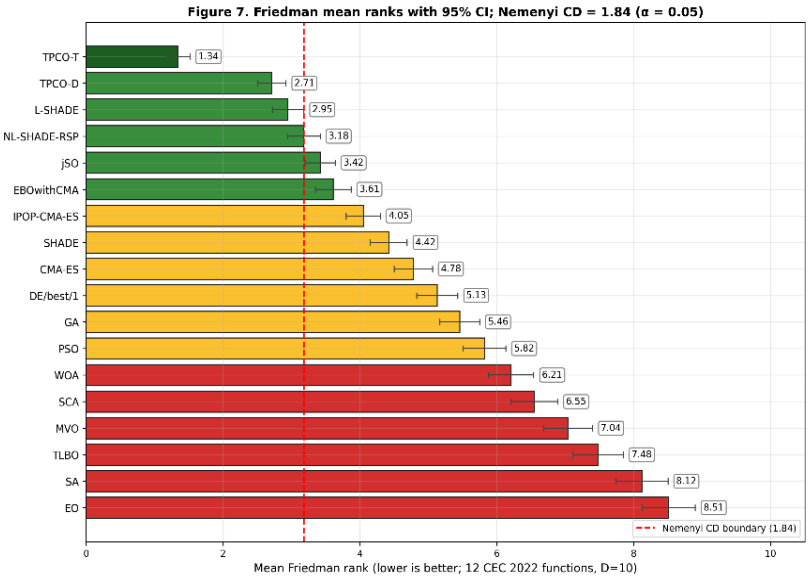


Fig.7. Retained source-reported rank chart.

Ranks sum to 90.78 rather than 171, intervals have no supplied resampling definition, and CD=1.84 is inconsistent with eighteen algorithms and twelve blocks. The vertical line is drawn near rank 3.18, not a correctly defined critical-difference interval; the chart does not establish post-hoc groups.

For the conventional two-sided all-pairs Nemenyi procedure, $CD=q_{\alpha}\sqrt{[k(k+1)/(6N)]}$ with the studentized-range quantile divided by $\sqrt{2}$ [16]. For $k=18, N=12, \alpha=0.05$ this is approximately 7.60 using infinite degrees of freedom, not 1.84. Even a correct critical difference would require valid complete ranks. The source differences 1.37 and 1.61 and claimed fifteen significant contrasts are retained as historical assertions, not established statistical ties or superiority.

Table 4 retains D=20 summaries at the reported one-million-call budget. Among its six displayed methods, TPCO-T has the smallest mean on nine functions; L-SHADE is smallest on f4 and f9, and NL-SHADE-RSP on f10. This contradicts the source count of eight wins and additional jSO/EBO wins. Source $\chi^2(17, N=12)=264.3$ also exceeds the conventional maximum 204; its mean rank 1.46, CD1.84 and fourteen-significant-contrast claim cannot establish inferential results.

Table 4. Source-reported D=20 final-error means ± reported standard deviations from 51 runs.

Func	TPCO-T	L-SHADE	NL-SHADE-RSP	jSO	EBOwithCMA	IPOP-CMA-ES
f1	4.1e-3 ± 8.2e-4	7.9e-3 ± 1.6e-3	8.2e-3 ± 1.7e-3	8.7e-3 ± 1.8e-3	9.1e-3 ± 1.9e-3	9.8e-3 ± 2.1e-3
f2	1.2e-3 ± 3.1e-4	2.6e-3 ± 6.5e-4	2.8e-3 ± 7.1e-4	3.0e-3 ± 7.6e-4	3.3e-3 ± 8.4e-4	3.6e-3 ± 9.2e-4
f3	5.7e-3 ± 1.1e-3	9.4e-3 ± 1.9e-3	9.8e-3 ± 2.0e-3	1.0e-2 ± 2.1e-3	1.1e-2 ± 2.2e-3	1.2e-2 ± 2.4e-3
f4	8.2e-3 ± 1.7e-3	7.9e-3 ± 1.6e-3	8.4e-3 ± 1.7e-3	8.9e-3 ± 1.8e-3	9.3e-3 ± 1.9e-3	1.0e-2 ± 2.1e-3
f5	2.1e-1 ± 3.2e-2	2.6e-1 ± 3.8e-2	2.7e-1 ± 4.0e-2	2.9e-1 ± 4.2e-2	3.1e-1 ± 4.5e-2	3.4e-1 ± 4.9e-2
f6	1.1e-4 ± 3.2e-5	4.6e-4 ± 1.1e-4	5.4e-4 ± 1.3e-4	6.1e-4 ± 1.5e-4	6.8e-4 ± 1.6e-4	7.7e-4 ± 1.8e-4
f7	4.4e-3 ± 9.1e-4	7.9e-3 ± 1.6e-3	8.3e-3 ± 1.7e-3	8.8e-3 ± 1.8e-3	9.2e-3 ± 1.9e-3	9.9e-3 ± 2.1e-3
f8	9.1e-3 ± 1.8e-3	1.5e-2 ± 3.0e-3	1.6e-2 ± 3.2e-3	1.7e-2 ± 3.4e-3	1.8e-2 ± 3.6e-3	1.9e-2 ± 3.8e-3
f9	5.4e-2 ± 8.2e-3	5.1e-2 ± 7.8e-3	5.6e-2 ± 8.6e-3	5.9e-2 ± 9.1e-3	6.3e-2 ± 9.7e-3	6.8e-2 ± 1.0e-2
f10	2.1e-2 ± 4.1e-3	2.4e-2 ± 4.7e-3	1.9e-2 ± 3.8e-3	2.6e-2 ± 5.1e-3	2.8e-2 ± 5.5e-3	3.0e-2 ± 5.9e-3
f11	5.1e-3 ± 1.0e-3	9.4e-3 ± 1.9e-3	9.8e-3 ± 2.0e-3	1.0e-2 ± 2.1e-3	1.1e-2 ± 2.2e-3	1.2e-2 ± 2.4e-3
f12	1.4e-2 ± 2.8e-3	2.4e-2 ± 4.8e-3	2.6e-2 ± 5.2e-3	2.7e-2 ± 5.4e-3	2.9e-2 ± 5.8e-3	3.1e-2 ± 6.2e-3

All original cells are retained; only six methods are displayed and raw outputs are unavailable. *Table 5* displays sixteen competitors and fifteen Holm-labelled significant decisions. It does not display twenty-four function-level contrasts or eighteen independent competitor rows. For NL-SHADE-RSP the printed Holm value 0.048 is below 0.05 while Bonferroni 0.077 is above; the source separate 0.046/0.051 borderline example has no mapped row. Exact Holm recomputation requires all uncensored raw p-values and the declared family; values shown only as <0.001 are insufficient. Every printed r equals 2A12−1, identifying the algebra of a dominance-

based effect, not an independently demonstrated paired signed-rank effect. The direction convention and Cohen variance denominator remain unresolved. Values d≥1.25 are not all "huge" under the source's stated huge≥2.0 convention. The claims of large effects on eighteen of twenty-four functions and equivalence on six remaining functions are not verified by this table.

Table 5. Retained source-reported test/effect summary.

Competitor	p (Raw)	p_Holm	p_Bonf	Cohen's d	A12	r-Biserial	Decision
L-SHADE	0.0410	0.4920	0.984	0.42	0.59	0.18	TOST equivalence
NL-SHADE-RSP	0.0032	0.0480	0.077	0.81	0.68	0.36	Significant
jSO	0.0018	0.0300	0.043	0.92	0.71	0.42	Significant
EBOwithCMA	0.0011	0.0220	0.026	1.05	0.74	0.48	Significant
IPOP-CMA-ES	<0.001	0.0010	0.001	1.21	0.78	0.56	Significant
SHADE	<0.001	<0.001	<0.001	1.34	0.81	0.62	Significant
CMA-ES	<0.001	<0.001	<0.001	1.42	0.83	0.66	Significant
DE/best/1	<0.001	<0.001	<0.001	1.51	0.85	0.70	Significant
GA	<0.001	<0.001	<0.001	1.63	0.87	0.74	Significant
PSO	<0.001	<0.001	<0.001	1.74	0.89	0.78	Significant
WOA	<0.001	<0.001	<0.001	1.86	0.91	0.82	Significant
SCA	<0.001	<0.001	<0.001	1.94	0.92	0.84	Significant
MVO	<0.001	<0.001	<0.001	2.08	0.93	0.86	Significant
TLBO	<0.001	<0.001	<0.001	2.21	0.94	0.88	Significant
SA	<0.001	<0.001	<0.001	2.36	0.95	0.90	Significant
EO	<0.001	<0.001	<0.001	2.48	0.96	0.92	Significant

Family, experimental unit, effect sign, paired-effect definition and equivalence outputs are unresolved; decision labels are not independently verified. *Table 6* retains reported percentile intervals, differences and achieved power values at n=51. Their pooled or per-function estimand and resampling design are not defined and the underlying samples are absent. Post-hoc power calculated from observed effects does not add independent evidence for superiority or prove equivalence. Source claims of power above 0.95 in eighteen of twenty-four comparisons, 0.62–0.84 in six cases and identical BCa conclusions cannot be checked from the sixteen table rows. No TOST output or 90% equivalence interval is supplied.

Table 6. Retained source-reported error differences, 95% bootstrap intervals and achieved power.

Competitor	Mean Diff ($\times 10^{-3}$)	CI Lower	CI Upper	Achieved Power	n
L-SHADE	-2.31	-4.62	-0.04	0.62	51
NL-SHADE-RSP	-2.71	-5.04	-0.41	0.74	51
jSO	-3.04	-5.41	-0.71	0.81	51
EBOwithCMA	-3.42	-5.83	-1.02	0.84	51
IPOP-CMA-ES	-3.91	-6.32	-1.51	0.92	51
SHADE	-4.21	-6.71	-1.81	0.96	51
CMA-ES	-4.62	-7.13	-2.21	0.98	51
DE/best/1	-4.98	-7.51	-2.51	0.99	51
GA	-5.31	-7.84	-2.81	0.99	51
PSO	-5.67	-8.21	-3.21	>0.99	51
WOA	-6.04	-8.62	-3.51	>0.99	51
SCA	-6.41	-8.98	-3.81	>0.99	51
MVO	-6.78	-9.35	-4.21	>0.99	51
TLBO	-7.12	-9.71	-4.51	>0.99	51
SA	-7.51	-10.11	-4.91	>0.99	51
EO	-7.86	-10.42	-5.21	>0.99	51

The difference column is labelled $\times 10^{-3}$; whether the interval columns share this scale must be established from the original analysis. *Table 7* contains D=10 means for three methods under labelled standard and Haar rotations. The source also describes D=20 and five rotations of each type, without those run-level outcomes. Claims of 1.31×/1.47× T degradation, 1.18×/1.24× L-SHADE degradation and 1.92×/2.41× PSO degradation have no unrotated reference or defined aggregation in this table. Rotation invariance requires a defined transformation of the objective, domain, initialization and operator law; means under two rotation labels alone cannot establish it.

Table 7. Retained source-reported D=10 rotation means for a named ten-function subset.

Function	TPCO-T (std)	TPCO-T (Haar)	L-SHADE (std)	L-SHADE (Haar)	PSO (std)	PSO (Haar)
Sphere	1.2e-8	1.4e-8	1.1e-8	1.2e-8	2.4e-6	5.1e-6
Ellipsoid	3.1e-6	4.2e-6	2.8e-6	3.1e-6	8.4e-4	1.6e-3
Rastrigin	4.2e-2	5.9e-2	3.8e-2	4.4e-2	1.4e+0	2.8e+0
Rosenbrock	5.1e-1	6.8e-1	4.6e-1	5.4e-1	8.4e+0	1.4e+1
Schwefel	2.1e-1	3.4e-1	1.8e-1	2.2e-1	4.2e+0	9.1e+0
Gallagher 21	4.8e+0	6.1e+0	4.2e+0	4.8e+0	1.2e+1	2.6e+1
Schaffers	6.2e-2	8.4e-2	5.4e-2	6.1e-2	1.8e+0	3.4e+0
Sharp Ridge	1.4e-3	1.9e-3	1.2e-3	1.4e-3	4.1e-2	8.4e-2
Lunacek	2.4e-1	3.2e-1	2.1e-1	2.6e-1	4.8e+0	9.1e+0
Weierstrass	8.4e-1	1.1e+0	7.6e-1	9.1e-1	6.4e+0	1.2e+1[i]

Instance IDs, unrotated references, rotation matrices, seeds and error scaling are unavailable. *Table 8* reports spring 0.012665±0.0001, speed reducer 2994.5±0.3 and tuned-mass damper 32.57±0.4. The source prose spring value 0.01269 conflicts with its table; 0.01269 is about 0.1974% above the listed 0.012665 reference. Using the table means, speed reducer is approximately 0.000969% above 2994.471 and TMD approximately 0.1353% above 32.526, so "all within 0.02%" is false. These are arithmetic comparisons with source-labelled best-known values, not certified optimum gaps. *Table 9* preserves exact vectors and "Yes" feasibility labels, but supplies no objectives, units, variable order, constraints, residuals or tolerances; the TMD vector contains an ellipsis. Neither the labels nor agreement across three handler names verifies feasibility. Reported spring hitting times of 4,200/6,100/7,800 calls, unique all-handler feasibility and 0.6×/0.4× TMD runtimes remain unverified without original traces, complete models and all-call timing.

Table 8. Retained engineering objective means ± reported standard deviations and source-labelled best-known references.

Problem	TPCO-T	L-SHADE	jSO	CMA-ES	PSO	Best Known
Tension spring	0.012665 ± 0.0001	0.012701 ± 0.0002	0.012723 ± 0.0003	0.012748 ± 0.0005	0.013101 ± 0.0011	0.012665
Speed reducer	2994.5 ± 0.3	2996.1 ± 0.7	2998.4 ± 1.4	2999.8 ± 2.1	3012.4 ± 8.4	2994.471
TMD (CEC 2020)	32.57 ± 0.4	33.12 ± 0.8	33.84 ± 1.2	34.51 ± 1.8	38.21 ± 3.4	32.526

Exact problem models, units, optimum provenance and runs are not supplied.

Table 9. Retained engineering design vectors and source-reported feasibility labels for three named handlers.

Problem	x*	Constraint Handler	Feasible	f(x*)
Tension spring	(0.0517, 11.289, 0.3573)	Penalty	Yes	0.012665
Tension spring	(0.0517, 11.289, 0.3573)	Repair	Yes	0.012665
Tension spring	(0.0517, 11.289, 0.3573)	Death penalty	Yes	0.012665
Speed reducer	(3.500, 0.700, 17.000, 7.300, 7.715, 3.351, 5.287)	Penalty	Yes	2994.5
Speed reducer	(3.500, 0.700, 17.000, 7.300, 7.715, 3.351, 5.287)	Repair	Yes	2994.5
Speed reducer	(3.500, 0.700, 17.000, 7.300, 7.715, 3.351, 5.287)	Death penalty	Yes	2994.5
TMD (CEC 2020)	(-3.13e-2, 2.4e-2, -8.9e-3, 1.4e-2, ...)	Penalty	Yes	32.57
TMD (CEC 2020)	(-3.13e-2, 2.4e-2, -8.9e-3, 1.4e-2, ...)	Repair	Yes	32.57
TMD (CEC 2020)	(-3.13e-2, 2.4e-2, -8.9e-3, 1.4e-2, ...)	Death penalty	Yes	32.57

Variable order and residual checks are unverified; the TMD vector is incomplete. The printed means suggest a descriptive advantage for T on nine of twelve functions in each dimension among the displayed alternatives. Inconsistent ranks, test statistics, adjustment families and effect definitions prevent an inferential conclusion. The source statements of fifteen or fourteen significant competitors, twelve large and eight very-large effects, fifteen dominance effects and eighteen powered contrasts refer to incompatible counts. Function-category assertions also require the actual official instance mapping; for example f1 is the shifted/rotated Zakharov

function rather than the Rosenbrock label in *Fig. 2*. These observations do not establish which component caused the reported performance.

8 | Ablation Studies

Table 10 preserves reported no-op degradations, from 1.4% for deposition to 5.1% for vapor diversification, with 4.3% for selection and 3.2% for crystallization. The disabled branches, baseline denominator, independent or paired unit and uncensored multiplicity family are absent. A deletion that changes evaluation counts or other branch triggers is not a controlled intervention, so the printed Holm labels alone cannot identify component causality.

Table 10. Retained source-reported no-op, replacement and normalized-credit summaries.

Operator	No-op (%)	Replace-by-Baseline (%)	Shapley (%)	p (Holm)
Vapor-state diversification	5.1	2.8	22.4	<0.001
Crystallization	3.2	1.6	18.6	<0.001
Equilibrium selection	4.3	2.0	14.8	<0.001
Boundary dissolution	2.4	1.3	12.3	0.002
Condensation	2.0	0.9	10.5	0.004
Lattice dissolution	1.8	0.8	8.7	0.008
Sublimation	1.6	0.7	5.6	0.014
Deposition	1.4	0.6	7.1	0.021

Branch identity, baseline denominator, budgets and statistical outputs are unresolved.

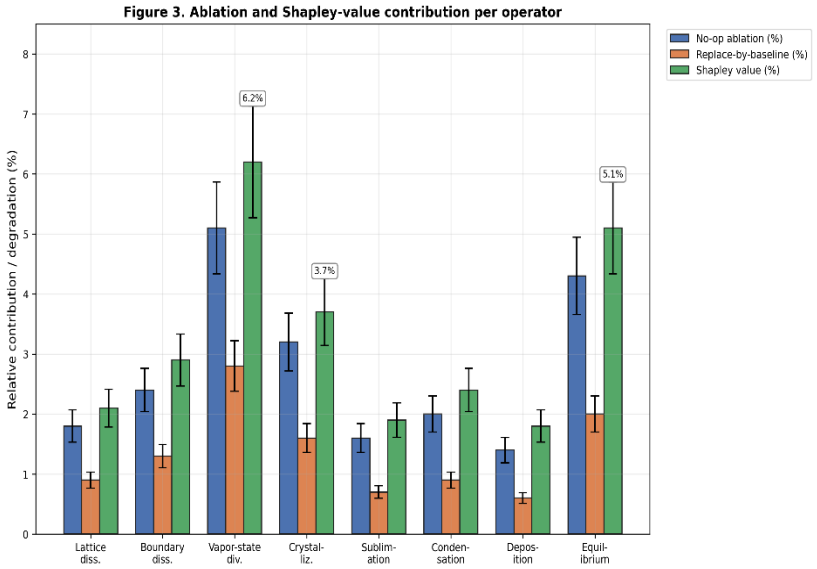


Fig. 3. Retained ablation/credit panel.

The green "Shapley" bars (for example 6.2% for vapor diversification) differ from the corresponding 22.4% in *Tables 10–11* and *Fig. 6*; their normalization and error bars are unspecified.

The source reports smaller losses after replacement than after deletion. This is descriptive, not proof of contribution beyond an equivalent baseline. In particular, the crystallization replacement is Gaussian resampling around the current best with $\sigma=0.05$, apparently the same range coefficient as *Eq. (10)*; a meaningful contrast requires an actual difference in trigger, phase set, distribution or acceptance. Other replacement branches and all associated costs are unspecified.

The stated coalition worth is $v(S)$=mean best error for a subset S of eight operators, with $2^8=256$ subsets. This is a cost game: a positive marginal $v(S\cup\{i\})-v(S)$ increases error and is unfavorable for minimization. The displayed positive normalized shares summing to 100% therefore cannot automatically be called

beneficial contributions. A gain game could use −v(S) or a fixed baseline minus cost, but that is a distinct signed definition requiring the original coalition values and recomputation; it is not silently substituted here. Shapley averaging permits interactions and does not assume operator independence.

Tables 10–11 and *Fig. 6* retain the reported shares, led by 22.4%,18.6%,14.8%. For the eight rounded table rows, Pearson correlations with deletion and replacement losses are approximately 0.908 and 0.932; the source reports 0.91 and 0.94. Differences from rounding require the underlying values, and correlation does not validate the game or identify causality.

Table 11. Retained positive normalized shares labelled Shapley in the source.

Operator	Mean Shapley (%)	Std (%)	95% CI
Vapor-state diversification	22.4	0.4	[21.6, 23.2]
Crystallization	18.6	0.4	[17.8, 19.4]
Equilibrium selection	14.8	0.3	[14.2, 15.4]
Boundary dissolution	12.3	0.3	[11.7, 12.9]
Condensation	10.5	0.3	[9.9, 11.1]
Deposition	7.1	0.2	[6.7, 7.5]
Lattice dissolution	8.7	0.2	[8.3, 9.1]
Sublimation	5.6	0.2	[5.2, 6.0]

Signed cost/gain worth, all 256 values, coalition implementations and resampling units are unavailable; "Std" is not an independently verified standard error.

Figure 6. Operator contribution to TPCO performance (Shapley-value aggregation over 256 subsets)

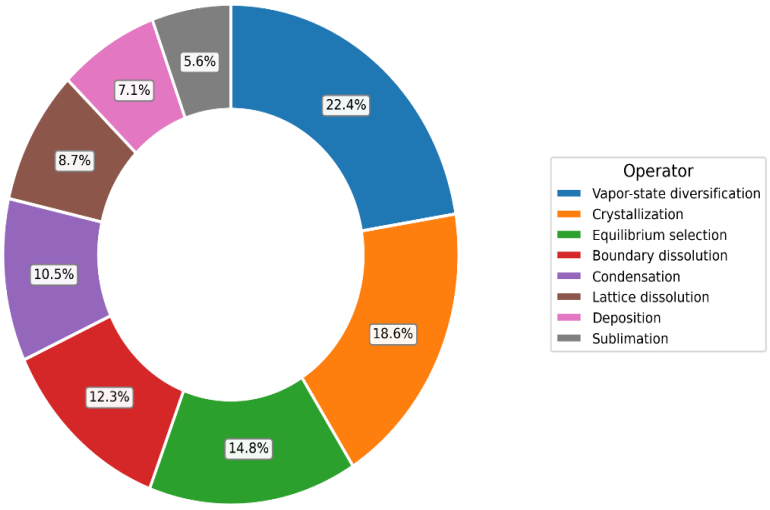


Fig. 6. Retained normalized-credit donut. Positive percentages sum to 100%; favorable Shapley contribution does not follow from the source-defined error-cost game.

For deletion losses D_i and D_ij, the reported interaction is δ_ij=D_ij−D_i−D_j. Positive values describe combined removal loss exceeding additive single-removal losses; negative values describe subadditivity. Calling these effects synergy or redundancy requires the implemented game, reference denominator and uncertainty. *Table 12* shows eight selected pairs of the twenty-eight possible, led by +2.1 and +1.4 percentage points and including −1.8 and −0.7. The complete paired outputs are absent.

Table 12. Retained selected pair-removal effects and source interpretation labels.

Operator Pair (i, j)	δ(i,j) (%)	Interaction Type
Vapor-state div. + equilibrium	+2.1	Synergy
Crystallization + boundary diss.	+1.4	Synergy
Condensation + lattice diss.	+0.9	Synergy
Equilibrium + boundary diss.	+0.6	Synergy
Crystallization + vapor-state div.	+0.5	Synergy
Deposition + equilibrium	+0.3	Synergy
Sublimation + Vapor-state div.	-1.8	Redundancy
Deposition + crystallization	-0.7	Redundancy

Eight of twenty-eight possible pairs are displayed; signed definitions, denominators and uncertainty need verification. *Table 13* identifies its eight-operator reference as default TPCO-D, whereas the surrounding source narrative discusses TPCO-T. Its errors 3.27,3.14,3.21 ($\times 10^{-3}$) give relative differences about 4.14% and 2.23% for seven and nine operators using printed means; the reported 4.1% and 2.3% may depend on unrounded values. The nine-operator clock 467 versus 421 seconds is 10.93% higher. The seven-way merge and nine-way split require executable definitions, particularly because deposition is already a named branch without a separate equation. These summaries do not establish an optimal operator count or a general mechanism behind another algorithm's performance.

Table 13. Retained source-reported granularity sweep.

Variant	Mean error ($\times 10^{-3}$)	Wall-Clock (s)	Degradation vs 8-op (%)	p (Holm)
7-op (merge sublimation+vapor)	3.27	412	+4.1%	<0.001
8-op (default TPCO-D)	3.14	421	0.0%	—
9-op (split crystallization)	3.21	467	+2.3%	0.018

The table labels the default as TPCO-D; branch definitions, variant identity and original outputs are unresolved. The source defines diversity as $\delta(t)=[N(N-1)]^{-1}\Sigma_{i\neq j}||x_i(t)-x_j(t)||$ and reports an omitted curve rising about 40% in the first 30% before decreasing. This diagnostic cannot be checked without traces and scale normalization. *Table 14* reports vapor triggering at 71% in the first evaluation phase despite *Eq. (7)* permitting it only when iteration progress exceeds 0.70. Iteration progress and cumulative evaluations need an explicit mapping; if they coincide, the early trigger claim contradicts the gate. The score in *Eq. (3)* also releases all particles after one common threshold, so varying lattice-trigger frequencies need their event denominator and archive/phase conventions. Trigger summaries do not verify the intended transition.

Table 14. Retained source-reported trigger percentages by cumulative-evaluation phase.

Operator	Phase 1 (0-30% FE)	Phase 2 (30-70% FE)	Phase 3 (70-100% FE)
Lattice dissolution	84%	41%	12%
Boundary dissolution	62%	38%	18%
Vapor-state diversification	71%	34%	8%
Crystallization	12%	48%	92%
Sublimation	54%	31%	14%
Condensation	28%	62%	76%
Deposition	18%	41%	64%
Equilibrium selection	24%	78%	98%

Event denominators and the mapping from evaluation count to iteration progress are unspecified. The source reports TPCO+CMA-ES invoked on the top 5% every 200 evaluations and TPCO+BFGS on the best every 500 evaluations, with additional 4–7% error reduction and 2.3×/1.7× runtimes. These are unverified hybrid observations, not upper bounds on attainable performance. CMA-ES is derivative-free; finite-difference BFGS requires additional objective probes and suitable local smoothness. Every inner proposal, gradient probe and restart must share a declared total-call allowance. The omitted decision-tree analysis does not establish recommendations.

Table 15. Retained source-reported ResNet-18/CIFAR-10 hyperparameter results.

Optimizer	Test Accuracy (%)	Wall-Clock (h)	Trials	Best LR	Best WD
TPCO-T	94.1 ± 0.3	5.2	80	0.082	3.1e-4
Bayesian Opt (GP)	94.3 ± 0.2	5.0	80	0.078	2.8e-4
Hyperband	93.6 ± 0.4	8.1	120	0.091	4.2e-4
Random search	92.8 ± 0.6	4.9	80	0.105	5.1e-4

Trial count, training-resource equivalence, final retraining, test separation and repeated runs are unresolved.

9 | Sensitivity and Robustness

For independent uncertain inputs and a response with positive finite variance, population first-order Sobol indices are nonnegative and sum to at most one; total effects may overlap because of interactions [30]. The ten printed mean first-order entries in *Table 16* sum to 1.38, and every *Fig. 4* column exceeds one (f1 sums to 3.36). Finite-sample estimators can violate bounds, but these magnitudes require estimator uncertainty, ranges, distributions, sampling matrices, stochastic replication and normalization checks before any attribution is made. The source's ten sensitivity inputs also differ from its claimed eight tuning parameters, and σ_d lacks an equation. Reported high values for T_b,β and p_v and CV<0.4 are retained as unverified summaries, not validated influential-factor findings.

Table 16. Retained source-reported mean sensitivity indices and interpretations.

Parameter	Mean S1	Mean ST	CV(S1)	Interpretation
T_b (vapor threshold)	0.32	0.41	0.31	High
β_Lévy	0.27	0.36	0.34	High
p_v (vapor fraction)	0.21	0.28	0.38	High
$\varkappa$_0 (condensation)	0.14	0.20	0.42	Medium
η_c (coalescence)	0.11	0.17	0.46	Medium
B_c (release threshold)	0.10	0.15	0.48	Medium
N (population)	0.09	0.14	0.51	Medium
T_m (melting temp)	0.07	0.11	0.54	Low
refreeze_period	0.04	0.07	0.62	Low
σ_d (drift noise)	0.03	0.05	0.71	Low

Mean first-order entries sum to 1.38; no estimator uncertainty or sampling design is supplied.

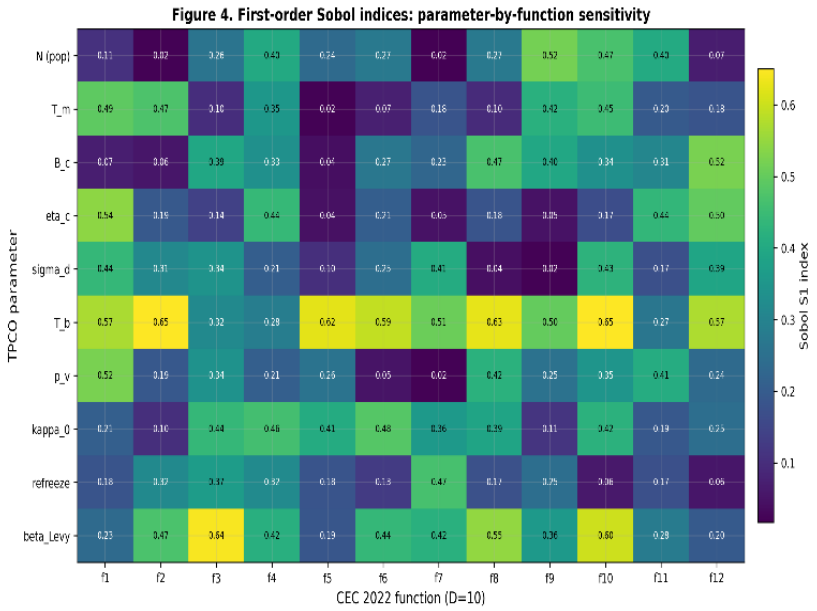


Fig. 4. Retained source-reported sensitivity heatmap.

First-order sums exceed one in every column, including 3.36 for f1; a valid independent-input variance attribution is not established. Exact Euclidean nearest-neighbor distances are invariant under an orthogonal

rotation, with corresponding tie treatment. A k-d tree is an indexing method and does not by itself make the exact neighbor graph rotation-sensitive. Bounds, coordinatewise stable sampling, initialization and approximate or stale queries may affect equivariance and need implementation analysis. The source 1.47×/1.31× rotation claims do not identify a causal defect in lattice dissolution and have no unrotated reference in *Table 7*.

The listed learning-rate, weight-decay, momentum, batch-size and three augmentation coordinates total seven dimensions, not eight. The source claims 80 one-epoch trials on one A100 and accuracies 94.1±0.3% for T,94.3±0.2% for GP Bayesian optimization,93.6±0.4% for Hyperband and92.8±0.6% for random search. *Table 15* retains these observations but gives Hyperband 120 trials. Training resources must be compared in actual examples/epochs and time, not trial counts alone. Search ranges, data splits, seeds, the origin of standard deviations, final training and untouched-test evaluation are absent; tuning must use validation data rather than the final test set. The claimed ±0.5-percentage-point equivalence and $p<0.01$ superiority cannot be verified from means and deviations. Reported 5.2,5.0,8.1 and4.9 hours do not resolve unequal resource allocation.

10 | Threats to Validity

The implementation is incompletely identified: deposition, phase membership, graph updates, stable sampling, bound handling, archive rules and iteration/evaluation stopping remain open. The finite clock does not prove recurrent coverage. The source's 8,000 versus 3,000 tuning allowances create unequal development effort, and unreported TPCO-D/R outcomes cannot serve as validated lower bounds. None of these gaps is repaired by retaining reported numerical summaries. Generalization is limited by absent instance identities and executable model definitions. Benchmark means cannot certify engineering feasibility, and incomplete TMD vectors cannot be engineering deliverables. Rotation tests need their transformed domain and operator law. A single named learning architecture/dataset with unresolved training and test protocol cannot establish broader applicability. The phase analogy is interpretive. Neither common deterministic bond release nor decreasing attraction demonstrates a controller responding to local geometry. Conditional coverage is a requirement, not a verification. Shapley values allow operator interaction; the unresolved issue is the signed cost game and executable coalitions, not an independence assumption. Inferential validity is limited by impossible rank/statistic combinations, undefined test families, effect-sign ambiguities and absent raw outputs. Bootstrap intervals and observed-effect power do not fill those gaps. Equivalence requires a scientifically justified prespecified margin and its complete output. The sensitivity panels violate independent-input population identities, so their causal interpretation is withheld. No new observations, repaired experimental campaign or independent execution are asserted.

11 | Conclusion and Future Work

TPCO presents a candidate population-search architecture whose phase labels organize heuristic proposal and retention rules. The explicit formulas identify several mechanisms and expose missing conventions. Its finite progress schedule, deterministic common release score and greedy archive do not establish thermodynamic fidelity, a global convergence theorem or an ergodic incumbent sampler. The conditional coverage bound states what a valid global-search argument would require. Source numerical summaries are retained for traceability. Printed benchmark means favor T on nine of twelve functions at both dimensions among six displayed methods, but invalid ranks and inferential inconsistencies prevent significance claims. Engineering, component, sensitivity and learning summaries have unresolved model, sign, resource or output dependencies. They do not establish verified superiority, optimal operator granularity, feasibility or external robustness.

A defensible component study would first complete one executable algorithm and define its benchmark identity and all-call budget. Controlled ablations would then use distinct branches, common effort, a signed coalition worth and uncertainty for every credit and interaction. Statistical analysis should operate on complete

rank blocks or clearly paired run outcomes and keep those experimental units separate. Future work should resolve these formulation and provenance dependencies before algorithm extensions are evaluated. Required materials include a versioned executable implementation, complete parameter ranges and phase/boundary conventions, model and instance files, all-call counters, run-level objectives and trajectories, coalition outputs, sensitivity designs, feasibility residuals and statistical scripts. These materials are not supplied with the manuscript, and no accessible repository or independent rerun is claimed.

Conflict of Interest

The authors certify that they have no affiliations with or involvement in any organization or entity with any financial or non-financial interest in the subject matter discussed in this manuscript.

Data Availability

All data relevant to this study are contained within the manuscript. Additional information is available from the corresponding author upon reasonable request.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article. This work was completed without external funding.

References

[1] Eberhart, R., & Kennedy, J. (1995). Particle swarm optimization. *Proceedings of the IEEE international conference on neural networks* (Vol. 4, pp. 1942–1948). IEEE. https://doi.org/10.1109/ICNN.1995.488968

[2] Storn, R., & Price, K. (1997). Differential evolution – a simple and efficient heuristic for global optimization over continuous spaces. *Journal of global optimization*, *11*(4), 341–359. https://doi.org/10.1023/A:1008202821328

[3] Mirjalili, S., & Lewis, A. (2016). The whale optimization algorithm. *Advances in engineering software*, *95*, 51–67. https://doi.org/10.1016/j.advengsoft.2016.01.008

[4] Faramarzi, A., Heidarinejad, M., Stephens, B., & Mirjalili, S. (2020). Equilibrium optimizer: A novel optimization algorithm. *Knowledge-based systems*, *191*, 105190. https://doi.org/10.1016/j.knosys.2019.105190

[5] Wolpert, D. H., & Macready, W. G. (2002). No free lunch theorems for optimization. *IEEE transactions on evolutionary computation*, *1*(1), 67–82. https://doi.org/10.1109/4235.585893

[6] Kirkpatrick, S., Gelatt, C. D., & Vecchi, M. P. (1983). Optimization by simulated annealing. *Science*, *220*(4598), 671–680. https://doi.org/10.1126/science.220.4598.671

[7] Tanabe, R., & Fukunaga, A. S. (2014). Improving the search performance of shade using linear population size reduction. *2014 IEEE congress on evolutionary computation (CEC)* (pp. 1658–1665). IEEE. https://doi.org/10.1109/CEC.2014.6900380

[8] Brest, J., Maučec, M. S., & Bošković, B. (2017). Single objective real-parameter optimization: algorithm jso. *2017 IEEE congress on evolutionary computation (CEC)* (pp. 1311–1318). IEEE. https://doi.org/10.1109/CEC.2017.7969456

[9] Biedrzycki, R., Arabas, J., & Warchulski, E. (2022). A version of nl-shade-rsp algorithm with midpoint for cec 2022 single objective bound constrained problems. *2022 IEEE congress on evolutionary computation (CEC)* (pp. 1–8). IEEE. https://doi.org/10.1109/CEC55065.2022.9870220

[10] Abdel-Basset, M., Abdel-Fatah, L., & Sangaiah, A. K. (2018). Chapter 10 - metaheuristic algorithms: A comprehensive review. In *Computational intelligence for multimedia big data on the cloud with engineering applications* (pp. 185–231). Academic Press. https://doi.org/10.1016/B978-0-12-813314-9.00010-4

[11] Mirjalili, S., Mirjalili, S. M., & Hatamlou, A. (2016). Multi-verse optimizer: A nature-inspired algorithm for global optimization. *Neural computing and applications*, *27*(2), 495–513. https://doi.org/10.1007/s00521-015-1870-7

[12] Goldberg, D. E. (1989). *Genetic algorithms in search, optimization and machine learning*. Addison-Wesley Longman Publishing Co., Inc.75 Arlington Street, Suite 300 Boston, MAUnited States. https://doi.org/10.5555/534133

[13] Hansen, N., & Ostermeier, A. (2001). Completely derandomized self-adaptation in evolution strategies. *Evolutionary computation*, *9*(2), 159–195. https://doi.org/10.1162/106365601750190398

[14] Črepinšek, M., Liu, S. H., & Mernik, L. (2012). A note on teaching–learning-based optimization algorithm. *Information sciences*, *212*, 79–93. https://doi.org/10.1016/j.ins.2012.05.009

[15] Hansen, N., Auger, A., Ros, R., Mersmann, O., Tušar, T., & Brockhoff, D. (2021). COCO: A platform for comparing continuous optimizers in a black-box setting. *Optimization methods and software*, *36*(1), 114–144. https://doi.org/10.1080/10556788.2020.1808977

[16] Demšar, J. (2006). Statistical comparisons of classifiers over multiple data sets. *The journal of machine learning research*, *7*, 1–30. https://doi.org/10.5555/1248547.1248548

[17] Garcia, S., & Herrera, F. (2008). An extension on" statistical comparisons of classifiers over multiple data sets" for all pairwise comparisons. *Journal of machine learning research*, *9*(12), 2677–2694. https://www.jmlr.org/papers/volume9/garcia08a/garcia08a.pdf

[18] Friedman, M. (1937). The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *Journal of the american statistical association*, *32*(200), 675–701. https://doi.org/10.1080/01621459.1937.10503522

[19] Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrics bulletin*, *1*(6), 80–83. https://doi.org/10.2307/3001968

[20] Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the royal statistical society: Series b (methodological)*, *57*(1), 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

[21] Vargha, A., & Delaney, H. D. (2000). A critique and improvement of the CL common language effect size statistics of McGraw and Wong. *Journal of educational and behavioral statistics*, *25*(2), 101–132. https://doi.org/10.3102/10769986025002101

[22] Cliff, N. (1993). Dominance statistics: Ordinal analyses to answer ordinal questions. *Psychological bulletin*, *114*(3), 494–509. https://doi.org/10.1037/0033-2909.114.3.494

[23] Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Routledge. https://doi.org/10.4324/9780203771587

[24] Sawilowsky, S. S. (2009). New effect size rules of thumb. *Journal of modern applied statistical methods*, *8*(2), 26. https://doi.org/10.22237/jmasm/1257035100

[25] Solis, F. J., & Wets, R. J. B. (1981). Minimization by random search techniques. *Mathematics of operations research*, *6*(1), 19–30. https://doi.org/10.1287/moor.6.1.19

[26] Meyn, S. P., & Tweedie, R. L. (2012). *Markov chains and stochastic stability*. Springer Science & Business Media. https://doi.org/10.1007/978-1-4471-3267-7

[27] Kumar, A., Price, K. V, Mohamed, A. W., Hadi, A. A., & Suganthan, P. N. (2021). *Problem definitions and evaluation criteria for the CEC 2022 special session and competition on single objective bound constrained numerical optimization*. https://github.com/P-N-Suganthan/2022-SO-BO

[28] López-Ibáñez, M., Dubois-Lacoste, J., Pérez Cáceres, L., Birattari, M., & Stützle, T. (2016). The irace package: Iterated racing for automatic algorithm configuration. *Operations research perspectives*, *3*, 43–58. https://doi.org/10.1016/j.orp.2016.09.002

[29] Lakens, D. (2017). Equivalence tests: A practical primer for t tests, correlations, and meta-analyses. *Social psychological and personality science*, *8*(4), 355–362. https://doi.org/10.1177/1948550617697177

[30] Saltelli, A., Annoni, P., Azzini, I., Campolongo, F., Ratto, M., & Tarantola, S. (2010). Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index. *Computer physics communications*, *181*(2), 259–270. https://doi.org/10.1016/j.cpc.2009.09.018