ISBN: 978-3-95450-261-5 ISSN: 2673-7000 doi: 10.18429/JACoW-NAPAC2025-WECN03

MULTI–OBJECTIVE BAYESIAN OPTIMIZATION OF AN ELECTRON INJECTOR LINAC FOR 4TH GENERATION LIGHT SOURCES: A COMPARATIVE STUDY WITH MOGA

C. S. Park*, Korea University, Sejong, South Korea

Abstract

The performance of electron injector linear accelerators (linacs) critically influences the beam brightness and stability in 4th generation light sources. In this study, we employ a multi-objective Bayesian optimization (MOBO) framework to optimize the injector linac design, targeting the simultaneous minimization of transverse emittances and energy spread at the linac exit. This data-efficient approach leverages Gaussian process regression and acquisition functions to navigate the high-dimensional design space with significantly fewer simulations than conventional methods. We compare the results of MOBO with those obtained from the well-established Multi-Objective Genetic Algorithm (MOGA), highlighting differences in convergence speed, solution diversity, and computational efficiency. Our findings demonstrate that MOBO achieves comparable or superior optimization outcomes with reduced computational cost, offering a powerful alternative for accelerator design and tuning in next-generation light source facilities.

INTRODUCTION

Fourth–generation synchrotron and FEL facilities require injector beams with ultra–low transverse emittance and also tightly controlled relative energy spread [1–3] For Korea–4GSR, the injector linac must minimize end–of–linac $\mathbf{f} = [\varepsilon_{n,x}, \varepsilon_{n,y}, \sigma_\delta]$ while meeting limits on beam size/divergence, bunch length, mean energy, and transmission [4]. This is challenging because the space–charge–dominated gun region couples strongly to downstream optics and RF phasing, so small changes in early knobs can propagate nonlinearly.

Multi–objective optimization (MOO) is well suited to expose these trade–offs via a Pareto set. Genetic algorithms (e.g., NSGA–II [5]) provide robust coverage on nonconvex fronts but typically require many high–fidelity simulations, which is costly when each evaluation is a 3D space–charge model.

Bayesian optimization (BO) offers a sample–efficient alternative by learning a probabilistic surrogate (Gaussian Process) and selecting new evaluations via an acquisition function that balances improvement and uncertainty [6]. While true MOBO methods based on expected hypervolume improvement exist, a practical route is to *scalarize* the objectives with normalized weights and run BO on the resulting single objective, $J(\mathbf{w}) = \sum_i w_i \tilde{f}_i$ with $\tilde{f}_i = f_i / f_{i,\mathrm{ref}}$.

In this work we study the Korea–4GSR injector with both approaches: NSGA–II as a high–quality Pareto baseline, and scalarized BO across nine weight triplets to reconstruct knees and extremes. The decision variables are eight operational knobs—solenoid field B_{sol}, focusing/defocusing quadrupoles $k_{\mathrm{QF}}, k_{\mathrm{QD}}$, gun phase ϕ_{gun}, and four cavity phases $\phi_{1..4}$—all evaluated at the linac end. BO uses GP surrogate and batched acquisition (qEI/qNEI).

Table 1: Decision Variables, Bounds, and Units Used in the Optimization

Variable	Range	Units
B_{sol}	[0.10, 0.30]	T
k_{QF}	[0, 10]	T/m
k_{QD}	[−10, 0]	T/m
ϕ_{gun}	[0, 360]	deg
ϕ_1, ϕ_2	[0, 360]	deg
ϕ_3, ϕ_4	[0, 360]	deg

INJECTOR AND OPTIMIZATION PROBLEM

Injector Overview and Evaluation Plane

The Korea–4GSR injector is a normal–conducting RF gun followed by a solenoid, four accelerating modules, and a short matching section with a focusing (QF) and defocusing (QD) quadrupole. The optimization is performed against high-fidelity 3D space–charge simulations (ASTRA [7]) and all figures of merit are evaluated at the end–of–linac diagnostics plane ($s \simeq 16.2$ m). Throughout this paper emittances are *normalized rms* and the energy spread is the rms fractional quantity $\sigma_\delta = \sigma_E / \langle E \rangle$.

Decision Variables

The decision vector $\mathbf{x} \in \mathbb{R}^8$ comprises eight operational knobs that are routinely accessible during tuning:

$$\mathbf{x} = [B_{\mathrm{sol}},\ k_{\mathrm{QF}},\ k_{\mathrm{QD}},\ \phi_{\mathrm{gun}},\ \phi_1,\ \phi_2,\ \phi_3,\ \phi_4].$$

Here B_{sol} is the on-axis solenoid field, $k_{\mathrm{QF/QD}}$ are the integrated gradients of the quadrupoles, and $\phi_{\mathrm{gun}}, \phi_{1\ldots4}$ are the RF phases of the gun and four downstream cavities (degrees). Hardware-realistic bounds are enforced.

Objectives

We target simultaneous reduction of the three end-of-linac objectives

$$\mathbf{f}(\mathbf{x}) = [\varepsilon_{n,x}(\mathbf{x}),\ \varepsilon_{n,y}(\mathbf{x}),\ \sigma_\delta(\mathbf{x})],$$

* kuphy@korea.ac.kr

ISBN: 978-3-95450-261-5 ISSN: 2673-7000 doi: 10.18429/JACoW-NAPAC2025-WECN03

each to be minimized. These objectives capture the transverse brightness and longitudinal quality required by the booster and storage rings.

Constraints and Feasibility

To ensure machine protection and ring acceptance, every candidate must satisfy the following end-of-linac constraints (treated as *hard* constraints in the comparison:

$$\begin{aligned} &\sigma_x, \sigma_y < 0.30 \text{ mm}, \qquad \sigma_{x'}, \sigma_{y'} < 0.30 \text{ mrad}, \\ &\sigma_z < 1.0 \text{ mm}, \qquad \langle E \rangle = 200 \text{ MeV } \pm 2.5\%, \\ &T > 99.99\% \, . \end{aligned}$$

Only constraint-satisfying samples are used for the Pareto set and hypervolume. During the search, MOGA applies Deb's constraint-domination rule, while BO uses a probabilistic feasibility filter (rejecting points with low feasibility) and a soft penalty near the boundary.

Scalarization for BO and Normalization

For Bayesian optimization we form a normalized scalar objective

$$J(\mathbf{w};\mathbf{x}) = \sum_{i=1}^{3} w_i \tilde{f}_i(\mathbf{x}), \; \tilde{f}_i = \frac{f_i}{f_{i,\text{ref}}}, \; \sum_i w_i = 1, \; w_i \geq 0, \quad (1)$$

where $f_{i,\text{ref}}$ are reference scales used to remove units and to prevent the weights from acting as unit conversions. We take $f_{i,\text{ref}}$ as the median per-objective value over the initial space-filling *feasible* design; results are insensitive to this choice provided the scales are $\mathcal{O}(1)$. To approximate a Pareto set with scalarized BO, we sweep nine weight triplets $\mathbf{w}$ covering the corners and knee region of the simplex, and take the union of the resulting optima.

METHODS: MOGA VS BO

Common Settings and Fairness

To ensure a one-to-one comparison: (i) identical variable bounds are used, as shown in Table 1; (ii) feasibility is evaluated with the same size/divergence, bunch length, mean energy window, and transmission limits; (iii) only feasible points contribute to Pareto sets and metrics (hypervolume, dominance counts). For convenience we normalize objectives by fixed scales $f_{i,\text{ref}}$ (median of the initial feasible pool) and use the normalized $\tilde{f}_i = f_i / f_{i,\text{ref}}$ when scalarizing.

MOGA (NSGA–II)

We adopt NSGA–II with standard variation operators and constraint handling:

- **Population and variation.** Simulated binary crossover (SBX) with probability p_c and index η_c, and polynomial mutation with probability p_m and index η_m. Offspring are created from the current population each generation.
- **Selection and diversity.** Non–dominated sorting by Pareto rank, with crowding distance for diversity preservation within each front.
- **Constraints.** Deb's constraint–domination principle: any feasible solution dominates an infeasible one; among infeasible solutions the one with smaller total violation is preferred. Infeasible individuals are carried only if the population lacks feasible solutions early on.
- **Termination.** The run stops at a fixed evaluation budget (MOGA total = $500 \times 300 \approx 150,000$ evaluations in this study).

The resulting non–dominated set provides a high–quality reference front for $\mathbf{f} = [\varepsilon_{n,x}, \varepsilon_{n,y}, \sigma_\delta]$.

Scalarized Bayesian Optimization

Because standard BO optimizes a single scalar, we combine the three objectives with nonnegative weights $\mathbf{w}$ that sum to one as shown in Eq. (1). We run BO independently for a small set of nine weight triplets that cover the simplex corners, edges, and a balanced knee, and then take the union of the per–weight optima as a BO–derived Pareto approximation.

Surrogate model. We model $J(\mathbf{w};\mathbf{x})$ with a Gaussian Process (GP). Inputs are scaled to $[0,1]$ per bound; the output is standardized per run. GP hyperparameters are fit by maximizing the log marginal likelihood with multi–start initialization.

Initialization and batching. Each BO run begins with a space–filling design of N_0 feasible points, followed by T BO iterations of batch size q (parallel evaluations). In this study N_0 and q were chosen so that each run consumed $\sim 10^2$ simulations (per–weight), giving a BO total of $\sim 9 \times 10^2$ simulations across all weights.

Acquisition and optimization. We *minimize* J; the batch acquisition is the q–point Expected Improvement for minimization,

$$\text{qEI}(X) = \mathbb{E}\Big[\max(0, \; J - \min_{x \in X} J(x))\Big], \qquad (2)$$

where J is the best (lowest) value observed so far and the expectation is under the joint GP posterior of $J(X)$. With observation noise we use *qNEI*, which integrates over the latent (noise–free) incumbent and predictions.

Constraints in BO. In the BO runs, constraints were not imposed as hard limits within the acquisition optimizer. Instead, we employed a posterior feasibility screen (reject if $\Pr(\text{feasible}) < \tau$) and, near the boundary, a large soft penalty on J; only feasible samples were retained for Pareto and hypervolume metrics.

ISBN: 978-3-95450-261-5 ISSN: 2673-7000 doi: 10.18429/JACoW-NAPAC2025-WECN03

Stopping criteria and outputs. Each run stops at its evaluation budget or when improvement in the best–seen J falls below a tolerance for several iterations. We record per–run best solutions, their objective triplets $(\varepsilon_{n,x}, \varepsilon_{n,y}, \sigma_\delta)$, and the full set of intermediate feasible points. The union across weights is compared against the MOGA front using Pareto dominance and hypervolume relative to a fixed reference point.

Compute Budgets and Complexity

MOGA requires $O(10^5)$ evaluations to densely cover the front for this problem, whereas the BO weight sweep uses $O(10^3)$ in total (~ 100 per weight $\times$ 9 weights). These $\sim$ two orders of magnitude reduction in simulator calls is the primary motivator for BO in commissioning scenarios, where repeated re–optimization is expected.

RESULTS

We summarize the comparative performance of MOGA (NSGA–II) and scalarized BO on the Korea–4GSR injector, using identical bounds, constraints, and diagnostics plane. All objectives are minimized at the linac end: $\mathbf{f} = [\varepsilon_{n,x}, \varepsilon_{n,y}, \sigma_\delta]$; feasibility requires compliance with size/divergence, bunch length, mean energy window, and transmission limits.

Pareto Baseline from MOGA

MOGA yields a smooth, nonconvex Pareto front with a pronounced knee where a small increase in σ_δ buys a large decrease in $(\varepsilon_{n,x}, \varepsilon_{n,y})$. Three reference points are retained for downstream checks: the σ_δ–minimum, an ε–lean solution (low $\varepsilon_{n,x}/\varepsilon_{n,y}$), and a knee solution (balanced trade). All are strictly feasible with margin, as shown in Fig. 1.

Scalarized BO: Per-Weight Behavior and Convergence

For each weight triplet $\mathbf{w}$, BO rapidly reduces the scalar objective $J(\mathbf{w})$ and then plateaus, with most gains in the first [50–60] iterations. Constraint handling keeps nearly all samples feasible after the initial design; the best-seen candidates per run satisfy all limits. Figure 2 shows best-seen J vs iteration and for all nine runs.

Pareto Knee from BO

Figure 3 overlays the per–weight incumbents from the nine BO runs on the MOGA reference front. The stitched

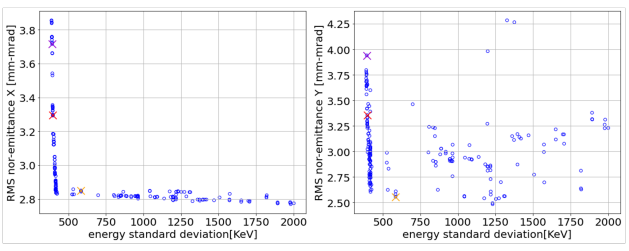

Figure 1: MOGA objective spaces across 9 weights.

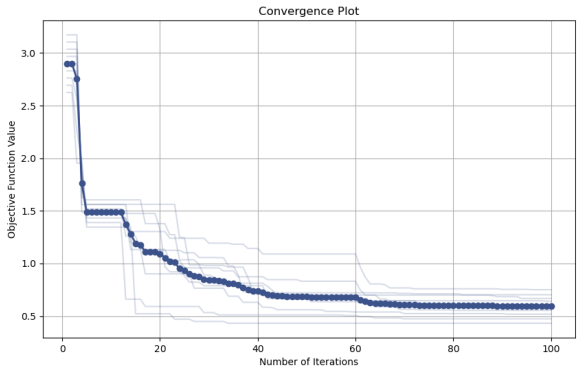

Figure 2: Scalarized BO convergence across 9 weights: best-seen $J(\mathbf{w})$ vs iteration.

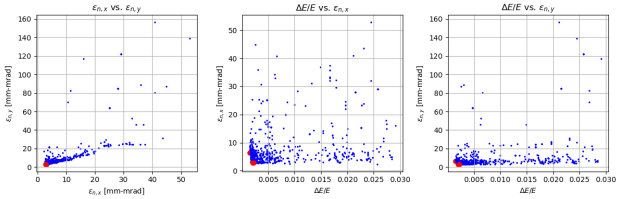

Figure 3: BO objective spaces.

BO set recovers the σ_δ–heavy and ε–lean extremes as well as the Pareto knee with comparable quality, while requiring roughly two orders of magnitude fewer simulations: $\sim 9 \times 10^2$ for BO versus $\sim 1.5 \times 10^5$ for MOGA. Across runs, BO approaches the reference front rapidly—most of the gain (e.g., in hypervolume) occurs within the first $\sim$ 50–60 iterations per weight.

Representative Operating Point and Beam Dynamics

For a σ_δ-heavy preference $\mathbf{w} = [0.1, 0.1, 0.8]$, we compare the best feasible solutions from MOGA and BO at the end-of-linac diagnostics plane ($s \simeq 16.2$ m). BO attains small but consistent gains in all three objectives (Table 2), consistent with slightly more off-crest late-cavity phases (reducing chirp) while maintaining emittance compensation.

DISCUSSION

Discussion

The comparison isolates algorithmic effects by holding the decision space, constraints, and diagnostics plane fixed. In this setting, MOGA (NSGA–II) provides a dense, high-quality Pareto baseline, while scalarized BO reaches Pareto-comparable knee and corner solutions with roughly two orders of magnitude fewer evaluations. The BO efficiency stems from (i) informative priors via GP surrogates; (ii) batched acquisitions (qEI/qNEI) that target regions of high expected improvement; and (iii) normalized scalarization that stabilizes learning across objectives of different scales.

The physics trends are consistent with injector intuition. σ_δ-heavy optima exploit *off-crest* phases in the later cavities

ISBN: 978-3-95450-261-5 ISSN: 2673-7000 doi: 10.18429/JACoW-NAPAC2025-WECN03

Table 2: Optimal Objectives for $\mathbf{w} = [0.1, 0.1, 0.8]$

Method	$\varepsilon_{n,x}$ [mm-mrad]	$\varepsilon_{n,y}$ [mm-mrad]	[%]
MOGA	3.473	3.297	0.197
BO	3.176	3.281	0.194

to suppress chirp, accepting a modest emittance penalty; ε-lean optima tighten the β-function in one plane, accepting slightly larger σ_δ. Early coupling between the solenoid field B_{sol} and the gun phase ϕ_{gun} remains decisive for emittance compensation. In the jitter study, knee solutions exhibit the highest pass rates and narrowest tails; σ_δ-extremals are most sensitive to RF phase jitter in the gun and late cavities, aligning with the knob patterns above.

Limitations of this study include the use of scalarization (rather than a direct multi-objective acquisition) and constraint handling via filtering and a soft penalty in BO. Both are practical for online use, but future work should adopt qEHVI/qNEHVI with feasibility weighting to allocate samples by expected hypervolume gain, and incorporate robustness directly.

CONCLUSION

We compared a broad-coverage multi-objective genetic algorithm (MOGA, NSGA-II) and a scalarized Bayesian optimization (BO) workflow on the Korea–4GSR injector with identical bounds, constraints, and diagnostics plane. The union of nine BO runs (Gaussian-process surrogate, qEI/qNEI, normalized scalarization) reproduced the MOGA Pareto knee and extremes with comparable quality while using $\mathcal{O}(10^3)$ simulations versus $\mathcal{O}(10^5)$ for MOGA. Representative solutions exhibited the expected physics trends (off-crest late cavities for low σ_δ; plane-selective focusing for low ε_n), and the Monte-Carlo jitter study indicated highest tolerance near the knee.

MOGA provides a reliable Pareto reference; scalarized BO recovers comparable trade-offs at two orders of magnitude fewer simulator calls, making frequent re-optimization feasible for commissioning. The proposed path to true MOBO and robustness-aware acquisition turns this into an operator-ready method: start from a balanced knee, bias toward σ_δ or ε_n as needed, and re-tune rapidly with bounded compute.

ACKNOWLEDGMENTS

The author thanks Dr. Chanmi Kim for her contributions to the MOGA simulations and for many helpful discussions.

REFERENCES

[1] P. F. Tavares, S. C. Leemann, M. Sjöström, and Å. Andersson, "The MAX IV storage ring project," *J. Synchrotron Radiat.*, vol. 21, no. 5, pp. 862–877, 2014. doi:10.1107/S1600577514011503

[2] S. Bettoni, M. Pedrozzi, and S. Reiche, "Low emittance injector design for free electron lasers," *Phys. Rev. Spec. Top. Accel. Beams*, vol. 18, p. 123403, Dec. 2015. doi:10.1103/PhysRevSTAB.18.123403

[3] C. Kim *et al.*, "Optimization study for RF cavity and beam dynamics simulation in the electron linear accelerator using the multi-objective genetic algorithm," *J. Instrum.*, vol. 19, no. 1, p. 01023, Jan. 2024. doi:10.1088/1748-0221/19/01/P01023

[4] G. S. Jang *et al.*, "Low emittance lattice design for Korea-4GSR," *Nucl. Instrum. Methods Phys. Res., Sect. A*, vol. 1034, p. 166779, 2022. doi:10.1016/j.nima.2022.166779

[5] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: NSGA-II," *IEEE Trans. Evol. Comput.*, vol. 6, no. 2, pp. 182–197, 2002. doi:10.1109/4235.996017

[6] D. R. Jones, M. Schonlau, and W. J. Welch, "Efficient global optimization of expensive black-box functions," *J. Global Optim.*, vol. 13, pp. 455–492, 1998. doi:10.1023/A:1008306431147

[7] K. Flöttmann, *ASTRA: A space charge tracking algorithm: User's manual, Version 3.2*, DESY, Zeuthen, Germany, Mar. 2017.