GUINEAGRAD: A Differentiable Beam-Beam Interaction Simulator for Gradient-Based Collider Optimization

Anonymous Authors[1]

Abstract

Making physical simulators differentiable unlocks gradient-based optimization. We present GUINEAGRAD, a JAX reimplementation of the GUINEA-PIG beam-beam interaction code that computes automatic derivatives of luminosity and beamstrahlung observables via reverse-mode automatic differentiation. Validated against GUINEA-PIG to within 2% across three collider configurations, GUINEAGRAD enables gradient-based multi-objective optimization in a 6-dimensional parameter space where Nelder–Mead collapses entirely, producing only degenerate or divergent solutions. Gradient-based optimization becomes wall-clock faster than derivative-free alternatives at ∼6–10 parameters, establishing differentiable simulation as a practical tool for high-dimensional collider design.

1. Introduction

Beam-beam interactions are among the most important phenomena in high-energy lepton colliders. The electromagnetic fields generated by each bunch distort the opposing bunch trajectory, affecting luminosity, beamstrahlung energy spread, and pair-production backgrounds. Accurate simulation of these effects is essential for the design and optimization of future colliders including C^3 (Bai et al., 2021), CLIC (Aicheler et al., 2012), and ILC (Adolphsen et al., 2013).

GUINEA-PIG (Schulte, 1997) has been the standard beam-beam simulation code used in collider design studies for over two decades. Written in C++, it models the strong-strong beam-beam interaction through a particle-in-cell (PIC) algorithm with longitudinal slicing, transverse charge deposition on a grid, FFT-based Poisson field solving, and electromagnetic kicks including quantum beamstrahlung. Although widely used, GUINEA-PIG provides no gradient information: derivatives of observables with respect to beam parameters must be estimated via finite differences, requiring $2N$ perturbed simulation evaluations for N parameters using central differences (plus the baseline evaluation, if not already available).

This limitation has significant practical consequences. Collider design involves optimizing beam parameters to maximize luminosity while constraining beamstrahlung energy loss and pair-production backgrounds, a multi-objective problem in a 20+ dimensional parameter space (beam sizes, emittances, offsets, crossing angles, waist shifts, energy spread, and bunch charge for both beams). Without gradients, optimization relies on derivative-free methods (simplex, genetic algorithms, grid scans) that scale poorly with dimensionality and provide no sensitivity information.

Recent advances in differentiable programming (Baydin et al., 2018) have demonstrated that automatic differentiation (AD) can be applied to complex scientific simulations, providing automatic gradients at a cost comparable to the forward evaluation. Frameworks such as JAX (Bradbury et al., 2018) enable AD through arbitrary Python/NumPy code with GPU acceleration, and have been applied to molecular dynamics (Schoenholz & Cubuk, 2020), fluid simulation (Kochkov et al., 2021), and accelerator lattice tracking (Gonzalez-Aguilera et al., 2023). Differentiating a particle-in-cell beam-beam code is nontrivial: scatter-type charge deposition is not natively differentiable (it requires custom vector-Jacobian products), the FFT-based Poisson solve must propagate adjoints through the full field computation, and the beamstrahlung spectrum involves special functions (Airy/Bessel) (Chen & Yokoya, 1988) whose AD paths must be carefully validated. Pair-production sampling introduces stochastic paths that are not everywhere differentiable.

In this work we present GUINEAGRAD, an implementation of the GUINEA-PIG beam-beam simulation algorithm in JAX. Our contributions are:

1. A differentiable beam-beam simulator that computes automatic derivatives of the deterministic expected-

[1]Anonymous Institution, Anonymous City, Anonymous Region, Anonymous Country. Correspondence to: Anonymous Author <anon.email@domain.com>.

Preliminary work. Under review by the AI4Physics Workshop at the International Conference on Machine Learning (AI4Physics@ICML 2026). Do not distribute.

value program for primary observables with respect to beam parameters via reverse-mode AD.

2. Comprehensive validation against GUINEA-PIG across three reference machine configurations, demonstrating agreement within 2% for primary observables, consistent with GUINEA-PIG run-to-run statistical fluctuation.

3. Demonstration of gradient-enabled applications: luminosity recovery optimization (wall-clock crossover at $N \approx 6$–10 over derivative-free methods), and 6-dimensional multi-objective Pareto analysis where derivative-free baselines fail entirely.

GUINEAGRAD does not yet reproduce all GUINEA-PIG outputs at full fidelity: pair production is preliminary, and low-resolution behavior differs from the high-fidelity regime. The focus of this work is on demonstrating the feasibility and utility of automatic differentiation through the beam-beam simulation. We note that finite-difference gradients can be trivially parallelized across independent perturbations; the advantage of AD is strongest when gradients are needed repeatedly inside iterative workflows or when finite-difference truncation noise is limiting.

2. Related Work

Beam-beam interaction simulation. Beam-beam effects in high-energy lepton colliders are traditionally modeled with specialized forward simulators such as GUINEA-PIG (Schulte, 1997), which has been used extensively in linear-collider design studies for TESLA/ILC, CLIC, and more recent compact collider concepts. These tools simulate the strong–strong interaction using macroparticle bunches, longitudinal slicing, transverse charge deposition, field solving, electromagnetic kicks, and beamstrahlung/pair-production models. Their strength is high-fidelity forward prediction of luminosity, beamstrahlung spectra, and detector-background rates. However, they are not designed to expose derivatives of observables with respect to beam or optics parameters. As a result, collider-design studies typically rely on parameter scans, finite differences, or derivative-free optimizers, which become expensive and noisy as the number of tunable parameters grows.

Differentiable accelerator modeling. Automatic differentiation has recently entered accelerator physics through differentiable lattice and beamline models. BmadX (Gonzalez-Aguilera et al., 2023) differentiates single-particle tracking and optics quantities, while Cheetah (Kaiser et al., 2024) provides fast differentiable beamline simulations for machine-learning and control applications. These approaches are complementary to GUINEAGRAD: they focus primarily on transport through accelerator elements, whereas beam-beam simulation requires differentiating through collective strong–strong interactions between two intense bunches. In particular, the beam-beam PIC algorithm introduces scatter/gather charge deposition, FFT-based field solves, slice-by-slice kicks, beamstrahlung special functions, and stochastic pair-production paths. These ingredients make the differentiability problem qualitatively different from single-particle lattice tracking.

Differentiable physical simulation. More broadly, differentiable simulation has enabled gradient-based optimization and inverse problems in molecular dynamics, fluids, rendering, and mechanics. Examples include JAX-MD (Schoenholz & Cubuk, 2020), DiffTaichi (Hu et al., 2020), Warp (Macklin, 2022), Mitsuba 3 (Jakob et al., 2022), and differentiable differential-equation solvers such as diffrax (Kidger, 2022). GUINEAGRAD extends this paradigm to beam-beam interactions in future collider design and is, to our knowledge, the first differentiable beam-beam interaction simulator. Its main novelty is not simply reimplementing a forward simulator in JAX, but exposing validated derivatives of luminosity and beamstrahlung observables through a PIC-based collective interaction model, enabling repeated gradient-based optimization and sensitivity analysis in regimes where finite differences are costly and derivative-free methods struggle.

3. Beam-beam scalings and differentiability target

Proposed e^+e^- Higgs factories and multi-TeV energy-frontier machines (ILC, CLIC, C^3, CEPC, FCC-ee) all require luminosities of order 10^{34}–10^{36} cm^{-2}s^{-1} to achieve their physics programmes. Luminosity optimization is therefore a central task in their design and operational tuning. We focus on *linear* collider options, where the single-pass nature of the beams produces large disruption parameters ($D_y \gg 1$). In this regime the electromagnetic pinch during the collision substantially enhances the luminosity beyond the geometric overlap, and the actual luminosity cannot be computed from a simple algebraic formula; instead, it must be obtained from a full beam-beam simulation.

Before describing the implementation, we review the analytic scalings that govern beam-beam observables in the low- and high-disruption limits. These establish the "target" elasticities against which the AD Jacobian can be cross-checked.

Geometric luminosity. For head-on collision of Gaussian bunches, the geometric luminosity is

$$\mathcal{L}_{\text{geom}} = \frac{N^2 f_{\text{rep}} n_b}{4\pi\sigma_x\sigma_y}, \tag{1}$$

where N is the bunch population, $\sigma_{x,y}$ the transverse RMS beam sizes at the interaction point, f_{rep} the repetition frequency, and n_b the number of bunches per train. Since each GUINEA-PIG/GUINEAGRAD simulation models a single bunch crossing, the per-crossing luminosity is obtained with $n_b = f_{\text{rep}} = 1$; these machine-level parameters do not enter the beam-beam dynamics and are not further discussed. In the absence of electromagnetic disruption, the elasticities are $\partial \ln \mathcal{L}_{\text{geom}}/\partial \ln N = 2$, $\partial \ln \mathcal{L}_{\text{geom}}/\partial \ln \sigma_x = -1$, and $\partial \ln \mathcal{L}_{\text{geom}}/\partial \ln \sigma_y = -1$.

Disruption enhancement. The electromagnetic interaction pinches the opposing bunch, increasing the effective overlap and luminosity by a factor H_D (Chen & Yokoya, 1988; Napoly, 1993; Ntounis et al., 2024). The disruption parameters

$$D_{x,y} = \frac{2Nr_e\sigma_z}{\gamma\,\sigma_{x,y}(\sigma_x + \sigma_y)} \tag{2}$$

characterize the strength of this pinch. We focus this study on flat beams ($\sigma_x \gg \sigma_y$), for which the vertical disruption D_y is large ($D_y \sim 10$–25) while $D_x \ll 1$. In this *disrupted regime* ($D_y \gg 1$), $H_D > 1$ and the effective luminosity exceeds $\mathcal{L}_{\text{geom}}$. Consequently the actual $|\partial \ln L/\partial \ln \sigma_x|$ is smaller than unity (the pinch partially compensates a wider beam), and the actual $\partial \ln L/\partial \ln N$ can deviate from the geometric value of 2 because the disruption enhancement H_D itself depends on N; in the configurations below the AD elasticities lie near 2 (1.9–2.1), consistent with the expected disrupted-beam behavior (Appendix D).

Beamstrahlung. The mean beamstrahlung parameter for a particle in the opposing bunch field is (Schulte, 1997)

$$\Upsilon \propto \frac{N\gamma}{\sigma_z(\sigma_x + \sigma_y)}. \tag{3}$$

For flat beams and up to order-unity geometry factors, the average fractional energy loss scales as $\delta_{\text{BS}} \propto \alpha r_e^2 N^2 \gamma/[\sigma_z^2(\sigma_x + \sigma_y)^2]$, and the mean number of emitted photons as $n_\gamma \propto \alpha\sigma_z\Upsilon/[(\sigma_x + \sigma_y)\sqrt{1 + \Upsilon^{2/3}}]$. These scalings predict strong sensitivity to N and $\sigma_x + \sigma_y$, with the expected signs matching the AD Jacobian.

Pair production. Coherent (e^+e^-) and incoherent pair production rates depend on the beamstrahlung photon spectrum and the local field strength through the differential cross sections for the Breit–Wheeler, Bethe–Heitler, and Landau–Lifshitz processes (Chen & Yokoya, 1988; Schulte, 1997). Closed-form elasticities are not available because the rates involve integrals over the photon spectrum that depend non-trivially on beam parameters. Qualitatively, pair rates increase with Υ and thus share the sensitivity pattern of beamstrahlung observables, but with stronger nonlinearity.

4. Method

GUINEAGRAD reimplements the deterministic core of the GUINEA-PIG algorithm (Schulte, 1997) in JAX (Bradbury et al., 2018), a Python framework for composable transformations including `jax.grad` (reverse-mode AD) and `jax.jit` (XLA compilation for CPU/GPU).

Algorithm. Two opposing bunches, each represented by N_{macro} macroparticles, are divided into N_{slice} longitudinal slices. For each colliding slice pair, the simulation (1) deposits charge onto a 2D transverse grid via cloud-in-cell (CIC) interpolation, (2) solves the Poisson equation via FFT, (3) applies electromagnetic kicks, (4) computes beamstrahlung energy loss and photon emission from the Sokolov–Ternov spectrum, and (5) evaluates pair-production cross sections. Luminosity $\mathcal{L}$ is accumulated from the overlap integral of colliding slice pairs.

Differentiable design. All state is stored in immutable `NamedTuple` pytrees. Macroparticle coordinates are generated via reparameterization ($x_i = \sigma_x \cdot z_i$ with fixed standard-normal samples z_i), ensuring smooth derivatives $\partial x_i/\partial\sigma_x = z_i$. The CIC deposit uses a custom VJP computing adjoint contributions to positions and charges in a single gather pass, validated against FD to $\lesssim 10^{-10}$. We use `jax.checkpoint` on the per-slice-pair function, reducing peak memory from $O(N_{\text{slice}}^2)$ to $O(N_{\text{slice}})$. The beamstrahlung spectrum uses the Airy function representation, differentiable via XLA primitives. All computations use float64 (`jax_enable_x64=True`).

Non-differentiable paths. Longitudinal slice assignment uses `floor`, making $\partial/\partial\sigma_z = 0$ by construction. The LL pair spectrum uses accept/reject sampling that is not differentiable. Coherent pairs are forward-only (see Section 8).

AD cost bound. Reverse-mode AD computes $\nabla_\theta f(\theta)$ for all N input parameters in a single backward pass at cost $\leq C \cdot \text{cost}(f)$ (Griewank & Walther, 2008), where $C \approx 4.5$ for this code (Section 6). The full Jacobian $J_{ij} = \partial f_i/\partial\theta_j$ is obtained via `jax.jacrev` with one reverse pass per observable.

5. Validation

We validate GUINEAGRAD in two stages: (1) parity of forward observables against GUINEA-PIG, and (2) correctness of AD gradients against finite-difference reference values.

5.1. Reference configurations

We use three reference configurations spanning the parameter space of proposed future lepton colliders:

- **C3-250:** Cool Copper Collider at $\sqrt{s} = 250$ GeV, with flat beams ($\sigma_x \approx 930$ nm, $\sigma_y \approx 5$ nm);
- **CLIC-380:** Compact Linear Collider at $\sqrt{s} = 380$ GeV, with very flat beams ($\sigma_x \approx 143$ nm, $\sigma_y \approx 3$ nm);
- **ILC-500:** International Linear Collider at $\sqrt{s} = 500$ GeV, with intermediate aspect ratio ($\sigma_x \approx 474$ nm, $\sigma_y \approx 6$ nm).

All simulations use $N_{\mathrm{macro}} = 10{,}000$ macroparticles per bunch, $N_{\mathrm{slice}} = 25$ longitudinal slices, and a 32×32 transverse grid, unless otherwise noted. Table 3 (Appendix) lists the full beam parameters for each configuration.

5.2. Parity with Guinea-Pig

Table 1 compares GUINEAGRAD and GUINEA-PIG observables at matched resolution ($N_{\mathrm{macro}} = 100{,}000$, 512×512 grid). The GG/GP ratio is shown alongside the GP relative standard deviation (RSD) from a 10-seed Monte Carlo ensemble, and the pull $\Delta/\sigma = (\text{ratio} - 1)/\text{RSD}$, which measures the discrepancy in units of GP statistical noise.

For the primary observables (luminosity $\mathcal{L}$, peak luminosity $\mathcal{L}_{\mathrm{pp}}$, beamstrahlung energy loss δ_{BS}, and photon number n_γ), all pulls are $\leq 2.3\sigma$, confirming that GG–GP differences are consistent with the GP statistical uncertainty. The pair-production observables show statistically significant discrepancies: n_{LL} (Landau–Lifshitz) is systematically 14–20% high (11–64σ), n_{BH} (Bethe–Heitler) ranges from -5% to $+20\%$ (3–12σ), and n_{BW} (Breit–Wheeler) shows order-of-magnitude discrepancies (8–91σ). These reflect known differences in the treatment of virtual photon Compton displacement kinematics and low-energy pair thresholds, discussed in Section 8.

Figure 1 provides a visual summary of the parity comparison, with GP multi-seed RSD shown as error bars for context.

GG/GP ratios improve monotonically with grid size, from 0.56–0.87 at 32×32 to 0.99–1.01 at 256×256. The gradient validation and application studies use 32×32 grids with $N_{\mathrm{macro}} = 10{,}000$ for computational feasibility; the AD gradients are correct derivatives of the deterministic program at whatever resolution is chosen.

5.3. Gradient correctness

We validate the AD gradients by comparing against central finite differences at multiple step sizes. For each parameter θ_j and observable f_i, the central FD estimate is:

$$\left.\frac{\partial f_i}{\partial \theta_j}\right|_{\mathrm{FD}} = \frac{f_i(\theta_j + h) - f_i(\theta_j - h)}{2h} + O(h^2). \quad (4)$$

The relative error between AD and FD,

$$\epsilon_{\mathrm{rel}} = \frac{|\partial f/\partial\theta_{\mathrm{AD}} - \partial f/\partial\theta_{\mathrm{FD}}|}{|\partial f/\partial\theta_{\mathrm{AD}}|}, \quad (5)$$

decreases as $O(h^2)$ when h is reduced from large values (truncation regime), plateaus at the optimal h where truncation and cancellation errors balance, and increases again for very small h due to floating-point cancellation in the FD numerator.

Figure 2 shows this convergence behavior for 4 representative parameters on ILC-500. The FD step h is a fractional perturbation of each parameter's nominal value. The AD values sit on the plateau of the FD convergence curve, confirming that the AD gradients are correct to within the floating-point precision of the FD reference. Several derivatives show the expected $O(h^2)$ trend in the large-step truncation regime; deviations reflect cancellation, finite macroparticle noise, and non-smooth CIC/slicing effects. For the core physical parameters (N, E, σ_x, σ_y, ε_x, ε_y, σ_E/E), best relative errors are typically below 10^{-4}, with a median of $\sim 3 \times 10^{-7}$ across all three configurations.

The AD Jacobian reveals the expected physical structure: N dominates luminosity ($E \approx 1.9$–2.1, consistent with $\mathcal{L} \propto N^2$), σ_x has large negative elasticity (-0.7 to -0.9), and σ_y shows a sign reversal for C3-250 ($+0.45$) due to strong disruption overshoot. The full Jacobian heatmap, analytic cross-checks, and detailed elasticity analysis appear in Appendix D.

Hessian diagonals $\partial^2\mathcal{L}/\partial\theta_j^2$ agree with FD-of-gradient to median 26% relative error. A grid-resolution convergence study (Appendix E) shows that bunch population N elasticities are stable to $< 17\%$ from $N_{xy} = 32$ to 512, while transverse-size elasticities shift by $\sim 118\%$, reflecting the resolution dependence of the beam-overlap integral. Because the gradients are correct automatic derivatives at each resolution, the application results in Section 7 should be interpreted as differentiable-workflow demonstrations at the chosen resolution, not as final high-fidelity design optimizations.

6. Performance

6.1. Hardware and software

All GPU benchmarks use JAX 0.9 with XLA compilation on NVIDIA A100-SXM4-40GB. GUINEA-PIG version 1.2.0 runs single-threaded on an AMD EPYC 7713 CPU. All times are median of 5 runs after 3 warmup runs (JIT compilation excluded). JIT compilation is a one-time cost per configuration: $\sim$30–50 s for the forward pass and $\sim$40 s for the gradient at $N_{\mathrm{macro}} = 10{,}000$ on A100; this cost is amortized over subsequent calls with the same static arguments (N_{macro}, grid size).

Table 1. Parity of GUINEAGRAD and GUINEA-PIG: ratio of observables (GG/GP) at matched resolution ($N_{\mathrm{macro}} = 100{,}000$, 512×512 grid). Parenthetical values are the GP 10-seed RSD; bracketed values are the pull Δ/σ. Primary observables (above the line) show pulls $\leq 2.3\sigma$; pair-production rates (below) are systematically discrepant (see Section 8).

Observable	C3-250	CLIC-380	ILC-500
$\mathcal{L}$	1.0005 (0.21%) [$+0.3\sigma$]	1.0078 (0.34%) [$+2.3\sigma$]	1.0161 (1.81%) [$+0.9\sigma$]
$\mathcal{L}_{\mathrm{pp}}$	1.0005 (0.22%) [$+0.2\sigma$]	1.0058 (0.35%) [$+1.7\sigma$]	1.0092 (1.78%) [$+0.5\sigma$]
δ_{BS}	0.9691 (1.65%) [-1.9σ]	0.9997 (0.42%) [-0.1σ]	1.0028 (1.11%) [$+0.3\sigma$]
n_γ	0.9923 (0.87%) [-0.9σ]	1.0008 (0.24%) [$+0.3\sigma$]	1.0012 (0.53%) [$+0.2\sigma$]
n_{BH}	1.0390 (1.18%) [$+3.3\sigma$]	0.9502 (0.44%) [-11σ]	1.1969 (1.68%) [$+12\sigma$]
n_{LL}	1.1432 (0.23%) [$+64\sigma$]	1.2016 (0.35%) [$+58\sigma$]	1.1796 (1.65%) [$+11\sigma$]
n_{BW}	0.4817 (3.13%) [-17σ]	2.1344 (1.25%) [$+91\sigma$]	0.8318 (2.03%) [-8σ]

6.2. Forward pass and gradient timing

Table 2 reports the luminosity-subgraph forward pass (the computation traced by `jax.grad`) and the value-and-gradient call.

The GPU gradient overhead is a stable $\sim 4.5\times$ across configurations and N_{macro} up to $50{,}000$. At higher macroparticle counts the GPU forward pass increasingly outperforms GUINEA-PIG on CPU, amplifying the practical advantage of AD gradients (Appendix F). For the full 9×3 Jacobian, AD provides a $1.4\times$ speedup over FD (1.4–$2.0\times$ vs. GP FD on CPU), matching the theoretical bound $(2N{+}1)/(M \cdot r) \approx 1.4$ (detailed timing in Appendix F).

7. Applications

Two applications illustrate the practical benefit of automatic gradients.

7.1. Luminosity recovery: from steering to joint retuning

A physically relevant use case for gradient-based optimization is luminosity recovery after beam mis-steering, precisely the task performed by beam-beam feedback systems during collider commissioning. We introduce deliberate misalignment of one beam ($3\sigma_x$ horizontal offset, $2\sigma_y$ vertical offset) relative to the opposing beam, and optimize four steering parameters (δ_x, δ_y, w_x, w_y) to recover nominal luminosity, with beam sizes *fixed* at their nominal values. The waist positions w_x, w_y are application-specific knobs not included in the 9-parameter sensitivity analysis of Table 3; they parameterize the longitudinal location of the beam focus relative to the IP. This avoids the unphysical regime of unconstrained beam shrinkage. Both optimizers operate without explicit parameter bounds; beam sizes use log-ratio parameterization ensuring positivity. We compare Adam (Kingma & Ba, 2015) (gradient-based, using GUINEAGRAD gradients) against Nelder–Mead (Nelder & Mead, 1965) (derivative-free simplex, forward-only evaluations).

We progressively enlarge the physically available retuning space from $N = 4$ (steering only) to $N = 12$ (adding beam sizes, emittances, energy, and crossing angles). Figure 3(a) shows the convergence traces at $N = 12$ for all three configurations: Adam reaches $>99\%$ recovery in ~ 300 s wall-clock, roughly half the time Nelder–Mead requires, while NM takes ~ 700 s or more. The objective minimized is $(\mathcal{L}/\mathcal{L}_{\mathrm{nom}} - 1)^2$; at large N Adam can exceed nominal luminosity ($\mathcal{L}/\mathcal{L}_0 > 1$) because the enlarged parameter space contains configurations that improve upon the design point. At $N = 4$ (steering only), both optimizers eventually converge, but NM is $\sim 9\times$ faster in wall-clock: each forward-only evaluation costs much less, and the 4-dimensional landscape is simple enough for direct search (Appendix I). Figure 3(b) demonstrates the crossover: as N grows, Adam's convergence time *decreases* (the richer gradient signal enables faster convergence at fixed per-step cost), while NM's time *increases* as the simplex grows.[1] The wall-clock crossover occurs at $N{\approx}6$–10, with Adam $\sim 2\times$ faster at $N{=}12$ across all three machines. This illustrates the core advantage of differentiable simulation: the $\mathcal{O}(1)$ gradient cost of reverse-mode AD amortizes for moderate-dimensional problems.

7.2. Multi-objective Pareto analysis ($N = 6$)

Beyond single-objective recovery, collider design requires balancing competing objectives: maximizing luminosity while minimizing beamstrahlung energy spread. We trace the Pareto front via weighted-sum scalarization (Eq. 6 in Appendix G), optimizing six beam parameters simultaneously (σ_x, σ_y, ε_x, ε_y, E, N) in log-ratio space, i.e., $\theta_i = \log(\mathrm{param}_i/\mathrm{nominal}_i)$.

Figure 4 shows the result for C^3-250. Adam traces a smooth Pareto front spanning the full luminosity–beamstrahlung tradeoff (41 points, 30 gradient steps each). Nelder–Mead on the same 6-dimensional problem fails catastrophically:

[1] We also tested SciPy's `differential_evolution` (DE) on C^3-250 at $N{=}10$ with a budget of 2000 function evaluations: DE reached only $\sim 8\%$ of nominal luminosity, confirming that the difficulty is not specific to the simplex algorithm.

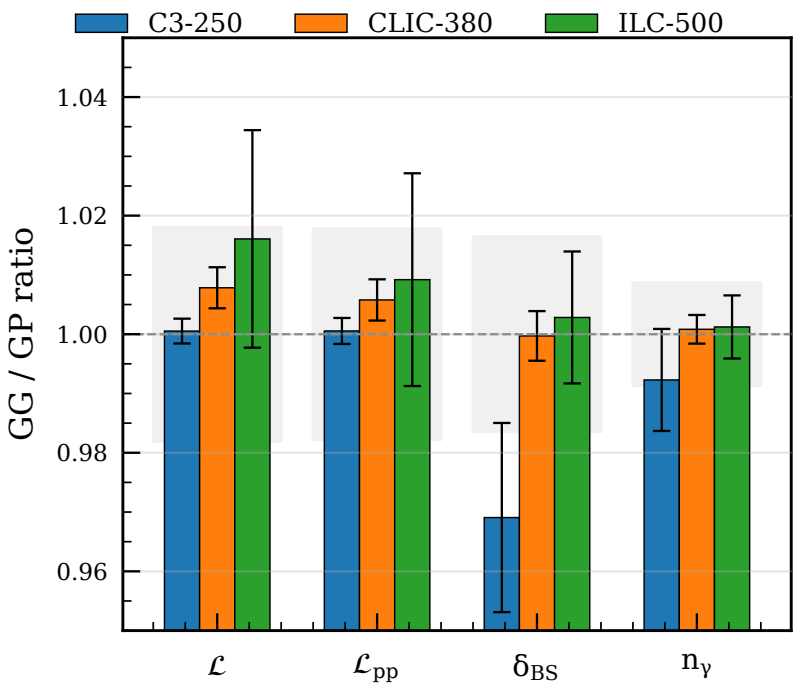

Figure 1. Parity of GUINEAGRAD and GUINEA-PIG: GG/GP observable ratios at nominal operating points for all three configurations ($N_{\text{macro}} = 100$k, $N_{xy} = 512$, exact mode). (a) Primary observables agree within 2%; grey bands show the maximum GP 10-seed relative standard deviation (RSD) envelope. (b) Pair-production rates show larger discrepancies from virtual photon kinematics differences (see §8). Error bars on all panels denote GP 10-seed RSD.

Table 2. Wall-clock times (seconds) at $N_{\text{macro}} = 10{,}000$, 64×64 grid (median of 5 runs, A100 GPU). GG rows measure the luminosity subgraph traced by `jax.grad`.

Method	C3-250	CLIC-380	ILC-500
GP forward (CPU, C++)	6.3	8.1	8.6
GG forward (GPU)	6.3	5.8	5.8
GG val+grad for $\mathcal{L}$ (GPU)	28.6	26.2	26.4
Gradient overhead (GPU)	4.5×	4.5×	4.5×

of 21 trials, 13 collapse to degenerate solutions ($\mathcal{L} \to 0$, beams completely defocused) and 8 diverge ($\mathcal{L} \to \infty$, unphysical beam compression). Zero NM trials produce a physically meaningful tradeoff point. The failure mechanism is geometric: the NM simplex ($N + 1 = 7$ vertices in 6D) cannot resolve the narrow manifold of Pareto-optimal solutions. This is consistent with known convergence failures of Nelder–Mead in moderate dimensions (Lagarias et al., 1998; McKinnon, 1998): the simplex can stagnate at non-stationary points or collapse to degenerate configurations. More broadly, derivative-free methods generally require $\mathcal{O}(N^2)$ evaluations per iteration and lose effectiveness rapidly above $N \approx 10$ (Rios & Sahinidis, 2013). Without gradient information to navigate the Pareto ridge, the simplex either collapses toward the trivial minimum or reflects into unbounded directions.

The original 2-parameter Pareto front ($\log \sigma_x, \log \sigma_y$) demonstrating the Pareto-elbow structure appears in Appendix G.

8. Discussion

Observable classification. We classify the simulation outputs into two tiers based on validation status. *Primary observables* (luminosity $\mathcal{L}$, $\mathcal{L}_{\text{pp}}$, mean beamstrahlung energy loss δ_{BS}, and photon number n_γ) show pulls $|\Delta/\sigma_{\text{GP}}| \leq 2.3$ across all three reference configurations (Table 1), establishing sub-2% agreement consistent with GUINEA-PIG's intrinsic statistical uncertainty. These are the observables for which gradient-based analysis is validated. *Secondary observables* (Landau–Lifshitz n_{LL}, Bethe–Heitler n_{BH}, and Breit–Wheeler n_{BW} pair-production rates) show statistically significant discrepancies (pulls $> 3\sigma$) and are reported as

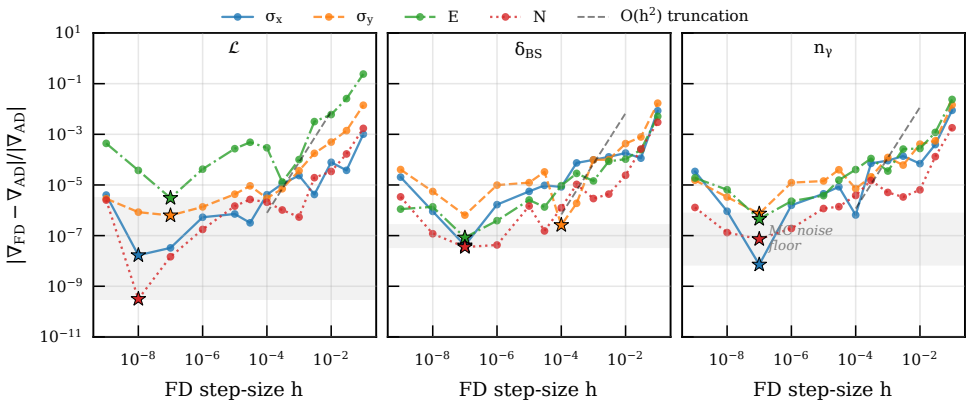

Figure 2. AD vs FD gradient convergence for 4 representative parameters (ILC-500, $N_{\text{macro}} = 10{,}000$, 32×32 grid). Each line traces relative error vs. FD step h; stars mark optimal h. Dashed line: expected $O(h^2)$ truncation slope. AD matches the FD plateau.

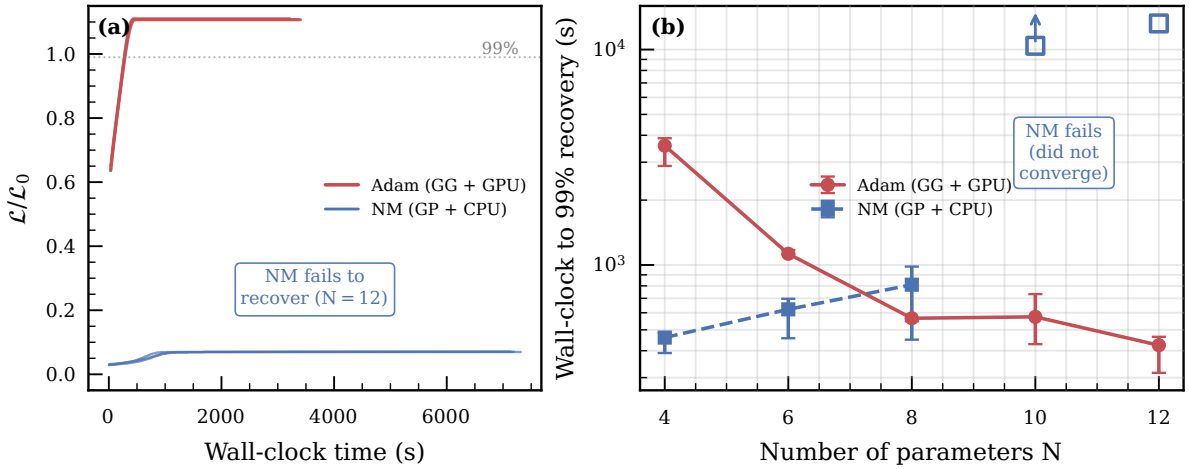

Figure 3. Gradient advantage for luminosity recovery. (a) Convergence at $N = 12$: $\mathcal{L}/\mathcal{L}_{\text{nom}}$ vs. wall-clock time for Adam (solid) and Nelder–Mead (dashed). Adam reaches 99% in $\sim$300 s; NM requires $\sim$700 s. (b) Time to 99% recovery vs. N. Adam improves with N; NM degrades. Crossover at $N \approx 6$–10. Open markers: non-convergence within budget. The physical problem changes with N: larger parameter sets provide additional recovery mechanisms.

preliminary.

Pair production discrepancies. The incoherent pair-production discrepancies (LL 14–20% high, BH -5% to $+20\%$, BW order-of-magnitude) are structural consequences of differences between GUINEAGRAD's grid-based deposit approach and GUINEA-PIG's stochastic per-particle pairing. A systematic study of virtual-photon displacement modes shows no single mode reproduces GP pair rates across all channels simultaneously. Importantly, pair rates are computed *after* the luminosity and δ_{BS} calculations and do not affect primary observables. Detailed GG/GP ratios and root-cause analysis appear in Appendix H.

Forward pass overhead. On GPU, GUINEAGRAD's forward pass (5.8–6.3 s) is comparable to GUINEA-PIG's CPU forward pass (6.3–8.6 s) at $N_{\text{macro}} = 10{,}000$ (Table 2). On CPU, GUINEAGRAD is 3–4$\times$ slower due to JAX's functional model; the advantage emerges in gradient-dependent workflows where the $O(1)$ gradient cost amortizes the per-evaluation overhead.

Limitations. Key limitations include: GPU memory restricting $N_{\text{macro}} \lesssim 50$k at 64×64; $\partial/\partial\sigma_z = 0$ from discrete slicing; stochastic emission producing noisy pair-rate gradients; and preliminary pair-production rates. Because the application studies use reduced resolution ($N_{xy} = 32$–64) for computational feasibility, their purpose is to validate differentiable optimization workflows rather than to propose final optimized beam parameters.

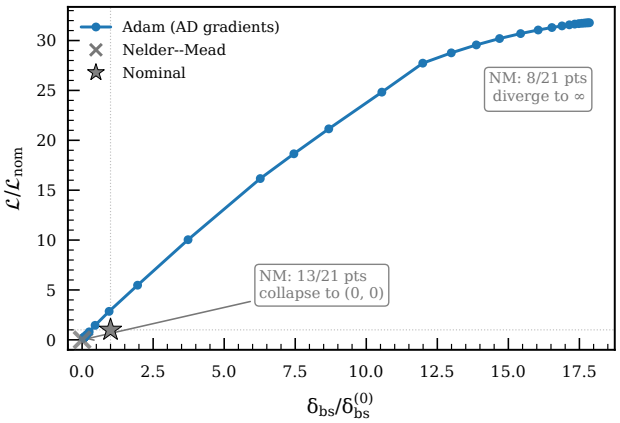

Figure 4. Multi-objective optimization in 6D beam-parameter space (C^3-250). Adam (blue, using AD gradients) traces a smooth Pareto front; Nelder–Mead (red crosses) produces only degenerate ($\mathcal{L} = 0$) or divergent solutions. The nominal operating point is marked with a star. NM failure counts: 13/21 collapse to the origin, 8/21 diverge to ∞.

9. Conclusion

We have presented GUINEAGRAD, a differentiable beam-beam interaction simulator that computes automatic derivatives. Validated against GUINEA-PIG across three reference collider configurations, GUINEAGRAD achieves percent-level agreement for primary observables. Two application results demonstrate the practical benefit of automatic derivatives: (1) gradient-based luminosity recovery crosses over a Nelder–Mead baseline at $N \approx 6$–10 parameters; and (2) 6-dimensional multi-objective Pareto optimization, where Nelder–Mead fails catastrophically while Adam produces smooth tradeoff curves. These results show that differentiable simulation makes moderate-dimensional beam-beam optimization substantially more robust than simplex-based alternatives. Future work will address continuous slicing, GUINEA-PIG-compatible pair production, and coupling to lattice tracking codes for end-to-end collider optimization.

Reproducibility

GUINEAGRAD will be released as open-source under the MIT license upon acceptance. All simulations use deterministic JAX PRNG seeds and Float64 precision.

Impact Statement

This work advances differentiable simulation tools for particle accelerator design. By enabling gradient-based optimization of beam-beam interaction parameters, it may accelerate the design cycle for future lepton colliders. We foresee no negative societal consequences from this research.

References

D. Schulte, *Study of Electromagnetic and Hadronic Background in the Interaction Region of the TESLA Collider*, PhD thesis, University of Hamburg, TESLA-97-08 (1997).

M. Bai *et al.*, "C^3: A 'Cool' Route to the Higgs Boson and Beyond," arXiv:2110.15800 (2021).

M. Aicheler *et al.*, "A Multi-TeV Linear Collider Based on CLIC Technology: CLIC Conceptual Design Report," CERN-2012-007 (2012).

C. Adolphsen *et al.*, "The International Linear Collider Technical Design Report – Volume 3.II: Accelerator Baseline Design," arXiv:1306.6328 (2013).

A. G. Baydin *et al.*, "Automatic differentiation in machine learning: a survey," *J. Mach. Learn. Res.* **18**(153), 1–43 (2018).

J. Bradbury *et al.*, "JAX: composable transformations of Python+NumPy programs," https://github.com/google/jax (2018).

A. Griewank and A. Walther, *Evaluating Derivatives: Principles and Techniques of Algorithmic Differentiation*, SIAM, 2nd edition (2008).

S. S. Schoenholz and E. D. Cubuk, "JAX MD: A Framework for Differentiable Physics," *Adv. Neural Inf. Process. Syst.* **33**, 11428–11441 (2020).

D. Kochkov *et al.*, "Machine learning–accelerated computational fluid dynamics," *PNAS* **118**, e2101784118 (2021).

J.-P. Gonzalez-Aguilera *et al.*, "Towards fully differentiable accelerator modeling," *Proc. IPAC'23*, 2797–2800 (2023), doi:10.18429/JACoW-IPAC2023-WEPA065.

D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *ICLR* (2015).

J. A. Nelder and R. Mead, "A simplex method for function minimization," *Comput. J.* **7**, 308–313 (1965).

P. Chen and K. Yokoya, "Disruption effects from the interaction of round e^+e^- beams," *Phys. Rev. D* **38**, 987–1000 (1988).

O. Napoly, "The luminosity for beam distributions with error and wake field effects in linear colliders," *Part. Accel.* **40**, 181–203 (1993).

D. Ntounis, E. A. Nanni and C. Vernieri, "Luminosity and beam-induced background studies for the Cool Copper Collider," *Phys. Rev. Accel. Beams* **27**, 061001 (2024).

Y. Hu *et al.*, "DiffTaichi: Differentiable programming for physical simulation," *ICLR* (2020).

M. Macklin, "Warp: A High-performance Python Framework for GPU Simulation and Graphics," NVIDIA GPU Technology Conference (2022).

W. Jakob *et al.*, "Mitsuba 3: A Retargetable Forward and Inverse Renderer," *ACM Trans. Graphics* **41**(6), 1–17 (2022).

P. Kidger, "On neural differential equations," PhD thesis, University of Oxford (2022).

J. Kaiser *et al.*, "Cheetah: Bridging the gap between machine learning and particle accelerator physics with high-speed differentiable simulations," *Phys. Rev. Accel. Beams* **27**, 054601 (2024).

J. C. Lagarias *et al.*, "Convergence properties of the Nelder–Mead simplex method in low dimensions," *SIAM J. Optim.* **9**(1), 112–147 (1998).

K. I. M. McKinnon, "Convergence of the Nelder–Mead simplex method to a nonstationary point," *SIAM J. Optim.* **9**(1), 148–158 (1998).

L. M. Rios and N. V. Sahinidis, "Derivative-free optimization: a review of algorithms and comparison of software implementations," *J. Global Optim.* **56**(3), 1247–1293 (2013).

A. Reference configurations

Table 3 lists the beam parameters for the three collider configurations used throughout this work.

B. Supplementary figures

Additional validation figures complement the main-text results. Figure 5 shows the full AD vs. FD relative error for all parameter–observable–configuration combinations.

C. Method details

CIC charge deposition uses a custom VJP that computes adjoint contributions to both particle positions and charges in a single gather pass (validated to $\lesssim 10^{-10}$). `jax.checkpoint` on the per-slice-pair function reduces peak memory from $O(N_{\text{slice}}^2)$ to $O(N_{\text{slice}})$. Grid extents ($\pm 5\sigma_{x,y}$) are treated as static constants during differentiation. Non-differentiable paths include: longitudinal slicing ($\partial/\partial\sigma_z = 0$), LL accept/reject sampling, stochastic photon emission in coarse mode, and coherent pair production (forward-only via Erber formula).

D. Jacobian analysis

Figure 6 shows the normalized Jacobian (elasticity) $E_{ij} = (\theta_j/f_i) \cdot \partial f_i/\partial\theta_j$ for the 10 sigma-mode beam parameters and 3 primary observables.

The Jacobian reveals the physical structure of beam-beam interactions: bunch population N has the strongest effect on luminosity ($E \approx 1.9$–2.1, reflecting $\mathcal{L} \propto N^2$). Horizontal beam size σ_x has large negative elasticity (-0.7 to -0.9), consistent with $\mathcal{L} \propto 1/\sigma_x$. The vertical size σ_y shows a sign reversal for C3-250 ($+0.45$) due to disruption overshoot, while CLIC-380 and ILC-500 have the expected negative sign.

Table 4 compares AD elasticities with analytic (undisrupted) predictions from Section 3.

Table 5 explains the zero or near-zero entries in the Jacobian heatmap.

E. Gradient resolution convergence

The gradient validation confirms that AD gradients are correct derivatives of the simulation at whatever resolution is chosen. A separate question is whether the elasticity values are stable across grid resolutions.

We address this with a convergence study at five grid sizes, $N_{xy} \in \{32, 64, 128, 256, 512\}$, computing 7 elasticities $\times$ 3 observables $\times$ 3 configurations with 5 seeds per point (Figure 7).

Bunch population N is stable to $< 17\%$ from $N_{xy} = 32$ to 512. Transverse sizes and emittances show median relative changes of $\sim 118\%$, physically expected because the beam-overlap integral depends on how well the transverse charge distribution is resolved. The gradients at each resolution are correct automatic derivatives of the deterministic program; the resolution dependence is a property of the physics.

F. Wall-clock comparison

Figure 8 compares wall-clock times across methods and hardware at two macroparticle counts. Table 6 reports the full Jacobian computation time via AD vs. finite differences.

G. Multi-objective Pareto front

Collider design requires balancing competing objectives: maximizing luminosity while minimizing beamstrahlung-induced energy spread δ_{BS}. We trace the Pareto front using weighted-sum scalarization with gradient descent.

For each weight $\lambda \in [0, 1]$, we minimize:

$$\mathcal{J}(\theta) = -\lambda \cdot \hat{\mathcal{L}}(\theta) + (1 - \lambda) \cdot \hat{\delta}_{\text{BS}}(\theta), \tag{6}$$

Table 3. Beam parameters for the three reference configurations. Values are per bunch. Units: E in GeV, N in 10^{10}, $\sigma_{x,y}$ in nm, σ_z in μm, $\varepsilon_{x,y}$ in μm·rad (normalized emittance). The AD column marks the 9 parameters with nonzero gradients.

Parameter	Symbol	C3-250	CLIC-380	ILC-500	AD
Energy	E	125	190	250	✓
Bunch pop.	N	0.24	0.52	2.0	✓
Hor. size	σ_x	930	143	474	✓
Vert. size	σ_y	4.9	2.8	5.9	✓
Bunch length	σ_z	600	70	300	†
Hor. emitt.	ε_x	9.96	0.95	10.0	✓
Vert. emitt.	ε_y	0.061	0.030	0.035	✓
Hor. offset	δ_x	0	0	0	✓
Vert. offset	δ_y	0	0	0	✓
Energy spread	σ_E/E	0.003	0.0035	0.001	✓

† σ_z: gradient is zero by construction under the current discrete slicing scheme (see Section 4).

Table 4. Analytic (undisrupted) vs. AD elasticities for selected parameter–observable pairs. Analytic values assume Gaussian beams with no disruption ($H_D = 1$); AD values are from the full self-consistent simulation at $N_{\mathrm{macro}} = 10{,}000$, $N_{xy} = 64$.

Elasticity	Analytic	C3	CLIC	ILC	Disruption effect
$\partial\ln\mathcal{L}/\partial\ln N$	$+2.0$	$+1.92$	$+2.09$	$+2.10$	$H_D(N)$ adds to N^2
$\partial\ln\mathcal{L}/\partial\ln\sigma_x$	-1.0	-0.72	-0.83	-0.84	Pinch compensates wider beam
$\partial\ln\mathcal{L}/\partial\ln\sigma_y$	-1.0	$+0.45$	-0.19	-0.26	C3: disruption overshoot reverses sign
$\partial\ln\delta_{\mathrm{BS}}/\partial\ln N$	$+2.0$	$+2.01$	$+1.60$	$+1.79$	Self-consistent field enhancement
$\partial\ln\delta_{\mathrm{BS}}/\partial\ln E$	$+1.0$	$+0.94$	$+0.44$	$+0.65$	$\Upsilon \propto \gamma$; pinch modifies scaling
$\partial\ln n_\gamma/\partial\ln N$	$+1.0$	$+1.01$	$+0.92$	$+1.00$	Near-linear; consistent with $n_\gamma \propto N$
$\partial\ln n_\gamma/\partial\ln\sigma_x$	-1.0	-0.54	-0.41	-0.44	Flat-beam suppression: $\sigma_x \gg \sigma_y$

where $\hat{\mathcal{L}}$ and $\hat{\delta_{\mathrm{BS}}}$ are normalized to their nominal values, and $\theta = (\log\sigma_x, \log\sigma_y)$ are the optimization variables in log-space. We use the Adam optimizer with 100 gradient steps at each of 21 coarse weights plus dense sampling near the Pareto elbow.

The nominal operating points lie near the elbow of the Pareto front, confirming that the machine designs are already approximately Pareto-optimal. A sharp optimizer transition occurs near $\lambda \approx 0.5$: below the transition, the beamstrahlung penalty dominates; above it, the luminosity reward overwhelms the penalty and the optimizer reduces σ_x to maximize overlap.

H. Pair-production details

Table 7 summarizes the pair-production observable status and GG/GP ratios.

The discrepancies are structural consequences of implementation differences: **LL** (14–20% high) is sensitive to the virtual-photon displacement mode and low-statistics spectral bins. **BH** (-5% to $+20\%$) reflects spatial overcounting from GUINEAGRAD's grid-based photon-pair overlap. **BW** (order-of-magnitude) is dominated by infrared divergence regularization and energy-group discretization. Pair-production rates are computed after the luminosity calculation and do not affect primary observables.

I. Optimizer scaling analysis

Figure 10 shows the 4-parameter (N=4) convergence traces including random search as a baseline. Table 8 gives detailed time-to-threshold data.

To quantify the scaling advantage, we repeat the steering-recovery experiment with $N = 4, 6, 8, 10$, and 12 beam parameters, progressively adding beam-size, emittance, energy, and crossing-angle knobs (Table 9). Figure 11 shows the wall-clock time to 99% luminosity recovery as a function of N. Adam's convergence time *decreases* with N (from ∼1200 s at N=4 to ∼300–400 s at N=12), because the per-step cost of `value_and_grad` is independent of N

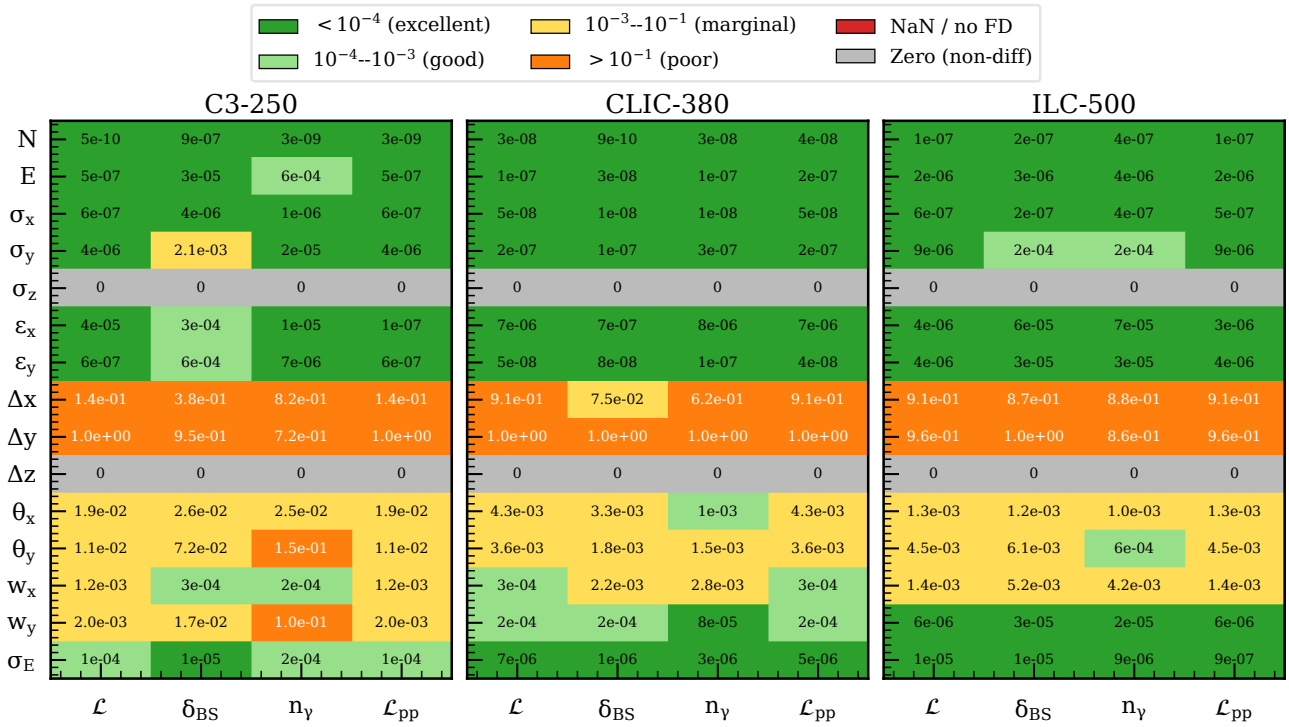

Figure 5. Full Jacobian verification: AD vs. finite-difference relative error for 15 beam parameters $\times$ 4 observables $\times$ 3 configurations ($\beta^*_{x,y}$ omitted as they are derived quantities in sigma-mode). Each cell shows the best relative error across step sizes $\epsilon \in \{10^{-3}, 10^{-5}, 10^{-7}\}$ at $N_{\text{macro}} = 5{,}000$. Green: agreement $< 10^{-4}$; light green: 10^{-4}–10^{-3}; yellow: 10^{-3}–10^{-1}; orange: $> 10^{-1}$; red: NaN/no FD data; grey: zero gradient (structurally non-differentiable).

while the richer gradient signal enables faster convergence. Nelder–Mead's time increases beyond $N \approx 6$–8 as the simplex grows.

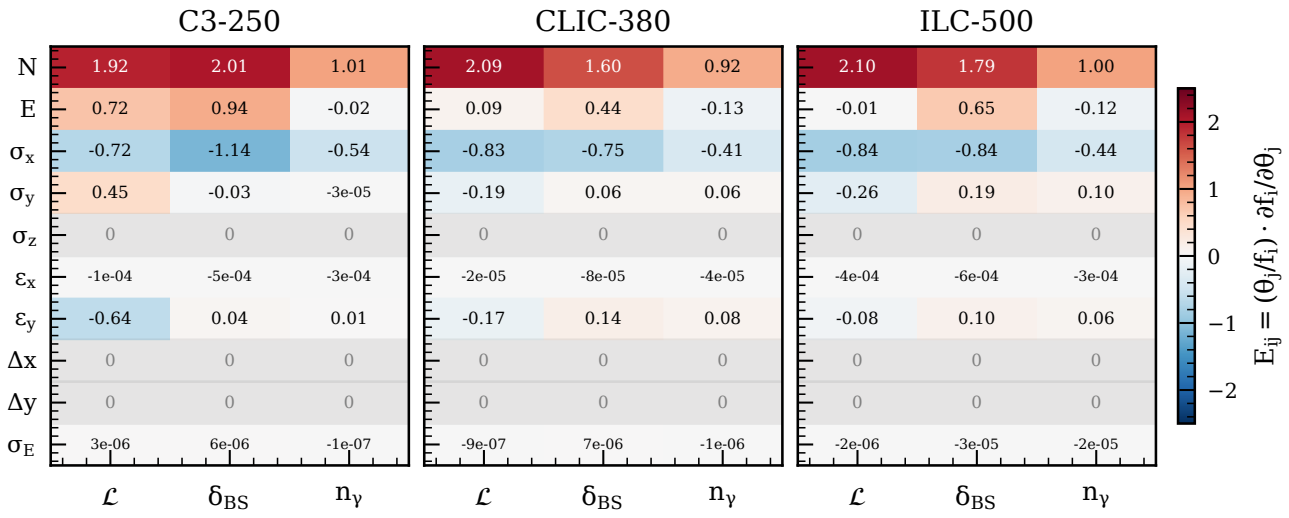

Figure 6. Normalized Jacobian (elasticity) heatmap for the 10 sigma-mode beam parameters and 3 primary observables across three configurations. Of the 10 parameters shown, 5 have visually significant elasticities ($|E| > 0.01$: N, E, σ_x, σ_y, ε_y); 2 have nonzero but negligible elasticities ($|E| \sim 10^{-4}$: ε_x, σ_E); and 3 are zero at the nominal point (shaded gray): σ_z (discrete slicing) and $\delta_{x,y}$ (elasticity normalization vanishes at $\delta = 0$; the raw gradients are large and nonzero). $N_{\text{macro}} = 10{,}000$, $N_{xy} = 64$.

Table 5. Explanation of zero or negligible elasticities in the Jacobian heatmap (Fig. 6). "Structural" zeros are artifacts of the current implementation; "physical" zeros reflect genuine weak sensitivity.

Parameter	Category	Explanation
σ_z	Structural	Discrete longitudinal slicing (`floor(z/Δz)`) makes slice assignment piecewise constant in σ_z; gradient is zero by construction.
ε_x	Physical	Enters only through the horizontal angular divergence $\sigma'_x = \varepsilon_x/(\gamma\sigma_x)$. For flat beams ($\sigma_x \gg \sigma_y$), σ'_x is small compared to the beam-beam electromagnetic kick. In contrast, ε_y enters through $\sigma'_y = \varepsilon_y/(\gamma\sigma_y)$ where $\sigma_y \sim$ 3–6 nm, yielding significant elasticities ($|E| \sim 0.1$–0.6).
δ_x, δ_y (at nominal)	Symmetry	At the nominal head-on point, luminosity is an even function of transverse offset; first derivatives vanish by symmetry. Offset sensitivity is second-order at the nominal point and first-order only away from zero offset, as in the mis-steered recovery experiment.
σ_E/E	Second-order	Gaussian symmetry cancels first-order effects; all elasticities are $\sim 10^{-5}$ at the nominal 0.3% spread.

Table 6. Full Jacobian computation time (9 sigma-mode params × 3 observables) via AD and finite differences at $N_{\text{macro}} = 10{,}000$, 64×64 grid. "FD (GPU)" uses GUINEAGRAD forward passes; "GP FD (CPU)" uses GUINEA-PIG.

Config	AD (s)	FD (GPU, s)	GP FD (CPU, s)	Speedup	
				vs. FD	vs. GP
C3-250	86.8	123.7	119	1.4×	1.4×
CLIC-380	79.5	111.1	154	1.4×	1.9×
ILC-500	80.0	111.2	164	1.4×	2.0×

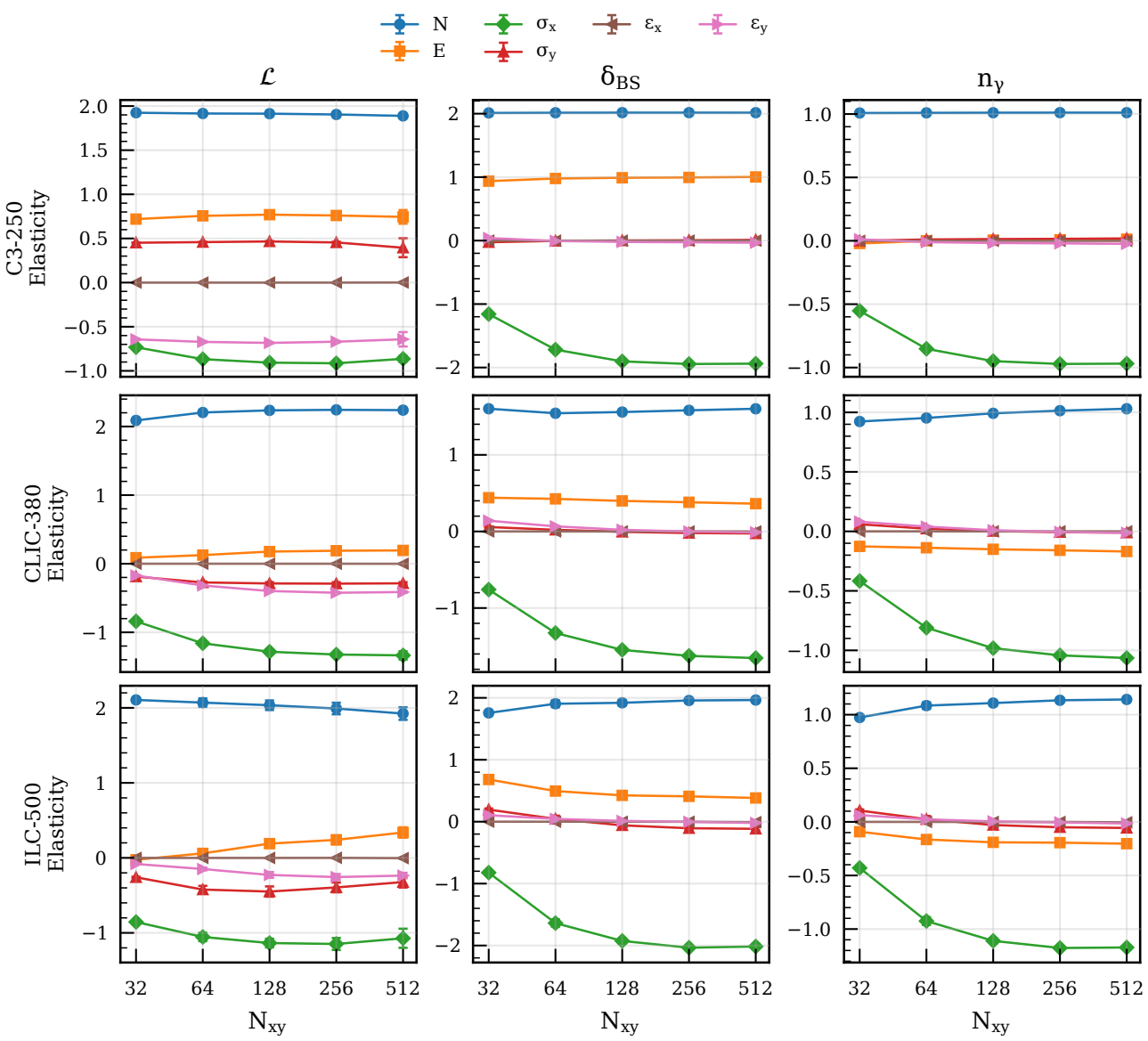

Figure 7. Gradient-resolution convergence: elasticity $E_{ij} = (\theta_j/f_i) \cdot \partial f_i/\partial\theta_j$ vs. transverse grid size N_{xy} for 7 beam parameters and 3 observables across all three configurations. Each point is the median over 5 seeds; error bars show ± 1 standard deviation. Bunch population N is stable across resolutions; transverse sizes (σ_x, σ_y) and emittances shift as the beam-overlap integral becomes better resolved. $N_{\text{macro}} = 10{,}000$ throughout.

Table 7. Pair-production observable status: GG/GP ratio per channel and configuration at matched resolution ($N_{\text{macro}} = 100\text{k}$, 512×512). "Gradient" indicates differentiability status.

Channel	Gradient	C3-250	CLIC-380	ILC-500	Suspected cause*
BH	Yes	1.04	0.95	1.20	Virtual photon Compton displacement*
LL	Rate only†	1.14	1.20	1.18	Gaussian beam-size smearing*
BW	Yes	0.48	2.13	0.83	IR regularization + energy groups*
Coherent	Fwd only‡		—		Computed via Erber formula; not GP-validated

* Root causes are based on displacement-mode scan evidence; these are leading hypotheses, not definitive attributions.
† LL rate gradient is computable via AD; the LL *spectrum* uses quasi-Monte Carlo accept/reject sampling that is not differentiable.
‡ Coherent pair rate is computed in the forward pass via the Erber asymptotic formula but is not included in the differentiable observable set.

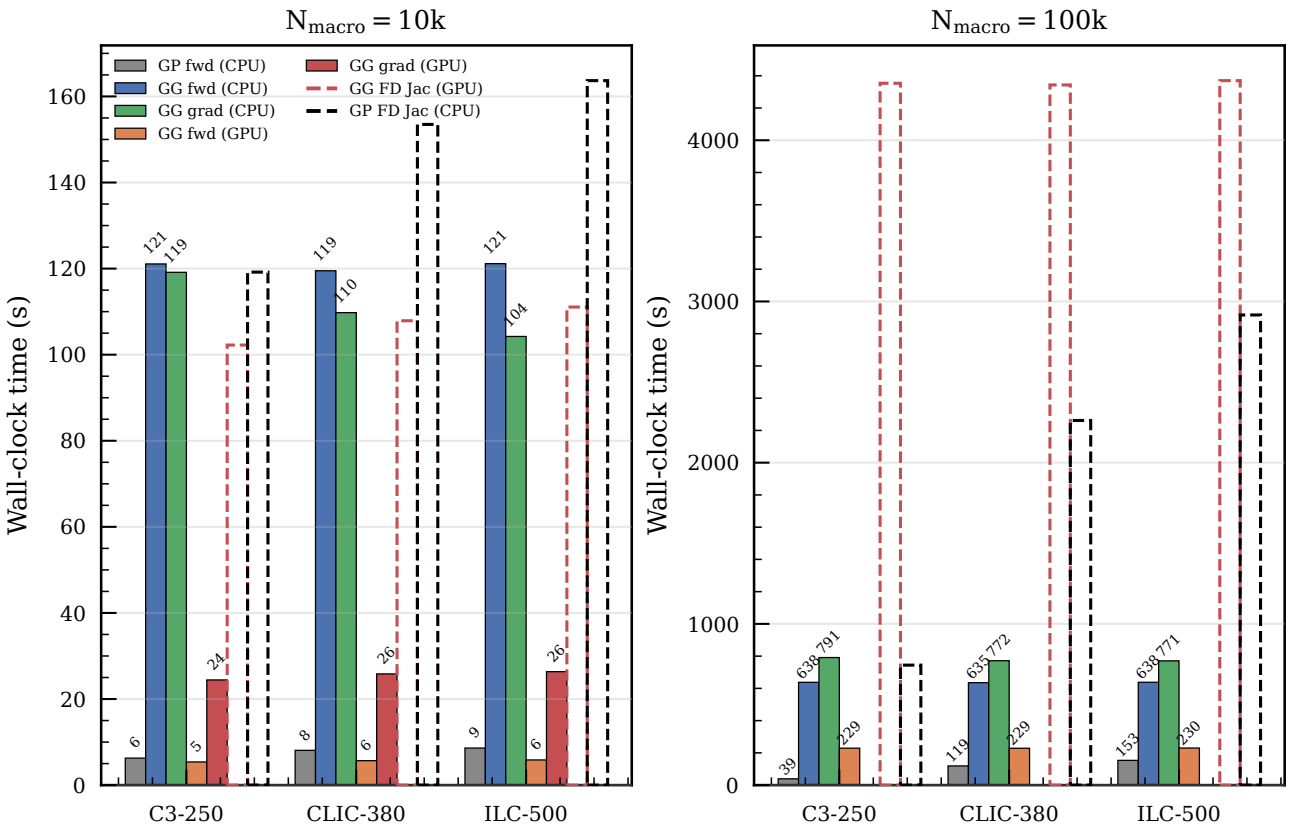

Figure 8. Wall-clock comparison across methods and hardware (64 × 64 grid). Left: $N_{\mathrm{macro}} = 10{,}000$. Right: $N_{\mathrm{macro}} = 100{,}000$. Solid bars show GP forward (CPU), GG forward and gradient on CPU and GPU; dashed-outline bars show the equivalent cost of a full 9-parameter Jacobian via finite differences ($2N+1 = 19$ forward evaluations). At $N_{\mathrm{macro}} = 100{,}000$ the GG GPU gradient exceeds 40 GB and is omitted; GG CPU gradient ($\sim$770 s, 1.2× the forward cost) remains the cheapest route to a full Jacobian compared with GP FD (19× the GP forward cost).

Table 8. Time-to-threshold for luminosity recovery. T_{95} and T_{99} denote the first evaluation at which the best-so-far luminosity exceeds 95% and 99% of nominal, respectively. Adam uses fewer function evaluations but each includes a gradient at $\sim$4.5× the forward cost; Nelder–Mead is faster in wall-clock for this 4-parameter problem. Random search never reaches 95% recovery.

Config	Method	T_{95} (evals)	T_{95} (s)	T_{99} (evals)	T_{99} (s)
C3-250	Adam	47	1097	52	1212
	Nelder–Mead	70	362	74	383
	Random	—		—	
CLIC-380	Adam	46	1086	49	1156
	Nelder–Mead	61	317	64	332
	Random	—		—	
ILC-500	Adam	46	1091	52	1232
	Nelder–Mead	59	306	86	443
	Random	—		—	

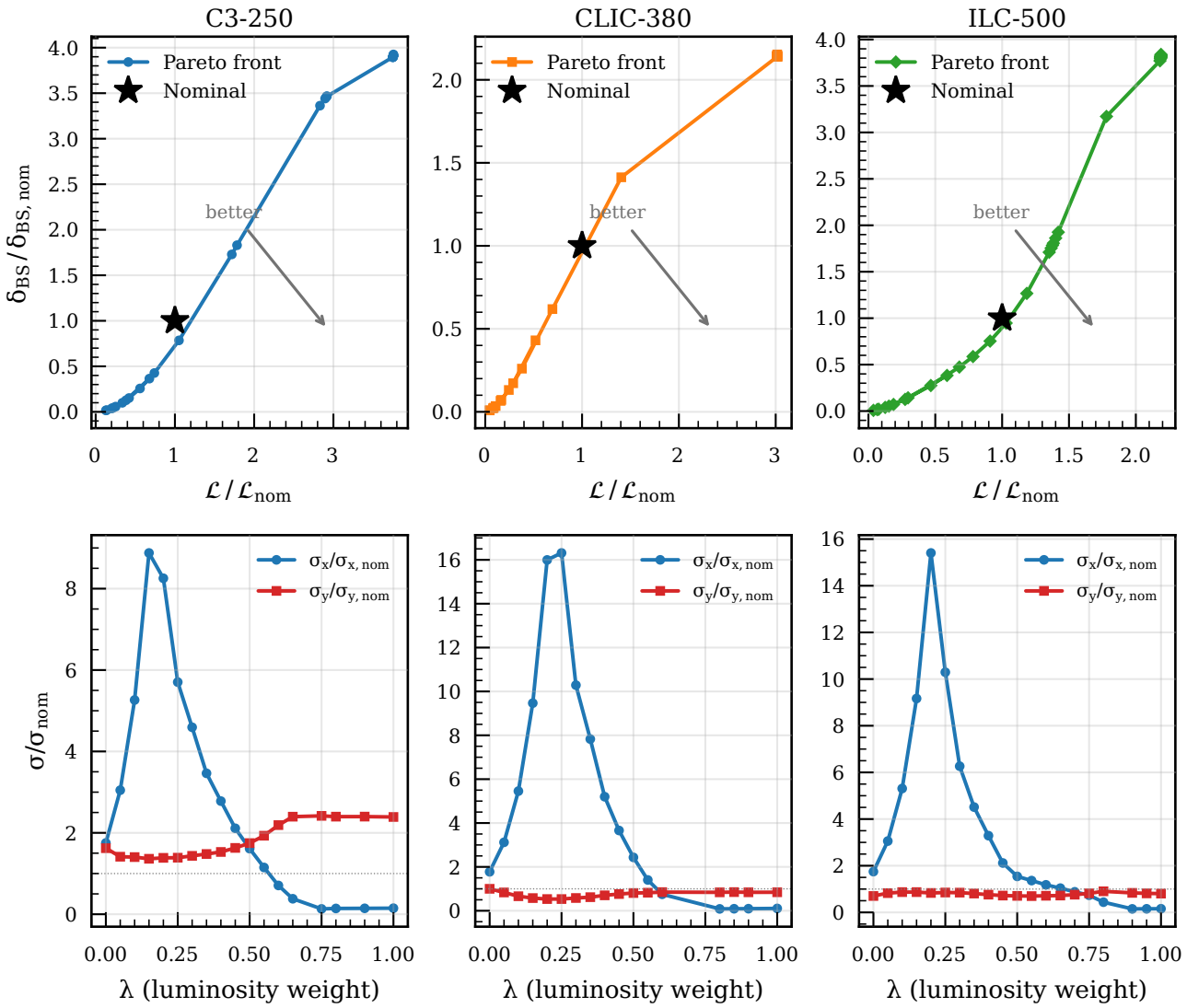

Figure 9. Multi-objective Pareto front: luminosity vs. beamstrahlung energy spread δ_{BS} for three configurations. Each point is the result of 100 Adam steps at a fixed weight λ optimizing $(\log \sigma_x, \log \sigma_y)$ at $N_{\mathrm{macro}} = 10{,}000$, $N_{xy} = 64$. Beam-size ratios are shown as $\sigma/\sigma_{\mathrm{nom}}$ (dimensionless). The nominal operating point is marked with a star. Lower panels show the optimized beam sizes as a function of λ.

Table 9. Beam parameters added at each optimisation tier. Parameters with zero nominal value (offsets, waist positions, crossing angles) are normalised by a physical reference scale; others are normalised by their nominal value.

N	Parameters added
4	$\delta_x,\ \delta_y,\ w_x,\ w_y$
6	$+\ \sigma_x,\ \sigma_y$
8	$+\ \varepsilon_y,\ E$
10	$+\ \theta_{cx},\ \theta_{cy}$
12	$+\ \varepsilon_x,\ \sigma_E$

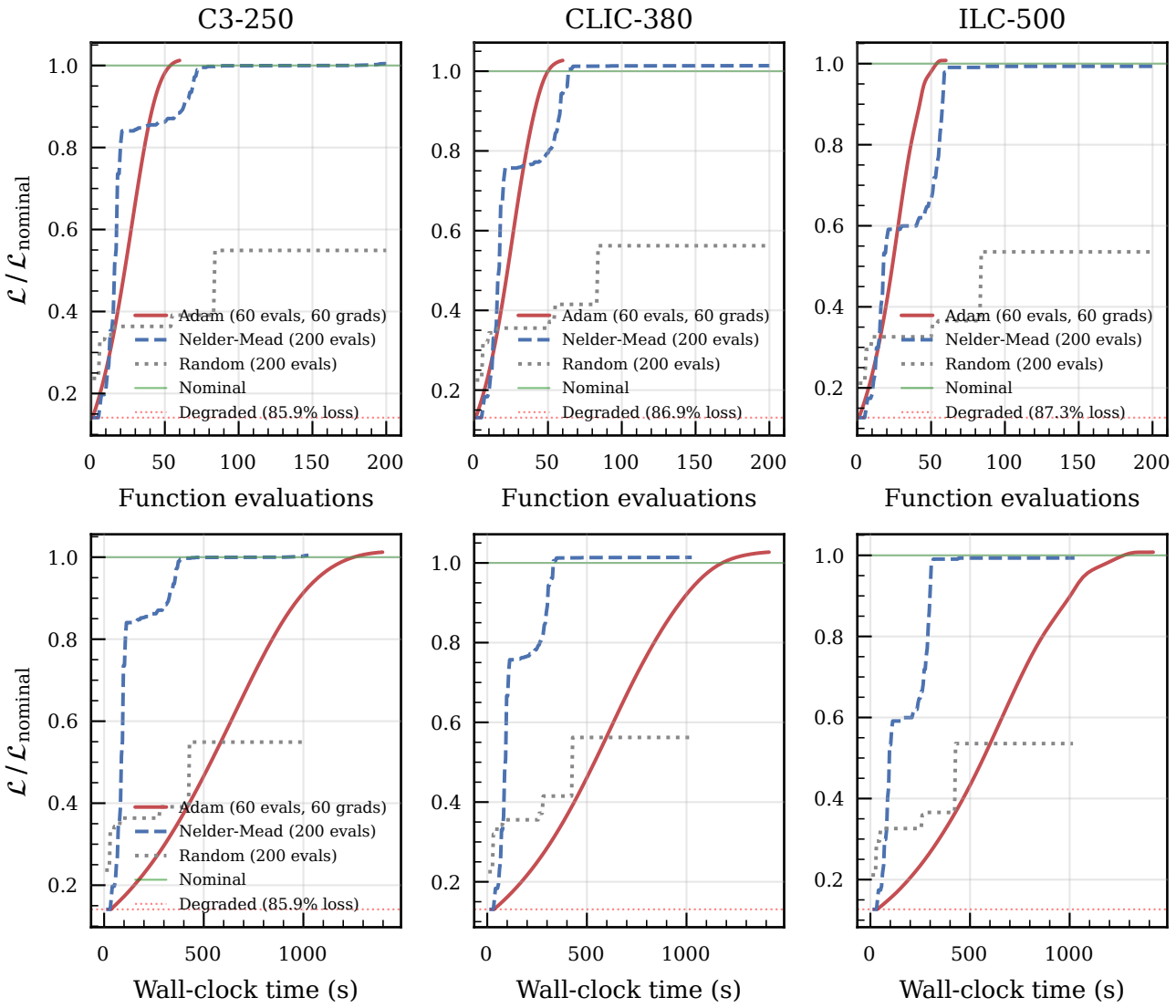

Figure 10. Luminosity recovery from beam mis-steering at $N = 4$ (offsets and waist positions only): luminosity (normalized to nominal) vs function evaluations (top) and wall-clock time (bottom) for Adam, Nelder–Mead, and random search across three configurations. For this low-dimensional problem, Nelder–Mead is faster in wall-clock despite requiring more function evaluations, because each forward-only evaluation costs $\sim 4.5\times$ less than Adam's value-plus-gradient step. Random search never exceeds $\sim 50\%$ recovery.

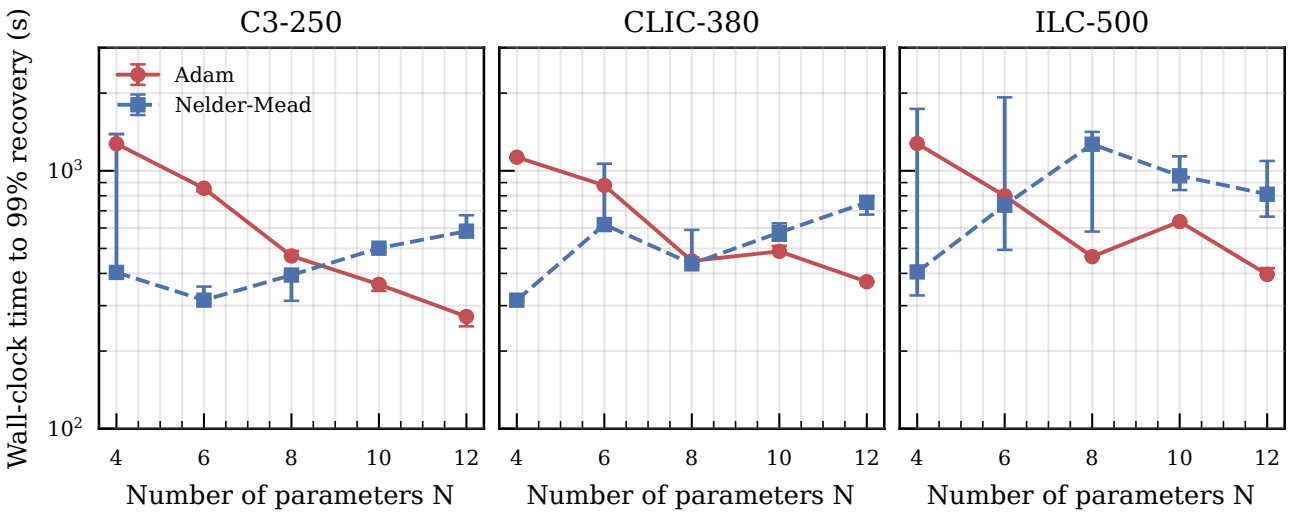

Figure 11. Wall-clock time to 99% luminosity recovery as a function of the number of optimised parameters N. Each panel corresponds to one collider configuration. Adam (red circles, solid) uses GUINEAGRAD reverse-mode gradients; Nelder–Mead (blue squares, dashed) uses forward-only evaluations. Error bars span the min/max over three random seeds. The crossover where gradient-based optimisation becomes wall-clock faster occurs at $N \approx 8$–10. Open markers with upward arrows indicate that the method did not reach 99% recovery within its evaluation budget.