

Exploring the Boundaries of Differentiable Radiation Transport and Detector Simulation

Jeffrey Krupa^{*1}, Yiyang Zhao², Mihaly Novak³, Max Aehle⁴, Max Sagebaum⁴, Long Chen⁴, Nicolas Gauger⁴, David Lange⁵, Vassil Vassilev⁵, Miaoyuan Liu⁶, Lukas Heinrich^{7,8}, and Michael Kagan^{*1}

¹SLAC National Accelerator Laboratory, Menlo Park, USA

²Tsinghua University, Beijing, China

³CERN, Geneva, Switzerland

⁴RPTU University Kaiserslautern-Landau, Kaiserslautern, Germany

⁵Princeton University, Princeton, USA

⁶Purdue University, West Lafayette, USA

⁷Technical University of Munich, Munich, Germany

⁸Munich Center for Machine Learning (MCML), Munich, Germany

Abstract

We present an application of automatic differentiation for particle transport through matter using a GEANT4-like radiation transport simulation with a full electromagnetic physics model. When differentiating this step-based transport, we observe exploding gradients driven by rare but extreme sensitivities at material boundaries, which propagate through subsequent transport and shower development. To obtain usable derivatives for optimization, we introduce a targeted mitigation strategy that stops gradient propagation through boundary-crossing operations under identifiable unstable conditions while leaving the forward (primal) simulation unchanged. We demonstrate that this enables stable, optimization-ready gradients in a detector-design problem.

1 Introduction

Radiation-transport simulations model the propagation of particles (such as photons, electrons, hadrons, etc.) through matter by combining deterministic particle navigation with stochastic sampling of physics interactions. High energy physics (HEP) leverages the high-fidelity radiation transport framework GEANT4 [1–3] extensively to model energy flow inside massive detectors. Outside of HEP, the same software is used extensively in applications including medical physics (e.g. DNA damage and radiation therapy [4–9]), space (e.g. instrumentation [10]), and nuclear (e.g. shielding [11, 12]).

High-energy physics detectors in particular have reached enormous complexity: modern collider experiments (e.g. the experiments at CERN’s Large Hadron Collider) contain on the order of 10^8 – 10^9 sensor elements. While these sensors are not tuned independently, their collective arrangement defines a hierarchically structured, high-dimensional design space spanning many design choices, such as geometry, materials, and operating conditions. Typically, the design of these detectors is informed by prior experience and hand-crafted heuristics, and tuned through resource-intensive simulation such as GEANT4. Informing the design of these high-dimensional systems using gradient-based

^{*}Corresponding authors: jkrupa@slac.stanford.edu, makagan@slac.stanford.edu

optimization would have significant utility, if accurate derivatives can be obtained for observables with respect to key parameters like geometry and composition.

More broadly, differentiable simulation—propagating derivatives through a physics simulator via automatic differentiation (AD) [13]—has emerged as a powerful paradigm across machine learning and computational science. By embedding a differentiable simulator inside a gradient-based optimization loop, one can solve inverse problems, tune design parameters, or train neural-network controllers end-to-end, without relying on surrogate models or finite-difference approximations. This differentiable programming approach has been used in differentiable rendering [14, 15], robotics and mechanics [16, 17], and scientific computing [18]. A recurring challenge in all these settings is that the underlying simulations contain discontinuities including visibility edges in rendering or boundary crossings in transport that can produce ill-defined or exploding gradients.

In this paper, we extend prior work [19] to present the first gradients obtained with full electromagnetic physics by applying AD to HEPEMSHOW [20], a lightweight transport application that implements the same stepping loop and physics engine (G4HEPEM [21]) used by GEANT4, but with a simplified planar geometry. In particular, we probe the accuracy, bias, and variance of derivatives by unpacking the stepper’s inner workings, yielding important insights into differentiable radiation transport.

Automating design with derivatives computed inside GEANT4 is a challenging technical task since the simulation itself is not built from the ground-up for propagating derivatives with AD. At its core, the simulation is non-smooth, relying on stochastic Monte Carlo (MC) sampling for modeling physics decays, and hard boundary decisions for modeling geometry. GEANT4 evolves showers by stepping individual particle tracks through volumes; at each step it samples interactions and may truncate the step at geometry boundaries. Indeed, as demonstrated in previous work [19], naïvely differentiating through geometry navigation in GEANT4 leads to exploding derivatives that are ill-suited for gradient-based optimization. For this reason, the physics engine used in these previous studies was incomplete. Specifically, multiple Coulomb scattering (MSC) and random fluctuations were not simulated.

We advance this prior work by extending it to the complete set of electromagnetic physics interactions, using MSC and fluctuations. To handle the exploding derivatives, we introduce a simple, physically-motivated mitigation rule that suppresses the dominant sources of gradient instability while leaving the forward simulation unchanged. The resulting derivatives are stable and low-variance enough to drive gradient-based optimization, though they carry a controlled bias from the suppressed geometry terms. We identify and characterize sources of gradient noise tied to geometry navigation, offering guidance for differentiable transport codes.

As a case study, we consider a *sampling calorimeter*, a device that measures particle energy through alternating layers of absorber and active material. Calorimeters are central to radiation measurement across many domains: sampling and homogeneous calorimeters are used throughout collider and astroparticle physics [22, 23] and water and graphite calorimeters serve as primary absorbed-dose standards in radiation therapy [24]. The layered geometry and well-understood physics of sampling calorimeters make them a canonical benchmark for detector optimization.

To demonstrate the complexity of this application, a single particle decay simulated by GEANT4 for a 50-layer sampling calorimeter comprised of lead (absorber layer) and liquid argon (active layer) is shown in Figure 1. The incoming particle, a 10 GeV electron, produces up to $\mathcal{O}(10^4)$ distinct secondary particles, shown as green lines.

In this paper, we contribute the following:

- **Identification and treatment of instabilities in differentiable particle transport.**

We demonstrate that certain interactions with boundaries in stepping schemes can induce

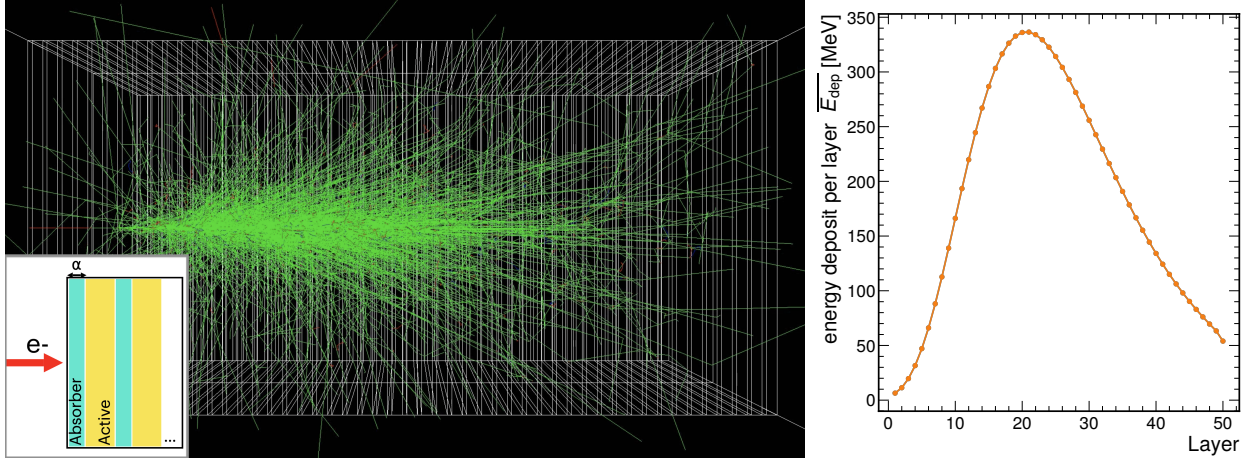


Figure 1: Left: Event display of a 50-layer sampling calorimeter illustrating an electromagnetic shower from an incident electron on the left side (inset: electron beam of energy E_{beam} entering a calorimeter comprised of alternating absorber/active layers with absorber thickness α). The complexity of the shower development motivates gradient-based optimization over manual tuning or grid search. Right: Mean longitudinal energy deposit per layer, showing the shower maximum in the middle layers.

singular derivatives of step length with respect to geometry parameters. These divergences propagate through the simulation chain and dominate event-level gradient variance.

- **A simple and physically consistent variance-control strategy.** We introduce a local *navigation-gradient stopping* rule that is applied only to geometry boundary navigation for tracks whose incidence angle with the local surface normal falls below a configurable threshold. This rule leaves the forward physics simulation unchanged while suppressing geometry-induced gradient noise by orders of magnitude.
- **Implications for differentiable detector simulation.** Our findings provide practical criteria for numerical stability in differentiable transport codes and concrete guidance for applying geometry-aware AD in more complex settings.

The remainder of the paper is organized as follows. We begin with a survey of related work on detector optimization and differentiable transport. We then describe the stepping structure of radiation-transport simulations and identify the general mechanisms that produce exploding gradients at material boundaries, using a sampling calorimeter as a concrete case study. Next, we present a variance-mitigation strategy based on selectively stopping gradients, which reduces variance at the cost of a small bias. Finally, we demonstrate the resulting derivatives and their utility in an optimization task.

2 Related Work

2.1 Detector Design Optimization

Traditional detector design in HEP relies on manual parameter tuning or grid searches over limited design spaces. Recent work has introduced a variety of data-driven approaches to systematically explore detector configurations. Several approaches focus on enabling gradient computations to

drive gradient-based optimization, such as creating differentiable simulations with differentiable programming [19, 25–32], as well as surrogate approaches [33–36] that train ML models to approximate simulators and enable differentiable optimization. The MODE collaboration is active in exploring end-to-end optimization of detectors with differentiable programming [37, 38] and co-design of detectors and software [39]. Bayesian optimization has also proven effective for low-dimensional problems with expensive simulations [40–42], offering sample-efficient black-box optimization without the need for gradient estimation. Additional approaches include evolutionary algorithms for multi-objective optimization [43] and reinforcement learning for sequential design decisions [44, 45]. Recent work also explores how Large Language Models, agentic AI, and diffusion models can be used in detector design [46–48]. Our work pursues direct differentiation of the underlying MC particle transport simulation.

2.2 Differentiable Particle Transport

Recent efforts have pursued multiple AD approaches directly on particle transport applications like GEANT4. Recent work has integrated forward- and reverse-mode operator-overloading AD into HEPEMSHOW [19] and forward-mode operator-overloading AD into GEANT4 [25]. Mean pathwise derivatives were computed for electromagnetic shower simulations in sampling calorimeters, achieving better performance than machine-code AD. A known limitation is the incompatibility of naïve AD application with MSC and fluctuations [19].

These efforts have established the technical feasibility of differentiating particle transport codes and have validated derivatives on simple geometric configurations. However, prior work has not investigated gradient behavior in realistic optimization scenarios or addressed gradient pathologies that may emerge when gradients must guide actual design decisions. Our work identifies systematic sources of exploding gradients driven by rare but extreme sensitivities in geometry navigation and boundary-limited stepping, and introduces a targeted mitigation strategy that enables stable gradients for detector optimization problems.

3 Background

3.1 Electromagnetic transport and stepping

Radiation-transport codes such as GEANT4 simulate the passage of particles through matter by tracking individual particles (“tracks”) through a sequence of discrete steps. The electromagnetic (EM) physics model encompasses all of the key interactions that charged particles and photons undergo in matter: ionization, bremsstrahlung, pair production, annihilation, Compton scattering, the photoelectric effect, as well as MSC, which produces angular deflection of charged tracks. Energy-loss fluctuations are also included. Previous work [19] demonstrated that AD can be applied to HEPEMSHOW, but MSC and energy-loss fluctuations had to be disabled because they produced excessive gradient variance. In this work, we include the complete set of EM interactions, including MSC and fluctuations, and address the resulting gradient instabilities.

At each step of the transport, physics imposes a constraint on the step length L_{phys} (e.g. due to discrete interactions, maximally allowed energy loss), while the geometry module computes the distance L_{bnd} to the nearest material boundary along the current direction. The realized step length L is the minimum of the two,

$$L = \min(L_{\text{phys}}, L_{\text{bnd}}), \quad (1)$$

so the update switches discontinuously between a physics-limited and a boundary-limited regime. This “min” operation is the critical point where geometry enters the transport: when $L = L_{\text{bnd}}$,

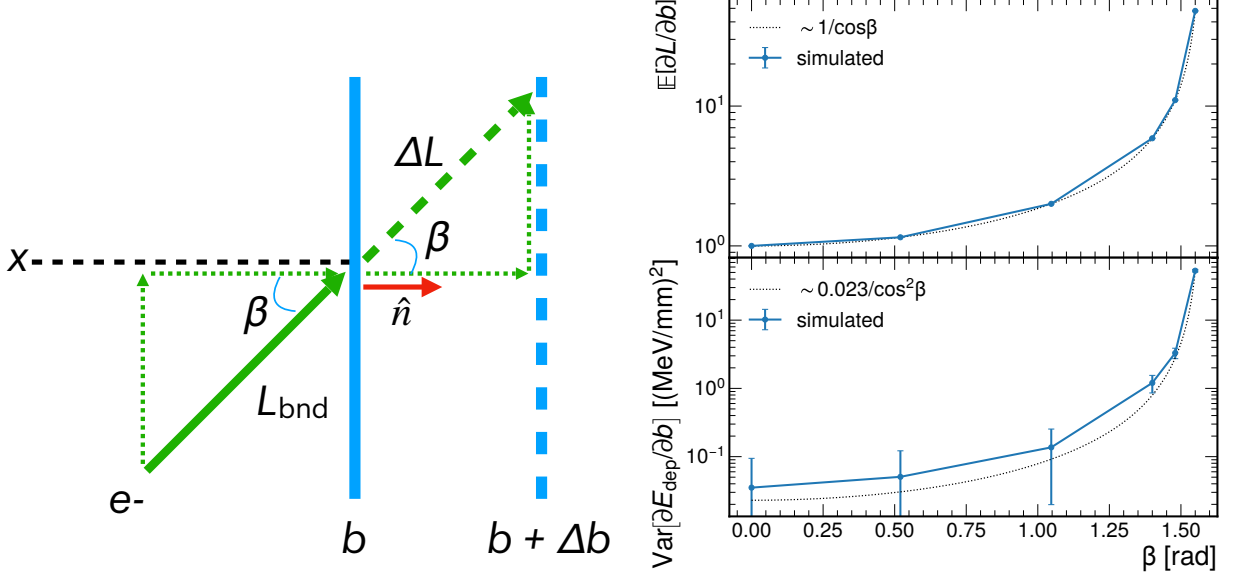


Figure 2: Left: a track of step length L incident on a boundary with angle β between the track direction $\hat{v}$ and the surface normal $\hat{n}$. Shifting the boundary by Δb along $\hat{n}$ changes the step length by $\Delta L \propto \Delta b / \cos \beta$. Right: the mean derivative of the step length with respect to boundary position (upper) and the variance of the per-step energy-deposit derivative (lower), both as a function of incidence angle β , confirming the $1/\cos \beta$ divergence as $\beta \rightarrow \pi/2$.

small perturbations of geometry parameters (e.g. boundary position or material thickness) or of the track direction can induce large changes in L and in the post-step state (position and material-region assignment). These sensitivities propagate to downstream steps and ultimately to any observable accumulated along the particle shower.

3.2 Gradient explosions at boundaries

When a step is boundary-limited ($L = L_{\text{bnd}}$), the realized step length depends directly on the intersection with the boundary. In general, the sensitivity of the step length to a boundary displacement depends on the angle between the track direction $\hat{v}$ and the outward surface normal $\hat{n}$ at the point of intersection. Defining $\cos \beta = \hat{v} \cdot \hat{n}$, a small shift Δb of the boundary along $\hat{n}$ changes the intersection distance by $\Delta L \approx \Delta b / \cos \beta$, implying

$$\left| \frac{\partial L}{\partial b} \right| \propto \frac{1}{|\cos \beta|}. \quad (2)$$

When a track travels nearly perpendicular to the normal vector ($|\cos \beta| \ll 1$, i.e. $\beta \rightarrow \pi/2$), this sensitivity diverges. An illustration of this geometric “slip” is shown in a planar geometry on the left side of Figure 2: a small normal displacement of the boundary produces a large tangential displacement of the post-step point, which can alter subsequent boundary-distance queries and transport decisions.

For sufficiently small steps, and in the absence of branching or splitting, the continuous energy deposited over a step scales approximately with step length, $E_{\text{dep}} \approx S(E) L$, where $S(E)$ is an energy-dependent stopping power. Boundary-limited steps inherit the same angular amplification in $\partial E_{\text{dep}} / \partial b$. Consequently, the variance of the per-step energy-deposit sensitivities grows rapidly as $\beta \rightarrow \pi/2$, scaling parametrically as $\sim 1/\cos^2 \beta$.

We validate this predicted scaling in a simplified two-layer geometry in which the beam position and boundary are arranged so that the first step is guaranteed to be boundary-limited. The experiment is repeated at six angles between 0 rad and 1.55 rad with 10 events per angle. For each event, we record the derivative at the first boundary-limited step only ($\partial L/\partial b$ and the corresponding $\partial E_{\text{dep}}/\partial b$). The measured mean of $\partial L/\partial b$ and the variance of $\partial E_{\text{dep}}/\partial b$ both diverge as $\beta \rightarrow \pi/2$, as shown in the right panels of Figure 2.

In addition to near-parallel incidence, we observe a second failure mode that is strongly correlated with extreme gradient outliers: *backward-going* boundary-limited steps. Using per-track diagnostics recorded during transport, we classify tracks by whether they undergo at least one backward boundary-limited step (moving opposite to the nominal beam direction). Although such steps are rare, they are disproportionately associated with exploding gradients. In our simulations, only a small fraction of tracks (a few percent) exhibit backward boundary-limited behavior, and they contribute a similarly small fraction of the total energy deposit. Nevertheless, these tracks account for more than 99% of the total squared derivative $\dot{E}_{\text{dep}}^2$ accumulated over charged tracks, dominating the gradient variance. This is consistent with the expectation that backward-going secondary particles at low energy undergo repeated boundary interactions, seeding large geometry-induced sensitivities that propagate through subsequent steps.

These two mechanisms—the $1/\cos\beta$ divergence at near-parallel incidence and the repeated-boundary instability for backward-going tracks—are general features of step-based transport through any geometry with material boundaries. They are not specific to a particular detector type.

3.3 Case study: sampling calorimeter

To study these gradient pathologies concretely, we consider a 50-layer sampling calorimeter consisting of alternating absorber (lead) and active (liquid argon) layers, a canonical detector-design benchmark. A mono-energetic electron beam enters normal to the layers, as shown in Figure 1, and the shower is simulated until all particle energy is deposited.

The parameters of interest are the incident beam energy E_{beam} , absorber layer thickness α , and active layer thickness g (in this setup, each of the 50 absorber and active layers has the same thickness). The key observable is the longitudinal energy-deposit profile: the total energy deposited in layer i by all tracks, $E_{\text{dep},i}$, averaged over events. The collection of these per-layer deposits, denoted E_{dep} , is shown in Figure 1 (right). The longitudinal profile directly reflects the shower development, sampling structure, and material budget, making it a natural objective for detector optimization. The baseline parameters are $E_{\text{beam}} = 10,000$ MeV, $\alpha = 2.3$ mm, and $g = 5.7$ mm, which provide near-complete containment of the shower.

We denote sensitivities of the mean longitudinal energy-deposit profile by $\partial \overline{E_{\text{dep},i}}/\partial E_{\text{beam}}$ and $\partial \overline{E_{\text{dep},i}}/\partial \alpha$. In forward-mode AD these per-layer derivatives are referred to as dot-values (tangent-linear sensitivities), written generically as $\dot{E}_{\text{dep},i}$.

Because all layer boundaries in this geometry are planar and perpendicular to the beam axis $\hat{x}$, the outward surface normal is $\hat{n} = \pm \hat{x}$ and $\cos\beta = v_x$, the x -component of the track direction. The near-parallel divergence of Equation 2 therefore corresponds to $|v_x| \ll 1$.

To illustrate how the general instabilities of subsection 3.2 manifest in this geometry, we examine how the deposited energy is distributed over track directions as a function of shower depth. Figure 3 shows the fraction of the charged-particle energy deposit as a function of $v_x = \hat{v} \cdot \hat{n}$. Tracks moving parallel to the incident beam have $v_x \approx +1$, while backward-going tracks have $v_x \approx -1$. Early layers are dominated by forward-going tracks, whereas later layers show an increasing contribution from large-angle and backward-going secondary particles, consistent with shower evolution toward lower energies and wider angular spread. Both near-parallel incidence ($|v_x| \ll 1$) and backward transport

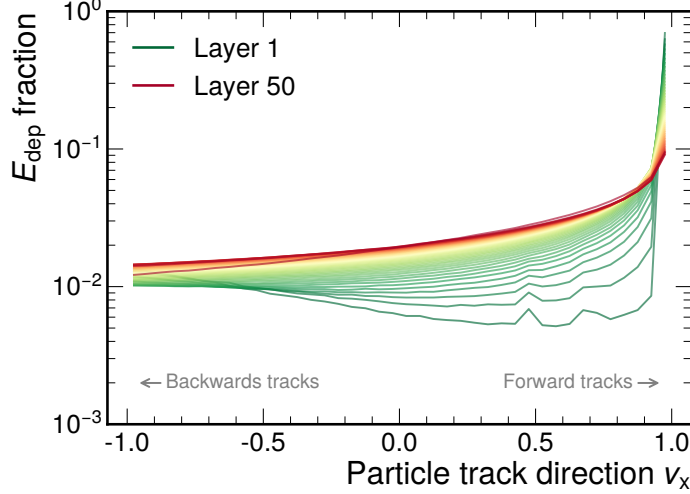


Figure 3: The fraction of track energy deposited as a function of the track direction projected onto the surface normal, $v_x = \hat{v} \cdot \hat{n}$. Each of the 50 layers is shown as a line in color gradients between green (layer 1) and red (layer 50). Later layers have a greater fraction of the deposit at small $|v_x|$ (near-parallel to the boundary) and $v_x < 0$ (backward-going), increasing the frequency of boundary-limited steps with large geometry-induced sensitivities.

($v_x < 0$) increase the probability of boundary-limited stepping and repeated boundary encounters, increasing exposure to the $1/\cos\beta$ divergence and the backward boundary-limited instability.

In the next section we describe a variance-mitigation strategy that selectively suppresses these geometry sensitivities while leaving the forward simulation and deposited-energy observables unchanged.

4 Methods

4.1 Differentiable simulation and AD

To explore these effects in a realistic particle simulator, we integrate an operator-overloaded AD package into a simplified framework with realistic electromagnetic physics simulation. In particular, the operator-overloaded AD tool CODIPACK [49–51] is integrated into a physics simulation toolkit, HEPEMSHOW [20, 52]. The methods here closely follow the original results presented in [19].

Electromagnetic interactions are modeled using the G4HEPEM physics engine [21], enabling a complete electromagnetic configuration in the studies presented here. We perform sensitivity analysis using forward-mode AD to measure the change in the simulation output when simulation parameters are varied. Reverse-mode AD, which is also supported by CODIPACK, is used to obtain adjoints for the optimization studies. The computational overhead introduced by forward and reverse mode AD is enumerated in [19].

4.2 Mitigating exploding gradients

A naïve integration of AD produces derivatives with extremely large variance (“exploding gradients”). As noted in [19], these gradients are not useful for downstream optimization because their magnitudes can fluctuate by many orders of magnitude across otherwise similar events. Motivated by the instabilities identified in subsection 3.2, we introduce a simple strategy to mitigate exploding

gradients that targets the dominant sources of instability. The approach reduces gradient variance substantially at the cost of introducing a small bias.

Our core tool is a *stop-gradient* operator (“stopgrad”), which blocks differentiation through a specific subset of the computation by setting the corresponding derivatives to zero. Critically, the forward (primal) simulation is left unchanged, allowing consistency and reproducibility. Such operators are widely used in machine-learning frameworks to prevent unstable or uninformative sensitivities from propagating.

In our implementation, we apply a stopgrad to the *geometry-induced* sensitivities (position and direction updates) for boundary-limited steps that satisfy simple kinematic criteria based on the incidence angle $\cos\beta = \hat{v} \cdot \hat{n}$ between the track direction and the surface normal. Specifically, we target:

- (i) **Near-parallel steps:** $|\hat{v} \cdot \hat{n}| < f$, where the threshold f controls how close to grazing incidence a track must be before triggering the stopgrad. These are steps for which the $1/\cos\beta$ sensitivity amplification identified in subsection 3.2 is largest.
- (ii) **Backward-going steps:** $\hat{v} \cdot \hat{n} < 0$, which are empirically associated with repeated boundary interactions and large accumulated sensitivities (cf. subsection 3.2).

In the planar geometry of our sampling calorimeter, $\hat{n} = \pm\hat{x}$ and $\hat{v} \cdot \hat{n} = v_x$, so these conditions reduce to $|v_x| < f$ and $v_x < 0$, respectively. More generally, for non-planar geometries, the local surface normal at each boundary intersection would be used.

Once a step triggers the stopgrad, the geometry-induced sensitivities for all subsequent steps of that track are also suppressed, since those steps inherit the perturbed post-step state and would otherwise reintroduce the same instability. A further choice concerns daughter particles created by the flagged track: their initial state also inherits the unstable sensitivities. We therefore consider two propagation rules and compare their effect:

- (i) **Track-only:** stopgrad is applied to the remainder of the flagged track; daughter particles are *not* affected.
- (ii) **Track + descendants:** stopgrad is applied to the remainder of the flagged track *and* to all particles it subsequently produces.

The track + descendants rule suppresses more derivative contributions and is used as the default configuration in this work (see Table 1 for a comparison). Further studies showing how the derivatives depend on f are presented in subsection 5.2. Note that choosing a parameter f depends on the bias and variance requirements of the application being considered.

The effect of the stopgrad is illustrated in Figure 4, which shows a representative low-energy charged track in the baseline AD configuration (left) and with our stopgrad protection enabled (right). The track originates near the bottom of the panel (with kinetic energy 1.6 MeV) and propagates in the negative- x direction toward the interface at $x = b$ (arrows indicate the direction of transport). Initially, the magnitude of the local sensitivity $|\partial E_{\text{dep}}/\partial\alpha|$ is modest, $\mathcal{O}(10^{-1})$. After several large-angle deflections from MSC—a common feature at low energies—the track repeatedly encounters the boundary and undergoes a sequence of boundary-limited steps near the interface. In the unprotected configuration, these repeated boundary interactions lead to large, intermittent spikes in $|\partial E_{\text{dep}}/\partial\alpha|$ that accumulate over many steps, reaching $\mathcal{O}(10^8)$ MeV mm $^{-1}$ by termination. With the stopgrad enabled, the primal trajectory is unchanged, but the unstable geometry-induced sensitivities are suppressed and the maximum derivative remains at $\mathcal{O}(10^2)$ MeV mm $^{-1}$.

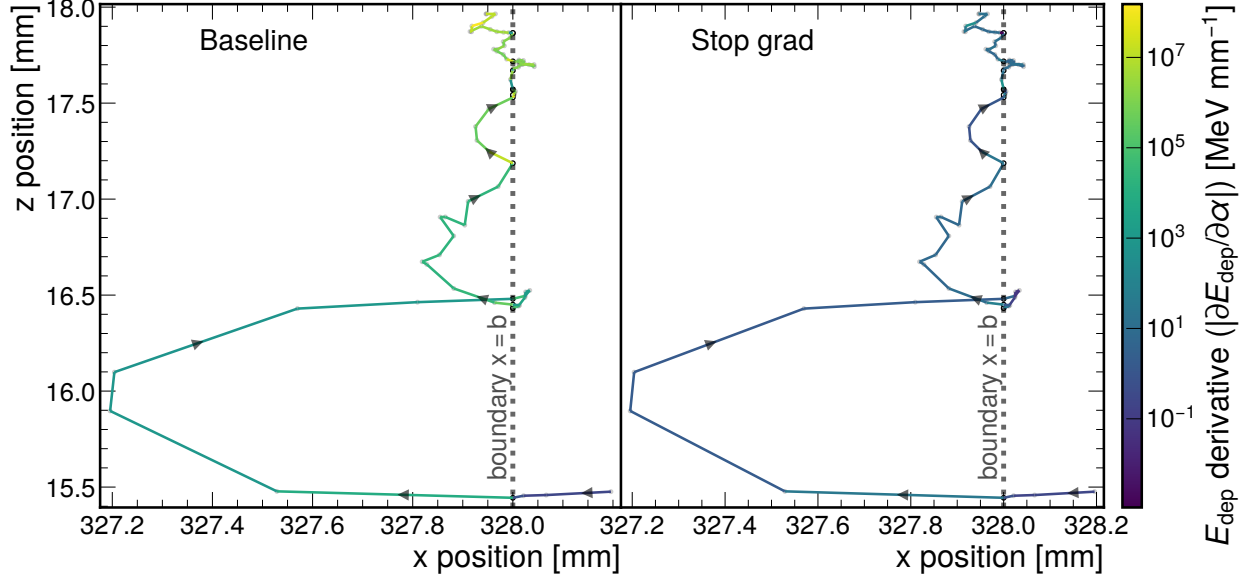


Figure 4: An example of a low-energy track exhibiting repeated boundary-limited steps near boundary $x = b$ in layer 41. Baseline AD (left) produces extreme $|\partial E_{\text{dep}}/\partial \alpha|$ spikes during backscatter/ping-pong near the interface, with values accumulating to 10^8 MeV mm^{-1} over 86 steps. Applying the geometric stopgrad for backward/near-parallel (right) steps suppresses these unstable geometry sensitivities by about 10^6 . The primal mode and trajectory remain unchanged.

Lastly, two additional *derivative-only regularizations* address near-singular Jacobian terms that arise in specific physics sub-models. In both cases we apply the same primal-preserving pattern: for a function $f(\mathbf{x})$ whose exact Jacobian ∇f may diverge when an intermediate quantity approaches zero, we write the regularized output as

$$\tilde{f}(\mathbf{x}) = \underbrace{\text{sg}(f(\mathbf{x}))}_{\text{exact primal}} + \underbrace{\nabla f_{\text{reg}}(\text{sg}(\mathbf{x})) \cdot (\mathbf{x} - \text{sg}(\mathbf{x}))}_{\text{regularized first-order correction}}, \quad (3)$$

where $\text{sg}(\cdot)$ denotes the stop-gradient operator (identity on values, zero on derivatives) and ∇f_{reg} is the Jacobian of f evaluated with near-singular denominators or arguments floored/capped at a threshold. By construction, $\tilde{f}$ returns the exact forward value $f(\mathbf{x})$ but injects a well-behaved Jacobian into the AD tape.

MSC path-length conversion. The MSC model is responsible for accounting for the net effects of the large number of small angular deflections along the condensed history step of charged particles as they traverse matter. The simulation therefore distinguishes between the *true path length*, t , and the *geometrical path length*, z , of a condensed history step. The former corresponds to the length of the actual path traversed between the pre- and post-step points if the many angular deflections are included, while the latter is the length of the projection of the vector pointing from the pre- to the post-step point along the original direction. In the MSC model used here [53], the two are related in the short-step regime (i.e. when energy loss along the step can be neglected) by

$$z = \lambda_1 \left(1 - e^{-t/\lambda_1}\right), \quad (4)$$

where λ_1 is the first transport mean free path for elastic scattering. When a boundary-limited step shortens the geometric path, the simulator must invert this relation to recover the true step length:

$$t = -\lambda_1 \ln\left(1 - \frac{z}{\lambda_1}\right), \quad \frac{dt}{dz} = \frac{1}{1 - z/\lambda_1}. \quad (5)$$

The derivative diverges as $z \rightarrow \lambda_1$. More generally, when energy loss along the step is accounted for, the conversion formula introduces additional rational functions of λ_1 in the denominator, all of which can become near-singular when the step length approaches the particle's remaining range. We regularize by evaluating the Jacobian with all such denominators floored at a threshold ϵ_{conv} , following Equation 3.

Photon step limit. The step length of a photon until the next physics interaction (Compton scattering, pair production, photoelectric absorption) follows an exponential distribution with the inverse macroscopic cross section, called the mean free path (MFP), as its parameter. Therefore, the step length until the next physics interaction is sampled by drawing the number of mean free paths to the next interaction, $n_{\text{left}} = -\ln(u)$ with $u \sim \text{Uniform}(0, 1)$, and converting it to a step length by $L_{\text{phys}} = \lambda n_{\text{left}}$ using the actual value of $\lambda(E, \text{Material})$, the energy- and material-dependent MFP. After each step of length ℓ and MFP of λ , the counter is decremented by removing the corresponding number of mean free paths:

$$n_{\text{left}} \leftarrow n_{\text{left}} - \frac{\ell}{\lambda}. \quad (6)$$

When λ is very large (i.e. the material is nearly transparent at a given photon energy), the step limit $L_{\text{phys}} = \lambda n_{\text{left}}$ amplifies any sensitivity in λ into a correspondingly large sensitivity in the proposed step length. Following Equation 3, we regularize by capping the MFP at λ_{clip} in the derivative evaluation.

Scans over both ϵ_{conv} and λ_{cap} are shown in Appendix A.

4.3 Experimental setup

In addition to computing derivatives of the HEPESHOW simulation using AD with CoDiPACK, we calculate derivatives with finite differences, providing independent validation of the derivatives. A central difference is computed for the energy deposited in each layer, $E_{\text{dep},\ell}$, with respect to a parameter. A common set of random numbers is used to seed the forward and backward differences. While finite differences serve as a useful cross-check, they become impractical in realistic settings: each additional parameter requires two extra simulation passes, making the cost scale linearly with the number of parameters; and achieving low-variance estimates requires large event samples for each evaluation, compounding the expense. AD, by contrast, obtains all derivatives in a single simulation pass (forward mode per parameter, or reverse mode for all parameters simultaneously), making it the only viable route for high-dimensional or large-sample sensitivity analyses.

To demonstrate the usefulness of these derivatives in an optimization setting, we apply the derivatives with AD in the context of a two-parameter optimization, closely following the method in [19]. First, a reference longitudinal energy-deposit profile is computed at the nominal parameters $(E_{\text{beam}}^*, \alpha^*) = (10^4 \text{ MeV}, 2.3 \text{ mm})$. The energy-deposit profile is estimated using 100k events. The optimization loop then proceeds by minimizing the mean-squared-error between the current longitudinal shower profile and the reference profile:

$$L(E_{\text{beam}}, \alpha) = \sum_{i=1}^{50} (\overline{E_{\text{dep}_i}}(E_{\text{beam}}, \alpha) - \overline{E_{\text{dep}_i}}(E_{\text{beam}}^*, \alpha^*))^2. \quad (7)$$

At each iteration of the optimization (epoch), we estimate $\overline{E_{\text{dep}_i}}(E_{\text{beam}}, \alpha)$ using 5k events, and compute gradients with the AD-enabled simulation. Our variance-mitigation rules are applied.

We use a mini-batch stochastic gradient descent (SGD) approach for the optimization. The parameter updates are clipped to stabilize occasional large steps. The initial clipping parameters are $(\Delta E_{\text{beam}}^{\text{max}}, \Delta \alpha^{\text{max}}) = (2000 \text{ MeV}, 1 \text{ mm})$. The learning rate is chosen such that typical parameter updates are comparable in scale across the two coordinates. The learning rate and clipping parameters are reduced by a factor of 2 after 30 epochs with no improvement in the best loss. The optimization terminates after 3 decays. A maximum of 200 epochs is considered.

5 Results

5.1 Gradients

Forward-mode derivatives from HEPEMSHOW after applying our variance-mitigation (stopgrad) rule are shown in Figure 5. The left and right panels show $\partial \overline{E_{\text{dep}}}/\partial E_{\text{beam}}$ and $\partial \overline{E_{\text{dep}}}/\partial \alpha$, respectively, as functions of calorimeter layer. The AD result (orange) applies the stopgrad to geometry-induced sensitivities on boundary-limited steps with $\hat{v} \cdot \hat{n} < 0.2$ (i.e. $v_x < 0.2$ in our planar geometry), covering all backward-going steps and near-parallel steps limited by a boundary. Derivatives are estimated from 10^8 events; error bars show the standard error on the mean ($\sigma/\sqrt{N}$) per layer. Finite-difference (blue) derivatives use a central difference with step sizes $h = 50 \text{ MeV}$ for E_{beam} and $h = 0.005 \text{ mm}$ for α (with common random numbers between $p \pm h$)¹. For reference, the unprotected AD estimator is shown in an inset with a different y-scale.

Without protection from the stopgrad, the AD estimator is dominated by rare outliers, producing layer-wise fluctuations that are 6 orders of magnitude larger than the stopgrad result. This prevents comparison on a single axis; a different y-axis is used in the inset (no stopgrad). The variance-mitigated AD results agree with finite differences within statistical uncertainty for E_{beam} . For α , the overall shape and peak location are consistent with finite differences, but AD under-estimates the peak magnitude by approximately 25%, indicating a residual bias introduced by suppressing geometry-driven sensitivities in the central region.

Uncertainties increase in the late layers ($\gtrsim 35$) since the deposited energy per event is small and the shower becomes more angularly diffused: a larger fraction of the energy deposit comes from large-angle and backward-going secondary particles (cf. Figure 3). In this regime, low-energy charged tracks undergo stronger MSC deflections and more frequent boundary interactions, leading to a heavier-tailed per-event derivative distribution and hence larger $\sigma/\sqrt{N}$. While the stopgrad rule suppresses catastrophic outliers by ~ 6 orders of magnitude, additional variance-reduction strategies may further stabilize the late-layer estimates. Overall, these derivatives are stable and well-suited for downstream optimization, as shown in subsection 5.3.

5.2 Bias analysis

To understand the bias in the derivative caused by the derivative regularization methods introduced in subsection 4.2, we measure the energy deposit derivative with respect to absorber thickness by scanning each regularization parameter independently while holding the others fixed. The resulting AD derivatives are then compared against central finite differences.

To study the dependence on f , the stopgrad threshold f is scanned for $f \in \{0.0, 0.1, 0.2, 0.5\}$. The results are shown in Figure 6. The central values of the AD derivatives are stable across this range with shapes generally close to the finite difference, most notably deviating by an underprediction of

¹Scans against step size show that derivatives are insensitive to choice of h .

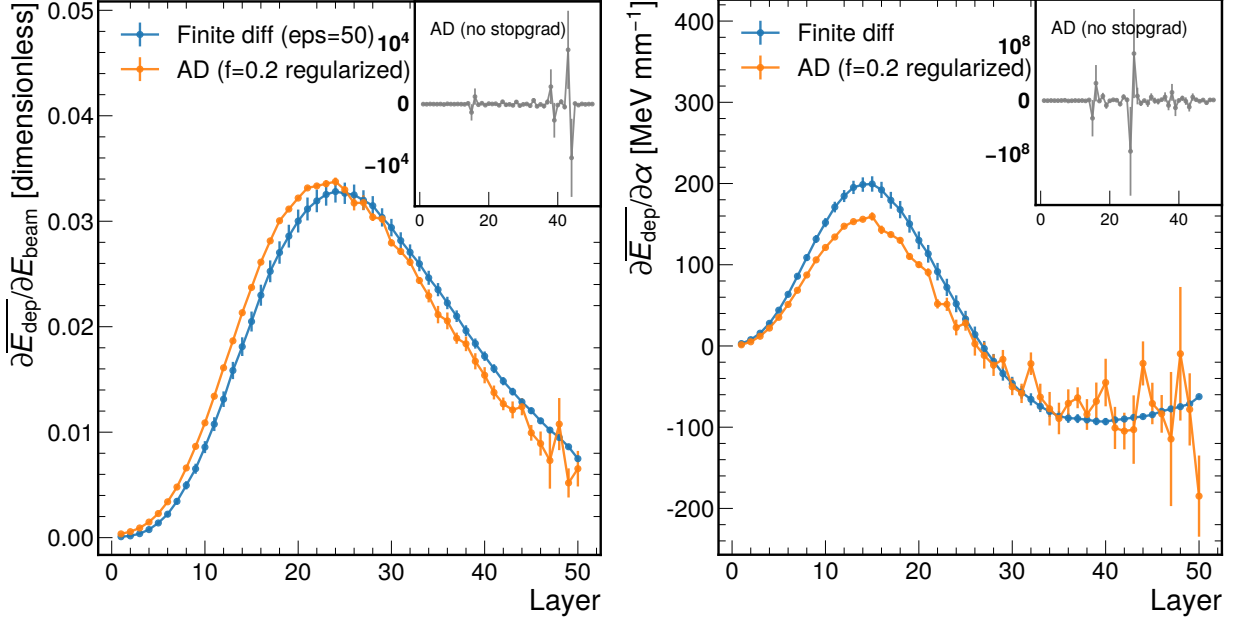


Figure 5: Mean longitudinal energy-deposit derivative per calorimeter layer comparing finite differences (blue) to AD (orange) with the navigation stop-gradient rule applied at boundary-limited steps with $\hat{v} \cdot \hat{n} < 0.2$ ($v_x < 0.2$). Left: derivative with respect to the beam energy, $\partial \overline{E}_{\text{dep}} / \partial E_{\text{beam}}$, showing close agreement in shape and scale between AD and finite differences across layers. Right: derivative with respect to absorber thickness, $\partial \overline{E}_{\text{dep}} / \partial \alpha$, where the stopgrad rule yields a stable, finite-variance estimate. The insets show the unprotected AD estimator (“no stopgrad”) exhibiting rare, catastrophic outliers and orders-of-magnitude larger fluctuations (note the difference in scales). Error bars show $\sigma / \sqrt{N}$ per layer.

25% at layer 14. Smaller values of f , such as $f = 0$, demonstrate large variance at specific layers. Larger values of f , such as $f = 0.5$ suppress more derivative contributions from near-boundary steps, slightly reducing variance at the cost of increased bias.

Additionally, the photon MFP cap is scanned over $\lambda_{\text{cap}} \in \{100, 1000, 10000\}$ mm, and the MSC conversion regularization parameter is scanned over $\epsilon_{\text{conv}} \in \{10^{-4}, 10^{-3}, 10^{-2}\}$. These results are shown in Appendix A (Figure 8). The choices $\lambda_{\text{cap}} = 1000$ mm and $\epsilon_{\text{conv}} = 10^{-3}$ exhibit minimal additional bias while providing further variance reduction. For these reasons, we adopt the working point $f = 0.2$, $\epsilon_{\text{conv}} = 10^{-3}$, and $\lambda_{\text{cap}} = 1000$ mm for the main text of this paper. A cross-check on the robustness of the stopgrad strategy is presented in Appendix A (Figure 9), using a looser threshold of $f = 0.05$. Good convergence of the optimization is obtained for both choices of f .

5.3 Optimization

We perform 50 optimization runs with identical hyperparameters and different random seeds. Their trajectories are shown in Figure 7. All optimization runs descend into the basin of the minimum (2D contours) and terminate near the target parameters ($E_{\text{beam}}^*, \alpha^*$) (shown as a purple diamond). We summarize the paths of each optimization using the median trajectory (blue) with a 5–95% band.

A stochastic gradient descent strategy is employed with a per-coordinate learning rate of $(1, 1 \times 10^{-7} \text{ mm}^2 \text{ MeV}^{-2})$, where the ratio is set by the natural scale a^2 / E_0^2 to equalize the fractional

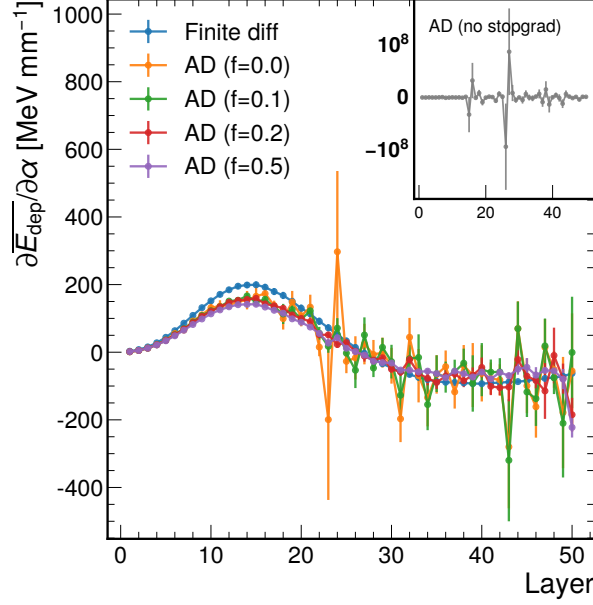


Figure 6: Mean longitudinal energy-deposit derivative per calorimeter layer comparing finite differences (blue) to AD with various choices of the stopgrad threshold f (i.e. stopgrad is applied when $|\hat{v} \cdot \hat{n}| < f$).

step sizes across the two parameters. Gradient clipping is found to be infrequent, triggered in only 0.02% of gradient updates (2 out of 1×10^4 total steps across 50 independent seeds), indicating that the gradients remain stable throughout the optimization. The final parameters of the optimizations are $E_{\text{beam}}^{\text{final}} = 9995 \pm 6$ MeV, and $\alpha^{\text{final}} = 2.307 \pm 0.007$ mm, consistent with E_{beam}^* and α^* .

An optimization with identical settings, except using a looser threshold of $f = 0.05$ is shown in Appendix A (Figure 9). Similar convergence quality with a larger fraction of clipped steps is observed.

6 Conclusion

We studied automatic differentiation in a GEANT4-like electromagnetic radiation transport simulation and identified rare, geometry-coupled code paths that produce exploding gradient variance, making naïve derivatives unusable for optimization. We observed that large-variance gradients were strongly correlated with a small subset of boundary-limited transport configurations, particularly those involving near-parallel incidence and backward-going intersections on boundaries, which can trigger repeated boundary interactions and amplify geometry-induced sensitivities.

To mitigate these instabilities, we introduced a targeted navigation stop-gradient rule applied only to geometry-induced updates during boundary-limited stepping under simple kinematic conditions. This intervention leaves the forward (primal) simulation unchanged while suppressing catastrophic outliers by 6 orders of magnitude, yielding stable, optimization-ready derivatives with a complete electromagnetic physics model including multiple Coulomb scattering.

Finally, we demonstrated the utility of the derivatives in a detector optimization, where ensembles of randomly seeded runs reliably converged to the target shower-profile parameters. These results establish practical criteria for numerical stability in differentiable particle transport. Furthermore, they suggest a path toward gradient-based detector and radiation-system design in high-dimensional

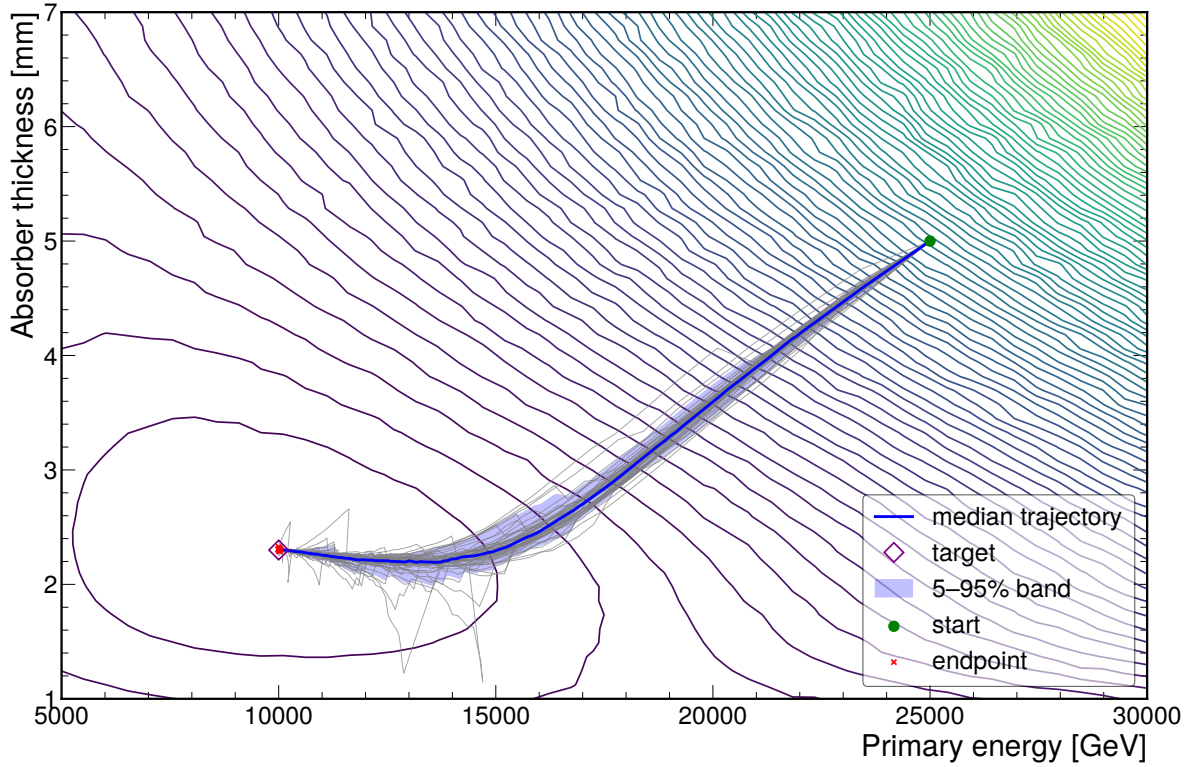


Figure 7: Trajectories of 50 randomly-seeded optimization runs minimizing the longitudinal-profile loss. Each grey curve shows one run, initialized at $(E_{\text{beam}}, \alpha) = (25000 \text{ MeV}, 5.0 \text{ mm})$ (green circle) and terminating at the final iterate (red cross). Contours show the loss landscape; the target $(E_{\text{beam}}^*, \alpha^*)$ is marked with a purple diamond. The blue curve shows the median trajectory with a 5–95% band.

settings, with potential applications including medical physics, shielding, and space instrumentation.

Ultimately, differentiable simulation has the potential to close the loop between detector design, simulation, and analysis, enabling end-to-end optimization pipelines that were previously intractable. Scaling to full hadronic physics and realistic detector geometries remain the key challenges, but the stability criteria identified here provide a foundation for systematically tackling these more complex settings.

Acknowledgements

J.K. and M.K. are supported by the US Department of Energy (DOE) under Grant No. DE-AC02-76SF00515. The authors from RPTU University Kaiserslautern-Landau appreciate funding from the German National High Performance Computing (NHR) association for the Center NHR South-West. L.H. is supported by BMFTR Project SciFM 05D2025 and by the Excellence Cluster ORIGINS, which is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany’s Excellence Strategy - EXC-2094-390783311. D.L. and V.V. are supported by the National Science Foundation (NSF) under Grant OAC-2311471.

Code availability

The simulation code and derivative-regularization implementations used in this work are publicly available. The modified G4HEPEM physics engine is at <https://github.com/jeffkrupa/g4hepem/releases/tag/v1.0-paper>, and the HEPEMSHOW application is at <https://github.com/jeffkrupa/hepemshow/releases/tag/v1.0-paper>.

References

- [1] Agostinelli S *et al.* 2003 Geant4—a simulation toolkit *Nucl. Instrum. Methods Phys. Res., Sect. A* **506** 250–303 doi:10.1016/S0168-9002(03)01368-8
- [2] Allison J *et al.* 2006 Geant4 developments and applications *IEEE Trans. Nucl. Sci.* **53** 270–278 doi:10.1109/TNS.2006.869826
- [3] Allison J *et al.* 2016 Recent developments in Geant4 *Nucl. Instrum. Methods Phys. Res., Sect. A* **835** 186–225 doi:10.1016/j.nima.2016.06.125
- [4] Zein S A, Bordage M C, Francis Z, Macetti G, Genoni A, Cappello C D, Shin W G and Incerti S 2021 Electron transport in DNA bases: An extension of the Geant4-DNA Monte Carlo toolkit *Nucl. Instrum. Methods Phys. Res., Sect. B* **488** 70–82 doi:10.1016/j.nimb.2020.11.021
- [5] Domínguez-Muñoz A D, Gallardo M I, Bordage M C, Francis Z, Incerti S and Cortés-Giraldo M A 2022 A model for Geant4-DNA to simulate ionization and excitation of liquid water by protons travelling above 100 MeV *Radiat. Phys. Chem.* **199** 110363 doi:10.1016/j.radphyschem.2022.110363
- [6] Ganjgah A A and Taherparvar P 2024 The effects of cell displacement on DNA damages in targeted radiation therapy using Geant4-DNA *Sci. Rep.* **14** 81863 doi:10.1038/s41598-024-81863-4
- [7] Tabbakh F 2024 Significance of the proton energy loss mechanism to gold nanoparticles in proton therapy: a Geant4 simulation *Scientific Reports* **14** 76244 doi:10.1038/s41598-024-76244-w
- [8] Shin W G *et al.* 2021 A Geant4-DNA Evaluation of Radiation-Induced DNA Damage on a Human Fibroblast *Cancers* **13** 4940 doi:10.3390/cancers13194940
- [9] Chappuis F, Tran H N, Zein S A, Bailat C, Incerti S, Bochud F and Desorgher L 2023 The general-purpose Geant4 Monte Carlo toolkit and its Geant4-DNA extension to investigate mechanisms underlying the FLASH effect in radiotherapy: Current status and challenges *Physica Med.* **110** 102601 doi:10.1016/j.ejmp.2023.102601
- [10] Simhony Y, Segal A, Amrani O and Etzion E 2025 High-end space electronics: Active shielding to mitigate catastrophic single-event effects *Aerosp. Sci. Technol.* **169** 111344 doi:10.1016/j.ast.2025.111344
- [11] Negm H and Ohgaki H 2024 Validation of GEANT4 simulation for active materials interrogation using nuclear resonance fluorescence *J. Radiat. Res. Appl. Sci.* **17** 101209 doi:10.1016/j.jrras.2024.101209
- [12] MacFadden N, Peggs S and Gulliford C 2018 Development and Validation of a Geant4 Radiation Shielding Simulation Framework doi:10.2172/1515417

- [13] Baydin A G, Pearlmutter B A, Radul A A and Siskind J M 2018 Automatic differentiation in machine learning: a survey *J. Mach. Learn. Res.* **18** 1–43 doi:10.48550/arXiv.1502.05767
- [14] Li T M, Aittala M, Durand F and Lehtinen J 2018 Differentiable Monte Carlo ray tracing through edge sampling *ACM Trans. Graphics* **37** 1–12 doi:10.1145/3272127.3275109
- [15] Nimier-David M, Vicini D, Zeltner T and Jakob W 2019 Mitsuba 2: A Retargetable Forward and Inverse Renderer *ACM Trans. Graphics* **38** 1–17 doi:10.1145/3355089.3356498
- [16] Hu Y, Anderson L, Li T M, Sun Q, Carr N, Ragan-Kelley J and Durand F 2020 DiffTaichi: Differentiable Programming for Physical Simulation *ICLR 2020* doi:10.48550/arXiv.1910.00935
- [17] Newbury R, Collins J, He K, Pan J, Posner I, Howard D and Cosgun A 2024 A Review of Differentiable Simulators *IEEE Access* **12** 97581–97604 doi:10.1109/ACCESS.2024.3425448
- [18] Chen R T Q, Rubanova Y, Bettencourt J and Duvenaud D K 2018 Neural Ordinary Differential Equations *Advances in Neural Information Processing Systems* vol 31 doi:10.48550/arXiv.1806.07366
- [19] Aehle M, Novák M, Vassilev V, Gauger N R, Heinrich L, Kagan M and Lange D 2024 Optimization Using Pathwise Algorithmic Derivatives of Electromagnetic Shower Simulations *Comput. Phys. Commun.* **309** 109491 doi:10.1016/j.cpc.2024.109491
- [20] Novák M 2023 HepEmShow: a compact EM shower simulation application documentation, accessed 2026-01-20 URL <https://hepemshow.readthedocs.io/en/latest/index.html>
- [21] Novák M, Hahnfeld J and Morgan B 2022 g4hepem: The G4HepEm R&D Project gitHub repository, accessed 2026-01-20 URL <https://github.com/mnovak42/g4hepem>
- [22] Fabjan C W and Gianotti F 2003 Calorimetry for particle physics *Rev. Mod. Phys.* **75** 1243–1286 doi:10.1103/RevModPhys.75.1243
- [23] Atwood W B *et al.* 2009 The Large Area Telescope on the Fermi Gamma-Ray Space Telescope Mission *Astrophys. J.* **697** 1071–1102 doi:10.1088/0004-637X/697/2/1071
- [24] Seuntjens J and Duane S 2009 Photon absorbed dose standards *Metrologia* **46** S39–S58 doi:10.1088/0026-1394/46/2/S04
- [25] Aehle M, Nguyen X T, Novák M, Dorigo T, Gauger N R, Kieseler J, Klute M and Vassilev V 2024 Efficient Forward-Mode Algorithmic Derivatives of Geant4 doi:10.48550/arXiv.2407.02966
- [26] Kagan M and Heinrich L 2023 Branches of a Tree: Taking Derivatives of Programs with Discrete and Branching Randomness in High Energy Physics doi:10.48550/arXiv.2308.16680
- [27] Cheong S, Frisch J C, Gasiorowski S, Hogan J M, Kagan M, Safdari M, Schwartzman A and Vandegar M 2022 Novel light field imaging device with enhanced light collection for cold atom clouds *J. Instrum.* **17** P08021 doi:10.1088/1748-0221/17/08/p08021
- [28] Strong G C *et al.* 2024 TomOpt: differential optimisation for task- and constraint-aware design of particle detectors in the context of muon tomography *Mach. Learn.: Sci. Technol.* **5** 035002 doi:10.1088/2632-2153/ad52e7

- [29] Gasiorowski S, Chen Y, Nashed Y, Granger P, Mironov C, Tsang K V, Ratner D and Terao K 2024 Differentiable simulation of a liquid argon time projection chamber *Mach. Learn.: Sci. Technol.* **5** 025012 doi:10.1088/2632-2153/ad2cf0
- [30] Alterkait O, Jesús-Valls C, Matsumoto R, de Perio P and Terao K 2026 End-to-end Differentiable Calibration and Reconstruction for Optical Particle Detectors doi:10.48550/arXiv.2602.24129
- [31] Aehle M *et al.* 2025 Progress in end-to-end optimization of fundamental physics experimental apparatus with differentiable programming *Reviews in Physics* **13** 100120 doi:10.1016/j.revip.2025.100120
- [32] Mokhtar F, Pata J, Garcia D, Wulff E, Zhang M, Kagan M and Duarte J 2025 Fine-tuning machine-learned particle-flow reconstruction for new detector geometries in future colliders *Phys. Rev. D* **111** doi:10.1103/physrevd.111.092015
- [33] Adelmann A *et al.* 2022 New directions for surrogate models and differentiable programming for High Energy Physics detector simulation *Snowmass 2021* doi:10.48550/arXiv.2203.08806
- [34] Shirobokov S, Belavin V, Kagan M, Ustyuzhanin A and Baydin A G 2020 Black-Box Optimization with Local Generative Surrogates *Advances in Neural Information Processing Systems* URL https://proceedings.neurips.cc/paper_files/paper/2020/file/a878dbebc902328b41dbf02aa87abb58-Paper.pdf
- [35] Wozniak K A, Mulligana S, Kieseler J, Klute M, Fleuret F and Golling T 2025 End-to-End Optimal Detector Design with Mutual Information Surrogates *Mach. Learn.: Sci. Technol.* **6** 045047 doi:10.1088/2632-2153/ae1acb
- [36] Schmidt K *et al.* 2025 End-to-End Detector Optimization with Diffusion Models: A Case Study in Sampling Calorimeters *Particles* **8** 47 doi:10.3390/particles8020047
- [37] Dorigo T (MODE Collaboration) 2023 Toward the end-to-end optimization of particle physics instruments with differentiable programming *Rev. Phys.* **10** 100085 doi:10.1016/j.revip.2023.100085
- [38] Dorigo T (MODE Collaboration) 2023 Toward artificial-intelligence assisted design of experiments *Nucl. Instrum. Methods Phys. Res., Sect. A* **1047** 167873 doi:10.1016/j.nima.2022.167873
- [39] Dorigo T (MODE Collaboration) 2026 On the Codesign of Scientific Experiments and Industrial Systems doi:10.48550/arXiv.2603.26613
- [40] Cisbani E *et al.* 2020 AI-optimized detector design for the future Electron-Ion Collider: the dual-radiator RICH case *J. Instrum.* **15** P05009 doi:10.1088/1748-0221/15/05/P05009
- [41] Diefenthaler M (AID(2)E Collaboration) 2024 AI-assisted detector design for the EIC (AID(2)E) *J. Instrum.* **19** C07001 doi:10.1088/1748-0221/19/07/C07001
- [42] Roussel R *et al.* 2024 Bayesian optimization algorithms for accelerator physics *Phys. Rev. Accel. Beams* **27**(8) 084801 doi:10.1103/PhysRevAccelBeams.27.084801
- [43] Fanelli C *et al.* 2023 AI-assisted optimization of the ECCE tracking system at the Electron Ion Collider *Nucl. Instrum. Methods Phys. Res., Sect. A* **1047** 167748 doi:10.1016/j.nima.2022.167748

- [44] Qasim S R, Owen P and Serra N 2025 Physics Instrument Design with Reinforcement Learning *Mach. Learn.: Sci. Technol.* 035033 doi:10.1088/2632-2153/adf7ff
- [45] Kortus T, Keidel R, Gauger N R and Kieseler J (Bergen pCT) 2026 Constrained collaborative optimization of charged particle tracking with multi-agent reinforcement learning *Mach. Learn.: Sci. Technol.* **7** 015021 doi:10.1088/2632-2153/ae352b
- [46] Zoccheddu S, Qasim S R, Owen P and Serra N 2026 Large Language Models for Physics Instrument Design doi:10.48550/arXiv.2601.07580
- [47] Nguyen X T, Chen L, Dorigo T, Gauger N R, Vischia P, Nardi F, Awais M, Hanif H, Abbas S and Kapoor R 2026 Differentiable Surrogate for Detector Simulation and Design with Diffusion Models doi:10.48550/arXiv.2601.07859
- [48] Hill J and Ryoo H J 2026 GRACE: an Agentic AI for Particle Physics Experiment Design and Simulation doi:10.48550/arXiv.2602.15039
- [49] SciCompKL 2024 CoDiPack (Code Differentiation Package) gitHub repository, accessed 2026-01-20 URL <https://github.com/SciCompKL/CoDiPack>
- [50] Sagebaum M, Albring T and Gauger N R 2018 Expression templates for primal value taping in the reverse mode of algorithmic differentiation *Optim. Methods Software* **33** 1207–1231 doi:10.1080/10556788.2018.1471140
- [51] M Sagebaum T Albring N G 2019 High-Performance Derivative Computations using CoDiPack *ACM Trans. Math. Software (TOMS)* **45** doi:10.1145/3356900
- [52] Novák M 2023 HepEmShow gitHub repository, accessed 2026-01-20 URL <https://github.com/mnovak42/hepemshow>
- [53] Urbán L 2006 A Model for Multiple Scattering in GEANT4 Tech. Rep. CERN-OPEN-2006-077 CERN URL <https://cds.cern.ch/record/1004190>

A Supplementary Results

f	Propagation	Stopgrad Fraction (%)		
		Energy	Steps	Tracks
0.0	Track-only	9.3	20.2	11.8
	Track + desc.	15.1	26.9	32.3
0.05	Track-only	9.6	20.4	11.9
	Track + desc.	15.5	27.2	32.8
0.1	Track-only	9.9	20.7	12.2
	Track + desc.	16.0	27.8	33.6
0.2	Track-only	10.8	21.6	12.9
	Track + desc.	17.6	29.3	36.0
0.5	Track-only	16.0	27.2	16.3
	Track + desc.	25.9	37.5	46.7

Table 1: From left to right, the columns show the percentages of energy deposited by stopgrad steps, stopgrad steps, and tracks containing a stopgrad step. *Track-only* applies stopgrad to the remainder of the affected track; *Track + desc.* additionally propagates stopgrad to all descendant particles.

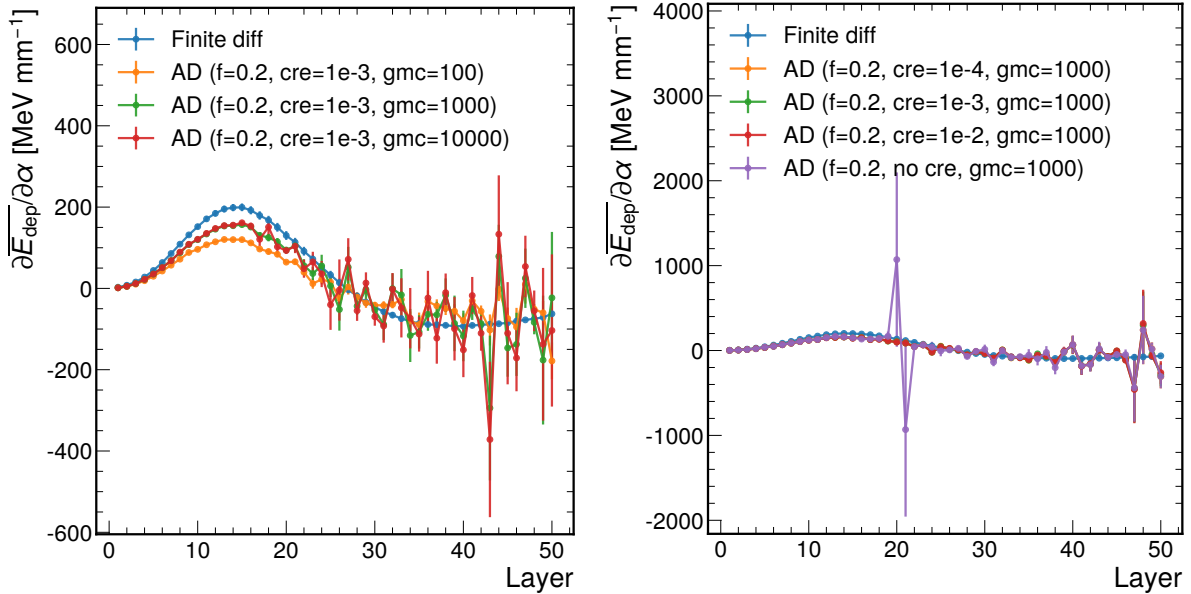


Figure 8: Mean longitudinal energy-deposit derivative per calorimeter layer comparing finite differences (blue) to AD with various settings. Left: a scan over parameter λ_{cap} . Right: a scan over parameter ϵ_{conv} .

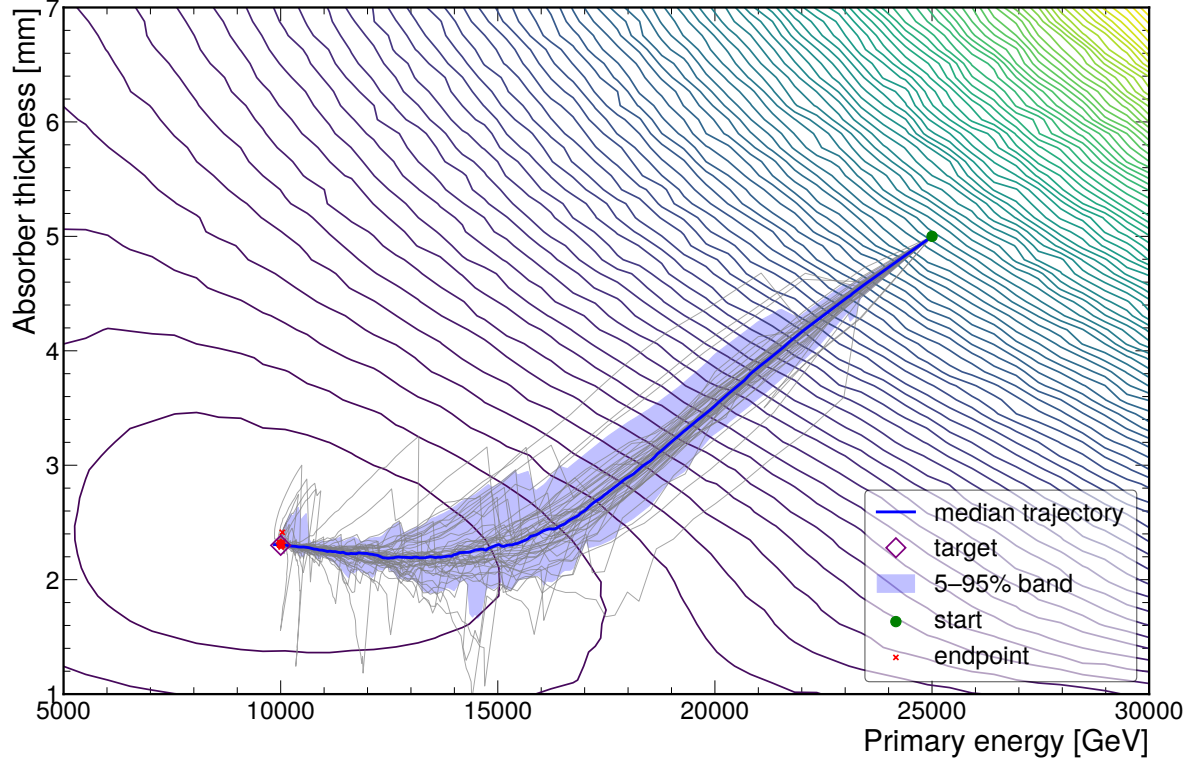


Figure 9: Trajectories of 50 randomly-seeded optimization runs minimizing the longitudinal-profile loss. Each grey curve shows one run, initialized at $(E_{\text{beam}}, \alpha) = (25000 \text{ MeV}, 5.0 \text{ mm})$ (green circle) and terminating at the final iterate (red cross). Contours show the loss landscape; the target $(E_{\text{beam}}^*, \alpha^*)$ is marked with a purple diamond. A regularization of $f = 0.05$, $\epsilon_{\text{conv}} = 10^{-3}$, and $\lambda_{\text{cap}} = 1000 \text{ mm}$ is used. The blue curve shows the median trajectory with a 5–95% band.