NOISE-DEBIASED THERMODYNAMIC VARIANCE FOR LOCAL LEARNING COEFFICIENT PROBES

Anonymous authors
Paper under double-blind review

ABSTRACT

Local learning coefficient (LLC) probes offer a singularity-aware view of neural-network training, but mean-energy methods require a local loss baseline that is ambiguous at transient checkpoints. Thermodynamic variance avoids this input; under mini-batch evaluation, however, direct variance mixes cross-state loss fluctuations with same-state noise. We operationalize this route with the *Shift-Invariant Variance Estimator* (SIVE), which estimates and subtracts the latter component using repeated evaluations. Conditional on any fixed retained path, unclipped SIVE is unbiased for noiseless path variance without requiring MCMC stationarity. The finite-scale diagnostic remains indexed by localization scale h—even a locally linear loss has tether-dependent variance—while interpretation as a Real Log Canonical Threshold (RLCT) requires additional stationary low-temperature conditions. Toy experiments recover calibrated finite-scale targets. At the primary localization scale, all five MNIST MLP trajectories exhibit a mid-training trough followed by a rebound in SIVE, while Raw Variance decreases from Epoch 40 to 100 in every trajectory. Across four localization scales, the joint early-drop/late-rise criterion is met in 19 of 20 trajectory–scale pairs. At Epoch 40, the estimated observation-noise correction accounts for 77.5% of Raw Variance. Same-state debiasing thus reveals a reproducible turning structure masked by time-varying observation noise.

1 INTRODUCTION

Training loss and accuracy summarize performance at a checkpoint but not how the loss responds to local parameter fluctuations. Such fluctuations may expose training-regime changes hidden by standard curves. In a stochastic pipeline, however, an apparent fluctuation has two sources: movement through parameter space and repeated evaluation of the same state on different mini-batches.

Singular Learning Theory (SLT) motivates local thermodynamic probes of neural-network loss landscapes through the Real Log Canonical Threshold (RLCT), commonly called a Local Learning Coefficient (LLC) in empirical work. Recent studies have used LLC trajectories to identify stagewise transformer development and refined LLCs to follow component specialization (Hoogland et al., 2025; Wang et al., 2025), making reliable longitudinal measurement a practical concern. Mean-energy probes conventionally reference the loss at a local solution; Imai's thermodynamic-variance route removes this reference and provides our starting point (Imai, 2019).

Deploying this route at transient checkpoints creates a separate measurement problem: finite localized SGLD paths are evaluated on mini-batches, so directly computed thermodynamic variance mixes loss fluctuations across visited parameter states with same-state evaluation noise. SIVE identifies and subtracts the latter through replicated evaluations, turning the inherited variance route into a scale-indexed training diagnostic. The estimator targets fixed-path noiseless variance; links to thermodynamic and RLCT limits remain scale- and sampling-dependent.

Our main contributions are:

- **Observation-noise correction:** Before optional clipping, SIVE uses same-state replicated evaluations to subtract the observation component and is conditionally unbiased for the fixed-path noiseless variance.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

- **Scale-aware interpretation:** We distinguish finite-path, finite-scale thermodynamic, and stationary LLC/RLCT targets, and make the h-dependent non-stationary slope contribution explicit.
- **Controlled validation:** Toy calibrations isolate anchoring and observation-noise errors; five neural trajectories and a four-scale audit quantify the practical effect of correction on longitudinal probing.

2 Background and Probe Setup

RLCT and Local Thermodynamic Probes. Standard statistical learning theory assumes regular models with identifiable parameters and non-degenerate local geometry. Deep neural networks are instead singular (Watanabe, 2009; Wei et al., 2023): distinct parameter values can encode the same function, producing degenerate low-loss regions. In SLT, the Real Log Canonical Threshold (RLCT) λ replaces ordinary parameter counting in the leading logarithmic complexity term of the asymptotic free energy and equivalently describes the local scaling of low-loss volume. Empirical local analyses commonly call the corresponding local quantity a Local Learning Coefficient (LLC). Computational studies probe it by simulating a localized tempered distribution, often with Stochastic Gradient Langevin Dynamics (SGLD) (Welling & Teh, 2011; Lau et al., 2023). We reserve λ for the stationary low-temperature target. Because neural-network training is a non-equilibrium process (Mandt et al., 2017; Chaudhari & Soatto, 2018) and a mid-training checkpoint need not be a local minimizer, the quantities measured along training are first treated as finite-path or finite-scale thermodynamic observables.

Mean-Energy Route and Anchoring. For a local chart defined by an energy $K(\theta) = L(\theta) - L^\star$ above a local optimum $L^\star$, the leading-order statistical-mechanical identity yields the mean-energy estimator

$$\hat{\lambda}_{\mathrm{mean}} \approx t\,\mathbb{E}_q[K(\theta)] = t\,\mathbb{E}_q[L(\theta) - L^\star], \tag{1}$$

where t is the chosen inverse-temperature scale (often written $t = n\beta$) and q is the localized tempered posterior.

At a reference local minimum θ_{t_k}, empirical LLC estimation uses the loss of the reference parameter (Lau et al., 2023), giving the checkpoint-anchored response

$$\lambda_{\mathrm{ckpt}}^{\mathrm{mean}} = t\left(\mathbb{E}_q[L(\theta)] - L(\theta_{t_k})\right). \tag{2}$$

This quantity is directly computable at any checkpoint, but away from a local minimum it is a finite-scale response rather than an LLC estimate: a localized path may have lower mean loss than its center. The neural experiment reports this conventional checkpoint reference together with path-minimum and final-checkpoint reference choices. These three curves share the same mean-energy statistic and differ only in the reference loss; they expose anchoring sensitivity rather than constituting three independent thermodynamic identities.

Thermodynamic-Variance Route and Floor Independence. Imai's variance route removes the reference value rather than estimating it more accurately. Let $\ell_n(\theta) = \log p(X^n \mid \theta)$ denote the total log likelihood and let

$$\pi_\beta(\theta) \propto \exp\{\beta\ell_n(\theta)\}\varphi(\theta) \tag{3}$$

be its power posterior. Its log normalizer

$$A_n(\beta) = \log \int \exp\{\beta\ell_n(\theta)\}\varphi(\theta)\,d\theta \tag{4}$$

satisfies the standard thermodynamic identities

$$A_n'(\beta) = \mathbb{E}_{\pi_\beta}[\ell_n(\theta)], \qquad A_n''(\beta) = \mathrm{Var}_{\pi_\beta}[\ell_n(\theta)]. \tag{5}$$

Thus variance is the second thermodynamic response of the same log normalizer, rather than an ad hoc replacement for mean energy. For $\beta = c/\log n$, Imai (2019) proposed

$$\widehat{\lambda}_{\mathbb{V}} = \beta^2\,\mathrm{Var}_{\pi_\beta}[\ell_n(\theta)] \tag{6}$$

and proved consistency for the RLCT under the corresponding SLT assumptions. With the average negative log likelihood $L_n(\theta) = -\ell_n(\theta)/n$ and $t = n\beta$, Eq. (6) is exactly

$$\widehat{\lambda}_{\mathbb{V}} = t^2 \operatorname{Var}_{\pi_\beta}[L_n(\theta)]. \tag{7}$$

This translation from Imai's notation to $t^2 \operatorname{Var}(L_n)$ is exact; the asymptotic SLT assumptions enter only when the quantity is identified with λ. Because variance is invariant to additive shifts, the route requires neither $L^\star$ nor an estimate of it. Its thermodynamic origin, RLCT consistency under Imai's assumptions, and floor independence make it attractive for transient local probing. SIVE adopts and preserves these inherited properties; its new contribution concerns the stochastic observation layer introduced in deployment.

Localized Checkpoint Probe and Dual Floor Elimination. Let $L(\theta)$ denote the reference loss and let θ_{t_k} be a training checkpoint, not necessarily a local minimizer. We define a local chart with the Gaussian tether

$$\phi_k(\theta) \propto \exp\left\{-\frac{\|\theta - \theta_{t_k}\|^2}{2h^2}\right\}. \tag{8}$$

For an intended local region U_k, write

$$L_k^\star = \inf_{\theta \in U_k} L(\theta), \qquad K_k(\theta) = L(\theta) - L_k^\star. \tag{9}$$

At inverse-temperature scale t, the normalized localized law satisfies

$$\begin{aligned} q_{k,t,h}(\theta) &= \frac{\exp\{-t(L(\theta) - L_k^\star)\}\phi_k(\theta)}{\int \exp\{-t(L(\theta') - L_k^\star)\}\phi_k(\theta')\, d\theta'} \\ &= \frac{\exp\{-tL(\theta)\}\phi_k(\theta)}{\int \exp\{-tL(\theta')\}\phi_k(\theta')\, d\theta'}. \end{aligned} \tag{10}$$

Thus the unknown floor cancels from the sampling law, and the localized SGLD drift can use the ordinary loss gradient. It cancels independently from the variance measurement:

$$\operatorname{Var}_{q_{k,t,h}}[K_k(\theta)] = \operatorname{Var}_{q_{k,t,h}}[L(\theta) - L_k^\star] = \operatorname{Var}_{q_{k,t,h}}[L(\theta)]. \tag{11}$$

The floor is therefore required neither to run the localized chain nor to compute thermodynamic variance. The Gaussian tether discourages an off-equilibrium probe from drifting away from its checkpoint, but it is a soft localization with full support; identifying its low-temperature limit with a local RLCT additionally requires concentration in the intended region. Its radius h remains part of the finite-scale measurement.

Stochastic Deployment Gap. Imai's estimator is formulated using the full-data log likelihood, which is deterministic conditional on θ. A practical stochastic neural-network probe instead observes mini-batch losses. For θ distributed according to the probe law, write $\widehat{L}(\theta, \xi) = L(\theta) + \epsilon(\theta, \xi)$ with $\mathbb{E}[\epsilon \mid \theta] = 0$. The population identity is

$$\operatorname{Var}[\widehat{L}(\theta, \xi)] = \operatorname{Var}[L(\theta)] + \mathbb{E}[\operatorname{Var}(\epsilon \mid \theta)] \tag{12}$$

This identity shows that additive-floor independence does not remove random within-state variation. Subtracting a within-group variance estimate from the between-group variation is the equal-replication form of a classical ANOVA variance-component construction (Searle et al., 1992); the same form appears in nested simulation estimators of $\operatorname{Var}(\mathbb{E}[X \mid Z])$ (Zouaoui & Wilson, 2003; Sun et al., 2011). Here we apply this standard construction conditionally to a fixed retained SGLD path, targeting its realized noiseless sample variance rather than an outer-population variance. Applying thermodynamic variance directly to stochastic group means mixes parameter-state fluctuations with evaluation noise, both amplified by t^2. Averaging independently generated datasets may reduce estimator variance but does not identify this same-state component for a given deployment. SIVE retains the variance route's independence from an unknown loss floor and uses same-state replicated evaluations to identify and subtract the observation component. The next section states the resulting finite-path target and separates this estimation guarantee from sampler convergence and the stronger stationary LLC/RLCT interpretation.

3 SIVE: FINITE-PATH ESTIMATION

3.1 FINITE-PATH TARGET AND SCALE

We distinguish the computed estimator introduced below, its fixed-path target, a finite-scale thermodynamic observable, and the stationary RLCT limit. Let $\mathcal{W} \subseteq \{1, \dots, M_{\mathrm{tot}}\}$ be a retention rule fixed before inspecting the probe losses (for example, a post-burn-in tail), and write $M_{\mathcal{W}} = |\mathcal{W}| > 1$. For values x_m indexed by $\mathcal{W}$, let $S^2_{\mathcal{W}}(x)$ denote their sample variance with denominator $M_{\mathcal{W}} - 1$. Given the retained parameter path $\theta_{\mathcal{W}}$, define the noiseless path fluctuation

$$\Lambda^{\mathrm{path}}_{\mathcal{W}} = t^2 S^2_{\mathcal{W}}\big(L(\theta_m) : m \in \mathcal{W}\big). \tag{13}$$

The empirical SIVE statistic $\widehat{\Lambda}_{\mathrm{SIVE}}$ estimates this finite-path target from noisy evaluations. Under mixing, a growing retained path, and vanishing sampler-discretization error, the path target approaches the finite-scale thermodynamic observable

$$\Lambda^{\mathrm{therm}}_{k,t,h} = t^2 \operatorname{Var}_{q_{k,t,h}}[L(\theta)]. \tag{14}$$

Only when the checkpoint defines a stationary local chart, the tempered law remains concentrated in the intended local region as $t \to \infty$, and the usual low-temperature SLT assumptions hold do we invoke

$$\Lambda^{\mathrm{therm}}_{k,t,h} = \lambda_k + o(1), \qquad t \to \infty. \tag{15}$$

Our checkpoint-level guarantee therefore concerns Eq. (13); the subsequent links require the mixing, concentration, and low-temperature conditions stated above.

Floor cancellation does not imply slope invariance. Let $x = \theta - \theta_{t_k}$ and consider the exactly linear local loss $L(\theta_{t_k} + x) = L(\theta_{t_k}) + g_k^\top x$, whose Hessian is zero. Under the Gaussian tether,

$$q_{k,t,h}(x) \propto \exp\left\{-t g_k^\top x - \frac{\|x\|^2}{2h^2}\right\}, \qquad x \sim \mathcal{N}(-th^2 g_k, h^2 I), \tag{16}$$

and hence

$$t\left(\mathbb{E}_{q_{k,t,h}}[L] - L(\theta_{t_k})\right) = -t^2 h^2 \|g_k\|^2, \tag{17}$$

$$t^2 \operatorname{Var}_{q_{k,t,h}}[L] = t^2 h^2 \|g_k\|^2. \tag{18}$$

Thus the same curvature-free non-stationary slope makes the checkpoint-anchored mean response negative and the variance response positive, both at a scale set by h. The localization radius is part of what the finite probe measures, not merely a numerical stabilizer. Appendix D.1 gives the corresponding curvature–slope decomposition.

3.2 OBSERVATION-NOISE DEBIASING

The thermodynamic variance target is formulated using the full-data log likelihood. In a stochastic deep-learning implementation, replacing it by mini-batch evaluations introduces a separate observation layer. We operationally decouple that layer from the SGLD parameter dynamics. Let B_{grad} denote the mini-batch size used to compute stochastic gradients for the Langevin drift, and B_{eval} denote the mini-batch size used to evaluate the loss at a given state.

For a single evaluation mini-batch of size B_{eval} at state θ_m, define the noisy observation as

$$\hat{L}_{m,i} = L(\theta_m) + \varepsilon_{m,i}, \qquad \mathbb{E}[\varepsilon_{m,i} \mid \theta_m] = 0, \quad \operatorname{Var}(\varepsilon_{m,i} \mid \theta_m) = \sigma^2_{\mathrm{eval}}(\theta_m, B_{\mathrm{eval}}). \tag{19}$$

At the population level, the Law of Total Variance gives

$$\operatorname{Var}_{q,\mathcal{B}}[\hat{L}_{m,i}] = \operatorname{Var}_q[L(\theta)] + \mathbb{E}_q[\sigma^2_{\mathrm{eval}}(\theta, B_{\mathrm{eval}})]. \tag{20}$$

By drawing N conditionally independent evaluation mini-batches at the same state θ_m, we compute the empirical mean over the evaluation group:

$$\bar{L}_m = \frac{1}{N} \sum_{i=1}^{N} \hat{L}_{m,i}. \tag{21}$$

Because the evaluations are conditionally independent given θ_m, the observation variance is reduced by a factor of N, yielding:

$$\mathrm{Var}_{q,\mathcal{B}}[\bar{L}_m] = \mathrm{Var}_q[L(\theta)] + \mathbb{E}_q\left[\frac{\sigma^2_{\mathrm{eval}}(\theta, B_{\mathrm{eval}})}{N}\right]. \tag{22}$$

Relying on the uncorrected empirical variance of $\bar{L}_m$ would still scale the residual observation noise by t^2. We debias this by computing the sample variance within each group to estimate the state-dependent noise profile:

$$s_m^2 = \frac{1}{N-1}\sum_{i=1}^{N}\left(\hat{L}_{m,i} - \bar{L}_m\right)^2. \tag{23}$$

Specializing this between/within decomposition to the retained path gives the estimator's primary guarantee.

Theorem 1 (Finite-path conditional unbiasedness)**.** *Let $M_{\mathcal{W}}, N > 1$. Conditional on a fixed retained path $\theta_{\mathcal{W}}$, suppose the evaluation errors $\{\varepsilon_{m,i} : m \in \mathcal{W}, 1 \le i \le N\}$ are independent, centered, and may have state-dependent variances σ_m^2. The evaluation draws must be fresh and independent of the transition randomness: neither they nor any functions of them are used to generate the parameter transitions. Then*

$$\mathbb{E}\left[S^2_{\mathcal{W}}(\bar{L}) - \frac{1}{M_{\mathcal{W}}}\sum_{m\in\mathcal{W}}\frac{s_m^2}{N}\,\middle|\,\theta_{\mathcal{W}}\right] = S^2_{\mathcal{W}}\big(L(\theta_m) : m \in \mathcal{W}\big). \tag{24}$$

Proof. Conditional on the path, $\mathrm{Var}(\bar{L}_m \mid \theta_m) = \sigma_m^2/N$. Direct expansion of the heteroscedastic sample variance gives $\mathbb{E}[S^2_{\mathcal{W}}(\bar{L}) \mid \theta_{\mathcal{W}}] = S^2_{\mathcal{W}}(L) + (M_{\mathcal{W}})^{-1}\sum_{m\in\mathcal{W}}\sigma_m^2/N$. Because $\mathbb{E}[s_m^2 \mid \theta_m] = \sigma_m^2$, the subtraction cancels the second term exactly. □

Accordingly, SIVE is the *unclipped*, observation-noise-debiased statistic

$$\widehat{\Lambda}^{\mathrm{unclip}}_{\mathrm{SIVE}} = t^2\left(S^2_{\mathcal{W}}(\bar{L}) - \frac{1}{M_{\mathcal{W}}}\sum_{m\in\mathcal{W}}\frac{s_m^2}{N}\right), \tag{25}$$

which targets Eq. (13). Regardless of whether the retained path is assigned an equilibrium or LLC interpretation, Theorem 1 identifies its noiseless cross-state loss variance from stochastic evaluations. SLT enters only in the subsequent relation between that path target, the finite-scale thermodynamic observable, and the asymptotic RLCT. For numerical reporting one may optionally use

$$\widehat{\Lambda}^{\mathrm{clip}}_{\mathrm{SIVE}} = \left(\widehat{\Lambda}^{\mathrm{unclip}}_{\mathrm{SIVE}}\right)_+. \tag{26}$$

Clipping enforces nonnegativity but introduces an upward finite-sample bias; the distinction is material when N is small. The estimator requires N independent evaluation mini-batches only at the retained states in $\mathcal{W}$, creating a direct precision–computation tradeoff. The reported implementation records these groups at every SGLD step and applies burn-in afterward; this uniform recording scheme is a computationally more expensive superset of the estimator's requirement.

The debiasing correction above acts only on the observation layer: it removes the conditional variance introduced by stochastic loss evaluations given the visited SGLD states. It does not correct bias in the sampling distribution itself, such as finite-step discretization error or stochastic-gradient-induced perturbations of the Langevin drift. In the deep-learning experiments, we therefore interpret SIVE as a stochastic-loss-compatible fixed-step local probe rather than an asymptotically exact posterior sampler.

4 Experiments

We evaluate the methodology in two stages: first, on analytically tractable toy models that isolate estimator error mechanisms under controlled conditions; second, by tracking deep-network SGD trajectories in a realistic off-equilibrium setting.

Algorithm 1 SIVE: Observation-Noise Debiasing on a Localized Finite Path

Input: Checkpoint θ_{t_k}; inverse-temperature scale t; localization radius h; SGLD step size η; total SGLD steps M_{tot}; prespecified retained set $\mathcal{W}$; evaluation group size N; evaluation mini-batch size B_{eval}; gradient mini-batch size B_{grad}
Initialize $\theta^{(0)} = \theta_{t_k}$
for MCMC step $m = 0$ to $M_{\mathrm{tot}} - 1$ **do**
  if $m \in \mathcal{W}$ **then**
    Draw N independent evaluation mini-batches $\mathcal{B}^{\mathrm{eval}}_{m,1}, \ldots, \mathcal{B}^{\mathrm{eval}}_{m,N}$, each of size B_{eval}, and compute $\hat{L}_{m,i} = \hat{L}_{\mathcal{B}^{\mathrm{eval}}_{m,i}}(\theta^{(m)})$
    Store $\bar{L}_m = N^{-1} \sum_{i=1}^{N} \hat{L}_{m,i}$ and $s_m^2 = (N-1)^{-1} \sum_{i=1}^{N} (\hat{L}_{m,i} - \bar{L}_m)^2$
  end if
  Compute stochastic gradient $g_m = \nabla \hat{L}(\theta^{(m)})$ using a mini-batch of size B_{grad} from a random stream independent of evaluation
  Draw $\xi_m \sim \mathcal{N}(0, I)$ from a third independent random stream
  Update $\theta^{(m+1)} = \theta^{(m)} - \eta t g_m - \eta(\theta^{(m)} - \theta_{t_k})/h^2 + \sqrt{2\eta}\xi_m$
end for
Compute $\widehat{\Lambda}^{\mathrm{unclip}}_{\mathrm{SIVE}}$ from the stored groups using Eq. (25); optionally report its positive part separately
Output: Finite-path SIVE estimate and optional clipped diagnostic

4.1 Controlled Validation on Toy Models

4.1.1 Experimental Setup

We consider two analytically tractable local energy models:

$$\text{Multiplicity } m = 1: \qquad K_1(u) = u^2, \tag{27}$$

$$\text{Multiplicity } m = 2: \qquad K_2(u, v) = u^2 v^2. \tag{28}$$

For both, $\min_\theta K(\theta) = 0$. The observed empirical loss is deliberately shifted by a hidden constant basin floor L_0 and corrupted by independent evaluation noise $\epsilon \sim \mathcal{N}(0, \sigma^2)$:

$$\hat{L}(\theta) = L_0 + K(\theta) + \epsilon. \tag{29}$$

We write $\theta = u$ for K_1 and $\theta = (u, v)$ for K_2, and use $K(\theta)$ to denote either energy when no distinction is needed. This formulation cleanly decouples three components: the true geometric signal $K(\theta)$, the unknown basin floor $L_0 = L^\star$, and the stochastic observation noise ϵ. Both models share an asymptotic continuous-time RLCT of $\lambda = 1/2$, though the $m = 2$ singularity introduces finite-temperature corrections that shift the practical target below $1/2$ at $t = 10^4$ (see Appendix A).

We simulate localized SGLD with inverse-temperature scale $t = 10^4$ and localization radius $h = 2$. To strictly isolate evaluation noise from dynamic optimization noise, SGLD transitions are driven by exact analytical gradients rather than stochastic mini-batches. At each sampled state θ_m, we draw $N = 64$ independent noisy observations $\hat{L}_{m,i}$ to compute the group mean $\bar{L}_m$ and within-group sample variance s_m^2. The SGLD learning rate is parameterized as $\eta = \texttt{base_lr}/t$ with $\texttt{base_lr} = 0.05$. Full parameter explorations are detailed in Appendix E.1.

4.1.2 Estimator Configurations

We compare four distinct estimator configurations to isolate the impacts of minimization bias and evaluation noise:

i. **Oracle Mean-based:** Requires the exact analytical basin floor $L^\star = L_0$ and evaluates strictly on the *noiseless* loss contour. This privileged end-to-end reference is unavailable in realistic off-equilibrium settings.

ii. **Naive Mean-based:** Does not require the true $L^\star$, instead substituting it with the empirical minimum along the noisy trajectory, $\hat{L}^\star = \min_m \hat{L}(\theta_m)$. It processes noisy loss evaluations without any noise removal, providing a deliberately implementable online anchoring baseline.

iii. **Raw Variance-based:** Implements the thermodynamic-variance estimator in Eq. (6) directly on noisy grouped losses. It bypasses $L^\star$ but contains the mini-batch evaluation variance and therefore represents the uncorrected stochastic deployment.

iv. **SIVE:** Retains the shift invariance of Raw Variance while subtracting the averaged within-group observation-noise profile in Eq. (25). The reported toy tables apply the optional positive-part safeguard; at small N this distinction creates an upward clipping bias.

The DNN runner also emits the uncorrected Raw Variance from the same retained histories at no additional sampling cost. Section 4.2 uses it to audit the practical effect of the within-state noise correction.

4.1.3 Results and Analysis

Table 1: Finite-scale probe estimates on controlled noisy toy models. SIVE denotes the noise-debiased variance estimator with the positive-part numerical safeguard used in the reported runs.

Estimator Configuration	$m = 1$: $K(u) = u^2$	$m = 2$: $K(u, v) = u^2 v^2$
Oracle Mean-based	0.5275 ± 0.0061	0.3980 ± 0.0424
Naive Mean-based	11.0155 ± 0.4489	10.9775 ± 0.3132
Raw Variance-based	6.8263 ± 0.0336	6.6818 ± 0.0919
SIVE	$\mathbf{0.5755 \pm 0.0355}$	$\mathbf{0.4310 \pm 0.0961}$

Table 1 reports the scaled estimates averaged over five independent trials. Both models have asymptotic RLCT $\lambda = 1/2$, but their finite-scale observables differ from this limit. For the regular quadratic model, the finite-step ULA references are approximately 0.5263 for the mean-energy observable and 0.5540 for the variance observable. For the singular model with the actual Gaussian localization $(t, h) = (10^4, 2)$ and chart center $(0.2, 0.2)$, numerical quadrature gives continuous-time targets 0.4296 for the mean observable and 0.4246 for the variance observable. The remaining finite-step target is not available in closed form for the singular chain. Appendix A and Appendix B separate these calibration levels.

Evaluating against these baselines, the empirical failure modes are consistent with our theoretical error decomposition, fully exposing the vulnerabilities of uncorrected estimators:

- **Minimization Bias (Naive Mean):** The naive mean estimator overestimates the geometric signal (yielding ≈ 11.0). Because $\hat{L}^\star$ is the minimum over M noisy evaluations along the trajectory, negative order-statistics inherently drive $\hat{L}^\star$ below the true $L^\star$. Even a microscopic downward shift $\Delta = L^\star - \hat{L}^\star \approx 10^{-3}$ induces an artificial error term $t\Delta \approx 10.0$ at $t = 10^4$, which entirely dominates the geometric signal.

- **Noise Amplification (Raw Variance):** Raw Variance absorbs additive evaluation noise by the Law of Total Variance. Scaled by t^2, the theoretical noise injection in our setting is $t^2\sigma^2/N = 10^8 \times (0.002)^2/64 = 6.25$. This explains why the uncorrected estimator outputs ≈ 6.7 despite its valid thermodynamic origin.

- **Signal Recovery (SIVE):** SIVE bypasses the anchored-mean failure through the inherited variance route and corrects its stochastic observation layer. Since s_m^2 estimates σ^2, the averaged correction $M^{-1}\sum_{m=1}^{M} t^2 s_m^2/N$ targets the same artifact $t^2\sigma^2/N = 6.25$. Subtracting it returns the two estimates to the finite-scale geometric regime without estimating $L^\star$. In particular, the singular-model estimate 0.4310 ± 0.0961 is close to the Gaussian-localized continuous-time variance target 0.4246; finite-step and finite-chain effects remain part of the sampler error.

A full derivation of the minimization-bias and noise-amplification mechanisms is provided in Appendix C.

4.2 Tracking SGD Dynamics in Deep Neural Networks

4.2.1 Experimental Setup

We train five independently initialized 3-layer Multi-Layer Perceptrons (MLPs) with ReLU activations on the MNIST image classification dataset using Stochastic Gradient Descent (SGD), with training seeds $1, \ldots, 5$. We save the initialization and every subsequent checkpoint through Epoch 100. At each checkpoint θ_{t_k}, we initialize five independent localized SGLD probe trials from the saved parameters and report the resulting finite-scale local fluctuation statistics. The main trajectory uses relative localization multiplier $c_h = 1$; a paired audit repeats Epochs $\{1, 40, 100\}$ at $c_h \in \{0.5, 1, 2, 4\}$. Implementation details are given in Appendix F.

4.2.2 Probe Configurations

The five plotted probes form two families. Let $\bar{L}_{\mathcal{W}} = M_{\mathcal{W}}^{-1} \sum_{m \in \mathcal{W}} \bar{L}_m$. The three mean-energy curves share

$$\widehat{\Lambda}_{\mathrm{mean}}(b) = t(\bar{L}_{\mathcal{W}} - b) \tag{30}$$

and differ only in their reference loss:

$$b_{\mathrm{ckpt}} = L(\theta_{t_k}), \qquad b_{\mathrm{path}} = \min_{m \in \mathcal{W}} \bar{L}_m, \qquad b_{\mathrm{final}} = \min_{m \in \mathcal{W}} \bar{L}_{m,K}, \tag{31}$$

where $K = 100$ is the final checkpoint and b_{final} is held fixed across checkpoints within a probe trial. We call these the *checkpoint-reference*, *path-minimum*, and *final-reference* mean probes. The last two coincide at Epoch 100 by construction.

The variance family instead uses

$$\widehat{\Lambda}_{\mathrm{Raw}} = t^2 S^2_{\mathcal{W}}(\bar{L}), \qquad \widehat{\Lambda}_{\mathrm{SIVE}} = \widehat{\Lambda}_{\mathrm{Raw}} - \hat{C}, \qquad \hat{C} = \frac{t^2}{M_{\mathcal{W}}} \sum_{m \in \mathcal{W}} \frac{s_m^2}{N}. \tag{32}$$

Raw Variance and SIVE use the same retained paths and evaluations, so they are the direct correction comparison. The three mean-energy probes provide reference-sensitivity context, not ablations of SIVE. We report unclipped SIVE; none of the DNN estimates in this run is negative.

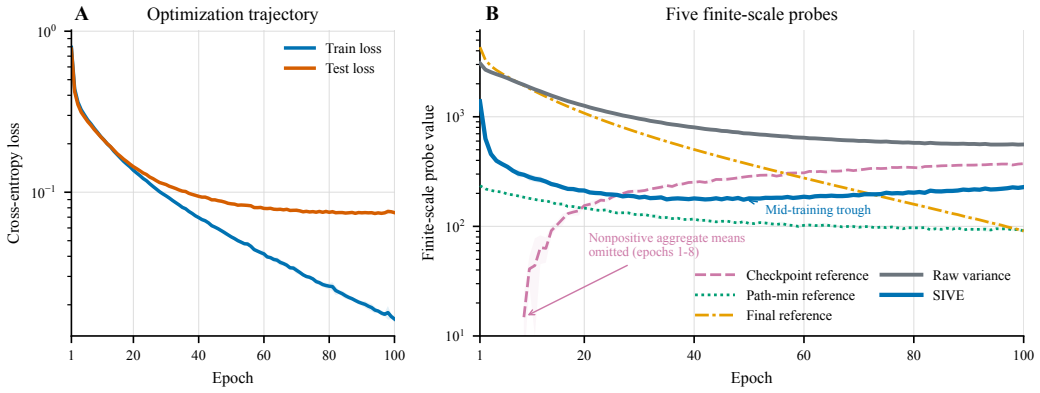


Figure 1: Five independently trained MNIST MLP trajectories, with five localized probe trials per checkpoint and $c_h = 1$. Curves and bands show the mean $\pm$ sample standard deviation across the five training-trajectory-level probe means; Epoch 0 is omitted. Panel A shows train and test loss. Panel B shows all five probes on a positive logarithmic scale and marks the mid-training SIVE trough. Nonpositive aggregate checkpoint-reference means at Epochs 1–8 are omitted rather than clipped or reflected; complete signed values are reported in Appendix G. Raw Variance and SIVE share the retained paths and evaluation groups; the three mean-energy probes differ only in reference loss but also change the statistic relative to the variance probes. Mean probes are compared for their longitudinal turning structure, not for numerical agreement with the variance probes.

4.2.3 Results and Analysis

For each checkpoint, we first average the five probes within each training trajectory; means and sample standard deviations are then computed across the five independent trajectories, while temporal contrasts are paired within trajectory.

Table 2: Five-trajectory DNN audit at the prespecified checkpoints for $c_h = 1$. Entries are mean $\pm$ sample standard deviation across five independently trained networks after first averaging five probe trials within each network.

			Mean energy: reference choice			Thermodynamic variance	
Epoch	Train Loss	Test Loss	Checkpoint	Path minimum	Final reference	Raw	SIVE
1	0.7929 ± 0.0393	0.7741 ± 0.0394	-3161.1 ± 314.7	235.0 ± 7.3	4322.5 ± 85.7	3083.6 ± 57.8	1385.0 ± 79.1
40	0.0696 ± 0.0022	0.0943 ± 0.0020	252.7 ± 8.5	116.8 ± 3.6	502.8 ± 12.8	800.9 ± 15.9	179.8 ± 3.7
100	0.0163 ± 0.0007	0.0747 ± 0.0012	373.4 ± 6.3	90.9 ± 4.1	90.9 ± 4.1	558.8 ± 15.5	228.1 ± 10.0

Observation-Noise Correction Changes the Inferred Turning Structure. All five SIVE trajectories decline to minima at Epochs 37–49 and then rebound, falling 86.1%–88.2% before rising 28.5%–39.4% by Epoch 100. Over Epochs 40–100, the paired changes are -242.2 ± 9.9 for Raw Variance, -290.5 ± 10.7 for the estimated correction, and $+48.3 \pm 8.5$ for SIVE; Raw falls and SIVE rises in every trajectory. The aggregate curves retain these opposite late directions for any intermediate checkpoint from Epoch 30 to 60 (Appendix G). The correction accounts for 55.1%, 77.5%, and 59.2% of Raw Variance at Epochs 1, 40, and 100, with the Epoch-40 trajectory-specific shares ranging only from 77.0% to 78.1%. Thus the time-varying observation component hides the direction reversal in Raw Variance. Same-state replications identify this component directly; without additional assumptions, it is not recoverable from stochastic group means alone.

Mean-Energy Reference Sensitivity. The aggregate checkpoint-reference response crosses zero between Epochs 8 and 9 and rises over both intervals, including a paired late increase of 120.7 ± 12.7. The path-minimum and final-reference responses instead decline over both intervals, with late changes of -25.9 ± 4.9 and -411.9 ± 9.0, respectively, in every trajectory. None changes direction at $c_h = 1$. These are comparisons of end-to-end diagnostic shapes rather than interchangeable estimators: the path minimum introduces an order statistic, while the final reference imports later optimization descent.

Trajectory and Scale Consistency. For $c_h \in \{0.5, 1, 2, 4\}$, the Raw/SIVE late contrasts are respectively $(-345.9, +5.5)$, $(-242.2, +48.3)$, $(-8.3, +101.1)$, and $(+76.1, +53.4)$. Together with the early SIVE drop in all 20 trajectory–scale pairs, the positive late contrast in 19 of 20 shows that the direction change is not an artifact of one aggregate curve. Linear residualization does not preserve it, limiting the interpretation to an h-indexed operational diagnostic rather than a stationary LLC or purely curvature-based quantity.

5 Conclusion and Future Work

Thermodynamic variance offers a theoretically grounded, baseline-free route for bringing LLC-inspired local analysis into the training process, but stochastic evaluation mixes cross-state fluctuations with same-state mini-batch noise. SIVE removes the identifiable observation component and is conditionally unbiased for finite-path noiseless variance, while sampler convergence and asymptotic RLCT interpretation remain separate questions. Across five MNIST MLP trajectories at the primary scale, training-dependent observation noise masked the mid-training trough and subsequent rebound revealed by SIVE; both the correction and the finite-scale signal remained h-dependent. These trajectory and scale checks concern a single MNIST MLP configuration; broader validation across models and tasks remains open. This makes SIVE a measurement tool for testing whether such stagewise changes align with delayed-generalization or grokking-like transitions (Power et al., 2022; Nanda et al., 2023), and for exploring scale-aware checkpointing, optimization, and regularization on larger models.

Reproducibility Statement. An anonymized source-code archive accompanies this submission as supplementary material. It contains the SIVE implementation, experiment configurations, experi-

ment and plotting scripts, and tests. Implementation details and the complete five-trajectory DNN results are provided in Appendices F and G, respectively.

AI USE STATEMENT

Generative AI tools were used to help develop and critically review the conceptual framing; formulate, check, and edit mathematical claims, derivations, and proofs; propose and refine hypotheses; provide feedback on experimental methodology and design; assist with software implementation and review; interpret and present results; identify and examine related literature; and draft and edit the manuscript. Generative AI was not used to generate the reported experimental data. The author reviewed all AI-assisted material, checked citations and reported numerical results against the underlying sources and experiment records, and takes responsibility for the final text, claims, code, and artifacts produced with generative-AI assistance.

REFERENCES

Pratik Chaudhari and Stefano Soatto. Stochastic gradient descent performs variational inference, converges to limit cycles for deep networks. In *2018 Information Theory and Applications Workshop*, pp. 1–10. IEEE, 2018.

Jesse Hoogland, George Wang, Matthew Farrugia-Roberts, Liam Carroll, Susan Wei, and Daniel Murfet. Loss landscape degeneracy and stagewise development in transformers, 2025. URL https://arxiv.org/abs/2402.02364. To appear in Transactions on Machine Learning Research.

Toru Imai. Estimating real log canonical thresholds, 2019. URL https://arxiv.org/abs/1906.01341.

Edmund Lau, Zach Furman, George Wang, Daniel Murfet, and Susan Wei. The local learning coefficient: A singularity-aware complexity measure, 2023. URL https://arxiv.org/abs/2308.12108.

Stephan Mandt, Matthew D. Hoffman, and David M. Blei. Stochastic gradient descent as approximate bayesian inference. *Journal of Machine Learning Research*, 18(134):1–35, 2017.

Neel Nanda, Lawrence Chan, Tom Lieberum, Jess Smith, and Jacob Steinhardt. Progress measures for grokking via mechanistic interpretability, 2023. URL https://arxiv.org/abs/2301.05217.

Alethea Power, Yuri Burda, Harri Edwards, Igor Babuschkin, and Vedant Misra. Grokking: Generalization beyond overfitting on small algorithmic datasets, 2022. URL https://arxiv.org/abs/2201.02177.

Shayle R. Searle, George Casella, and Charles E. McCulloch. *Variance Components*. Wiley, New York, 1992. doi: 10.1002/9780470316856.

Yunpeng Sun, Daniel W. Apley, and Jeremy Staum. Efficient nested simulation for estimating the variance of a conditional expectation. *Operations Research*, 59(4):998–1007, 2011. doi: 10.1287/opre.1110.0932.

George Wang, Jesse Hoogland, Stan van Wingerden, Zach Furman, and Daniel Murfet. Differentiation and specialization of attention heads via the refined local learning coefficient. In *The Thirteenth International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=SUc1UOWndp.

Sumio Watanabe. *Algebraic Geometry and Statistical Learning Theory*. Cambridge University Press, 2009.

Susan Wei, Daniel Murfet, Mingming Gong, Hui Li, Jesse Gell-Redman, and Thomas Quella. Deep learning is singular, and that's good. *IEEE Transactions on Neural Networks and Learning Systems*, 34(12), 2023.

Max Welling and Yee Whye Teh. Bayesian learning via stochastic gradient langevin dynamics. In *Proceedings of the 28th International Conference on Machine Learning*, pp. 681–688, 2011.

Faker Zouaoui and James R. Wilson. Accounting for parameter uncertainty in simulation input modeling. *IIE Transactions*, 35(9):781–792, 2003. doi: 10.1080/07408170304413.

A Free-Energy Identities and Gaussian-Localized Finite-Temperature Calibration

This appendix separates the asymptotic RLCT from the finite-temperature target induced by the actual Gaussian reference measure. The distinction is essential for the multiplicity-2 toy model, whose convergence is logarithmically slow and explicitly depends on the localization scale h.

A.1 Free-Energy Identities

Let $K(\theta) \geq 0$ define a local energy function (typically the loss gap $L(\theta) - L^\star$), and let $\phi(\theta)\, d\theta$ denote a localized reference measure. For inverse-temperature scale t, define

$$Z(t) = \int \exp\{-tK(\theta)\}\phi(\theta)\, d\theta, \tag{33}$$

and the corresponding free energy is:

$$F(t) = -\log Z(t). \tag{34}$$

The associated localized tempered posterior is:

$$q_t(\theta) = \frac{\exp\{-tK(\theta)\}\phi(\theta)}{Z(t)}. \tag{35}$$

According to the standard asymptotic expansion in Singular Learning Theory (SLT), as $t \to \infty$, the local Laplace integral is dominated by the most singular region of the parameter space:

$$Z(t) \asymp C\, t^{-\lambda} (\log t)^{m-1}, \tag{36}$$

where λ is the Real Log Canonical Threshold (RLCT) and m is the multiplicity of the singularity. Equivalently, the leading-order free energy is:

$$F_{\text{lead}}(t) = \lambda \log t - (m-1) \log\log t. \tag{37}$$

By standard statistical-mechanical identities, the posterior mean and variance of the local energy correspond exactly to the first and second derivatives of the free energy with respect to the inverse temperature:

$$\mathbb{E}_{q_t}[K(\theta)] = F'(t), \qquad \mathrm{Var}_{q_t}[K(\theta)] = -F''(t). \tag{38}$$

Thus the mean- and variance-energy observables are distinct derivatives of the same thermodynamic system. Their common asymptotic coefficient does not make the mean and variance observables identical at finite t.

A.2 Exact Centered-Gaussian Calibration for u^2v^2

For $K(u,v) = u^2v^2$ and the normalized centered Gaussian reference

$$\phi_h(u,v) = \frac{1}{2\pi h^2} \exp\left[-\frac{u^2+v^2}{2h^2}\right], \tag{39}$$

direct integration gives

$$Z_h(t) = \frac{e^z K_0(z)}{2\sqrt{\pi}\, h^2 \sqrt{t}}, \qquad z = \frac{1}{8th^4}, \tag{40}$$

where K_ν is the modified Bessel function of the second kind. Writing $r(z) = K_1(z)/K_0(z)$ and differentiating Eq. (40) yields the exact centered-reference observables

$$t\,\mathbb{E}[K] = \frac{1}{2} + z\big(1 - r(z)\big), \tag{41}$$

$$t^2\,\mathrm{Var}(K) = \frac{1}{2} + 2z - zr(z) + z^2\big(1 - r(z)^2\big). \tag{42}$$

For small z, define

$$A(t,h) = \log(16th^4) - \gamma, \tag{43}$$

where γ is Euler's constant. Then

$$t\,\mathbb{E}[K] = \frac{1}{2} - \frac{1}{A(t,h)} + o\big(A^{-1}\big)\,, \tag{44}$$

$$t^2\,\mathrm{Var}(K) = \frac{1}{2} - \frac{1}{A(t,h)} - \frac{1}{A(t,h)^2} + o\big(A^{-2}\big)\,. \tag{45}$$

The finite-temperature correction is therefore controlled by $\log(th^4)$, not by $\log t$ alone. The localization radius is part of the finite-scale estimand even though the leading RLCT remains $1/2$.

The reported toy chain is centered at $(0.2, 0.2)$ rather than exactly at the singularity. Applying deterministic quadrature to that actual shifted Gaussian reference with $(t, h) = (10^4, 2)$ gives

$$t\,\mathbb{E}[K] = 0.42957, \qquad t^2\,\mathrm{Var}(K) = 0.42461. \tag{46}$$

These are continuous-time localized targets. The finite-step singular ULA target remains distinct, as discussed next. Finally, the exact shift identity $\mathrm{Var}[K] = \mathrm{Var}[L - L^\star] = \mathrm{Var}[L]$ continues to hold for every reference measure and every finite t.

B Finite-Step ULA Bias and the Moment-Closure Barrier

The continuous-time reference values derived in Appendix A formally assume exact sampling from the target thermodynamic posterior. In practice, however, the finite-scale probes are driven by the discrete-time Unadjusted Langevin Algorithm (ULA). This appendix demonstrates how finite step sizes alter the stationary distribution, injecting a systematic discretization bias. We first separate the mathematically tractable regular case—where exact moment calculations are explicitly solvable—from the singular case, where structural non-linearities obstruct closed-form solutions.

B.1 Finite-Step Bias in Regular Quadratic Geometries

To isolate the precise mechanics of ULA discretization bias, we analyze the strictly regular quadratic toy model, $K_1(u) = u^2$. Because the gradient is linear ($\nabla K_1 = 2u$), the infinite hierarchy of moments collapses into a solvable algebraic system.

Let the dimensionless effective step size be denoted by $a = \eta t = \texttt{base_lr}$, and let c be the one-dimensional chart center. Including the same Gaussian tether used by the experiment, the discrete ULA update is the affine AR(1) process

$$u_{m+1} = r u_m + \delta c + \sqrt{2\eta}\,\xi_m, \qquad r = 1 - 2a - \delta, \qquad \delta = \frac{\eta}{h^2}, \qquad \xi_m \sim \mathcal{N}(0,1). \tag{47}$$

For $|r| < 1$, its exact stationary mean and variance are

$$\mu_u = \frac{\delta c}{1 - r}, \qquad V_u = \frac{2\eta}{1 - r^2}. \tag{48}$$

Because the stationary distribution is Gaussian, the exact discrete-time targets for the implemented tethered chain are

$$t\,\mathbb{E}_{\mathrm{ULA}}[u^2] = t(V_u + \mu_u^2), \qquad t^2\,\mathrm{Var}_{\mathrm{ULA}}(u^2) = t^2(2V_u^2 + 4\mu_u^2 V_u). \tag{49}$$

For the reported $(t, h, c) = (10^4, 2, 0.2)$, Eq. (49) rounds to the values shown below. The tether and center corrections are smaller than the displayed precision, so the familiar untethered expressions $[2(1-a)]^{-1}$ and $[2(1-a)^2]^{-1}$ give the same four-decimal references. As $a \to 0$, Eq. (49) approaches the continuous Gaussian-localized target, not an automatically untethered quantity.

Table 3 compares empirical measurements against these exact ULA targets for the regular K_1 model. We intentionally evaluate a relatively large step size ($a = 0.10$) to amplify the discretization artifact.

The quantitative agreement is consistent with the estimator tracking the distribution of the discrete samples even when finite-step Langevin dynamics widen that distribution. It does not remove the finite-step change of target.

Table 3: Finite-step calibration on the regular quadratic toy model $K_1(u) = u^2$ at $t = 10^4$. "Exact Target" is computed from the tethered affine-Gaussian chain in Eq. (49); the remaining discrepancies are finite-chain and observation-estimation variation.

$\texttt{base_lr} = a$	Oracle Mean	Exact Target	SIVE	Exact Target
0.05	0.5275 ± 0.0061	0.5263	0.5755 ± 0.0355	0.5540
0.10	0.5549 ± 0.0040	0.5555	0.6307 ± 0.0285	0.6173

B.2 The Moment-Closure Barrier in Singular Geometries

For singular geometries, the corresponding elementary moment calculation does not close. To display the obstruction while retaining localization, consider a centered Gaussian tether and the u update for $K_2(u, v) = u^2v^2$:

$$u_{m+1} = r_0 u_m - 2\eta t u_m v_m^2 + \sqrt{2\eta}\xi_m, \qquad r_0 = 1 - \frac{\eta}{h^2}. \tag{50}$$

Squaring and taking expectations at stationarity gives

$$0 = (r_0^2 - 1)\mathbb{E}[u^2] - 4r_0\eta t\mathbb{E}[u^2v^2] + 4\eta^2 t^2\mathbb{E}[u^2v^4] + 2\eta. \tag{51}$$

Equivalently,

$$r_0 t\mathbb{E}[u^2v^2] = \frac{1}{2} + \eta t^2\mathbb{E}[u^2v^4] - \frac{1 - r_0^2}{4\eta}\mathbb{E}[u^2]. \tag{52}$$

In the continuous-step limit, Eq. (52) retains the finite-h identity $t\mathbb{E}[u^2v^2] = \frac{1}{2} - \mathbb{E}[u^2]/(2h^2)$; only the subsequent low-temperature asymptotics approach the RLCT $1/2$. At finite step size, the target is additionally coupled to $\mathbb{E}[u^2v^4]$. An off-center tether, as used in the reported experiment, introduces further first and mixed moments rather than restoring closure.

Consequently, this low-order moment method does not yield a closed-form finite-step target for the singular chain. We therefore report controlled step-size sensitivity in Appendix E and treat $\texttt{base_lr} = 0.05$ as the selected finite-run operating point rather than an exact calibration.

C Finite-Path Error Decomposition for Anchored Mean Estimators

This appendix gives an exact decomposition for a mean statistic anchored to the lowest noisy loss on the same finite path. It separates incomplete path coverage from an order-statistics effect without assigning either a universal finite-realization sign.

C.1 General Error Decomposition

Write the grouped loss at step m as

$$\bar{L}_m = L_k^\star + K_m + \bar{\varepsilon}_m, \tag{53}$$

where $K_m = L(\theta_m) - L_k^\star \geq 0$. Define

$$\min_m \bar{L}_m = L_k^\star + g_M - \nu_M, \qquad g_M = \min_m K_m, \qquad \nu_M = g_M - \min_m(K_m + \bar{\varepsilon}_m). \tag{54}$$

The realized ν_M is signed. Nevertheless, because the minimum is a concave functional of the centered evaluation noise,

$$\mathbb{E}[\nu_M \mid K_{1:M}] \geq 0. \tag{55}$$

Substitution gives the exact finite-path identity

$$\begin{aligned} \widehat{\Lambda}_{\text{online}} &= t\left(\frac{1}{M}\sum_m \bar{L}_m - \min_m \bar{L}_m\right) \\ &= t\left(\frac{1}{M}\sum_m K_m - g_M\right) + t\nu_M + t\frac{1}{M}\sum_m \bar{\varepsilon}_m. \end{aligned} \tag{56}$$

After averaging over evaluation noise, the first term is the noiseless path thickness above its lowest visited state and the second is the expected noisy-minimum inflation. Replacing either by a stationary LLC expression requires the additional assumptions in Section 3.1.

C.2 Noise-Dominated Toy Regime

The controlled low-dimensional paths repeatedly visit states close to the exact minimum, so the order-statistics effect dominates the anchored result. Since $K_m \geq g_M$,

$$\nu_M \leq -\min_m \bar{\varepsilon}_m. \tag{57}$$

For independent Gaussian evaluation noise, the corresponding pure-noise extreme-value scale is

$$\frac{\sigma}{\sqrt{N}}\sqrt{2\log M_{\text{eff}}}. \tag{58}$$

It is an approximate upper scale, not a general equality for $\mathbb{E}[\nu_M]$, because only states sufficiently near the path minimum can realize the pure-noise extreme. With $t = 10^4$, $\sigma = 0.002$, $N = 64$, and $M_{\text{eff}} \approx 7 \times 10^4$, the scaled value is approximately 11.8, consistent in magnitude with the empirical Naive Mean result ≈ 11.0 in Table 1.

C.3 Path-Relative Anchoring in High Dimensions

In a high-dimensional non-stationary chart, a finite localized path need not reach the unknown basin floor. Anchoring to its lowest visited loss therefore reports the path-relative thickness in Eq. (56), rather than an independently calibrated thermodynamic target. The compressed path-minimum curve in Figure 1 is consistent with this interpretation. Without access to $L_k^\star$, however, neither the magnitude of g_M nor a universal direction of error relative to an intrinsic LLC can be identified from the experiment alone.

C.4 Capabilities and Limitations of SIVE

- **Inherited shift invariance:** Thermodynamic variance removes the additive baseline $L_k^\star$ and never performs a noisy minimum operation; SIVE preserves this property.
- **Conditional observation-noise debiasing:** Theorem 1 removes the expected variance contributed by conditionally independent mini-batch evaluations, including state-dependent noise levels.
- **Unaddressed path and sampler effects:** SIVE does not correct finite-step discretization, finite-chain Monte Carlo error, stochastic-gradient perturbations, or the dependence of the finite-scale estimand on (t, h).
- **No slope invariance:** At a non-stationary checkpoint, parameter fluctuations along a local slope induce loss fluctuations even though the additive floor has canceled. The exact quadratic decomposition is given next.

D Localization Radius, Chain Length, and Tail-Window Consistency

While the variance operator structurally eliminates the absolute baseline $L_k^\star$, the localized SGLD probe remains dependent on the scale of the local chart (indexed by h) and the convergence properties of the Markov chain (controlled in part by the total steps M_{tot}). This appendix separates these two roles and makes explicit how the localized posterior interacts with off-equilibrium gradients.

D.1 A Local Gaussian Tether Model on a Non-Equilibrium Slope

In practical deep learning, stochastic gradient descent operates as a non-equilibrium process; mid-training checkpoints θ_{t_k} are rarely true stationary points ($\nabla L \neq 0$). Consequently, localized MCMC chains explore gradient "slopes" rather than symmetric minima.

To model this analytically, we approximate the local excess loss by a surrogate that explicitly retains both its checkpoint offset and the non-vanishing linear drift:

$$K_k(\theta) \approx K_{k,0} + g_k^\top x + \frac{1}{2} x^\top H_k x, \qquad x = \theta - \theta_{t_k}, \tag{59}$$

where $K_{k,0} = L(\theta_{t_k}) - L_k^\star$, $g_k = \nabla L(\theta_{t_k})$, and H_k is the local Hessian. The offset $K_{k,0}$ does not affect the localized law, and it is removed exactly by a variance. The slope g_k, in contrast, changes the law and therefore remains part of the finite-scale estimand.

Under the Gaussian prior $\exp\{-\|x\|^2/(2h^2)\}$, the tethered local Gibbs law becomes:

$$q_{k,t,h}(x) \propto \exp\left\{-t\left(K_{k,0} + g_k^\top x + \frac{1}{2} x^\top H_k x\right) - \frac{1}{2h^2}\|x\|^2\right\}. \tag{60}$$

Let $A_{k,t,h} = tH_k + h^{-2} I$. If $A_{k,t,h} \succ 0$, this quadratic surrogate defines the Gaussian $\mathcal{N}(\mu_{k,t,h}, \Sigma_{k,t,h})$ with

$$\Sigma_{k,t,h} = A_{k,t,h}^{-1}, \qquad \mu_{k,t,h} = -t\Sigma_{k,t,h} g_k. \tag{61}$$

The positive-definiteness condition is essential: a Gaussian tether does not make an indefinite quadratic approximation normalizable for arbitrary h. The original nonlinear localized law may nevertheless be normalizable outside the range where this local quadratic surrogate is valid.

D.2 Curvature–Slope Decomposition of the Finite-Scale Variance

The following proposition makes explicit what shift invariance does—and does not—remove.

Proposition 1 (Exact quadratic tether decomposition)**.** *Assume H_k is symmetric and $A_{k,t,h} \succ 0$. Under the Gaussian quadratic surrogate,*

$$\begin{aligned}\mathbb{E}[K_k] = K_{k,0} - t g_k^\top \Sigma_{k,t,h} g_k + \frac{t^2}{2} g_k^\top \Sigma_{k,t,h} H_k \Sigma_{k,t,h} g_k \\ + \frac{1}{2}\operatorname{tr}(H_k \Sigma_{k,t,h}),\end{aligned} \tag{62}$$

whereas the scaled variance is

$$t^2 \operatorname{Var}(K_k) = \frac{1}{2}\operatorname{tr}\big[(tH_k\Sigma_{k,t,h})^2\big] + t^2 h^{-4} g_k^\top \Sigma_{k,t,h}^3 g_k. \tag{63}$$

For a tight tether, $\Sigma_{k,t,h} = h^2 I + \mathcal{O}(h^4)$, and therefore

$$t\mathbb{E}[K_k] = tK_{k,0} - t^2 h^2 \|g_k\|^2 + \frac{th^2}{2}\operatorname{tr}(H_k) + \mathcal{O}(h^4), \tag{64}$$

$$t^2 \operatorname{Var}(K_k) = t^2 h^2 \|g_k\|^2 + \mathcal{O}(h^4). \tag{65}$$

Proof. For $x \sim \mathcal{N}(\mu, \Sigma)$ and symmetric H,

$$\operatorname{Var}\left(g^\top x + \frac{1}{2} x^\top H x\right) = \frac{1}{2}\operatorname{tr}[(H\Sigma)^2] + (g + H\mu)^\top \Sigma (g + H\mu). \tag{66}$$

Substitution of $\mu = -t\Sigma g$ gives the stated mean. Because $I - tH_k\Sigma_{k,t,h} = h^{-2}\Sigma_{k,t,h}$, the second variance term becomes $h^{-4} g_k^\top \Sigma_{k,t,h}^3 g_k$. Finally, expanding $\Sigma_{k,t,h} = h^2(I - th^2 H_k + \mathcal{O}(h^4))$ proves the tight-tether expressions. □

The flat-slope case recovers Eq. (18): if $H_k = 0$, then $t^2 \operatorname{Var}(K_k) = t^2 h^2 \|g_k\|^2$ exactly. Thus the variance removes the unknown additive floor but does not remove non-stationary slope effects. If $g_k = 0$, only the curvature term in Eq. (63) remains. Neither case supplies a universal h-independent plateau in singular or off-equilibrium directions; h indexes the finite spatial scale being probed.

D.3 Finite Chain Length and Tail Windows

The second critical dependency is the total chain length M_{tot}. Let F_m be an observable evaluated along the SGLD chain (for the path target, $F_m = L(\theta_m)$), and let $\mu_j = \mathbb{E}_q[F^j]$ for $j \in \{1, 2\}$. Controlling a variance statistic requires control of both coordinates of the observable $G_m = (F_m, F_m^2)$; consistency of the sample mean alone is insufficient.

To mitigate the transient drift, we define a tail window that discards the initial burn-in phase:

$$\mathcal{W}_\alpha = \{\lfloor \alpha M_{\mathrm{tot}} \rfloor, \dots, M_{\mathrm{tot}} - 1\}, \qquad 0 < \alpha < 1, \tag{67}$$

and compute the empirical average:

$$\overline{F^j}_{\mathcal{W}_\alpha} = \frac{1}{|\mathcal{W}_\alpha|} \sum_{m \in \mathcal{W}_\alpha} F_m^j, \qquad j \in \{1, 2\}. \tag{68}$$

Proposition 2 (Tail mean and variance consistency)**.** *Assume F has a finite fourth moment under q, the post-burn-in process $G_m = (F_m, F_m^2)$ is geometrically mixing, and the two transient moment biases satisfy $|\mathbb{E}[F_m^j] - \mu_j| \le C_j \rho^m$ for $j \in \{1, 2\}$ and some $0 < \rho < 1$. Let $\tau_\star = \max\{\tau_{\mathrm{int}}(F), \tau_{\mathrm{int}}(F^2)\}$. Then, for $j \in \{1, 2\}$,*

$$\overline{F^j}_{\mathcal{W}_\alpha} - \mu_j = \mathcal{O}(\rho^{\alpha M_{\mathrm{tot}}}) + \mathcal{O}_p\Big(|\mathcal{W}_\alpha|_{\mathrm{eff}}^{-1/2}\Big), \tag{69}$$

where

$$|\mathcal{W}_\alpha|_{\mathrm{eff}} \asymp \frac{(1-\alpha) M_{\mathrm{tot}}}{\tau_\star}. \tag{70}$$

Consequently,

$$S^2_{\mathcal{W}_\alpha}(F) - \mathrm{Var}_q(F) = \mathcal{O}(\rho^{\alpha M_{\mathrm{tot}}}) + \mathcal{O}_p\Big(|\mathcal{W}_\alpha|_{\mathrm{eff}}^{-1/2}\Big) + \mathcal{O}(|\mathcal{W}_\alpha|^{-1}). \tag{71}$$

Proof. Apply the geometric-tail bias bound and the usual mixing variance bound separately to F and F^2; the slower coordinate determines $\tau_\star$. The Bessel-corrected sample variance obeys

$$S^2_{\mathcal{W}_\alpha}(F) = \frac{|\mathcal{W}_\alpha|}{|\mathcal{W}_\alpha| - 1} \left(\overline{F^2}_{\mathcal{W}_\alpha} - \big(\overline{F}_{\mathcal{W}_\alpha}\big)^2\right). \tag{72}$$

Substituting the two moment bounds and using $|\mathcal{W}_\alpha| = (1-\alpha) M_{\mathrm{tot}} + \mathcal{O}(1)$ gives the result. □

D.4 Separating Scale Dependence from Estimation Error

The dependence on h is part of the target, not a generic error term. For a tail-window implementation, a schematic decomposition is

$$\widehat{\Lambda}^{\mathrm{unclip}}_{\mathrm{tail}} = \Lambda^{\mathrm{therm}}_{k,t,h} + b_{\mathrm{step}} + b_{\mathrm{grad}} + \mathcal{O}(\rho^{\alpha M_{\mathrm{tot}}}) + \mathcal{O}_p\Big(|\mathcal{W}_\alpha|_{\mathrm{eff}}^{-1/2}\Big) + \varepsilon_{\mathrm{eval}}, \tag{73}$$

where b_{step} and b_{grad} denote discretization and stochastic-gradient biases. Proposition 2 supplies the displayed Monte Carlo order for the variance statistic, with the effective sample size governed by the slower of F and F^2. Conditional on a fixed retained path, Theorem 1 gives $\mathbb{E}[\varepsilon_{\mathrm{eval}} \mid \theta_{\mathcal{W}_\alpha}] = 0$. The remaining terms require sampler diagnostics or limits; no universal rate in h follows from shift invariance.

The decomposition separates three conceptually different choices:

- **Probe scale** (t, h)**:** changing either parameter generally changes $\Lambda^{\mathrm{therm}}_{k,t,h}$ itself.
- **Sampler approximation:** finite step size, stochastic gradients, burn-in, and incomplete mixing separate the implemented chain from its intended localized law.
- **Observation and Monte Carlo uncertainty:** replicated evaluation removes the conditional mean contribution of observation noise, while finite effective sample size still produces random error.

D.5 Operational Consequence

The localization radius h is therefore a *scale index*, whereas the total chain length M_{tot} is a *consistency control*. A single-h curve can be informative as an operational, end-to-end diagnostic, but a multi-h profile is needed to identify features that persist across scales and to expose tether- and slope-dependent effects. Increasing M_{tot} and examining multiple tail windows assess a different question: whether the chosen finite-scale target is being estimated stably. The following appendix reports the existing sensitivity sweeps under this interpretation.

E Finite-Step Operating Point and Scale Diagnostics

This appendix reports empirical sensitivity diagnostics for the finite-step and finite-scale effects discussed in Appendices B and D. The sweeps provide an operating point for the reported experiments; they do not turn h into a nuisance parameter or certify sampler convergence in a neural network.

E.1 Stable Operating Point and Diagnostic Protocol

To prevent confounded measurements, we anchor the diagnostics around an empirically established stable operating baseline. By freezing the system at this baseline, we sweep one structural parameter at a time across a wide numerical range, while holding the remaining parameters fixed. This protocol isolates the physical role of each component of the localized SGLD probe.

Unless otherwise specified, the controlled diagnostics are executed on the singular toy model $K_2(u, v) = u^2v^2$ with a hidden basin floor $L_0 = 100$ and independent Gaussian evaluation noise $\sigma = 0.002$. The inverse-temperature scale is fixed explicitly at $t = 10^4$, with total SGLD steps $M = 100{,}000$ and burn-in ratio $\alpha = 0.3$.

Table 4 summarizes the stable operating point used throughout the toy-model diagnostics, together with the corresponding deep-learning batch-allocation parameters. The toy diagnostics use $N = 64$ as the default evaluation-group size in order to expose the statistical mechanics of the debiasing correction under controlled noise. In the neural-network experiments, we increase the evaluation group to $N = 512$ using vectorized forward passes, as described in Appendix F.3, to further suppress the residual uncertainty of the noise correction.

Table 4: Default probe configuration. Unless otherwise stated, these values are held fixed when sweeping a single structural parameter.

Category	Quantity	Value
Toy model	Localization and N diagnostics	$K_2(u, v) = u^2v^2$
Toy model	Finite-step calibration	$K_1(u) = u^2$ and $K_2(u, v) = u^2v^2$
Toy model	Hidden basin floor L_0	100
Toy model	Evaluation noise standard deviation σ	0.002
Common	Inverse-temperature scale t	10,000
Common	Total SGLD steps M	100,000
Common	Burn-in ratio α	0.3
Common	Independent MCMC trials	5
Default toy setting	Localization radius h	2.0
Default toy setting	Base step size `base_lr`	0.05
Default toy setting	Evaluation group size N	64
DNN setting	Relative localization multiplier c_h	1.0
DNN setting	Base step size `base_lr`	0.05
DNN setting	Evaluation group size N	512
DNN setting	Evaluation mini-batch size B_{eval}	64
DNN setting	Gradient mini-batch size B_{grad}	1024

These values define the operating point of the reported study. The subsequent sweeps perturb h, `base_lr`, and N around this baseline to show their observed sensitivities.

E.2 Localization Sensitivity to h

The localization radius h should not be interpreted as a purely numerical hyperparameter. In the off-equilibrium setting, it defines the spatial scale of the local chart. If h is too small, the resulting estimate is dominated by the tight tether and its interaction with the local slope. If h is too large, the chain can move toward lower-loss regions outside the intended observation window. A checkpoint-local finite-scale diagnostic therefore requires declaring h and inspecting whether the qualitative trajectory persists across a useful range of scales.

Having fixed the remaining configuration in Table 4, we first vary the localization radius h. Table 5 is an empirical multi-scale profile of the end-to-end probe, including its finite-step sampler; it should not be read as a theorem that the continuous finite-temperature target is independent of h.

Table 5: Localization sensitivity to h on the singular toy model $K_2(u, v) = u^2v^2$. SIVE values reproduce the original end-to-end implementation, including its nonnegative numerical clipping.

h	Oracle Mean-based	SIVE
0.05	1.7305 ± 0.1120	1.7330 ± 0.1469
0.10	0.4942 ± 0.0857	0.5077 ± 0.1172
0.20	0.3264 ± 0.1011	0.3376 ± 0.1462
0.50	0.3763 ± 0.0760	0.4264 ± 0.0969
1.00	0.3866 ± 0.0576	0.4427 ± 0.0957
2.00	0.3970 ± 0.0698	0.4547 ± 0.1040
5.00	0.3944 ± 0.0665	0.4551 ± 0.1043
10.00	0.3938 ± 0.0657	0.4549 ± 0.1041
20.00	0.3938 ± 0.0658	0.4550 ± 0.1042
50.00	0.3938 ± 0.0658	0.4550 ± 0.1042

The sweep displays two useful operational regimes, but their interpretation is empirical:

- **Tight-tether sensitivity ($h \leq 0.5$):** The non-monotone values at very small h show that localization, local slope, discrete step size, and incomplete exploration interact strongly in this regime. Proposition 1 rules out assigning these changes a universal $\mathcal{O}(h^{-2})$ law.
- **Empirical stability region ($h \geq 1.0$):** Within the resolution of these five-trial runs, the implemented SIVE values remain near 0.45 from $h = 1$ to $h = 50$. This is a useful robustness observation for this sampler and toy problem. It is not proof of h-invariance or of a topological plateau; indeed, the exact continuous target in Appendix A retains a slow logarithmic dependence on h.

Operational Translation for DNNs. The toy sweep uses an absolute Gaussian width in a two-parameter chart. The DNN probe instead uses a relative multiplier c_h whose physical width depends on the checkpoint norm, as defined in Eq. (75). Thus the two numerical values called "1" are not a cross-model calibration. We declare $c_h = 1$ as the main DNN convention and report a paired scale profile rather than selecting a favorable scale after inspection.

Table 6: Paired DNN localization audit. Values and contrasts are mean $\pm$ sample standard deviation across five training-trajectory-level quantities, after first averaging five probe trials per trajectory. $D^{\mathrm{late}}_{\mathrm{Ckpt}}$, $D^{\mathrm{late}}_{\mathrm{Raw}}$, and $D^{\mathrm{late}}_{\mathrm{SIVE}}$ are the corresponding Epoch-40-to-100 contrasts; a positive late contrast is called a rebound only when preceded by the reported early drop. "SIVE Late +" counts positive SIVE late contrasts across the five trajectories. The final column is the ratio of aggregate observation-noise correction to aggregate Raw Variance at Epoch 40. No unclipped SIVE estimate was negative.

c_h	SIVE_1	SIVE_{40}	SIVE_{100}	$D^{\mathrm{drop}}_{\mathrm{SIVE}}$	$D^{\mathrm{late}}_{\mathrm{Ckpt}}$	$D^{\mathrm{late}}_{\mathrm{Raw}}$	$D^{\mathrm{late}}_{\mathrm{SIVE}}$	SIVE Late +	$\hat{C}/\mathrm{Raw}_{40}$
0.5	1169.6 ± 113.6	64.5 ± 2.6	70.0 ± 2.7	1105.1 ± 113.1	47.5 ± 10.5	-345.9 ± 11.8	5.5 ± 2.5	5/5	88.2%
1	1385.0 ± 79.1	179.8 ± 3.7	228.1 ± 10.0	1205.2 ± 80.1	120.7 ± 12.7	-242.2 ± 9.9	48.3 ± 8.5	5/5	77.5%
2	1485.4 ± 73.8	463.1 ± 15.6	564.3 ± 33.8	1022.2 ± 83.7	261.1 ± 15.6	-8.3 ± 39.0	101.1 ± 33.2	5/5	67.2%
4	1246.1 ± 64.8	840.3 ± 50.1	893.8 ± 53.1	405.7 ± 81.1	417.0 ± 22.6	76.1 ± 71.8	53.4 ± 73.8	4/5	62.5%

The SIVE early-to-middle drop is positive in every one of the 20 paired trajectory/scale comparisons. Its late rebound is positive for all trajectories at $c_h \in \{0.5, 1, 2\}$ and for four of five at $c_h = 4$, while the checkpoint-anchored late contrast is positive in all 20 comparisons; both magnitudes are plainly scale dependent. A positive checkpoint late contrast is not the same turning evidence: at the main scale, that response also rises over the early interval. The aggregate noise shares ($\hat{C}/\mathrm{Raw}$) at Epochs $(1, 40, 100)$ are $(57.0\%, 88.2\%, 65.0\%)$, $(55.1\%, 77.5\%, 59.2\%)$, $(53.2\%, 67.2\%, 59.8\%)$, and $(51.7\%, 62.5\%, 61.4\%)$ for increasing c_h. In particular, the Epoch-40 share decreases with c_h in every training trajectory. This variation is expected because changing c_h changes the localized path and therefore both terms in the decomposition; it also makes clear why the 77.5% value is a property of the declared protocol rather than an empirical constant that can be transferred to another probe. Because the fixed SGLD step also makes the tether coefficient $\eta/h_k(c_h)^2$ vary with c_h, this is an end-to-end scale profile rather than an experiment that separately identifies thermodynamic scale dependence and discretization effects. The normalized RMS displacement remains close to one physical localization width throughout the audited points, but this diagnostic is not a step-size convergence certificate.

The full-data gradient diagnostic supplies a deliberately conservative counterpoint. The residualized late contrasts are -45.4 ± 45.3, -64.8 ± 118.7, -818.3 ± 875.9, and -8998.6 ± 8404.3 for $c_h \in \{0.5, 1, 2, 4\}$, with positive signs in $0/5$, $2/5$, $1/5$, and $0/5$ trajectories, respectively. At $c_h = 1$, the average Epoch-40-to-100 change decomposes exactly along the retained path as a residual change of -64.8, a linear-path-variance change of -166.6, and a change of $+279.7$ in the $2t^2\,\mathrm{Cov}_{\mathcal{W}}(R, \ell)$ term, summing to the observed SIVE change of $+48.3$. Because the covariance is substantial and the residualized statistic is highly scale sensitive, linear detrending is a diagnostic decomposition rather than a curvature estimator. The scale profile supports the non-monotonicity of the reported finite-scale statistic while explicitly limiting a stronger stationary-geometric interpretation.

E.3 Finite-Step Calibration of `base_lr`

Because the Moment-Closure barrier (Appendix B.2) prevents calculating the exact Unadjusted Langevin Algorithm (ULA) target for singular geometries, we empirically ablate the dimensionless base step size (`base_lr` $= \eta t$). We select a reported operating point using the cross-trial standard deviation of the privileged Oracle Mean estimator as one finite-run stability diagnostic.

Table 7: Finite-step sensitivity on the singular toy model $K_2(u, v) = u^2v^2$. Among the tested values, `base_lr` $= 0.05$ gives the smallest Oracle cross-trial standard deviation and is used as the reported operating point.

`base_lr`	Regime Description	Oracle Mean-based	SIVE
0.0001	Severe Finite-Run Bias	3.3124 ± 0.7967	4.0592 ± 2.9188
0.0010	Very Slow Exploration	0.2959 ± 0.2465	0.2349 ± 0.1881
0.0050	Slow Exploration	0.3325 ± 0.1669	0.3747 ± 0.3230
0.0100	Improved Finite-Run Stability	0.3544 ± 0.0791	0.4046 ± 0.2046
0.0500	**Selected Operating Point**	$\mathbf{0.3980 \pm 0.0424}$	$\mathbf{0.4310 \pm 0.0961}$
0.1000	Moderate Discretization Shift	0.4538 ± 0.0682	0.4808 ± 0.1063
0.5000	Severe Discretization Instability	0.5203 ± 0.1248	3.0707 ± 4.8986
1.0000	Numerical Divergence	0.5718 ± 0.1379	9.4489 ± 17.1410

As shown in Table 7, very small tested step sizes (≤ 0.005) exhibit poor finite-run mixing, while large values (≥ 0.5) exhibit severe discrete-time instability (Appendix B.1). Among this finite grid, `base_lr` $= 0.05$ has the smallest Oracle cross-trial standard deviation. We therefore use it as an empirical operating point, not as a universal optimum or a convergence certificate.

E.4 Evaluation-Group Sensitivity to N

SIVE dynamically subtracts the empirical noise floor using the within-group sample variance s_m^2. By the Law of Total Variance, the raw variance of the grouped empirical losses absorbs a structural artifact equal to $\mathbb{E}[t^2 s_m^2/N] = t^2\sigma^2/N$. For $t = 10^4$ and $\sigma = 0.002$, this uncorrected artifact scales to $400/N$, significantly dominating the $\mathcal{O}(1)$ geometric signal.

Our debiasing step subtracts this average noise profile $\hat{C} = t^2 M_{\mathcal{W}}^{-1} \sum_{m \in \mathcal{W}} s_m^2 / N$, where $M_{\mathcal{W}}$ is the number of retained post-burn-in steps. Under the idealized homoscedastic Gaussian and independent-step calculation, the standard deviation of the correction term alone is

$$\mathrm{Std}[\hat{C}] = \left(\frac{400}{N}\right) \sqrt{\frac{2}{M_{\mathcal{W}}(N-1)}}. \tag{74}$$

The reported sweep retains $M_{\mathcal{W}} = 70{,}000$ states. This idealized formula describes only correction-term uncertainty, not the full uncertainty of SIVE.

Table 8 records an end-to-end sweep of SIVE and the privileged Oracle Mean statistic across varying N. Within each row and trial, both use the same MCMC path, but Oracle Mean targets $t\bar{K}_{\mathcal{W}}$ whereas unclipped SIVE targets $t^2 S^2_{\mathcal{W}}(K)$; agreement between them is therefore not a calibration criterion. Across different N, evaluation and transition draws shared a pseudorandom stream, so changing N also changed the sampled path. The table summarizes the resulting end-to-end sensitivity.

Table 8: End-to-end evaluation-group sensitivity on the singular toy model $K_2(u, v) = u^2 v^2$. Entries are mean $\pm$ sample standard deviation across five trials. SIVE includes the positive-part numerical safeguard used in the reported runs.

N	Oracle Mean	SIVE
2	0.4057 ± 0.0483	1.8260 ± 1.1970
4	0.3651 ± 0.0571	0.5277 ± 0.4193
8	0.4300 ± 0.0772	0.4323 ± 0.3739
16	0.4300 ± 0.0772	0.3856 ± 0.1033
32	0.3899 ± 0.0474	0.3834 ± 0.0880
64	$\mathbf{0.3980 \pm 0.0424}$	$\mathbf{0.4310 \pm 0.0961}$
128	0.4286 ± 0.0820	0.4522 ± 0.1262
256	0.4214 ± 0.0699	0.4613 ± 0.1098
512	0.4960 ± 0.0545	0.5207 ± 0.0930
1024	0.3872 ± 0.0657	0.3664 ± 0.1059
2048	0.3941 ± 0.0636	0.3874 ± 0.0804

F Implementation and Scale Translation for Deep Neural Networks

Operationalizing the thermodynamic-variance target on overparameterized neural networks introduces practical challenges: the true local minimum $L_k^\star$ is inaccessible mid-training, the SGLD path is finite and localized, and loss measurements are stochastic. To deploy the estimator under a declared finite-step protocol, we rely on an operating point informed by the controlled ablations, explicit decoupling of stochastic noise sources, and hardware-accelerated vectorization.

The trained classifier has architecture 784–128–128–10, with ReLU activations after both hidden affine layers. We use the standard MNIST split (60,000 training and 10,000 test images), map pixels to $[0, 1]$ with `ToTensor`, and optimize mean cross-entropy for 100 epochs using plain SGD with learning rate 0.01, batch size 64, no momentum, and no weight decay. Train and test losses stored with the current checkpoints are full-dataset mean cross-entropies evaluated at the exact saved parameters.

F.1 Thermodynamic Translation and Adaptive Localization

The DNN experiments use a declared fixed-step protocol rather than calibrating the Unadjusted Langevin Algorithm (ULA) discretization bias. We use the empirical operating value `base_lr` $= 0.05$ identified in the controlled ablations (Appendix E.3) as a fixed convention across the deep-learning experiments, not as a convergence guarantee. This choice is independent of the unknown additive floor, which cancels from the sampling law; the estimation-layer guarantee does not cover ULA discretization bias.

We distinguish the dimensionless relative multiplier c_h from the physical Gaussian width used at checkpoint k:

$$h_k(c_h) = c_h \frac{\|\theta_{t_k}\|}{\sqrt{d}}, \tag{75}$$

where d is the total number of parameters. This RMS convention makes c_h a relative per-coordinate scale, while the actual width $h_k(c_h)$ remains checkpoint dependent. It does not confer general reparameterization invariance: the finite-scale probe is defined relative to the chosen Euclidean parameterization.

The five main trajectories in Section 4.2 use $c_h = 1$. The paired audit evaluates

$$c_h \in \{0.5, 1, 2, 4\} \tag{76}$$

at the checkpoints $\{1, 40, 100\}$. These values change the estimand and are therefore reported as a profile in Table 6, rather than pooled into a single estimate. The paired contrasts are the early-to-middle drop $D_{\mathrm{drop}}(c_h) = \widehat{\Lambda}_{1,c_h} - \widehat{\Lambda}_{40,c_h}$ and the late change $D_{\mathrm{late}}(c_h) = \widehat{\Lambda}_{100,c_h} - \widehat{\Lambda}_{40,c_h}$. For SIVE, the latter is a rebound when it is positive following the early drop. The primary correction comparison is Raw Variance versus SIVE under shared retained paths and evaluation groups. All three mean-energy reference choices are reported in the main trajectory; the multi-scale audit additionally tracks the checkpoint-reference contrast.

At each audited checkpoint, the implementation computes the gradient g_k of the full empirical training loss in memory-bounded batches, avoiding the positive noise floor in the squared norm of a stochastic gradient estimate. For $\ell_m = g_k^\top (\theta_m - \theta_{t_k})$ and $R_m = \bar{L}_m - \ell_m$, the retained path obeys the exact algebraic decomposition

$$t^2 S^2_{\mathcal{W}}(\bar{L}) = t^2 S^2_{\mathcal{W}}(R) + t^2 S^2_{\mathcal{W}}(\ell) + 2t^2 \operatorname{Cov}_{\mathcal{W}}(R, \ell). \tag{77}$$

Because subtracting ℓ_m does not change the within-state evaluation variance s^2_m, the same correction gives a residual SIVE statistic. Accordingly, the runner records the full-data $\|g_k\|^2$, $t^2 h_k(c_h)^2 \|g_k\|^2$, the three pathwise terms in Eq. (77), the uncorrected Raw Variance, and both original and residual SIVE values. This is a diagnostic decomposition of the realized path, not an exact nonlinear slope correction.

F.2 Decoupling Gradient Dynamics and Observation Noise

Neural network experiments rely on mini-batch subsampling, which injects stochasticity into both the SGLD parameter dynamics and the geometric loss measurements. To control these effects independently, we separate the gradient mini-batch size B_{grad} (used strictly for the Langevin drift) from the evaluation mini-batch size B_{eval} (used strictly to measure the local energy state). Gradient and evaluation indices are sampled uniformly with replacement from the 60,000 training examples. The experiment assigns independent pseudorandom streams to gradient-batch indices, Langevin Gaussian draws, and evaluation-batch indices. Stream seeds are deterministic functions of training trajectory, probe trial, and checkpoint, so the multi-c_h audit reuses common random numbers without allowing N or evaluation draws to alter the parameter dynamics.

To reduce stochastic-gradient perturbations of the intended drift, we use a relatively large gradient mini-batch, fixing $B_{\mathrm{grad}} = 1024$ (approximately 1.7% of the MNIST training data). This does not make the drift equal to the full empirical-loss gradient, so stochastic-gradient bias remains one of the sampler effects in Eq. (73). In contrast, B_{eval} is controlled independently, as it governs the observation noise injected into the variance-based finite-scale measurement.

F.3 Vectorized Scaling of the Evaluation Group N

At each retained state used by SIVE, the estimator requires N independent evaluation mini-batches, each of size B_{eval}. The reported runner records these groups at every SGLD step and discards burn-in only during estimation; this uniform implementation is not a mathematical requirement of the estimator. If $\sigma^2_{\mathrm{single}}(\theta_m)$ denotes the conditional variance of the per-sample loss at state θ_m, the additive observation variance embedded within the grouped empirical mean scales according to:

$$\operatorname{Var}(\bar{L}_m \mid \theta_m) = \frac{\sigma^2_{\mathrm{single}}(\theta_m)}{N B_{\mathrm{eval}}}. \tag{78}$$

Consequently, the raw observation-noise floor is controlled by the total number of evaluated examples, NB_{eval}. Under the idealized homoscedastic Gaussian calculation in Appendix E.4, the within-group sample variance s_m^2 has $N-1$ degrees of freedom, so increasing N improves the precision of the empirical correction. For general non-Gaussian evaluation noise, that uncertainty also depends on higher moments and is not determined by $N-1$ alone. In the reported DNN protocol, B_{eval} is fixed and increasing N trades additional evaluation cost for more precise within-state replication.

Because all N evaluations share the same parameter state, the implementation uses `torch.func.functional_call` to flatten indices of shape (N, B_{eval}) into a batched evaluation dimension and then reshape the unreduced losses, avoiding N sequential model invocations. This reduces launch overhead while retaining the underlying evaluation and memory costs; the toy diagnostics use $N = 64$, whereas the neural experiments use $N = 512$ with $B_{\mathrm{eval}} = 64$.

The code stores all three mean-energy reference choices, Raw Variance, and unclipped SIVE from the same retained group means; the clipped positive part is a separate field. The CSV fields `Online` and `Retrospective` are the path-minimum and final-reference mean probes, respectively. The logger also records negative-estimate frequencies, empirical quantiles, $h_k(c_h)$, $\eta/h_k(c_h)^2$, the RMS parameter displacement, and displacement divided by $h_k(c_h)$. Train and test loss are evaluated at the exact saved parameters, including the random-initialization checkpoint, rather than writing an Epoch-0 placeholder. Each output directory contains the resolved configuration, seeds, Python/platform information and Git state, source-file hashes, raw per-trial CSV files, and compressed SGLD histories. The reported five-trajectory run contains no negative unclipped SIVE values and the RMS displacement remains within approximately 0.99–1.01 times $h_k(c_h)$ at the audited checkpoints. These records define the reported protocol.

G Complete Five-Trajectory DNN Results

The table below reports every checkpoint from the five-trajectory experiment used in Figure 1. Each entry is the mean $\pm$ sample standard deviation across the five independently trained networks after first averaging five probe trials within each network. The three mean-energy columns share Eq. (31); the final two columns are the direct Raw Variance–SIVE pair in Eq. (32). The implementation fields `Online` and `Retrospective` correspond to "Path min" and "Final ref." below.

For the aggregate curves, the Epoch-100 contrast from any intermediate checkpoint $k = 30, \ldots, 60$ ranges from $+41.8$ to $+53.0$ for SIVE and from -400.8 to -85.3 for Raw Variance, computed from the unrounded trajectory means. Thus the opposite late directions do not depend on choosing Epoch 40, consistent with a broad mid-training trough in SIVE; this aggregate check does not establish every corresponding within-trajectory comparison.

Table 9: Complete five-trajectory DNN results at $c_h = 1$.

			Mean energy: reference choice			Thermodynamic variance	
Epoch	Train loss	Test loss	Ckpt.	Path min	Final ref.	Raw	SIVE
0	2.3027 ± 0.0011	2.3029 ± 0.0013	-15513.1 ± 35.0	319.0 ± 10.9	7068.5 ± 40.6	5709.3 ± 89.8	4650.1 ± 92.2
1	0.7929 ± 0.0393	0.7741 ± 0.0394	-3161.1 ± 314.7	235.0 ± 7.3	4322.5 ± 85.7	3083.6 ± 57.8	1385.0 ± 79.1
2	0.4382 ± 0.0064	0.4209 ± 0.0062	-647.6 ± 55.5	219.7 ± 4.6	3289.1 ± 17.1	2687.3 ± 23.5	630.1 ± 36.1
3	0.3626 ± 0.0021	0.3499 ± 0.0026	-274.7 ± 31.9	212.7 ± 4.1	2905.5 ± 20.7	2552.9 ± 25.0	457.2 ± 22.6
4	0.3255 ± 0.0016	0.3139 ± 0.0030	-144.9 ± 22.6	208.3 ± 2.3	2664.8 ± 30.6	2446.5 ± 57.3	396.9 ± 17.5
5	0.3021 ± 0.0023	0.2934 ± 0.0035	-96.3 ± 15.9	205.7 ± 1.7	2479.4 ± 38.5	2351.5 ± 56.1	371.1 ± 10.8
6	0.2811 ± 0.0047	0.2746 ± 0.0061	-44.8 ± 11.7	197.1 ± 6.3	2320.3 ± 43.4	2251.4 ± 55.3	349.5 ± 6.8
7	0.2657 ± 0.0045	0.2603 ± 0.0052	-34.4 ± 24.3	193.8 ± 3.3	2176.6 ± 46.1	2140.2 ± 55.4	321.8 ± 8.5
8	0.2494 ± 0.0050	0.2451 ± 0.0050	-3.3 ± 8.6	189.4 ± 5.1	2045.2 ± 46.9	2040.7 ± 70.2	303.2 ± 24.7
9	0.2355 ± 0.0046	0.2327 ± 0.0049	14.8 ± 9.5	184.4 ± 5.3	1924.1 ± 45.2	1954.7 ± 51.3	295.5 ± 10.0
10	0.2217 ± 0.0050	0.2208 ± 0.0056	41.0 ± 12.4	180.0 ± 6.2	1812.7 ± 42.6	1862.0 ± 53.6	281.3 ± 20.7
11	0.2111 ± 0.0061	0.2103 ± 0.0069	43.9 ± 33.9	176.2 ± 4.2	1709.1 ± 39.1	1778.1 ± 48.3	271.5 ± 17.2
12	0.1996 ± 0.0044	0.2000 ± 0.0050	64.0 ± 18.8	171.7 ± 3.7	1614.1 ± 35.1	1706.5 ± 47.0	266.5 ± 18.6
13	0.1909 ± 0.0039	0.1919 ± 0.0025	63.0 ± 16.4	171.2 ± 4.1	1526.7 ± 32.3	1632.5 ± 32.0	256.3 ± 9.6
14	0.1800 ± 0.0020	0.1826 ± 0.0021	91.4 ± 9.8	165.8 ± 3.6	1446.1 ± 30.2	1562.5 ± 34.2	243.3 ± 10.7
15	0.1715 ± 0.0029	0.1738 ± 0.0028	102.6 ± 7.0	159.5 ± 4.4	1372.4 ± 28.0	1502.0 ± 32.5	237.0 ± 9.6
16	0.1634 ± 0.0038	0.1672 ± 0.0035	116.0 ± 13.9	160.0 ± 5.7	1304.4 ± 26.9	1443.4 ± 30.9	227.6 ± 11.7
17	0.1556 ± 0.0029	0.1604 ± 0.0030	130.7 ± 4.9	152.1 ± 1.5	1241.3 ± 25.1	1389.3 ± 23.3	219.7 ± 1.2
18	0.1493 ± 0.0033	0.1549 ± 0.0036	135.1 ± 9.9	151.6 ± 5.2	1182.7 ± 24.1	1345.0 ± 32.1	218.8 ± 10.7
19	0.1435 ± 0.0019	0.1492 ± 0.0024	138.8 ± 13.3	148.2 ± 4.9	1128.5 ± 23.5	1299.5 ± 32.0	213.3 ± 10.1
20	0.1369 ± 0.0026	0.1441 ± 0.0023	154.8 ± 2.2	146.6 ± 4.3	1077.8 ± 22.9	1261.0 ± 24.1	212.3 ± 6.2
21	0.1317 ± 0.0026	0.1401 ± 0.0019	159.1 ± 7.9	144.7 ± 4.9	1030.3 ± 22.1	1217.0 ± 27.9	204.0 ± 5.2
22	0.1268 ± 0.0027	0.1359 ± 0.0016	163.1 ± 12.6	141.0 ± 3.5	985.8 ± 21.1	1180.9 ± 26.2	201.6 ± 7.7
23	0.1216 ± 0.0025	0.1316 ± 0.0020	173.5 ± 8.9	139.7 ± 2.7	944.2 ± 20.7	1148.7 ± 22.9	200.7 ± 6.9
24	0.1179 ± 0.0025	0.1288 ± 0.0030	171.7 ± 5.4	141.6 ± 4.6	904.9 ± 19.9	1114.1 ± 24.2	196.1 ± 7.3
25	0.1129 ± 0.0024	0.1243 ± 0.0015	184.3 ± 9.6	134.9 ± 4.3	868.1 ± 19.7	1085.9 ± 27.3	195.4 ± 6.1
26	0.1086 ± 0.0021	0.1215 ± 0.0024	192.8 ± 6.7	132.6 ± 3.6	833.6 ± 19.0	1057.8 ± 19.7	193.5 ± 4.2
27	0.1047 ± 0.0024	0.1180 ± 0.0016	199.3 ± 5.7	133.1 ± 2.1	800.7 ± 18.8	1028.2 ± 23.0	187.8 ± 3.0
28	0.1008 ± 0.0024	0.1149 ± 0.0026	206.9 ± 8.2	132.8 ± 4.5	769.8 ± 18.2	1006.8 ± 25.9	190.4 ± 5.9
29	0.0977 ± 0.0021	0.1132 ± 0.0021	209.1 ± 9.0	129.0 ± 4.3	740.7 ± 17.7	984.1 ± 19.6	188.4 ± 6.3
30	0.0950 ± 0.0023	0.1116 ± 0.0024	209.4 ± 8.3	127.1 ± 3.9	713.3 ± 17.8	959.6 ± 19.8	183.4 ± 4.3
31	0.0920 ± 0.0020	0.1089 ± 0.0021	212.5 ± 7.3	128.6 ± 5.5	687.3 ± 17.1	942.3 ± 18.5	186.0 ± 3.0
32	0.0889 ± 0.0020	0.1072 ± 0.0017	219.1 ± 4.9	124.4 ± 3.2	662.2 ± 16.5	917.8 ± 19.5	180.9 ± 2.6
33	0.0861 ± 0.0015	0.1054 ± 0.0012	223.3 ± 6.4	123.8 ± 6.5	638.6 ± 15.7	901.7 ± 21.7	182.0 ± 6.7
34	0.0834 ± 0.0016	0.1033 ± 0.0014	227.9 ± 12.4	120.1 ± 4.7	616.5 ± 15.3	886.9 ± 19.8	183.8 ± 5.4
35	0.0805 ± 0.0018	0.1016 ± 0.0018	235.6 ± 5.3	120.4 ± 4.5	595.4 ± 14.8	864.6 ± 16.7	176.3 ± 4.0
36	0.0789 ± 0.0021	0.1010 ± 0.0027	231.3 ± 10.6	117.8 ± 3.7	574.7 ± 14.6	850.9 ± 19.5	178.9 ± 4.3
37	0.0760 ± 0.0023	0.0984 ± 0.0021	241.6 ± 8.7	118.7 ± 3.1	555.9 ± 14.4	839.2 ± 19.2	180.0 ± 5.9
38	0.0738 ± 0.0017	0.0969 ± 0.0019	244.1 ± 7.6	115.7 ± 1.7	536.8 ± 13.6	824.3 ± 17.9	178.7 ± 7.8
39	0.0714 ± 0.0014	0.0957 ± 0.0024	251.4 ± 8.7	115.5 ± 3.0	519.4 ± 13.5	810.2 ± 18.8	177.2 ± 3.8
40	0.0696 ± 0.0022	0.0943 ± 0.0020	252.7 ± 8.5	116.8 ± 3.6	502.8 ± 12.8	800.9 ± 15.9	179.8 ± 3.7
41	0.0676 ± 0.0016	0.0932 ± 0.0011	255.9 ± 6.1	113.2 ± 3.6	486.7 ± 13.0	787.4 ± 14.7	178.2 ± 3.3
42	0.0665 ± 0.0020	0.0932 ± 0.0025	252.0 ± 7.9	113.6 ± 3.9	471.0 ± 12.5	773.6 ± 17.6	176.1 ± 5.7
43	0.0645 ± 0.0013	0.0918 ± 0.0024	256.9 ± 14.1	111.6 ± 3.3	456.7 ± 12.1	764.3 ± 15.7	177.2 ± 3.0
44	0.0620 ± 0.0017	0.0897 ± 0.0010	268.1 ± 8.0	109.9 ± 2.4	442.3 ± 11.7	757.7 ± 15.3	180.3 ± 2.2
45	0.0607 ± 0.0008	0.0897 ± 0.0021	267.5 ± 9.8	111.2 ± 2.8	428.7 ± 11.5	743.4 ± 15.6	176.5 ± 2.1
46	0.0593 ± 0.0018	0.0886 ± 0.0013	268.9 ± 12.8	111.6 ± 2.4	415.7 ± 10.9	739.1 ± 13.7	181.1 ± 3.4
47	0.0576 ± 0.0015	0.0879 ± 0.0023	272.8 ± 5.8	109.2 ± 2.7	403.3 ± 10.7	726.7 ± 12.1	177.9 ± 4.3
48	0.0560 ± 0.0022	0.0868 ± 0.0017	277.0 ± 13.5	107.3 ± 3.2	391.4 ± 10.6	721.4 ± 16.1	181.6 ± 3.7
49	0.0544 ± 0.0013	0.0853 ± 0.0015	281.7 ± 2.2	110.5 ± 3.3	379.7 ± 10.2	707.0 ± 13.6	175.1 ± 2.4
50	0.0528 ± 0.0010	0.0847 ± 0.0016	286.1 ± 6.1	106.3 ± 3.8	368.3 ± 10.2	703.9 ± 14.6	180.0 ± 4.3
51	0.0515 ± 0.0016	0.0843 ± 0.0018	288.1 ± 7.8	106.1 ± 2.5	357.6 ± 10.0	695.8 ± 17.3	179.5 ± 4.9
52	0.0500 ± 0.0015	0.0831 ± 0.0010	292.6 ± 4.2	108.3 ± 2.8	347.4 ± 9.6	688.0 ± 12.5	179.5 ± 3.2
53	0.0490 ± 0.0012	0.0833 ± 0.0016	293.3 ± 6.7	107.3 ± 3.4	337.4 ± 9.2	684.0 ± 14.0	182.2 ± 3.2
54	0.0479 ± 0.0011	0.0831 ± 0.0007	294.4 ± 9.3	105.5 ± 4.4	327.6 ± 9.3	675.9 ± 14.0	181.8 ± 5.2

Continued on next page

Table 9 (continued)

			Mean energy: reference choice			Thermodynamic variance	
Epoch	Train loss	Test loss	Ckpt.	Path min	Final ref.	Raw	SIVE
55	0.0477 ± 0.0016	0.0828 ± 0.0019	287.4 ± 3.3	104.1 ± 1.9	318.5 ± 8.9	670.1 ± 12.5	182.5 ± 3.6
56	0.0462 ± 0.0010	0.0824 ± 0.0016	293.2 ± 8.4	105.8 ± 4.9	309.2 ± 9.0	663.4 ± 13.8	182.4 ± 4.7
57	0.0440 ± 0.0011	0.0803 ± 0.0008	305.8 ± 5.6	103.0 ± 3.7	300.6 ± 8.6	660.7 ± 12.3	185.4 ± 3.7
58	0.0431 ± 0.0009	0.0802 ± 0.0017	306.8 ± 5.3	101.9 ± 1.7	292.1 ± 8.6	650.4 ± 13.8	181.6 ± 4.2
59	0.0422 ± 0.0014	0.0805 ± 0.0021	307.7 ± 4.9	99.9 ± 4.2	283.9 ± 8.6	646.9 ± 10.1	183.6 ± 3.9
60	0.0414 ± 0.0022	0.0793 ± 0.0017	307.7 ± 10.4	102.2 ± 3.2	276.2 ± 8.1	644.1 ± 12.8	186.4 ± 4.4
61	0.0397 ± 0.0011	0.0792 ± 0.0009	317.7 ± 3.5	102.8 ± 3.6	268.7 ± 8.1	636.6 ± 11.2	184.0 ± 6.1
62	0.0392 ± 0.0010	0.0795 ± 0.0016	314.6 ± 8.5	101.7 ± 1.8	261.3 ± 8.0	635.1 ± 11.3	187.9 ± 3.6
63	0.0382 ± 0.0013	0.0784 ± 0.0013	317.1 ± 10.1	101.7 ± 5.0	253.9 ± 7.8	630.7 ± 16.1	189.3 ± 6.3
64	0.0379 ± 0.0016	0.0782 ± 0.0015	313.7 ± 16.4	103.3 ± 4.2	246.8 ± 7.8	625.6 ± 14.4	188.9 ± 3.7
65	0.0367 ± 0.0016	0.0784 ± 0.0020	319.4 ± 6.7	100.6 ± 3.1	240.5 ± 7.8	622.7 ± 12.3	190.6 ± 5.9
66	0.0360 ± 0.0012	0.0778 ± 0.0007	318.8 ± 5.8	100.5 ± 3.0	233.5 ± 7.5	616.5 ± 11.8	189.0 ± 3.0
67	0.0352 ± 0.0014	0.0783 ± 0.0011	320.9 ± 8.0	98.9 ± 1.2	227.4 ± 7.4	612.2 ± 14.1	189.0 ± 6.6
68	0.0339 ± 0.0010	0.0771 ± 0.0007	327.5 ± 4.4	99.8 ± 1.7	221.3 ± 7.4	609.2 ± 12.8	190.3 ± 3.7
69	0.0338 ± 0.0012	0.0772 ± 0.0008	322.6 ± 2.4	99.7 ± 4.6	215.1 ± 7.0	607.3 ± 12.5	193.6 ± 3.3
70	0.0328 ± 0.0016	0.0773 ± 0.0024	327.2 ± 15.5	98.2 ± 2.5	209.1 ± 6.8	601.8 ± 15.0	192.0 ± 5.3
71	0.0318 ± 0.0010	0.0766 ± 0.0011	331.7 ± 10.0	99.5 ± 3.2	203.7 ± 6.8	602.3 ± 11.3	196.3 ± 6.3
72	0.0307 ± 0.0008	0.0756 ± 0.0014	336.9 ± 7.5	99.4 ± 3.1	198.2 ± 6.5	598.7 ± 12.8	196.3 ± 3.9
73	0.0300 ± 0.0008	0.0761 ± 0.0013	338.8 ± 8.0	96.8 ± 3.0	192.8 ± 6.5	594.3 ± 13.6	196.1 ± 7.5
74	0.0296 ± 0.0009	0.0754 ± 0.0013	336.9 ± 5.2	98.6 ± 3.0	187.5 ± 6.4	592.2 ± 10.6	197.0 ± 4.4
75	0.0290 ± 0.0007	0.0758 ± 0.0006	338.6 ± 9.3	97.3 ± 3.9	182.5 ± 6.2	594.5 ± 12.8	202.5 ± 7.3
76	0.0279 ± 0.0010	0.0751 ± 0.0015	344.6 ± 6.3	98.7 ± 2.3	177.6 ± 6.1	587.7 ± 11.1	199.9 ± 6.2
77	0.0274 ± 0.0010	0.0760 ± 0.0014	344.1 ± 7.7	97.0 ± 3.2	172.7 ± 5.9	585.6 ± 13.3	201.1 ± 5.2
78	0.0267 ± 0.0010	0.0750 ± 0.0016	346.6 ± 8.9	95.1 ± 2.2	168.1 ± 5.9	581.5 ± 11.7	200.0 ± 3.0
79	0.0263 ± 0.0011	0.0760 ± 0.0022	345.8 ± 7.4	97.2 ± 2.8	163.6 ± 6.0	583.0 ± 11.2	204.2 ± 6.6
80	0.0261 ± 0.0012	0.0750 ± 0.0011	344.0 ± 6.2	97.0 ± 2.5	159.2 ± 5.6	577.8 ± 13.0	202.9 ± 5.0
81	0.0260 ± 0.0016	0.0757 ± 0.0024	341.3 ± 18.2	96.9 ± 1.7	155.1 ± 5.9	578.2 ± 10.3	205.7 ± 3.0
82	0.0245 ± 0.0015	0.0746 ± 0.0018	351.4 ± 6.7	97.5 ± 3.5	150.9 ± 5.7	571.8 ± 12.2	202.3 ± 4.1
83	0.0239 ± 0.0007	0.0748 ± 0.0012	352.8 ± 6.5	93.2 ± 2.2	146.6 ± 5.5	578.2 ± 11.7	211.7 ± 3.9
84	0.0234 ± 0.0010	0.0748 ± 0.0011	354.0 ± 5.7	95.2 ± 1.5	142.6 ± 5.2	572.7 ± 12.8	208.7 ± 6.8
85	0.0228 ± 0.0010	0.0742 ± 0.0014	356.9 ± 9.5	94.0 ± 2.7	138.8 ± 5.0	569.7 ± 12.9	208.3 ± 6.5
86	0.0222 ± 0.0007	0.0745 ± 0.0012	358.1 ± 8.0	96.2 ± 3.0	134.7 ± 4.9	569.3 ± 13.3	210.4 ± 6.0
87	0.0218 ± 0.0007	0.0749 ± 0.0010	358.7 ± 6.0	94.3 ± 2.6	131.1 ± 5.3	567.0 ± 12.3	210.6 ± 5.4
88	0.0213 ± 0.0006	0.0739 ± 0.0014	361.0 ± 6.2	96.9 ± 1.9	128.0 ± 4.9	566.5 ± 12.2	212.3 ± 4.2
89	0.0208 ± 0.0007	0.0748 ± 0.0010	361.9 ± 6.8	92.8 ± 1.6	124.1 ± 5.2	568.0 ± 8.8	216.5 ± 2.3
90	0.0204 ± 0.0007	0.0745 ± 0.0017	362.0 ± 7.9	94.3 ± 0.9	120.5 ± 4.7	564.2 ± 14.0	214.7 ± 6.3
91	0.0201 ± 0.0010	0.0747 ± 0.0017	362.3 ± 11.9	93.8 ± 3.2	117.8 ± 4.8	564.9 ± 14.0	217.4 ± 5.9
92	0.0196 ± 0.0010	0.0745 ± 0.0015	364.2 ± 4.3	94.0 ± 1.7	114.5 ± 4.7	561.1 ± 14.4	216.0 ± 7.3
93	0.0192 ± 0.0010	0.0746 ± 0.0014	364.7 ± 8.6	94.3 ± 5.4	111.2 ± 4.5	564.9 ± 14.9	221.3 ± 7.3
94	0.0185 ± 0.0008	0.0738 ± 0.0012	368.0 ± 10.7	92.7 ± 2.2	107.8 ± 4.7	557.7 ± 10.3	217.0 ± 4.6
95	0.0183 ± 0.0004	0.0744 ± 0.0009	368.0 ± 10.0	95.2 ± 2.1	105.1 ± 4.4	557.6 ± 8.7	218.2 ± 2.1
96	0.0178 ± 0.0007	0.0739 ± 0.0011	370.4 ± 6.1	94.8 ± 2.8	102.4 ± 4.5	557.8 ± 16.4	220.3 ± 9.7
97	0.0174 ± 0.0005	0.0745 ± 0.0017	370.3 ± 9.5	95.2 ± 2.2	99.0 ± 4.4	558.7 ± 10.5	222.8 ± 5.3
98	0.0179 ± 0.0017	0.0760 ± 0.0032	363.1 ± 9.6	92.7 ± 2.5	96.4 ± 4.4	561.3 ± 14.6	226.9 ± 8.2
99	0.0169 ± 0.0007	0.0756 ± 0.0014	370.1 ± 6.2	93.9 ± 2.3	93.7 ± 4.1	556.9 ± 16.0	224.4 ± 9.8
100	0.0163 ± 0.0007	0.0747 ± 0.0012	373.4 ± 6.3	90.9 ± 4.1	90.9 ± 4.1	558.8 ± 15.5	228.1 ± 10.0