EQUICONFIG: EFFICIENT MITIGATION OF DISPARATE IMPACT IN DIFFERENTIALLY PRIVATE LEARNING

Anonymous authors
Paper under double-blind review

ABSTRACT

Differentially private stochastic gradient descent (DP-SGD) enables neural network training with record-level privacy by clipping per-example gradients and adding calibrated noise to their aggregate. These operations can disproportionately weaken underrepresented groups' contributions, reducing their accuracy. Mitigation strategies rely on adaptive clipping and introduce extra parameters, increasing the computational and privacy costs of calibration. We propose EQUICONFIG, which screens DP-SGD's existing batch-size, clipping, and noise settings offline to construct a targeted shortlist for private calibration. Privacy accounting pairs each batch size with the noise required at the target privacy budget, forming a privacy ridge. Along each ridge, larger batches can strengthen group contributions relative to noise but leave fewer optimization steps under fixed privacy and epoch budgets. To assess this trade-off, EQUICONFIG defines a cumulative group-information score that combines per-update group signal relative to noise with expected group participation over training. Privacy-frontier elasticity analysis identifies batch increases that provably improve this score for every target group, while minimum total-update and group-activity requirements exclude configurations with too few training opportunities. We apply the screen to candidate grids spanning Gaussian, Laplace, and mixture-Laplace noise and multiple clipping bounds, yielding alternative configurations without adaptive training controls or gradient checkpoints. We compare against state-of-the-art mitigation methods both without private calibration and with DP-HyPO accounting for selection costs. Across vision, tabular, and language tasks, the results demonstrate substantial improvements in accuracy–disparity trade-offs at matched privacy targets, with fewer candidate evaluations under private calibration.

1 INTRODUCTION

Private learning should protect individual records without concentrating utility loss on smaller or historically underrepresented groups. Differentially private stochastic gradient descent (DP-SGD) (Abadi et al., 2016) provides record-level privacy by clipping per-example gradients and adding calibrated noise to their aggregate. These actions are known to impact different groups in unequal ways. On imbalanced MNIST, a standard benchmark for disparate impact, Bagdasaryan et al. (2019) observed an accuracy gap exceeding 20 percentage points under DP-SGD, compared with approximately 3 points under non-private training. Existing mitigation methods, including DPSGD-F, DPSGD-GLOBAL-ADAPT, and bounded adaptive clipping, address group disparities by modifying the clipping mechanism. Their adaptive rules introduce controls such as a target quantile γ, adaptation rate η_C, and minimum clipping threshold C_{LB}, whose private calibration can increase computational and privacy costs (Table 1). Beyond calibration, these methods consume a portion of the privacy budget during training to estimate the gradient statistics used to adjust clipping thresholds. These costs motivate a *complementary approach*: improving accuracy–disparity trade-offs through joint configuration of

Table 1: Existing methods add parameters requiring calibration.

Existing Methods	Additional Parameters
DP-SGD-F Xu et al. (2021)	σ_{count}
DP-SGD-G-A Esipova et al. (2023)	Z_0, η_Z, τ, σ_{count}
Bounded-Adaptive Zhao et al. (2026)	C_{LB}, γ, τ, η_C, σ_{count}
EquiConfig (ours)	None

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.

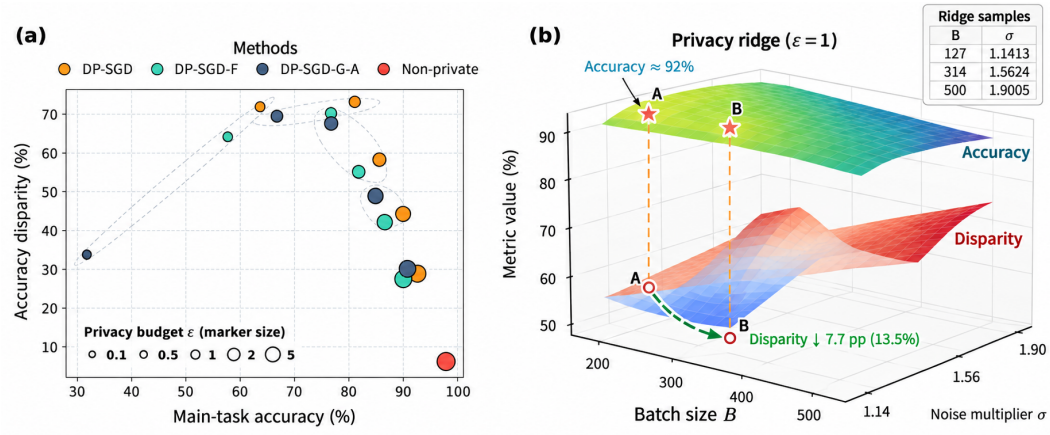


Figure 1: Improving accuracy–disparity trade-offs through DP-SGD configuration. (a) Standard DP-SGD and clipping-based mitigation methods, with a non-private reference. Marker size denotes the privacy budget ε for private methods; dashed ellipses group results at matched budgets. Higher accuracy ($\rightarrow$) and lower disparity ($\downarrow$) are preferred. (b) Gaussian DP-SGD along a batch–noise privacy ridge at $\varepsilon = 1$ and a fixed epoch budget. Increasing batch size and recalibrating noise between the highlighted configurations reduces disparity by 7.7 percentage points (13.5% relative) while maintaining accuracy near 92%.

standard DP-SGD parameters without introducing additional mechanism hyperparameters. Our key observation is that batch size, clipping, and noise jointly govern the strength and frequency of group contributions under fixed privacy and epoch budgets. Larger batches increase expected group representation per update but yield fewer updates and alter the required noise scale, while clipping determines how much of each contribution is retained. Figure 1(b) illustrates this opportunity: adjusting batch size and Gaussian noise under fixed privacy and epoch budgets preserves accuracy near $\mathbf{92}\%$ while reducing disparity by $\mathbf{7.7}\%$. However, finding such configurations through repeated training and private evaluation can be costly. We propose EQUICONFIG, which screens DP-SGD configurations offline to focus the available private-calibration budget on a promising shortlist. Along the privacy ridge illustrated by the dashed ellipses in Figure 1(b), the screen balances **potential group-signal gains from larger batches** against **fewer group-active updates** under fixed epochs.

Motivated by **the fixed-epoch trade-off studied by Rehn et al. (2026)**, EQUICONFIG floors (filters) minimums on the total number of optimizer updates and the expected number of updates containing examples from each target group. Among configurations meeting minimum total and group-active update counts, privacy-frontier elasticity analysis (PAN) tests whether a larger batch improves the cumulative group-information score for every target group. The score combines group participation over training with retained signal relative to privacy noise. Using public or separately privacy-accounted group metadata, the offline screen constructs a shortlist across Gaussian, Lap_2, and PLRV-O noise choices without training candidate models or inspecting their gradients. Within each noise, clipping, and learning-rate branch, it preserves nondominated trade-offs between group information and update count. DP-HYPO (Wang et al., 2023) then trains shortlisted candidates and uses privacy-accounted validation to select a model with low measured disparity subject to a predeclared accuracy requirement. PAN certifies improvement in the screening score under its stated conditions; the measured outcomes determine the final choice. Section 3 details both stages.

Contributions.

- We introduce EQUICONFIG[1], an offline screen that narrows existing DP-SGD settings before privacy-accounted calibration. It supports Gaussian, Lap_2, and PLRV-O noise without adaptive training controls or gradient checkpoints.
- We derive a cumulative group-information score and a finite PAN test for batch increases that improve it for every target group at fixed privacy and epoch budgets. Minimum total and group-active update counts exclude configurations with too few training opportunities.
- We evaluate six vision, tabular, and language benchmarks. Screening reduces the CNN–MNIST grid from 324 to 72 configurations (Section 4.3). On TinyViT–CIFAR-10, exhaustive selection raises minority accuracy from 46.90% to 71.30% and lowers disparity from 50.50 to 26.60 percentage points relative to standard DP-SGD at the same per-model privacy target (Table 3). Prompt-tuned RoBERTa also improves accuracy and disparity on SST-2 and QNLI, with smaller gains on QNLI.

[1] Anonymized source-code artifact.

2 BACKGROUND

Differential privacy and DP-SGD. Differential privacy (DP) limits the influence of one training record on a released model. A randomized mechanism $\mathcal{M}$ satisfies (ε, δ)-DP (Dwork, 2006) if, for neighboring datasets D, D' [2] and every measurable output event S, $\Pr[\mathcal{M}(D) \in S] \leq e^{\varepsilon} \Pr[\mathcal{M}(D') \in S] + \delta$. Here ε controls the multiplicative term and δ allows additive slack; smaller values give stronger privacy guarantees. Differentially private stochastic gradient descent (DP-SGD) (Abadi et al., 2016) clips per-example loss gradients and adds calibrated noise to the average.

Consider a model $f_{\theta_t} : \mathcal{X} \to \mathcal{Y}$ with d parameters $\theta_t \in \mathbb{R}^d$ at training step t, where $\mathcal{X}$ and $\mathcal{Y}$ denote the input and output spaces, respectively. Let $\mathcal{D} = \{z_i\}_{i=1}^N$ be the training dataset, $g_{t,i} = \nabla_\theta \ell(\theta_t; z_i)$ the loss gradient for record i, and $C > 0$ the ℓ_2 clipping bound. Under Poisson sampling, each record is independently included with probability q at each step. Thus, the batch index set S_t has expected size $B = qN$. The DP-SGD update is

$$\begin{aligned} \bar{g}_{t,i} &= \frac{g_{t,i}}{\max\{1, \|g_{t,i}\|_2 / C\}}, \qquad \|\bar{g}_{t,i}\|_2 \leq C, \\ \theta_{t+1} &= \theta_t - \eta_t \left(\frac{1}{B} \sum_{i \in S_t} \bar{g}_{t,i} + \frac{\sigma C}{B} \xi_t \right), \qquad \xi_t \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, I_d), \end{aligned} \tag{1}$$

where η_t is the learning rate, σ is the noise multiplier, and I_d is the $d \times d$ identity matrix. For a nominal epoch budget E, we schedule $T(B) = \lceil EN/B \rceil$ parameter updates. For a specified δ, PLD accountant (Doroshenko et al., 2022) tightly composes the per-update privacy-loss distributions to bound ε for the full training run. In practice, the accountant is inverted to select the noise multiplier that attains a prescribed target (ε, δ)-DP. However, the PLD bound certifies one training run; it does not automatically certify a search over configurations. If several candidates are trained and compared using private data, end-to-end accounting must also cover candidate training and the private measurements used to select the final model.

Disparate Impacts of DP-SGD. Clipping and noise can reduce accuracy disproportionately for underrepresented classes and attribute-defined groups (Bagdasaryan et al., 2019; Esipova et al., 2023). Let $\mathcal{G}$ be the predeclared groups considered in screening and evaluation, and let n_ν be the number of training records in group $\nu \in \mathcal{G}$. Here, underrepresented means that n_ν is small relative to the comparison groups. Let $\mathrm{Acc}_\nu(f_\theta)$ denote accuracy on group ν in a specified held-out split: validation data for configuration selection or test data for final reporting. Using private validation data for selection requires privacy-accounted measurements. We measure the group-accuracy gap as $\mathrm{Disp}(f_\theta) = \max_{\nu \in \mathcal{G}} \mathrm{Acc}_\nu(f_\theta) - \min_{\nu \in \mathcal{G}} \mathrm{Acc}_\nu(f_\theta)$.

Existing mitigation approaches adaptively tune clipping during training to preserve group contributions. DPSGD-F (Xu et al., 2021) adjusts contributions according to group clipping bias; DPSGD-G-A (Esipova et al., 2023) scales gradients to reduce clipping-induced misalignment and privately adapts a clipping threshold. Bounded adaptive clipping (Zhao et al., 2026) places a tunable floor on that threshold. Evaluating candidate models on private validation data consumes both privacy and computational budgets. Calibrating these additional parameters therefore incurs costs without necessarily improving the accuracy–disparity trade-off.

In the exhaustive-search CNN–MNIST comparison at the same per-run privacy target ($\varepsilon = 3$), DPSGD-G-A achieves 90% overall accuracy with a group-accuracy gap of 41 percentage points, compared with 95% accuracy and a 20-point gap for standard DP-SGD (Table 2). Appendix A includes an extended discussion of related work. Appendix B details PLD and Rényi accounting, private calibration with DP-HYPO (Wang et al., 2023), and the Gaussian and non-Gaussian noise mechanisms considered here, including Lap_2 (Mohammady et al., 2026) and coordinatewise PLRV-O (Yang et al., 2025).

3 METHODOLOGY: SCREENING AND CALIBRATION WITH EQUICONFIG

Figure 2 shows the two roles of EQUICONFIG. Its offline screen asks where increasing batch size at a fixed training privacy target improves cumulative group information as the update count falls

[2] e.g., a *sample-level add-remove* adjacency, under which D and D' differ by one record.

or stays constant. It then passes the survivors to private calibration, which chooses a model using measured accuracy and disparity. The screening score guides where to train; it does not predict either measured outcome.

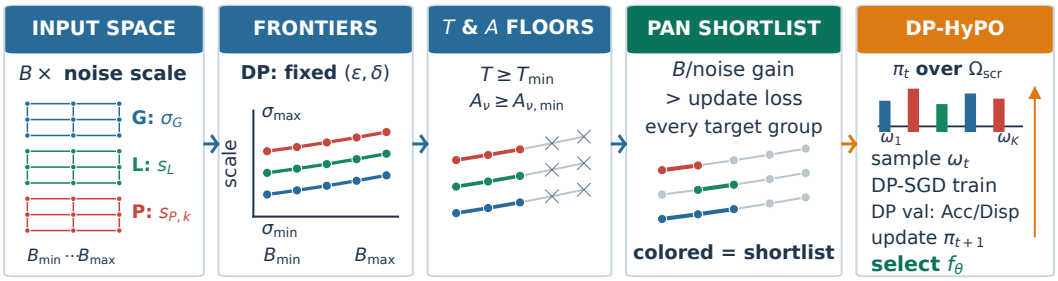


Figure 2: EQUICONFIG workflow. The declared grid contains batch sizes, clipping bounds, learning rates, and Gaussian (G), Lap$_2$ (L), or PLRV-O (P) noise choices. For each noise choice and batch size, mechanism-specific accounting sets the scale required by the training privacy target; scale is not an independently tuned grid parameter. Minimum total and group-active update counts remove infeasible points. PAN retains feasible batch increases that improve the group-information score for every target group, and within-branch Pareto pruning forms the shortlist. DP-HYPO samples this fixed menu, trains and privately evaluates models, updates its proposal π_t, and selects by measured disparity subject to an accuracy requirement. Curves are schematic and have incomparable vertical scales.

3.1 OFFLINE CONSTRUCTION OF THE SHORTLIST

Fix N training records, an epoch budget E, and a per-candidate training target $(\varepsilon_{\text{train}}, \delta_{\text{train}})$. Before examining outcomes, declare a finite grid Ω of configurations $\omega = (j, B, C, \eta)$: noise choice j specifies a family and any discrete shape, B is expected batch size, C is the clipping bound, and η is the learning rate. The choices include Gaussian, Lap$_2$ (Mohammady et al., 2026), and declared PLRV-O shapes (Yang et al., 2025). For fixed (j, C, η), varying B defines a *branch*. Under Poisson sampling, $q(B) = B/N$ and $T(B) = \lceil EN/B \rceil$; the declared batches satisfy $0 < B \le N$.

Let s be the coordinate standard deviation of noise added to the clipped gradient sum, normalized by C; for Gaussian noise, $s = \sigma$. For each j and B, invert the mechanism's accountant:

$$s_{j,\star}(B) = \inf\left\{ s > 0 : \varepsilon_{\text{acc},j}\big(q(B), T(B), s; \delta_{\text{train}}\big) \le \varepsilon_{\text{train}} \right\}. \tag{2}$$

An implemented candidate uses a scale verified to meet this bound. The pairs $(B, s_{j,\star}(B))$ form a fixed-privacy frontier for noise choice j. Mechanism parameterizations and accounting appear in Appendix B.

Larger batches may improve group signal relative to noise in an update, but fixed E leaves fewer updates. For each target group $\nu \in \mathcal{G}$ with $n_\nu > 0$ training examples, let $K_\nu \sim \text{Binomial}(n_\nu, q(B))$ count its examples in an update. The probability of a *group-active* update, its expected number over training, and the expected group count when active are

$$\begin{aligned} p_\nu(B) &= \Pr(K_\nu > 0) = 1 - \big(1 - q(B)\big)^{n_\nu}, \\ A_\nu(B) &= T(B) p_\nu(B), \qquad \bar{K}_\nu(B) = \frac{n_\nu q(B)}{p_\nu(B)}. \end{aligned} \tag{3}$$

Thus A_ν counts opportunities for the group to affect distinct updates, whereas $\bar{K}_\nu$ measures its contribution within an active update.

Let $\mu_\nu(C) > 0$ denote a branch-fixed proxy, in gradient units, for the coherent retained contribution of one group example. Its value is not estimated during screening. We define the *cumulative group-information score*

$$\begin{aligned} \widetilde{\mathcal{I}}_{\nu,j}(B, C) &= A_\nu(B) \left[\frac{\bar{K}_\nu(B) \mu_\nu(C)}{C s_{j,\star}(B)} \right]^2 \\ &= \frac{T(B)[n_\nu q(B)]^2}{p_\nu(B) s_{j,\star}(B)^2} \left[\frac{\mu_\nu(C)}{C} \right]^2. \end{aligned} \tag{4}$$

It combines group-active updates with a squared per-active-update signal-to-noise proxy. Because screening compares batch sizes only within a fixed (j, C, η) branch, the positive $[\mu_\nu(C)/C]^2$ factor cancels. The implemented comparisons therefore use only counts and accountant-calibrated noise; they do not require gradients or checkpoints. The score is a screening diagnostic, not an accuracy or disparity guarantee. Appendix C.1 proves the cancellation.

The score also explains why a favorable batch range depends on group size. For interpretation, replace $T(B)$ by EN/B and $p_\nu(B)$ by $1 - e^{-x_\nu}$, where $x_\nu = n_\nu B/N$. If $a_j(B) = d\log s_{j,\star}(B)/d\log B$ is the local noise-frontier elasticity, the resulting smooth score obeys

$$\frac{d\log \mathcal{I}^{\mathrm{sm}}_{\nu,j}}{d\log B} = 1 - \frac{x_\nu}{e^{x_\nu} - 1} - 2a_j(B). \tag{5}$$

For a sparsely active group ($x_\nu \ll 1$), a batch increase helps only if normalized noise grows very slowly or falls. When a group appears in nearly every update ($x_\nu \gg 1$), the score can improve while noise grows more slowly than $\sqrt{B}$. If elasticity is constant with $0 < a_j < 1/2$, the smooth score improves for every group only beyond $B > \max_{\nu\in\mathcal{G}} Nx_\star/n_\nu$, where $x_\star/(e^{x_\star} - 1) = 1 - 2a_j$; the update floors impose an upper batch limit (Corollary 2). These are predictions about the score, not model accuracy; the finite test below uses exact Poisson participation and integer update counts. The derivation is in Appendix C.

This favorable range need not persist as training grows longer. For the analytic accountants studied in Theorem 3, at each fixed declared batch size, the calibrated squared noise scale grows asymptotically in proportion to EB, while the group-information score approaches a plateau proportional to $1/p_\nu(B)$. Thus, for two fixed batch sizes in the same branch, the larger one does not improve the limiting score. This result delimits the finite-epoch opportunity that PAN tests; it is not a substitute for calibrating and comparing the actual grid points.

Predeclare minimum total updates $T_{\min}$ and minimum expected group-active updates $A_{\nu,\min}$ for each target group. After accountant calibration, retain only configurations satisfying

$$T(B) \geq T_{\min}, \qquad A_\nu(B) \geq A_{\nu,\min} \quad \text{for every } \nu \in \mathcal{G}. \tag{6}$$

Call this feasible set Ω_{req}; a point without a certified noise scale is excluded. These are predeclared training-opportunity requirements, not accuracy guarantees. They also constrain the admissible grid before any score comparison: $A_\nu(B) < n_\nu(E+q(B)) \leq n_\nu(E+1)$, so an activity floor at or above $n_\nu(E+1)$ is impossible for that group (Lemma 3).

Within each branch, order the feasible batches and compare adjacent $B^- < B^+$. PAN is the finite log-score slope

$$\mathrm{PAN}_{\nu,j}(B^-, B^+; C) = \frac{\log \widetilde{\mathcal{I}}_{\nu,j}(B^+, C) - \log \widetilde{\mathcal{I}}_{\nu,j}(B^-, C)}{\log(B^+/B^-)}. \tag{7}$$

The pair passes when this value is positive for every $\nu \in \mathcal{G}$. Both endpoints of each passing pair enter Ω_{Eq}; consecutive passing pairs trace a favorable frontier segment.

Theorem 1 (Exact finite PAN criterion)**.** *For adjacent feasible batches $B_m < B_{m+1}$ within a branch, write $s_m = s_{j,\star}(B_m)$, $T_m = T(B_m)$, and $p_{\nu,m} = p_\nu(B_m)$, with analogous notation at $m+1$. The pair passes PAN if and only if*

$$\frac{s_{m+1}}{s_m} < \min_{\nu\in\mathcal{G}} \frac{B_{m+1}}{B_m} \sqrt{\frac{T_{m+1}p_{\nu,m}}{T_m p_{\nu,m+1}}}. \tag{8}$$

For Gaussian noise, $s_m = \sigma_m$.

The criterion states how slowly calibrated noise must grow for a batch increase to offset the change in group-active updates. The proof is in Appendix C.3.

Finally, within each branch, remove points dominated in every group score and total update count, with at least one strict inequality. The remaining points form Ω_{scr}. Pruning stays within branches because scores across noise choices and clipping bounds do not rank measured training outcomes. Every passing pair remains represented by a shortlisted point with at least the larger endpoint's group scores and update count (Proposition 7). This preserves a score gain, not a measured-performance

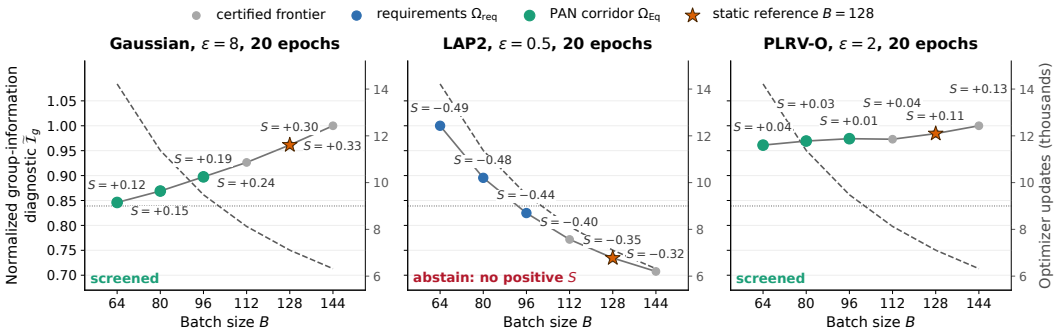


Figure 3: MNIST reference-population screening example. The reference population has $N = 54{,}207$ records, including $n_8 = 58$ class-8 records, with $E = 20$ and training target $(\varepsilon, \delta) = (3, 10^{-5})$. The left panel shows each mechanism's group-information score relative to its own value at $B = 128$; scores across mechanisms are not compared. Gaussian, Lap$_2$, and PLRV-O with $k = 16$ are shown. The middle and right panels show total and class-8-active update counts; dashed lines mark $T_{\min} = 2{,}000$ and $A_{8,\min} = 800$. The shaded $B = 1024$ point fails both floors. Every feasible batch increase lowers the class-8 score, so PAN retains no alternative and the static Gaussian $B = 128$ reference remains. These offline calculations use the illustrative population in Table 5, not the executed experiment's preprocessing counts or measured model outcomes.

gain. If the shortlist is empty, EQUICONFIG does not launch private calibration. The batch grid is fixed before screening, so it can miss a favorable interval between declared sizes. Noise calibration is reusable across the (C, η) choices at each (j, B) pair; after those accountant inversions, floors and group scores require $O(|\Omega||\mathcal{G}|)$ arithmetic, apart from sorting and Pareto pruning. A denser offline grid improves the chance of resolving a favorable interval without requiring every grid point to be trained. Appendix C.3 gives the complete procedure and metadata assumptions.

3.2 PRIVATE CALIBRATION OF THE SHORTLIST

Screening cannot determine which surviving clipping bound, learning rate, or noise family yields the lowest measured disparity. For a predeclared minimum overall accuracy $U_{\min}$, the selection target is

$$\min_{\omega \in \Omega_{\text{scr}}} \text{Disp}(f_\omega) \quad \text{subject to} \quad U(f_\omega) \geq U_{\min}, \tag{9}$$

where f_ω is the model trained under ω and U is validation accuracy. The accuracy gate and disparity ranking use privacy-protected validation measurements. If no evaluated model meets the gate, no model is selected.

DP-HYPO (Wang et al., 2023) samples candidates from the fixed Ω_{scr}, trains them, uses private validation to update its proposal π_t, and selects from the evaluated models. Predeclare an end-to-end target $(\varepsilon_{\text{hypo}}, \delta_{\text{hypo}})$ and allocation ratio $r \in (0, 1)$ with $\varepsilon_{\text{train}} = r\varepsilon_{\text{hypo}}$. Its positive-geometric search horizon T_{H} has mean $1/\gamma$. The joint privacy bound for candidate training and private accuracy–disparity evaluation, together with proposal constraints and any separately accounted group metadata, must certify γ within the end-to-end budget. The ratio r alone does not certify that guarantee. If no horizon is certified, private calibration does not run. The declared batch grid sets offline resolution; T_{H} limits how many shortlisted models can be evaluated. Appendix F.1 gives the evaluator and privacy-accounting details.

The shortlist also changes how that limited search is spent. If the full-menu privacy bound is feasible and the evaluator bounds, privacy budget, and proposal constraints are held fixed, restricting to a nonempty menu cannot shorten the certified expected search horizon; it may leave the horizon unchanged (Proposition 11). With the same lower proposal density bound, fewer shortlisted candidates raise a lower bound on the chance of visiting any favorable configuration that survived screening (Proposition 12). Neither statement guarantees that screening retained the best model or that a smaller menu alone reduces privacy cost.

4 EXPERIMENTS

We next conduct experiments on vision, language, and tabular classification tasks. We briefly describe key aspects of our experiments. Further experimental details are provided in Appendix F.3.

4.1 EXPERIMENT SETUP

Baselines & Datasets: We compare EQUICONFIG with DP-SGD (Abadi et al., 2016) and two clipping-based disparity-mitigation methods, DPSGD-F (Xu et al., 2021) and DPSGD-G-A (Esipova et al., 2023) with evaluation spanning over imbalanced MNIST, CIFAR-10, SST-2, and QNLI, together with CelebA and Adult. Following Bagdasaryan et al. (2019), imbalanced MNIST retains $\sim$15% of class 8 (minority) against class 2 (majority); CIFAR-10 retains $\sim$20% of the cat class against automobile. SST-2 and QNLI retain 30% of negative-sentiment and not-entailment examples, respectively. CelebA and Adult use eyeglasses and gender as protected attributes.

	Private Calibration				Exhaustive Search			
Method	**Minority Acc. ↑**	**Majority Acc. ↑**	**Overall Acc. ↑**	**Disparity ↓**	**Minority Acc. ↑**	**Majority Acc. ↑**	**Overall Acc. ↑**	**Disparity ↓**
EQUICONFIG (ours)	**66.66**	**95.47**	**93.56**	**28.81**	**87.10**	**97.28**	**97.85**	**10.18**
DP-SGD	53.35	93.31	92.43	39.96	75.05	94.86	94.67	19.81
DPSGD-F	52.11	90.01	85.81	37.90	74.12	92.33	92.18	18.21
DPSGD-G-A	38.72	89.52	86.25	50.80	54.03	92.14	90.05	38.11
Non-private	93.42	98.44	98.39	5.02	93.42	98.44	98.39	5.02

Table 2: Performance comparison of EQUICONFIG on CNN-MNIST under a fixed model privacy budget ($\varepsilon_{train} = 3, \delta = 10^{-5}$). Results are reported for *Private Calibration* ($\varepsilon_{\text{hypo}} = 6.6, r = 0.45$) and *Exhaustive Search*, where selection carries no extra privacy cost. Bold marks the best result among private methods.

	Private Calibration				Exhaustive Search			
Method	**Minority Acc. ↑**	**Majority Acc. ↑**	**Overall Acc. ↑**	**Disparity ↓**	**Minority Acc. ↑**	**Majority Acc. ↑**	**Overall Acc. ↑**	**Disparity ↓**
EQUICONFIG (ours)	**62.78**	**96.29**	**92.24**	**33.51**	**71.30**	**97.90**	**92.92**	**26.60**
DP-SGD	27.92	77.18	74.10	49.26	46.90	97.40	90.48	50.50
DPSGD-F	52.72	86.71	82.61	33.99	63.80	94.80	90.06	31.00
DPSGD-G-A	27.46	83.41	75.99	55.95	48.60	94.80	87.52	46.20
Non-private	81.80	97.70	94.94	15.90	81.80	97.70	94.94	15.90

Table 3: Performance comparison of EQUICONFIG on TinyViT-CIFAR10 ($\varepsilon_{train} = 5$, 3-epoch fine-tuning) under a fixed model privacy budget. Results are reported for *Private Calibration* ($\varepsilon_{\text{hypo}} = 13.9, r = 0.45$) and *Exhaustive Search*, where selection carries no extra privacy cost. Bold marks the best result among methods.

Values (%)

	ε = 0.5								**ε = 1**							
	Fine-tuning				**From Scratch**				**Fine-tuning**				**From Scratch**			
	Minority Acc. ↑	**Majority Acc. ↑**	**Overall Acc. ↑**	**Disparity ↓**	**Minority Acc. ↑**	**Majority Acc. ↑**	**Overall Acc. ↑**	**Disparity ↓**	**Minority Acc. ↑**	**Majority Acc. ↑**	**Overall Acc. ↑**	**Disparity ↓**	**Minority Acc. ↑**	**Majority Acc. ↑**	**Overall Acc. ↑**	**Disparity ↓**
EquiConfig	74	95	85	21	80	95	88	14	76	95	86	19	82	95	89	13
DP-SGD	3	100	52	97	66	94	81	28	59	95	77	36	68	95	82	28
DPSGD-F	0	100	51	100	76	91	84	15	67	94	80	27	78	91	85	13
DPSGD-G-A	46	97	72	52	59	96	78	37	62	95	79	34	63	96	80	34
Non-private	85	96	91	11	83	96	90	13	85	96	91	11	83	96	90	13

Figure 4: Performance comparison of EQUICONFIG on imbalanced SST-2 using BERT-base-uncased, with 5-epoch fine-tuning and 20-epoch training from scratch. Results are reported for $\varepsilon_{train} \in \{0.5, 1\}$ for exhaustive search with reduced configuration space. Values are rounded to the nearest integer.

Implementation & Privacy budgets: We used PyTorch on NVIDIA A100 GPUs with $\delta = 10^{-5}$. Training budgets $(\varepsilon_{\text{train}}, E)$ are $(3, 20)$ for MNIST, $(5, 3)$ for CIFAR-10, $(4, 20)$ for Adult, and $(0.5, 1)$ for CelebA, following prior disparate-impact and private fine-tuning conventions for from-scratch versus pretrained settings. BERT-based SST-2 uses $\varepsilon_{train} \in \{0.5, 1\}$ (5-epoch fine-tuning / 20-epoch from scratch); prompt-based RoBERTa experiments use $\varepsilon_{train} = 8$ (SST-2) and $\varepsilon_{train} = 20$ (QNLI) for 100 epochs. The full configuration space, including method-specific parameters and the noise/batch/clipping grid used by EQUICONFIG, is given in Table 9 (Appendix F.3).

Selection Regimes. We evaluate the screened shortlist under two regimes. *Exhaustive search* trains and evaluates every shortlisted candidate without an additional selection-privacy cost, isolating the

Method	SST-2 ($\varepsilon = 8$) Minority Acc. ↑	Majority Acc. ↑	Overall Acc. ↑	Disparity ↓	QNLI ($\varepsilon = 20$) Minority Acc. ↑	Majority Acc. ↑	Overall Acc. ↑	Disparity ↓
EQUICONFIG (ours)	**83.18**	95.72	**89.56**	**12.54**	**59.55**	96.30	**76.72**	**36.75**
DP-SGD	77.80	95.05	86.58	17.24	55.85	**96.97**	76.19	41.12
DPSGD-F	76.64	94.82	85.89	18.18	50.96	91.86	71.19	40.90
DPSGD-G-A	73.13	**95.95**	84.75	22.82	32.81	95.08	63.61	62.26
Non-private	89.95	94.59	92.32	4.64	72.84	96.74	84.66	23.91

Table 4: Performance comparison of EQUICONFIG under prompt-based training on imbalanced SST-2 and QNLI. Results are reported for SST-2 ($\varepsilon_{train} = 8$) and QNLI ($\varepsilon_{train} = 20$) for 100 epochs. Selection *exhaustive search* over the reduced configuration spaces described in Appendix F.3.1. Bold marks the best result among private methods.

benefit of offline screening. *Private calibration* uses privacy-accounted DP-HyPO (Wang et al., 2023), with selection cost charged against an end-to-end budget $\varepsilon_{\text{hypo}}$; this is the setting relevant to deployment. We set $\varepsilon_{\text{train}} = r\,\varepsilon_{\text{hypo}}$ with $r \in \{0.45, 0.36, 0.36, 0.36\}$ for MNIST, CIFAR-10, Adult, and CelebA, giving $\varepsilon_{\text{hypo}} \in \{6.6, 13.9, 11.4, 1.4\}$. Large language model tasks are evaluated via exhaustive search on their reduced configuration space.

4.2 RESULTS

Exhaustive Search. We first evaluate the screened configuration space using exhaustive search. On CNN-MNIST, Table 2 shows that EQUICONFIG achieves 97.85% overall accuracy with 10.18% disparity, compared with 94.67% and 19.81% for DP-SGD. Whereas DPSGD-F and DPSGD-G-A achieve 18.211% and 38.11% disparity, respectively. On TinyViT-CIFAR10, Table 3 shows a similar advantage for EQUICONFIG, which achieves 92.92% overall accuracy and 26.60% disparity, compared with 90.48% and 50.50% for DP-SGD. Here, DPSGD-F reduces disparity to 31.00%, whereas DPSGD-G-A obtains 46.20%. Figure 4 further evaluates SST-2 using BERT under pretrained fine-tuning and training from scratch at $\varepsilon_{train} \in \{0.5, 1\}$. At $\varepsilon_{train} = 0.5$, EQUICONFIG achieves 74% minority accuracy and 21% disparity under fine-tuning, and 80% minority accuracy and 14% disparity when training from scratch. At $\varepsilon_{train} = 1$, the corresponding values are 76%, 19% and 82%, 13%, respectively. For prompt-based language classification, Table 4 shows that EQUICONFIG achieves 89.56% overall accuracy and 12.54% disparity on SST-2, compared with 86.58% and 17.24% for DP-SGD. On QNLI, EQUICONFIG achieves 76.72% overall accuracy and 36.75% disparity, compared with 76.19% and 41.12% for DP-SGD, 71.19% and 40.90% for DPSGD-F, and 63.61% and 62.26% for DPSGD-G-A, respectively. These results show that EQUICONFIG maintains a favorable accuracy–disparity trade-off under prompt-based language-model training. Additional results for MLP-Adult and ViT-CelebA are reported in Appendix F.4.

We note that prior evaluations of DPSGD-F/DPSGD-G-A use a single fixed clipping bound (e.g., C=1), whereas our search sweeps clipping for every method, including DP-SGD, and restricts each baseline's additional parameters to two values to bound training cost (Table 9). This wider sweep is why the relative ordering of these baselines against DP-SGD can differ from their original single-clip comparisons. EQUICONFIG instead jointly searches noise mechanism, batch, and clipping, with PAN and the update/group-activity floors reducing this larger space to a tractable shortlist (Table 12).

Private Calibration. On CNN-MNIST, Table 2 shows that EQUICONFIG achieves approximately 94% overall accuracy and 28% disparity, compared with 91% and 47% for DP-SGD. On TinyViT-CIFAR10, Table 3 reports 92.24% overall accuracy and 33.51% disparity for EQUICONFIG, compared with 90.48% and 50.50% for DP-SGD. The Adult and CelebA results in Appendix F.4 Tables 11 and 10 remain close to their exhaustive-search counterparts. Overall, these results show that EQUICONFIG retains strong accuracy–disparity trade-offs even with private candidate selection.

4.3 ABLATION STUDIES

We examine the robustness of EQUICONFIG across different imbalance, minority-identity, and calibration settings, and analyze the contribution of its screening components and computational cost. We first examine the sensitivity of EQUICONFIG to the severity and composition of class imbalance. Figure 5 varies the minority-class proportion from 1% to 50% and shows that minority accuracy generally increases and disparity decreases as the minority proportion grows across methods. The differences are most pronounced under severe imbalance, where EQUICONFIG maintains a favor-

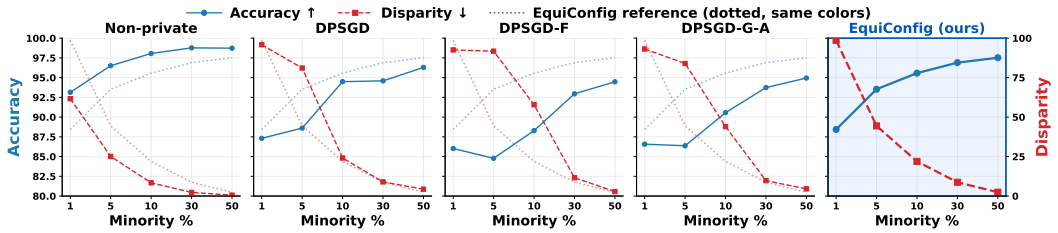
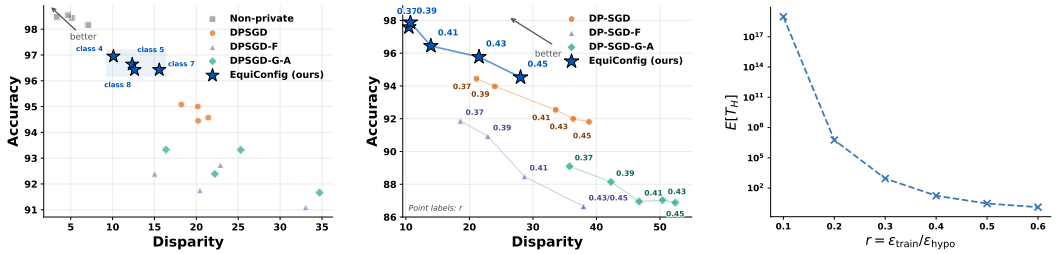


Figure 5: Effect of varying the minority-class proportion on accuracy and disparity on CNN-MNIST($\varepsilon_{train} =$ 3,E=20). Each panel shows a different method, with dotted curves as EQUICONFIG reference.



(a) Effect of minority-class identity under a fixed imbalance ratio.

(b) Accuracy–disparity trade-off under different privacy-budget allocation ratios.

(c) Expected search horizon $E[T_H]$ under different privacy-budget allocation ratios.

Figure 6: Evaluating the effect of minority-class identity and privacy-budget allocation on EQUICONFIG performance on CNN-MNIST ($\varepsilon_{train} = 3, E = 20$) and private calibration expected search horizons.

able accuracy–disparity trade-off. Figure 6(a) further varies the minority-class identity under a fixed imbalance ratio. Across classes 4, 5, 7, and 8, EQUICONFIG exhibits similar accuracy–disparity trade-offs, although the absolute performance varies with the minority class, indicating that its behavior is not tied to a particular minority-class identity.

We next examine how the privacy budget is allocated between private model training and private configuration selection. We fix $\varepsilon_{\text{train}}$ and vary r, so that $\varepsilon_{\text{hypo}} = \varepsilon_{\text{train}}/r$ and the remaining $(1 - r)\varepsilon_{\text{hypo}}$ is allocated to DP-HyPO. As shown in Figure 6(b), changing r produces a corresponding accuracy–disparity trade-off by shifting privacy budget between training and candidate selection. Figure 6(c) shows that the expected DP-HyPO search horizon decreases as a larger fraction of the budget is allocated to training. Together, these results characterize the trade-off between model-training utility and the extent of private configuration search under a fixed total privacy budget.

Screening Components and Computational Cost. The full pipeline of EQUICONFIG reduces the initial candidate space to 72 on CNN-MNIST; removing PAN or the update constraints degrades private-calibration performance, while exhaustive search with either component removed converges to the same utility as described in Table 12. The exhaustive-search runtime is 3.21 GPU-hours for EQUICONFIG, compared with 2.76 for DP-SGD, 10.15 for DPSGD-F, and 73.47 for DPSGD-G-A. Although EQUICONFIG evaluates fewer candidates, its search includes both Gaussian and PLRV-O noise, so the total runtime reflects both the number of candidates and the training cost of the configurations explored. Further details are provided in Table 12, Appendix F.4.

5 CONCLUSION

We introduced EQUICONFIG, which screens existing DP-SGD batch-size, clipping, and noise choices before private model calibration. It retains configurations with sufficient training opportunities and improved cumulative group information, then selects a model using measured accuracy and disparity. Across the evaluated vision, tabular, and language tasks, EQUICONFIG improves accuracy–disparity trade-offs at matched training privacy targets while narrowing the set of candidates for private evaluation. Its screening score directs calibration toward favorable noise–sampling balances without introducing additional training-time controls, expanding the practical design space for reliable and useful private learning.

AI USE STATEMENT

We used generative AI tools, including ChatGPT/Codex, to help develop and critique the EQUICONFIG framework, review mathematical claims and proof drafts, refine experimental methodology, and interpret results. These tools also assisted with implementation, schematic figure design, and drafting and editing the manuscript. The authors take responsibility for the final methods, proofs, code, figures, and reported results.

ETHICS STATEMENT

This work studies group-accuracy disparities in private learning. The reported disparity measures differences only among the specified groups on the evaluated tasks; reducing it does not establish fairness for unmeasured groups or settings. We report aggregate group results for benchmarks that include attribute-defined groups. The privacy claims depend on the stated adjacency and accounting assumptions, including the treatment of group counts and private model selection.

REPRODUCIBILITY STATEMENT

Section 3 specifies the screening and calibration procedure. Appendix B gives privacy-accounting details and mathematical derivations; Section 4 and Appendix F.3 describe the datasets, model settings, and evaluation protocol. An anonymized source-code artifact is available at anonymous.4open.science.

REFERENCES

Martin Abadi, Andy Chu, Ian Goodfellow, H. Brendan McMahan, Ilya Mironov, Kunal Talwar, and Li Zhang. Deep learning with differential privacy. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, CCS '16, pp. 308–318, New York, NY, USA, 2016. Association for Computing Machinery. ISBN 9781450341394. doi: 10.1145/2976749.2978318. URL https://doi.org/10.1145/2976749.2978318.

Brendan Avent, Javier González, Tom Diethe, Andrei Paleyes, and Borja Balle. Automatic discovery of privacy-utility pareto fronts. *Proc. Priv. Enhancing Technol.*, 2020(4):5–23, 2020. doi: 10.2478/POPETS-2020-0060. URL https://doi.org/10.2478/popets-2020-0060.

Eugene Bagdasaryan, Omid Poursaeed, and Vitaly Shmatikov. Differential privacy has disparate impact on model accuracy. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett (eds.), *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019. URL https://proceedings.neurips.cc/paper_files/paper/2019/file/fc0de4e0396fff257ea362983c2dda5a-Paper.pdf.

Alex Beutel, Jilin Chen, Zhe Zhao, and Ed H. Chi. Data decisions and theoretical implications when adversarially learning fair representations, 2017. URL https://arxiv.org/abs/1707.00075.

Zhiqi Bu, Yu-Xiang Wang, Sheng Zha, and George Karypis. Automatic clipping: Differentially private deep learning made easier and stronger. In A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine (eds.), *Advances in Neural Information Processing Systems*, volume 36, pp. 41727–41764. Curran Associates, Inc., 2023. doi: 10.52202/075280-1808. URL https://proceedings.neurips.cc/paper_files/paper/2023/file/8249b30d877c91611fd8c7aa6ac2b5fe-Paper-Conference.pdf.

Mateusz Buda, Atsuto Maki, and Maciej A. Mazurowski. A systematic study of the class imbalance problem in convolutional neural networks. *Neural Netw.*, 106(C):249–259, October 2018. ISSN 0893-6080. doi: 10.1016/j.neunet.2018.07.011. URL https://doi.org/10.1016/j.neunet.2018.07.011.

Toon Calders, Faisal Kamiran, and Mykola Pechenizkiy. Building classifiers with independency constraints. In *Proceedings of the 2009 IEEE International Conference on Data Mining Workshops*, ICDMW '09, pp. 13–18, USA, 2009. IEEE Computer Society. ISBN 9780769539027. doi: 10.1109/ICDMW.2009.83. URL https://doi.org/10.1109/ICDMW.2009.83.

Chen Chen and Jaewoo Lee. Stochastic adaptive line search for differentially private optimization. In Xintao Wu, Chris Jermaine, Li Xiong, Xiaohua Hu, Olivera Kotevska, Siyuan Lu, Weija Xu, Srinivas Aluru, Chengxiang Zhai, Eyhab Al-Masri, Zhiyuan Chen, and Jeff Saltz (eds.), *2020 IEEE International Conference on Big Data (IEEE BigData 2020), Atlanta, GA, USA, December 10-13, 2020*, pp. 1011–1020. IEEE, 2020. doi: 10.1109/BIGDATA50022.2020.9378011. URL https://doi.org/10.1109/BigData50022.2020.9378011.

Sai Venkatesh Chilukoti, Md Imran Hossen, Liqun Shan, Vijay Srinivas Tida, Mahathir Mohammad Bappy, Wenmeng Tian, and Xiali Hei. Dp-sgd-global-adapt-v2-s: Triad improvements of privacy, accuracy and fairness via step decay noise multiplier and step decay upper clipping threshold. *Electron. Commer. Rec. Appl.*, 70(C), March 2025. ISSN 1567-4223. doi: 10.1016/j.elerap.2025.101476. URL https://doi.org/10.1016/j.elerap.2025.101476.

Youlong Ding and Xueyang Wu. Revisiting hyperparameter tuning with differential privacy, 2025. URL https://arxiv.org/abs/2211.01852.

Friedrich Dörmann, Osvald Frisk, Lars Nørvang Andersen, and Christian Fischer Pedersen. Not all noise is accounted equally: How differentially private learning benefits from large sampling rates. In *2021 IEEE 31st International Workshop on Machine Learning for Signal Processing (MLSP), Gold Coast, Australia, October 25-28, 2021*, pp. 1–6. IEEE, 2021. doi: 10.1109/MLSP52302.2021.9596307. URL https://doi.org/10.1109/MLSP52302.2021.9596307.

Vadym Doroshenko, Badih Ghazi, Pritish Kamath, Ravi Kumar, and Pasin Manurangsi. Connect the dots: Tighter discrete approximations of privacy loss distributions. *Proc. Priv. Enhancing Technol.*, 2022(4):552–570, 2022. doi: 10.56553/POPETS-2022-0122. URL https://doi.org/10.56553/popets-2022-0122.

Cynthia Dwork. Differential privacy. In Michele Bugliesi, Bart Preneel, Vladimiro Sassone, and Ingo Wegener (eds.), *Automata, Languages and Programming*, pp. 1–12, Berlin, Heidelberg, 2006. Springer Berlin Heidelberg. ISBN 978-3-540-35908-1. URL https://doi.org/10.1007/11787006_1.

Maria S. Esipova, Atiyeh Ashari Ghomi, Yaqiao Luo, and Jesse C. Cresswell. Disparate impact in differential privacy from gradient misalignment. In *The Eleventh International Conference on Learning Representations, ICLR 2023, Kigali, Rwanda, May 1-5, 2023*. OpenReview.net, 2023. URL https://openreview.net/forum?id=qLOaeRvteqbx.

Tom Farrand, Fatemehsadat Mireshghallah, Sahib Singh, and Andrew Trask. Neither private nor fair: Impact of data imbalance on utility and fairness in differential privacy. In *Proceedings of the 2020 Workshop on Privacy-Preserving Machine Learning in Practice*, PPMLP'20, pp. 15–19, New York, NY, USA, 2020. Association for Computing Machinery. ISBN 9781450380881. doi: 10.1145/3411501.3419419. URL https://doi.org/10.1145/3411501.3419419.

Michael Feldman, Sorelle A. Friedler, John Moeller, Carlos Scheidegger, and Suresh Venkatasubramanian. Certifying and removing disparate impact. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '15, pp. 259–268, New York, NY, USA, 2015. Association for Computing Machinery. ISBN 9781450336642. doi: 10.1145/2783258.2783311. URL https://doi.org/10.1145/2783258.2783311.

Antti Koskela and Antti Honkela. Learning rate adaptation for differentially private learning. In Silvia Chiappa and Roberto Calandra (eds.), *The 23rd International Conference on Artificial Intelligence and Statistics, AISTATS 2020, 26-28 August 2020, Online [Palermo, Sicily, Italy]*, volume 108 of *Proceedings of Machine Learning Research*, pp. 2465–2475. PMLR, 2020. URL http://proceedings.mlr.press/v108/koskela20a.html.

Antti Koskela and Tejas D Kulkarni. Practical differentially private hyperparameter tuning with subsampling. In A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine (eds.), *Advances in Neural Information Processing Systems*, volume 36, pp. 28201–28225. Curran Associates, Inc., 2023. doi: 10.52202/075280-1226. URL https://proceedings.neurips.cc/paper_files/paper/2023/file/59b9582cd35f555ea8415030073e7b22-Paper-Conference.pdf.

Antti Koskela, Joonas Jälkö, and Antti Honkela. Computing tight differential privacy guarantees using fft. In Silvia Chiappa and Roberto Calandra (eds.), *Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics*, volume 108 of *Proceedings of Machine Learning Research*, pp. 2560–2569. PMLR, 26–28 Aug 2020. URL https://proceedings.mlr.press/v108/koskela20b.html.

Ruixuan Liu and Zhiqi Bu. Towards hyperparameter-free optimization with differential privacy. In *The Thirteenth International Conference on Learning Representations, ICLR 2025, Singapore, April 24-28, 2025*. OpenReview.net, 2025. URL https://openreview.net/forum?id=2kGKsyhtvh.

Ryan McKenna, Yangsibo Huang, Amer Sinha, Borja Balle, Zachary Charles, Christopher A. Choquette-Choo, Badih Ghazi, Georgios Kaissis, Ravi Kumar, Ruibo Liu, Da Yu, and Chiyuan Zhang. Scaling laws for differentially private language models. In Aarti Singh, Maryam Fazel, Daniel Hsu, Simon Lacoste-Julien, Felix Berkenkamp, Tegan Maharaj, Kiri Wagstaff, and Jerry Zhu (eds.), *Forty-second International Conference on Machine Learning, ICML 2025, Vancouver, BC, Canada, July 13-19, 2025*, volume 267 of *Proceedings of Machine Learning Research*. PMLR / OpenReview.net, 2025. URL https://proceedings.mlr.press/v267/mckenna25a.html.

Ilya Mironov, Kunal Talwar, and Li Zhang. Rényi differential privacy of the sampled gaussian mechanism, 2019a. URL https://arxiv.org/abs/1908.10530.

Ilya Mironov, Kunal Talwar, and Li Zhang. Rényi differential privacy of the sampled gaussian mechanism. *arXiv preprint arXiv:1908.10530*, 2019b. URL https://arxiv.org/abs/1908.10530.

Meisam Mohammady, Qin Yang, Nicholas Stout, Ayesha Samreen, Han Wang, Christopher J Quinn, and Yuan Hong. LAP2: Revisiting Laplace DP-SGD for High Dimensions via Majorization Theory . In *2026 IEEE 39th Computer Security Foundations Symposium (CSF)*, pp. 189–204, Los Alamitos, CA, USA, July 2026. IEEE Computer Society. doi: 10.1109/CSF68417.2026.00013. URL https://doi.ieeecomputersociety.org/10.1109/CSF68417.2026.00013.

Felix Morsbach, Jan Reubold, and Thorsten Strufe. R+R: understanding hyperparameter effects in DP-SGD. In *Annual Computer Security Applications Conference, ACSAC 2024, Honolulu, HI, USA, December 9-13, 2024*, pp. 1217–1230. IEEE, 2024. doi: 10.1109/ACSAC63791.2024.00097. URL https://doi.org/10.1109/ACSAC63791.2024.00097.

Ashwinee Panda, Xinyu Tang, Saeed Mahloujifar, Vikash Sehwag, and Prateek Mittal. A new linear scaling rule for private adaptive hyperparameter optimization. In Ruslan Salakhutdinov, Zico Kolter, Katherine A. Heller, Adrian Weller, Nuria Oliver, Jonathan Scarlett, and Felix Berkenkamp (eds.), *Forty-first International Conference on Machine Learning, ICML 2024, Vienna, Austria, July 21-27, 2024*, volume 235 of *Proceedings of Machine Learning Research*, pp. 39364–39399. PMLR / OpenReview.net, 2024. URL https://proceedings.mlr.press/v235/panda24a.html.

Nicolas Papernot and Thomas Steinke. Hyperparameter tuning with renyi differential privacy. In *The Tenth International Conference on Learning Representations, ICLR 2022, Virtual Event, April 25-29, 2022*. OpenReview.net, 2022. URL https://openreview.net/forum?id=-70L8lpp9DF.

Nicolas Papernot, Steve Chien, Shuang Song, Abhradeep Thakurta, and Ulfar Erlingsson. Making the shoe fit: Architectures, initializations, and tuning for learning with privacy, 2020. URL https://openreview.net/forum?id=rJg851rYwH.

Felix Petersen, Debarghya Mukherjee, Yuekai Sun, and Mikhail Yurochkin. Post-processing for individual fairness. In Marc'Aurelio Ranzato, Alina Beygelzimer, Yann N. Dauphin, Percy Liang, and Jennifer Wortman Vaughan (eds.), *Advances in Neural Information Processing Systems 34: Annual Conference on Neural Information Processing Systems 2021, NeurIPS 2021, December 6-14, 2021, virtual*, pp. 25944–25955, 2021. URL https://proceedings.neurips.cc/paper/2021/hash/d9fea4ca7e4a74c318ec27c1deb0796c-Abstract.html.

Aman Priyanshu, Rakshit Naidu, Fatemehsadat Mireshghallah, and Mohammad Malekzadeh. Efficient hyperparameter optimization for differentially private deep learning, 2021. URL https://arxiv.org/abs/2108.03888.

Aki Rehn, Linzh Zhao, Mikko Heikkilä, and Antti Honkela. On optimal hyperparameters for differentially private deep transfer learning. In C. Vondrick, B. Hariharan, C. Raffel, L. Pinto, D. Yang, and A. Faust (eds.), *International Conference on Learning Representations*, volume 2026, pp. 85224–85261, 2026. URL https://proceedings.iclr.cc/paper_files/paper/2026/file/895834beb08f2806edf3e998851ec2ff-Paper-Conference.pdf.

Yuji Roh, Kangwook Lee, Steven Euijong Whang, and Changho Suh. Fairbatch: Batch selection for model fairness. In *9th International Conference on Learning Representations, ICLR 2021, Virtual Event, Austria, May 3-7, 2021*. OpenReview.net, 2021. URL https://openreview.net/forum?id=YNnpaAKeCfx.

A. Samreen. Efficient fair learning with subset selection. Master's thesis, The University of Texas at Dallas, 2023. URL https://utd-ir.tdl.org/handle/10735.1/10034.

Tom Sander, Pierre Stock, and Alexandre Sablayrolles. TAN without a burn: Scaling laws of DP-SGD. In Andreas Krause, Emma Brunskill, Kyunghyun Cho, Barbara Engelhardt, Sivan Sabato, and Jonathan Scarlett (eds.), *International Conference on Machine Learning, ICML 2023, 23-29 July 2023, Honolulu, Hawaii, USA*, volume 202 of *Proceedings of Machine Learning Research*, pp. 29937–29949. PMLR, 2023. URL https://proceedings.mlr.press/v202/sander23b.html.

Getoar Sopa, Juraj Marusic, Marco Avella Medina, and John P. Cunningham. Differentially private hyperparameter tuning using local bayesian optimization, 2026. URL https://arxiv.org/abs/2502.06044.

Cuong Tran, My Dinh, and Ferdinando Fioretto. Differentially private empirical risk minimization under the fairness lens. In M. Ranzato, A. Beygelzimer, Y. Dauphin, P.S. Liang, and J. Wortman Vaughan (eds.), *Advances in Neural Information Processing Systems*, volume 34, pp. 27555–27565. Curran Associates, Inc., 2021. URL https://proceedings.neurips.cc/paper_files/paper/2021/file/e7e8f8e5982b3298c8addedf6811d500-Paper.pdf.

Khang Tran, Ferdinando Fioretto, Issa Khalil, My T. Thai, Linh Thi Xuan Phan, and NhatHai Phan. Fairdp: Achieving fairness certification with differential privacy. In *IEEE Conference on Secure and Trustworthy Machine Learning, SaTML 2025, Copenhagen, Denmark, April 9-11, 2025*, pp. 956–976. IEEE, 2025. doi: 10.1109/SATML64287.2025.00058. URL https://doi.org/10.1109/SaTML64287.2025.00058.

Koen Lennart van der Veen, Ruben Seggers, Peter Bloem, and Giorgio Patrini. Three tools for practical differential privacy, 2018. URL https://arxiv.org/abs/1812.02890.

Marten van Dijk, Nhuong V. Nguyen, Toan N. Nguyen, Lam M. Nguyen, and Phuong Ha Nguyen. Proactive DP: A multiple target optimization framework for DP-SGD. In Ruslan Salakhutdinov, Zico Kolter, Katherine A. Heller, Adrian Weller, Nuria Oliver, Jonathan Scarlett, and Felix Berkenkamp (eds.), *Forty-first International Conference on Machine Learning, ICML 2024, Vienna, Austria, July 21-27, 2024*, volume 235 of *Proceedings of Machine Learning Research*, pp. 49029–49077. PMLR / OpenReview.net, 2024. URL https://proceedings.mlr.press/v235/van-dijk24a.html.

Hua Wang, Sheng Gao, Huanyu Zhang, Weijie Su, and Milan Shen. Dp-hypo: An adaptive private framework for hyperparameter optimization. In A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine (eds.), *Advances in Neural Information Processing Systems*, volume 36, pp. 41868–41891. Curran Associates, Inc., 2023. doi: 10.52202/075280-1814. URL https://proceedings.neurips.cc/paper_files/paper/2023/file/82d7d58cba24731c0ca952dff1de46ae-Paper-Conference.pdf.

Chengkun Wei, Weixian Li, Chen Gong, and Wenzhi Chen. DC-SGD: differentially private SGD with dynamic clipping through gradient norm distribution estimation. *IEEE Trans. Inf. Forensics*

Secur., 20:4498–4511, 2025. doi: 10.1109/TIFS.2025.3557755. URL https://doi.org/10.1109/TIFS.2025.3557755.

Ruicheng Xian, Lang Yin, and Han Zhao. Fair and optimal classification via post-processing. In Andreas Krause, Emma Brunskill, Kyunghyun Cho, Barbara Engelhardt, Sivan Sabato, and Jonathan Scarlett (eds.), *International Conference on Machine Learning, ICML 2023, 23-29 July 2023, Honolulu, Hawaii, USA*, volume 202 of *Proceedings of Machine Learning Research*, pp. 37977–38012. PMLR, 2023. URL https://proceedings.mlr.press/v202/xian23b.html.

Zihang Xiang, Tianhao Wang, Cheng-Long Wang, and Di Wang. Revisiting differentially private hyper-parameter tuning. In *Proceedings of the Network and Distributed System Security Symposium*, 2026. doi: 10.14722/ndss.2026.230447. URL https://doi.org/10.14722/ndss.2026.230447.

Depeng Xu, Wei Du, and Xintao Wu. Removing disparate impact on model accuracy in differentially private stochastic gradient descent. In Feida Zhu, Beng Chin Ooi, and Chunyan Miao (eds.), *KDD '21: The 27th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, Virtual Event, Singapore, August 14-18, 2021*, pp. 1924–1932. ACM, 2021. doi: 10.1145/3447548.3467268. URL https://doi.org/10.1145/3447548.3467268.

Qin Yang, Nicholas Stout, Meisam Mohammady, Han Wang, Ayesha Samreen, Christopher J. Quinn, Yan Yan, Ashish Kundu, and Yuan Hong. Plrv-o: Advancing differentially private deep learning via privacy loss random variable optimization. In *Proceedings of the 2025 ACM SIGSAC Conference on Computer and Communications Security*, CCS '25, pp. 306–320, New York, NY, USA, 2025. Association for Computing Machinery. ISBN 9798400715259. doi: 10.1145/3719027.3765151. URL https://doi.org/10.1145/3719027.3765151.

Linzh Zhao, Aki Rehn, Mikko A. Heikkilä, Razane Tajeddine, and Antti Honkela. Mitigating disparate impact of differentially private learning through bounded adaptive clipping. *Transactions on Machine Learning Research*, 2026. ISSN 2835-8856. URL https://openreview.net/forum?id=UlzcKSHVoN.

Yuqing Zhu and Yu-Xiang Wang. Poisson subsampled Rényi differential privacy. In *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, 2019. URL https://proceedings.mlr.press/v97/zhu19c.html.

Yuqing Zhu, Jinshuo Dong, and Yu-Xiang Wang. Optimal accounting of differential privacy via characteristic function. In Gustau Camps-Valls, Francisco J. R. Ruiz, and Isabel Valera (eds.), *Proceedings of The 25th International Conference on Artificial Intelligence and Statistics*, volume 151 of *Proceedings of Machine Learning Research*, pp. 4782–4817. PMLR, 28–30 Mar 2022. URL https://proceedings.mlr.press/v151/zhu22c.html.

A Extended Related Work

The first systematic evidence that DP-SGD induces group-dependent accuracy loss was shown by Bagdasaryan et al. (2019), who demonstrated that privacy noise disproportionately harms underrepresented or high-variance groups. Farrand et al. (2020) further examined this phenomenon under varying levels of group imbalance. A major underlying cause, also highlighted in subsequent studies, is the interaction between *subsampling*, *clipping*, and *noise*—all of which suppress minority-group signal more aggressively than majority-group signal.

Fairness-driven modifications to DP-SGD. A line of work proposes algorithmic adjustments to the DP-SGD procedure. Xu et al. (2021) introduced DPSGD-F, which reweights per-group contributions to compensate for systematic clipping of large-norm gradients from minority or complex groups. Esipova et al. (2023) identified *gradient misalignment*—the deviation between clipped and unclipped gradient directions—as a key source of disparity and proposed DPSGD-G-A, a global adaptive clipping scheme. Follow-up work incorporated noise decay and adaptive clipping schedules (Chilukoti et al., 2025). These methods operate primarily through *clipping adaptation*, and their fairness improvements depend heavily on how well clipping norms align with per-group gradient geometry.

Certified fairness under DP. Other approaches aim to guarantee fairness notions under DP constraints. FairDP (Tran et al., 2025) partitions data into k groups, independently trains DP-SGD models per group, and aggregates them to certify demographic parity or equalized odds. Tran et al. (2021) studied group-wise excessive risk under DP noise and clipping, proposing group-normalized sensitivity and noise to equalize learning difficulty across groups. While effective, these approaches incur nontrivial computational costs due to per-group training or multiple DP runs.

Fairness in non-private ML. Outside of the DP setting, fairness research spans three major categories: (i) *pre-processing* (e.g., reweighting (Calders et al., 2009), disparate impact remover (Feldman et al., 2015), resampling (Buda et al., 2018)), (ii) *in-processing* (e.g., adaptive batch selection (Roh et al., 2021; Samreen, 2023), adversarial fairness objectives (Beutel et al., 2017)), and (iii) *post-processing* (Petersen et al., 2021; Xian et al., 2023). However, these methods generally cannot be applied under DP-SGD, as they either (i) require access to sensitive group information inside the optimization loop, or (ii) violate DP-SGD's strict gradient-sensitivity constraints.

DP-SGD calibration and parameter tuning. A related line of work studies the selection and calibration of DP-SGD training and privacy parameters. van der Veen et al. (2018) proposed adaptive clipping and large-batch training, while Papernot et al. (2020) examined architectural and optimization choices for private learning. The adaptation of the private learning-rate was subsequently explored through extrapolation and stochastic line-search methods (Koskela & Honkela, 2020; Chen & Lee, 2020). More recent methods automate clipping or learning-rate selection (Bu et al., 2023; Wei et al., 2025; Liu & Bu, 2025). Empirical and theoretical studies have also characterized interactions among clipping, learning rate, sampling rate, batch size, training duration, and noise (Dörmann et al., 2021; Morsbach et al., 2024; Rehn et al., 2026; van Dijk et al., 2024), while scaling-law analyses provide guidance for compute-efficient configurations (Sander et al., 2023; McKenna et al., 2025). Search-based approaches use Bayesian, evolutionary, and related optimization methods to select private-training configurations (Avent et al., 2020; Priyanshu et al., 2021). Since evaluating configurations on sensitive data can introduce additional privacy loss, other work develops specialized accounting mechanisms (Papernot & Steinke, 2022; Ding & Wu, 2025), performs tuning on a private-data subsample followed by extrapolation (Koskela & Kulkarni, 2023), or reduces the cost of adaptive private search (Wang et al., 2023; Panda et al., 2024; Sopa et al., 2026). Recent work further audits and tightens the privacy analysis of DP-SGD hyperparameter tuning (Xiang et al., 2026). However, these approaches primarily target aggregate utility, privacy accounting, or computational efficiency and generally do not incorporate group-level disparity into DP-SGD configuration.

B Noise mechanisms and privacy accounting

This section gives one accounting path for the three noise families. Each update clips an example gradient to norm C; for each mechanism we bound a shifted *base moment* $H_{j,\ell}$, apply Poisson subsampling once to the entire record, compose T updates, and invert the resulting privacy bound to calibrate noise at $(\varepsilon_{\text{train}}, \delta_{\text{train}})$. Gaussian uses an exact base moment; Lap_2 uses a sharp base

moment; coordinatewise PLRV-O uses a conservative majorization envelope. The reference numerical lattice below uses Rényi bounds for all three; the PLD discussion explains how a separately implemented PLD accountant relates to those bounds.

B.1 Privacy loss distributions and composition

For an ordered pair of output laws P, Q with densities p, q, the *privacy loss* is $L_{P\|Q}(y) = \log[p(y)/q(y)]$. Its distribution under $Y \sim P$ is the privacy loss distribution (PLD). If $p(y) > 0 = q(y)$, the loss is $+\infty$. The directional privacy profile is $\delta_{P\|Q}(\varepsilon) = \sup_S\{P(S) - e^{\varepsilon}Q(S)\}$. Both ordered neighboring directions must be bounded for add/remove DP.

Lemma 1 (Privacy profile from a PLD)**.** *For $\varepsilon \geq 0$,*

$$\delta_{P\|Q}(\varepsilon) = \mathbb{E}_{Y\sim P}\left[(1 - e^{\varepsilon - L_{P\|Q}(Y)})_+\right]. \tag{10}$$

Proof. The event $\{p > e^{\varepsilon}q\}$ maximizes the integral of $p - e^{\varepsilon}q$. Integrating its positive part and factoring out p gives Equation 10, with $e^{-\infty} = 0$. □

For independent product pairs, privacy losses add and their PLDs convolve. Adaptive DP-SGD needs an additional condition: for each update, the accounting pair must dominate the conditional mechanism uniformly over previous outputs and neighboring datasets. Here domination means an upper bound on every hockey-stick divergence $\int(p - aq)_+$, $a \geq 0$. Products of such dominating pairs bound adaptive composition (Zhu et al., 2022). Thus convolution applies to the accounting distributions; it does not assume that the actual gradients or training losses are independent.

A numerical PLD accountant discretizes these distributions and composes them, often using FFT convolution (Koskela et al., 2020). A certified calculation must upper-bound the effects of discretization, truncation, and numerical approximation; pessimistic PLD discretization provides one approach (Doroshenko et al., 2022). The accountant then checks both adjacency directions and returns an upper bound on $\delta(\varepsilon)$, or inverts that bound for a requested δ. Repeating this calculation over the noise scale calibrates a training run at its declared q, T and privacy target. Tightness depends on the dominating pairs and the numerical bounds; it is not automatic for every noise family.

B.2 Rényi accounting and its relation to PLD accounting

For $\alpha > 1$, Rényi divergence satisfies

$$D_\alpha(P\|Q) = \frac{1}{\alpha - 1}\log \mathbb{E}_{Y\sim P} e^{(\alpha-1)L_{P\|Q}(Y)}. \tag{11}$$

A mechanism is $(\alpha, \rho(\alpha))$-RDP if this divergence is at most $\rho(\alpha)$ for every ordered neighboring pair. Uniform conditional RDP bounds compose additively. A total curve ρ_{tot} yields the basic conversion

$$\varepsilon(\delta) \leq \inf_{\alpha>1}\left\{\rho_{\text{tot}}(\alpha) + \frac{\log(1/\delta)}{\alpha - 1}\right\}. \tag{12}$$

These are standard RDP properties (Mironov et al., 2019a). PLD accounting retains the privacy-loss distribution; RDP accounting uses its exponential moments. They are related representations, but their numerical bounds need not coincide. The reference lattice in Appendix E uses the explicitly stated RDP construction and improved conversion in Equation 23. It is not a PLD computation. Either valid accountant can supply the fixed-privacy frontier in Equation 2; changing accountants requires recalibrating the candidate noise values.

B.3 Gaussian and non-Gaussian noise mechanisms

Each update clips every example gradient to an ℓ_2 norm at most C and adds noise Z to the clipped sum. Its add/remove sensitivity is at most C. Poisson sampling uses one inclusion indicator per record, shared by all gradient coordinates. Sampling probability, schedule, and normalization are fixed conditional on public or separately privatized design inputs. A private dataset-size estimate used to set these quantities must be included in that metadata cost; the accountant cannot silently recompute them from unprotected counts.

For Gaussian noise, $Z_i \overset{\text{iid}}{\sim} \mathcal{N}(0, C^2\sigma^2)$. The Lap2 convention used here is independent coordinate Laplace noise, $Z_i \overset{\text{iid}}{\sim} \text{Laplace}(0, b)$, with joint density proportional to $\exp(-\|z\|_1/b)$ despite the ℓ_2 clipping constraint (Mohammady et al., 2026). It is distinct from radial noise with density proportional to $\exp(-\|z\|_2/b)$.

For the coordinatewise Gamma–Laplace generator used in the public PLRV-O optimizer[3], independently draw an inverse scale $U_i \sim \text{Gamma}(k, \theta)$ for each coordinate, in the shape–scale convention, then draw $Z_i \mid U_i \sim \text{Laplace}(0, 1/U_i)$. Its coordinate density is $k\theta/[2(1+\theta|z|)^{k+1}]$. We require $k > 2$ so that the variance exists. The U_i are independent across coordinates and are refreshed at each update. The Poisson inclusion decision for a record is nevertheless shared across all its gradient coordinates; the multivariate accountant below retains that dependence. All PLRV-O calculations and experiments claimed here must use this coordinatewise generator. The shared-scale vector density in the formal PLRV-O definition (Yang et al., 2025) is a different mechanism and is not covered by the factorized bounds below.

The normalized per-coordinate standard deviations are

$$s_{\text{G}} = \sigma, \qquad s_{\text{L}} = \frac{\sqrt{2}\, b}{C}, \qquad s_{\text{P}} = \frac{\sqrt{2}}{C\theta\sqrt{(k-1)(k-2)}}. \tag{13}$$

These formulas follow from the Gaussian and Laplace variances and $\mathbb{E}[U^{-2}] = 1/[\theta^2(k-1)(k-2)]$ for the mixture. Appendix B derives the corresponding moment bounds. None of these parameterizations alone determines the best clipping threshold or learning rate for a task.

The following bounds use the add/remove convention and noise generators defined above. Each is uniform over clipped-sum shifts of norm at most C.

B.4 Base moments and the non-Gaussian envelopes

For a noise density f and shift v, define the order-ℓ base moment as $\int f(z)[f(z-v)/f(z)]^\ell\, dz$. Let $H_{j,\ell}$ upper-bound its logarithm uniformly over $\|v\|_2 \le C$, and put $H_{j,0} = H_{j,1} = 0$. Gaussian noise gives

$$H_{\text{G},\ell} = \frac{\ell(\ell-1)}{2\sigma^2}, \tag{14}$$

by completing the square in the Gaussian density ratio. Consequently its calibrated fixed-privacy frontier is

$$\sigma_\star(B) = \inf\{\sigma > 0 : \varepsilon_{\text{acc,G}}(q(B), T(B), \sigma; \delta_{\text{train}}) \le \varepsilon_{\text{train}}\}. \tag{15}$$

Lemma 2 (Prefix envelope for a sensitivity ball)**.** *Let $w_i = \sqrt{i} - \sqrt{i-1}$. For any $v \in \mathbb{R}^d$ with $\|v\|_2 \le C$ and any convex, nondecreasing $g : [0, \infty) \to \mathbb{R}$ with $g(0) = 0$, $\sum_i g(|v_i|) \le \sum_i g(Cw_i)$.*

Proof. Cauchy–Schwarz gives $\sum_{i=1}^m |v|_{(i)} \le C\sqrt{m} = \sum_{i=1}^m Cw_i$, where the coordinates are sorted in decreasing order. For every $t \ge 0$, the identity $\sum_i (x_i - t)_+ = \max_{0 \le m \le d}(\sum_{i=1}^m x_{(i)} - mt)$ therefore bounds the hinge sum for $|v|$ by that for Cw. A convex nondecreasing function is a nonnegative linear combination (or integral) of a linear function and these hinges, proving the claim. □

The dimension factor associated with this envelope is

$$S_d = \sum_{i=1}^d w_i^2 = \tfrac{1}{4}\log d + O(1), \tag{16}$$

whose asymptotics are proved in Appendix D.

For scalar Laplace noise, splitting the likelihood-ratio integral at 0 and at the nonnegative normalized shift t gives

$$F_\ell(t) = \frac{\ell}{2\ell-1} e^{(\ell-1)t} + \frac{\ell-1}{2\ell-1} e^{-\ell t}. \tag{17}$$

[3] https://github.com/datasec-lab/plrvo/blob/main/Computer_Vision/opacus/optimizers/LMODPOptimizer.py

Indeed the three integrals are $e^{-\ell t}/2$, $(e^{(\ell-1)t} - e^{-\ell t})/[2(2\ell-1)]$, and $e^{(\ell-1)t}/2$. Independence makes the vector log-moment $\sum_i \log F_\ell(|v_i|/b)$. The function $\log F_\ell$ is convex by log-sum-exp, and nondecreasing because its derivative is zero at zero and increasing. Lemma 2 yields

$$H_{\mathrm{L},\ell}^{\mathrm{prefix}} = \sum_{i=1}^{d} \log F_\ell(Cw_i/b). \tag{18}$$

Proposition 1 (Exact base moment for coordinate Laplace noise)**.** *For $\ell > 1$, independent coordinate Laplace noise of scale $b > 0$ has worst-case log-moment over $\|v\|_2 \le C$ equal to*

$$H_{\mathrm{L},\ell}^{\mathrm{sharp}} = d \log F_\ell\left(\frac{C}{b\sqrt{d}}\right). \tag{19}$$

An equal-magnitude shift $|v_i| = C/\sqrt{d}$ attains the bound.

Proof. Write $g(t) = \log F_\ell(t)$ and $r(t) = \ell e^{(\ell-1)t}/[(2\ell-1)F_\ell(t)]$. For $t \ge 0$, $r(t) > 1/2$ and increases. Thus $g''(t) = (2\ell-1)^2 r(t)(1-r(t))$ is positive and decreasing. Since $g'(0) = 0$, integration gives $g'(t) \ge t g''(t)$. For $\psi(u) = g(\sqrt{u})$, $\psi''(u) = [t g''(t) - g'(t)]/(4t^3) \le 0$ for $t = \sqrt{u} > 0$. Its continuous extension to zero is concave and nondecreasing. Jensen's inequality and the sensitivity constraint give $\sum_i g(|v_i|/b) \le d\, g(\|v\|_2/(b\sqrt{d})) \le d\, g(C/(b\sqrt{d}))$. Equal-magnitude coordinates attain equality. □

This sharp result is for the base product mechanism. Subsampling and conversion to approximate DP still use the stated accounting bounds.

Proposition 2 (Independent-scale mixture envelope)**.** *For the coordinatewise independent Gamma–Laplace mixture in Appendix B.3, define*

$$G_\ell(x) = \frac{\ell}{2\ell-1}[1-(\ell-1)\theta x]^{-k} + \frac{\ell-1}{2\ell-1}(1+\ell\theta x)^{-k}. \tag{20}$$

For an integer $\ell \ge 2$ with $(\ell-1)C\theta < 1$, a valid uniform log-moment bound is

$$H_{\mathrm{P},\ell} = \sum_{i=1}^{d} \log G_\ell(Cw_i). \tag{21}$$

Proof. Revealing the data-independent latent scales cannot reduce Rényi divergence. Conditional on a scale U_i, the scalar moment is $F_\ell(xU_i)$. The Gamma moment-generating function gives $\mathbb{E}F_\ell(xU_i) = G_\ell(x)$ in the stated finite domain. Because the U_i and conditional Laplace draws are independent across coordinates, the augmented base moment is $\prod_i G_\ell(|v_i|)$, and its log is $\sum_i \log G_\ell(|v_i|)$. Moreover $G_\ell(x)$ is a moment-generating function of a signed, scaled Gamma variable with mean zero. Its logarithm is convex, equals zero at zero, and has derivative zero there; it is therefore nondecreasing for $x \ge 0$. Lemma 2 proves the bound. □

Orders outside this sufficient domain receive an infinite envelope and are excluded from conversion. This restriction belongs to the latent-scale bound: the marginal density has bounded likelihood ratios under fixed shifts, so its privacy need not fail outside the domain. A tighter marginal accountant can change the calibrated frontier.

Proposition 3 (Clipping homogeneity)**.** *Replacing C by aC, $a > 0$, while leaving σ unchanged, replacing b by ab, or replacing θ by θ/a, respectively, preserves each mechanism's normalized noise and the privacy bounds above. Thus their normalized ridges are independent of C.*

Proof. The bounds depend only on σ, b/C, or $(k, C\theta)$, respectively. These quantities and Equation 13 are invariant under the stated transformations. Equivalently, scaling the query and noise together is an invertible output transformation. □

In the calculations below, $H_{j,\ell}$ is the Gaussian value in Equation 14, the sharp Lap_2 value in Equation 19, or the coordinatewise PLRV-O envelope in Equation 21. The alternative Lap_2 prefix bound in Equation 18 is used only for the stated bound comparison. These family-specific base moments all enter the same vector-level subsampling and composition calculation next.

B.5 SHARED-RECORD SUBSAMPLING AND COMPOSITION

Proposition 4 (Two-direction Poisson-subsampled bound)**.** *For the centrally symmetric additive mechanisms above and integer $\alpha \geq 2$ with finite base bounds through order α, define*

$$\mathcal{L}_{j,\alpha}(q) = \log \sum_{\ell=0}^{\alpha} \binom{\alpha}{\ell} (1-q)^{\alpha-\ell} q^{\ell} e^{H_{j,\ell}}. \tag{22}$$

One Poisson-sampled update is $(\alpha, \mathcal{L}_{j,\alpha}(q)/(\alpha-1))$-RDP under add/remove adjacency. Uniform conditional bounds for T updates compose to $T\mathcal{L}_{j,\alpha}(q)/(\alpha-1)$.

Proof. Let $f_0(z) = f(z)$, $f_1(z) = f(z-v)$ and $r = f_1/f_0$, with $f(-z) = f(z)$. Reflection $z \mapsto v - z$ exchanges the densities. For odd $\ell \geq 3$, pairing the integrand of $\int f_0(r-1)^{\ell}$ over $r \geq 1$ with its reflection gives $f_0(r-1)^{\ell}(1-r^{1-\ell}) \geq 0$. The reverse ordered pair has the same property. The nonnegative odd-moment condition of Zhu & Wang (2019, Theorem 8) therefore holds. That theorem bounds both sampling directions by the binomial moment expansion of $(1-q)f_0 + qf_1$ against f_0. For the product noise mechanisms considered here, the base ℓth moment factors across coordinates; it is $\prod_i F_{\ell}(|v_i|/b)$ for Lap_2 and $\prod_i G_{\ell}(|v_i|)$ for coordinatewise PLRV-O. The single Poisson inclusion indicator shifts the entire vector, so the binomial expansion is taken *after* this product, not separately for each coordinate. Substituting uniform bounds $e^{H_{j,\ell}}$ preserves the inequality because the expansion coefficients are nonnegative. Conditional RDP composition then gives the training bound. □

Coordinate log-moments are summed *before* the binomial expansion: a record's inclusion is shared across its coordinates. The distinction is visible already at order two. For two independent noise coordinates, write $m_i = \mathbb{E}_{f_{0,i}}(f_{1,i}/f_{0,i})^2 \geq 1$. The shared-inclusion moment is $1 + q^2(m_1 m_2 - 1)$, whereas multiplying separately subsampled coordinate moments gives $[1+q^2(m_1-1)][1+q^2(m_2-1)]$. The former exceeds the latter by $q^2(1-q^2)(m_1-1)(m_2-1)$ for $0 < q < 1$. Thus coordinatewise noise independence does not justify coordinatewise subsampling accounting. Gaussian substitution recovers the integer-order sampled Gaussian expression (Mironov et al., 2019b). This RDP certificate does not require constructing a PLD; the fixed-privacy frontier also accepts a separately valid PLD accountant.

B.6 CONVERSION AND NUMERICAL CONSTRUCTION

The reference calculations use $\mathcal{A} = \{2, \ldots, 128\}$ and $c_{\alpha}(\delta) = \log[(\alpha-1)/\alpha] - (\log\delta + \log\alpha)/(\alpha-1)$ in the improved conversion

$$\varepsilon_{\text{acc}} = \max\left\{0, \min_{\alpha \in \mathcal{A}} \left[\frac{T\mathcal{L}_{j,\alpha}}{\alpha-1} + c_{\alpha}(\delta)\right]\right\}. \tag{23}$$

For completeness, the scalar inequality $(t - e^{\varepsilon})_+ \leq [(\alpha-1)^{\alpha-1}/\alpha^{\alpha}] e^{-(\alpha-1)\varepsilon} t^{\alpha}$ follows by maximizing $(t - e^{\varepsilon})/t^{\alpha}$ for $t \geq e^{\varepsilon}$. Integrating with $t = p/q$ against Q and using the RDP bound gives $\delta(\varepsilon) \leq [(\alpha-1)^{\alpha-1}/\alpha^{\alpha}] e^{(\alpha-1)(\rho-\varepsilon)}$. Solving for ε gives the conversion above. Restricting the order grid remains conservative. Calibration uses bisection and checks the unrounded feasible noise value.

For large-dimensional prefix envelopes, the first 512 coordinates are summed exactly. Subsequent intervals $[a, b]$ with $b = \min\{d, \lfloor 1.01a \rfloor\}$ use their largest weight w_a for every term. Monotonicity makes this an upper bound on the analytic sum. Direct checks at $d = 1{,}000, 90{,}000, 400{,}000$ confirmed this direction; among the checked finite orders and shapes $16, 1024$ at normalized RMS noise 3, the largest relative moment increase was below 0.36%. The sixteen Gaussian points agreed with `dp-accounting` 0.6.0 within 2×10^{-14} in ε. These are numerical consistency checks, not rigorous floating-point error certificates.

C GROUP INFORMATION AND SHORTLIST GUARANTEES

We first establish the exact identities used by the finite screen. We then analyze a smooth approximation to explain its task-dependent behavior. The approximation is not used to certify grid comparisons.

C.1 Participation and proxy invariance

Lemma 3 (Group participation and exposure). *Fix $N, n_\nu \in \mathbb{N}$ with $1 \le n_\nu \le N$, $0 < B \le N$, $q = B/N$, and $T = \lceil E/q \rceil$. Under independent Poisson inclusion, $K_\nu \sim \mathrm{Binomial}(n_\nu, q)$ and*

$$p_\nu = 1 - (1-q)^{n_\nu}, \qquad A_\nu = Tp_\nu, \qquad \mathbb{E}[K_\nu \mid K_\nu > 0] = \frac{n_\nu q}{p_\nu}. \tag{24}$$

Moreover,

$$A_\nu \le Tn_\nu q < n_\nu(E+q) \le n_\nu(E+1). \tag{25}$$

Proof. The count is a sum of n_ν independent Bernoulli variables. The probability that all are absent is $(1-q)^{n_\nu}$, and linearity of expectation gives $A_\nu = Tp_\nu$. Since K_ν vanishes when the group is absent, $\mathbb{E}K_\nu = p_\nu \mathbb{E}[K_\nu \mid K_\nu > 0]$. Finally, the union bound gives $p_\nu \le n_\nu q$, while $\lceil E/q \rceil < E/q + 1$. □

The score uses the square of the conditional mean count, not its conditional second moment. It measures a specified mean-contribution proxy; group counts do not determine evolving gradient directions or clipping retention. The exposure bound also provides a basic feasibility check: $n_\nu = 58, E = 20$ cannot support an expected activity floor of $5{,}000$.

Proposition 5 (Invariance to the retained-signal proxy). *Within a fixed (j, C, η) branch, batch-to-batch ratios of $\widetilde{\mathcal{I}}_{\nu,j}$ and the branchwise Pareto ordering under the score and update count are independent of the values of the positive, batch-invariant proxies $\mu_\nu(C)$.*

Proof. Write $F_{\nu,j}(B) = T(B)(n_\nu B/N)^2/[p_\nu(B)s_{j,\star}(B)^2]$ and $c_\nu = [\mu_\nu(C)/C]^2 > 0$. Within a branch, $\widetilde{\mathcal{I}}_{\nu,j}(B, C) = c_\nu F_{\nu,j}(B)$, so

$$\frac{\widetilde{\mathcal{I}}_{\nu,j}(B^+, C)}{\widetilde{\mathcal{I}}_{\nu,j}(B^-, C)} = \frac{F_{\nu,j}(B^+)}{F_{\nu,j}(B^-)}. \tag{26}$$

Thus logarithmic ratios and their signs are unchanged. Positive coordinate scaling preserves each weak and strict inequality defining Pareto dominance; the update-count coordinate is unchanged. □

C.2 Smooth score laws and the feasible batch interval

Let $x_\nu = n_\nu B/N$ and define the smooth surrogate by replacing $T(B)$ with EN/B and $p_\nu(B)$ with $1 - e^{-x_\nu}$:

$$\mathcal{I}^{\mathrm{sm}}_{\nu,j}(B, C) = \frac{En_\nu^2 B}{N(1-e^{-x_\nu})s_{j,\star}(B)^2}\left[\frac{\mu_\nu(C)}{C}\right]^2. \tag{27}$$

The exponential participation approximation is appropriate at small q; integer schedules and the exact probability remain in the screen.

Proposition 6 (Group-information elasticity). *On an interval where $s_{j,\star}(B) > 0$ is differentiable, set $a_j(B) = d\log s_{j,\star}(B)/d\log B$. Holding the other inputs fixed,*

$$\frac{d\log \mathcal{I}^{\mathrm{sm}}_{\nu,j}}{d\log B} = 1 - \frac{x_\nu}{e^{x_\nu} - 1} - 2a_j(B). \tag{28}$$

For a local law $s_{j,\star} \propto B^{a_B} E^{a_E} \varepsilon_{\mathrm{train}}^{-a_\varepsilon}$, the leading-order sparse and saturated regimes are

$$\mathcal{I}^{\mathrm{sm}}_{\nu,j} \propto \left[\frac{\mu_\nu(C)}{C}\right]^2 E^{1-2a_E} \varepsilon_{\mathrm{train}}^{2a_\varepsilon} \begin{cases} n_\nu B^{-2a_B}, & x_\nu \ll 1, \\ (n_\nu^2/N)B^{1-2a_B}, & x_\nu \gg 1. \end{cases} \tag{29}$$

Proof. Logarithmic differentiation of Equation 27 gives $1 - d\log(1-e^{-x_\nu})/d\log B - 2a_j(B)$. Since $dx_\nu/d\log B = x_\nu$,

$$\frac{d\log(1-e^{-x_\nu})}{d\log B} = \frac{x_\nu e^{-x_\nu}}{1-e^{-x_\nu}} = \frac{x_\nu}{e^{x_\nu}-1}. \tag{30}$$

For the regime laws, substitute the local noise law and use $1 - e^{-x} \sim x$ as $x \to 0$ and $1 - e^{-x} \to 1$ as $x \to \infty$. □

If only the schedule is smoothed, the exact participation contribution is instead

$$\frac{d \log p_\nu(B)}{d \log B} = \frac{n_\nu q (1-q)^{n_\nu - 1}}{1-(1-q)^{n_\nu}}, \qquad q \in (0,1), \tag{31}$$

obtained by differentiating $p_\nu = 1-(1-B/N)^{n_\nu}$. These identities predict changes in the diagnostic, not accuracy.

Corollary 2 (Smooth screening interval)**.** *Suppose the noise law has a constant exponent $0 < a_B < 1/2$ on the interval under consideration. Define $h(x) = x/(e^x - 1)$, $\phi(x) = (1 - e^{-x})/x$, and let $x_\star > 0$ be the unique solution of $h(x_\star) = 1 - 2a_B$. For every positive activity floor satisfying $u_\nu = A_{\nu,\min}/(En_\nu) \in (0,1)$, set $B_{A,\nu} = (N/n_\nu)\phi^{-1}(u_\nu)$; a zero floor gives $B_{A,\nu} = +\infty$. The smooth score increases for every group while meeting the smooth update floors precisely on the interval*

$$\max_{\nu \in \mathcal{G}} \frac{N x_\star}{n_\nu} < B \le \min\left\{ B_{\text{resource}}, \frac{EN}{T_{\min}}, \min_{\nu \in \mathcal{G}} B_{A,\nu} \right\}, \tag{32}$$

intersected with the interval where the constant-exponent law holds. Here $B_{\text{resource}} \le N$ and $T_{\min} > 0$.

Proof. For $x > 0$, $h'(x) = [e^x(1-x) - 1]/(e^x - 1)^2 < 0$; the numerator is zero at zero and has derivative $-xe^x < 0$. Thus h decreases from 1 to 0, giving a unique $x_\star$ and a positive slope exactly when $x_\nu > x_\star$. Also $A_\nu^{\text{sm}} = En_\nu\phi(x_\nu)$ and $\phi'(x) = [(x+1)e^{-x} - 1]/x^2 < 0$, so each activity floor is an upper batch bound. The total-update and resource constraints supply the other upper bounds. Intersecting these conditions proves the claim. □

If a positive floor has $u_\nu \ge 1$, no positive batch satisfies it in this smooth model. If $a_B \le 0$, the smooth score slope is positive for every finite $x > 0$; if $a_B \ge 1/2$, it is negative. For $0 < a_B < 1/2$, the transition is a score minimum, not an optimal batch size. Finite PAN comparisons remain necessary when slopes vary, sampling is not sparse, or integer update counts matter.

C.3 Finite shortlist and metadata assumptions

Offline screening procedure. Given the predeclared finite grid Ω, update floors, group counts, and reference ω_{ref}, construct the shortlist without training candidate models:

1. For each declared noise choice j and batch size B, invert its mechanism-specific accountant at $(\varepsilon_{\text{train}}, \delta_{\text{train}})$. Discard a grid point if no valid noise scale is certified.
2. For every remaining configuration, compute $T(B)$ and $A_\nu(B)$ for all target groups. Keep those meeting Equation 6 in Ω_{req}.
3. Within each fixed (j, C, η) branch, order its feasible batch sizes. Test each adjacent pair with Equation 7; put both endpoints of every passing pair into Ω_{Eq}.
4. Remove configurations Pareto-dominated within their own branch under group-information scores and update count. Add the static reference, whether or not it passed the preceding tests, to obtain Ω_{scr}. If no pair passed, return only the reference.

The positive, branch-fixed retention proxy cancels from the PAN tests and within-branch dominance comparisons, so the procedure does not estimate gradients or fit a model during screening.

Proof of Theorem 1. For a fixed group and branch, Equation 4 and $q(B) = B/N$ give

$$\frac{\widetilde{\mathcal{I}}_{\nu,j}(B_{m+1}, C)}{\widetilde{\mathcal{I}}_{\nu,j}(B_m, C)} = \frac{T_{m+1}}{T_m} \left(\frac{B_{m+1}}{B_m}\right)^2 \frac{p_{\nu,m}}{p_{\nu,m+1}} \left(\frac{s_{j,\star}(B_m)}{s_{j,\star}(B_{m+1})}\right)^2.$$

Both endpoints are in the feasible list by assumption. The PAN denominator is positive, so the pair passes for every group precisely when all these ratios exceed one. Taking positive square roots and then the minimum over groups gives Equation 8; for Gaussian noise $s_{j,\star} = \sigma$. □

Proposition 7 (Representation of passing edges)**.** *For every accepted edge (ω^-, ω^+), the branch-wise nondominated shortlist contains a configuration with at least the scores and update count of ω^+. It therefore has a strictly higher score than ω^- for every target group.*

Proof. Both endpoints of each accepted pair belong to Ω_{req}, and the strictly positive finite PAN values imply a strict score increase for every group. The branchwise nondominated set is a subset of these feasible endpoints, so all its elements meet the floors. For an accepted upper endpoint ω^+, either it is nondominated or another endpoint dominates it. Following strict domination in the finite branch must terminate at a nondominated configuration $\widehat{\omega}$. Transitivity preserves every score inequality and $T(\widehat{\omega}) \geq T(\omega^+)$. Its scores therefore strictly exceed those of ω^-. This representative is retained in the union, and the reference is then added explicitly. □

These score guarantees are relative to the counts and fixed proxies defining the diagnostic. With privately estimated counts it applies to that stated input, not automatically to the unknown true counts. A guarantee for true counts can instead require the inequalities uniformly over a count-confidence set, with coverage stated separately. The metadata mechanism is included in the overall privacy accounting.

D Dimension, shape, and long-training limits

These results explain the task-dependent predictions in Section 3.1. They concern the specified analytic accountants and diagnostic, not optimal accuracy or a lower bound over all possible accountants.

D.1 Dimension and noise-shape limits

Lemma 4 (Prefix-weight growth)**.** *For $w_i = \sqrt{i} - \sqrt{i-1}$, $S_d = \frac{1}{4}\log d + O(1)$ and $\sum_{i=1}^{\infty} w_i^3 < \infty$.*

Proof. Rationalization gives $w_i = 1/(\sqrt{i} + \sqrt{i-1})$, hence $w_i^2 = 1/(4i) + O(i^{-2})$ and $w_i^3 = O(i^{-3/2})$. Summing these expansions proves both claims. □

Proposition 8 (Dimension dependence of the calibrated bounds)**.** *Fix q, T, a finite integer order grid, a positive privacy target, and the conversion in Equation 23, with a finite positive Gaussian calibration root $\sigma_\star$. Let $d \to \infty$ and hold the mixture shape $k > 2$ fixed. For the ungrouped analytic envelopes,*

$$\frac{s_{\mathrm{L},\star}^{\mathrm{prefix}}}{\sigma_\star} \sim \sqrt{2S_d}, \qquad \frac{s_{\mathrm{P},\star}^{\mathrm{prefix}}}{\sigma_\star} \sim \sqrt{2S_d\kappa_k}, \qquad \kappa_k = \frac{k(k+1)}{(k-1)(k-2)}. \tag{33}$$

For the sharp Lap2 base-moment bound, $s_{\mathrm{L},\star}^{\mathrm{sharp}}/\sigma_\star \to \sqrt{2}$.

Proof. For each fixed moment order, Taylor expansion at zero gives

$$\log F_\ell(t) = \tfrac{1}{2}\ell(\ell-1)t^2 + O_\ell(t^3), \tag{34}$$

$$\log G_\ell(Cw) = \tfrac{1}{2}\ell(\ell-1)k(k+1)\xi^2 w^2 + O_{\ell,k}(\xi^3 w^3), \qquad \xi = C\theta. \tag{35}$$

Set $b/C = \sigma_{\mathrm{eff}}\sqrt{S_d}$ for the prefix Lap2 bound. Lemma 4 then yields $H_{\mathrm{L},\ell}^{\mathrm{prefix}} = \ell(\ell-1)/(2\sigma_{\mathrm{eff}}^2) + O_\ell(S_d^{-3/2})$. For the mixture, set $\xi = [\sigma_{\mathrm{eff}}\sqrt{k(k+1)S_d}]^{-1}$ and obtain the same limit. The expansions are uniform over the finite order grid and compact positive ranges of σ_{eff}. The finite subsampling sums and conversion therefore converge uniformly near the Gaussian root. Continuity and strict monotonicity there bracket the calibrated roots by $\sigma_\star \pm \epsilon$ for every sufficiently small $\epsilon > 0$ and large enough d. Converting to normalized RMS noise gives the first two limits.

For sharp Lap2, $d \log F_\ell(C/(b\sqrt{d})) = \ell(\ell-1)C^2/(2b^2) + O_\ell(d^{-1/2})$ at fixed $b/C > 0$. The same root argument gives $b_\star/C \to \sigma_\star$, hence the factor $\sqrt{2}$ in RMS units. □

Thus the logarithmic dimension factor belongs to the prefix bound and is avoidable for Lap2. Upper-rectangle coordinate grouping adds a separate numerical overestimate; it is not part of this asymptotic claim.

Proposition 9 (Laplace limit of the mixture envelope)**.** *At fixed dimension and finite order grid, hold $r_0 = Ck\theta > 0$ fixed and let $k \to \infty$. The mixture prefix moments converge to the Lap2 prefix moments with $b/C = 1/r_0$, and $s_{\mathrm{P}} \to \sqrt{2}/r_0$.*

Proof. For each fixed w and ℓ, the two terms in $G_\ell(Cw)$ obey $[1-(\ell-1)r_0w/k]^{-k} \to e^{(\ell-1)r_0w}$ and $[1+\ell r_0 w/k]^{-k} \to e^{-\ell r_0 w}$. Thus $G_\ell(Cw) \to F_\ell(r_0 w)$. There are finitely many coordinates and orders, so their sums converge. Substitution into Equation 13 gives the RMS limit. □

This limit retains the prefix certificate; it does not become the sharp Lap2 certificate. A finite grid's largest shape can therefore give its smallest calibrated RMS without establishing a universally optimal finite shape or superior measured utility.

D.2 Long-training ridge law and information plateau

Fix the conversion c_α in Equation 23, and define

$$\zeta = \max_{\substack{\alpha \in \mathcal{A} \\ \varepsilon_{\mathrm{train}} > c_\alpha(\delta_{\mathrm{train}})}} \frac{\varepsilon_{\mathrm{train}} - c_\alpha(\delta_{\mathrm{train}})}{\alpha}. \tag{36}$$

Theorem 3 (Long-training limit for the declared accountants)**.** *Fix N, B, d, C, target-group counts and positive proxies, a finite integer order grid $\mathcal{A}$, and a training privacy target for which $\zeta > 0$. For the mixture, also fix $k > 2$. As $E \to \infty$, the calibrated ungrouped analytic bounds satisfy*

$$s_{j,\star}^2 \sim \frac{D_j EB}{2N\zeta}, \qquad D_{\mathrm{G}} = 1, \quad D_{\mathrm{L}} = 2, \quad D_{\mathrm{P}} = 2S_d\kappa_k, \tag{37}$$

where Lap2 uses the sharp bound and the mixture uses the prefix bound. Writing $R_j = s_{j,\star}\sqrt{T(B)}$, this implies

$$R_j \sim E\sqrt{\frac{D_j}{2\zeta}}, \qquad \widetilde{\mathcal{I}}_{\nu,j} \longrightarrow \frac{2\zeta n_\nu^2}{D_j p_\nu}\left[\frac{\mu_\nu(C)}{C}\right]^2. \tag{38}$$

The Lap2 prefix bound has the same form with $D_{\mathrm{L}}^{\mathrm{prefix}} = 2S_d$.

Proof. At large noise, the base expansions give $H_{j,\ell} = \ell(\ell-1)D_j/(2s_j^2) + o(s_j^{-2})$ uniformly over the finite order grid. In the binomial sum, $\mathbb{E}[L(L-1)] = \alpha(\alpha-1)q^2$ for $L \sim \mathrm{Binomial}(\alpha, q)$. Expanding the exponential and logarithm therefore gives $\mathcal{L}_{j,\alpha}/(\alpha-1) = \alpha D_j q^2/(2s_j^2) + o(s_j^{-2})$. With $T = \lceil E/q \rceil$, at s_j^2 proportional to E the total bound converges to $\alpha D_j EB/(2Ns_j^2)$. For a fixed order with positive conversion slack, the asymptotic noise requirement is $s_j^2 \geq D_j EB/[2N\{(\varepsilon_{\mathrm{train}} - c_\alpha)/\alpha\}]$. Minimizing this requirement over a finite grid gives ζ. Bracketing the root with any fixed multiplicative factors above and below one proves Equation 37. Since $T \sim EN/B$ and p_ν is fixed, substitution into the definitions of R_j and the score proves Equation 38. □

For intuition, under the basic conversion $c_\alpha = \log(1/\delta)/(\alpha-1)$ and unrestricted continuous orders, maximizing the quadratic surrogate gives $\zeta = (\sqrt{\log(1/\delta) + \varepsilon_{\mathrm{train}}} - \sqrt{\log(1/\delta)})^2$. To verify this, set $t = \alpha - 1$, differentiate $(\varepsilon_{\mathrm{train}} - \log(1/\delta)/t)/(t+1)$, and solve the positive root of $\varepsilon_{\mathrm{train}} t^2 - 2\log(1/\delta)t - \log(1/\delta) = 0$. The numerical lattice instead uses its finite grid and improved conversion.

The theorem takes only E to infinity. It is not uniform over varying orders, privacy targets, shape, batch size, or dimension. At finite epochs, the calibrated ridge can still admit a positive PAN interval even though the quadratic surrogate has batch exponent $1/2$.

D.3 Implications for model size and clipping

Independence and zero mean give the exact identity $\sqrt{\mathbb{E}\|Z/B\|_2^2} = Cs_j\sqrt{d}/B$ by summing coordinate variances. Substituting the quadratic ridge law gives $C\sqrt{dD_jE/(2N\zeta B)}$. Holding this noise

Table 5: Reference calibration inputs. The target minority is class 8, female, cat, eyeglasses, negative sentiment, and not-entailment, respectively. $B_{\mathrm{occ}} = N/n_\nu$ marks the group-occupancy scale. Model dimensions are approximate; h denotes the trainable language classification head. The frozen RoBERTa backbone has approximately 125M parameters and is excluded from d.

Dataset	Model	N	n_ν	Trainable d	E	$\varepsilon_{\mathrm{train}}$	B_{occ}
MNIST	CNN	54,207	58	400,000	20	3	934.6
Adult	MLP	32,561	10,771	90,000	20	4	3.0
CIFAR-10	TinyViT	45,918	918	5,500,000	3	5	50.0
CelebA	ViT-base	162,770	10,521	85,800,000	1	0.5	15.5
SST-2	RoBERTa	46,503	8,934	$7{,}680 + h$	E_{S}	$5, 8, 50$	5.2
QNLI	RoBERTa	68,086	15,709	$7{,}680 + h$	E_{Q}	$12, 20, 50$	4.3

norm constant would suggest $B \propto dD_j$ at fixed C, E, N, ζ. This is a conditional noise-scaling relation: it ignores signal geometry and can violate update requirements. It does not identify an optimal batch for a new architecture.

Where the proxy is differentiable, clipping homogeneity gives $d \log \widetilde{\mathcal{I}}_{\nu,j}/d \log C = 2(d \log \mu_\nu(C)/d \log C - 1)$. Neither model dimension nor privacy accounting determines this proxy curve. Accordingly, clipping and learning-rate choices remain for outcome-based calibration.

Across the four reference vision/tabular models, $\sqrt{2S_d} \approx 2.77$–3.34, whereas sharp Lap2 approaches $\sqrt{2}$ relative to the Gaussian RMS ridge. These are accounting comparisons. For prompt tuning, the relevant dimension is $7{,}680 + h$, counting the prompts and trainable head; the frozen backbone is excluded. Language-task epochs and head sizes remain inputs to their numerical ridges.

E Reference frontiers and worked shortlists

Table 5 specifies worked accounting examples, not the populations used to generate the empirical tables. MNIST retains 1% of class 8; CIFAR-10 retains cats at approximately 2% of its final training set; SST-2 and QNLI retain 30% of their respective minority classes. Adult uses the unfiltered UCI split and CelebA the official training partition. An executed calculation uses its actual record counts and perturbed parameter dimension; the language-task epochs and head sizes remain inputs to their ridges.

The lattice records $N, n_\nu, d, E, B, q, T, p_\nu, A_\nu$, mechanism and shape, normalized RMS noise, physical parameters at the stated reference clipping bounds, the minimizing RDP order, and the resulting privacy bound. The normalized diagnostic column divides out $[\mu_\nu(C)/C]^2$; it does not estimate that retention factor. Within-branch score ratios and the worked Pareto ordering therefore require no fitted retention values. The example shortlist checks both the minority and its complement.

E.1 Numerical ridges and executable parameter examples

We compute reference ridges on $B \in \{128, 256, 512, 1024\}$ using $\delta_{\mathrm{train}} = 10^{-5}$, Poisson sampling, the integer schedule, and RDP orders $2, \ldots, 128$. Gaussian and refined Lap2 base moments are exact worst-case expressions; PLRV-O uses the conservative prefix envelope in Appendix B. PLRV-O shapes $k \in \{4, 16, 64, 256, 1024\}$ give 112 calibrated lattice points across the four fully specified workloads; 16 additional Lap2-prefix points measure the cost of the looser bound. These calculations involve no model fitting or private validation.

From normalized noise to executable parameters. A ridge specifies noise after B and the mechanism shape are fixed; the declared clipping grid supplies C. For example, at $B = 256$, the CIFAR-10 reference yields the following lattice elements, all with learning rate $\eta = 10^{-3}$:

$$\begin{aligned} \omega_{\mathrm{G}} &= (B = 256, C = 1, \sigma \approx 0.617610), \\ \omega_{\mathrm{L}} &= (B = 256, C = 1, b \approx 0.617538), \\ \omega_{\mathrm{P},16} &= (B = 256, C = 0.1, k = 16, \theta \approx 0.472190), \\ \omega_{\mathrm{P},1024} &= (B = 256, C = 0.1, k = 1024, \theta \approx 0.007626675). \end{aligned} \tag{39}$$

Table 6: Accountant-derived ridge examples. Left: normalized per-coordinate RMS noise $s_{j,\star}$ at $B = 256$. Right: the minority score ratio $\widetilde{\mathcal{I}}_{\nu,j}(512, C)/\widetilde{\mathcal{I}}_{\nu,j}(128, C)$ within each fixed mechanism and shape. Ratios hold clipping fixed and cancel the positive retention proxy. Values are rounded for display; the accompanying lattice records the unrounded calibrated parameters.

	$s_{j,\star}(256)$				Score ratio: $128 \to 512$			
Workload	Gaussian	Lap2	P_{16}	P_{1024}	Gaussian	Lap2	P_{16}	P_{1024}
MNIST	0.819	1.158	2.573	2.263	0.698	0.698	0.657	0.652
Adult	0.818	1.156	2.444	2.150	2.213	2.212	2.075	2.063
CIFAR-10	0.618	0.873	2.067	1.813	2.774	2.774	2.720	2.715
CelebA	1.165	1.648	4.117	3.595	3.064	3.064	3.036	3.032

Table 7: Worked shortlists under explicitly chosen example floors. A_ν refers to the minority; the same activity floor is also imposed on its complement. All four branches pass the update requirements at $B = 128, 256, 512$ and fail them at 1024. PAN then determines whether the feasible increases qualify. These are predicted menus under the stated illustrative floors.

Workload	$T_{\min}$	$A_{\nu,\min}$	$T(512)$	$A_\nu(512)$	Retained batches	Menu size
MNIST	2,000	800	2,118	896.5	Reference only	1
Adult	1,000	1,000	1,272	1,272.0	128, 256, 512 per branch	12
CIFAR-10	250	200	270	270.0	128, 256, 512 per branch	12
CelebA	300	250	318	318.0	128, 256, 512 per branch	12

Their unrounded parameters meet the same training privacy target. For Lap2, $b = Cs_{\mathrm{L}}/\sqrt{2}$; for PLRV-O, $\theta = \sqrt{2}/[Cs_{\mathrm{P}}\sqrt{(k-1)(k-2)}]$. The displayed clipping bounds and learning rate specify a worked lattice, rather than inferred utility optima. For example, changing the Lap2 element to $C = 2$ and $b \approx 1.235077$ leaves its privacy certificate unchanged but can alter clipping retention and learning.

The broader search can use the declared clipping domains and an independent geometric learning-rate grid. Their declared domains and discretization are discussed in Section E.3. Accounting supplies neither grid's utility optimum; private outcome evaluation selects among their retained settings. Likewise, PLRV-O's smallest RMS value on the tested shape grid occurs at $k = 1024$, near the Lap2-prefix accounting limit. This is a boundary choice on a finite grid, not a claim of an optimal PLRV-O shape or a universal advantage over Gaussian noise. At $B = 256$, the Lap2 prefix envelope gives noise levels $2.258, 2.145, 1.810, 3.588$ for the four workloads, compared with $1.158, 1.156, 0.873, 1.648$ under the refinement. The reduction comes entirely from the certificate; the underlying Laplace distribution and clipping convention are unchanged.

E.2 Worked reference shortlists

Table 7 applies the finite-grid screen to a fully specified example: the four batches above, the clipping bounds and learning rate from Equation 39, and branches $\{\mathrm{G}, \mathrm{L}, \mathrm{P}_{16}, \mathrm{P}_{1024}\}$. The target groups are the designated minority and its complement, with fixed positive retention proxies in each branch. Their magnitudes cancel from the within-branch score ratios. The table declares illustrative update floors to expose the screening logic; an executed run uses its declared floors. The reference is the Gaussian $B = 128$ element for each workload.

For Adult, CIFAR-10, and CelebA, every feasible adjacent increase improves both groups' scores. The endpoints remain nondominated: larger batches offer more information, while smaller batches retain more updates. For CIFAR-10, for example, the menu is exactly the twelve accountant-calibrated elements indexed by $\{128, 256, 512\} \times \{\mathrm{G}, \mathrm{L}, \mathrm{P}_{16}, \mathrm{P}_{1024}\}$ with the stated C and η. Keeping all five tested PLRV-O shapes would instead produce 21 elements; privacy calibration decides among these additional choices.

E.3 Declared parameter grids

The worked menu uses the Cartesian grid $B \in \{128, 256, 512, 1024\}$, $\eta = 10^{-3}$, $C = 1$ for Gaussian and Lap2, and $C = 0.1$ with $k \in \{16, 1024\}$ for PLRV-O. The additional shape calculations use $k \in \{4, 64, 256\}$. Every noise parameter is supplied by accounting rather than independently crossed into this grid. These choices and the illustrative floors specify the reference lattice. Executed workloads must separately declare their complete batch, clipping, shape, and learning-rate grids before screening; the accountant supplies each noise scale rather than crossing it independently into the grid.

F Private calibration and experimental protocol

F.1 Private hyperparameter calibration with DP-HyPO

DP-HyPO adaptively samples configurations and returns the best evaluated output under a predeclared score ordering (Wang et al., 2023). For a fixed finite menu with J elements, its proposals satisfy $\underline{c}/J \le \pi_t(\omega) \le \bar{c}/J$, where $0 < \underline{c} \le 1 \le \bar{c}$; an unconstrained adaptive proposal must be projected into these bounds. The positive-geometric specialization samples an independent horizon with $\Pr(T_{\mathrm{H}} = t) = \gamma(1-\gamma)^{t-1}$, $t \ge 1$. Its expected number of evaluations is $1/\gamma$.

For our application, the base evaluator must cover the joint output of candidate training and the private utility/disparity measurements used for selection. Training DP does not privatize a subsequent query to private validation records. Computing disparity, applying the accuracy gate, and ranking candidates are post-processing only once their data-dependent inputs have been privatized. Shared or overlapping groups require accounting for the joint measurement. A separately evaluated private reference also requires accounting.

If metadata, reference evaluation, and calibration have respective conditional guarantees $(\varepsilon_m, \delta_m)$, $(\varepsilon_r, \delta_r)$, and $(\varepsilon_{\mathrm{cal}}, \delta_{\mathrm{cal}})$, basic composition gives the sufficient overall allocation $\varepsilon_m + \varepsilon_r + \varepsilon_{\mathrm{cal}} \le \varepsilon_{\mathrm{hypo}}$ and $\delta_m + \delta_r + \delta_{\mathrm{cal}} \le \delta_{\mathrm{hypo}}$. A cost already included in the calibration evaluator is not charged twice. Appendix F.2 specializes the published DP-HyPO bound and derives the horizon and menu results used here.

F.2 DP-HyPO horizon and menu guarantees

We now specialize the published DP-HyPO theorem to the evaluator and bounded proposals defined above. The consequences of menu restriction retain those assumptions.

F.2.1 Calibrating the geometric horizon

For a fixed finite menu Ω_{scr}, let $\rho_\omega(\alpha)$ uniformly bound the conditional RDP cost of the joint trained-model/private-score evaluator. Define $\bar{\rho}_{\Omega_{\mathrm{scr}}}(\alpha) = \max_{\omega \in \Omega_{\mathrm{scr}}} \rho_\omega(\alpha)$. Fix a finite grid of order pairs $\alpha, \hat{\alpha} > 1$ with finite bounds, proposal constraints $\underline{c}, \bar{c}$, and remaining budget $(\varepsilon_{\mathrm{cal}}, \delta_{\mathrm{cal}})$. Write

$$H_{\alpha,\hat{\alpha}} = \bar{\rho}_{\Omega_{\mathrm{scr}}}(\alpha) + 2\left(1 - \frac{1}{\hat{\alpha}}\right)\bar{\rho}_{\Omega_{\mathrm{scr}}}(\hat{\alpha}) + \left(\frac{\alpha}{\alpha - 1} + 2\right)\log\frac{\bar{c}}{\underline{c}} + \frac{\log(1/\delta_{\mathrm{cal}})}{\alpha - 1}, \tag{40}$$

$$L_{\alpha,\hat{\alpha}} = \frac{2}{\hat{\alpha}} + \frac{1}{\alpha - 1}. \tag{41}$$

Proposition 10 (Horizon calibration from DP-HyPO)**.** *Under the DP-HyPO assumptions, a positive-geometric horizon with parameter $\gamma \in (0,1)$ is certified whenever an available order pair satisfies $H_{\alpha,\hat{\alpha}} + L_{\alpha,\hat{\alpha}} \log(1/\gamma) \le \varepsilon_{\mathrm{cal}}$. If some pair has $H < \varepsilon_{\mathrm{cal}}$, the smallest parameter certified by these pairs is*

$$\gamma^\star = \min_{H_{\alpha,\hat{\alpha}} < \varepsilon_{\mathrm{cal}}} \exp\left[-\frac{\varepsilon_{\mathrm{cal}} - H_{\alpha,\hat{\alpha}}}{L_{\alpha,\hat{\alpha}}}\right]. \tag{42}$$

It maximizes the certified expected horizon $1/\gamma$. If no such pair exists, this bound certifies no $0 < \gamma < 1$.

Proof. Set the truncated negative-binomial parameter to one in Wang et al. (2023, Theorem 1), so $\mathbb{E}T_{\mathrm{H}} = 1/\gamma$. Substituting the uniform evaluator bounds and adding the basic RDP conversion term gives $H + L\log(1/\gamma)$. Since $L > 0$, each feasible pair requires $\gamma \geq \exp[-(\varepsilon_{\mathrm{cal}} - H)/L]$. Certification by any pair is sufficient, so the smallest threshold over the finite grid is admissible and minimizes γ. □

Corollary 4 (Necessary adaptivity budget)**.** *If $\bar{c} > \underline{c}$, certification by Proposition 10 requires*

$$\frac{\bar{c}}{\underline{c}} < \exp(\varepsilon_{\mathrm{cal}}/3). \tag{43}$$

Proof. Both evaluator terms and the conversion term are nonnegative, and $\alpha/(\alpha - 1) > 1$. Thus $H > 3\log(\bar{c}/\underline{c})$, whereas a geometric horizon requires $H < \varepsilon_{\mathrm{cal}}$ for at least one pair. □

Uniform proposals have $\bar{c} = \underline{c} = 1$ and remove this term, but the evaluator and conversion costs must still fit. If the full bound is infeasible, a one-run reference alternative needs its own accounting; it is not an automatic consequence of this certificate.

F.2.2 Consequences of restricting the menu

Proposition 11 (Menu restriction and the certified horizon)**.** *Suppose $\Omega_{\mathrm{scr}} \subseteq \Omega$ and the full-menu bound is feasible. Hold fixed the per-candidate RDP curves, remaining budget, order grid, and proposal density bounds. Then*

$$\gamma^{\star}_{\Omega_{\mathrm{scr}}} \leq \gamma^{\star}_{\Omega}, \qquad \mathbb{E}[T_{\mathrm{H}}^{\Omega_{\mathrm{scr}}}] \geq \mathbb{E}[T_{\mathrm{H}}^{\Omega}]. \tag{44}$$

Proof. Taking a maximum over a subset cannot increase $\bar{\rho}(\alpha)$. Its coefficients in H are nonnegative, so each previously feasible pair remains feasible with a no-larger threshold for γ. Minimizing over the pairs preserves this inequality; taking reciprocals gives the horizon claim. □

If the privacy envelope does not decrease, the horizon bound need not improve. Cardinality alone does not lower this privacy cost. Extra private metadata can also reduce the remaining budget, so it cannot be ignored in an end-to-end comparison.

Proposition 12 (Conditional discovery bound)**.** *Let $\mathcal{S} \subseteq \Omega_{\mathrm{scr}}$ be a fixed nonempty subset, $J = |\Omega_{\mathrm{scr}}|$, and $h = \underline{c}|\mathcal{S}|/J$. If every proposal, conditional on its history, assigns each candidate probability at least $\underline{c}/J$, then an independent positive-geometric horizon satisfies*

$$\Pr(\textit{visit } \mathcal{S} \textit{ by } T_{\mathrm{H}}) \geq \frac{h}{\gamma + h - \gamma h}. \tag{45}$$

Proof. The conditional miss probability at each evaluation is at most $1 - h$. Induction bounds the probability of missing for t steps by $(1 - h)^t$. Averaging over the horizon gives $\Pr(\text{miss } \mathcal{S}) \leq \sum_{t\geq 1} \gamma(1-\gamma)^{t-1}(1-h)^t = \gamma(1-h)/(\gamma + h - \gamma h)$. Subtracting from one proves the result. □

At fixed $|\mathcal{S}|$, the bound increases when J or γ decreases. This statement is conditional on retaining that subset; screening can remove configurations with good empirical performance. Visiting a configuration also does not guarantee a favorable training realization or its eventual selection after noisy validation.

F.3 Experimental settings and calibration inputs

F.3.1 Experimental settings

Table 8 summarizes the per-dataset model, dataset size, training horizon, and privacy budgets used throughout Section 4. The target $\varepsilon_{\mathrm{train}}$ applies to each candidate training run, while $\varepsilon_{\mathrm{hypo}}$ is the overall privacy budget for the private-calibration procedure (candidate training together with data-dependent configuration selection); the two values are not additive. All training calculations use $\delta_{\mathrm{train}} = 10^{-5}$; end-to-end certification of the private-calibration procedure additionally requires

Table 8: Experimental settings used in the main evaluation. The overall private-calibration target $\varepsilon_{\text{hypo}}$ is reported only for experiments that use DP-HyPO; SST-2 and QNLI prompt-based experiments use exhaustive search without private calibration.

Dataset	Model	N	Epochs	$\varepsilon_{\text{train}}$	$\varepsilon_{\text{hypo}}$	Training
MNIST	CNN	55,000	20	3	6.6	Scratch
CIFAR-10	TinyViT	46,000	3	5	13.9	Pretrained
Adult	MLP	32,561	20	4	11.4	Scratch
CelebA	ViT-base	162,770	1	0.5	1.4	Pretrained
SST-2	BERT-base	–	5/20	0.5/1	–	Pretrained/Scratch
SST-2	RoBERTa-base	–	100	8	–	Prompt-based
QNLI	RoBERTa-base	–	100	12	–	Prompt-based

Table 9: Configuration spaces used for candidate generation for CV tasks.

Category	Parameter	Values
Common	Batch size B	$\{128, 192, 256, 512, 768, 1024, 1280, 1536, 2048, 2560, 3072, 3584\}$
	Clipping C	$\{1, 10, 20\}$
	Learning rate η	$[10^{-4}, 10^{-3}, 10^{-2}]$
DPSGD-F	σ_{count}	$\{10, 15\}$
DPSGD-G-A	σ_{count}	$\{10, 15\}$
	Z_0	$\{50, 100\}$
	η_Z	$\{10^{-3}, 10^{-2}\}$
	τ	$\{0.7, 1.0\}$
PLRV-O Noise	Clipping C	$\{0.05, 0.1\}$
	Shape parameter θ	3 geometric values in $[0.005, 0.08]$

accounting for private evaluation, metadata, and reference costs, together with a stated allocation of δ_{hypo}.

Table 9 lists the declared configuration space for the vision and tabular tasks. Common parameters (batch size, clipping bound, learning rate) are shared across all methods so that every method is evaluated over the same batch and clipping grid; each baseline's additional method-specific parameters (e.g., σ_{count} for DPSGD-F; Z_0, η_Z, τ for DPSGD-G-A) are restricted to two values each to keep the overall training cost of the comparison tractable, while EQUICONFIG's PLRV-O branch additionally varies clipping and the shape parameter θ. This is the grid referenced in Section 4.2 when discussing why the relative ordering of DPSGD-F and DPSGD-G-A against DP-SGD can differ from their original, single-clipping-bound comparisons. For the language-model experiments, we use a reduced configuration space with batch size $B \in \{1024, 1280\}$, clipping $C = 0.05$, and a single value for each of the remaining parameters, reflecting the narrower search.

Before accountant calibration, a shared, method-agnostic heuristic pre-filter narrows the declared batch and clipping grid using the privacy target, epoch budget, and backbone/dataset difficulty. The clipping and batch-size rules follow the tuning guidance of Rehn et al. (2026): a tighter privacy target increases the retained clipping values and decreases the retained batch values, a weaker backbone or harder dataset also shifts the retained clipping values upward, and fewer epochs shifts the retained batch values downward. This pre-filter is applied identically to DP-SGD, DPSGD-F, DPSGD-G-A, and EQUICONFIG before each method's own privacy calibration and, for EQUICONFIG, its screening components. It is an implementation-level convenience to keep the declared grid a manageable size for accountant calibration, and is separate from the formal components Section 3 defines and analyzes: privacy-frontier calibration, the update and group-activity floors, PAN, and within-branch nondominance.

Trainable dimensions are 406,714 for the MNIST CNN, about 5.5M for TinyViT on CIFAR-10, 91.7k for the Adult MLP, and 85.8M for ViT-base on CelebA. The compared groups are MNIST

Table 10: Performance comparison of EQUICONFIG on ViT-CelebA under a fixed model privacy budget ($\varepsilon_{train} = 0.5, \delta = 10^{-5}$) with 1-epoch fine-tuning. Results are reported for *Private Calibration* with $\varepsilon_{\text{hypo}} = 1.4, r = 0.36$ and *Exhaustive Search* where selection carries no extra privacy cost. Bold marks the best value per column.

	Private Calibration				Exhaustive Search			
Method	**Minority Acc. ↑**	**Majority Acc. ↑**	**Overall Acc. ↑**	**Disparity ↓**	**Minority Acc. ↑**	**Majority Acc. ↑**	**Overall Acc. ↑**	**Disparity ↓**
EQUICONFIG (ours)	**94.30**	**97.89**	**97.66**	**3.58**	**95.14**	**98.43**	**98.05**	**3.29**
DP-SGD	93.80	97.74	97.48	3.93	93.80	97.74	97.48	3.93
DPSGD-F	90.91	94.78	94.53	4.21	93.63	97.93	97.66	4.30
DPSGD-G-A	83.62	88.70	88.37	5.09	92.55	97.56	97.29	5.01
Non-private	95.41	98.53	98.31	3.12	95.41	98.53	98.31	3.12

Table 11: Performance comparison of EQUICONFIG on MLP-Adult under a fixed model privacy budget ($\varepsilon_{train} = 4, \delta = 10^{-5}$) with 20-epoch training from scratch. Results are reported for *Private Calibration* with $\varepsilon_{\text{hypo}} = 11.4, r = 0.36$ and *Exhaustive Search* where selection carries no extra privacy cost. Bold marks the best value.

	Private Calibration				Exhaustive Search			
Method	**Minority Acc. ↑**	**Majority Acc. ↑**	**Overall Acc. ↑**	**Disparity ↓**	**Minority Acc. ↑**	**Majority Acc. ↑**	**Overall Acc. ↑**	**Disparity ↓**
EQUICONFIG (ours)	**81.57**	92.36	**85.10**	10.79	81.62	92.33	**85.22**	**10.71**
DP-SGD	81.32	**92.38**	84.94	11.05	81.36	92.42	84.98	11.06
DPSGD-F	81.37	92.06	84.83	**10.69**	**81.72**	**92.46**	85.20	10.74
DPSGD-G-A	81.00	91.95	84.54	10.95	81.46	92.19	84.94	10.73
Non-private	82.27	92.81	85.45	10.54	82.27	92.81	85.45	10.54

classes 8 and 2, CIFAR-10 cats and automobiles, Adult gender groups, and CelebA eyeglasse wearers and non-wearers.

BERT-base-uncased has about 110M parameters in the SST-2 fine-tuning study; RoBERTa-base has about 125M, of which only ten soft prompts and a classification head are trained in the prompt-based SST-2 and QNLI studies. The LLM studies are for a fixed batch of 1024. The additional BERT-based SST-2 settings use 5 epochs with pretraining or 20 from random initialization, $\varepsilon \in \{0.5, 1\}$, and $\delta = 10^{-5}$.

F.3.2 DP-HYPO CALIBRATION DETAILS

For the private-calibration experiments, the per-candidate training target is specified as a fraction of the overall DP-HyPO budget, $\varepsilon_{\text{train}} = r\,\varepsilon_{\text{hypo}}$, with $r \approx 0.45, 0.36, 0.36, 0.36$ for MNIST, CIFAR-10, Adult, and CelebA, respectively (Table 8). The reported proposal bounds are $\underline{c} = 1/1.1$ and $\bar{c} = 1.1$; the adaptive DP-HyPO proposal is constrained to remain within these bounds relative to its initial proposal at every trial, and the stopping rule uses the positive-geometric horizon specialization described in Section 4.1. The selection score used by DP-HyPO to rank candidates is

$$U = 0.5\,\text{Accuracy} + 0.5\,(1 - \text{Disparity}).$$

Here, accuracy and disparity are normalized to the scale required by the corresponding benchmark implementation. Minority- and majority-group accuracies are reported separately for evaluation but are not included as additional terms in this score.

F.4 ADDITIONAL RESULTS

This section reports full results for MLP-Adult and ViT-CelebA, together with the complete component-ablation study on CNN-MNIST referenced in Section 4.3.

MLP-Adult and ViT-CelebA. Table 11 reports results on MLP-Adult under both selection regimes. Gaps between methods are substantially smaller here than on the vision benchmarks: since the Adult protected attribute (gender) is far less severely imbalanced than the minority classes used in MNIST or CIFAR-10, all methods, including vanilla DP-SGD, already achieve comparatively low disparity (∼10.7–11.1%), and EQUICONFIG performs on par with or marginally better than the baselines in both regimes. Table 10 reports the corresponding results on ViT-CelebA, where EQUICONFIG achieves the best overall accuracy and lowest disparity among private methods in both private calibration (97.66% / 3.58%) and exhaustive search (98.05% / 3.29%), while DPSGD-G-A

Table 12: Component ablation of EQUICONFIG on CNN-MNIST. Each ablation removes one screening component while keeping the remaining components fixed. After candidate generation, private calibration is performed using DP-HyPO with $r = 0.4$ and $\varepsilon_{\text{hypo}} = 7.5$. Results are reported for both private calibration and exhaustive search, with runtimes reported in GPU-hours. The baseline methods are shown below the dashed separator.

	Screening Components				Private Calibration			Exhaustive Search		
Variant	**PAN**	**Update Constraints**	**Nondominance**	**# Candidates**	**Acc. ↑**	**Disp. ↓**	**Runtime ↓**	**Acc. ↑**	**Disp. ↓**	**Runtime ↓**
EQUICONFIG	✓	✓	✓	**72**	**97.3**	**10.9**	**0.96**	**97.45**	**10.18**	**3.21**
w/o PAN	×	✓	✓	81	97.1	13.1	0.97	97.45	10.18	3.71
w/o Update constraints	✓	×	✓	261	95.1	16.3	0.97	97.45	10.18	12.11
w/o Nondominance	✓	✓	×	72	97.3	10.9	0.96	97.45	10.18	3.21
DPSGD-F	–	–	–	174	90.2	25.1	1.24	92.18	18.21	10.15
DPSGD-G-A	–	–	–	1392	87.8	43.9	1.12	90.05	38.11	73.47
DP-SGD	–	–	–	87	94.0	24.6	0.68	94.67	19.81	2.76

shows a comparatively larger private-calibration drop relative to its exhaustive-search result, consistent with its additional adaptively-calibrated parameters increasing sensitivity to the private-search budget.

Component Ablation. Table 12 evaluates the three EQUICONFIG-specific screening components – PAN, the update and group-activity floors, and within-branch nondominance – by removing each in turn from the full pipeline on CNN-MNIST, after the shared heuristic pre-filter and privacy-frontier calibration have already been applied to the declared grid of 324 candidates. The full pipeline, with all three components active, retains 72 candidates. Removing the update and group-activity floors alone has the largest effect, leaving 261 candidates; removing PAN alone leaves 81; removing nondominance alone leaves 72, unchanged from the full pipeline, though it can reduce other configuration spaces.

Under private calibration, removing the update and group-activity floors causes the largest degradation (95.1% accuracy, 16.3% disparity, versus 97.3% and 10.9% for the full method), since the resulting shortlist retains configurations with too few group-active updates for DP-HyPO to reliably distinguish; removing PAN alone has a smaller but still consistent effect (97.1%, 13.1%). Unlike the main private-calibration results in Section 4.2, which use $\varepsilon_{\text{hypo}} = 6.6$ and $r = 0.45$, this ablation uses a separate calibration budget of $\varepsilon_{\text{hypo}} = 7.5$ and $r = 0.4$, chosen to additionally test the robustness. All four variants are run with the same expected DP-HyPO search horizon $E[T_H]$, so their private-calibration runtimes remain close (0.96–0.97 GPU-hours) even though the underlying shortlist sizes differ by more than a factor of three: DP-HyPO samples the same expected number of candidates from each menu regardless of the menu's size, so the private-calibration cost tracks the fixed search horizon.

Under exhaustive search, all four variants converge to identical accuracy and disparity, confirming that none of the screening components discards a high-utility configuration from the search space – their effect is purely on the size of the space that must be privately searched, not on which optima it contains. This is reflected in the exhaustive-search runtimes: the full EQUICONFIG pipeline requires 3.21 GPU-hours, close to vanilla DP-SGD (2.76) and substantially below DPSGD-F (10.15) and DPSGD-G-A (73.47), even though EQUICONFIG's search additionally spans both Gaussian and PLRV-O noise families, while the baselines search Gaussian noise only. The reported exhaustive-search runtime therefore reflects both the number of retained candidates and the per-candidate training cost of the configurations explored, rather than candidate count alone.