# SlaClip: Gradient Norm Slacks can be Indicator for Adaptive Clipping in DP-SGD

ICML 2026 Spotlight

Shuyan Zou[1], Shaowei Wang[2], Zhanxing Zhu[1], Jin Li[2], Changyu Dong[2], Vladimiro Sassone[1], Han Wu[1]

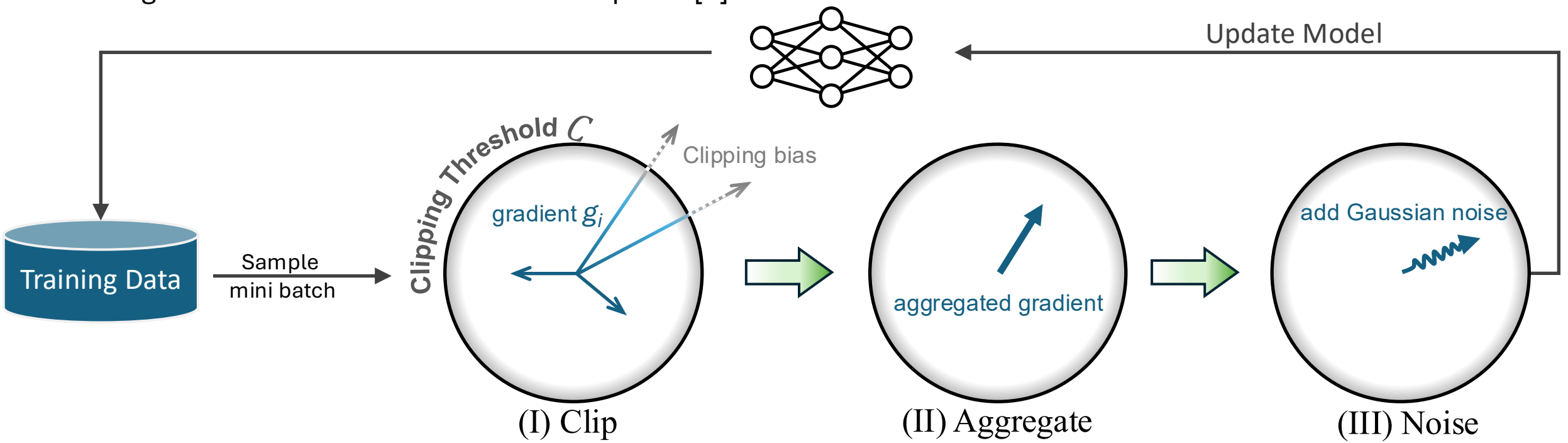
[1] University of Southampton
[2] Guangzhou University

✉ wangsw@gzhu.edu.cn
h.wu@soton.ac.uk

# I. Motivation: Vanilla DP-SGD and the clipping problem

DP-SGD protects training data by

- Clipping each per-sample gradient;
- Adding Gaussian noise before the model update [1].



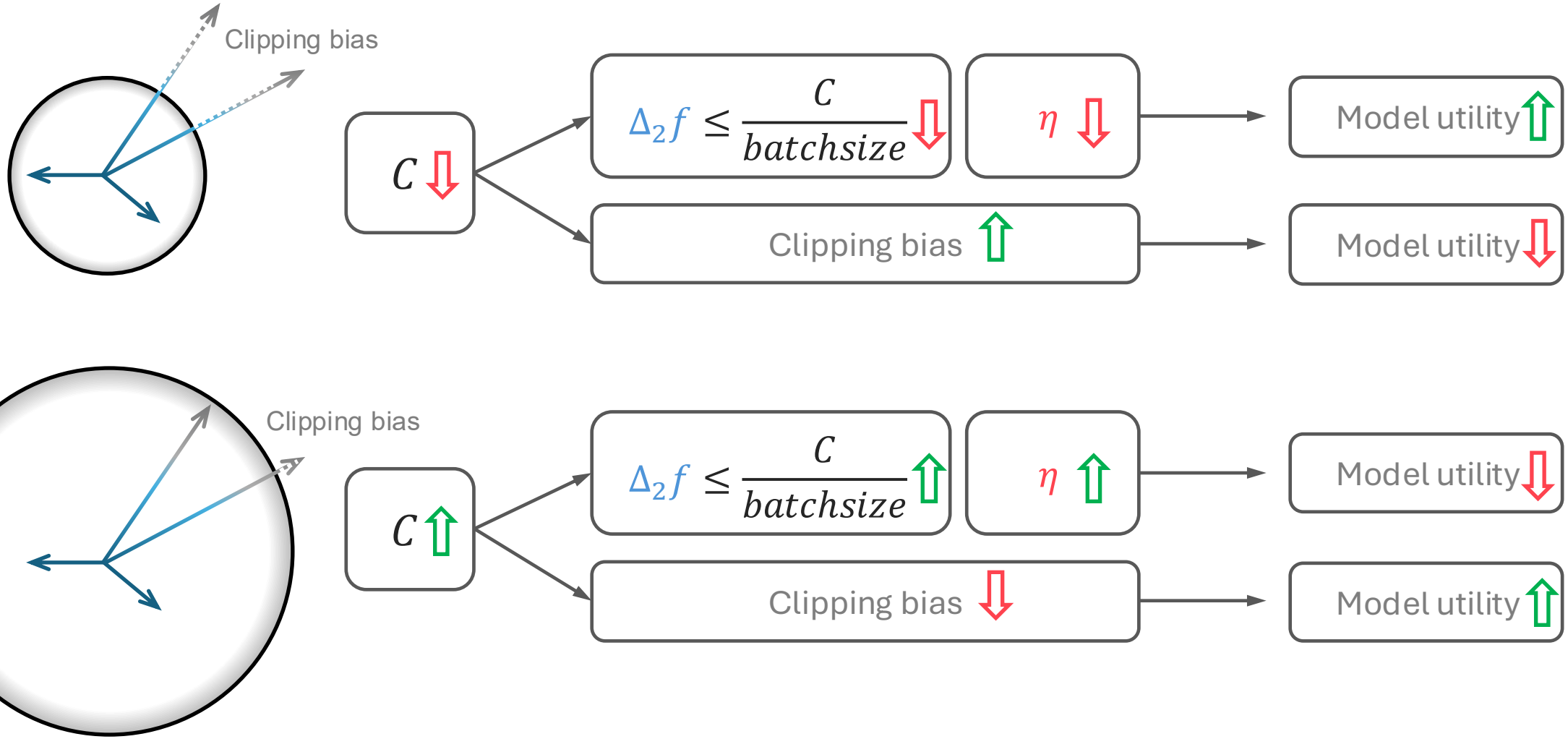

$$Clip_{C_t}(\mathbf{g}_{t,i}) = \begin{cases} \mathbf{g}_{t,i}, & \|\mathbf{g}_{t,i}\| \le C_t, \\ C_t \cdot \mathbf{g}_{t,i}/\|\mathbf{g}_{t,i}\|, & \|\mathbf{g}_{t,i}\| > C_t. \end{cases}$$

$$\tilde{\mathbf{g}}_t = \frac{1}{B}\sum_{i\in\mathcal{B}_t} Clip_{C_t}(\mathbf{g}_{t,i}) + \mathcal{N}\left(\mathbf{0}, \left(\frac{\sigma C_t}{B}\right)^2 \mathbf{I}_d\right)$$

[1] Abadi, Martin, et al. "Deep learning with differential privacy." *CCS* 2016.

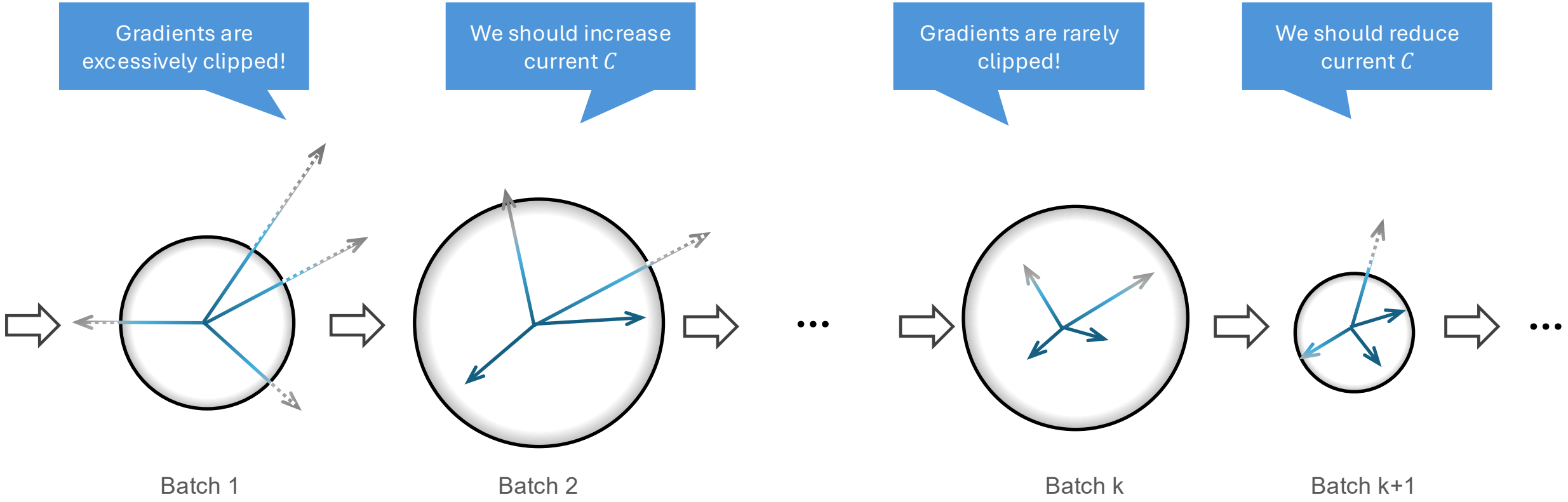
# I. Motivation: DP-SGD and the clipping problem

- Clipping threshold $\mathbf{C}$: caps each per-sample gradient's $\ell_2$ norm.
- If Clipping threshold $\mathbf{C}$ is too small, useful gradient directions are over-clipped.
- If Clipping threshold $\mathbf{C}$ is too large, the Gaussian noise becomes too large.



# II. Prior adaptive clipping: useful but costly or rigid

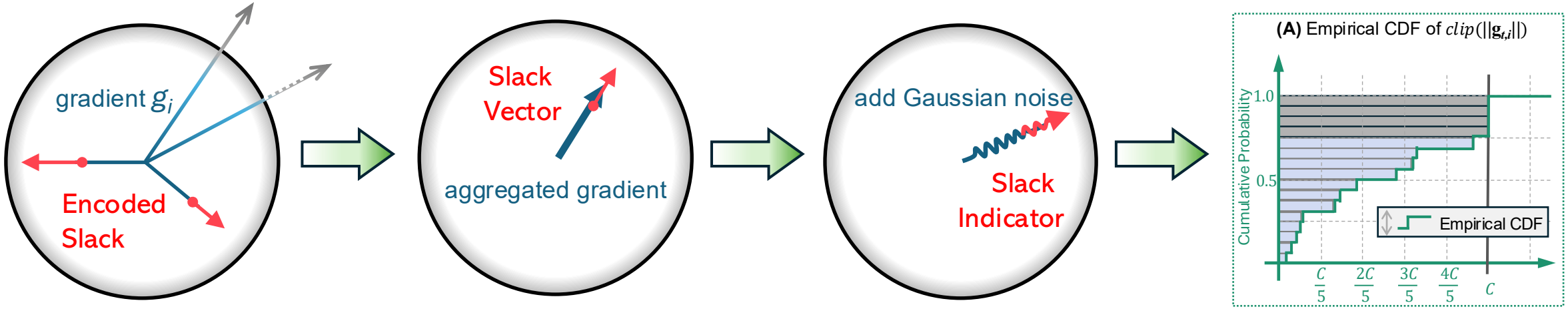
- Adap-Clip [2] tracks the unclipped ratio and adjusts the threshold toward a target quantile (50%).
- Limitation 1: estimating this statistic typically requires an additional private measurement.
- Limitation 2: A fixed target can become suboptimal when the gradient norm distribution changes, especially when many small norm gradients appear later in training.




[2] Thakkar, O., Andrew, G., & McMahan, H. B. (2019). Differentially private learning with adaptive clipping. *arXiv e-prints*, arXiv-1905.

# III. SlaClip: use clipping slack as a no-extra-privacy signal

- SlaClip asks whether the **Clipping Slack** can be encoded into extra coordinates, while keeping the extended gradient within the same norm bound $C$.
- This makes it possible to release a noisy binned Cumulative Distribution Function (CDF) signal through the same Gaussian mechanism, without an additional private query.




**SlaClip:** (I) Clip (II) Aggregate (III) Noise

# III. SlaClip: use clipping slack as a no-extra-privacy signal

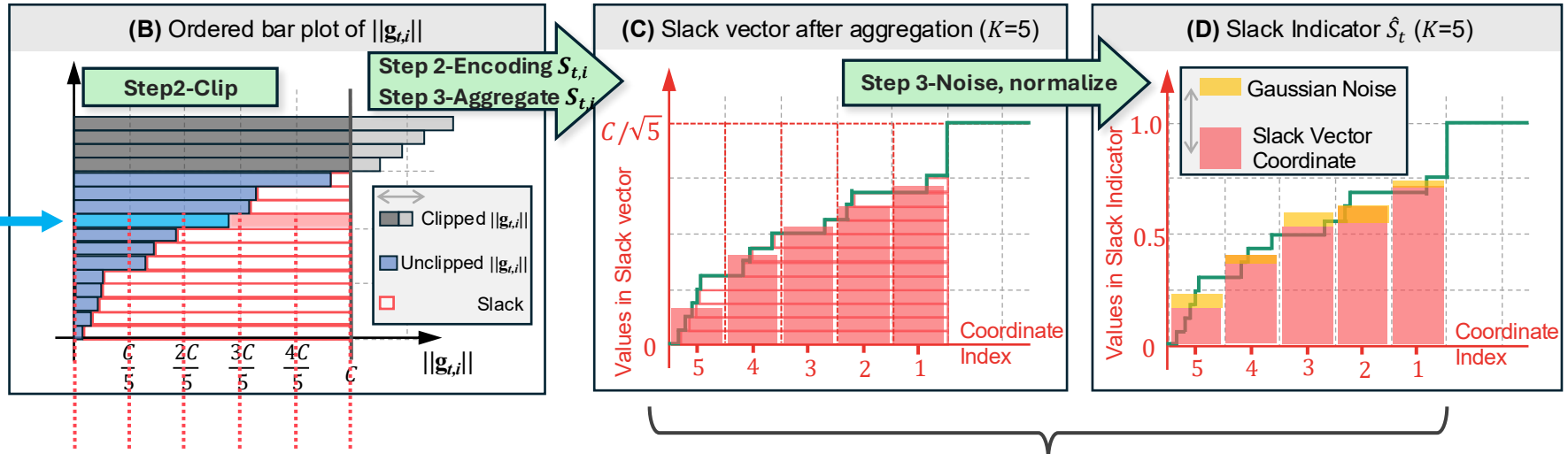
- The extended gradient is constructed to make sure its norm is still bounded by the same clipping threshold $\mathbf{C}_t$.
- Therefore, the extended vector preserves the same global $\ell_2$ sensitivity bound as vanilla DP−SGD.

**Slack encoding**

$$\mathbf{g}_{t,i}^{+} = \left[\, Clip_{C_t}(\mathbf{g}_{t,i});\ \mathbf{s}_{t,i} \,\right]$$

$$\mathbf{s}_{t,i} \triangleq \left[\lambda\, \mathbf{1}^{(a)};\ \ b;\ \mathbf{0}\right]_{t,i}$$

$$\sqrt{K} \cdot \max\{\, C_t - \|\mathbf{g}_{t,i}\|,\ 0 \,\} = a\lambda + b.$$




Slack vector example: $[0,\ 0,\ b,\ \lambda,\ \lambda]$

$$\lambda = c/\sqrt{K}$$

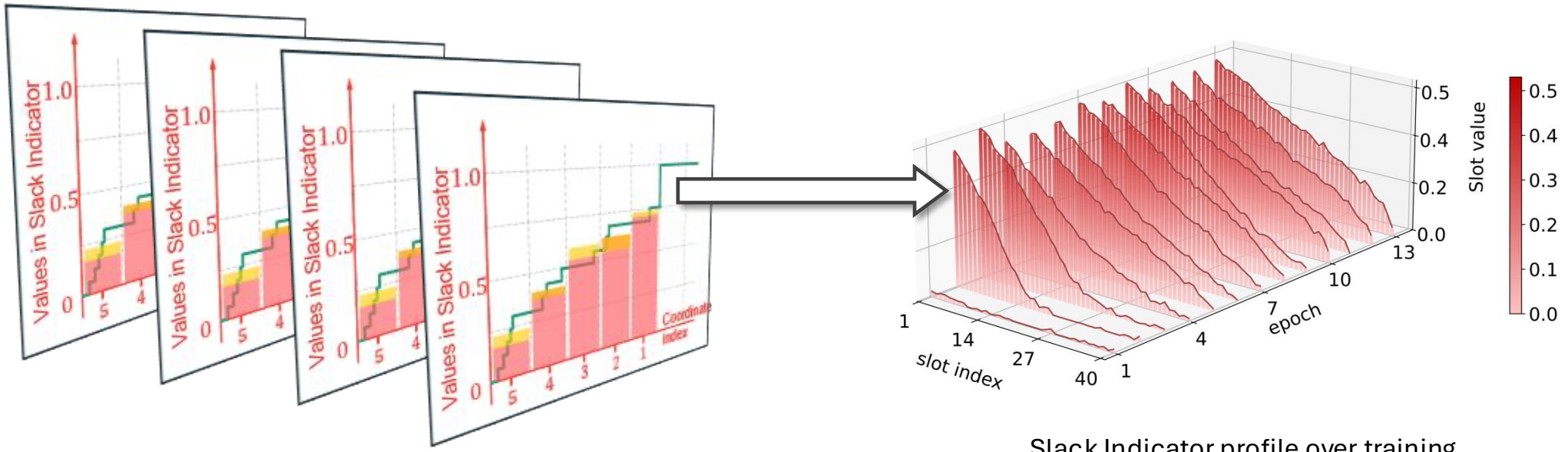
$$\widetilde{\mathbf{g}}_t^{+} = \overbrace{\frac{1}{B}\sum_{i\in\mathcal{B}_t}\mathbf{g}_{t,i}^{+}}^{f_{avg}^{+}} + \mathcal{N}\left(\mathbf{0},\ (\frac{\sigma C_t}{B})^2\mathbf{I}_{d+K}\right)$$

# IV. What information does the Slack Indicator provide?

- Slack Indicator can be interpreted as a noisy, binned CDF estimate of $clip(||\mathbf{g}_{t,i}||)$ below the current threshold.
- SlaClip adapts the clipping level using both near-threshold and near-zero CDF information.

$$\text{Dynamic target clipping ratio:} \quad \gamma_t \triangleq \Pi_{[0,1]}(1-(1-\hat{s}_{t,K}/C_t)/2)$$

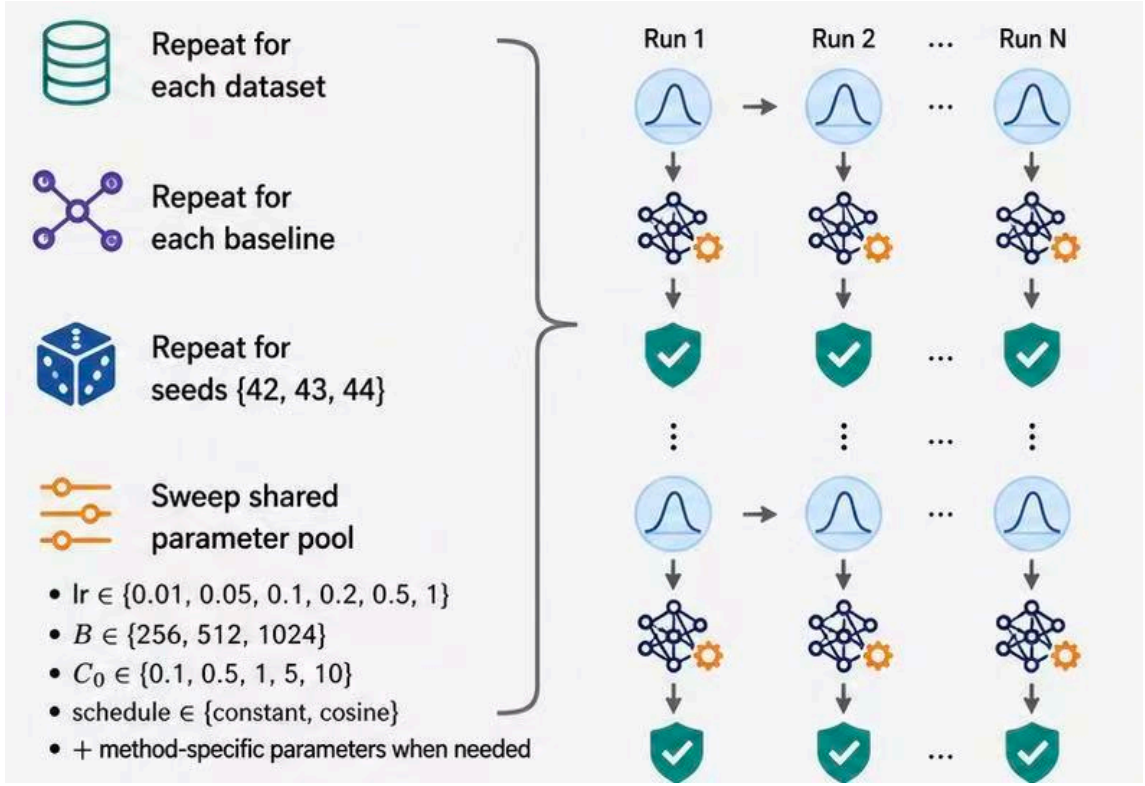
$$\text{Clipping threshold adaptation:} \quad C_{t+1} \leftarrow C_t \exp(\eta\,(\gamma_t - \hat{s}_{t,1}))$$

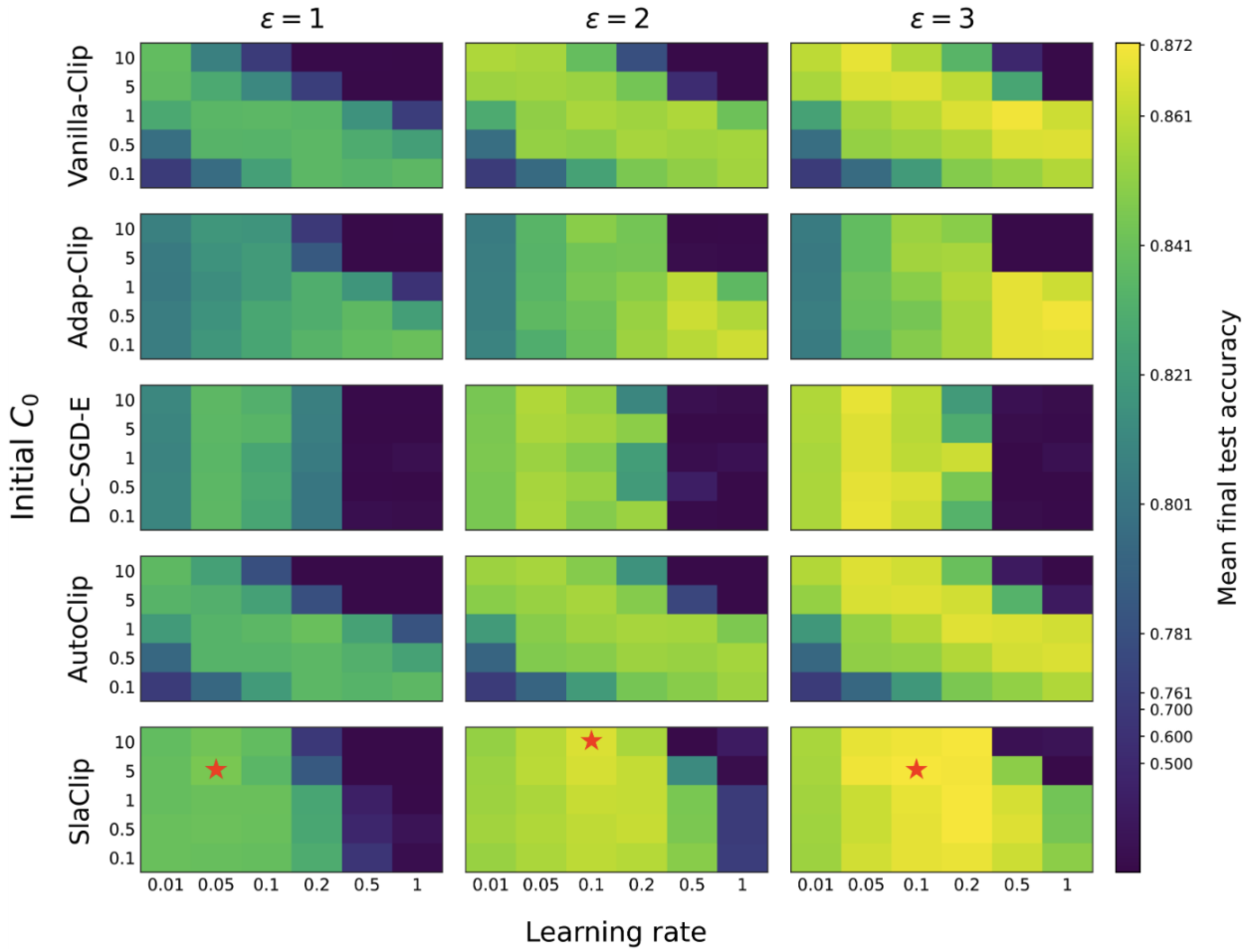



Slack Indicator profile over training

# V. Evaluation:

- For the main comparison, we use a fairly tuned protocol under matched privacy budgets.
- Across these settings, SlaClip achieves the best or second-best private accuracy in every setting.
- SlaClip has a broader high-accuracy region around $\mathbf{lr}$ near 0.1, and is less sensitive to the initial threshold $\mathbf{C_0}$.



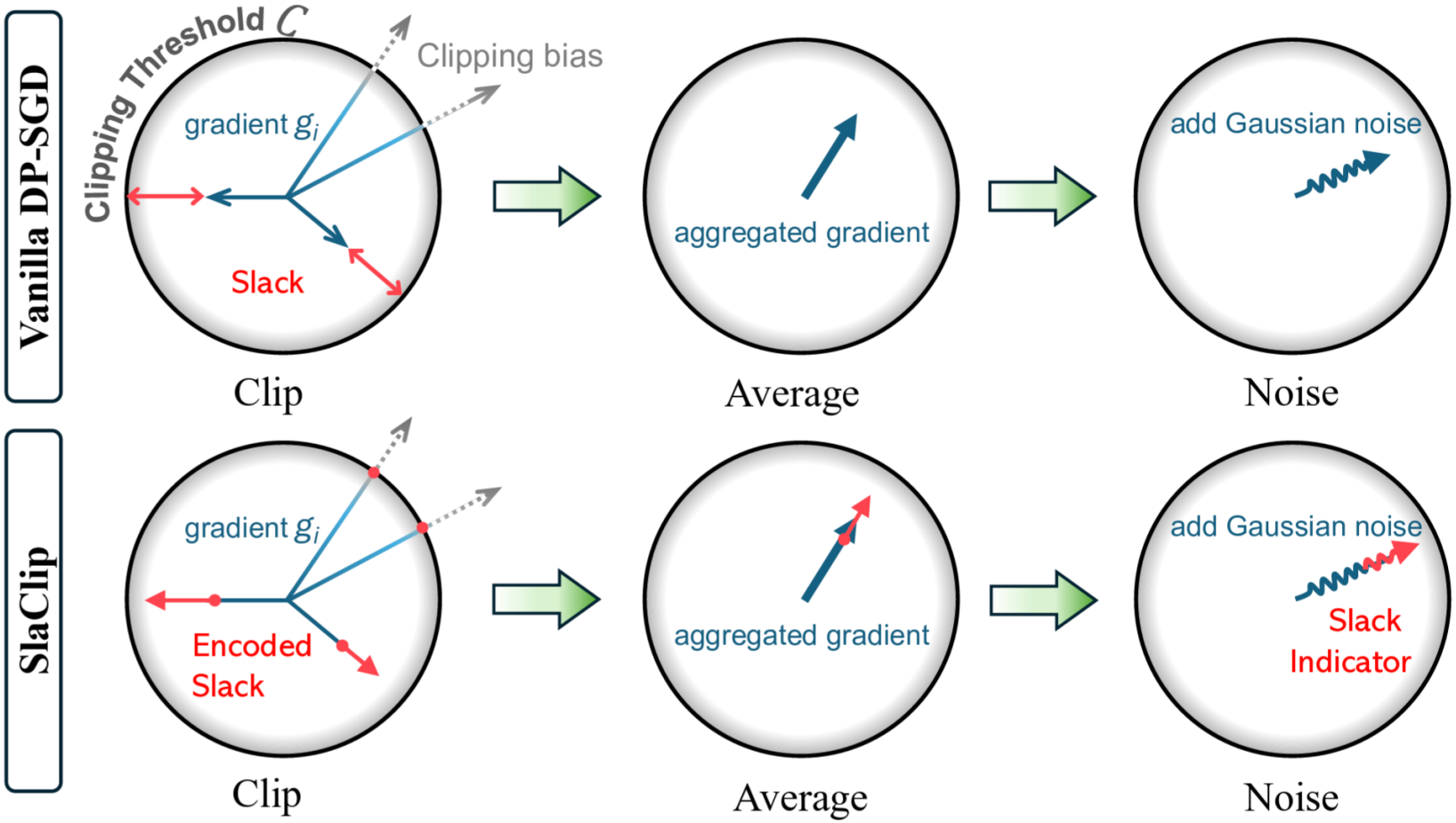



# VI. Takeaway:

Code link:
https://github.com/ZsyRock/SlaClip

SlaClip turns the clipping slack into a privacy-preserving CDF signal, released through the same Gaussian mechanism as vanilla DP-SGD.

SlaClip is plug-and-play with the standard DP-SGD pipeline, and provides a richer adaptive clipping signal than fixed-target clipping rules.




**SlaClip: Gradient Norm Slacks can be Indicator for Adaptive Clipping in DP-SGD**