OPTIMAL OBSERVABLES FROM OPTIMAL BINNING

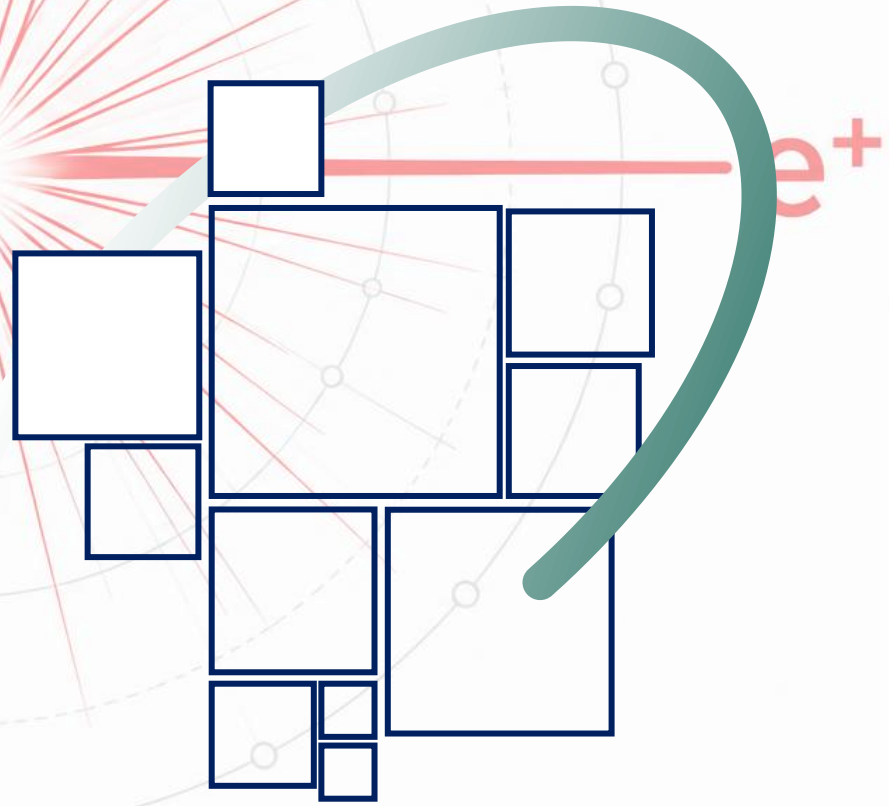
Lingfeng Li

李凌风

Nanjing Normal U.

Oct. 23 . 2026, Shanghai

The 2026 International Workshop on the High-Energy Circular Electron-Positron Collider

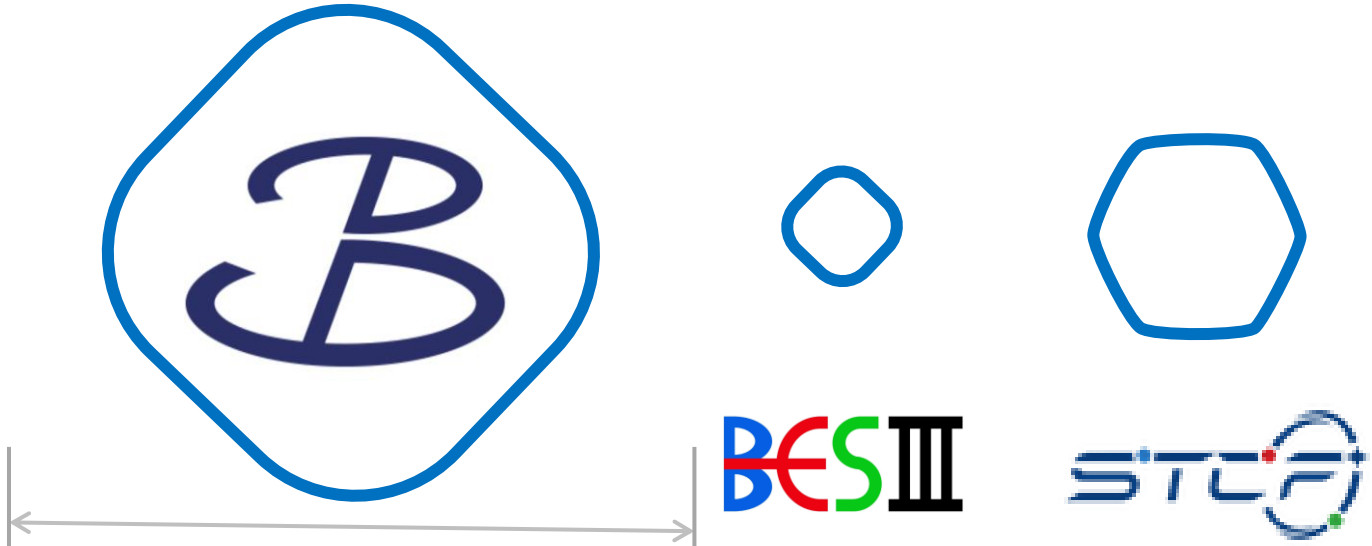


THE PRECISION ERA IS COMING (?)
THE PRECISION ERA IS NOW



~100/π km





~1 km

- A continuous timeline
- Imagine precision neutrino, gravitational wave, cosmology…

EFT Predictions for CEPC:

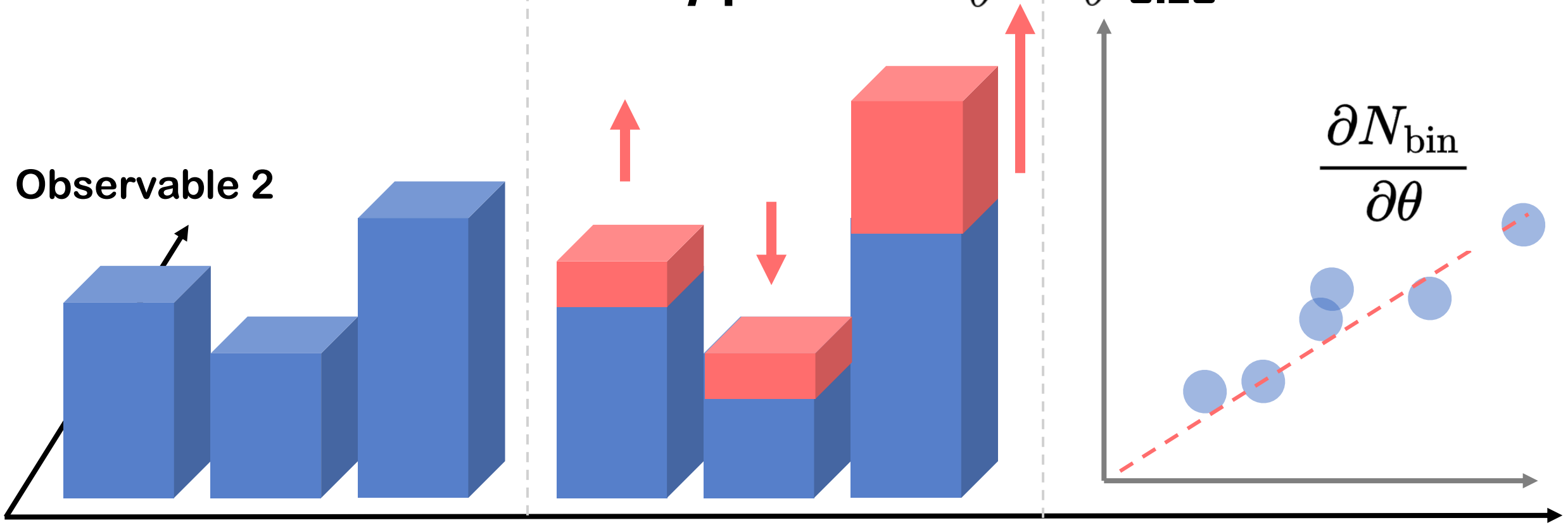
de Blas, Du, Grojean, Gu, Miralles, Peskin, Tian, Vos, Vryonidou, 2206.08326; The SMEFiT Collaboration 2105.00006; Brivio, Bruggisser, Elmer, Geoffray, Luchmann, Plehn, 2208.08454; Allwicher, Cornella, Isidori, Stefanek, 2311.00020; Ellis, Madigan, Mimasu, Sanz, You, 2012.02779 ……

Differential Measurement from Histograms

Cut SR into several bins

Counts change with theory parameter θ

(Linear) Fit to find out θ size



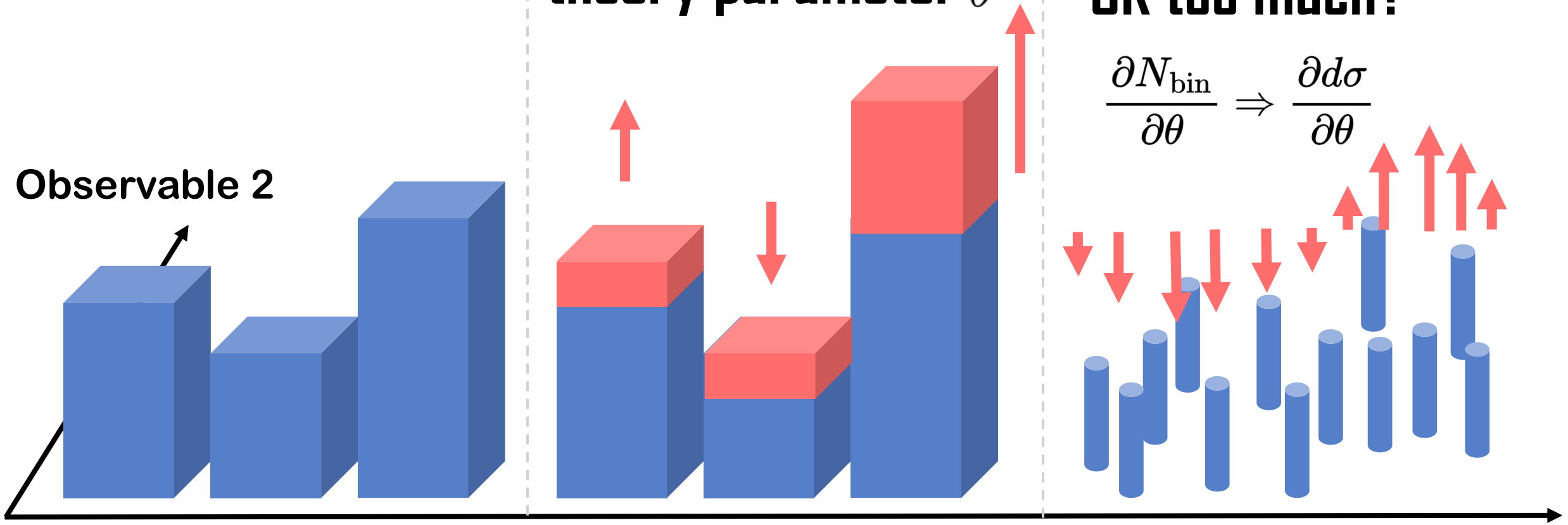
Ultimate Differential Measurement

Cut SR into several bins

Counts change with theory parameter θ

What if you dice your SR too much?

$$\frac{\partial N_{\text{bin}}}{\partial \theta} \Rightarrow \frac{\partial d\sigma}{\partial \theta}$$



From Bins to Optimal Observables

Promote into event-by-event vectors:

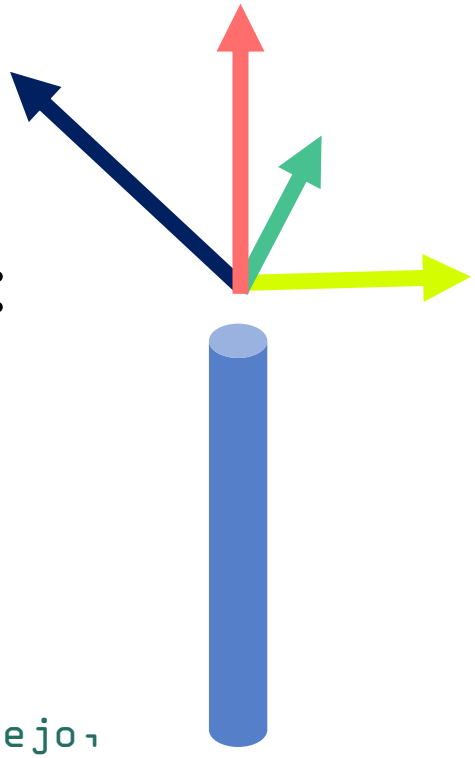
$$\vec{\alpha}_i(x) \propto \frac{\partial |\mathcal{A}(x)|^2}{\partial \theta_i}$$

Linearized likelihood:

$$L \propto \sum_{i \in \text{events}} (1 + \vec{\alpha}_i \cdot \vec{\theta}_i)$$

Can be achieved during MC:

J. Brehmer, F. Kling, I. Espejo, K. Cranmer, 1907.10621……

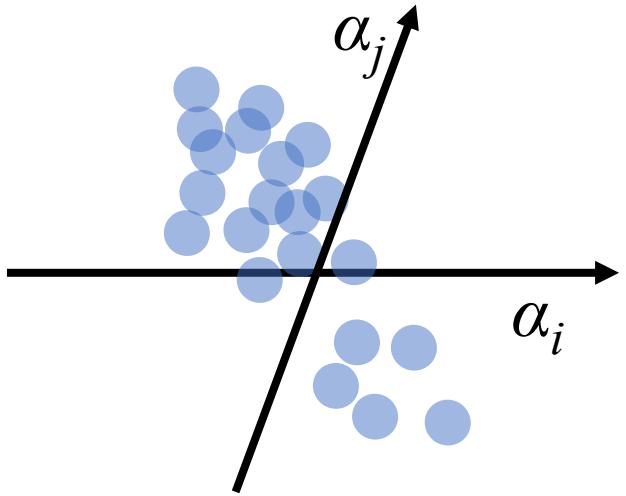


The Optimal Observable (OO)

M. Diehl and O. Nachtmann, Z.Phys.C 62 (1994)

$$V \propto \left(\det \sum_i \alpha_i^T \alpha_i \right)^{-\frac{1}{2}}$$

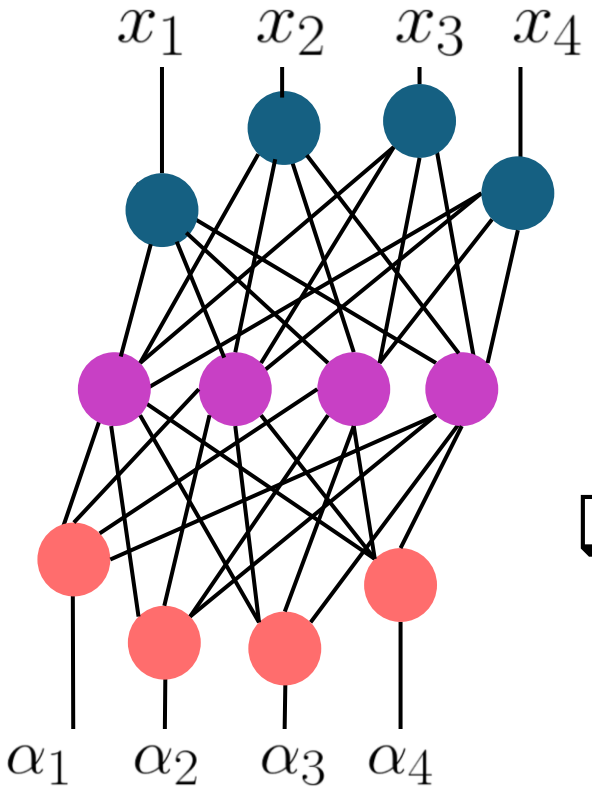
~Fisher Information Matrix from the full observable



Saturating the Cramér–Rao bound

H. Cramér, C.R. Rao, 1946

SBI for the Precision Task



From Classification to Regression

- SALLY (Score Approximates Likelihood Locally) strategy, Optimal Observable extension in ML

  J. Brehmer, K. Cranmer, G. Louppe, J. Pavez 1805.00013; J. Brehmer, K. Cranmer, G. Louppe, J. Pavez 1805.00020……

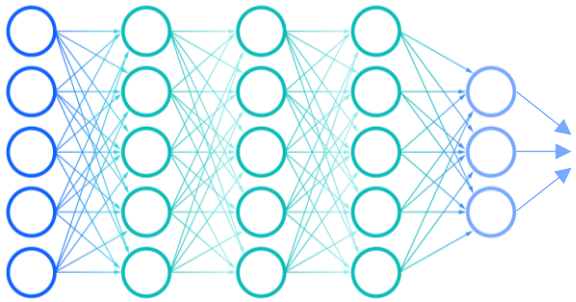
- In the precision context, the loss takes the form:

$$L = \sum_i (\alpha_i - \bar{\alpha}_i)^T G (\alpha_i - \bar{\alpha}_i)$$

*Other SBI frameworks also apply, but not optimized for ultra-high precision: Chen, Glioti, Panico, Wulzer, 2007.10356; Ambrosio, Hoeve, Madigan, Rojo, Sanz, 2211.02058; R. Mastandrea, B. Nachman, T. Plehn, 2405.15847……

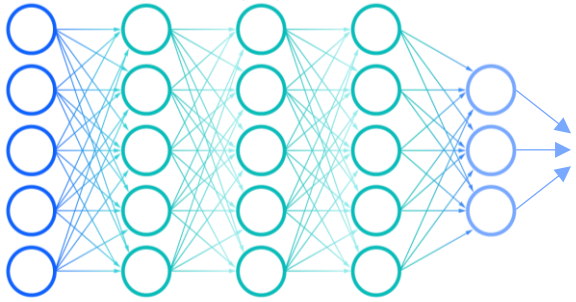
See also
LFL, T. Liu, S.J. Xu, 2004.15013

Precision ML for ATGC Measurements

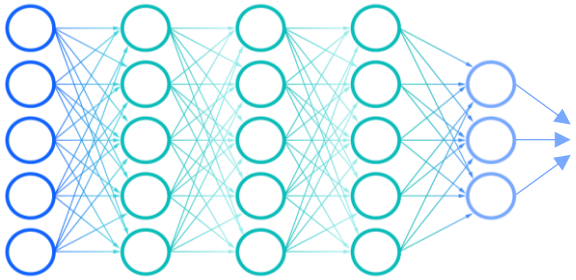


Classifier Network × Sally Regressor

(cross entropy loss)



OR

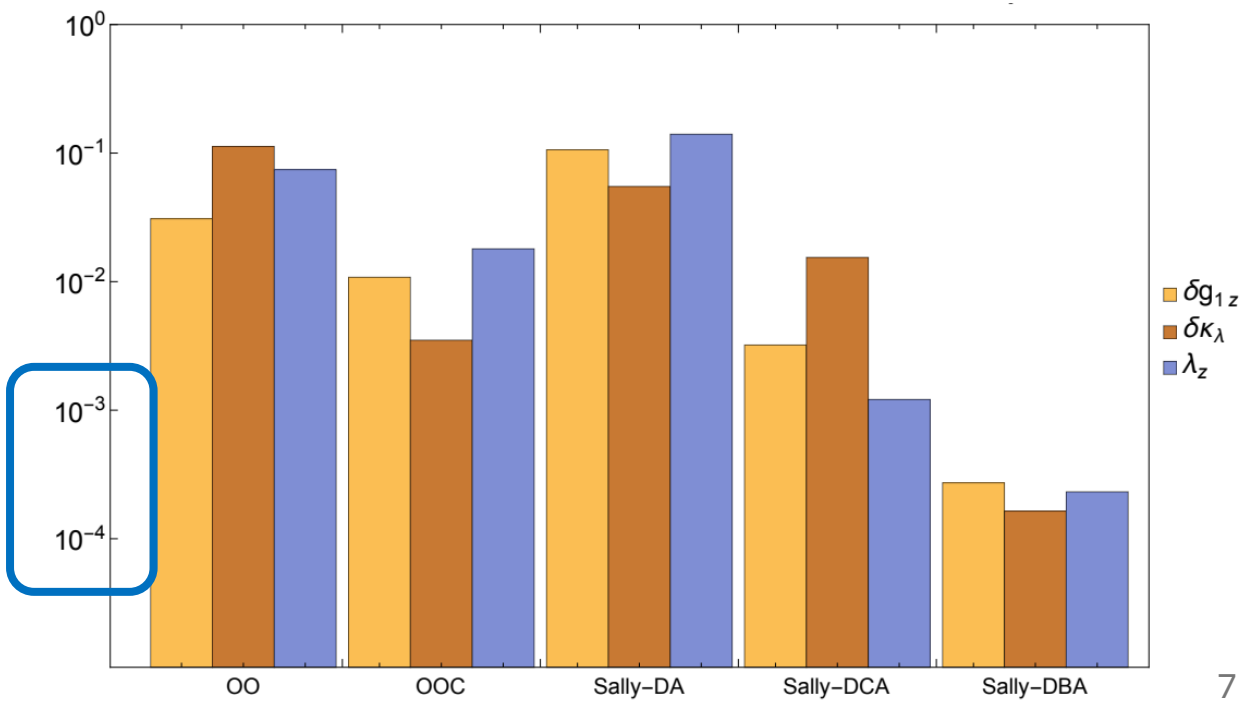


Sally Regressor

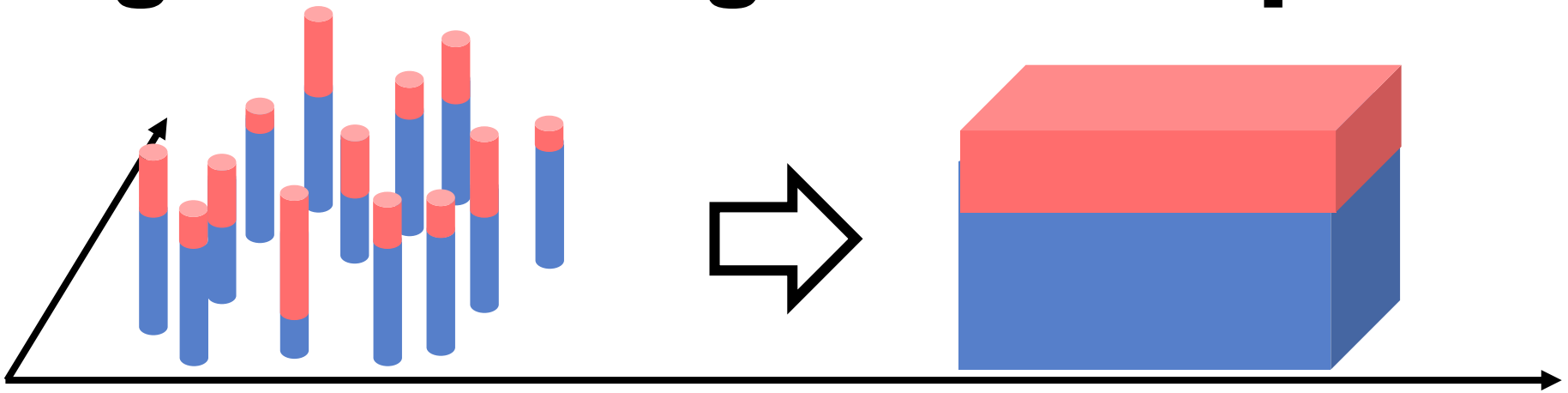
(training on reweighted samples)

Two equivalent frameworks to remove bias when there are irreducible background noises

arXiv:2401.02474
S. Chai, J. Gu, LFL

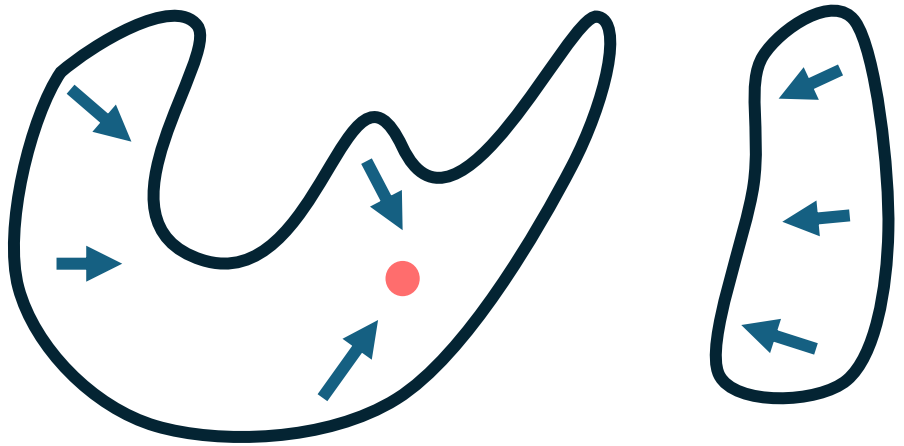


Unified Language: Histogram-binning in the α-space



Binned statistics is just averaging the OO vector inside each the bin:

$$\frac{\partial N_{\text{bin}}}{\partial \theta} \propto \sum_{i \in \text{bin}} \frac{\partial L_i}{L_i \partial \theta} = \sum_{i \in \text{bin}} \alpha_i$$

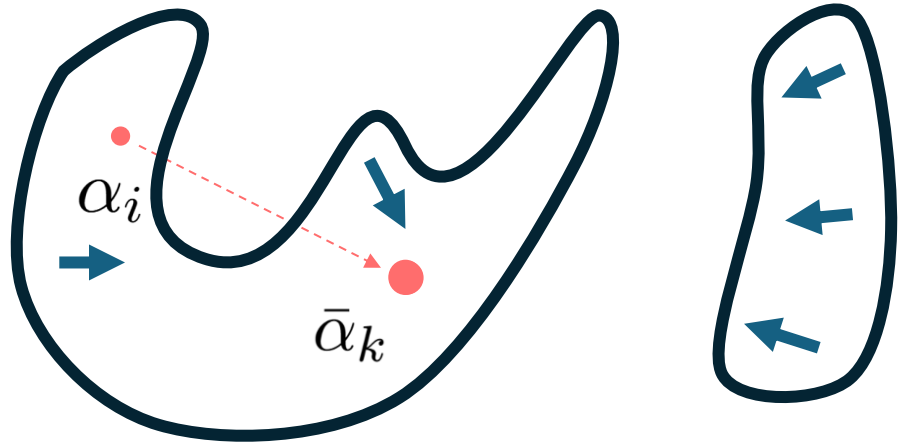


...but according to different criteria

Averaging a Single Event into a Bin

To evaluate the loss from α-space contraction, recall the matrix determinant lemma:

$$\det(A + \mathbf{u}\mathbf{v}^T) = (1 + \mathbf{u}^T A^{-1}\mathbf{v}) \det(A)$$



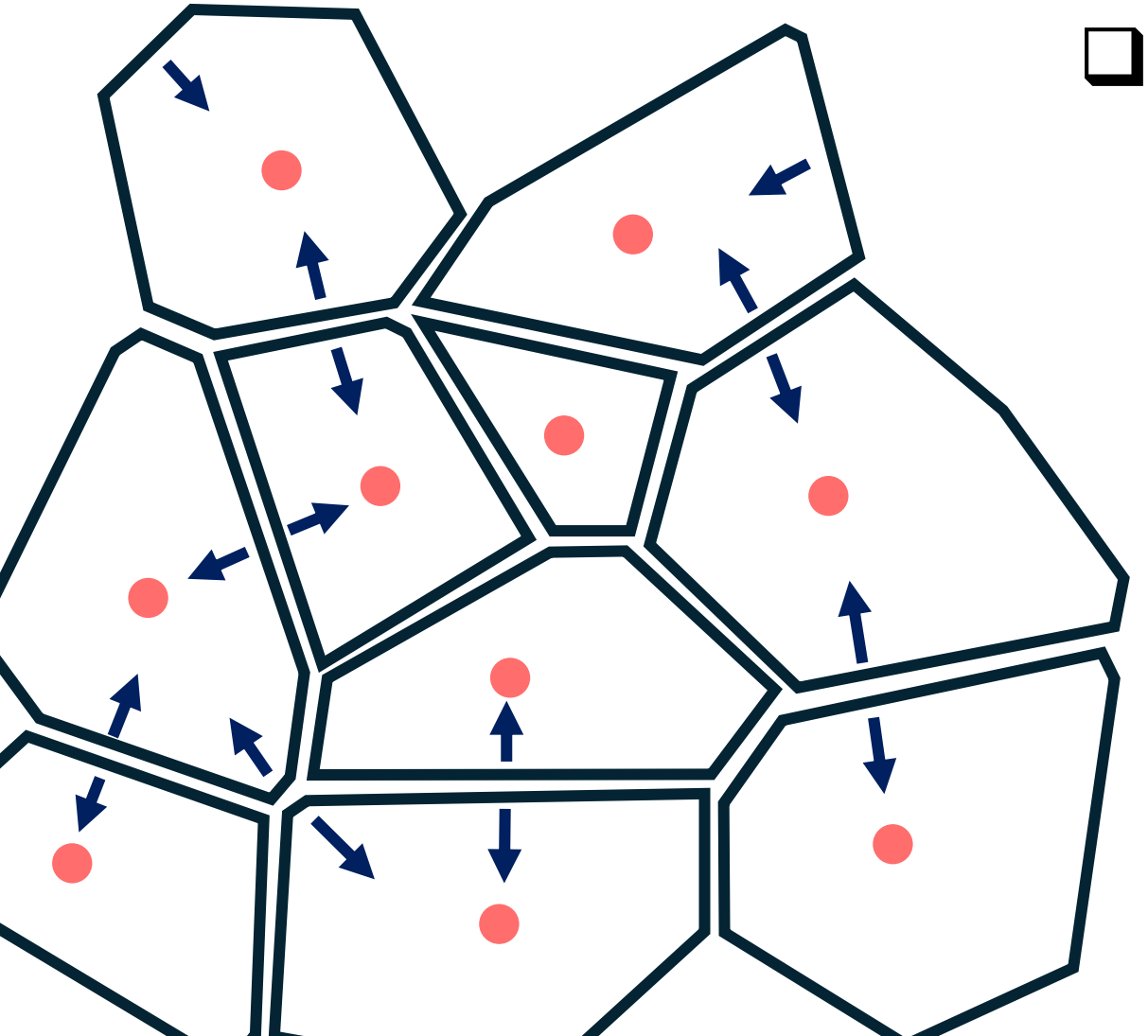
$$F \Rightarrow F'$$

$$\begin{aligned}
\det(F') &\simeq \det(F_{\bar{i}}) \left[1 + \delta_{ki}^T F_{\bar{i}}^{-1} V_k + V_k^T F_{\bar{i}}^{-1} \delta_{ki}\right] \\
&\simeq \det(F_{\bar{i}}) \left(1 + 2\,\bar{\alpha}_k^T F_{\bar{i}}^{-1} \alpha_i - \bar{\alpha}_k^T F_{\bar{i}}^{-1} \bar{\alpha}_k\right) \\
&= \det(F_{\bar{i}}) \left[1 - (\alpha_i - \bar{\alpha}_k)^T F_{\bar{i}}^{-1} (\alpha_i - \bar{\alpha}_k) + \alpha_i^T F_{\bar{i}}^{-1} \alpha_i\right]
\end{aligned}$$

The (only) binning dependent term: a norm-2 distance measure

Using Tr(F) gives similar results

“Optimal Transportation” of the Optimal Observable

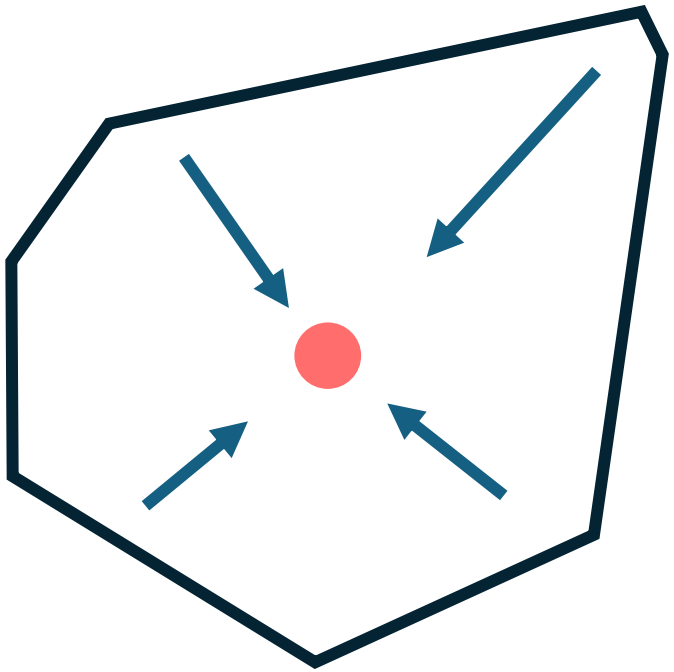


- The ideal binning would be an approximate Voronoi tessellation (or binning) in the theory space

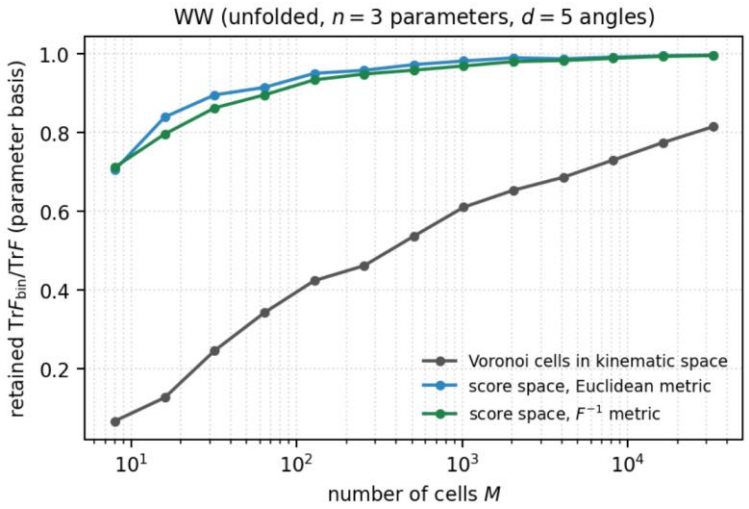
- In the small-cell limit, such an optimal binning will behave just like the optimal observable…

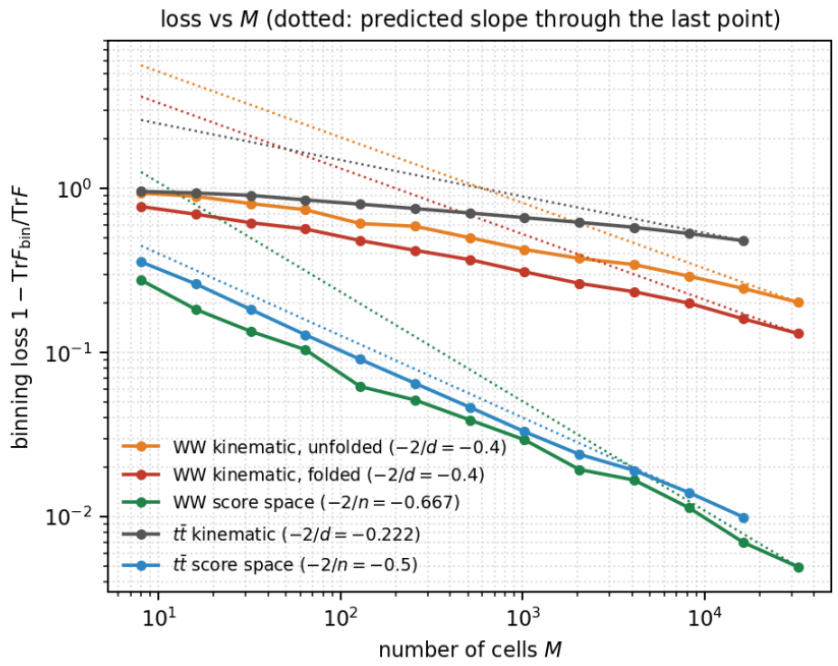
Let’s reverse the thinking!

Degenerate Loss/Noise



- Fisher information loss (also means additional noise) due to the averaging process

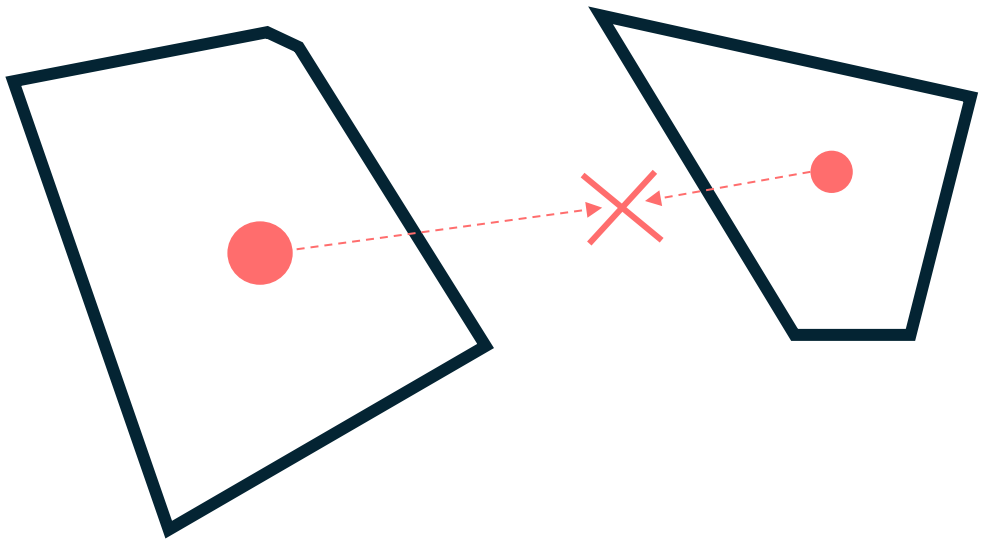




In the large number of cell (M) limit:

$$1 - \frac{\det F_{\mathrm{bin}}}{\det F_{\mathrm{full}}} \sim nM^{-2/n}$$

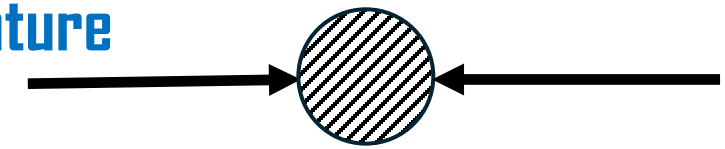
(Orbit)fold Noise/Loss



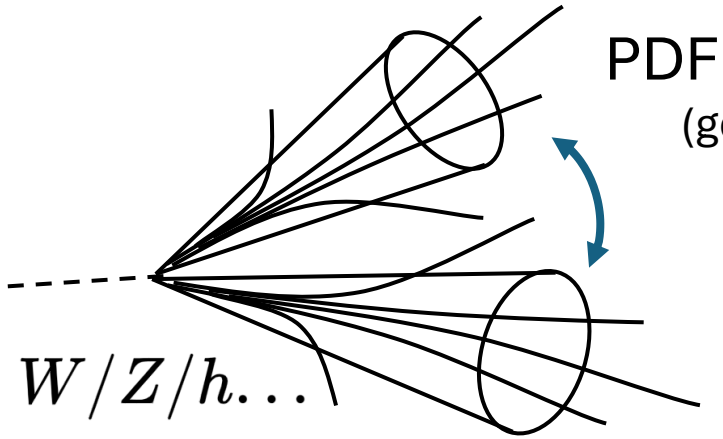
Multiple α configurations give the same observable, the no-bias α must be their WEIGHTED average:

$$\bar{\alpha} \simeq \frac{d\sigma_1\alpha_1 + d\sigma_2\alpha_2}{d\sigma_1 + d\sigma_2}$$

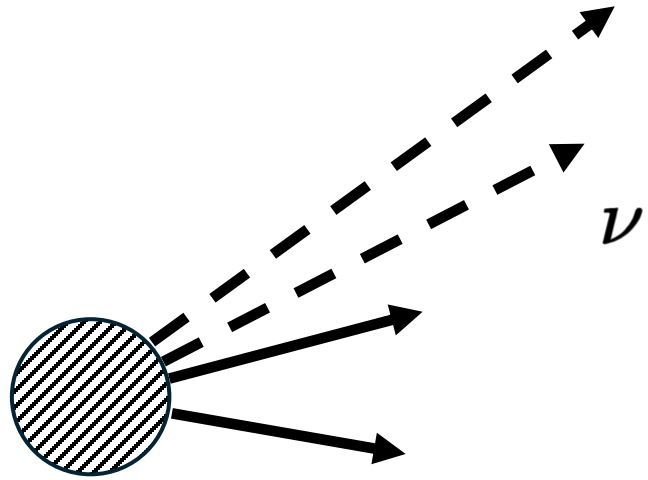
Typical in particle physics from indiscernible* final states or detector limitations due to the system's quantum nature



PDF ambiguity in $q_i\bar{q}_i \to X$
(generators only give you one)



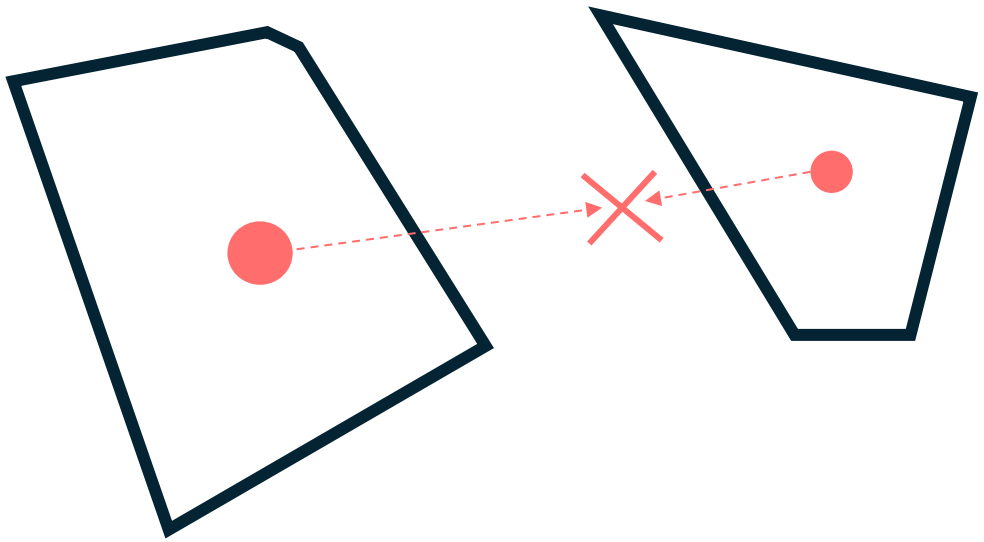
Flipped dijets



Multiple final state neutrinos

See also:
T. Ai, Q. Bi, Y. He, Z. Li, J. Liu, X.P. Wang, 2607.08346

(Orbit)fold Noise/Loss



- In the effective optimal binning perspective, less cells means more statistics within each, reducing the noise from such averaging.

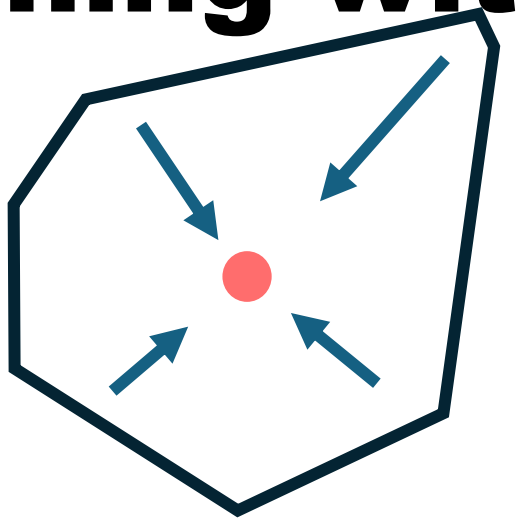
Introduces the folding term:

$$\Sigma_\Omega \sim \mathrm{Var}[\alpha|x] > 0$$

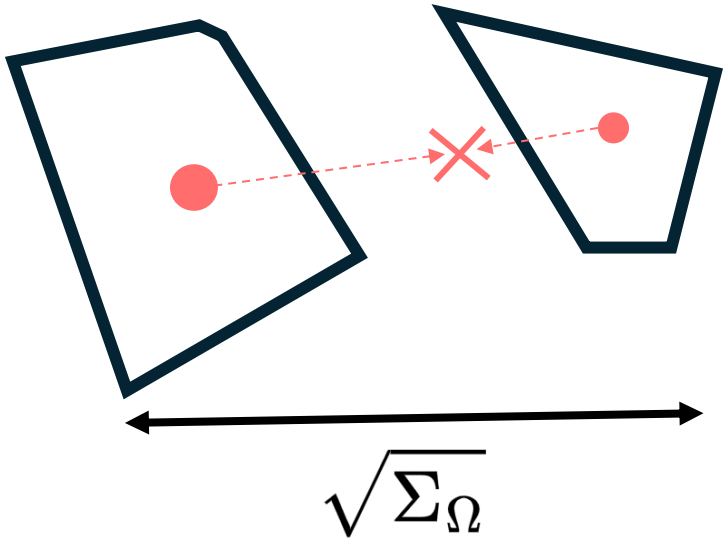
- Contradict the need for reducing the degeneracy loss.

Permanent Fisher info. Loss: $F_{\mathrm{obs}} \simeq F_{\mathrm{truth}} - \Sigma$

Scaling Behavior from Optimal Binning with Dynamical Bins



$$V_{\text{eff}}(h) \simeq C_d h^d$$

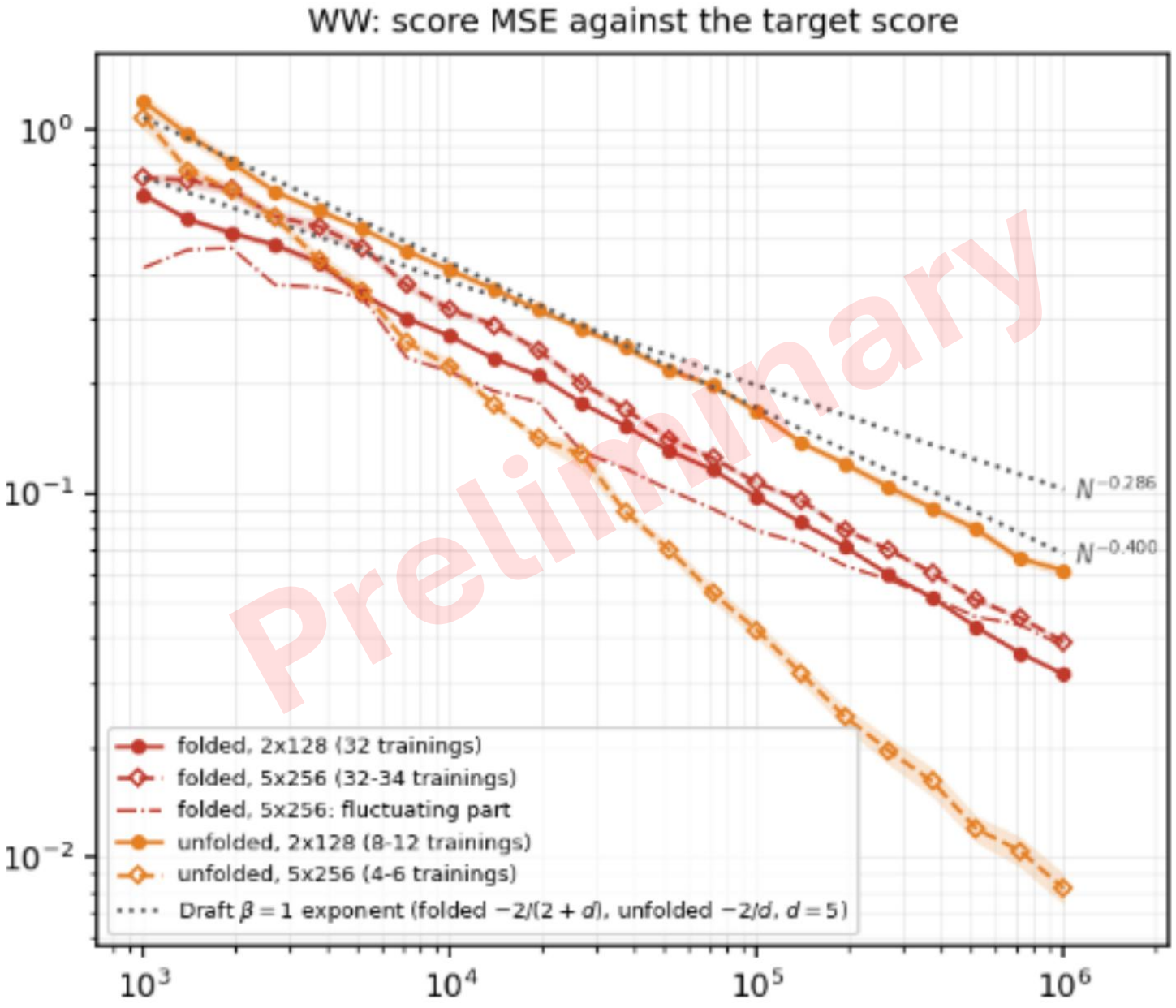


Size~ h

$\sqrt{\Sigma_\Omega}$

$$\|\bar{\alpha}(x+\delta x) - \bar{\alpha}(x)\| \sim L_\beta(x)\|\delta x\|^\beta$$

$$\text{Noise}^2 \sim |L_\beta^2 h^{2\beta}| + \left|\frac{\Sigma_\Omega}{N\rho(x)C_d h^d}\right| \propto N^{-2\beta/(2\beta+d)}$$



Preliminary

- Reduce to classic statistics (N^{-1}) when d → 0: inference being "local", no cell needed

arXiv:2xyz.abcde
Y. Fei, J. Gu, LFL, et al.

$$\propto N^{-2\beta/(2\beta+d)}$$

See also
H. Bahl, V.~Breso-Pla, A. Butter and J.I. Ramirez, 2601.13308

Summary

- The Maximal Power of a Dataset is encoded in its Fisher information and thus α-space distribution
- Biasless learning framework developed and demonstrated
- An optimal binning will be a binning in the α-space
- The learning process still applies when there is irreducible noise, at the price of FIM loss
- The network scaling behavior can be understood as a dynamical binning process.

THANK YOU!