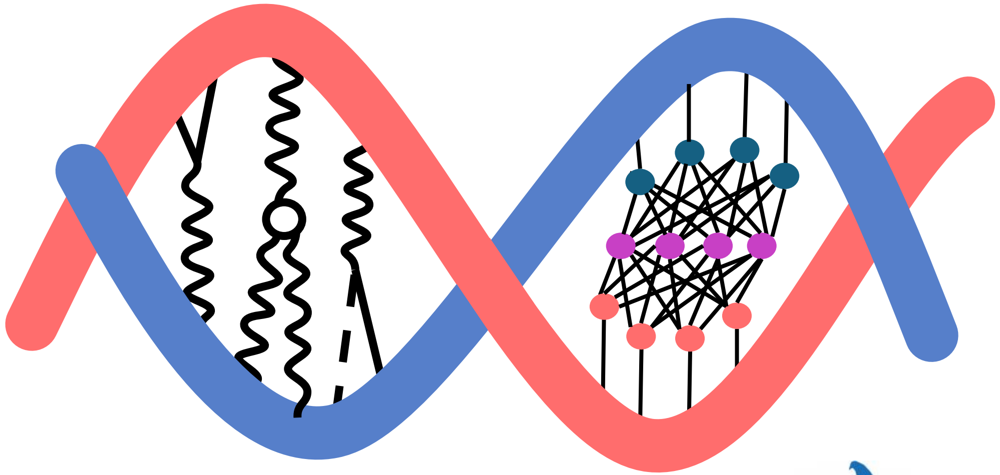
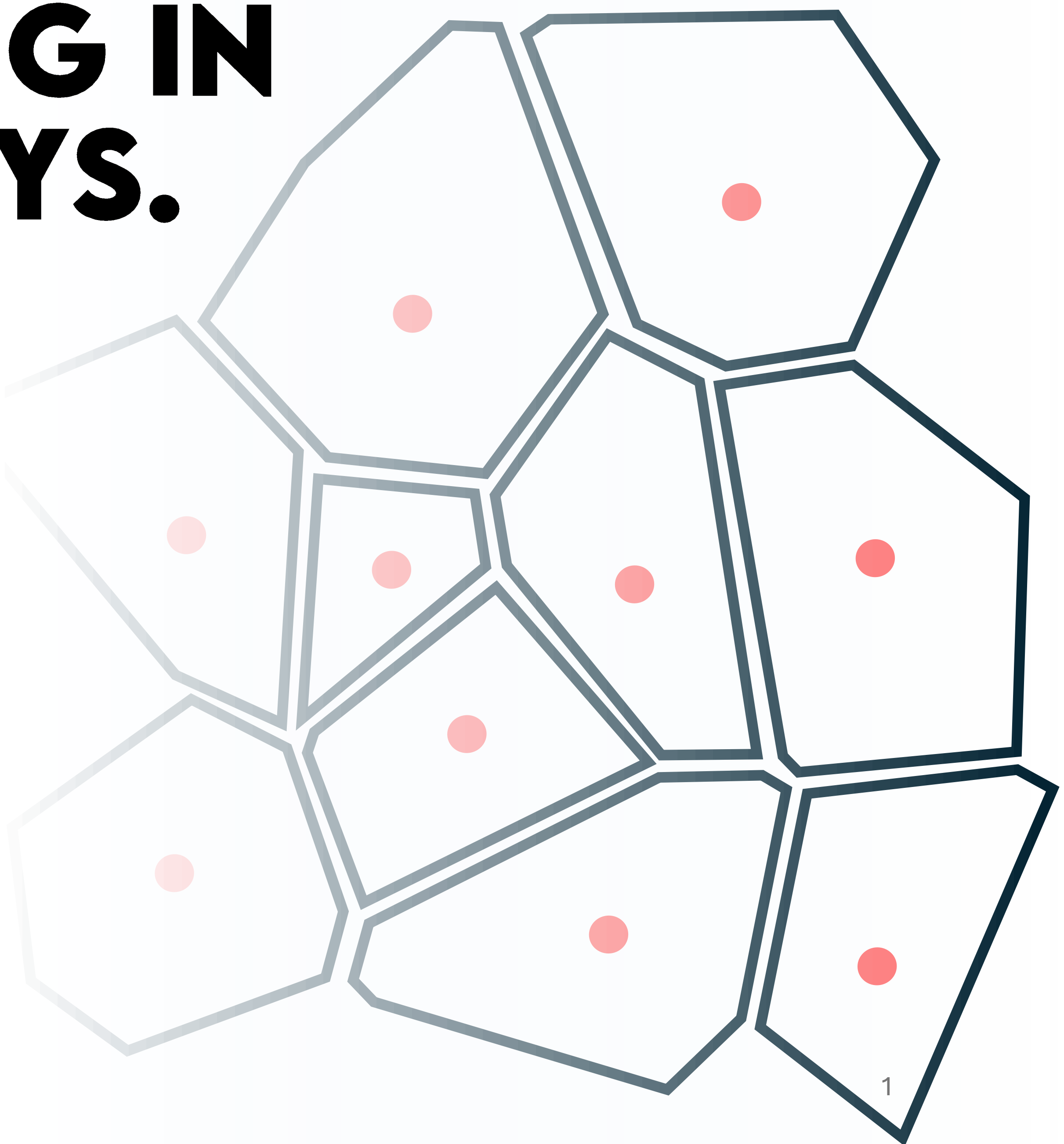
MACHINE LEARNING IN THEORETICAL PHYS.

Of the Precision Era

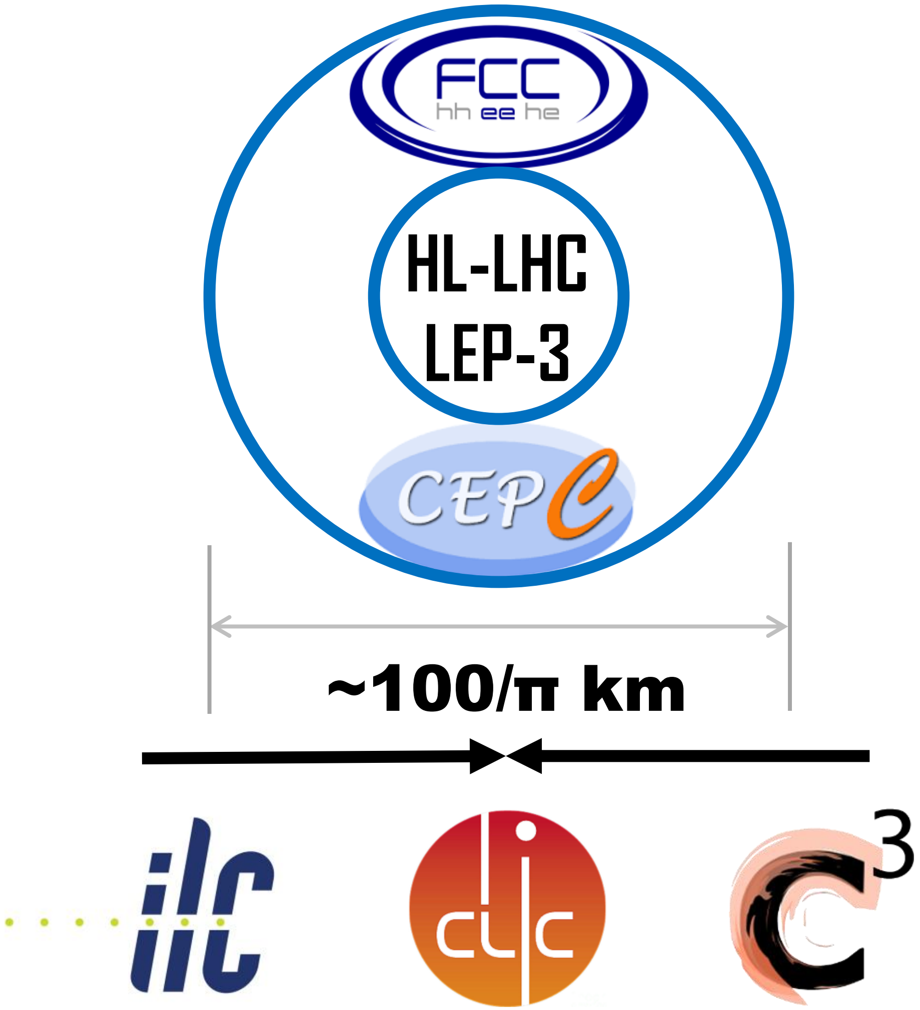


Lingfeng Li
李凌风

Apr. 16 . 2026



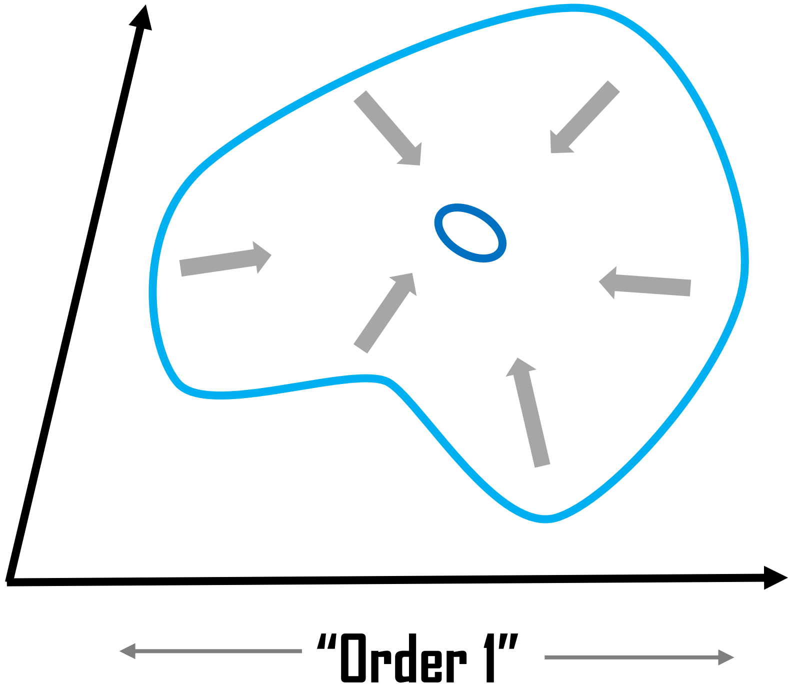
THE PRECISION ERA IS COMING (?)
THE PRECISION ERA IS NOW



- ➢A continuous timeline
- ➢Imagine precision neutrino, gravitational wave, cosmology...

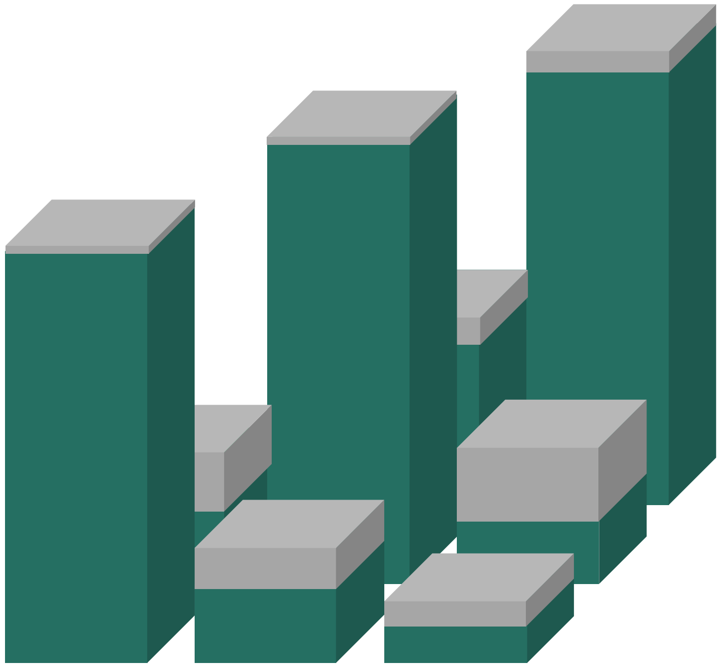
See also: Jiang Hao Yu's talk

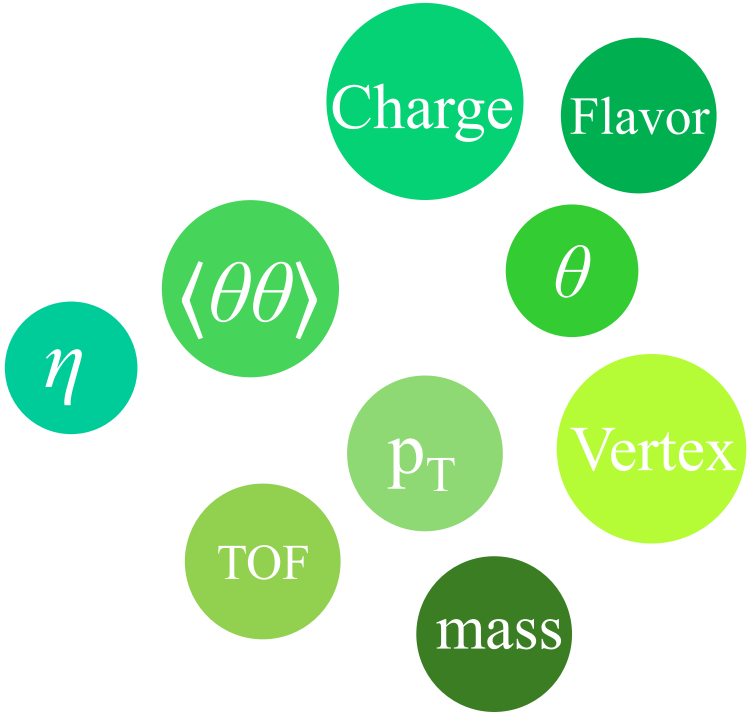
Key Features of the Precision Era



Theory space greatly shrinks to below "O(1)" (and likely linearized)

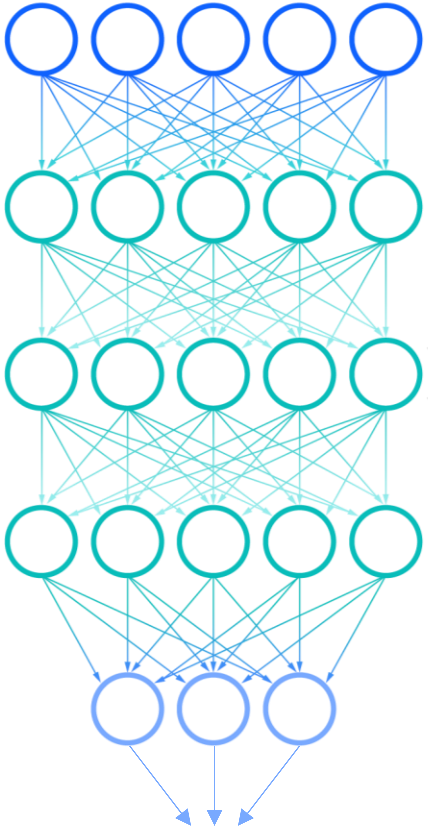
"Background" may not exist, only difference in differential distributions



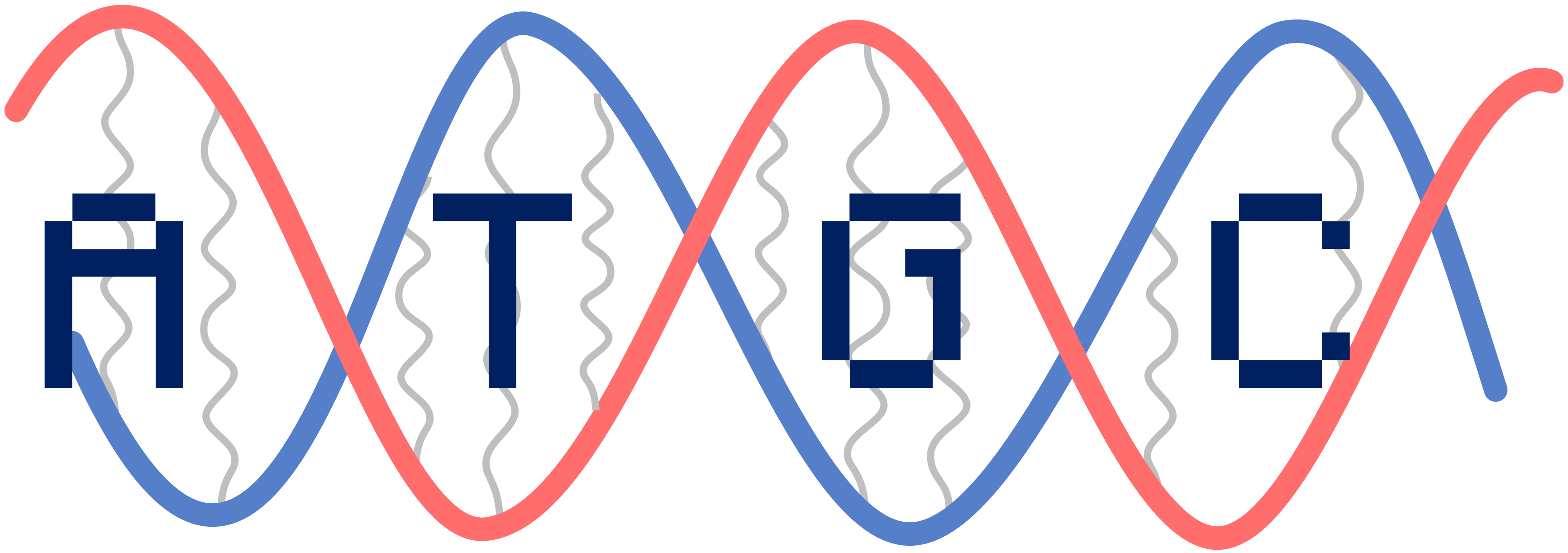


Input information lives in high dimensionality with noise

Precision SMEFT with Simulation Based Inference



$$L = F(x_i, l_i) + \alpha(x_i) + \ldots$$



Anomalous Triple Guage Coupling

arXiv:2401.02474

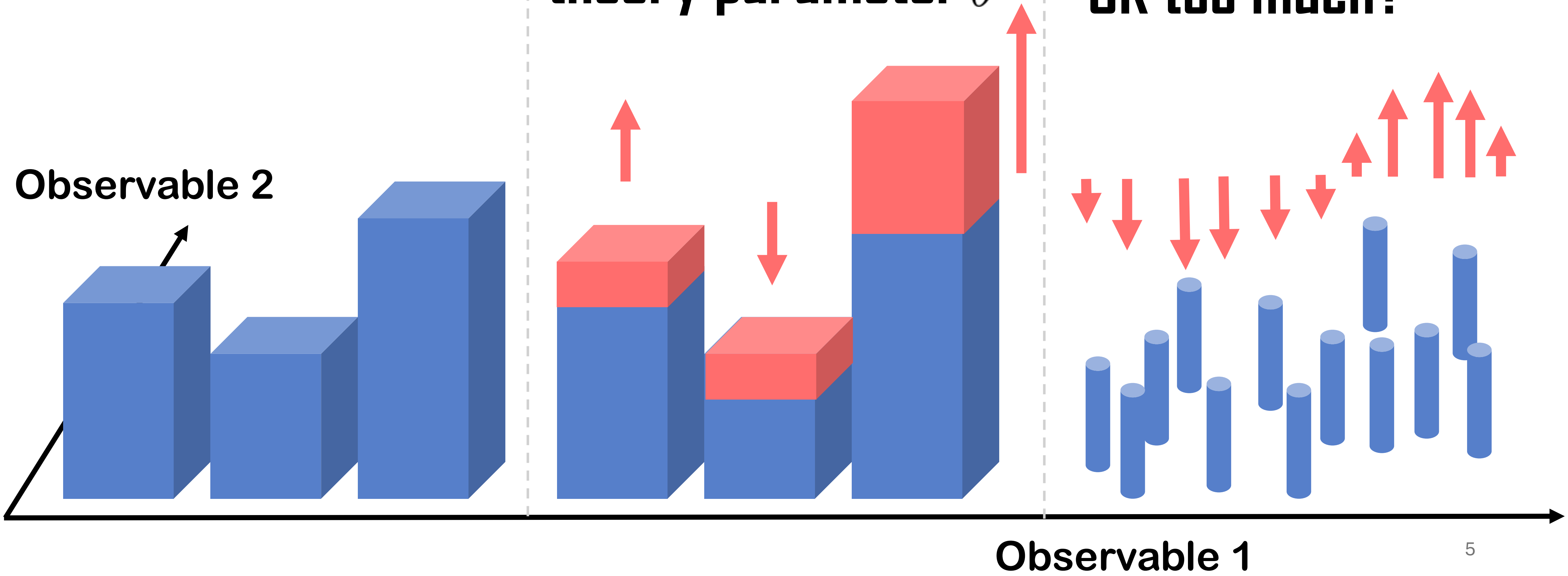
W/ Shengdu Chai and Jiayin Gu

SBI as Differential Measurement

Cut SR into several bins

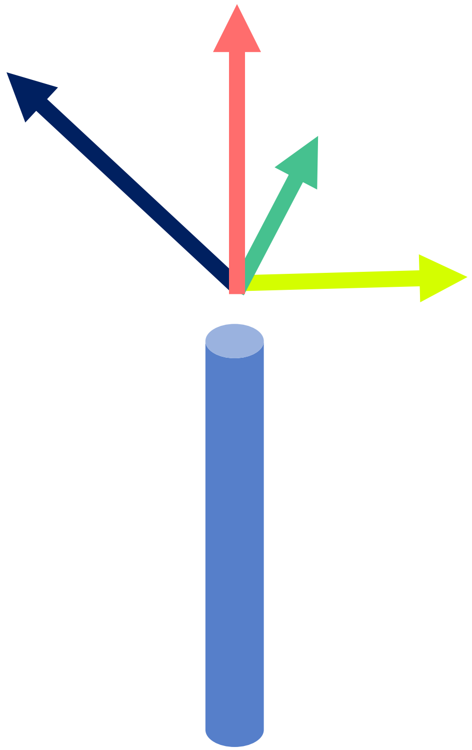
Counts change with theory parameter θ

What if you dice your SR too much?



From Bins to Optimal Observables

Promote into event-by-event vectors



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

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) 397-412

Expectation & limits on θ are extracted from the likelihood

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

*Have a spirit similar to unfolding

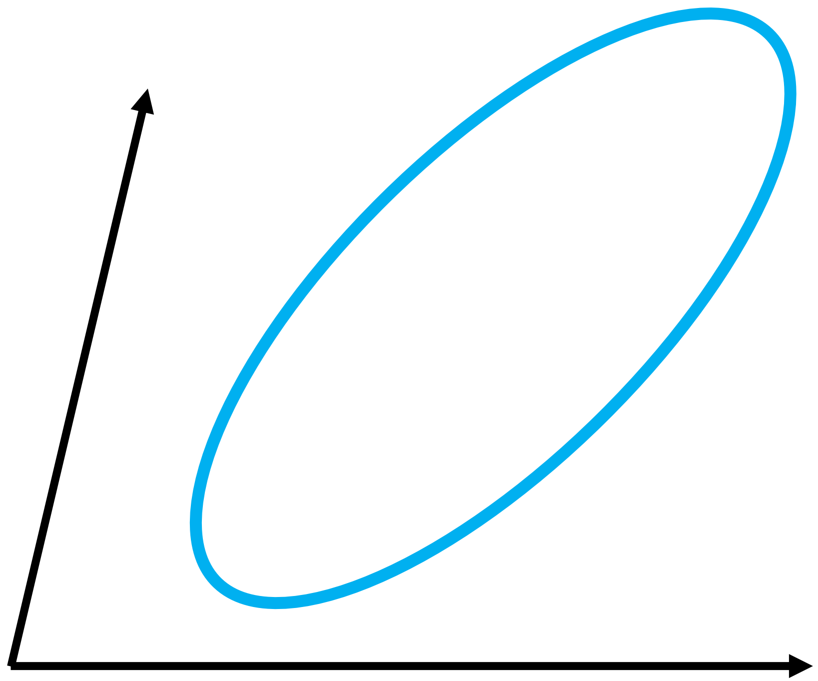
A. Andreassen, P.T. Komiske, E.M. Metodiev, B. Nachman, J. Thaler, 1911.09107

Evaluation of the Result

The constraining power can be evaluated by the "volume" of the remaining parameter space

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

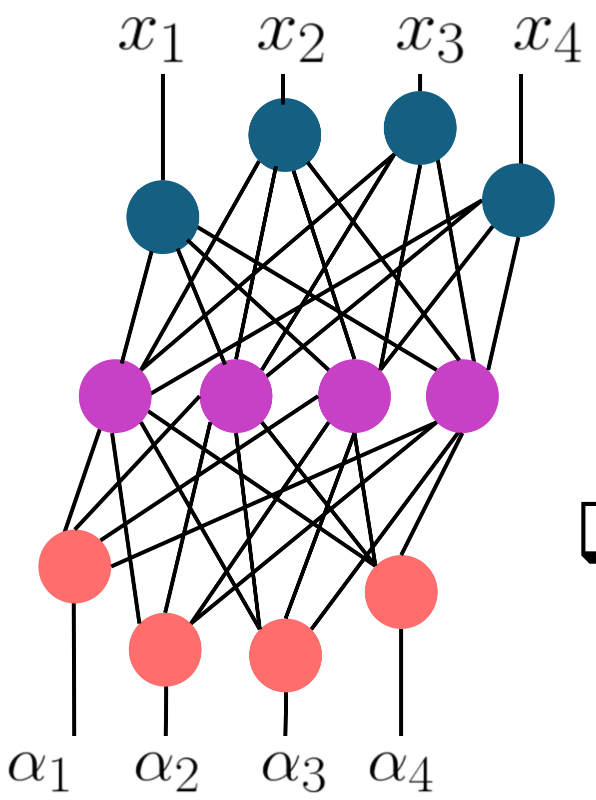
~Fisher Information Matrix



Saturating the Cramér–Rao bound with optimal observable

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:

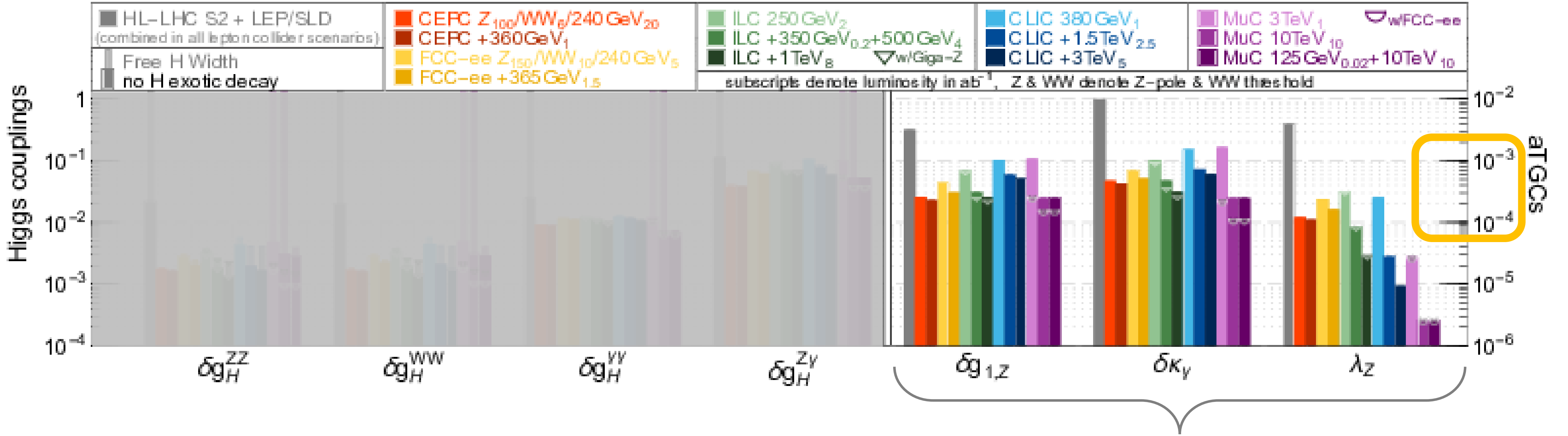
$$L = \sum_i (\alpha_i - \bar{\alpha}_i)^T G (\alpha_i - \bar{\alpha}_i)$$

*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……

de Blas, Du, Grojean, Gu, Miralles, Peskin, Tian, Vos, Vryonidou, 2206.08326

“Promised” precision of aTGC



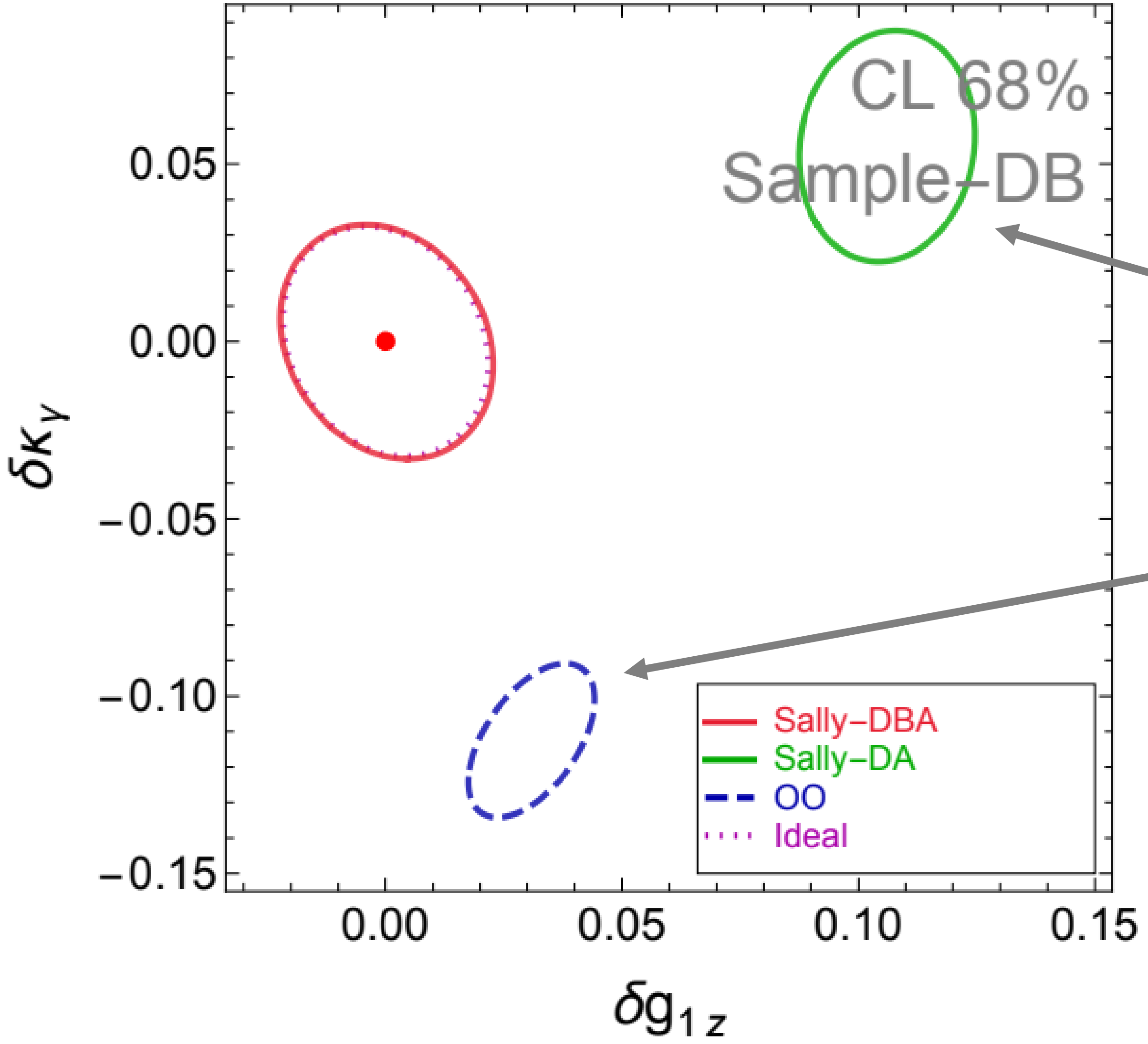
Anomalous Triple Gauge Couplings

See also:

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 ……

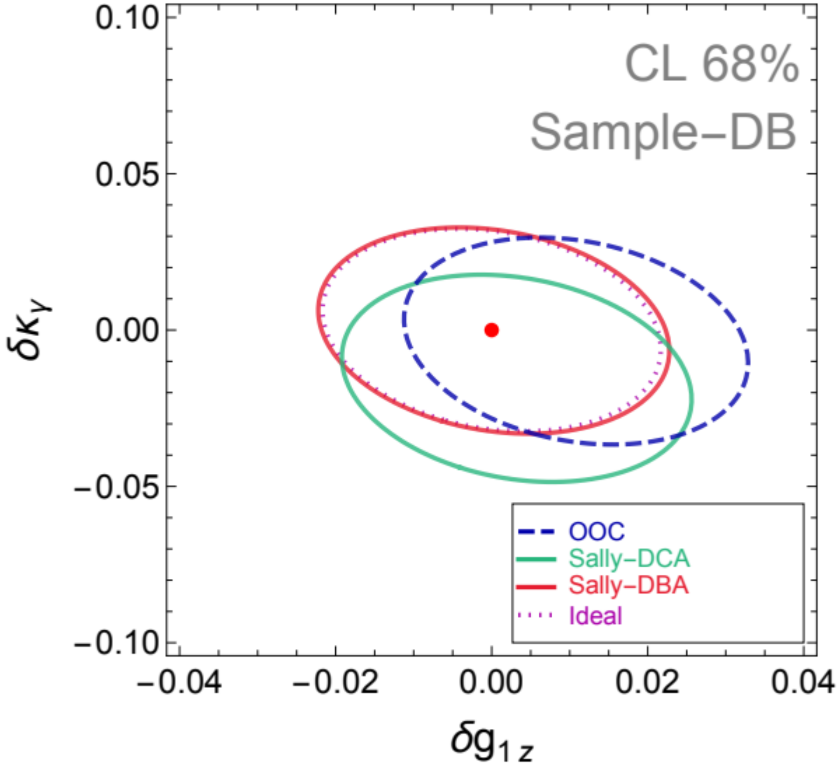
From differential measurements of 10^8 diboson events @240 GeV

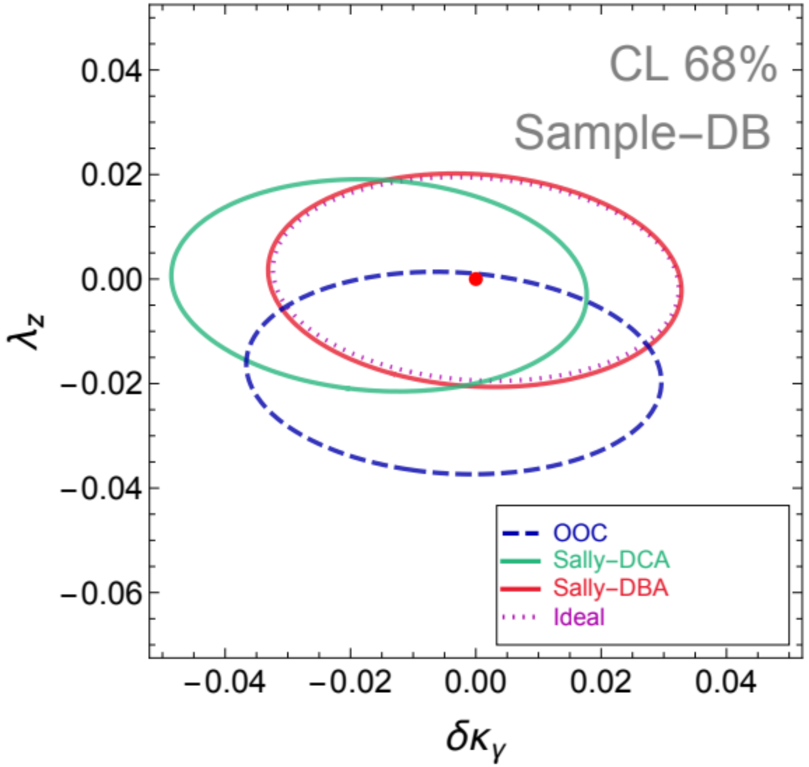
Injecting Background Noise

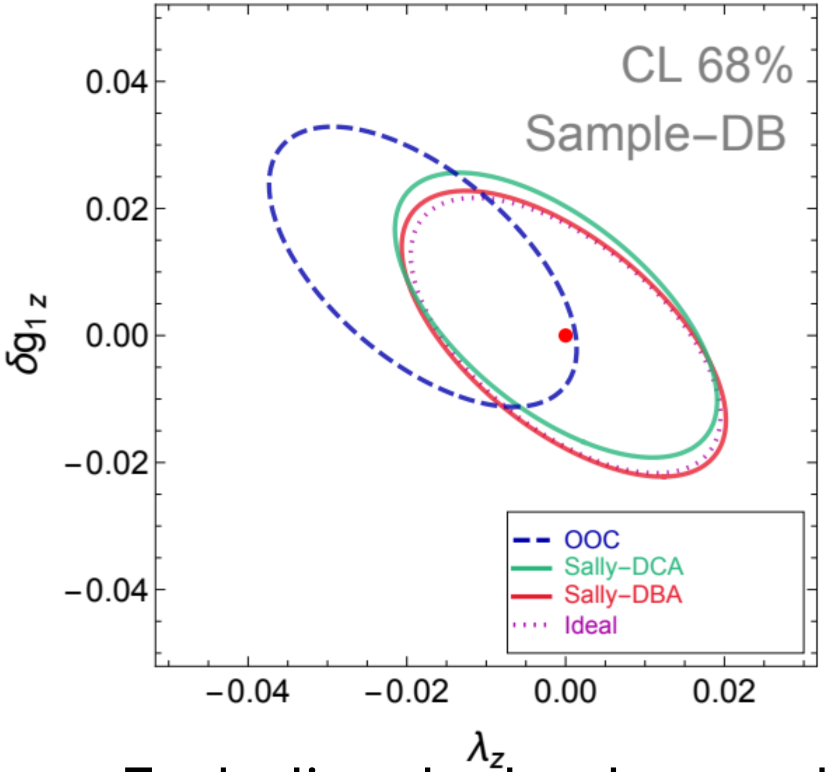


Fail to catch up with backgrounds

Two ways of Mitigation







Training with proper sample and loss functions

One single network system

Equivalence*

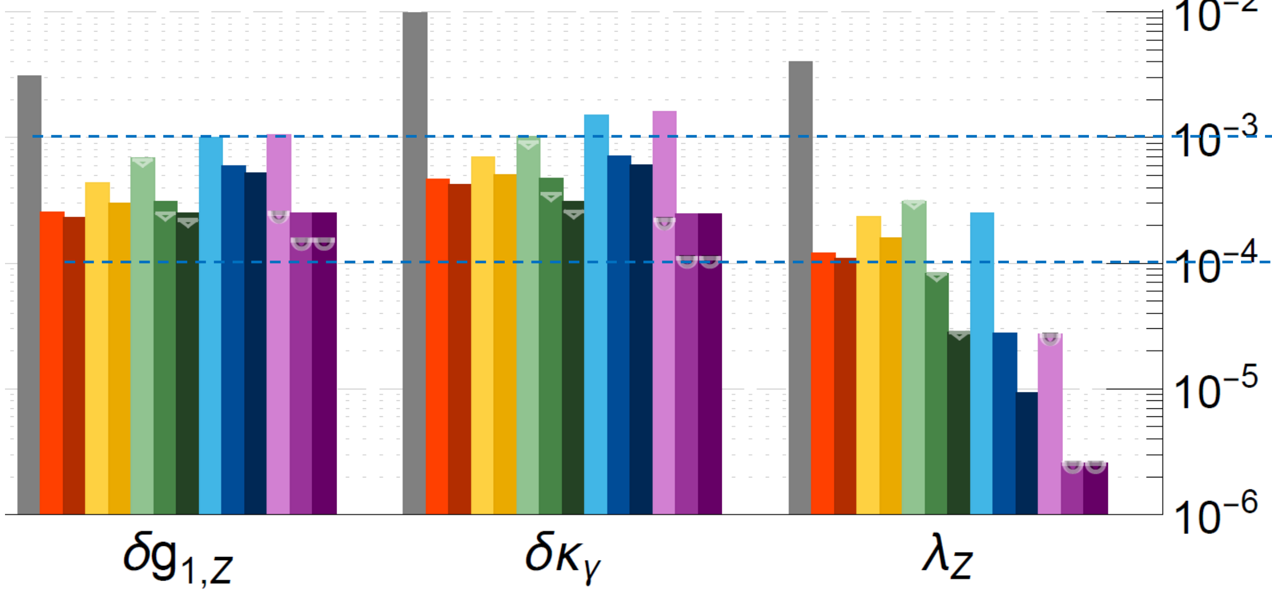
Excluding the background by a classifier

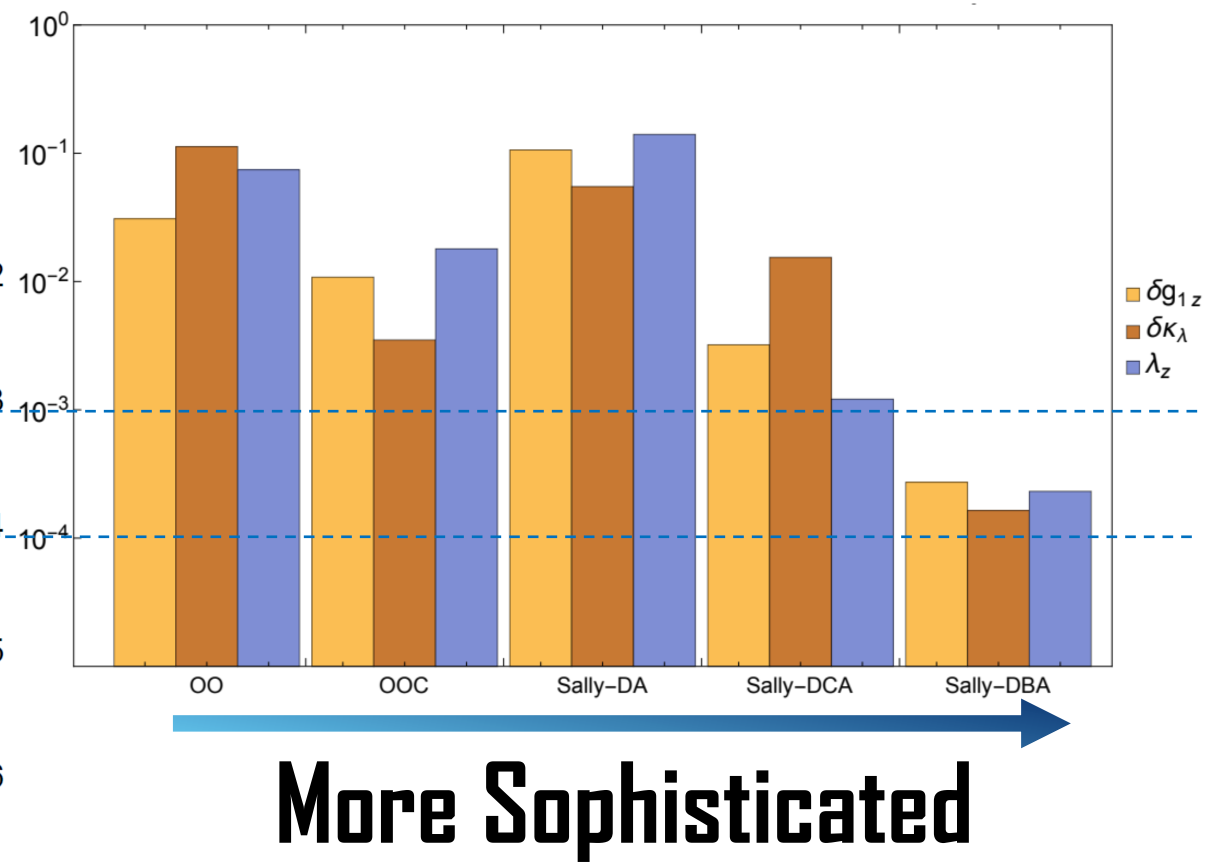
Work separated in two networks

*with conditions

Getting closer to the commitment

Training sample size: $2 \cdot 10^6$, much smaller than 10^8 expected at Higgs factories



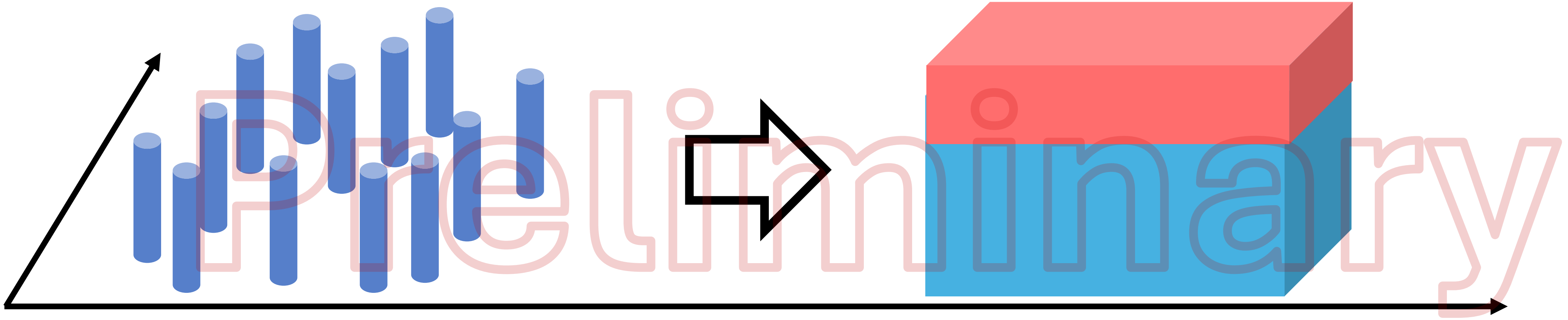


Theoretical Update: Unified Language in the Linear Regime

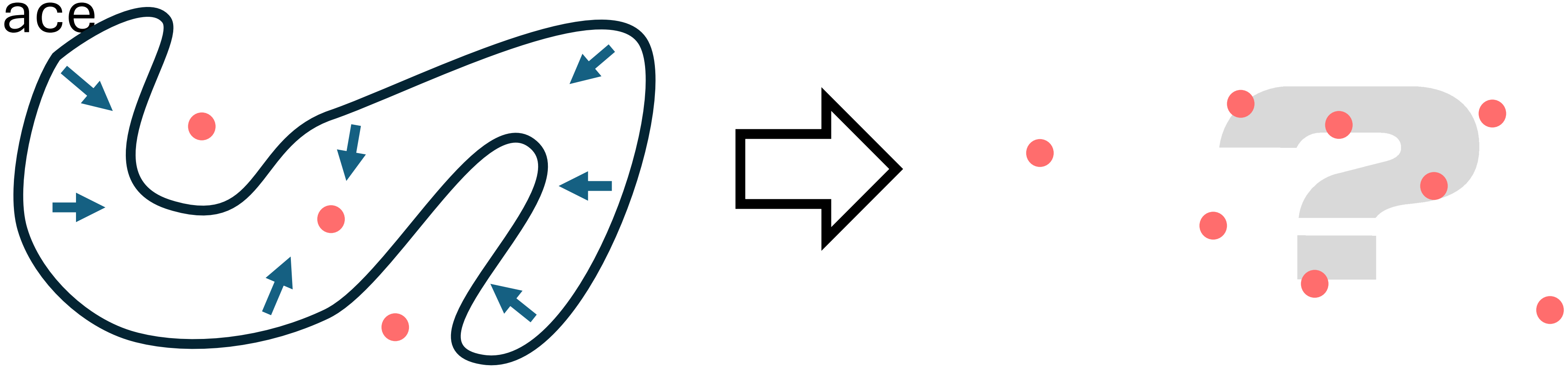
arXiv:25ab.uvxyz

Jingxuan Bu, Yifan Fei, Jiayin Gu, LFL

Binned statistics is just averaging some regions of the dual space



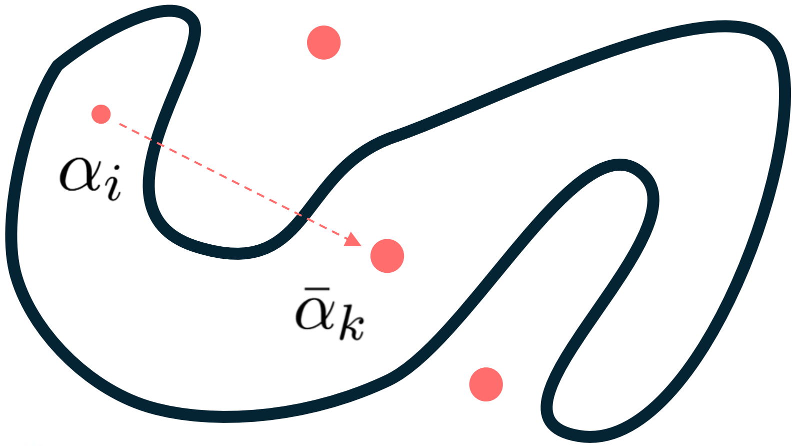
...but according to some criteria in the observable space, not the dual space



Averaging a Single Event into a Bin

Using matrix determinant lemma:

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

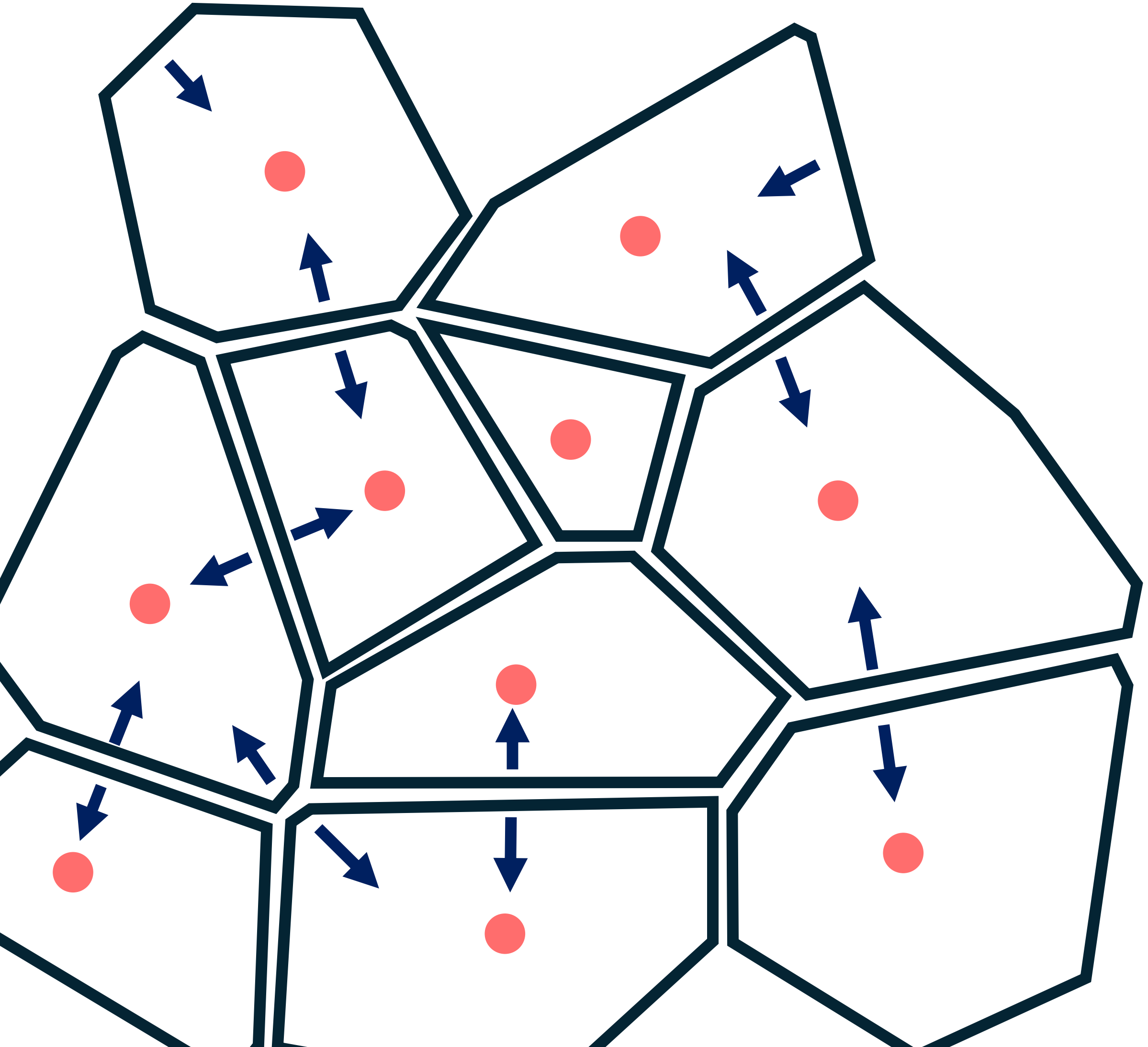


$$F \Rightarrow F'$$

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

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

Optimal Transportation of the Optimal Observable



- ❑ The ideal binning would be an approximate Voronoi tessellation in the theory space
- ❑ Reduction of information loss dual to a optimal transportation problem

Form like "Jet Clustering"

More Open Problems (in Progress)

- Generalize the above to arbitrary estimator of any dimension
- The scaling law from the complexity of OO
- The non-Gaussian feature of the data and its implication
- The "quantumness" of such ML-measurement complex