GIVING ML A BOOST TOWARDS RESPECTING (APPROXIMATE) SYMMETRIES

11th LITP Symposium on Theoretical Physics and AI

INBAR SAVORAY

UC Berkeley & Lawrence Berkeley National Lab

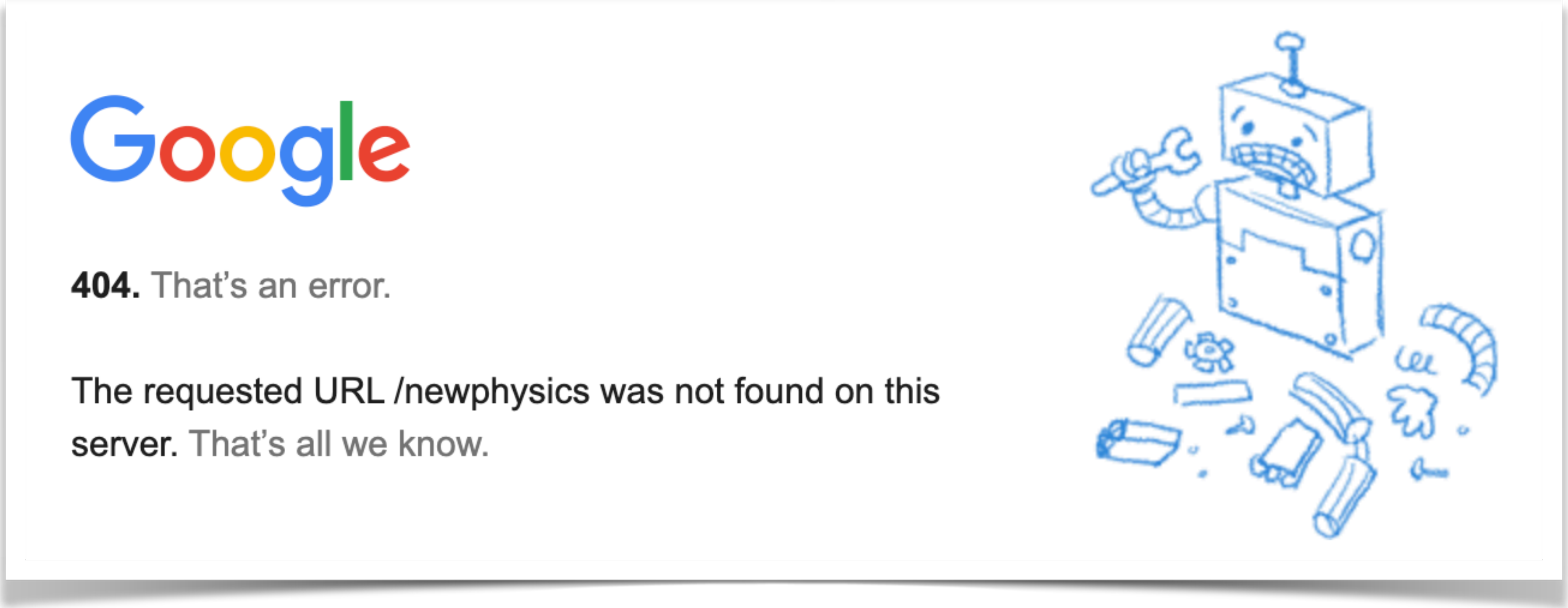
WITH: PRADYUN HEBBAR, THANDIKIRE MADULA, VINICIUS MIKUNI, BENJAMIN NACHMAN & NADAV OUTMEZGUINE

2511.01982

INTRODUCTION

PARTICLE PHYSICS 404

- Despite great theoretical and experimental effort, no evidence of **New Physics** has been found to date.
- Many dedicated searches ruled out a significant portion of the parameter space of theoretically motivated models.
- However, there is still much more to explore:
  - New theoretical models.
  - A lot of data.



404. That's an error.

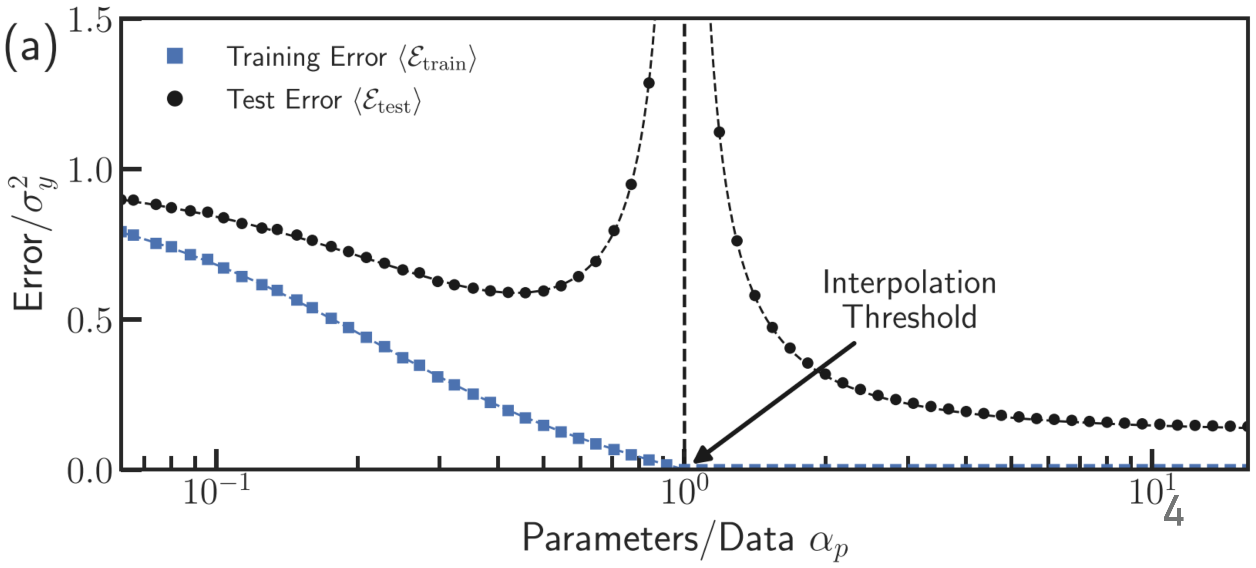
The requested URL /newphysics was not found on this server. That's all we know.

MORE DATA/PARAMETERS VS. MORE STRUCTURE

- **Modern Machine Learning is "more is more" -**
  - More data, more parameters, more compute.
  - Better capabilities, but also better generalization.
- **Modern ML is less specialized -**
  - Transformers perform well on a wide range of tasks.
  - Shift from carefully designing models for specific tasks to fine tuning foundational models.

Scalable

Flexible



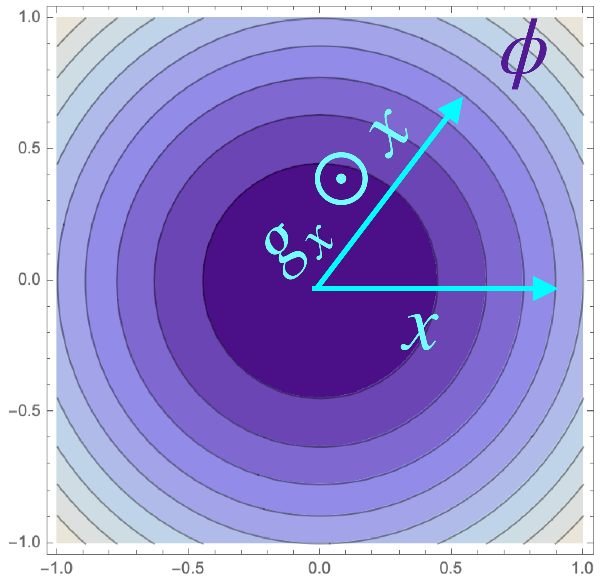
MORE DATA/PARAMETERS VS. MORE STRUCTURE

- On the other hand, more information is also more - especially for scientific applications
- **Data**
  - Noisy data can "trick" over-parameterized models
  - Might require more precision than language or images
- **Theory**
  - Often the underlying truth is "simple" - Ockham's razor
  - We have **guiding theoretical principles** that can be easily phrased as clear mathematical/logical statements

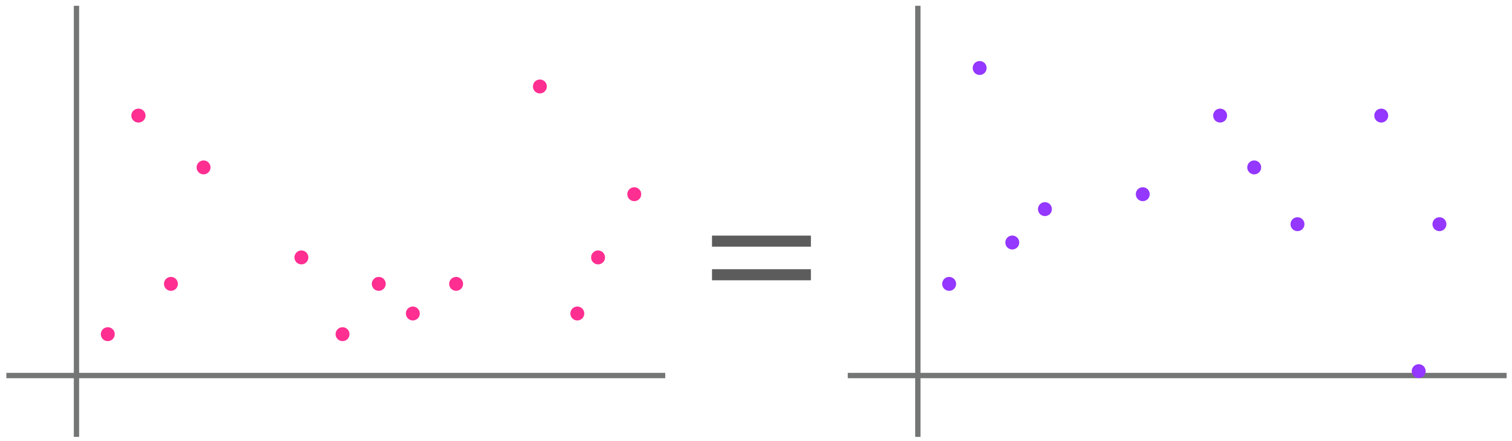
Physical~structure

SYMMETRIES

- **Symmetries as theoretical input** - physical information about the system we are trying to describe



Constraints



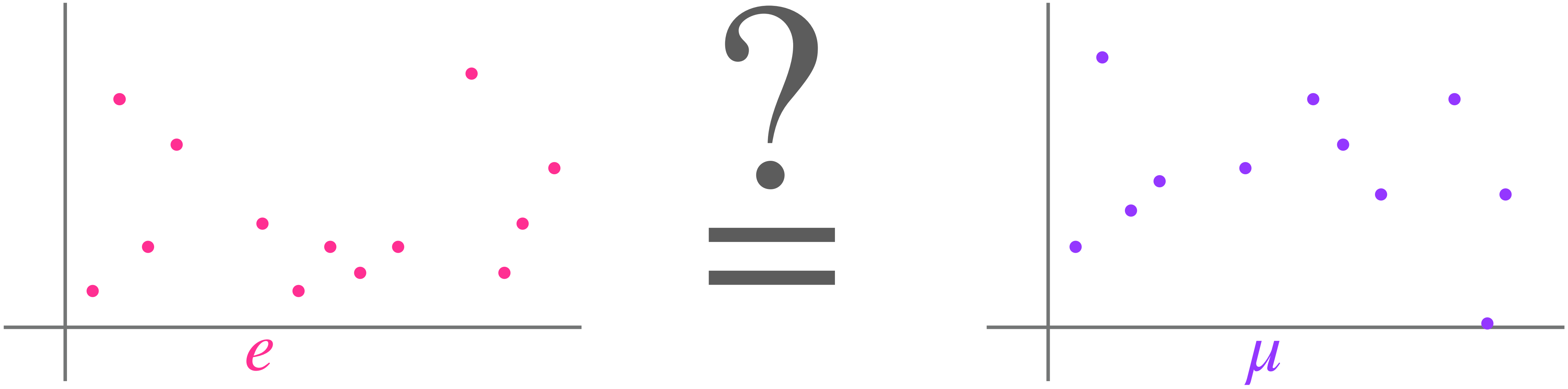
Data relations

- Ubiquitous in particle physics - flavor, P/CP, rotations, translations…
- Often *approximate* - either theoretically or practically

SYMMETRIES AS RELATIONS BETWEEN DATA

- **Symmetries** imply **relations between different regions of the data**, can be utilized for **model-agnostic searches for NP**.

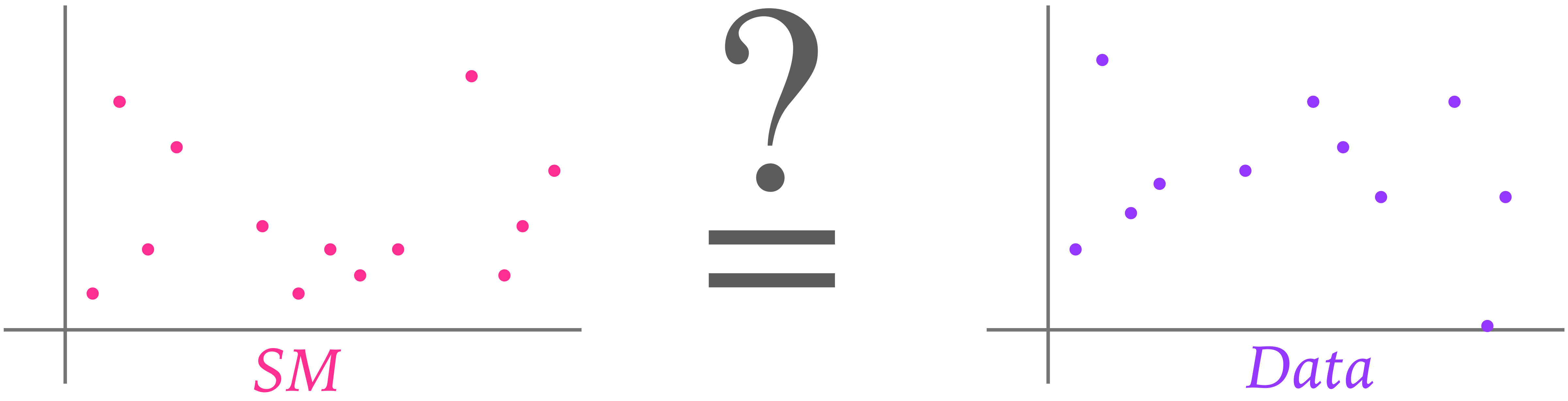
- Lepton flavor universality: $e/\mu/\tau$ **should be interchangeable*** (*up to H+phase space).



S. Bressler, IS, Y. Zurgil, 2401.09530

SYMMETRIES AS RELATIONS BETWEEN DATA

- Symmetries imply **relations between different regions of the data**, can be utilized for **model-agnostic anomaly detection**.
- Any two sample comparison: **SM simulation vs. data, control vs. signal, etc.**

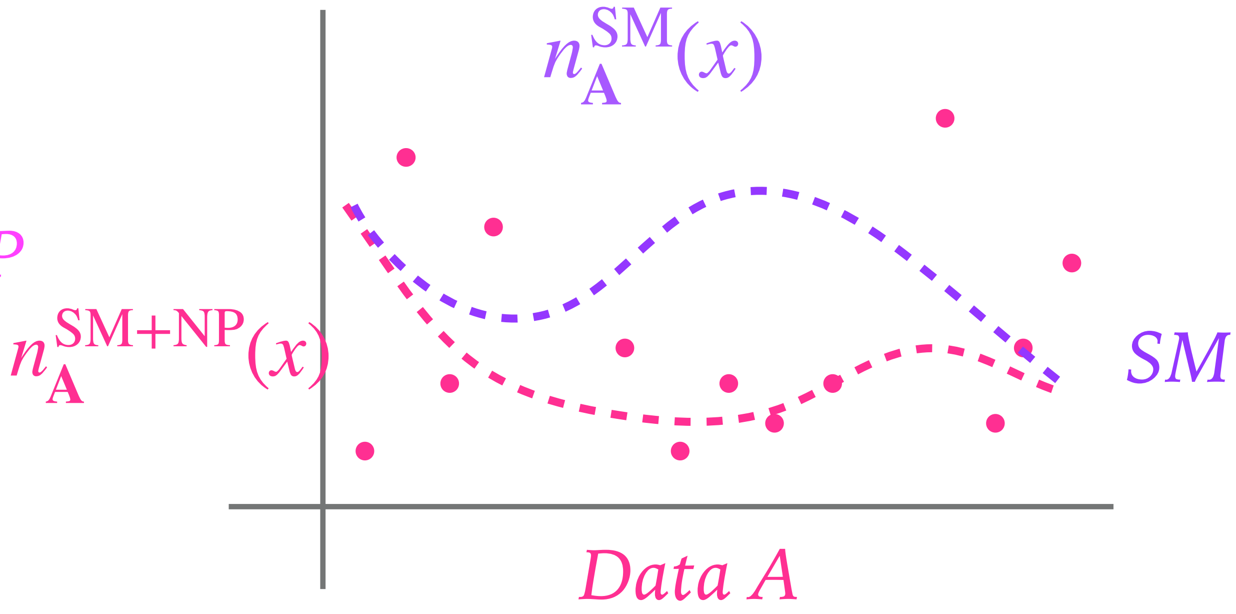


S. Bressler, IS, Y. Zurgil, 2401.09530

NEW PHYSICS LEARNING MACHINE

- Determine if sample **A** is drawn from SM or SM+NP distribution.
- Unbinned profile likelihood test

$$t = 2\log\left(\frac{max\left(\mathscr{L}\left(\mathscr{H}_1|\mathbf{A}\right)\right)}{max\left(\mathscr{L}\left(\mathscr{H}_0|\mathbf{A}\right)\right)}\right), \quad \begin{matrix}\mathit{SM+NP}\\ \\ \mathit{SM}\end{matrix}$$



R. T. D'Agnolo & A. Wulzer, [1806.02350].

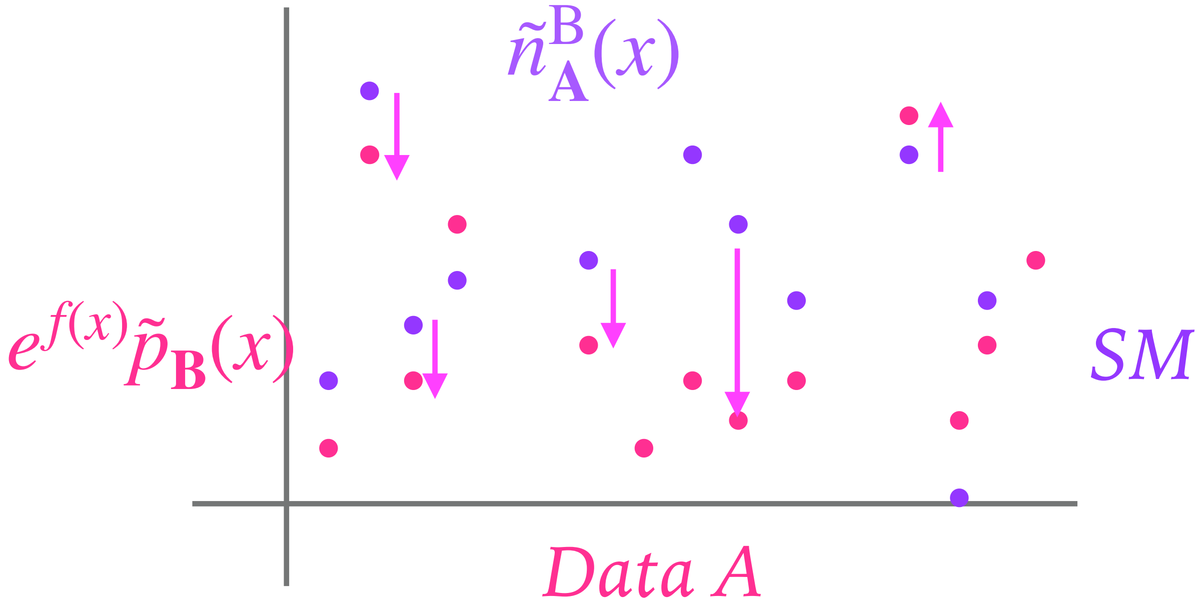
NEW PHYSICS LEARNING MACHINE

- Determine if sample **A** is drawn from SM or SM+NP distribution.
- Unbinned profile likelihood test

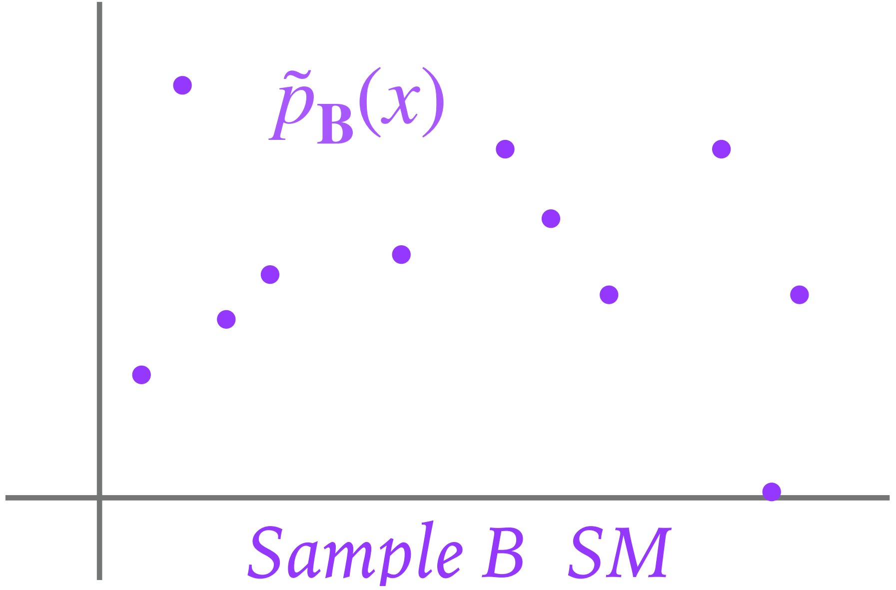
$$t = 2\log\left(\frac{max_{p_A}\left(\mathscr{L}\left(p_A\,|\,\mathbf{A}\right)\right)}{\mathscr{L}\left(\tilde{p}_B\,|\,\mathbf{A}\right)}\right),$$

SM+NP

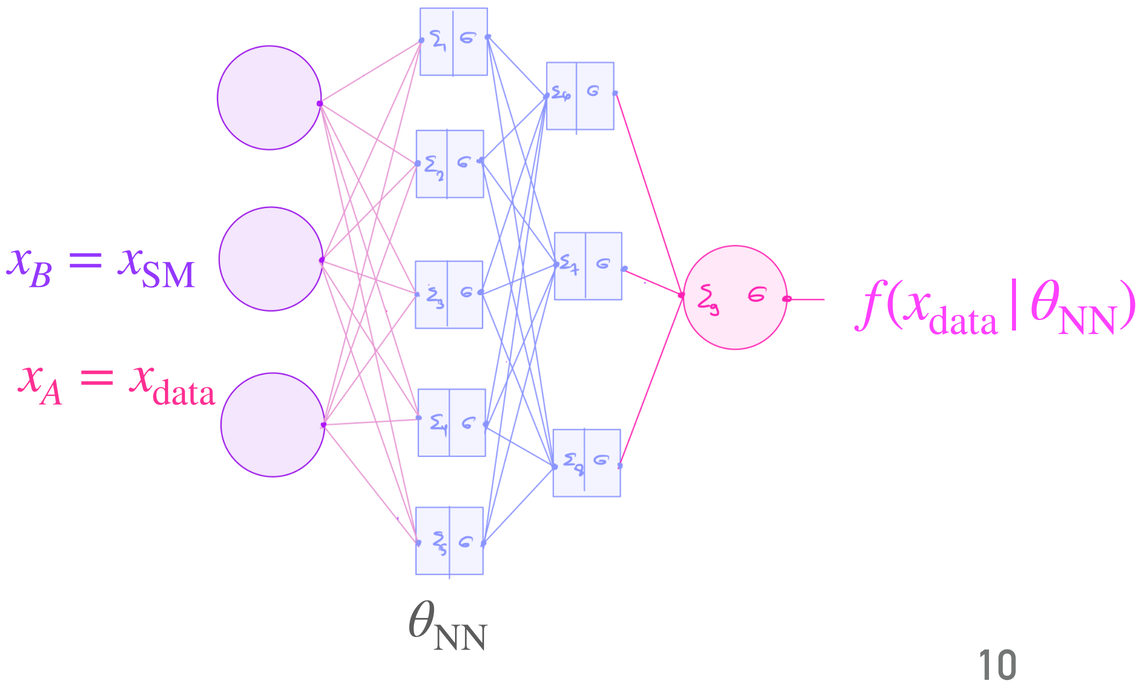
SM



SM: large sample B



NP: output of a NN maximizing t



R. T. D'Agnolo & A. Wulzer, [1806.02350].

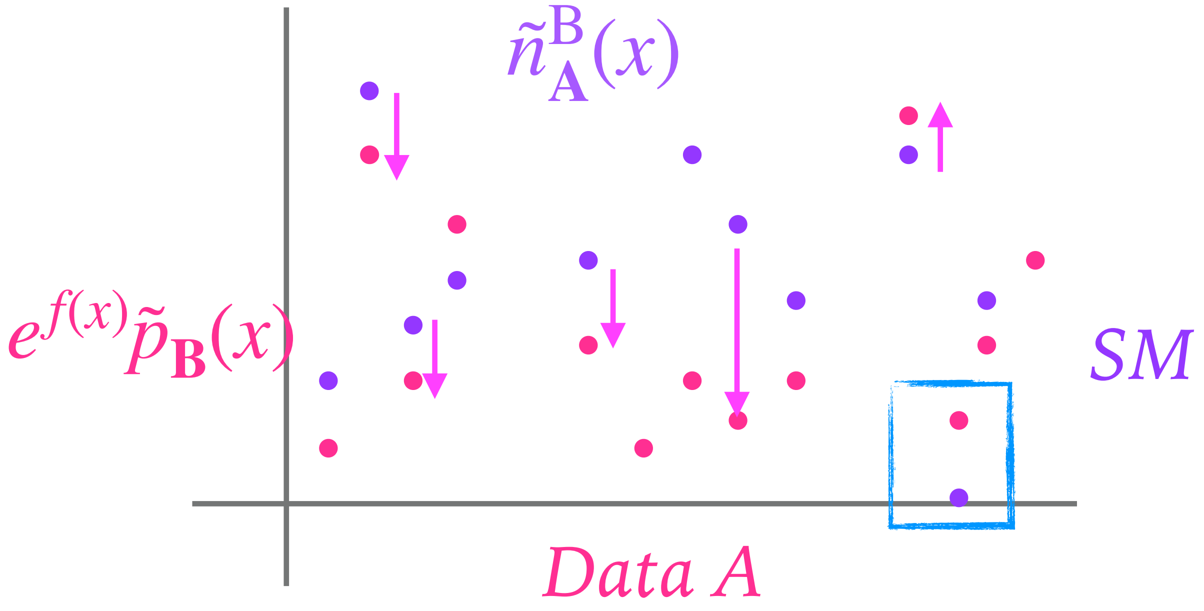
NEW PHYSICS LEARNING MACHINE

- Determine if sample **A** is drawn from SM or SM+NP distribution.
- Unbinned profile likelihood test

$$t = 2\log\left(\frac{max_{p_A}\left(\mathscr{L}\left(p_A\,|\,\mathbf{A}\right)\right)}{\mathscr{L}\left(\tilde{p}_B\,|\,\mathbf{A}\right) = 0}\right),$$

SM+NP

SM

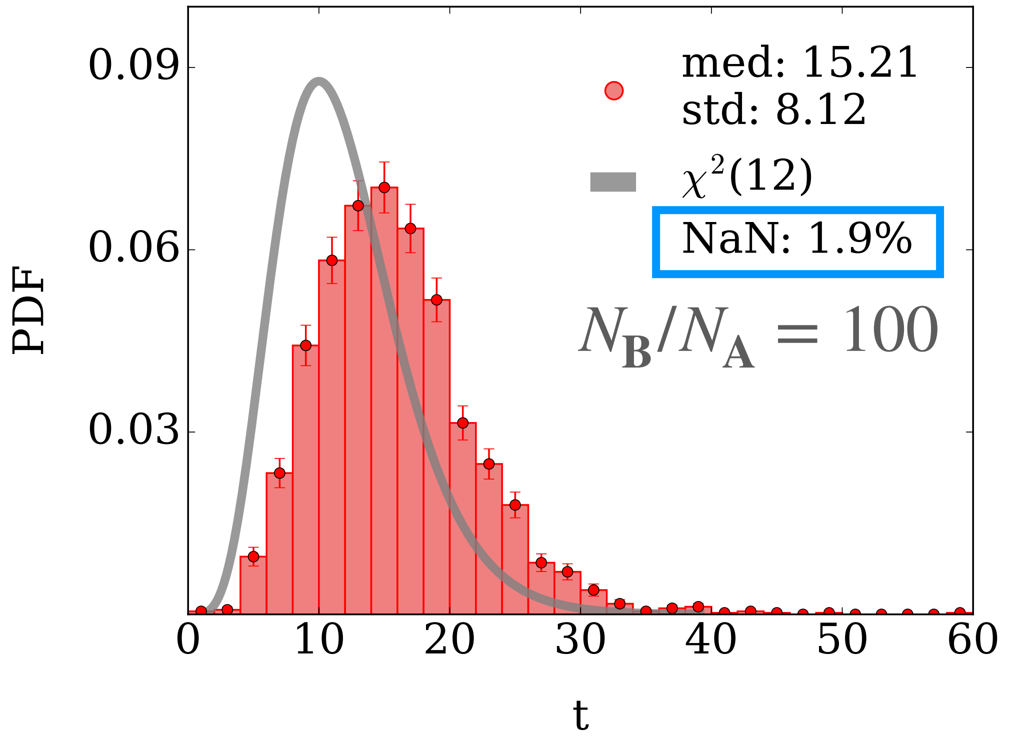


NP: output of a NN maximizing t

Unbounded

Statistical fluctuations easily result in false $\mathscr{H}_0$!

Regularize?



R. T. D'Agnolo & A. Wulzer, [1806.02350].

THE SYMMETRIZED FORMALISM

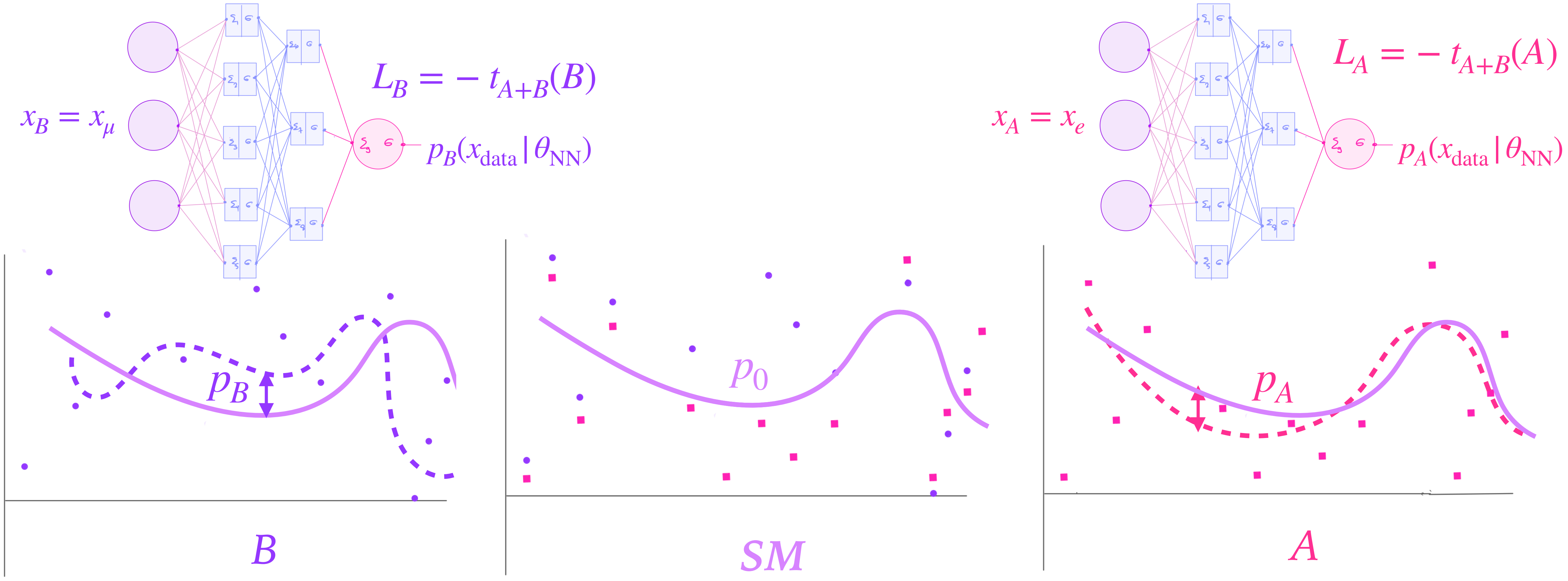
S. Bressler, IS, Y. Zurgil, 2401.09530

- ➤ Determine if samples **A** and **B** are drawn from the same distribution.
- ➤ Symmetric test - learn common PDF from both samples, test on both

$p_e \neq p_\mu$ SM+NP

$$\boxed{t = 2\log\left(\frac{\max_{p_\mathbf{A},p_\mathbf{B}}\left(\mathscr{L}\left(N_\mathbf{A}, p_\mathbf{A}(x)\,|\,\mathbf{A}\right)\mathscr{L}\left(N_\mathbf{B}, p_\mathbf{B}(x)\,|\,\mathbf{B}\right)\right)}{\max_{p_0}\left(\mathscr{L}\left(N_\mathbf{A}, p_0(x)\,|\,\mathbf{A}\right)\mathscr{L}\left(N_\mathbf{B}, p_0(x)\,|\,\mathbf{B}\right)\right)}\right)}$$

$p_e = p_\mu = p_0$ SM



THE SYMMETRIZED FORMALISM

S. Bressler, IS, Y. Zurgil, 2401.09530

- Determine if samples **A** and **B** are drawn from the same distribution.
- Symmetric test - learn common PDF from both samples, test on both

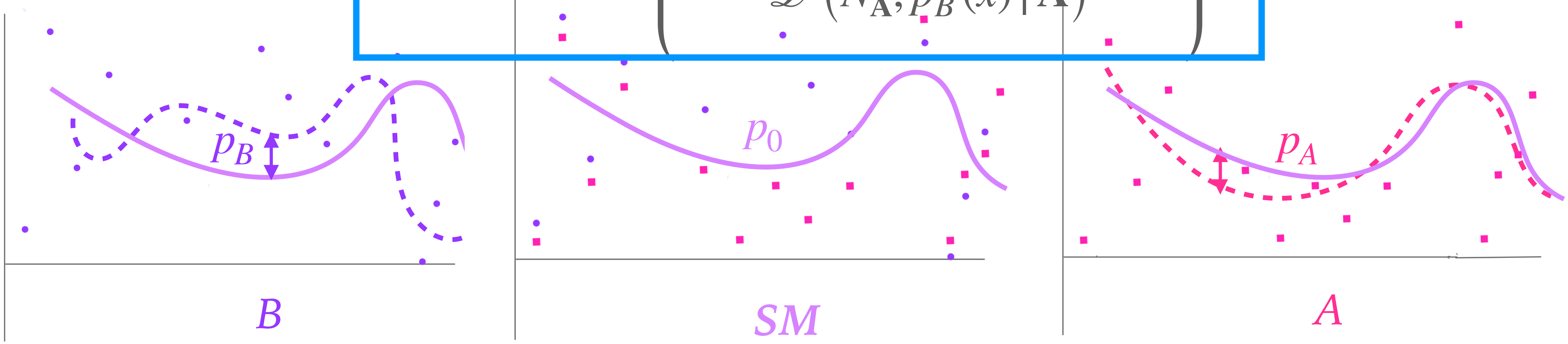
$p_e \neq p_\mu$ SM+NP

$$\boxed{t = 2\log\left(\frac{\max_{p_\mathbf{A},p_\mathbf{B}}\left(\mathscr{L}\left(N_\mathbf{A},p_\mathbf{A}(x)\,|\,\mathbf{A}\right)\mathscr{L}\left(N_\mathbf{B},p_\mathbf{B}(x)\,|\,\mathbf{B}\right)\right)}{\max_{p_0}\left(\mathscr{L}\left(N_\mathbf{A},p_0(x)\,|\,\mathbf{A}\right)\mathscr{L}\left(N_\mathbf{B},p_0(x)\,|\,\mathbf{B}\right)\right)}\right)}$$

$p_e = p_\mu = p_0$ SM

- **NPLM**: if $\tilde{N}_\mathbf{B} \gg \tilde{N}_\mathbf{A}$, learn common PDF from **B** - $\hat{p}_0 \approx \hat{p}_\mathbf{B}$, test on **A**

$$\boxed{t_{N_\mathbf{B} \gg N_\mathbf{A}} \to 2\log\left(\frac{\max_{p_\mathbf{A}}\left(\mathscr{L}\left(N_\mathbf{A},p_\mathbf{A}(x)\,|\,\mathbf{A}\right)\right)}{\mathscr{L}\left(N_\mathbf{A},\hat{p}_B(x)\,|\,\mathbf{A}\right)}\right)}$$



RESULTS - THE SYMMETRIZED FORMALISM

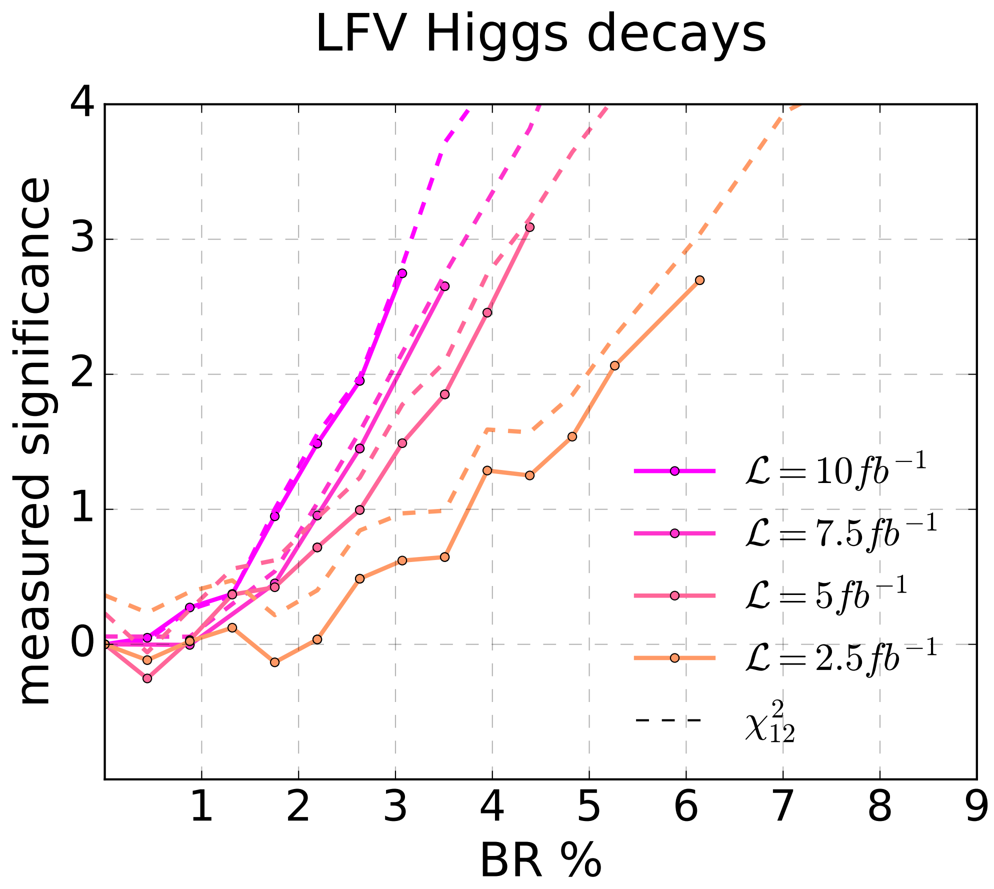
S. Bressler, IS, Y. Zurgil, 2401.09530

- Current bound: $BR\,(H \to \tau e) \sim 0.2\,\%$ at $L = 138\,\mathrm{fb}^{-1}$

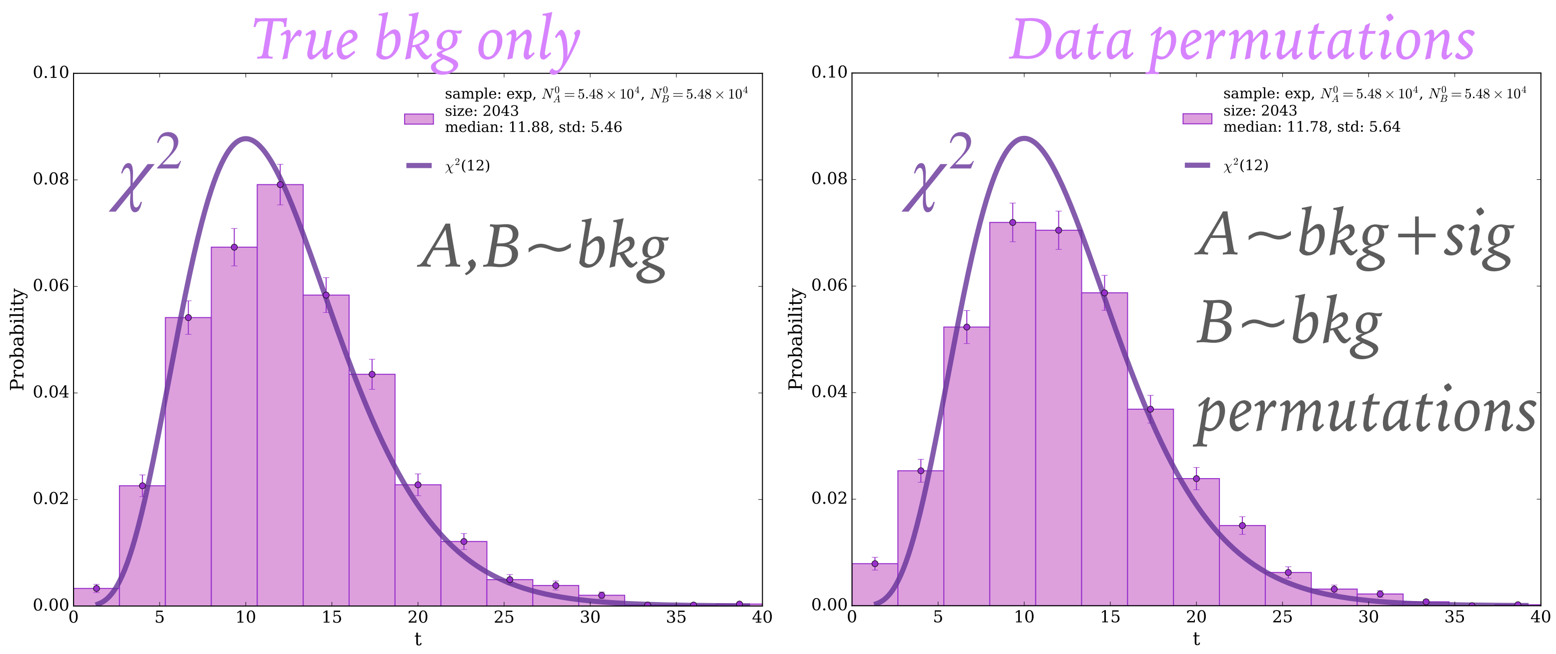
ATLAS, [2302.05225], CMS , [2105.03007]

- **Expected sensitivity:** $BR\,(H \to \tau e) \sim 0.7\,\%$ at $L = 138\,\mathrm{fb}^{-1}$

- **Robust significances via predictable background-only distributions** through asymptotic χ^2/permutations.



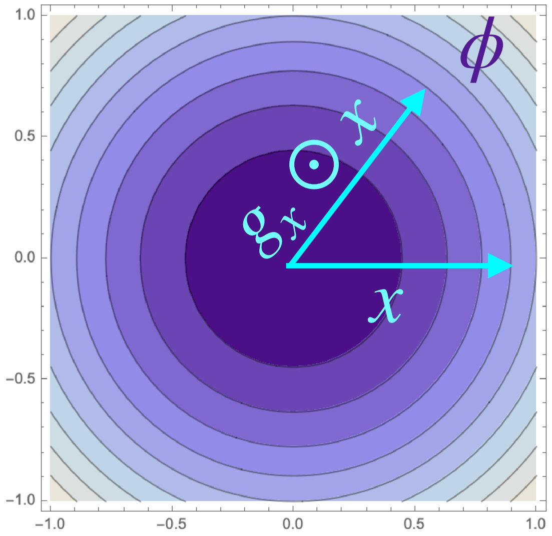
No regularization!



Symmetry assumptions only, no need for detailed SM simulation!

SYMMETRIES AS PHYSICAL CONSTRAINTS

- **Symmetries** imply a **constraint** on our target functions -



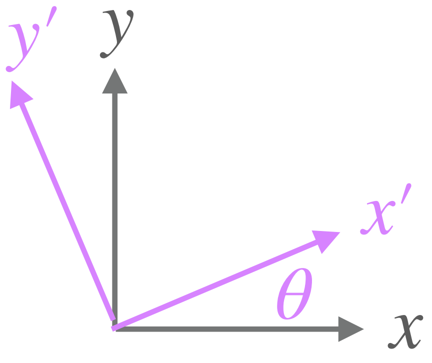
$$\phi\left(g_x \odot x\right) = g_\phi \odot \phi\left(x\right)$$

property of input data

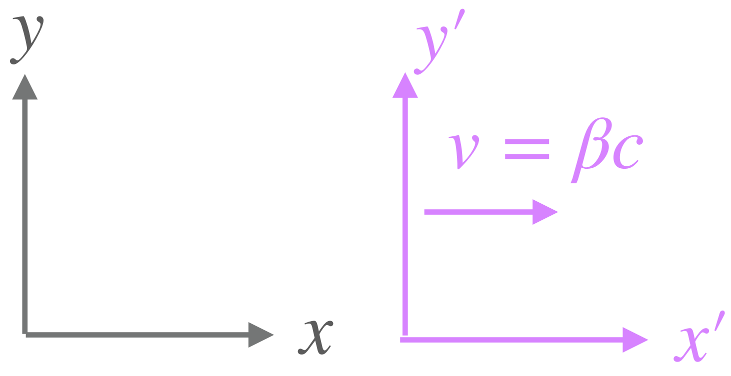
desired property of output

$$g \in G_{\text{symm}}$$

- Lorentz invariance - theoretically exact, space-time symmetry, continuous and non-compact.



$$g = \Lambda\left(\vec{\beta}, \vec{\theta}\right)$$



Scalar $(m^2, k \cdot p)$: $g_\phi = 1$

Covariant (p^μ): $g_\phi = g_x = g$

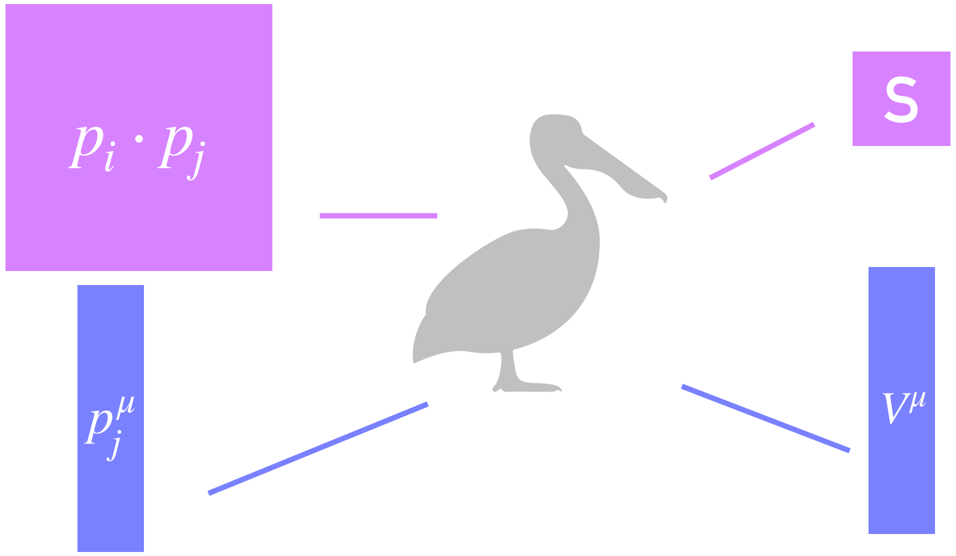
IMPOSING LORENTZ INVARIANCE THROUGH ARCHITECTURE

- Equivariant architecture - output functions transform in the correct way by construction.

$$\phi_{\mathrm{ML}}\left(g_x \odot x\right) = g_\phi \odot \phi_{\mathrm{ML}}\left(x\right)$$

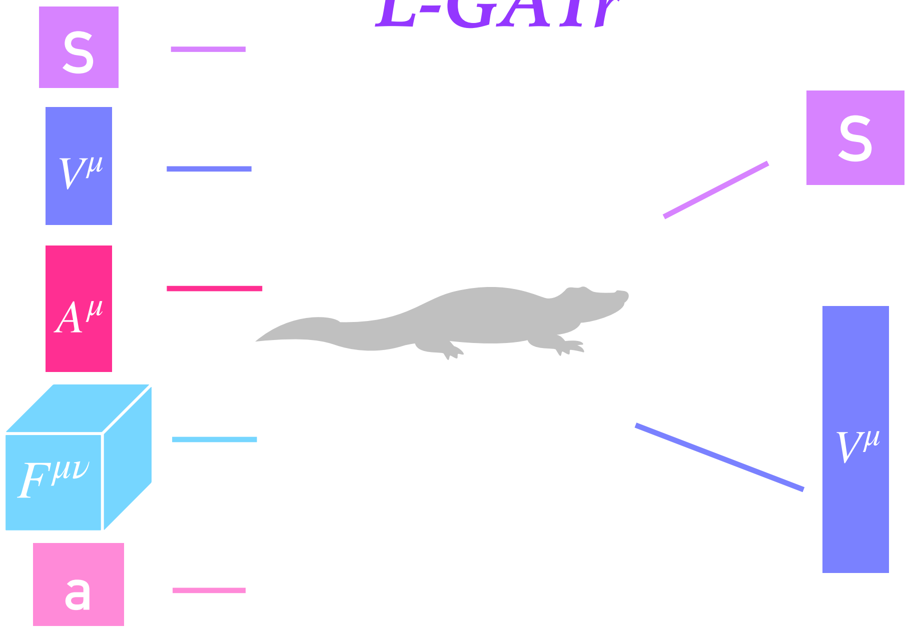
- Systematically build representations

PELICAN



A. Bogatskiy, T. Hoffman, D. W. Miller, J. T. Offermann, X. Liu, [2307.16506]

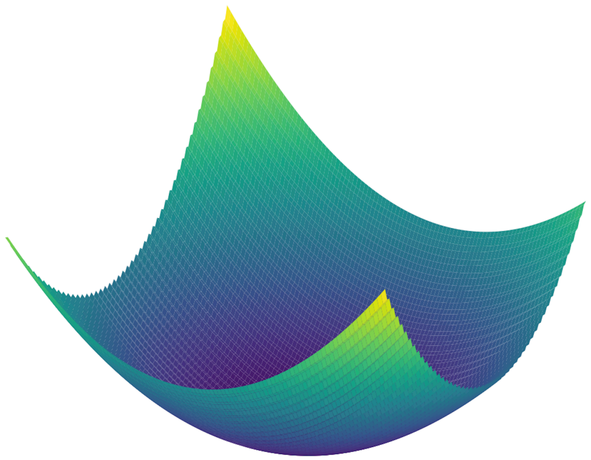
L-GATr



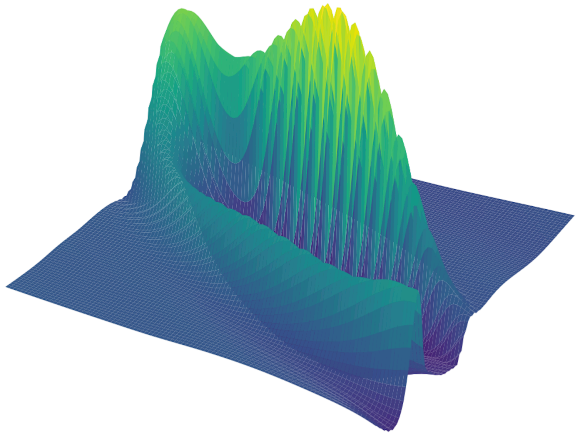
J. Spinner, V. Bresó, P. De Hann, T. Plehn, J. Thaler, J. Brehmer, [2405.14806]

CHALLENGES OF EQUIVARIANT MODELS

- Equivariant models have shown to improve performance on particle physics tasks.
- Expressivity could be challenging due to limited “building blocks”.
- Can be more compute intensive - overhead evaluation time and more FLOPs per parameter.
- Trainability - less smooth loss surface.

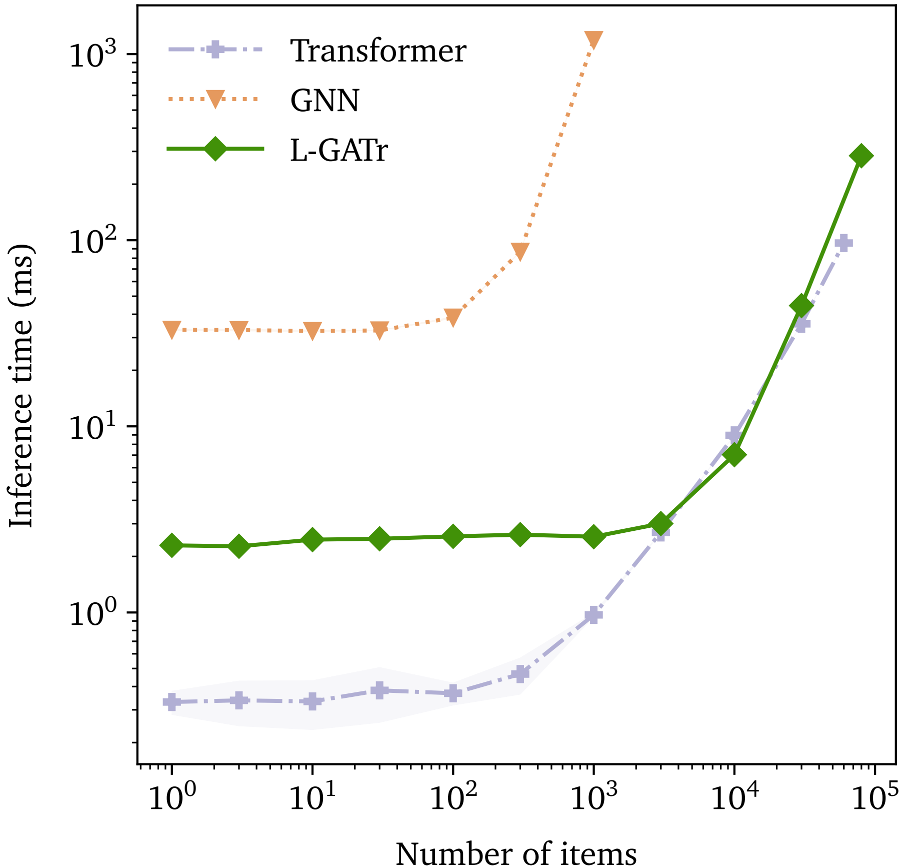


Transformer



GATr

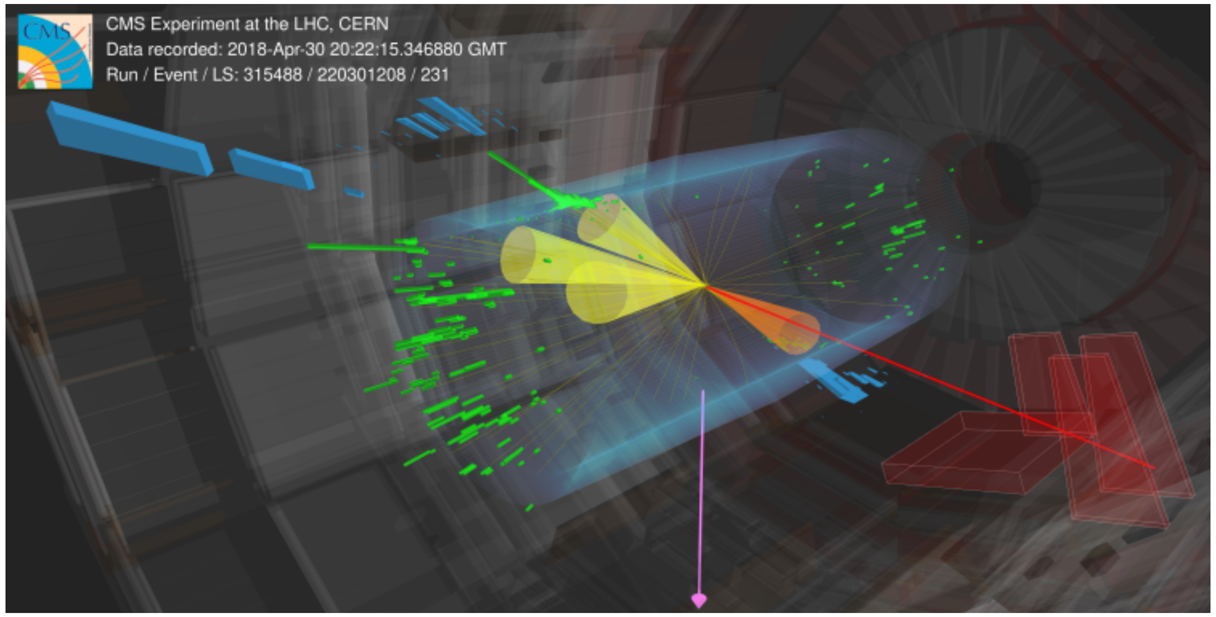
A. Elhag, T. Rusch, F. Di Giovanni and M. Bronstein, [2410.17878]



J. Spinner, V. Bresó, P. De Hann, T. Plehn, J. Thaler, J. Brehmer, [2405.14806]

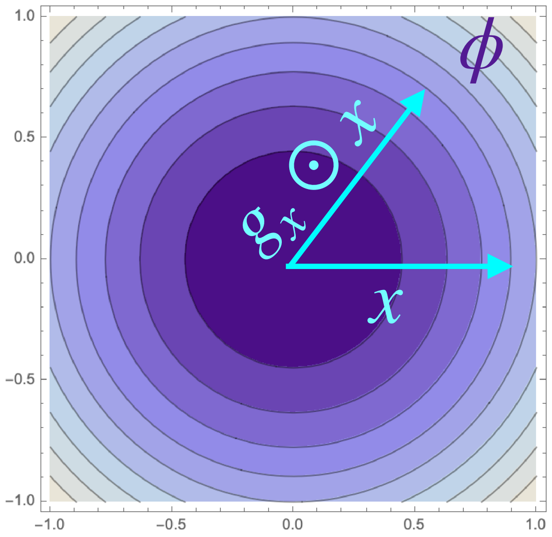
APPROXIMATE SYMMETRIES

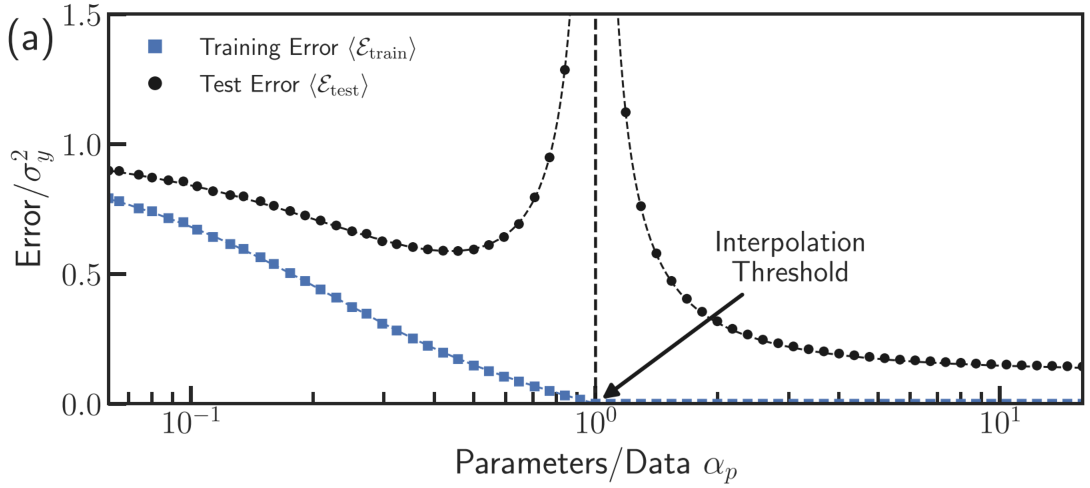
- Often physical symmetries are only approximate.
- Although Lorentz invariance is exact, it is effectively broken if one only transforms the final state momenta
  - Beam - introduces a preferred direction.
  - Detector - different energy efficiencies and spatial coverage/ sensitivity.
  - Clustering - algorithm takes into account euclidean distances.



SOFT SYMMETRIES?

- We want flexible and easy to train models, that are aware of symmetries but can choose how to use that information.
- Instead of imposing symmetries, specify a preference towards respecting them.





Physical

Scalable

Flexible

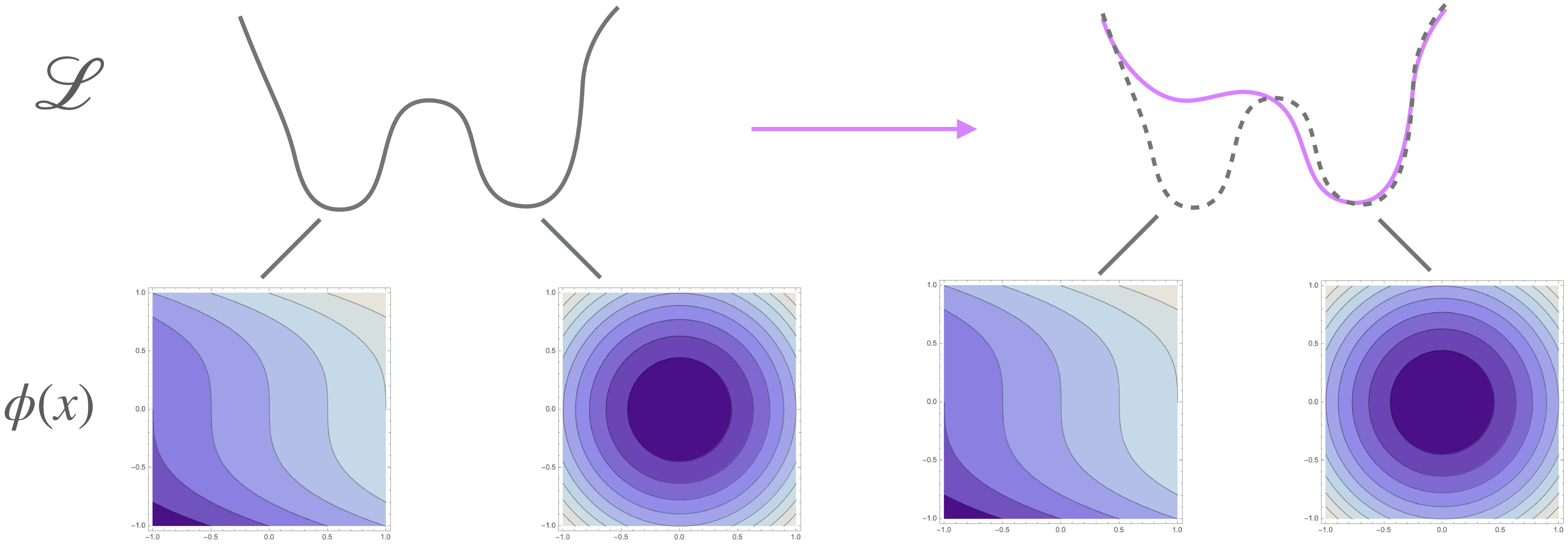
SYMMETRY ENCOURAGING LOSS

SYMMETRY ENCOURAGING LOSS

- A symmetry-encouraging term added to the loss

$$\mathscr{L} = \mathscr{L}_{\text{task}} + \lambda \mathscr{L}_{\text{SEAL}}$$

$$\mathscr{L}_{\text{SEAL}} = \| \phi_{\text{ML}} \left(g_x \odot x \right) - g_\phi \odot \phi_{\text{ML}} (x) \|^2$$

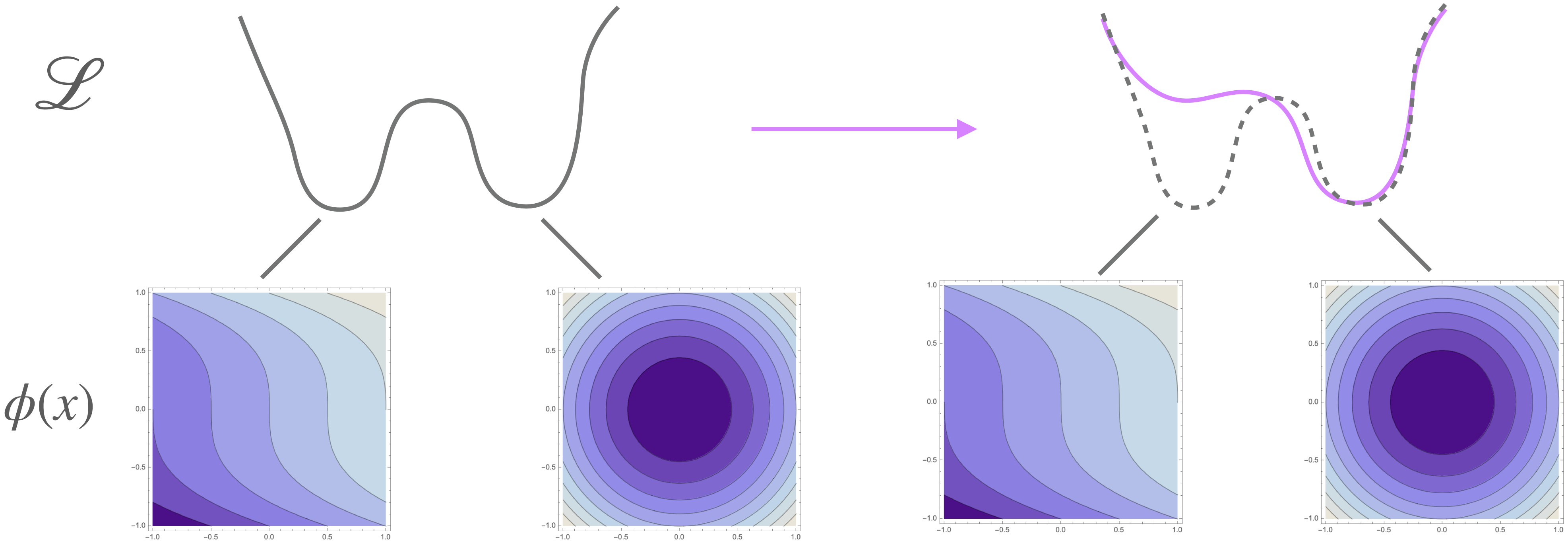


SEAL - SYMMETRY ENCOURAGING LOSS

- A symmetry-encouraging term added to the loss

$$\mathscr{L} = \mathscr{L}_{\text{task}} + \lambda \mathscr{L}_{\text{SEAL}}$$

$$\mathscr{L}_{\text{SEAL}} = \|\phi_{\text{ML}}\left(g_x \odot x\right) - g_\phi \odot \phi_{\text{ML}}(x)\|^2$$



SEAL - SYMMETRY ENCOURAGING LOSS

- A symmetry-encouraging term added to the loss

$$\mathscr{L} = \mathscr{L}_{\text{task}} + \lambda \mathscr{L}_{\text{SEAL}}$$

$$\mathscr{L}_{\text{SEAL}} = \|\phi_{\text{ML}}\left(g_x \odot x\right) - g_\phi \odot \phi_{\text{ML}}(x)\|^2$$

- **Relax hard constraints** -
  - Allow for approximate symmetries (and even no symmetries at all).
  - Bias is tunable and controllable.
- **Flexible** - can be added to any model.

Scalable

Physical

Flexible

SEAL - SYMMETRY ENCOURAGING LOSS

- A symmetry-encouraging term added to the loss

$$\boxed{\begin{aligned} \mathscr{L} &= \mathscr{L}_{\text{task}} + \lambda \mathscr{L}_{\text{SEAL}} \\ \mathscr{L}_{\text{SEAL}} &= \|\phi_{\text{ML}}\left(g_x \odot x\right) - g_\phi \odot \phi_{\text{ML}}(x)\|^2 \end{aligned}}$$

- $\mathscr{L}_{\text{SEAL}} \to 0$ if ϕ is in the desired representation for any group element g and any input x.
- In practice:
  - Average over the data.
  - Or choose any region!
    No need to know truth labels!

GSEAL - GROUP SAMPLE

- Compare output for original vs. boosted inputs

GSEAL:

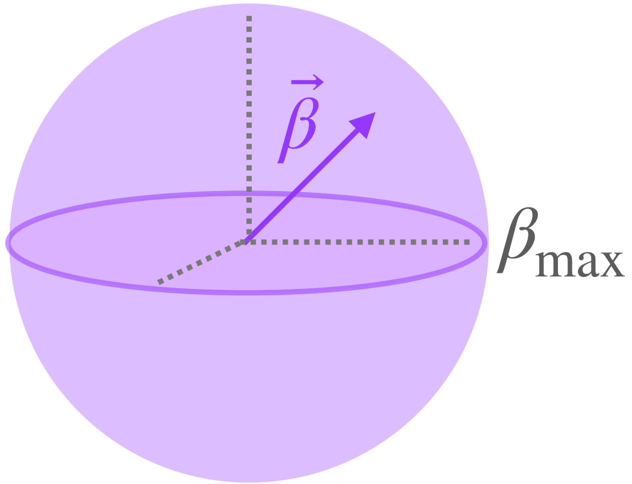
$$\mathscr{L}_G = \frac{1}{N}\sum_{i=1}^{N} \left\| \phi_{ML}\left(\Lambda\left(\vec{\beta}_i\right) \odot x_i\right) - g_i^{\phi} \odot \phi_{ML}\left(x_i\right) \right\|^2$$

Sample $\Lambda(\vec{\beta}_i)$

scalar: $g_i^{\phi} = 1$

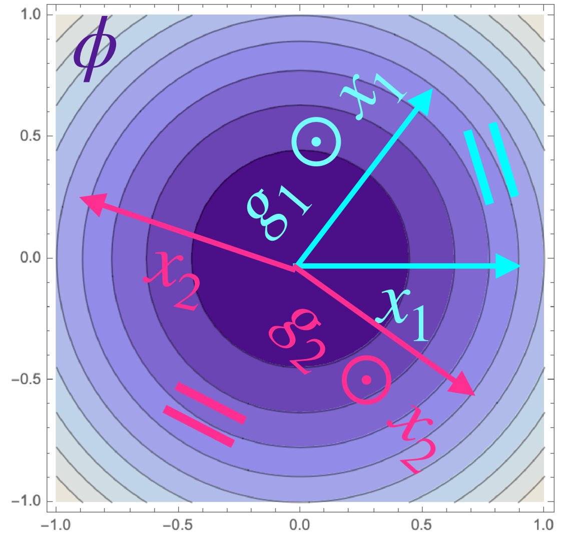
4-vector: $g_i^{\phi} = \Lambda\left(\vec{\beta}_i\right)$

- **Randomly sample group elements.**
- Cheap to calculate.
- **Lorentz**: boost vector $\vec{\beta}$ uniformly sampled from a sphere of radius $\beta_{\max}$.



GSEAL - GROUP SAMPLE

- Compare output for original vs. boosted inputs



GSEAL:

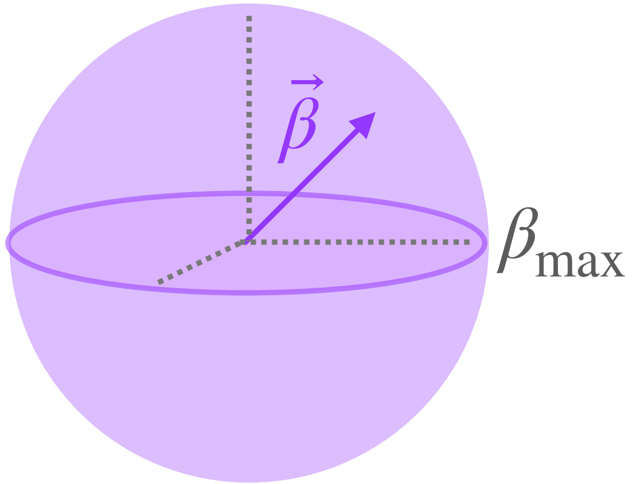
$$\mathscr{L}_G = \frac{1}{N}\sum_{i=1}^{N} \left\| \phi_{ML}\left(\Lambda\left(\vec{\beta}_i\right)\odot x_i\right) - g_i^{\phi}\odot \phi_{ML}\left(x_i\right) \right\|^2$$

Sample $\Lambda(\vec{\beta}_i)$

scalar: $g_i^{\phi} = 1$

4-vector: $g_i^{\phi} = \Lambda\left(\vec{\beta}_i\right)$

- **Randomly sample group elements.**
- Cheap to calculate.
- **Lorentz**: boost vector $\vec{\beta}$ uniformly sampled from a sphere of radius $\beta_{\max}$.



δSEAL – INFINITESIMAL SYMMETRY ACTION

- Infinitesimal transformations by generator L^a:

$$\delta^a \phi(x) = \partial_x \phi(x)\, \delta^a \vec{x}$$

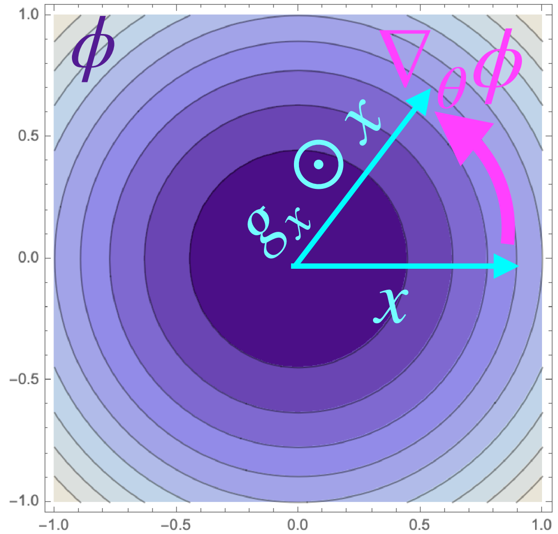
δSEAL: $$\mathscr{L}_\delta = \left\| \sum_{j=1}^{\text{features}} \left(L^a_\phi(\phi) - L^a_x x_j \cdot \partial_{x_j} \phi \right) \right\|^2_{\text{gens, data}}$$

scalar: 0

4-vector: Lϕ

4-vector: Lx

- **Only uses generators - no sampling.**
- Gradients analytically calculated with autograd.
- **Lorentz**: 6 generators: $K_x, K_y, K_z, L_x, L_y, L_z$



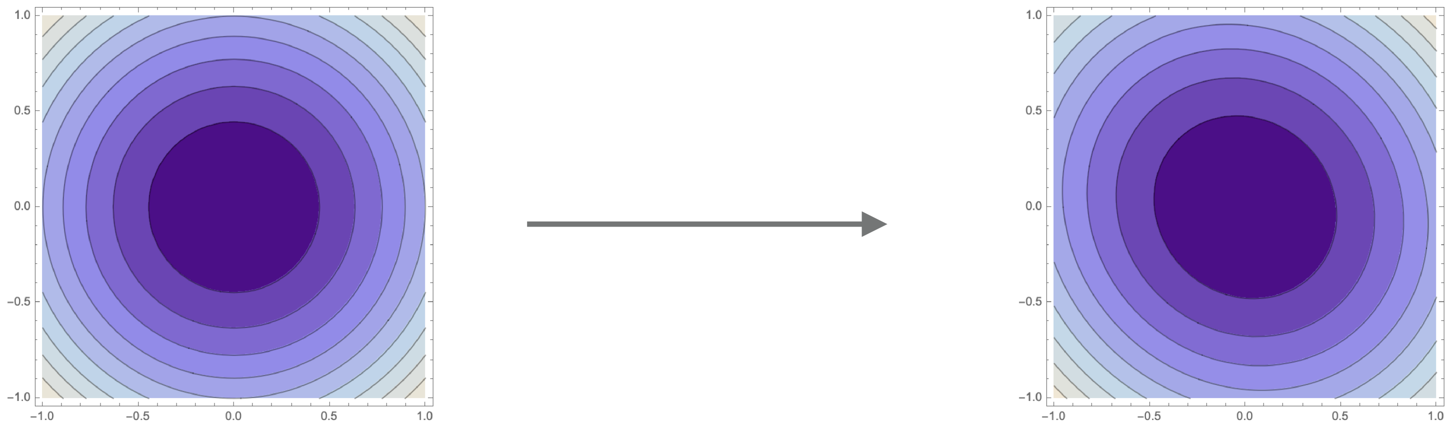
EXPERIMENTS & RESULTS

TOY EXPERIMENTS - REGRESSION

- Input: list of 4-momenta $p_i^\mu \in [-1,1]$

$$\mathscr{L} = \mathscr{L}_{\mathrm{MSE}} + \lambda \mathscr{L}_{\mathrm{SEAL}}$$

- NN with 3 hidden layers of width 300, GeLU activation.
- **Exact Symmetry:** $f_{\mathrm{truth}}\left(p_i^\mu\right) = \mathrm{poly}\left(\mathrm{p_i} \cdot \mathrm{p_j}\right)$
- **Approximate Symmetry:** $f_{\mathrm{truth}}\left(p_i^\mu\right) = \mathrm{poly}\left(\mathrm{p_i} \cdot \mathrm{p_j}, \mathrm{p_i} \cdot \mathrm{s}\right)$
  - “Spurion” $s = \begin{pmatrix} 0 & 0 & 0 & 10^{-3} \end{pmatrix}$

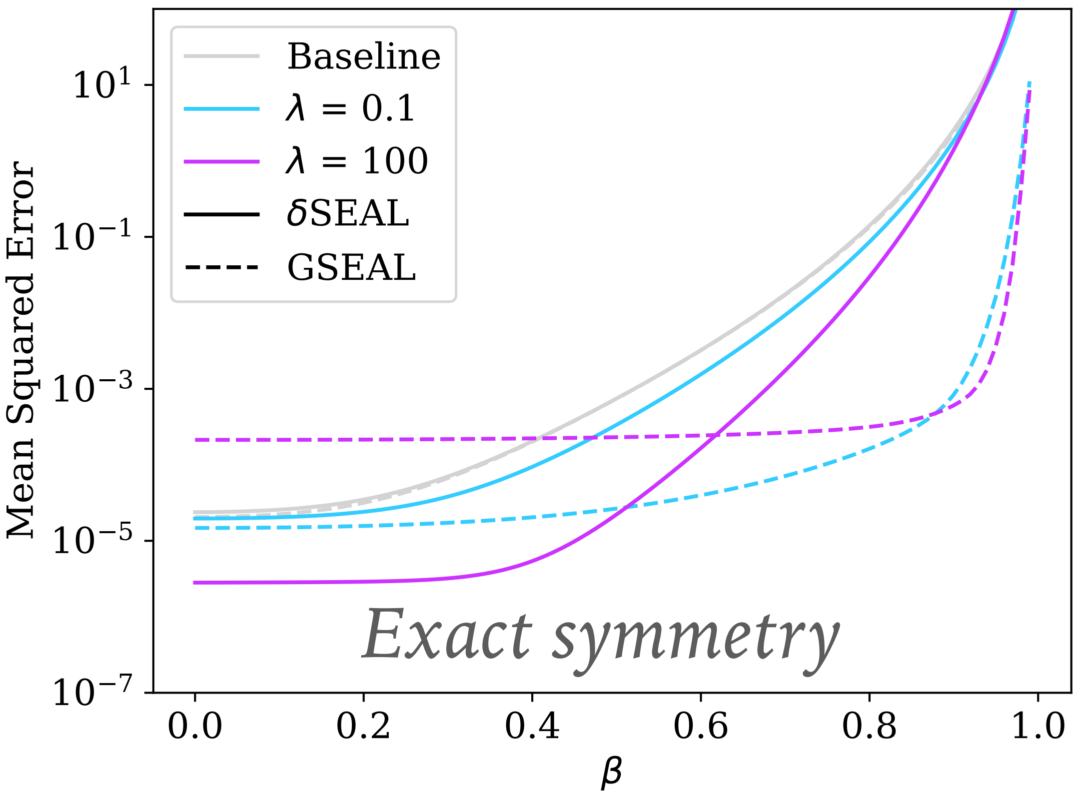


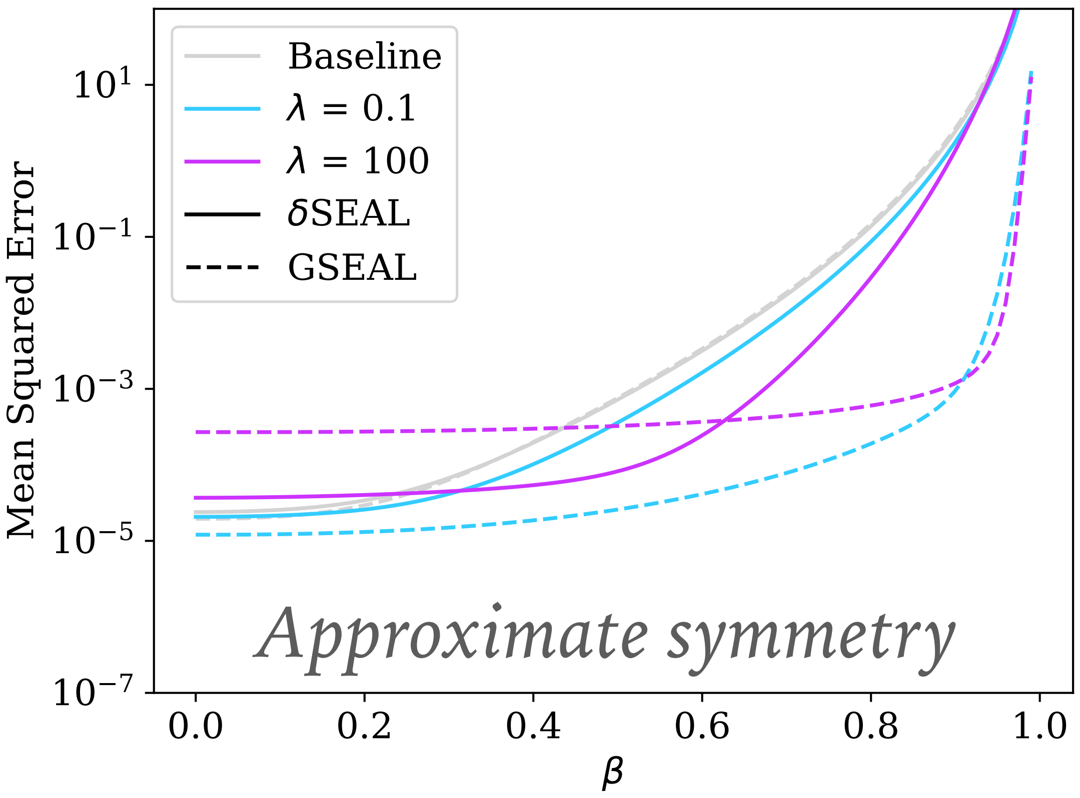
TOYS - RESULTS

$$\mathscr{L}_G = \left\| \phi_{ML}\left(B_i\left(x_i\right)\right) - \phi_{ML}\left(x_i\right) \right\|^2 \qquad \mathscr{L}_\delta = \left\| \frac{\partial \phi_{ML}}{\partial p_\mu} \cdot \left(L_{\mu\nu} p_i^\nu\right) \right\|^2$$

- **SEAL** outperforms baseline in-distribution and on boosted inputs.
- **GSEAL**: flatter as a function of boost, beneficial for large boosts.
- **δSEAL**: extends to non-infinitesimal boosts!
- Can still gain for broken symmetries.

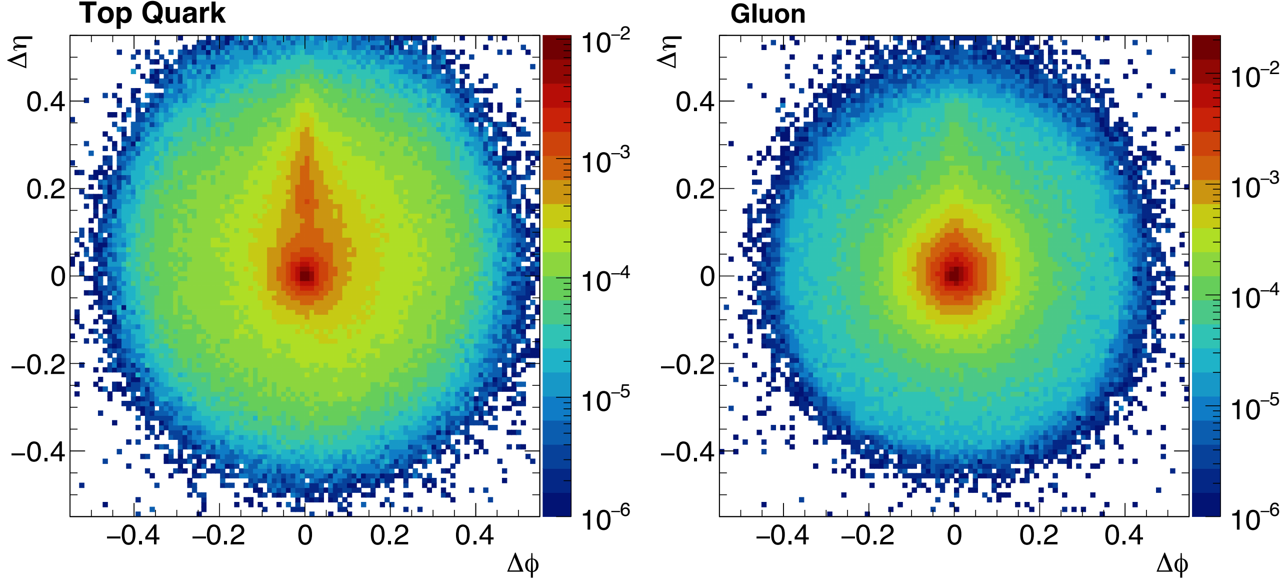
$$\mathscr{L} = \mathscr{L}_{\text{MSE}} + \lambda \mathscr{L}_{\text{SEAL}}$$



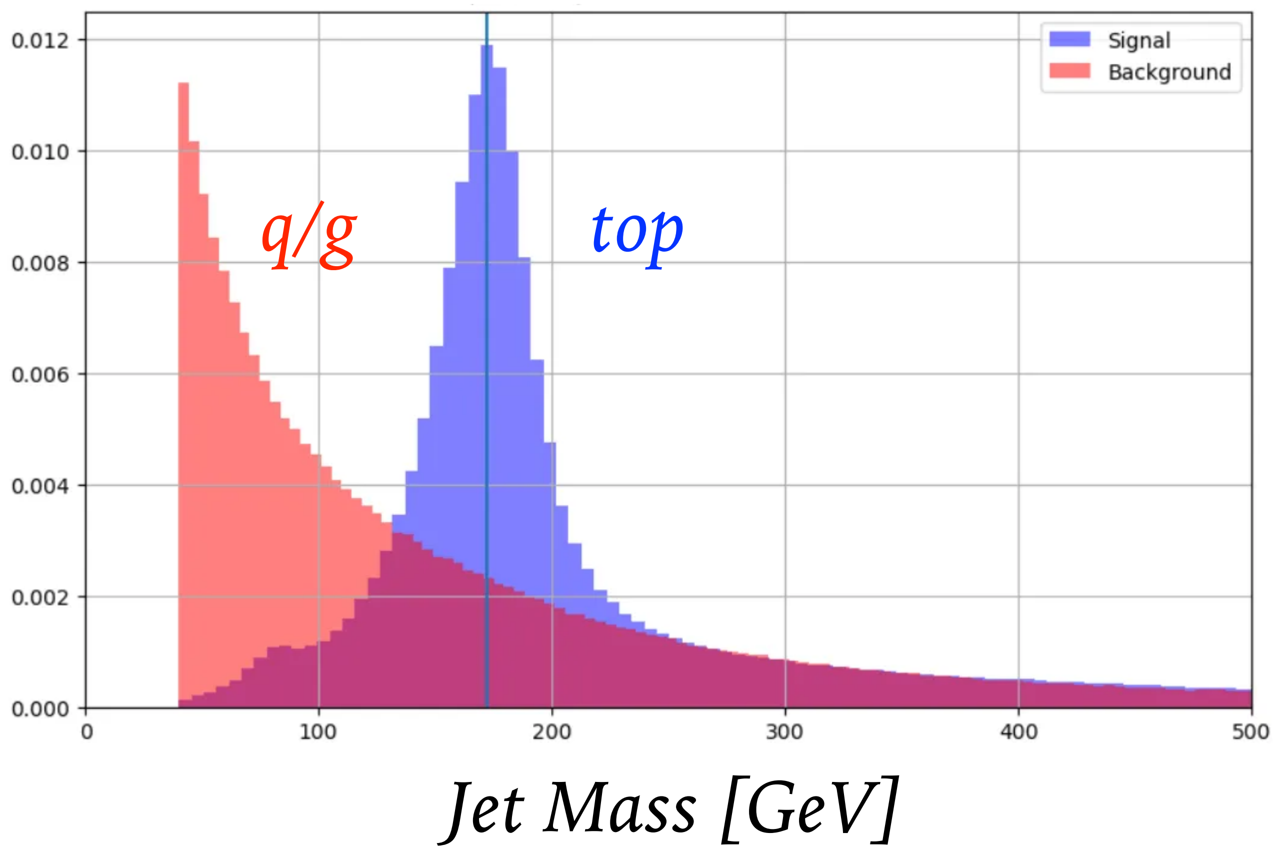


TOP JET TAGGING

- **Goal - learn $p(x\,|\,\text{top})$ vs. $p(x\,|\,\text{QCD}) \rightarrow$ classify jet.**
  - Precision measurements
  - BSM studies
- **ATLAS top tagging dataset**
  - **Most realistic dataset**
  - Full LHC Run-2 conditions (including pile-up)
  - Full detector simulation
  - Event reconstruction



V. Mikuni, F. Canelli,[2102.05073]



ATLAS collaboration (2022),
https://opendata.cern.ch/record/15013

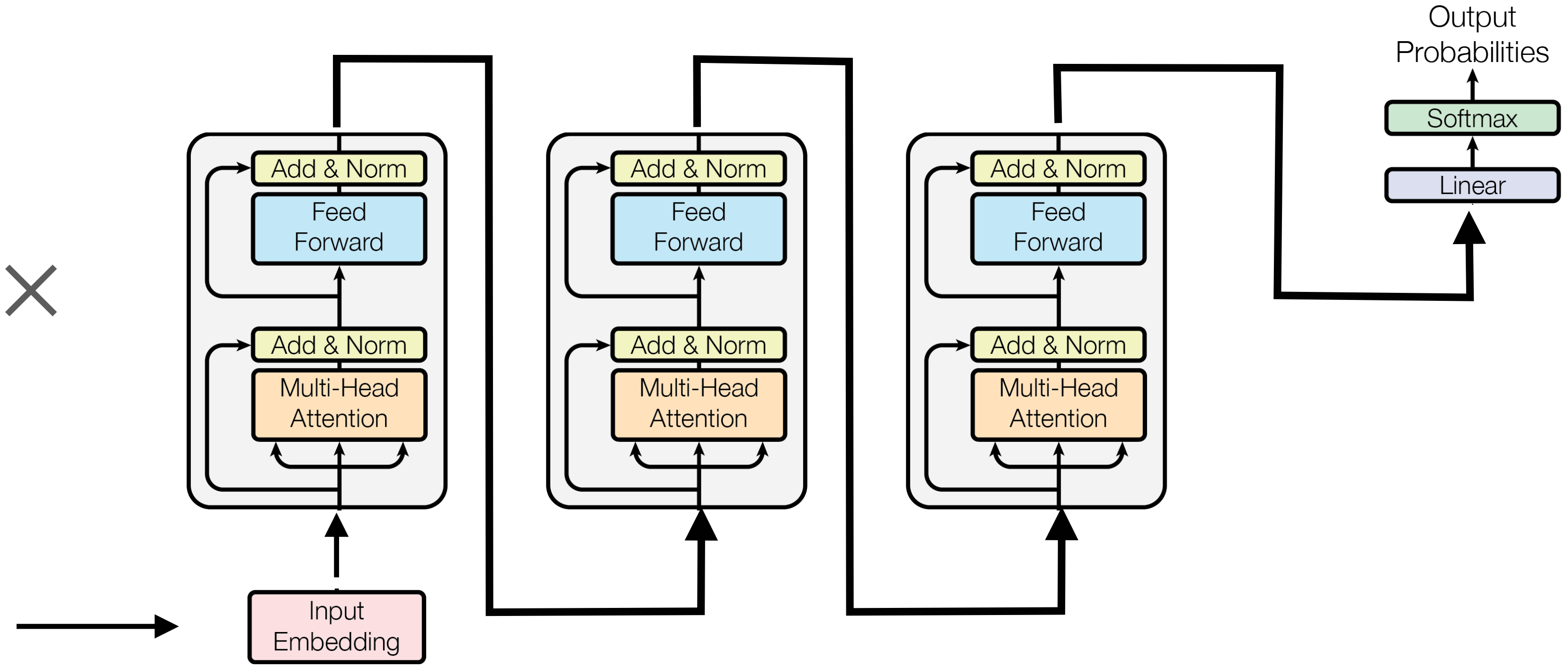
TOP TAGGING - MODEL

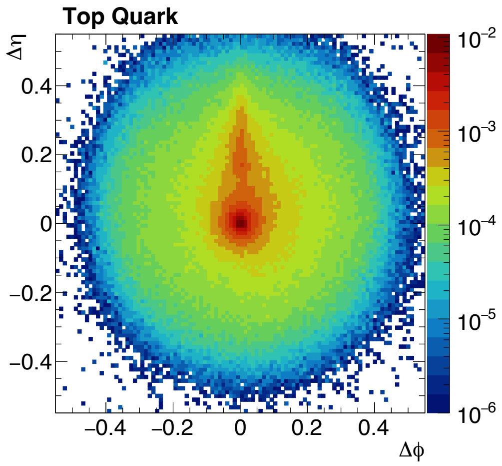
- Input - jet constituents 4-momenta

$$\left\{ p_T^i, E^i, \frac{p_T^i}{p_T^{\text{jet}}}, \frac{E^i}{E^{\text{jet}}}, \Delta\phi^i, \Delta\eta^i, \Delta R^i = \sqrt{\Delta\eta^2 + \Delta\phi^2} \right\}$$

- Transformer - *Lorentz invariant* ≈ $p\left(x \,|\, \text{top}\right)$

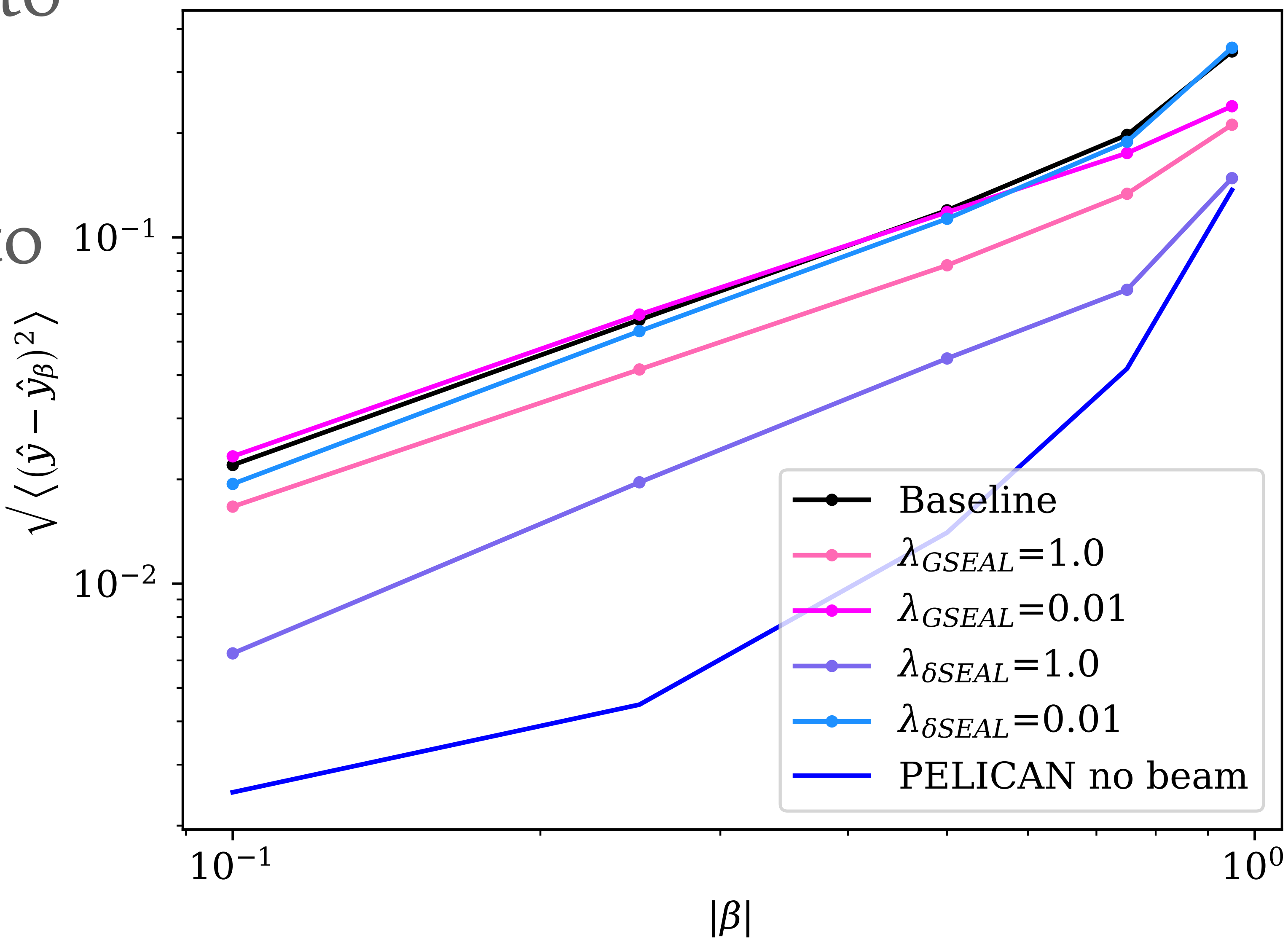
$4\times$





TOP TAGGING RESULTS

- Invariance check - boosting inputs and checking the similarity between prediction for jet and prediction for boosted jet.
- **SEAL** improves invariance to any boost compared to baseline!
- **GSEAL** more sensitive to large boosts.
- δ**SEAL** more sensitive to infinitesimal boosts.



TOP TAGGING RESULTS

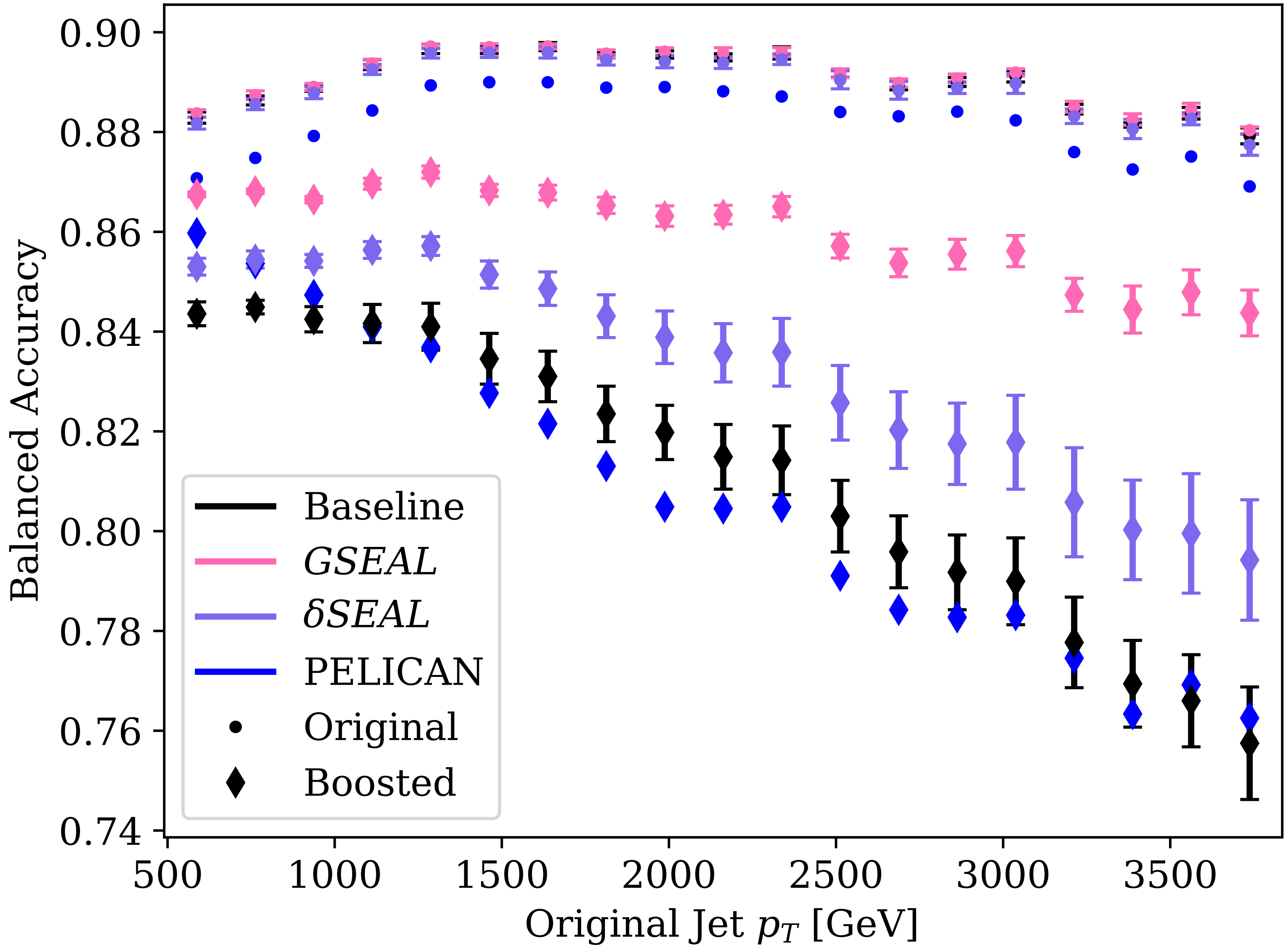
- ➤ SEAL - **improved invariance** while **maintaining performance.**

- Balanced accuracy on test jets.

$$acc\left(y_{ML}(j), y_{\text{truth}}(j)\right)$$

◆ Balanced accuracy on randomly boosted jets with truth labels given by original jets.

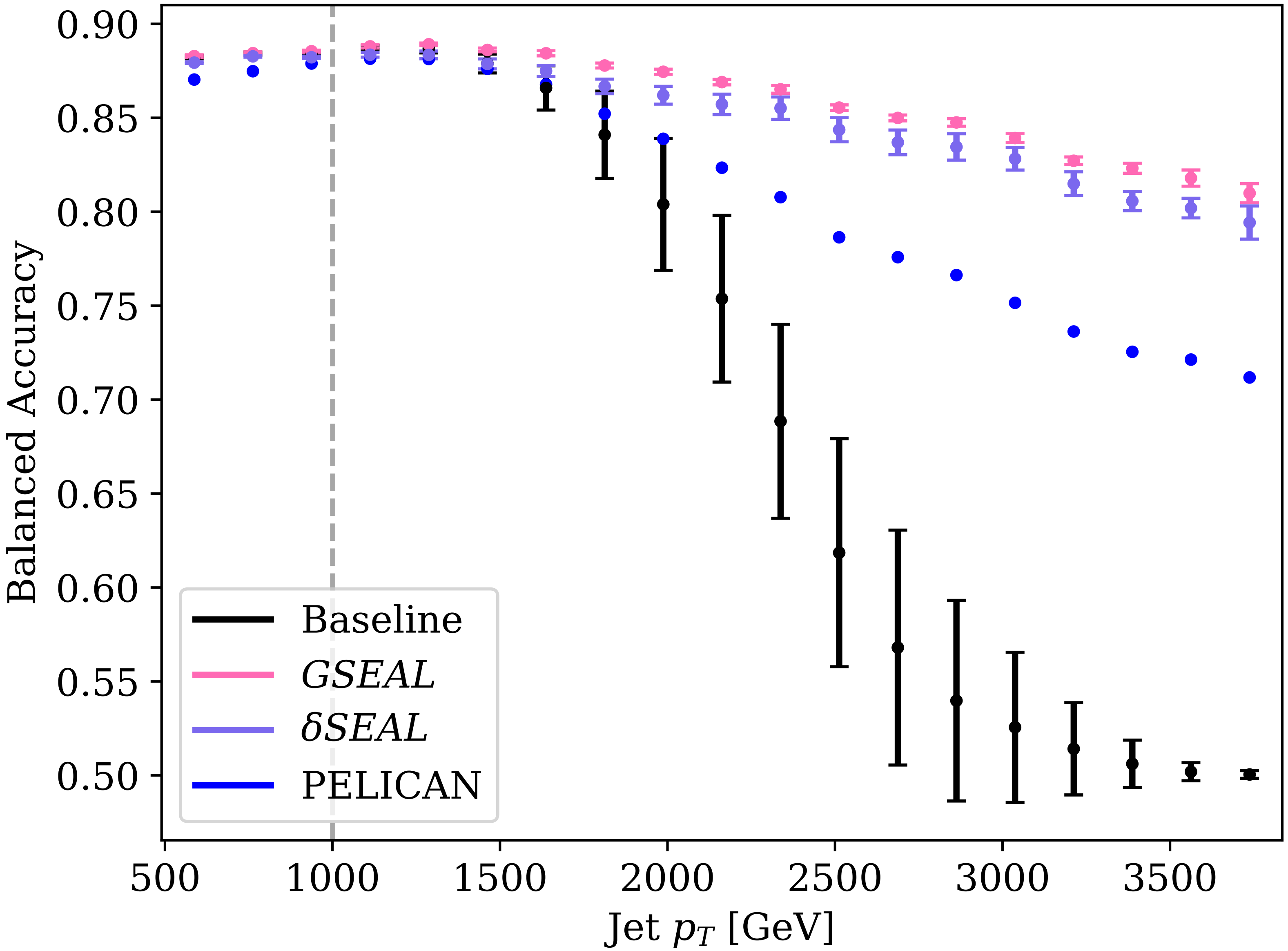
$$acc\left(y_{ML}(Bj), y_{\text{truth}}(j)\right)$$



Model	Bl. Accuracy	AUC	$1/\epsilon_b^{0.3}$	Train Step [s]	Eval Step [s]
Baseline	0.891 ± 0.001	0.959 ± 0.001	652 ± 37	0.03	0.005
+ GSEAL	0.891 ± 0.001	0.959 ± 0.001	638 ± 37	0.06	0.005
+ δSEAL	0.890 ± 0.002	0.959 ± 0.001	620 ± 48	0.06	0.005
PELICAN	0.890	0.959	630	0.4	0.2

TOP TAGGING RESULTS

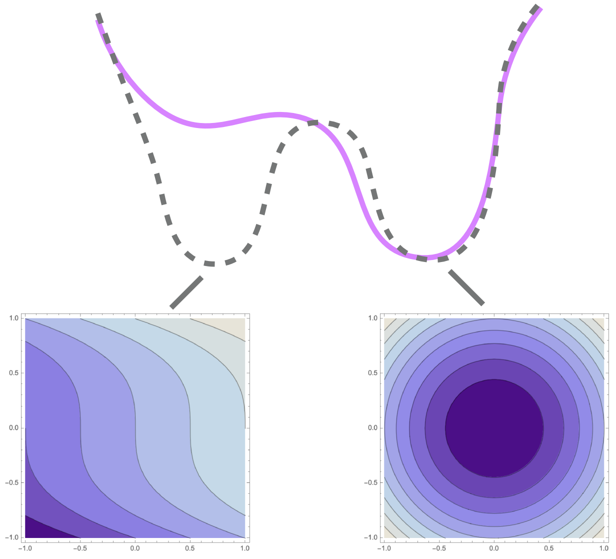
- Extrapolation test: train only on $p_T \leq 1$ TeV
- **SEAL** have smaller uncertainties and better accuracy compared to baseline on unseen p_T!



CONCLUSIONS

- **SEAL - bias the model towards respecting symmetries.**

$$\mathscr{L} = \mathscr{L}_{\text{task}} + \lambda \mathscr{L}_{\text{SEAL}}$$

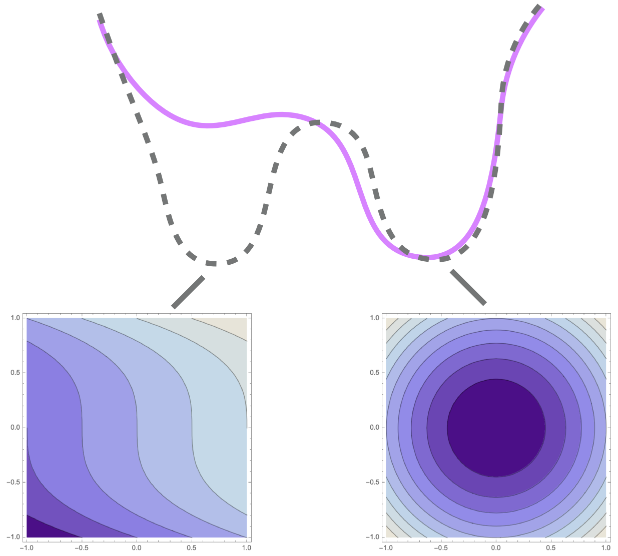


- **Flexible - can be added to any model, easy to implement.**
- Bias is **tunable** and controllable.
- **Multi-purpose** - accommodates approximate symmetries (and no symmetries).
- **Better results for symmetric problems, even if the symmetry is broken.**

CONCLUSIONS

- SEAL - bias the model towards respecting symmetries.

$$\mathscr{L} = \mathscr{L}_{\text{task}} + \lambda \mathscr{L}_{\text{SEAL}}$$



- **Flexible** - can be added to any model, easy to implement.
- Bias is **tunable** and controllable.
- **Multi-purpose** - accommodates approximate symmetries (and no symmetries).
- **Better results for symmetric problems, even if the symmetry is broken.**

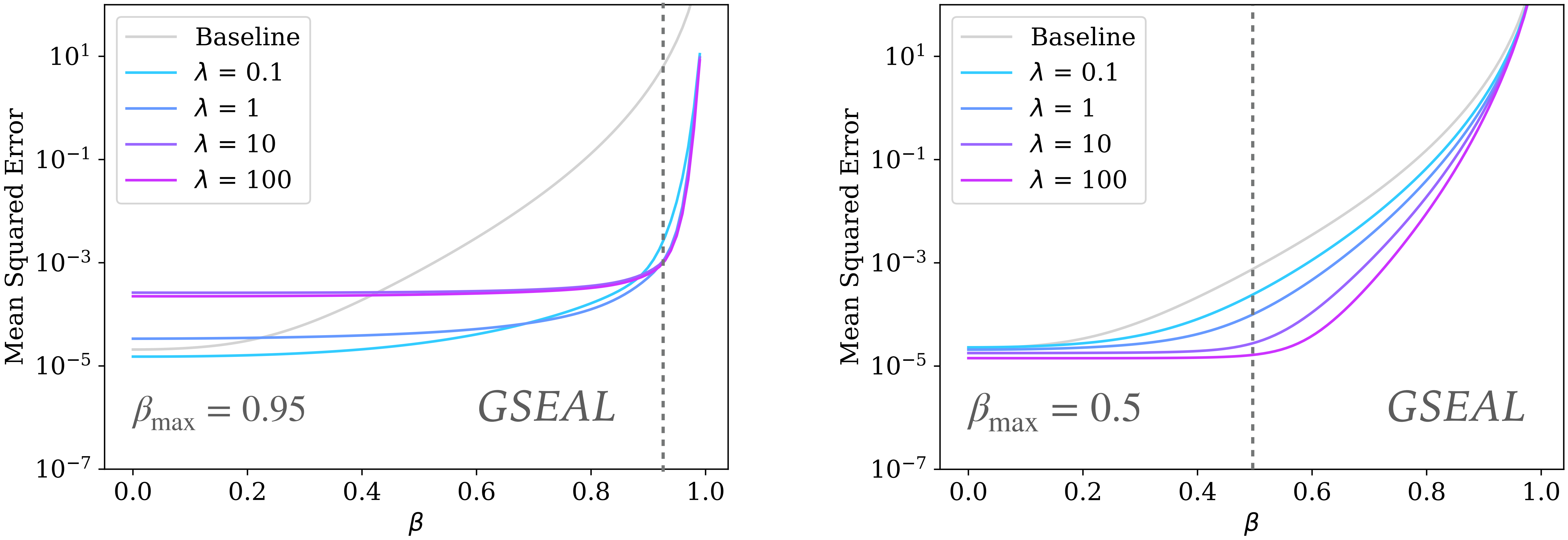
THANK YOU!

BACKUP SLIDES

TOYS - EXACT SYMMETRY

$$\mathscr{L} = \mathscr{L}_{\text{MSE}} + \lambda \mathscr{L}_{SEAL} \qquad \mathscr{L}_G = \left\| \phi_{ML}\left(B_i\left(x_i\right)\right) - \phi_{ML}\left(x_i\right) \right\|^2$$

- **GSEAL** can achieve better performance than baseline on boosted inputs.
- Larger training $\beta_{\max}$ - flatter as function of boost, but can under-perform for small boosts.

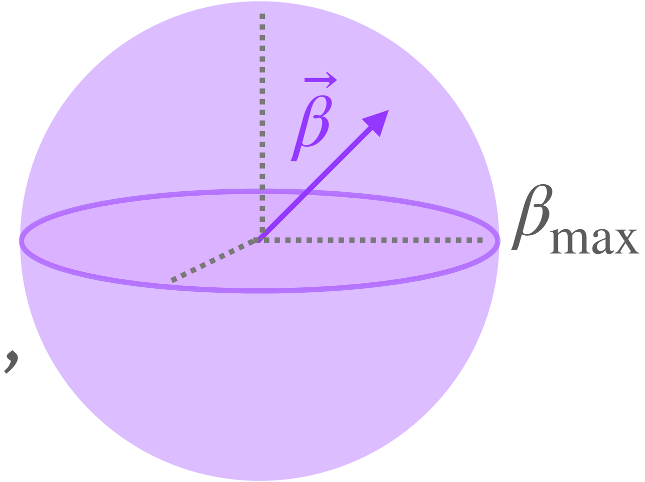


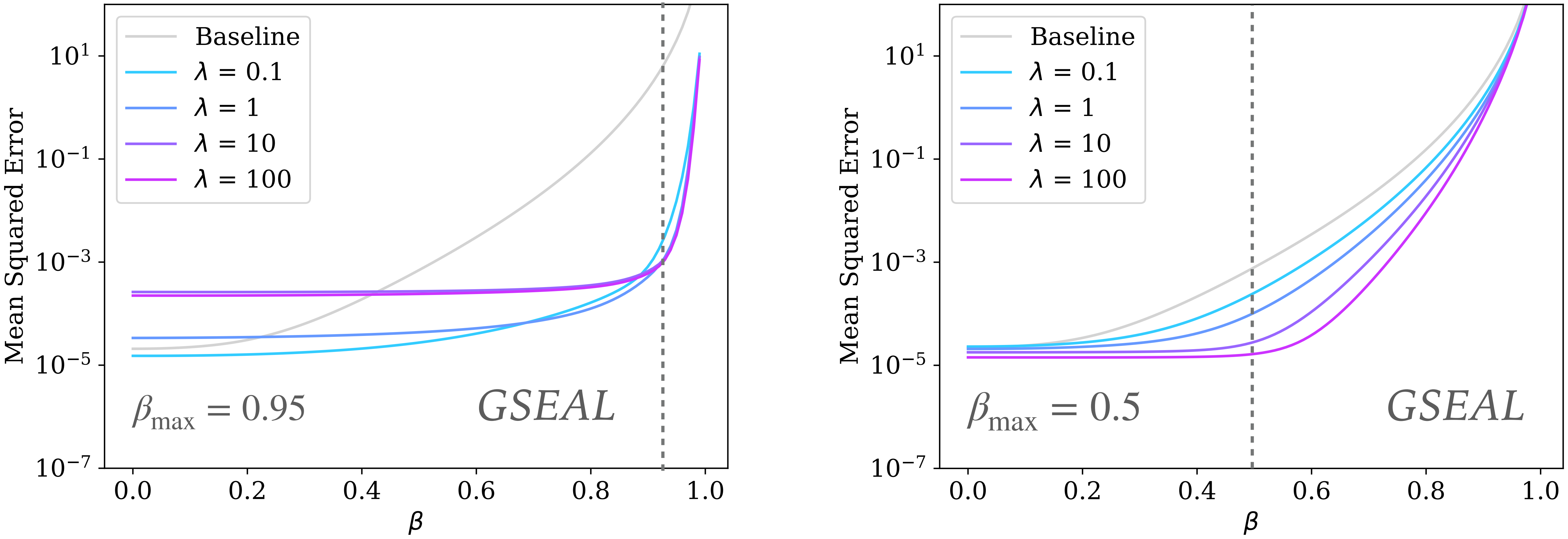
TOYS - EXACT SYMMETRY

$$\mathscr{L} = \mathscr{L}_{\mathrm{MSE}} + \lambda \mathscr{L}_{SEAL}$$

$$\mathscr{L}_G = \left\| \phi_{ML}\left(B_i\left(x_i\right)\right) - \phi_{ML}\left(x_i\right) \right\|^2$$

- **GSEAL** can achieve better performance than baseline on boosted inputs.
- Larger training $\beta_{\max}$ - flatter as function of boost, but can under-perform for small boosts.





TOYS - EXACT SYMMETRY

$$\mathscr{L} = \mathscr{L}_{\text{MSE}} + \lambda \mathscr{L}_{SEAL}$$

$$\mathscr{L}_{\delta} = \left\| \frac{\partial \phi_{ML}}{\partial p_{\mu}} \cdot \left(L_{\mu\nu} p_i^{\nu} \right) \right\|^2$$

- Even infinitesimal loss achieves better performance than baseline, and can extend to non-infinitesimal boosts!
- **δSEAL** better at smaller β.
- Big λ doesn't hurt for small transformations.