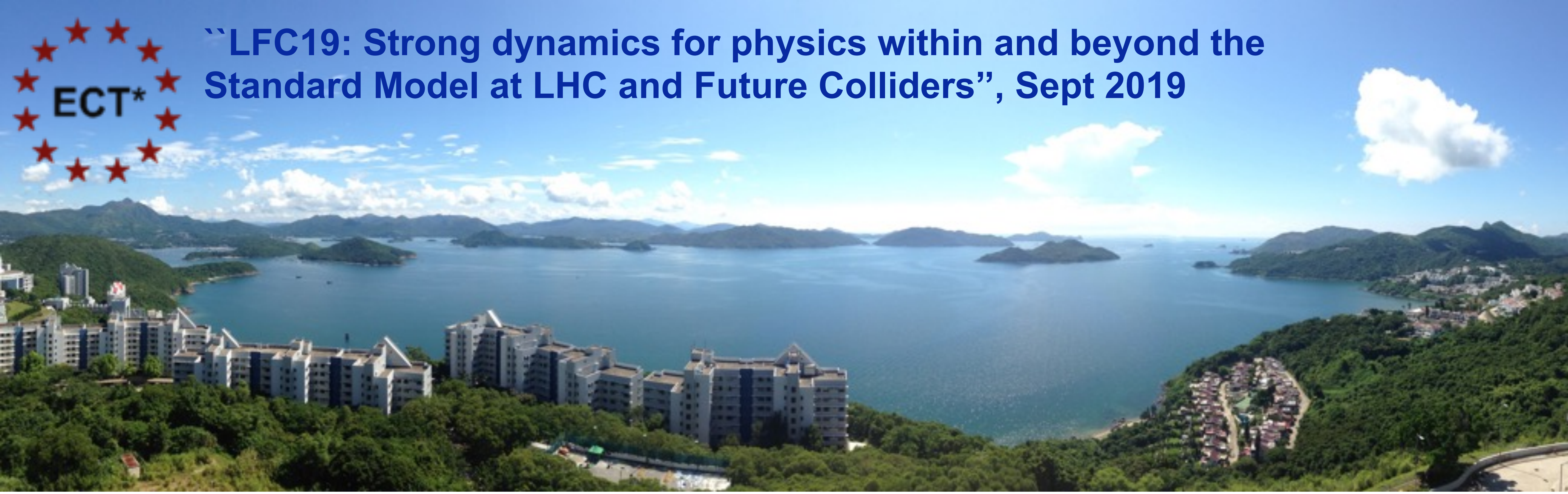


NOVELTY DETECTION MEETS COLLIDER PHYSICS

Tao Liu

The Hong Kong University of Science and Technology

Based on arXiv: 1807.10261 and follow-up project
in collaboration with Jan Hajer, Ying-Ying Li, He Wang, Xu-Hui Jiang and Aurelio Juste

History of Supervised Learning

- High Energy Physics (HEP) is a big data science and has a long history of using supervised ML for data analysis
  - Neural network for top quark search @D0 (1990)
  - BDT was first used by MiniBooNe for neutrino data (2004)
  - BDT has become very popular in HEP data analysis. E.g. in TOP2018, more than 50% of the exp results presented were based on BDT analysis
- Despite its high efficiency in analyzing signal events with complex topologies, the supervised ML method is challenged by some other tasks at colliders

[Talk by Ying-Ying Li, 3rd IML Machine Learning Workshop]

Con of Supervised Learning

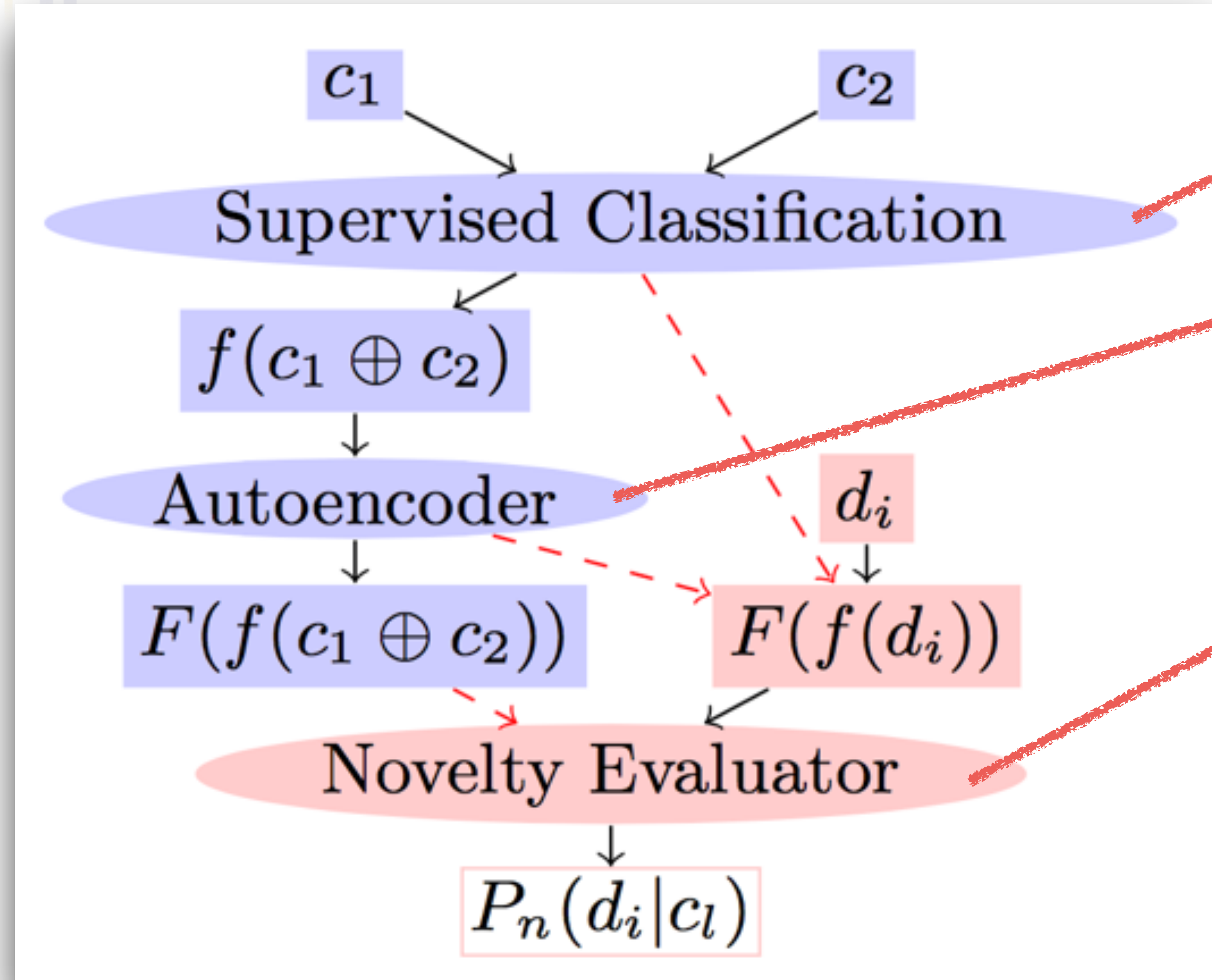
- For new physics processes sharing similar final states but with different kinematics, can we search for them in a universal way?
  - Case I: di-top partner production vs. Z’ production (decaying to top pair)
  - Case II: exotic Higgs decays (rich topologies): h -> Z a and h -> a + DM
- Given the null results at LHC, how to search for new physics which could be highly unexpected?

Supervised learning: model-dependent, incapable for these tasks

Novelty detection

A task of detecting novel events without a prior knowledge
(no data of the signal pattern available for model training).
``Model”- independent, and complementary to supervised learning.

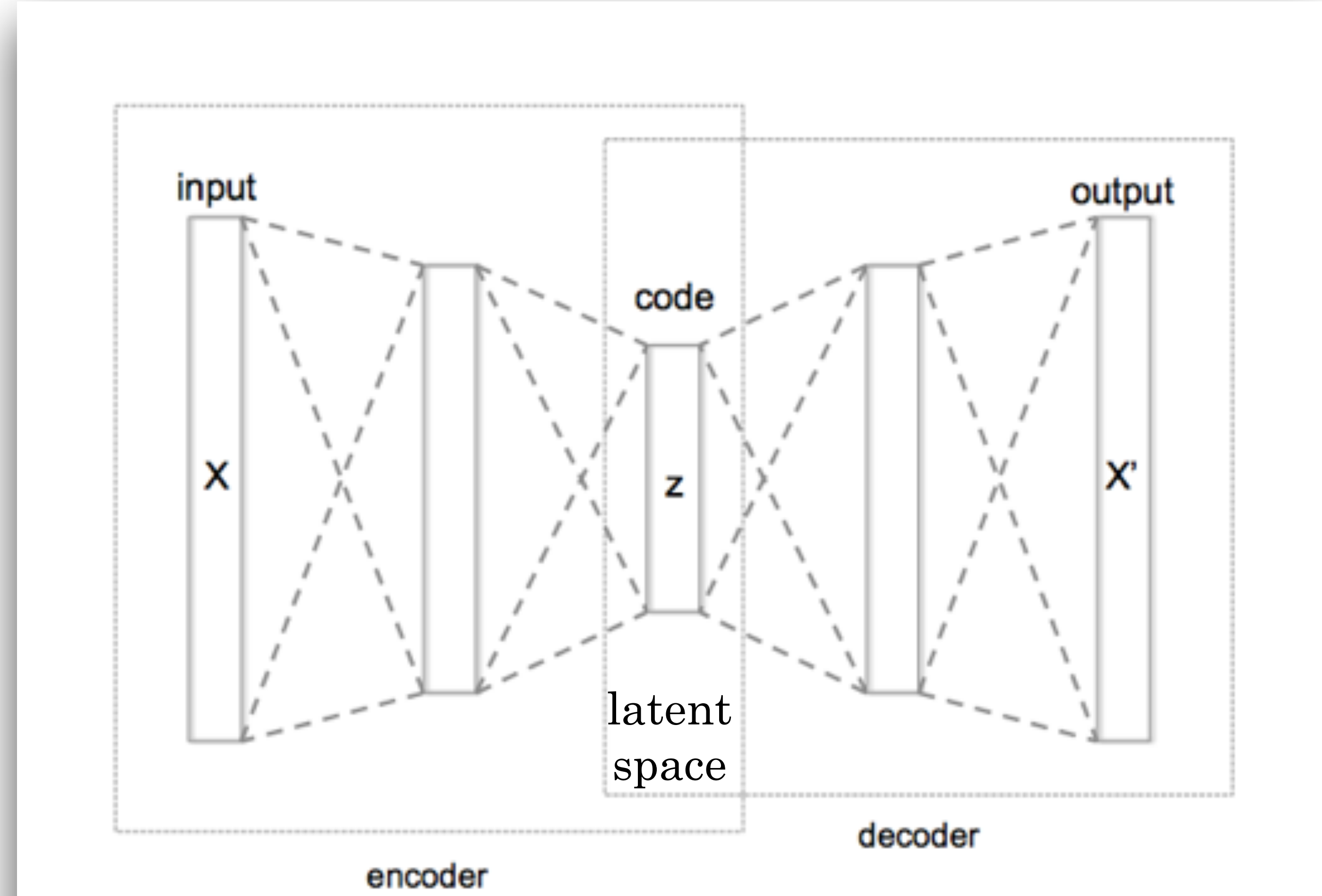
Workflow



- Step 1: (SM/background) feature learning
- Step 2: dimension reduction of feature space (auto-encoder)
- Step 3: novelty evaluation of testing data
- Analyze detection sensitivity based on novelty response of testing data

With this algorithm, new physics can be searched for without a priori knowledge!

Dimensionality Reduction



Compress information with a demand of minimizing reconstruction error

Our Proposal: Autoencoder

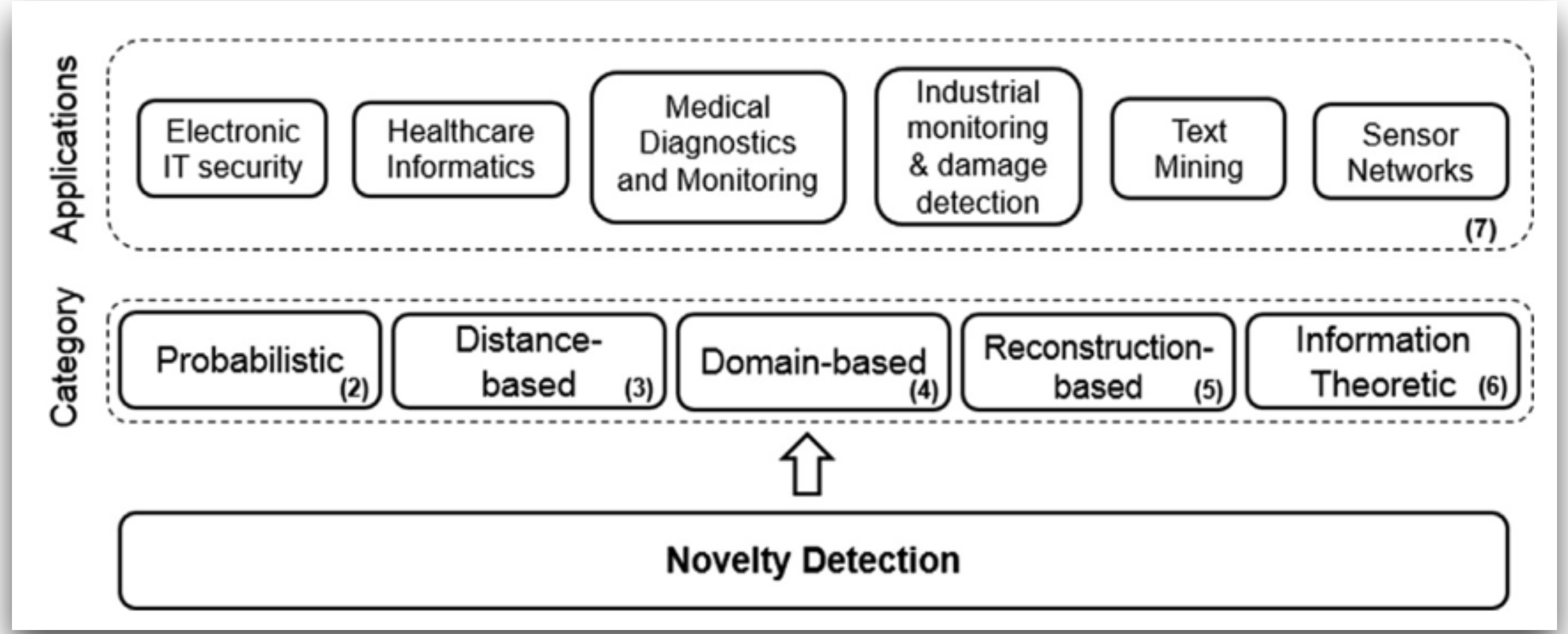
- novelty evaluation is subsequently pursued in its latent space

Dimensionality Reduction

To our knowledge, we were the first to introduce auto-encoder for novelty (anomaly) detection at colliders. After our work, many others came out:

- arXiv:1808.08992: ``Searching for New Physics with Deep Autoencoders'', Marco Farina, Yuichiro Nakai, and David Shih
- arXiv:1808.08992: ``QCD or What?'', Theo Heimel, Gregor Kasieczka, Tilman Plehn, and Jennifer M Thompson
- arXiv:1811.10276, ``Variational Autoencoders for New Physics Mining at the Large Hadron Collider'', Olmo Cerri, Thong Q. Nguyen, Maurizio Pierini, Maria Spiropulua and Jean-Roch Vlimant
- arXiv:1903.02032, ``A robust anomaly finder based on autoencoder'', Tuhin S. Roy and Aravind H. Vijay
- arXiv:1905.10384, ``Adversarially-trained autoencoders for robust unsupervised new physics searches'', Andrew Blance, Michael Spannowsky, and Philip Waite
- … … … …

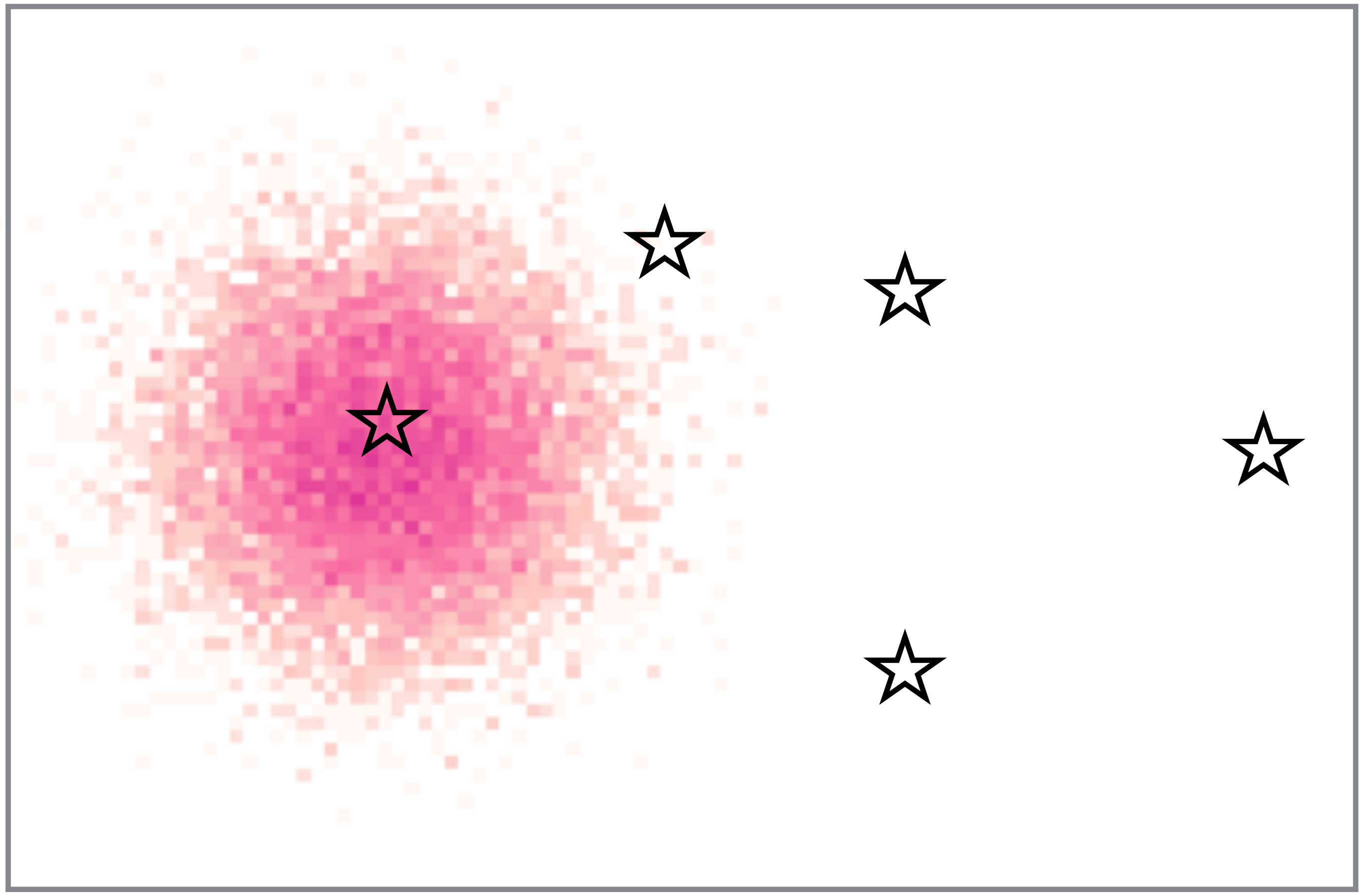
Novelty Evaluation



The history of novelty detection is basically a history of developing novelty evaluators

[M. Pimentel, D. Clifton, L. Clifton, and L. Tarassenko, 2014]

Traditional Wisdom: Isolation-based



Scoring according to the distance or isolation of individual testing point from the training data distribution in a feature space

Traditional Wisdom: Isolation-based

$$\Delta_{\text{trad}} = \frac{d_{\text{train}} - \langle d'_{\text{train}} \rangle}{\langle d'^2_{\text{train}} \rangle^{1/2}} \qquad \mathcal{O} = \frac{1}{2}\left(1 + \operatorname{erf}\left(\frac{c\Delta}{\sqrt{2}}\right)\right)$$

Novelty measure: range unnormalized

Novelty evaluator: $0 \leq \mathcal{O} \leq 1$

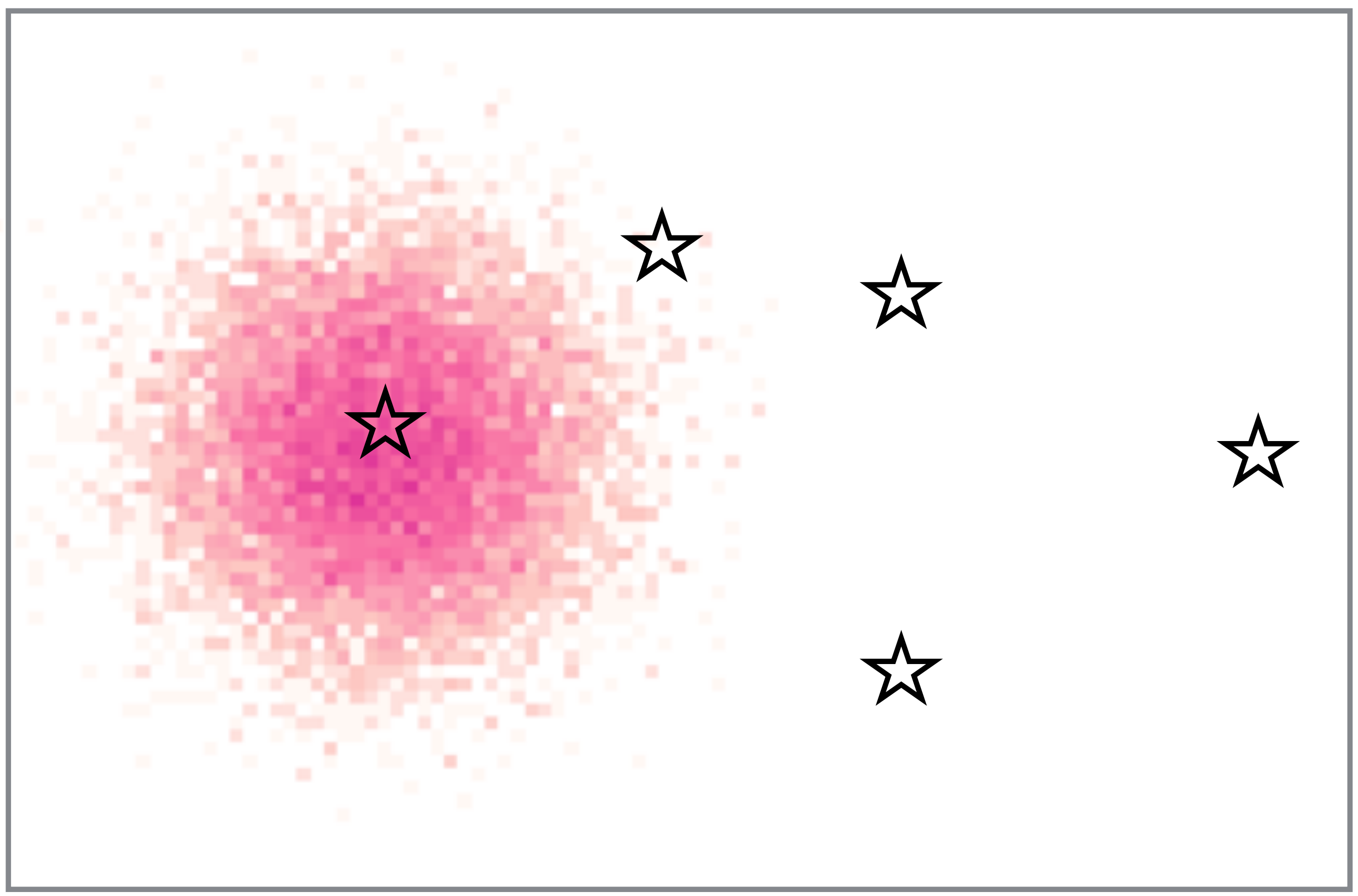
- d_{train} : mean distance of a testing data point to its k nearest neighbors
- $\langle d'_{\text{train}} \rangle$: average of the mean distances defined for its k nearest neighbors
- $\langle d'^2_{\text{train}} \rangle^{1/2}$: standard deviation of the latter
- All quantities are defined wrt the training dataset

[H. Kriegel, P. Kroger, E. Schubert, and A. Zimek, 2009]

[R. Socher, M. Ganjoo, C. D. Manning, and A. Ng , 2013]

Traditional Wisdom: Isolation-based

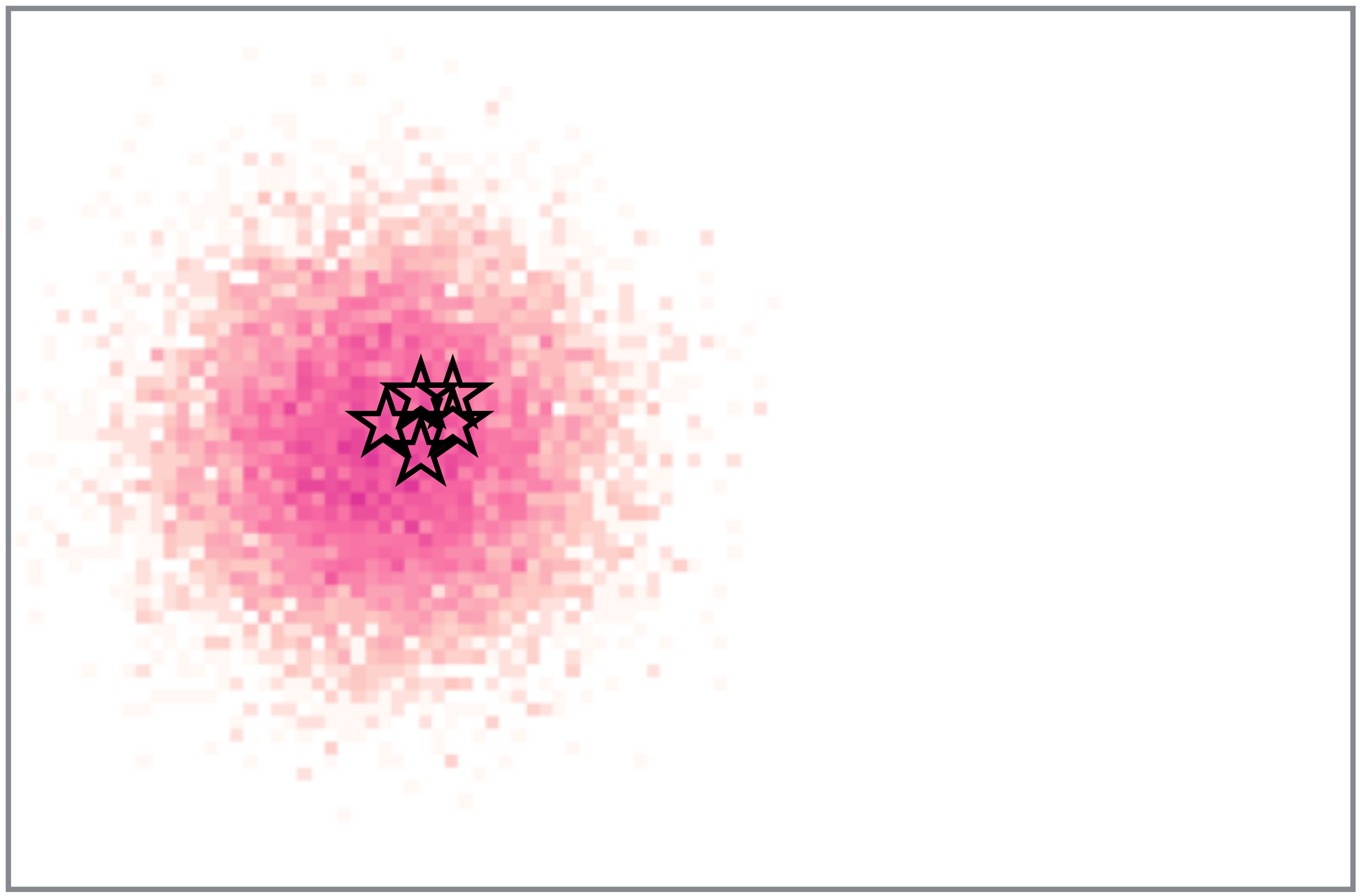
$$\Delta_{\text{trad}} = \frac{d_{\text{train}} - \langle d'_{\text{train}} \rangle}{\langle d'^2_{\text{train}} \rangle^{1/2}} \qquad \mathcal{O} = \frac{1}{2}\left(1 + \operatorname{erf}\left(\frac{c\Delta}{\sqrt{2}}\right)\right)$$



- Large distance => high score
- Short distance => low score
- => a measure of isolation
- Note: reconstruction error as a novelty evaluator is isolation-based in essence
- Successful while being applied to recognize, e.g., anomalous finger print or face!

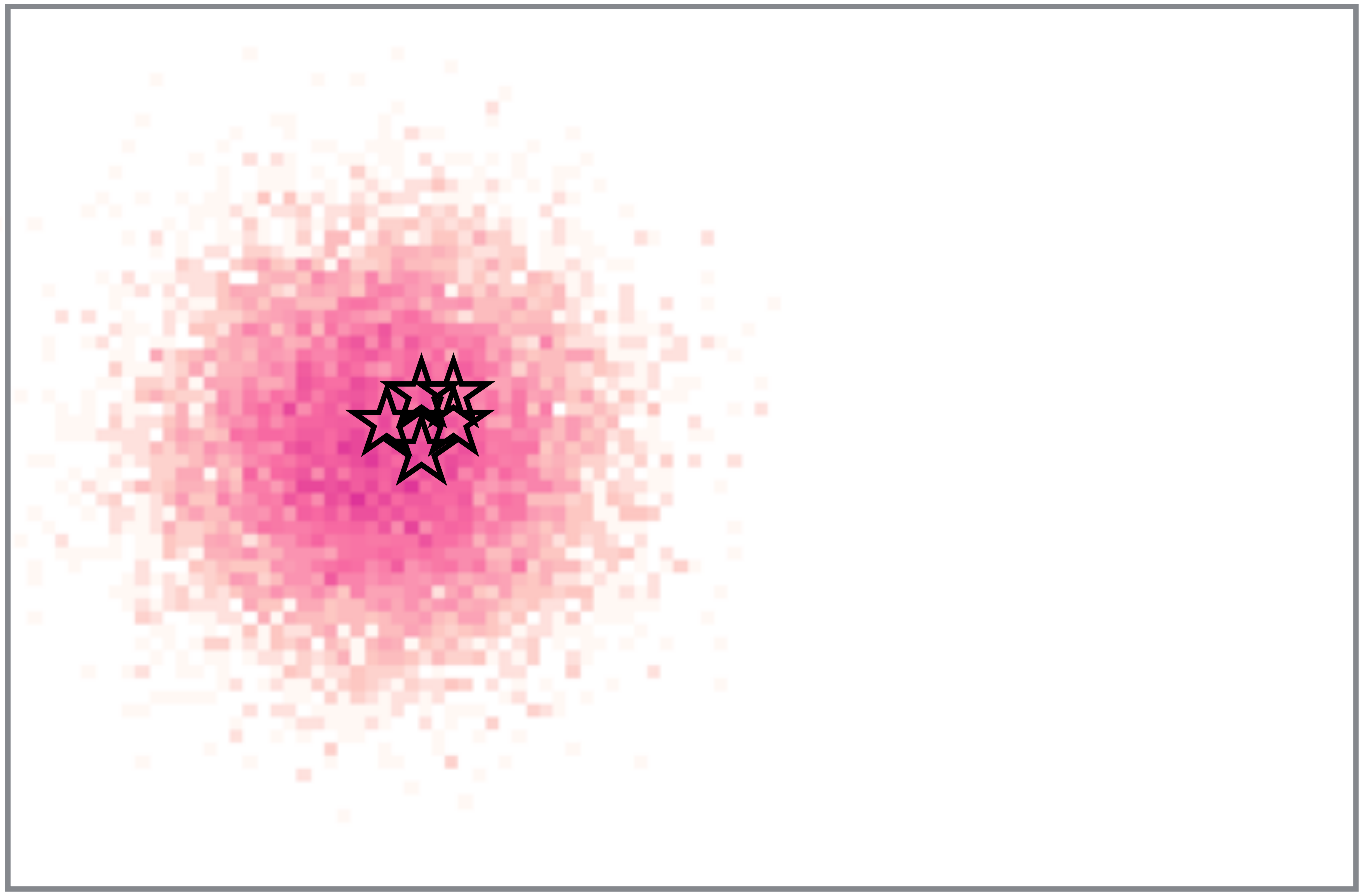
Traditional Wisdom: Isolation-based

$$\Delta_{\mathrm{trad}} = \frac{d_{\mathrm{train}} - \langle d'_{\mathrm{train}} \rangle}{\langle d'^2_{\mathrm{train}} \rangle^{1/2}} \qquad \mathcal{O} = \frac{1}{2}\left(1 + \mathrm{erf}\left(\frac{c\Delta}{\sqrt{2}}\right)\right)$$



- However, this design is insensitive to the clustering structure of the testing data with unknown pattern
- The clustering features such as resonance, shape, etc., are generally important for BSM physics detection

Calling for: Clustering-based



Scoring according to the clustering around each testing point on top of the training data distribution in a feature space

Novelty Evaluators: New Input

$$\Delta_{\mathrm{trad}} = \frac{d_{\mathrm{train}} - \langle d'_{\mathrm{train}} \rangle}{\langle d'^2_{\mathrm{train}} \rangle^{1/2}} \quad \Delta_{\mathrm{new}} = \frac{d_{\mathrm{test}}^{-m} - d_{\mathrm{train}}^{-m}}{d_{\mathrm{train}}^{-m/2}}$$

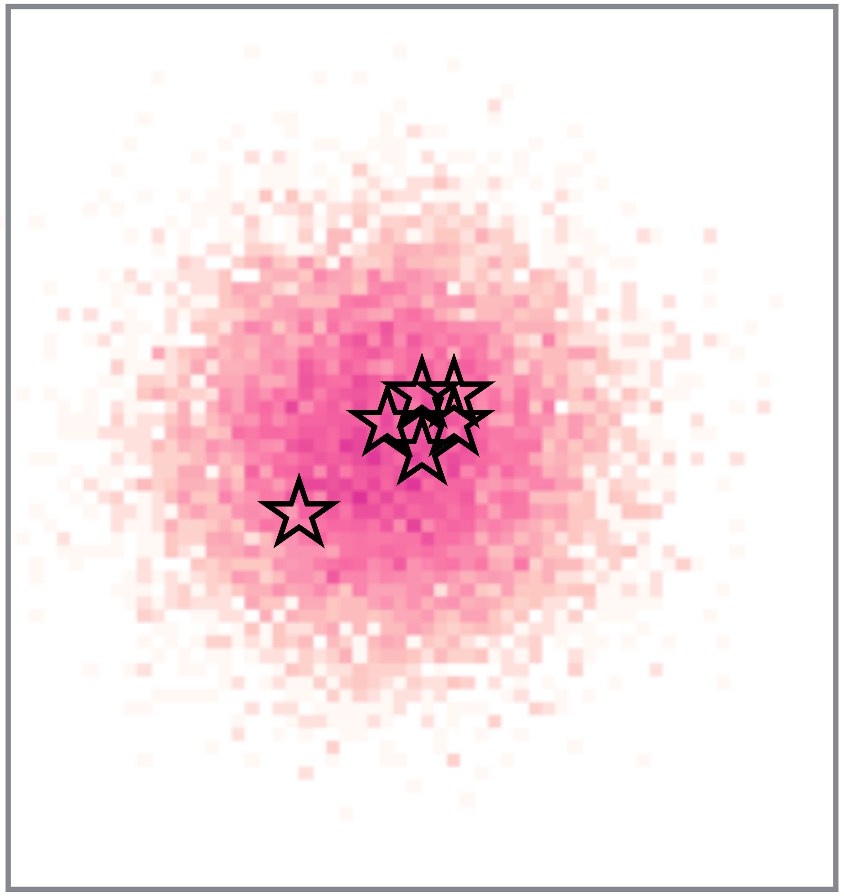
- d_{train}: mean distance of a testing data point to its k nearest neighbors in the training dataset
- d_{test} : mean distance of a testing data point to its k nearest neighbors in the testing dataset
- m: dimension of the feature space
- Novelty response is evaluated by comparing local densities of the testing point in the training and testing datasets
- Approximately statistical interpretation : $\Delta_{\mathrm{new}} \propto \frac{S}{\sqrt{B}}\Big|_{\mathrm{local\ bin}}$

Novelty Evaluators: New Input

$$\Delta_{\text{trad}} = \frac{d_{\text{train}} - \langle d'_{\text{train}} \rangle}{\langle d'^2_{\text{train}} \rangle^{1/2}} \qquad \Delta_{\text{new}} = \frac{d_{\text{test}}^{-m} - d_{\text{train}}^{-m}}{d_{\text{train}}^{-m/2}}$$

Training dataset

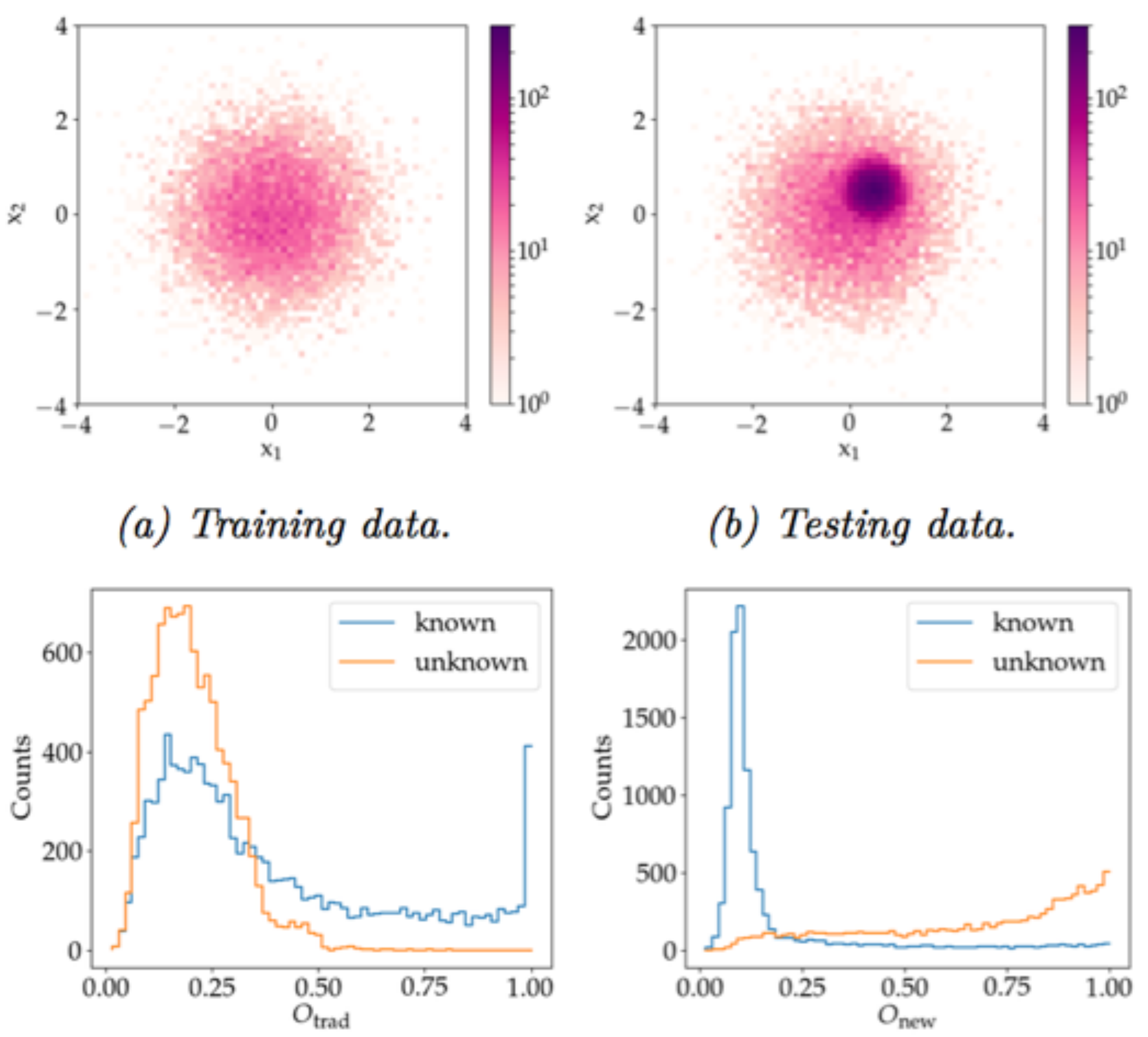
VS



Testing dataset

- Big density difference => high score
- Small density difference => low score
- => a measure of clustering

Novelty Evaluators: Performance Comparison



(a) Training data. (b) Testing data.

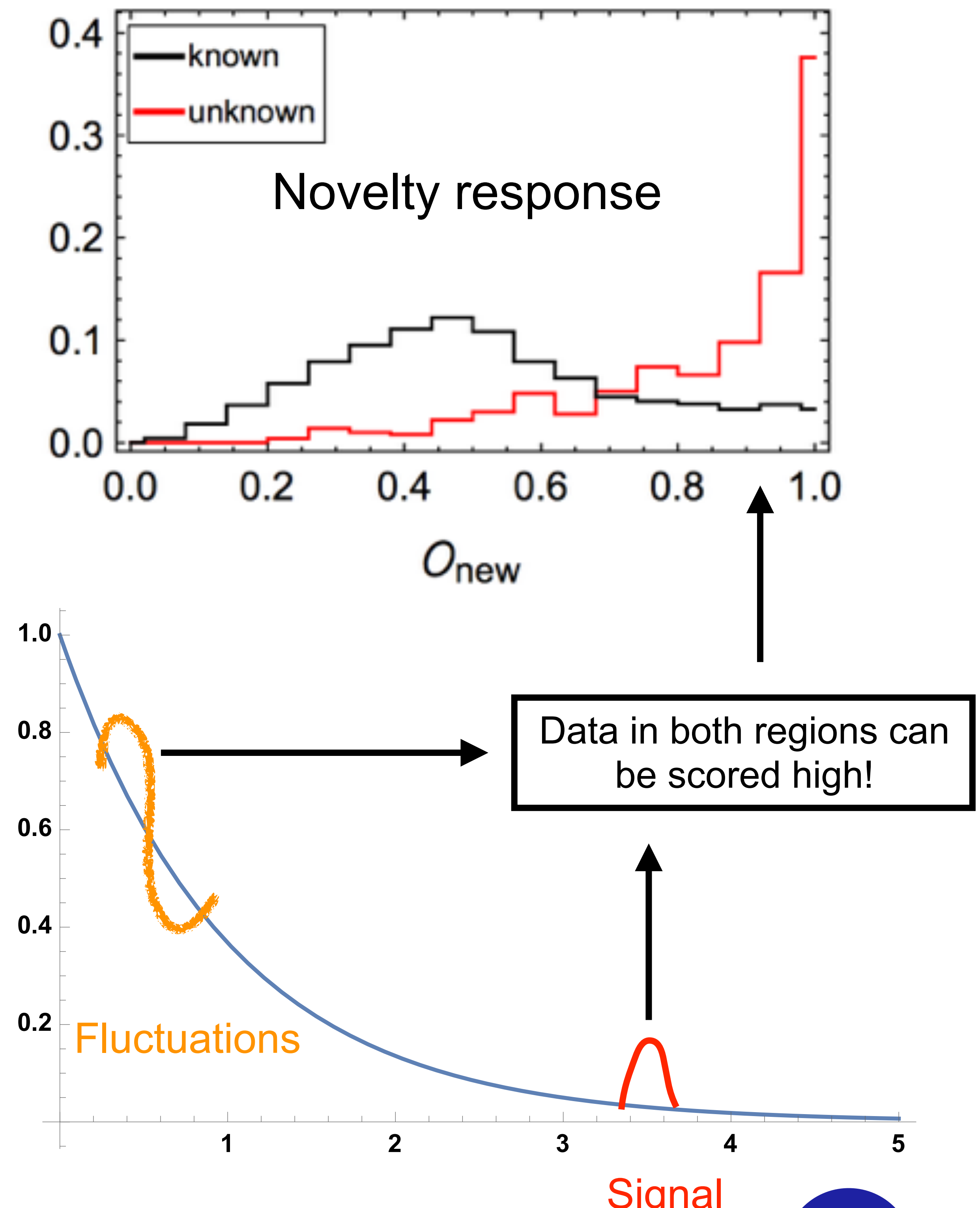
(c) O_{trad} performance. (d) O_{new} performance.

- Consider 2D Gaussian samples
- Training dataset: known pattern only
- Testing dataset: known + unknown patterns
- Compared to O_trad, the novelty response of unknown-pattern data is much stronger for O_new
- => A well-separation between the known- and unknown-pattern data distributions

``Look Elsewhere Effect''

$$\Delta_{\text{new}} = \frac{d_{\text{test}}^{-m} - d_{\text{train}}^{-m}}{d_{\text{train}}^{-m/2}}$$

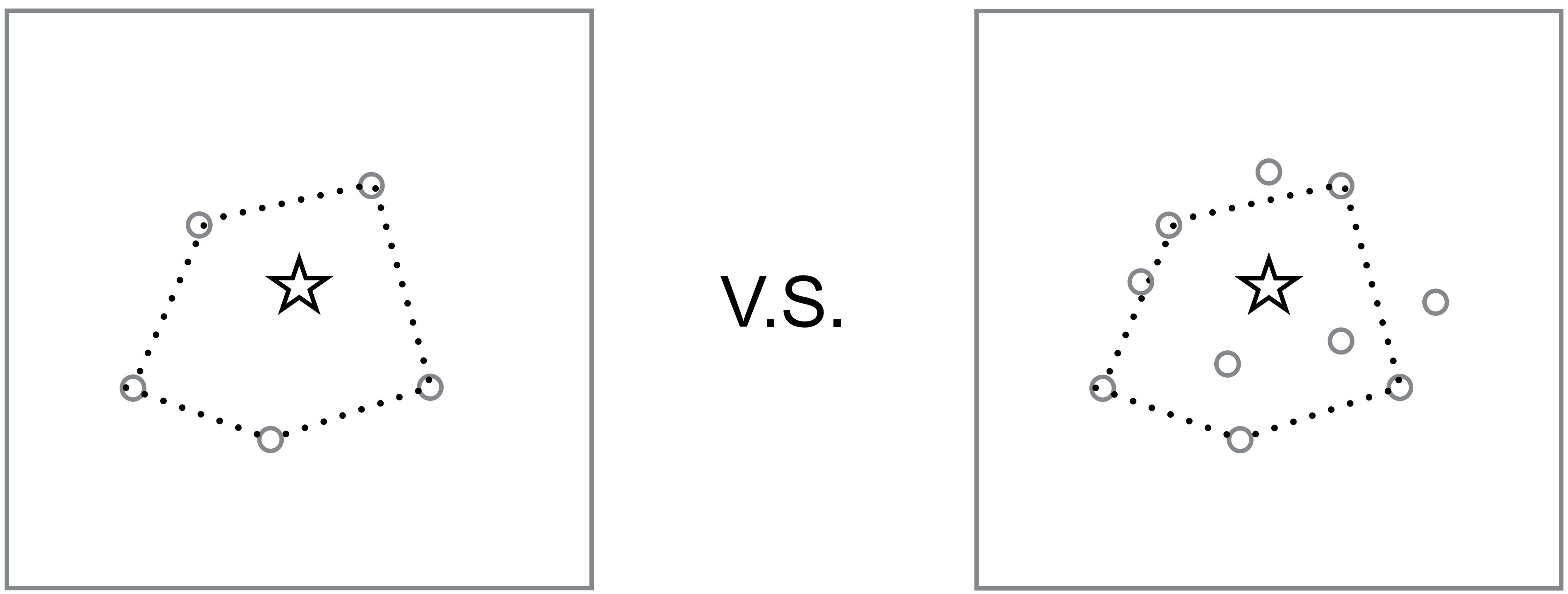
Without a priori knowledge on signal, novelty detection might suffer from a large ``Look Elsewhere Effect (LEE)'', given the feature space to probe!



``Look Elsewhere Effect'' - Central Limit Theorem

The influence of fluctuations for detection sensitivity can be compensated for as the luminosity L increases, if k scales with L.

This can be understood since more and more data are used to calculate d_{test} in the local bin which is barely changed.



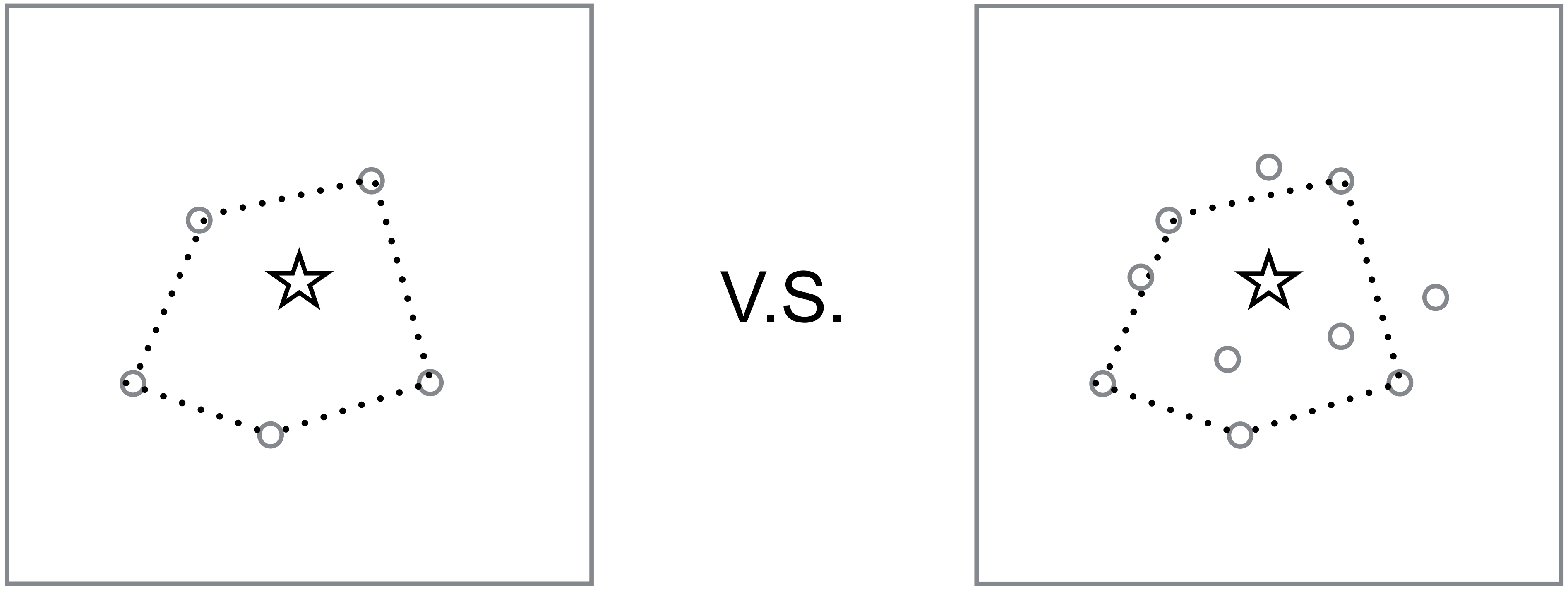
L

2 * L

``Look Elsewhere Effect'' - Central Limit Theorem

Central Limit Theorem

The standard deviation of the novelty response based on Delta_new scales with 1/sqrt{k} or 1/sqrt{L}, for the testing data with known patterns only.



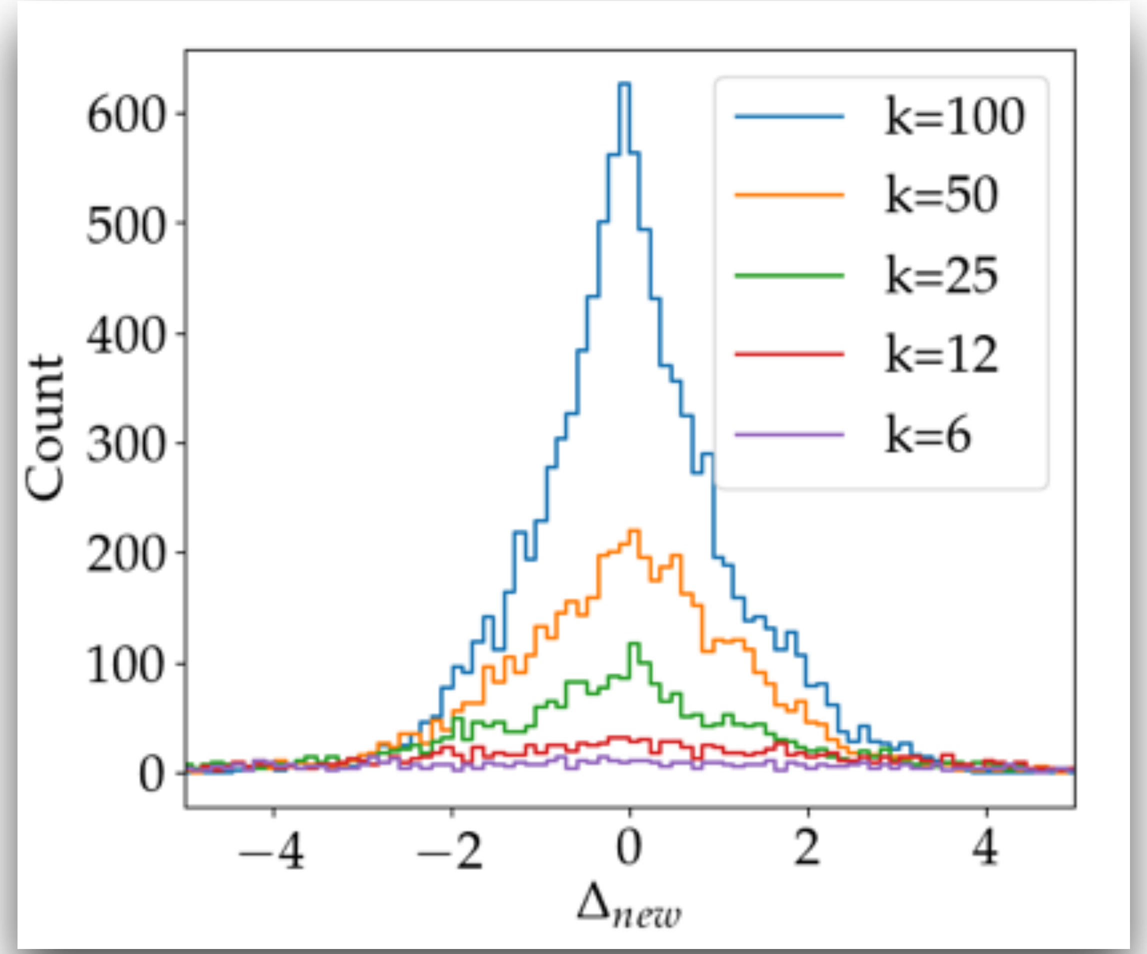
L

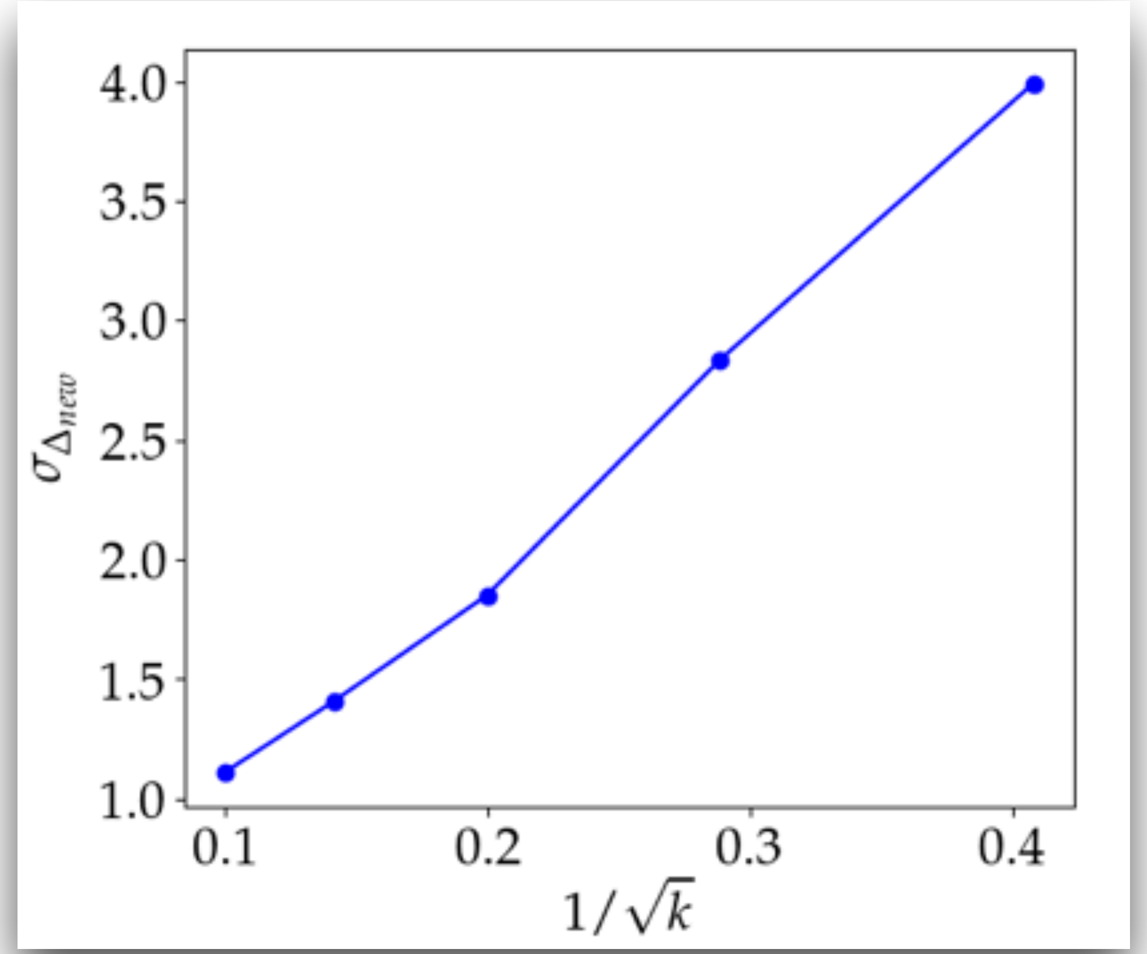
2 * L

``Look Elsewhere Effect'' - Central Limit Theorem

Central Limit Theorem

The standard deviation of the novelty response based on Delta_new scales with 1/sqrt{k} or 1/sqrt{L}, for the testing data with known patterns only.

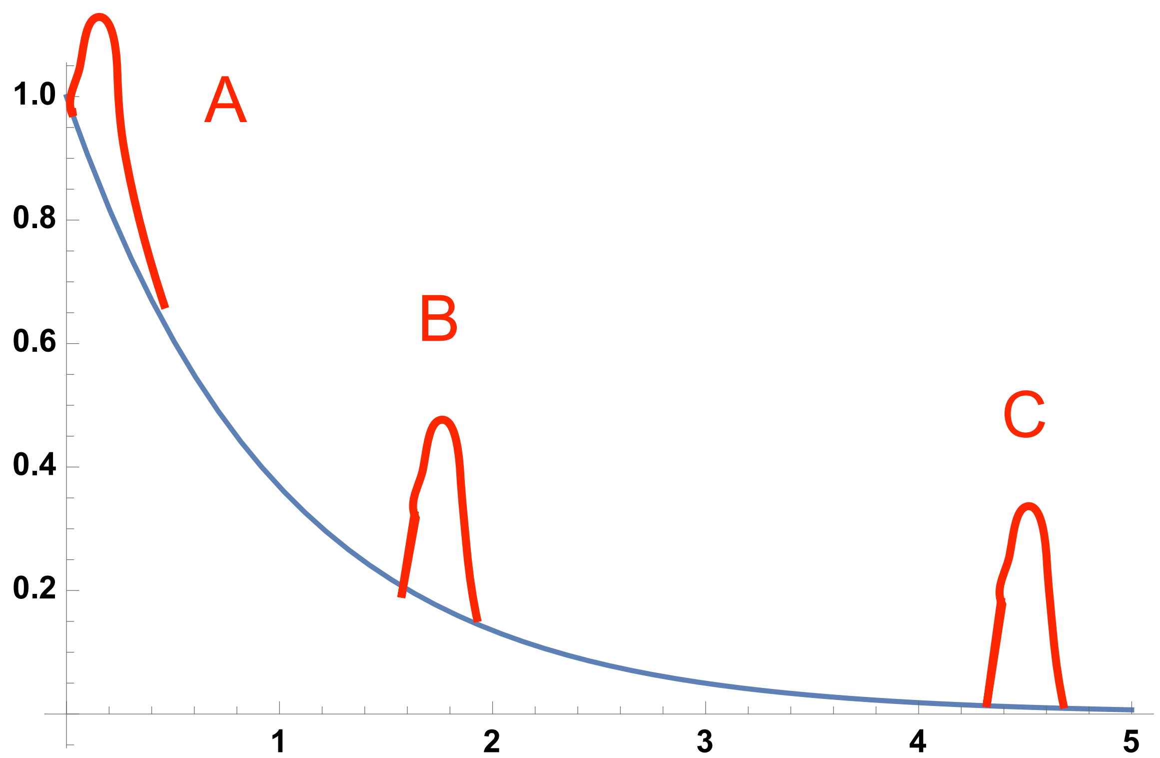




Strategy to Address Large LEE

The suppression of LEE by luminosity might not be sufficient if S/B is small.

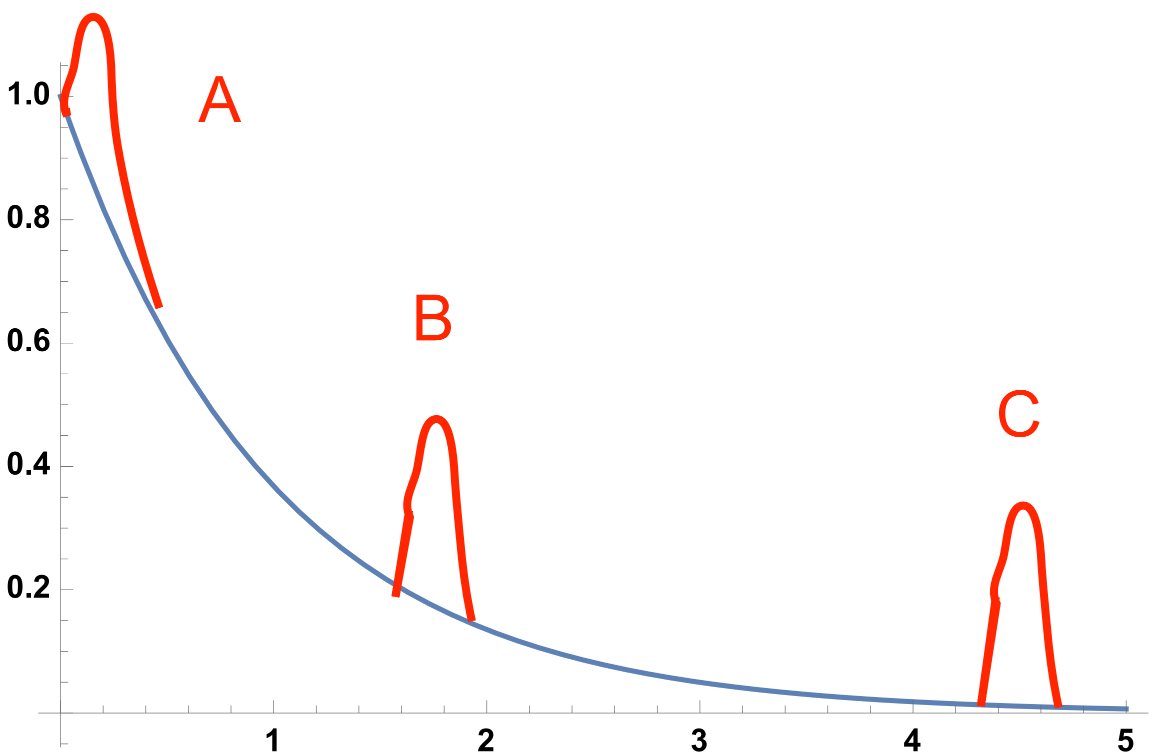
To find a way to address this problem, consider three cases A, B and C (given the fixed number of background and signal events): which ones suffer more from LEE?

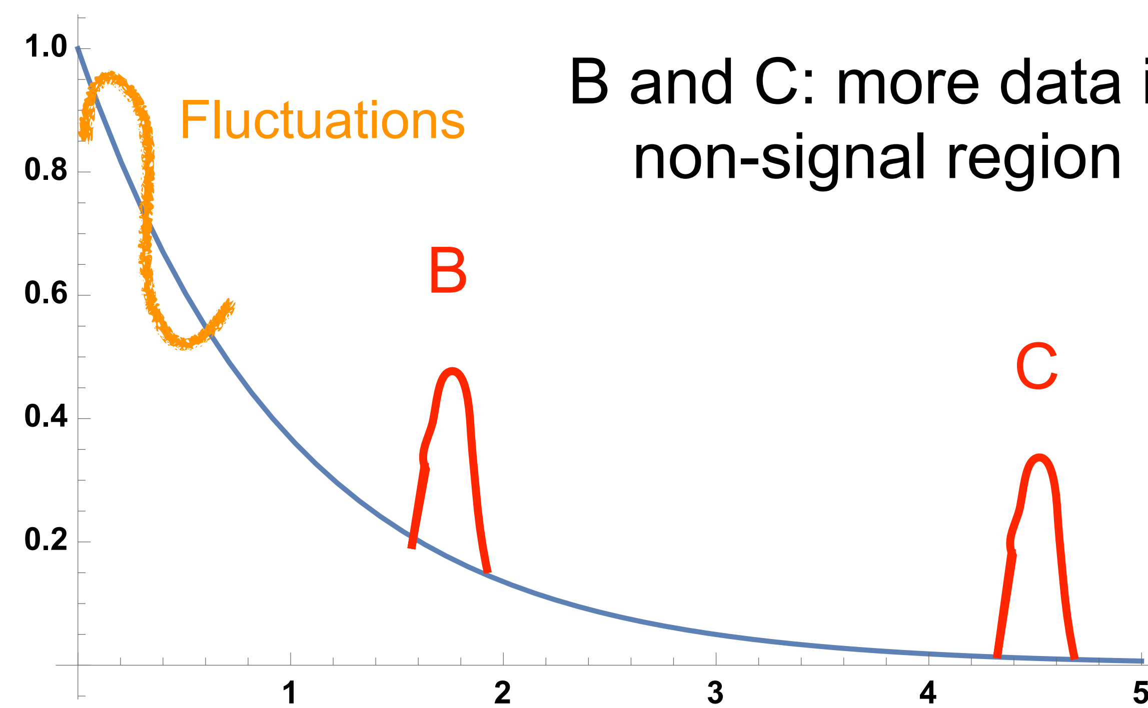


Strategy to Address Large LEE

The suppression of LEE by luminosity might not be sufficient if S/B is small.

To find a way to address this problem, consider three cases A, B and C (given the fixed number of background and signal events): which ones suffer more from LEE?



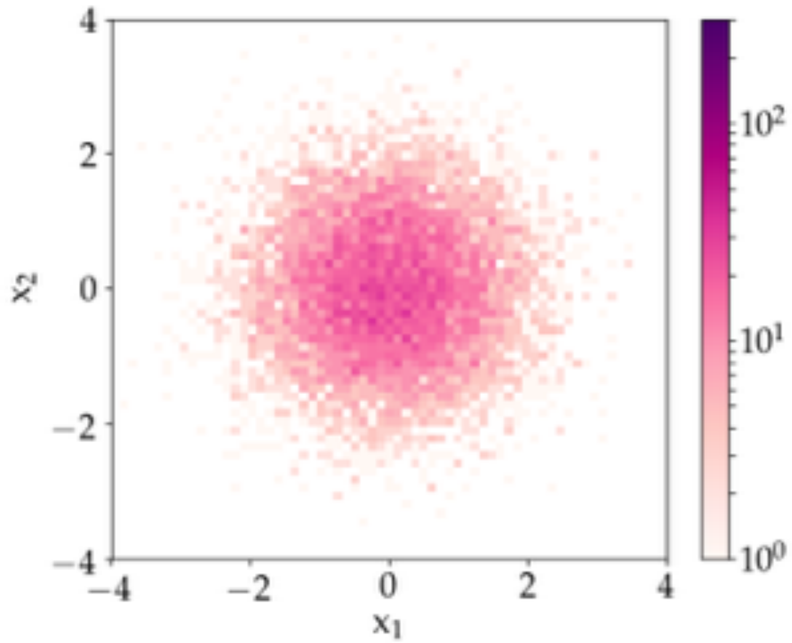


B and C: more data in non-signal region

To compensate for high-scoring (by O_new) of known-pattern data from high-density non-signal region

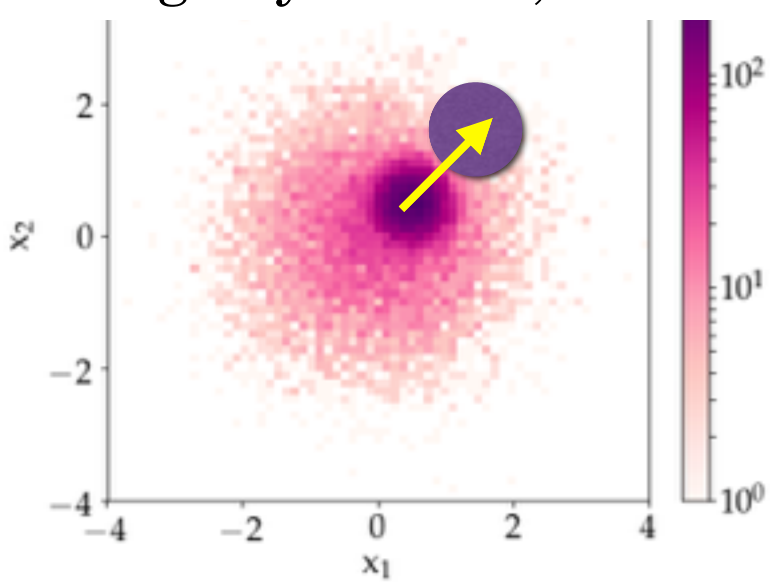
=> $$\mathcal{O}_{\rm comb} = \sqrt{\mathcal{O}_{\rm trad}\mathcal{O}_{\rm new}}$$

Strategy to Address Large LEE

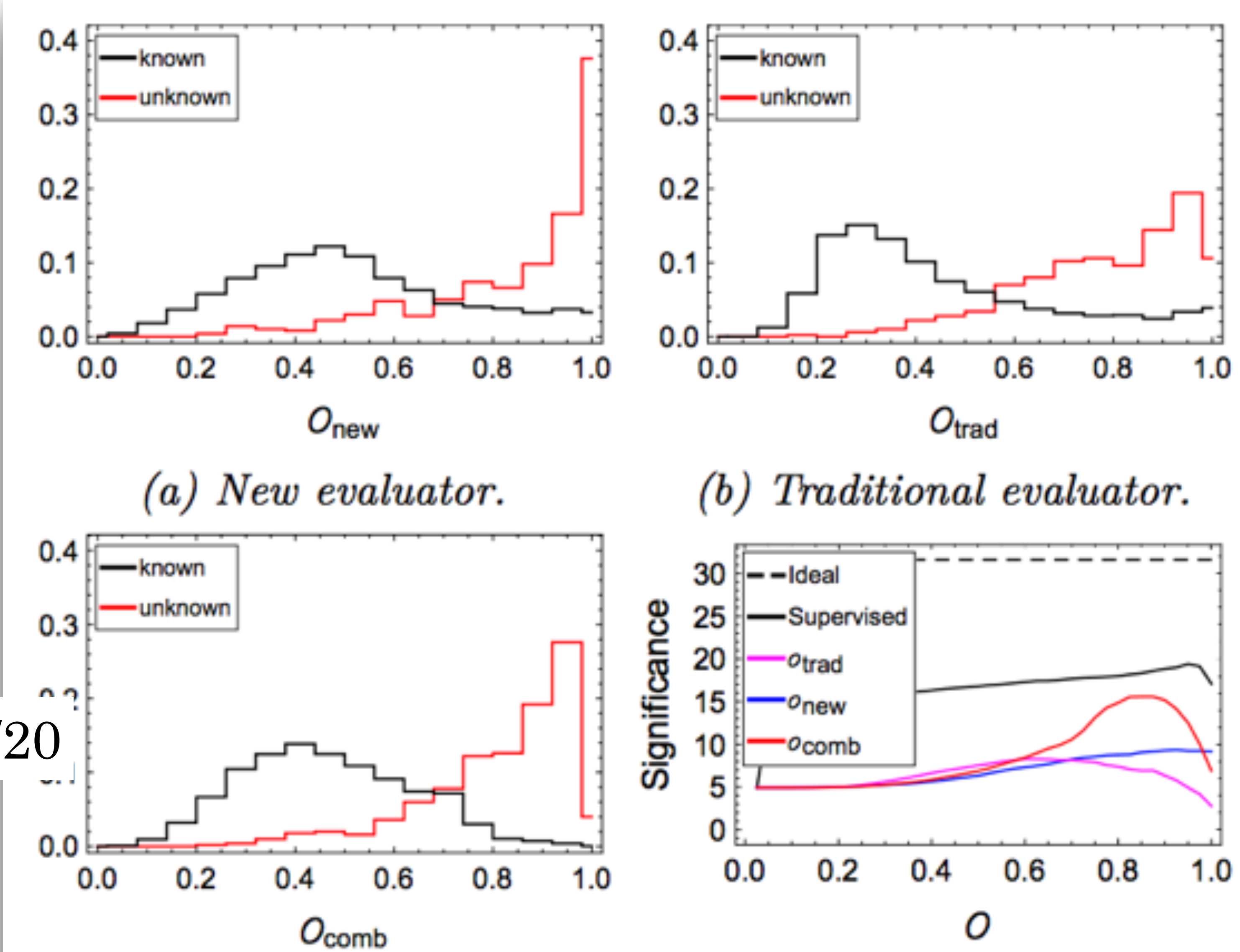


(a) Training data.

Center slightly shifted, with S/B = 1/20



(b) Testing data.



(a) New evaluator. *(b) Traditional evaluator.*

(c) Combined evaluator. *(d) Significance.*

O_comb based analysis yields more than 50% improvement in detection sensitivity!

Parton-level Benchmark Study

Analysis one: di-top (leptonic) production at LHC (the SM cross sections have been scaled by a factor 1/2000, for simplification)

- $pp \to \bar{t}_l t_l$, $\sigma = 11.5\,\mathrm{fb}$,
- $pp \to t_l \bar{b} W_l^{\pm}$, $\sigma = 0.365\,\mathrm{fb}$,
- $pp \to Z_b Z_l$, $\sigma = 0.0765\,\mathrm{fb}$.

$\mathbf{X_1}$: $pp \to \overline{T}T \to W_l^+ W_l^- \bar{b}b$

$\mathbf{X_2}$: $pp \to Z' \to \bar{t}t$

Analysis two: exotic Higgs decays at e+e- collider

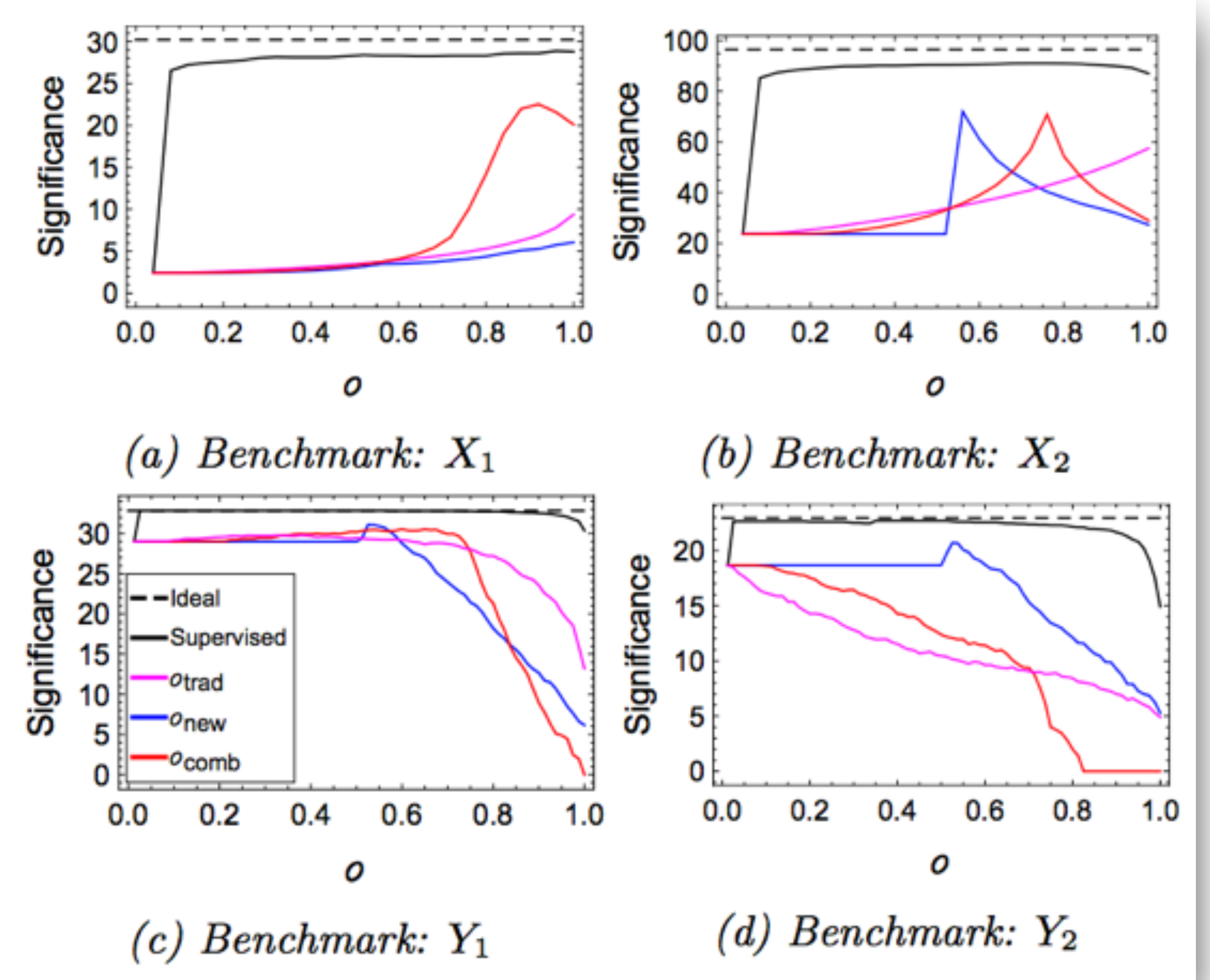
- $e^+e^- \to hZ \to Z^*_{\rm inv} Z_{\bar{b}b} l^+ l^-$ $\sigma = 0.00686\,\mathrm{fb}$,
- $e^+e^- \to hZ \to Z^*_{\bar{b}b} Z_{\rm inv} l^+ l^-$ $\sigma = 0.00259\,\mathrm{fb}$.

$\mathbf{Y_1}$: $h \to \tilde{\chi}_1 \tilde{\chi}_2 \to \tilde{\chi}_1 \tilde{\chi}_1 a$.

$\mathbf{Y_2}$: $h \to Za$

	Parameter values	$\sigma(fb)$
$X1$	$m_T = m_{\overline{T}}$ 1.2 TeV, BR$(T \to W_l^+ b) = 50\,\%$	0.152
$X2$	$m_{Z'} = 3\,\mathrm{TeV}$, $g_{Z'} = g_Z$, BR$(Z' \to \bar{t}t) = 16.7\,\%$	1.55
$Y1$	$m_{N_1} = \frac{m_{N_2}}{9} = \frac{m_a}{4} = 10\,\mathrm{GeV}$, BR$(h \to \bar{b}bE_T^{\rm miss}) = 1\,\%$	0.108
$Y2$	$m_a = 25\,\mathrm{GeV}$, BR$(h \to \bar{b}bE_T^{\rm miss}) = 1\,\%$	0.053

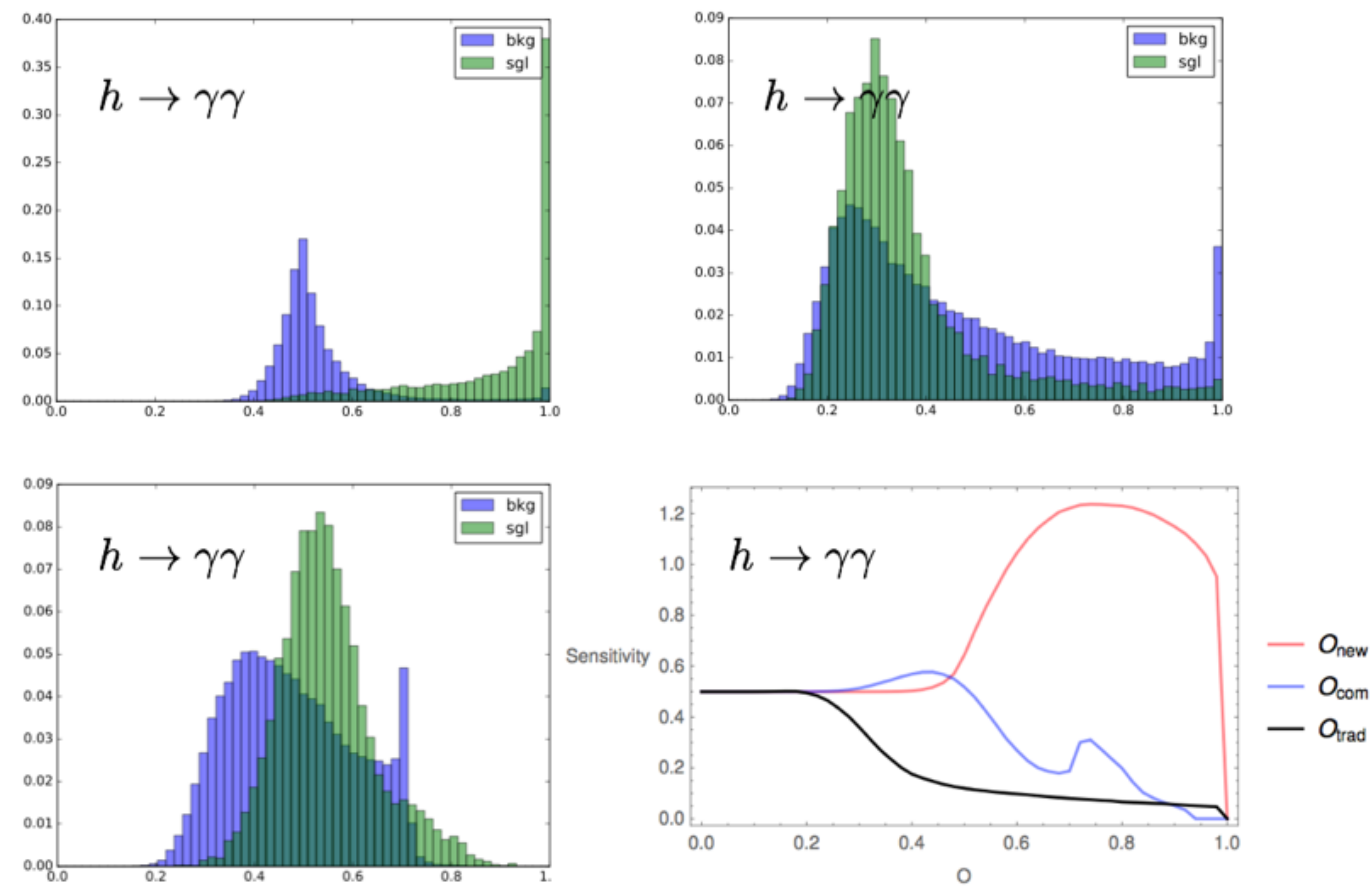
Parton-level Benchmark Study



(a) *Benchmark:* X_1 (b) *Benchmark:* X_2 (c) *Benchmark:* Y_1 (d) *Benchmark:* Y_2

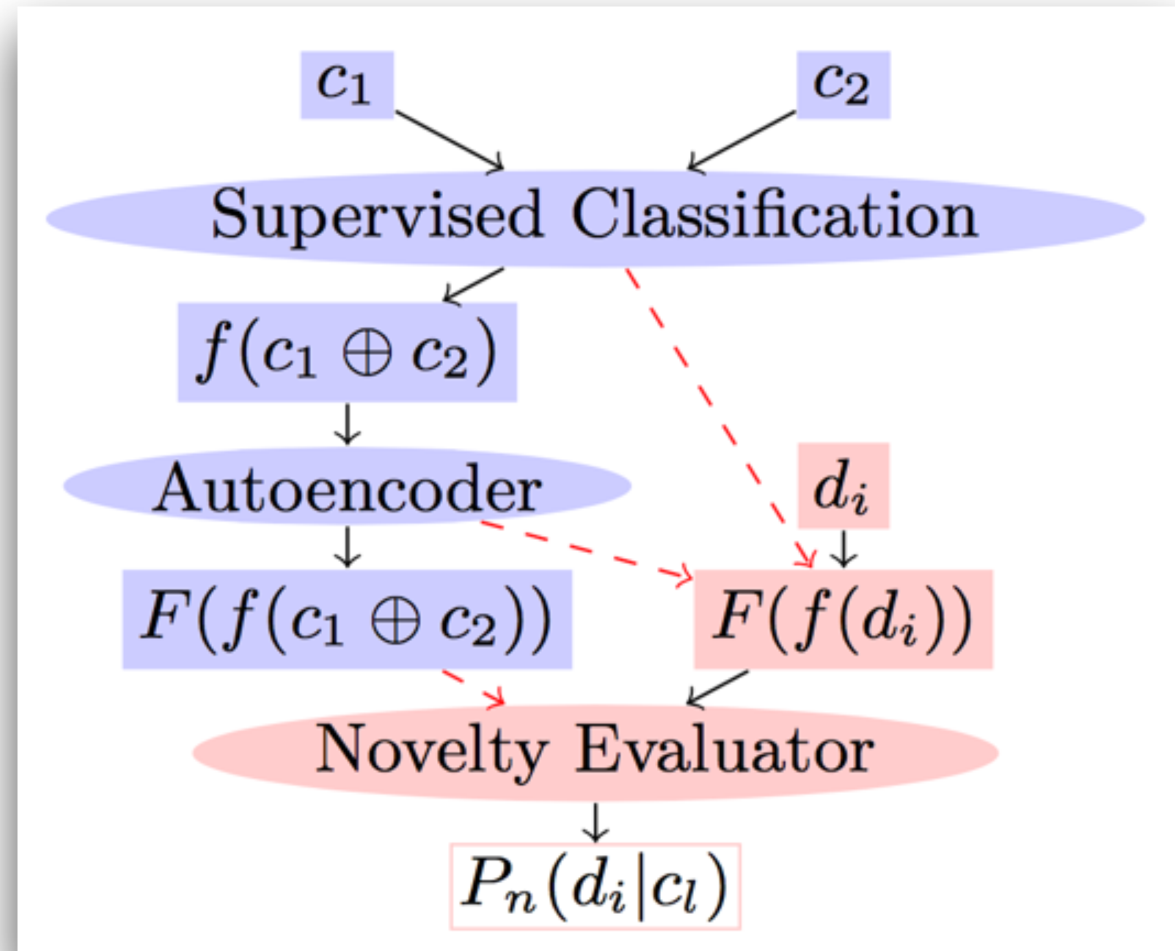
- X1: well-modeled by the 2D Gaussian sample!
- X2: O_comb less efficient due to one-order larger S/B
- X3 and X4: O_new performs universally better than the others, due to large S/B
- The sensitivities based on the algorithm designed are not far below the ones set by supervised learning

Rediscover Higgs Boson with Novelty Detection - tth



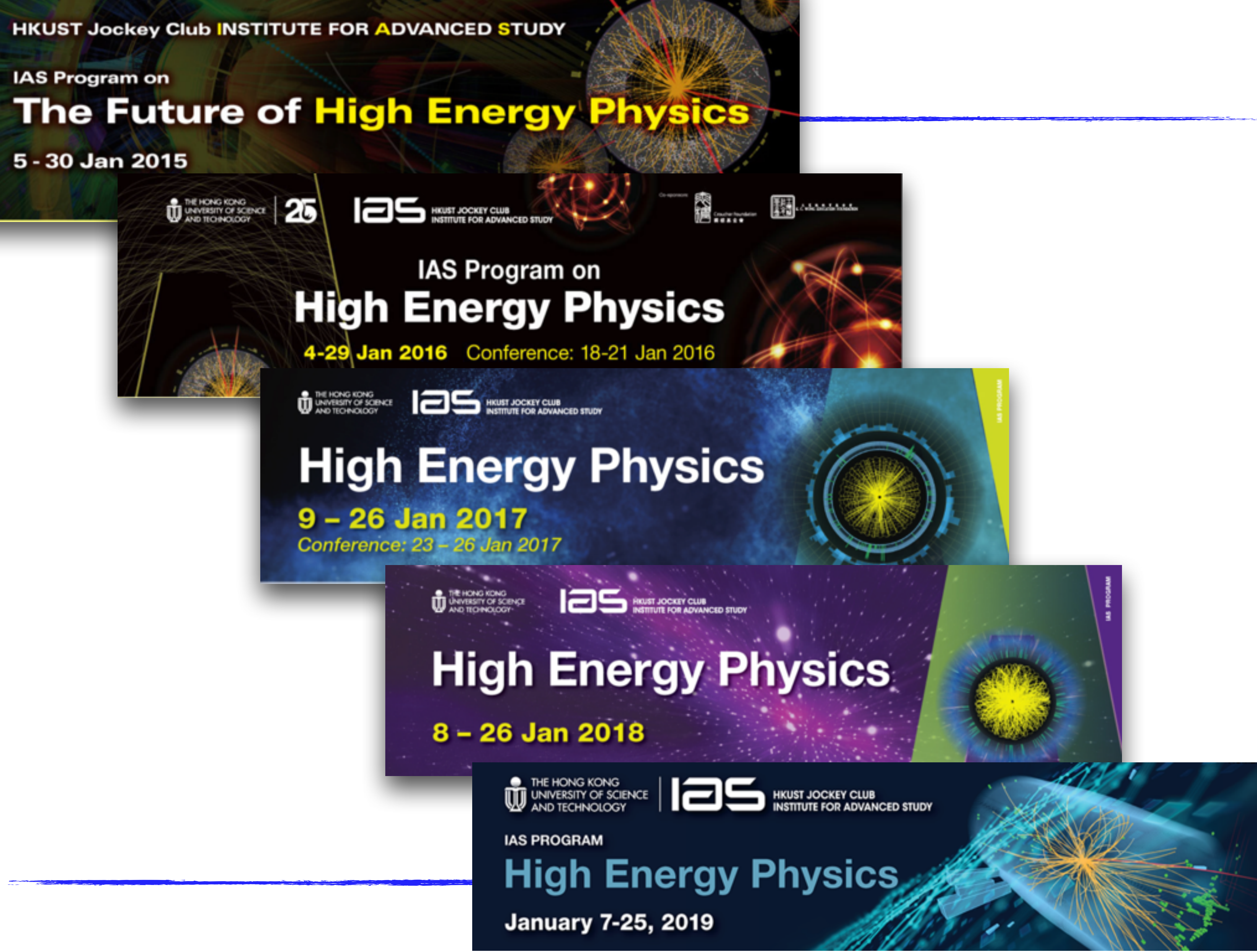
Background events: top pair with two real, one real and one faked, and two faked photons. Analyzed at hadron level, with 36/fb data for sensitivity.

Summary

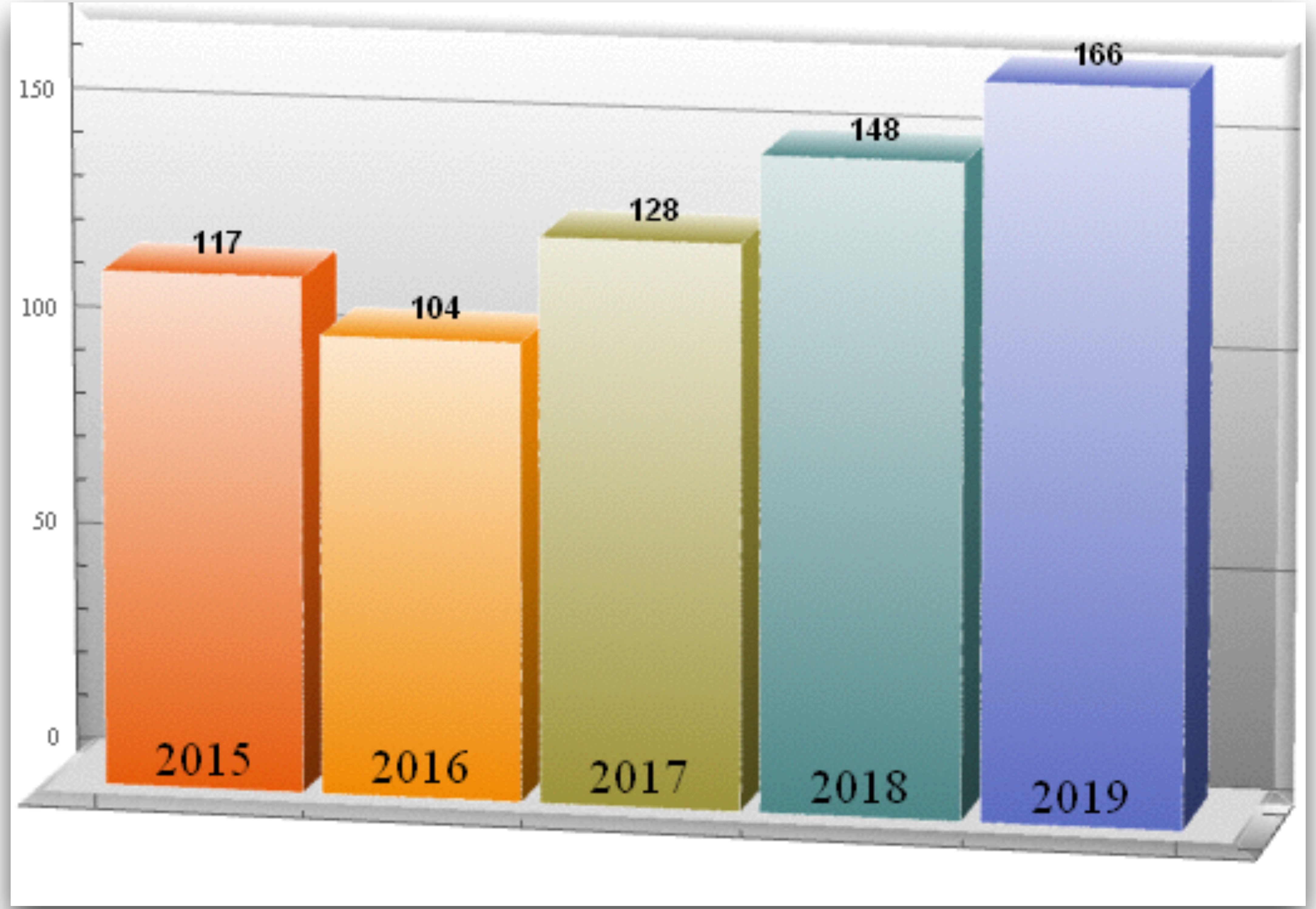


- Rapid development of the DNN techniques is bringing far-reaching impact for particle physics
- A combination of supervised learning and novelty detection may lay out the framework for future data analysis
- By properly designing novelty evaluators (clustering sensitive, LEE suppressed, etc.), encouragingly high sensitivity could be achieved in detecting unexpected BSM physics
- More efforts are needed to filling up the gap between proof of concept and real data analysis

HKIAS

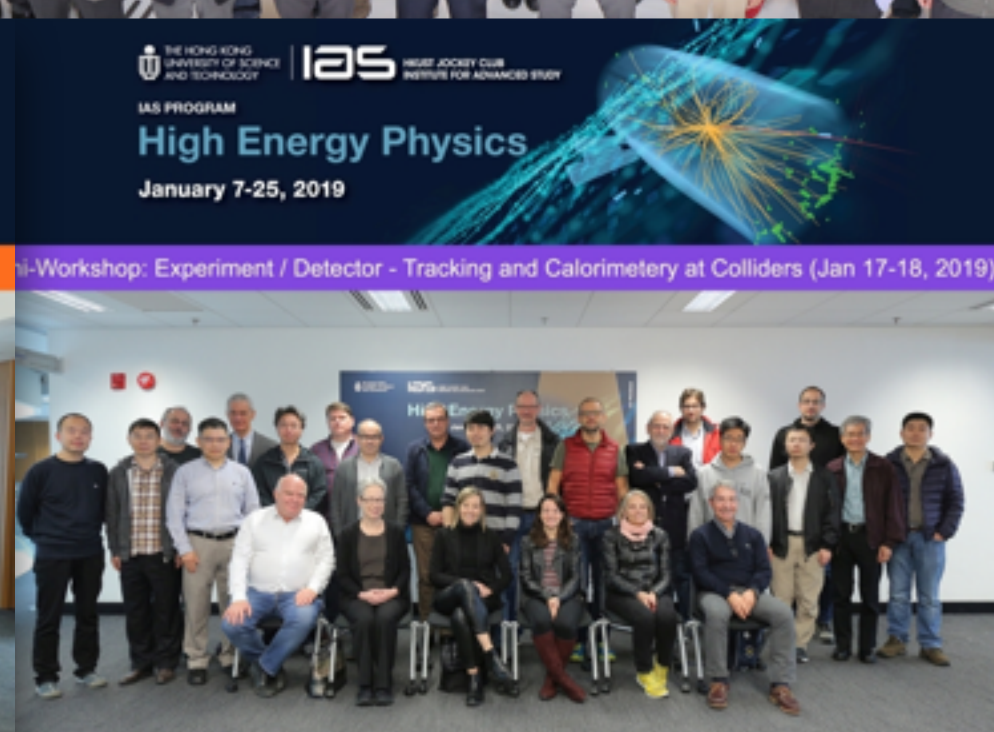

HKUST Jockey Club INSTITUTE FOR ADVANCED STUDY
IAS Program on
The Future of High Energy Physics
5 - 30 Jan 2015
THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY
ias HKUST JOCKEY CLUB INSTITUTE FOR ADVANCED STUDY
IAS Program on
High Energy Physics
4-29 Jan 2016
Conference: 18-21 Jan 2016
THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY
ias HKUST JOCKEY CLUB INSTITUTE FOR ADVANCED STUDY
High Energy Physics
9 – 26 Jan 2017
Conference: 23 – 26 Jan 2017
THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY
ias HKUST JOCKEY CLUB INSTITUTE FOR ADVANCED STUDY
High Energy Physics
8 – 26 Jan 2018
THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY
ias HKUST JOCKEY CLUB INSTITUTE FOR ADVANCED STUDY
IAS PROGRAM
High Energy Physics
January 7-25, 2019

Increasing Attendance




UNIVERSITY OF SCIENCE
AND TECHNOLOGY
ias
HKUST JOCKEY CLUB
INSTITUTE FOR ADVANCED STUDY
IAS PROGRAM
High Energy Physics
January 7-25, 2019
Conference Week (Jan 21-24, 2019)


ias
IAS PROGRAM
High Energy Physics
January 7-25, 2019
Mini-Workshop: Theory - Physics Opportunities and Advanced Tools (Jan 10-11, 2019)


ias
IAS PROGRAM
High Energy Physics
January 7-25, 2019
i-Workshop: Experiment / Detector - Tracking and Calorimetery at Colliders (Jan 17-18, 2019)


ias
IAS PROGRAM
High Energy Physics
January 7-25, 2019
i-Workshop: Accelerator - Beam Polarization in Future Colliders (Jan 17-18,

Conference Forum (2019 program)

2020 HKIAS Program on Future Colliders (Jan 6-24)

(Preliminary)

Week 1: Theory mini-workshop (Jan 9-10)

Week 2: CEPC detector/physics workshop (Jan 13-15)

Week 2: Exp/Detector mini-workshop (Jan 16-17)

Week 2: Accelerator physics mini-workshop (Jan 16-17)

Week 3: Conference week (Jan 20-23)

Conference chair: Tao Liu (taoliu@ust.hk)


See you in HK in Jan 2020!