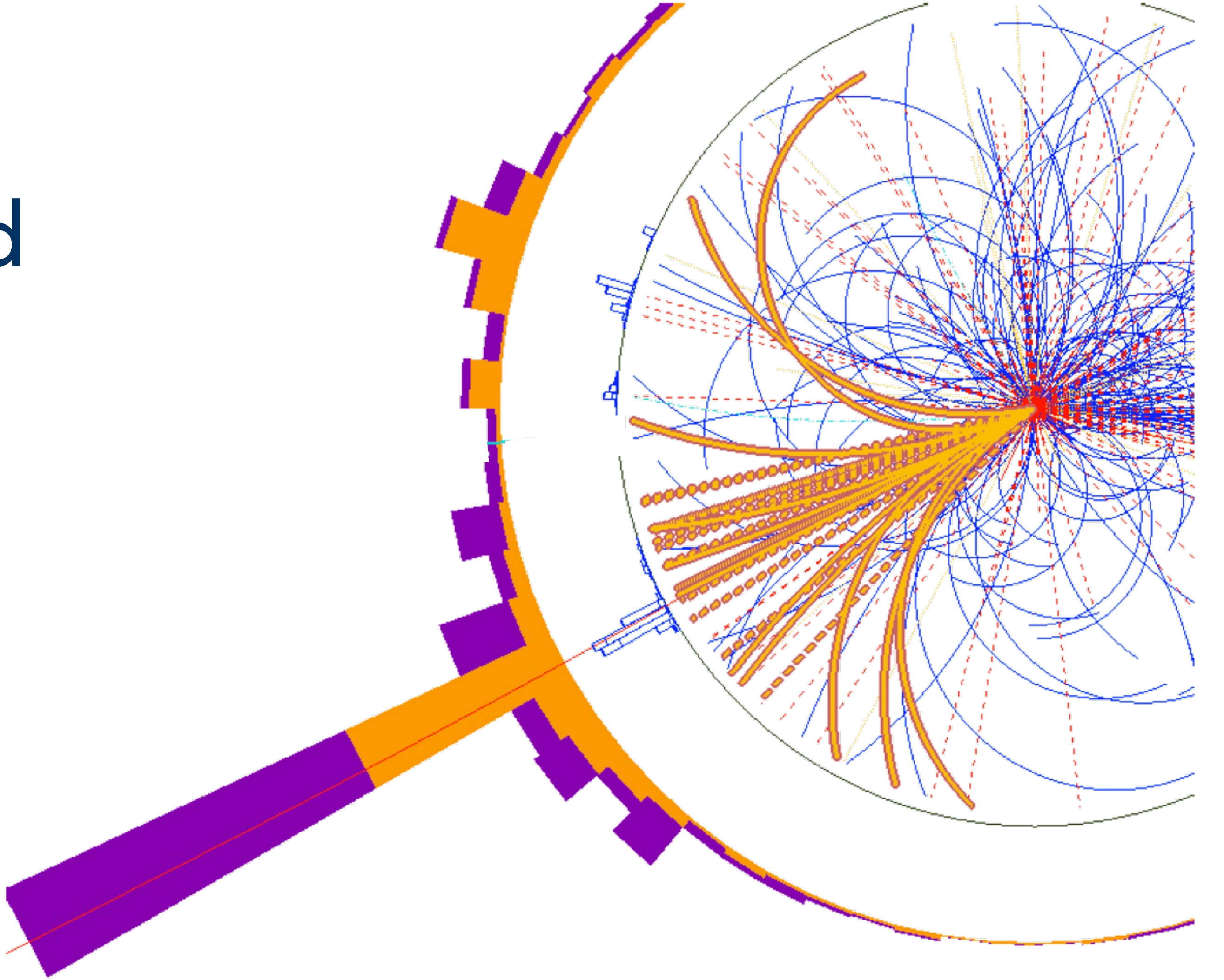
Perturbatively Regularized Neural Networks



Arianna Garcia Caffaro

Ian Moult

Chase Shimmin

Yale University

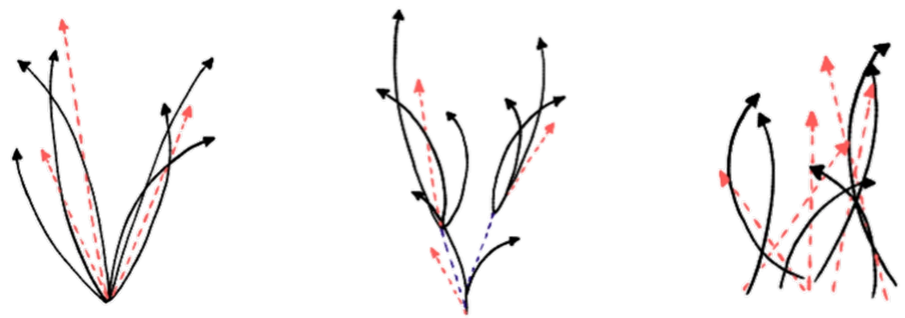
Jets Substructure

- Improve tagging for boosted objects
- Many successful observable measurements
- Gain insight into QCD

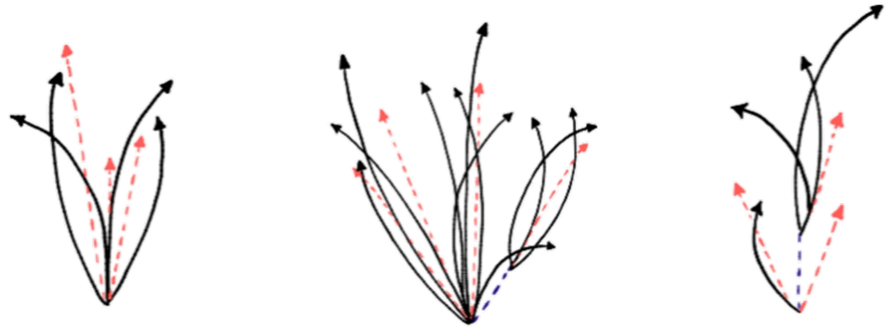
⇢ Done through NN jet taggers

Higher jet detail → Higher sensitivity to theoretical uncertainties *

✻ Rooted in modeling of non-perturbative processes: hadronization



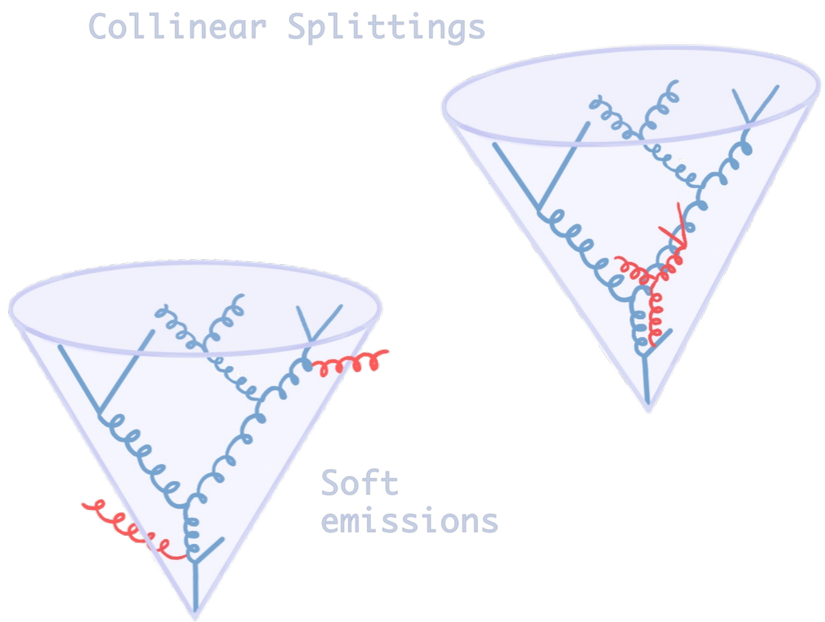
W or Z jet | Higgs jet | pileup



up, down, or strange jet | top jet | charm or bottom jet

To take full advantage of jet substructure measurements we must have a good handle on the uncertainties

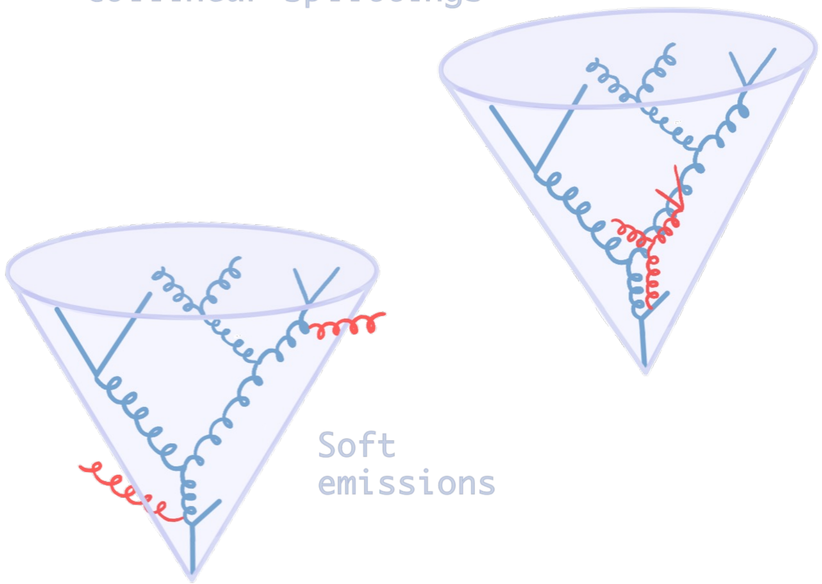
Energy Flow Networks (EFNs)



- Introduced by Komiske, Metodiev, Thaler
  - ⇢ To control non-perturbative corrections
- IRC safe
  - ⇢ Observable is unchanged under soft emissions and collinear splittings

$$EFN = F\left(\sum_{i}^{N} e_i \Phi\left(\hat{n}_i\right)\right)$$

Energy Flow Networks (EFNs)



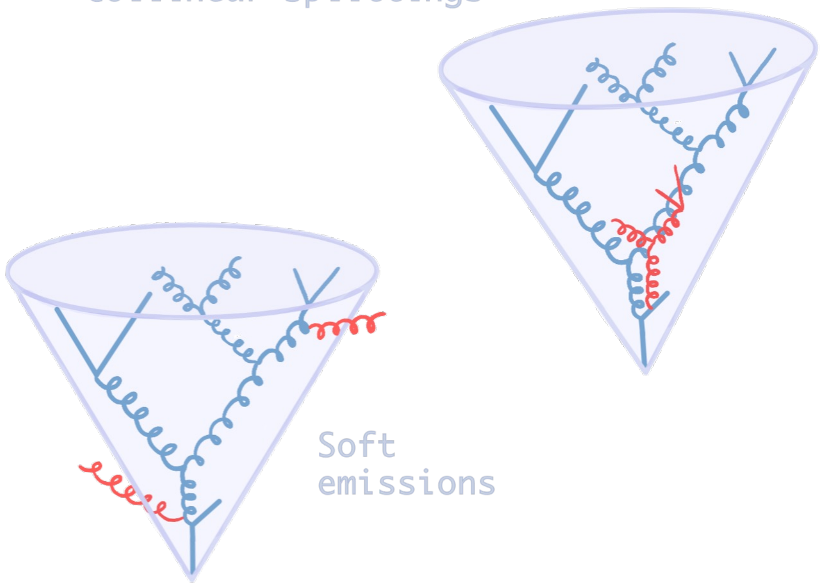
- Introduced by Komiske, Metodiev, Thaler
  - ⇢ To control non-perturbative corrections

- IRC safe
  - ⇢ Observable is unchanged under soft emissions and collinear splittings

$$EFN = F\left(\sum_{i}^{N} e_i \Phi\left(\hat{n}_i\right)\right)$$

- Comparatively poor performance

Energy Flow Networks (EFNs)



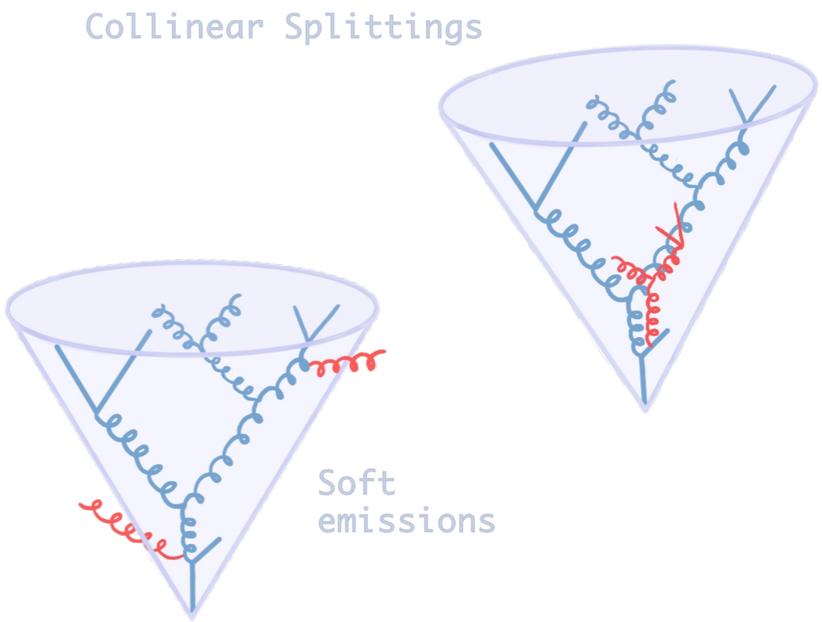
- Introduced by Komiske, Metodiev, Thaler
  - ⇢ To control non-perturbative corrections

- IRC safe
  - ⇢ Observable is unchanged under soft emissions and collinear splittings

$$EFN = F\left(\sum_{i}^{N} e_i \Phi\left(\hat{n}_i\right)\right)$$

- Comparatively poor performance
- IRC benefits?

Energy Flow Networks (EFNs)



- Introduced by Komiske, Metodiev, Thaler
  ⇢ To control non-perturbative corrections

- IRC safe
  ⇢ Observable is unchanged under soft emissions and collinear splittings

$$EFN = F\left(\sum_{i}^{N} e_i \Phi\left(\hat{n}_i\right)\right)$$

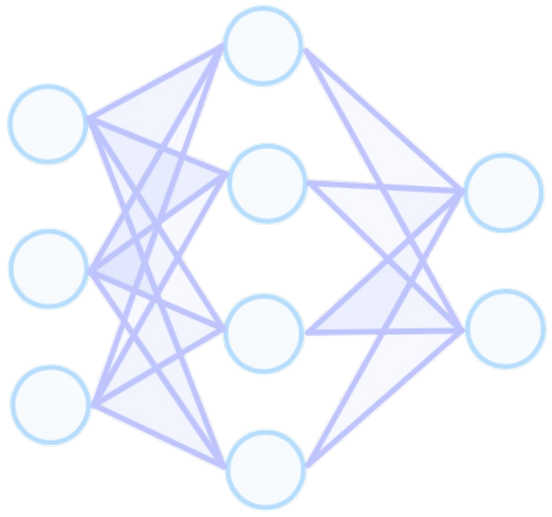
- Comparatively poor performance
- IRC benefits?

⁕ Introducing better performing IRC safe NN

⁕ Quantifying benefits IRC safety

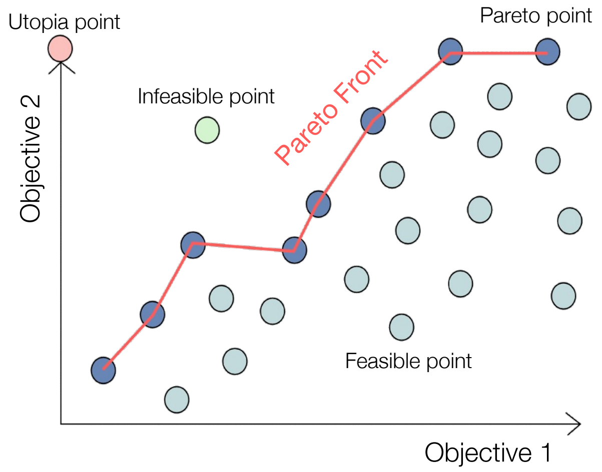
Outline

Beyond the EFN



- EFN and IRC safety
- E2FN architecture and features
- Performance Studies

Quantifying perturbative regularization



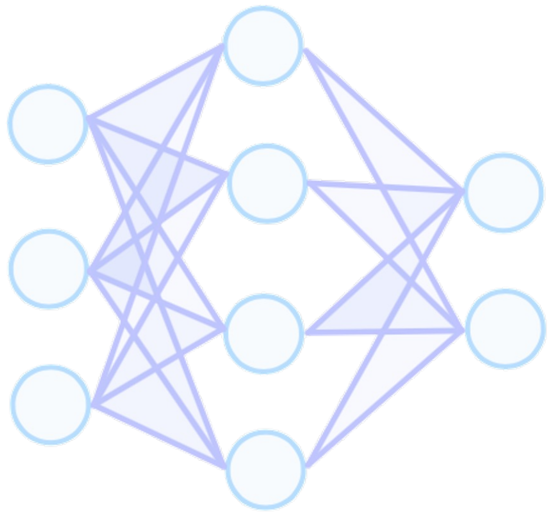
- Introduction to perturbative regularization
- Overview of study
- Pareto Front

A Toy Study

- NNs significance comparison on toy boosted Z boson search
- Effects of theoretical uncertainties on significance

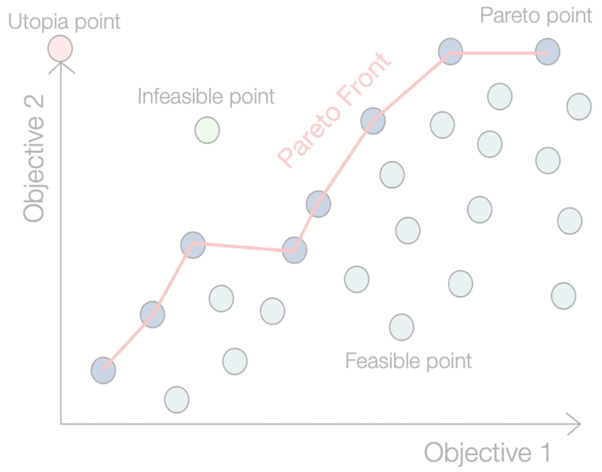
Outline

Beyond the EFN



- EFN and IRC safety
- E2FN architecture and features
- Performance Studies

Quantifying perturbative regularization



- Introduction to perturbative regularization
- Overview of study
- Pareto Plots

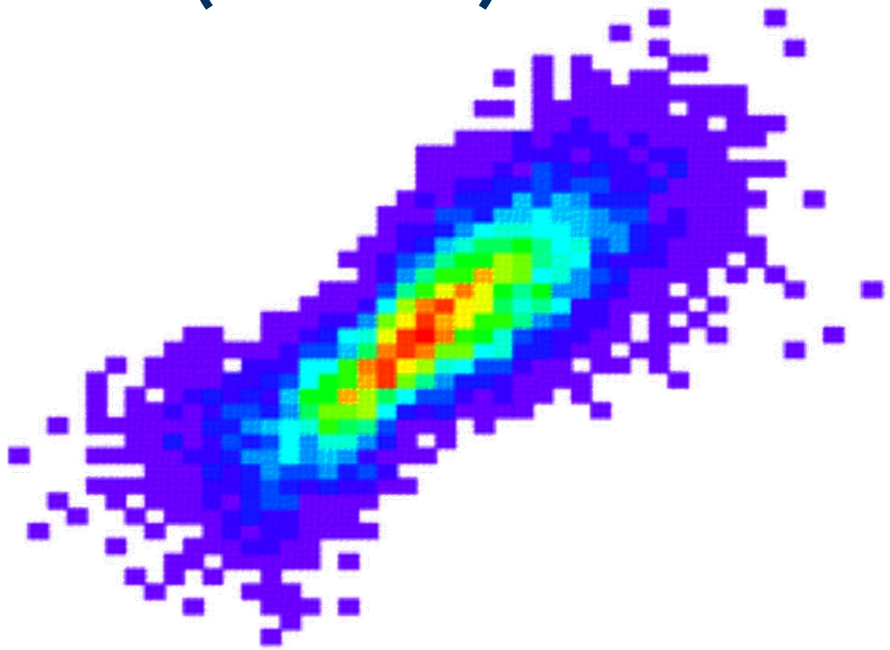
A Toy Study

- NNs significance comparison on toy boosted Z boson search
- Effects of theoretical uncertainties on significance

Introducing Energy-Energy Flow Networks (E2FNs)

$$EFN = F\left(\sum_{i}^{N} e_i \Phi\left(\hat{n}_i\right)\right)$$

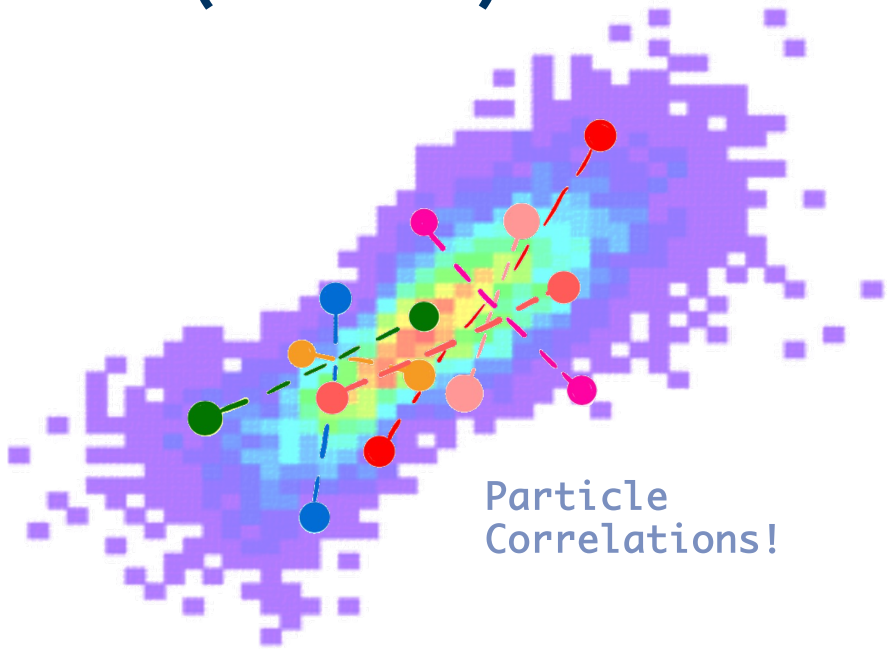
- Comparatively poor performance



Introducing Energy-Energy Flow Networks (E2FNs)

$$EFN = F\left(\sum_{i}^{N} e_i \Phi\left(\hat{n}_i\right)\right)$$

- Comparatively poor performance

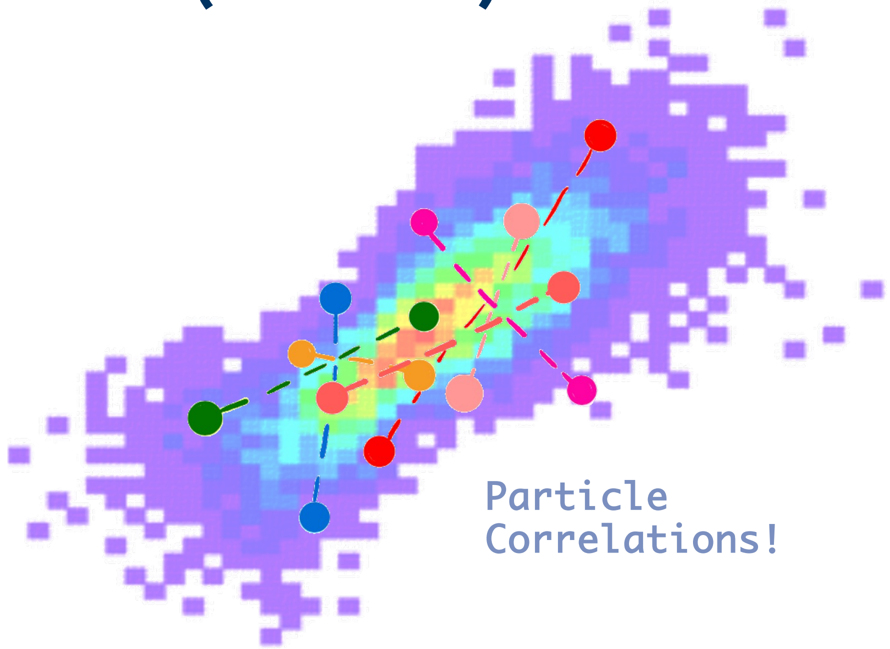


Introducing Energy-Energy Flow Networks (E2FNs)

$$EFN = F\left(\sum_{i}^{N} e_i \Phi\left(\hat{n}_i\right)\right)$$

- Comparatively poor performance

$$E2FN = F\left(\sum_{i,j}^{N} e_i e_j \Phi\left(\hat{n}_i, \hat{n}_j\right)\right)$$

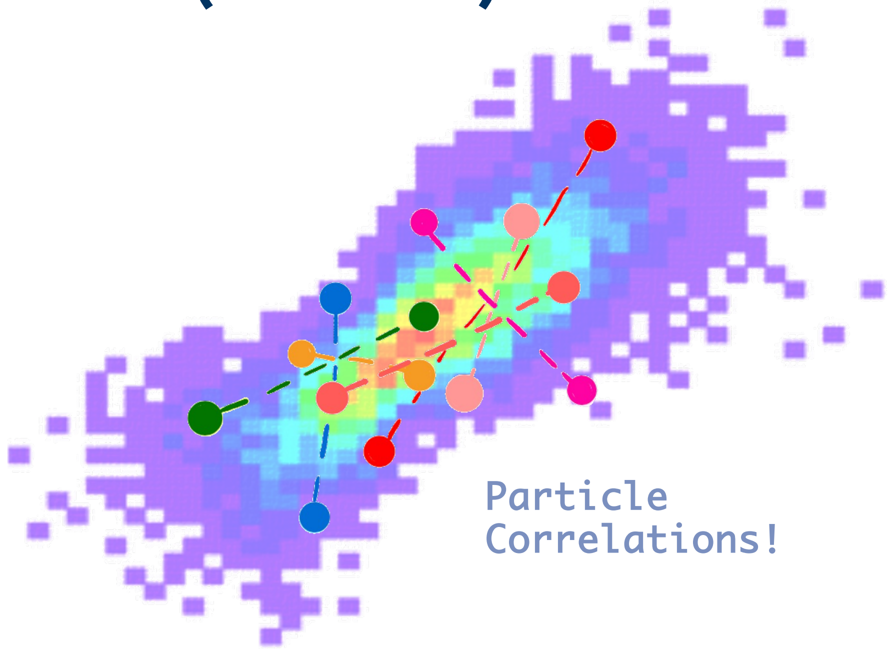


Introducing Energy-Energy Flow Networks (E2FNs)

$$EFN = F\left(\sum_{i}^{N} e_i \Phi\left(\hat{n}_i\right)\right)$$

- Comparatively poor performance

$$E2FN = F\left(\sum_{i,j}^{N} e_i e_j \Phi\left(\hat{n}_i, \hat{n}_j\right)\right)$$



Inspiration:

- Tkachov's work on C-continuous observables
  - ⇢ Energy Correlator (n=1) is analogous to an EFN
  - ⇢ Set n=2 (EEC) and make F a NN ⟶ E2FN

Energy correlators 3.39

These have the form

$$f(\mathbf{P}) = \sum_{a_1 \ldots a_n} E_{a_1} \ldots E_{a_n} f_n(\hat{\boldsymbol{p}}_{a_1}, \ldots, \hat{\boldsymbol{p}}_{a_n}) \,, \qquad 3.40$$

where f_n is a symmetric continuous function of n arguments. Basic shape observables are special cases corresponding to $n = 1$.

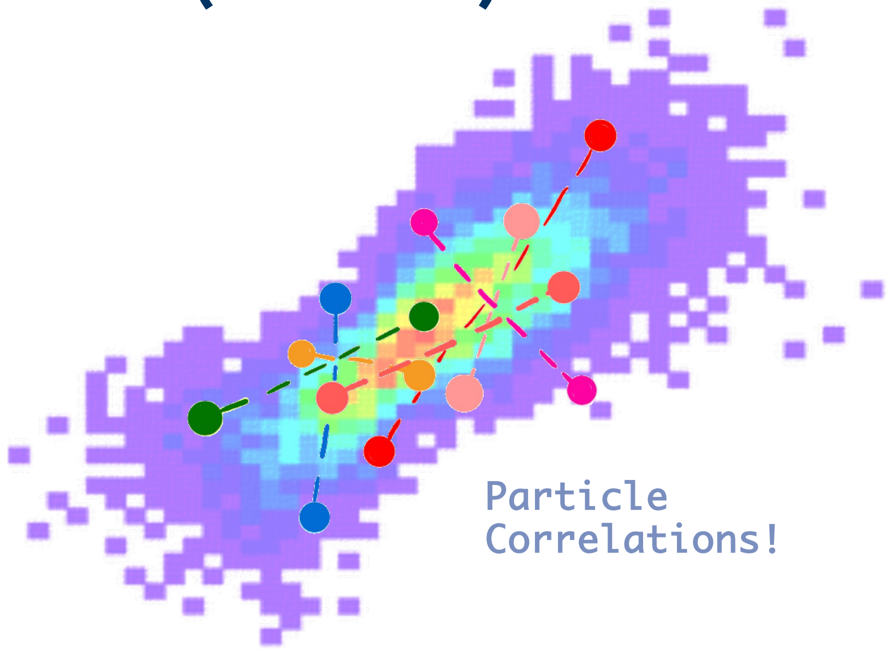
Introducing Energy-Energy Flow Networks (E2FNs)

$$EFN = F\left(\sum_{i}^{N} e_i \Phi\left(\hat{n}_i\right)\right)$$

- Comparatively poor performance

$$E2FN = F\left(\sum_{i,j}^{N} e_i e_j \Phi\left(\hat{n}_i, \hat{n}_j\right)\right)$$

$$P2FN = F\left(\sum_{ij}^{N} \Phi\left(p_i, p_j\right)\right)$$



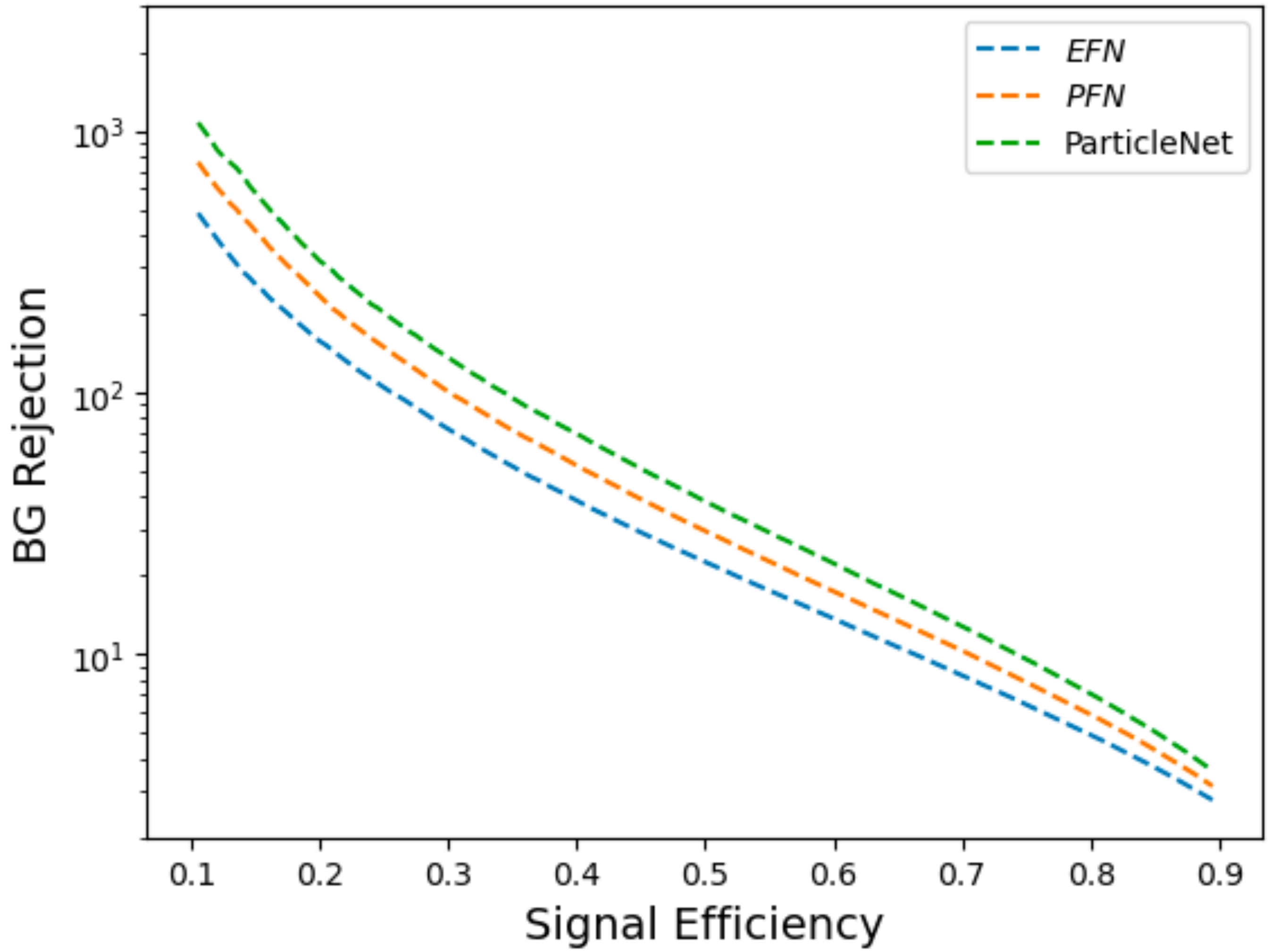
Energy correlators 3.39

These have the form

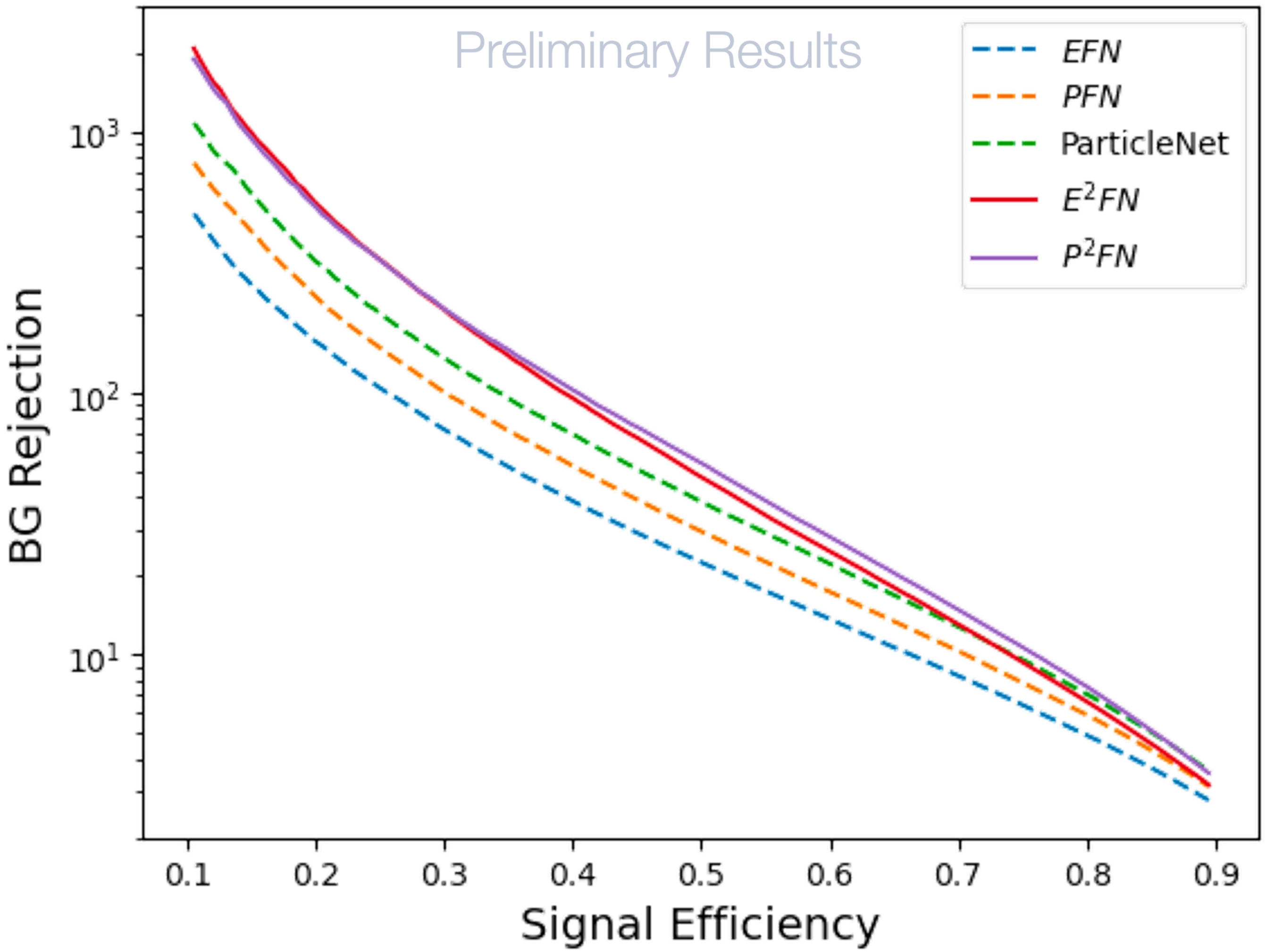
$$f(\mathbf{P}) = \sum_{a_1 \ldots a_n} E_{a_1} \ldots E_{a_n} f_n(\hat{\boldsymbol{p}}_{a_1}, \ldots, \hat{\boldsymbol{p}}_{a_n}), \qquad 3.40$$

where f_n is a symmetric continuous function of n arguments. Basic shape observables are special cases corresponding to $n = 1$.

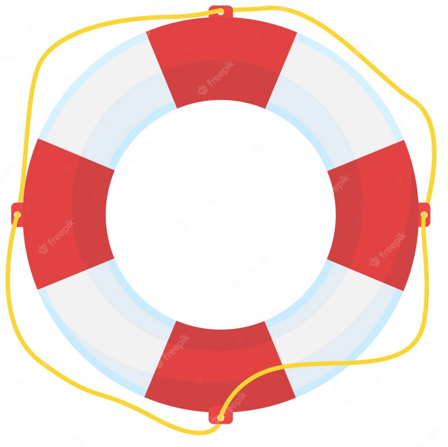
Raw Performance



Raw Performance



IRC Safety in E2FNs



- IRC safety conditions:
  1. Φ network is continuous
  2. $\Phi \rightarrow 0$ as the two studied particles become collinear

$$E2FN = F\left(\sum_{i,j}^{N} e_i e_j \Phi\left(\hat{n}_i, \hat{n}_j\right)\right)$$

IRC Safety in E2FNs

- IRC safety conditions:
  1. Φ network is continuous
  2. $\Phi \to 0$ as the two studied particles become collinear

$$E2FN = F\left(\sum_{i,j}^{N} e_i e_j \Phi\left(\hat{n}_i, \hat{n}_j\right)\right)$$

✻ IRC safety is a good first step but not enough!

Non-perturbative corrections of an IRC safe object can be arbitrarily large!

IRC Safety in E2FNs

- IRC safety conditions:
  1. Φ network is continuous
  2. $\Phi \to 0$ as the two studied particles become collinear

$$E2FN = F\left(\sum_{i,j}^{N} e_i e_j \Phi\left(\hat{n}_i, \hat{n}_j\right)\right)$$

✻ IRC safety is a good first step but not enough!

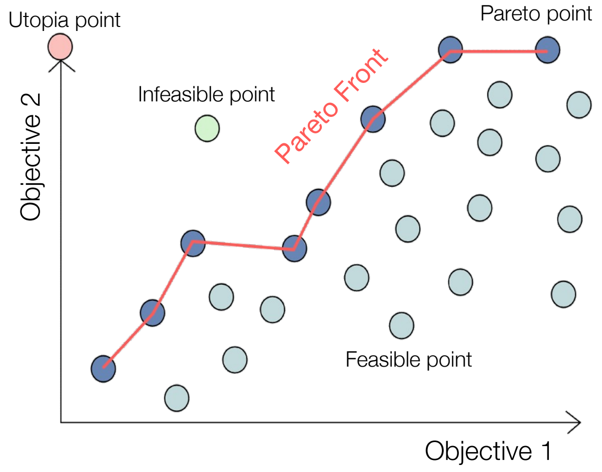
Non-perturbative corrections of an IRC safe object can be arbitrarily large!

→ Quantitative measurement of robustness of NN to hadronization effects

Beyond the EFN

- EFN and IRC safety
- E2FN architecture and features
- Performance Studies

Quantifying perturbative regularization



- Introduction to perturbative regularization
- Overview of study
- Pareto Plots

A Toy Study

- NNs significance comparison on toy boosted Z boson search
- Effects of theoretical uncertainties on significance

Perturbatively Regularized NNs

- Quantify the robustness of a NN to non-perturbative information:
  - ⇢ How much NN's output changes when non-perturbative info is present in the input

Perturbatively Regularized NNs

- Quantify the robustness of a NN to non-perturbative information:
  - ⇢ How much NN's output changes when non-perturbative info is present in the input

arXiv:1803.07977v1 [hep-ph] 21 Mar 2018

Les Houches 2017: Physics at TeV Colliders Standard Model Working Group Report

2 Performance versus robustness: Two-prong substructure taggers for the LHC [14]

Perturbatively Regularized NNs

- Quantify the robustness of a NN to non-perturbative information:
  - ⇢ How much NN's output changes when non-perturbative info is present in the input

"I envisioned something a little more... robust... I guess."

arXiv:1803.07977v1 [hep-ph] 21 Mar 2018

Les Houches 2017: Physics at TeV Colliders Standard Model Working Group Report

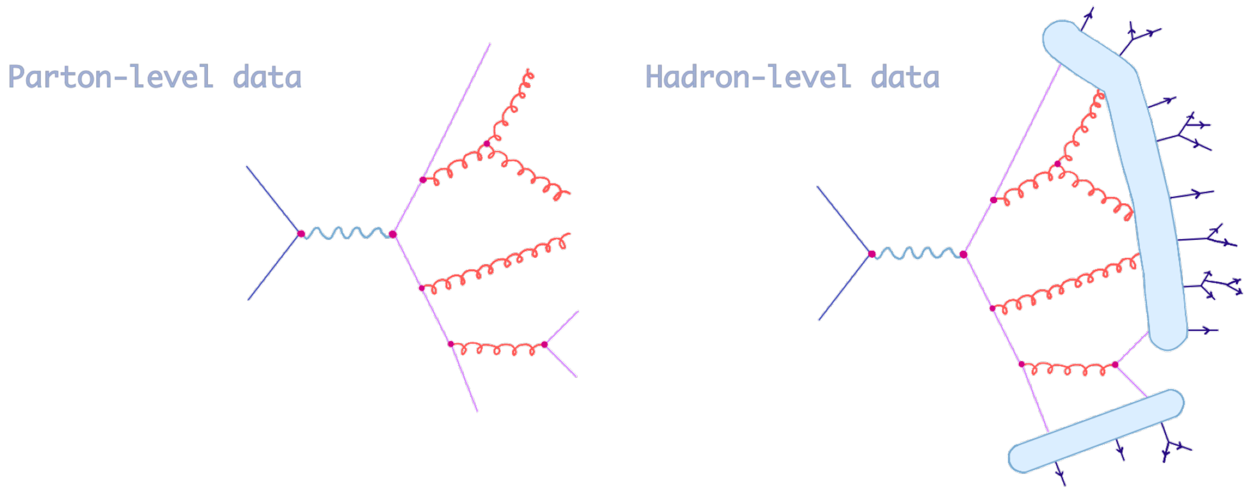
2 Performance versus robustness: Two-prong substructure taggers for the LHC 14

Control the type of information the NNs are sensitive to &

restrict their access to non-perturbative information

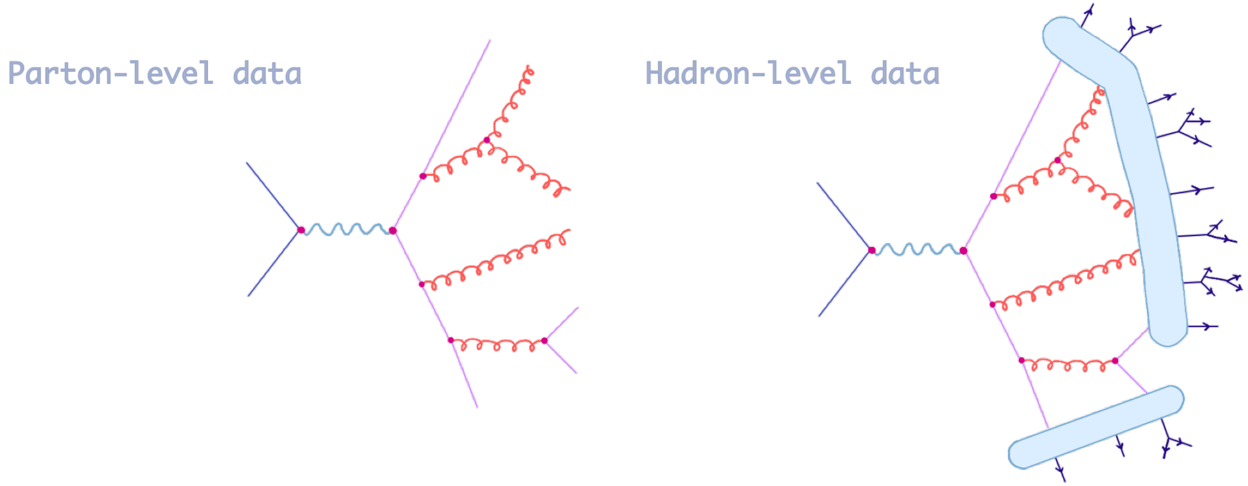
Test Data

- PYTHIA data
  - ⇢ Each event generated once
  - ⇢ Data collected twice per event:
    - ⇢ Parton-level
    - ⇢ Hadron-level



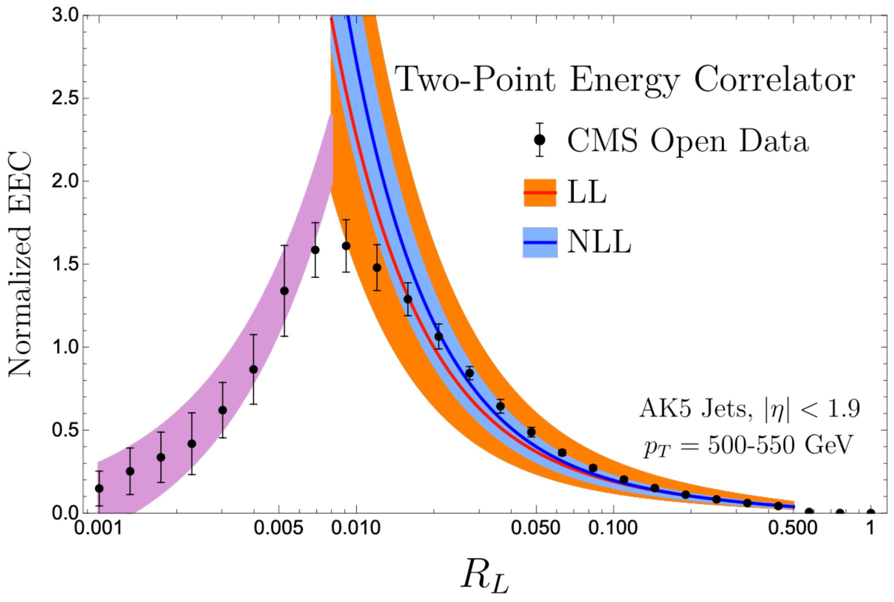
Test Data

- PYTHIA data
  - ⇢ Each event generated once
  - ⇢ Data collected twice per event:
    - ⇢ Parton-level
    - ⇢ Hadron-level



- Data pre-processing:
  - ⇢ Eta, phi coordinates centered to jet's eta, phi
  - ⇢ Constituent's p_T normalized to jet p_T
  - ⇢ Exclusive kt reclustering by momentum splitting

$$k_T = {p_T}_{jet} \times \Delta\Theta$$

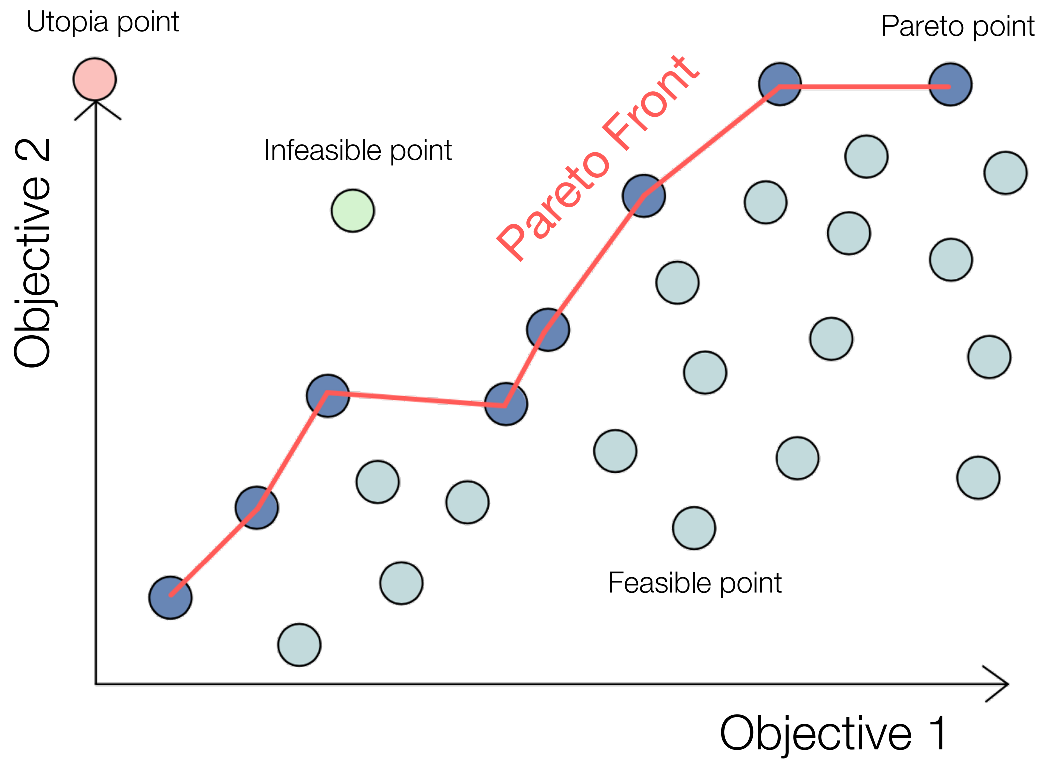


Regularization Metric & Pareto Plots

- Train on hadron-level data, test on parton-level data

- RMS as regularization metric:

$$RMS = \sqrt{\frac{\sum_{i}^{N} \Big(P_i(x \mid \textbf{parton-data}) - P_i(x \mid \textbf{hadron-data}) \Big)^2}{N}}$$



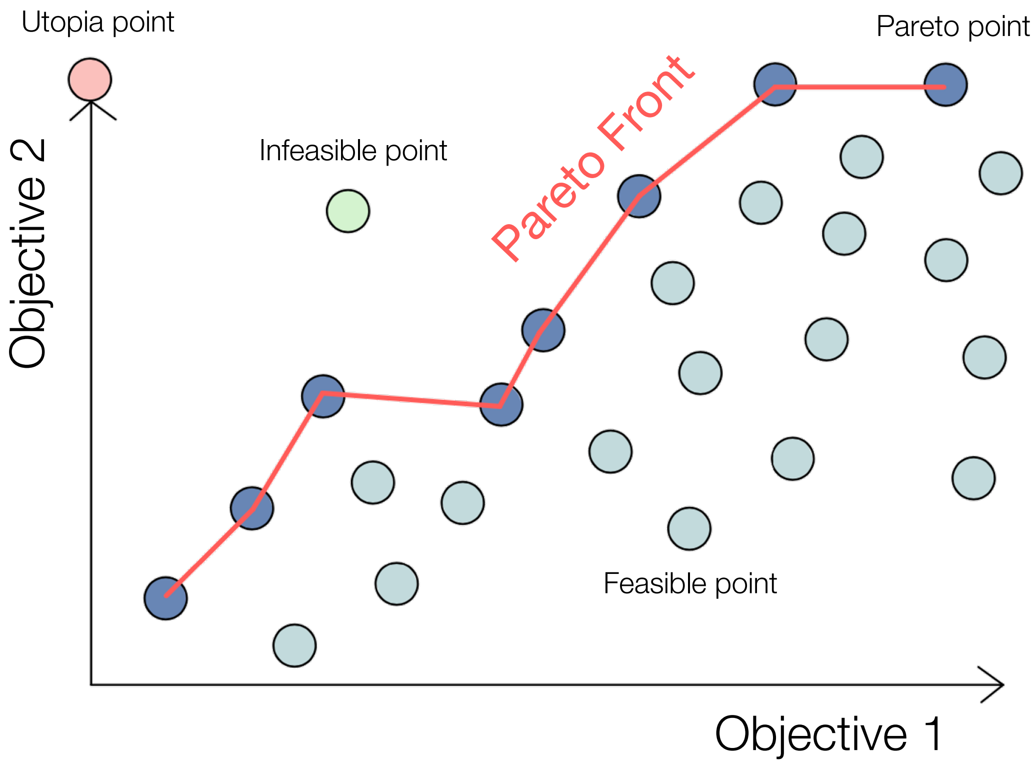
Regularization Metric & Pareto Plots

- Train on hadron-level data, test on parton-level data

- RMS as regularization metric:

$$RMS = \sqrt{\frac{\sum_{i}^{N}\left(P_i(x \mid \textbf{parton-data}) - P_i(x \mid \textbf{hadron-data})\right)^2}{N}}$$

- Objectives:
  - ⇢ Better performance: higher rejection
  - ⇢ Better regularization: lower RMS



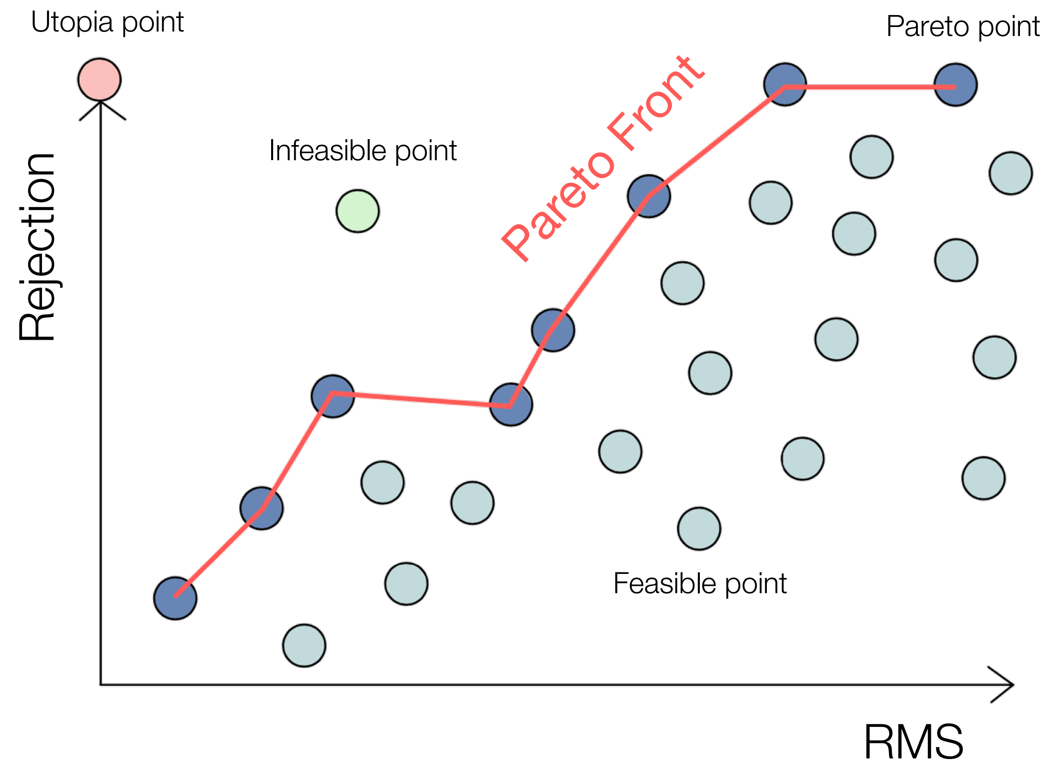
Regularization Metric & Pareto Plots

- Train on hadron-level data, test on parton-level data

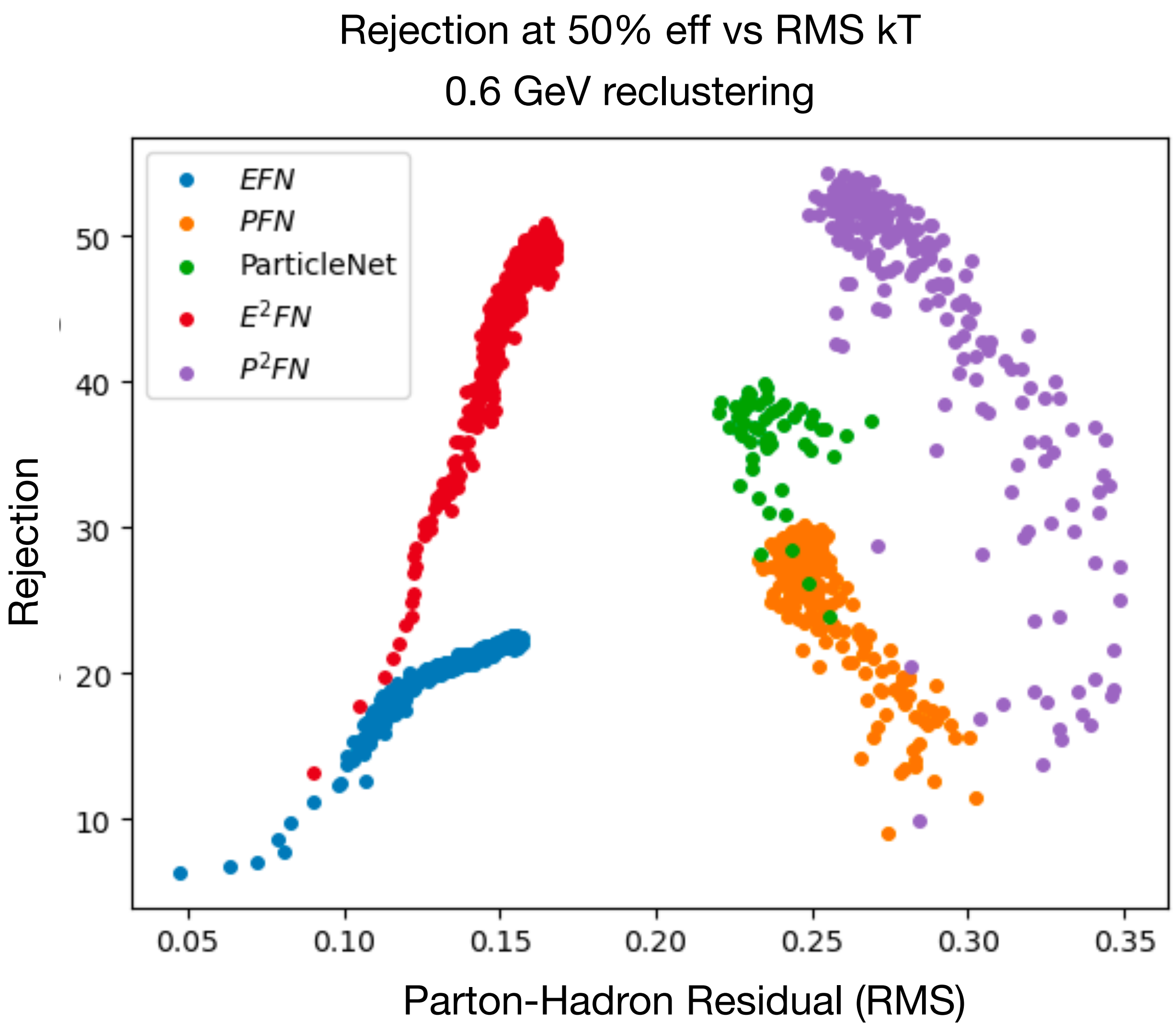
- RMS as regularization metric:

$$RMS = \sqrt{\frac{\sum_{i}^{N} \left(P_i(x \mid \textbf{parton-data}) - P_i(x \mid \textbf{hadron-data}) \right)^2}{N}}$$

- Objectives:
  - ⇢ Better performance: higher rejection
  - ⇢ Better regularization: lower RMS



Regularization vs. Performance



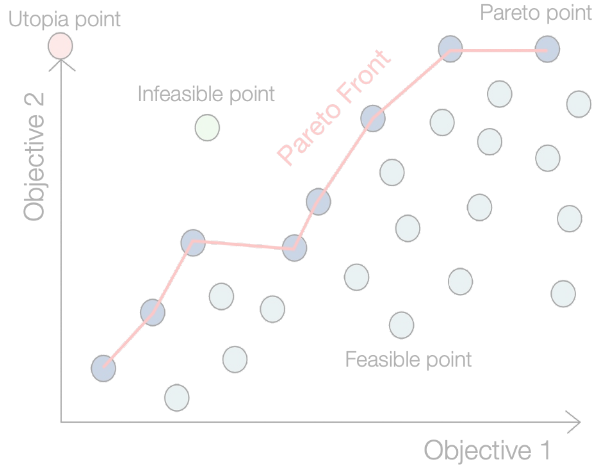
- Pareto Front is dominated by EFN, E2FN at low RMS and high rejection
- P2FN contribution at higher RMS and rejection

To understand the effect of non-perturbative uncertainties on searches we want to relate RMS metric back to a physical quantity

Beyond the EFN

- EFN and IRC safety
- E2FN architecture and features
- Performance Studies

Quantifying perturbative regularization



- Introduction to perturbative regularization
- Overview of study
- Pareto Plots

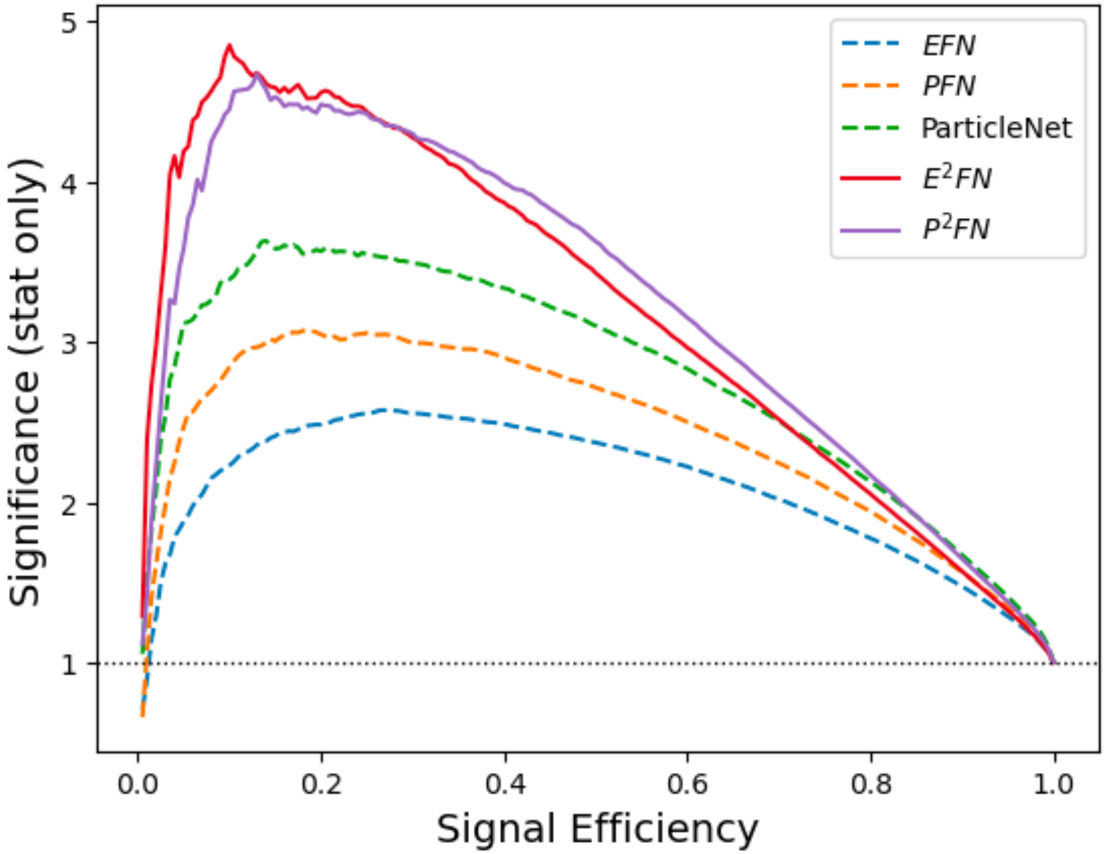
A Toy Study

- NNs significance comparison on toy boosted Z boson search
- Effects of theoretical uncertainties on significance

Raw Discovery Significance in Toy Search

- Tagging boosted Z Bosons in QCD background
- NN represents the entire analysis work flow
  - Cut on NN signal score
  - Single bin counting experiment

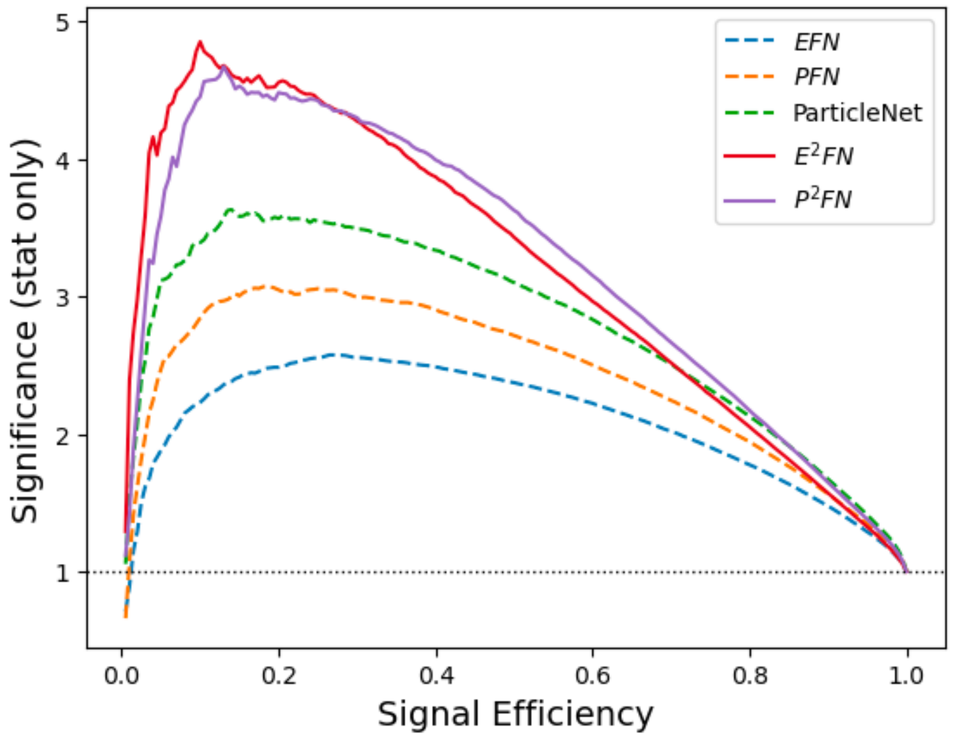
Optimize analysis for best discovery significance



Significance with Theoretical Uncertainties

- Background Uncertainties:

$$\frac{1}{10} * [B(\text{hadron}) - B(\text{parton})]$$

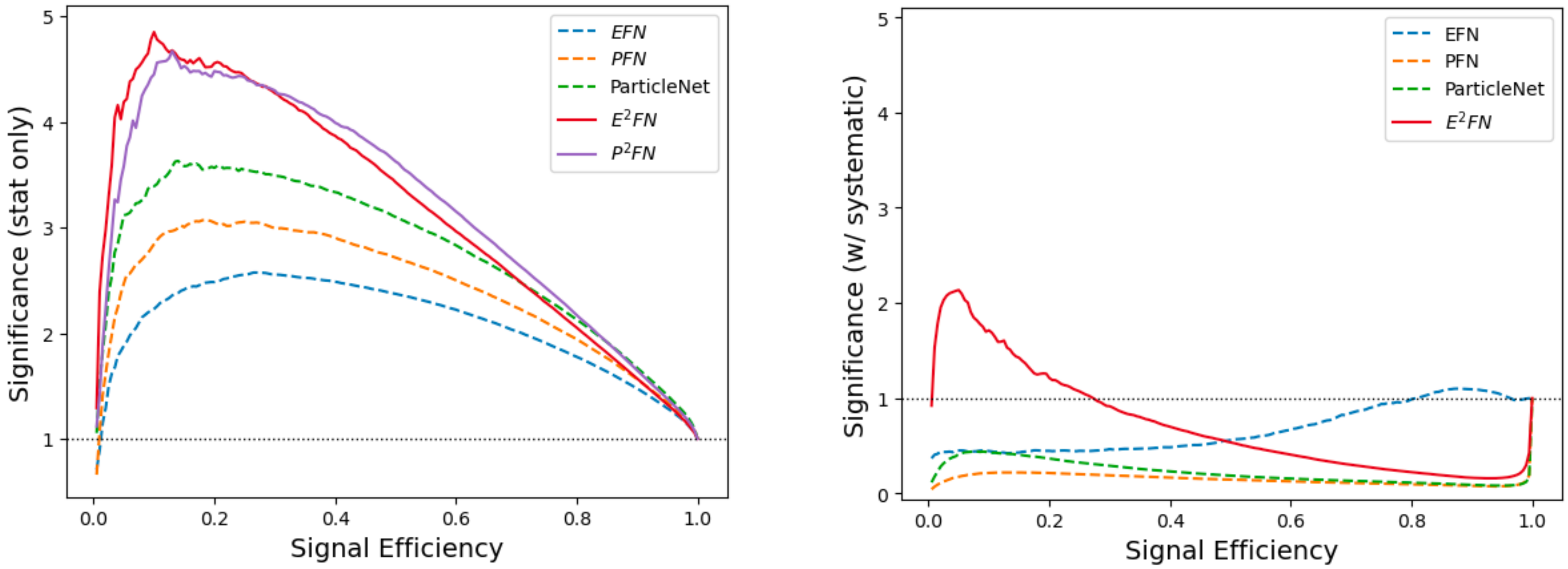


Significance with Theoretical Uncertainties

- Background Uncertainties:

$$\frac{1}{10} * [B(\text{hadron}) - B(\text{parton})]$$

⁕ IRC safety can help protect our measurements from theoretical systematics



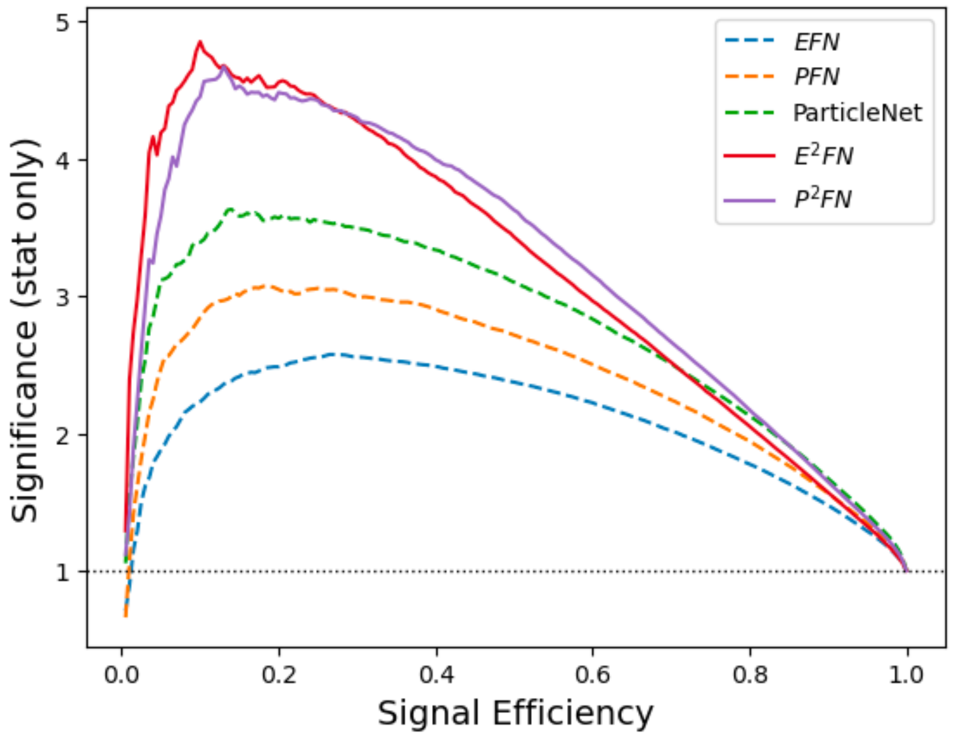
Significance with Theoretical Uncertainties

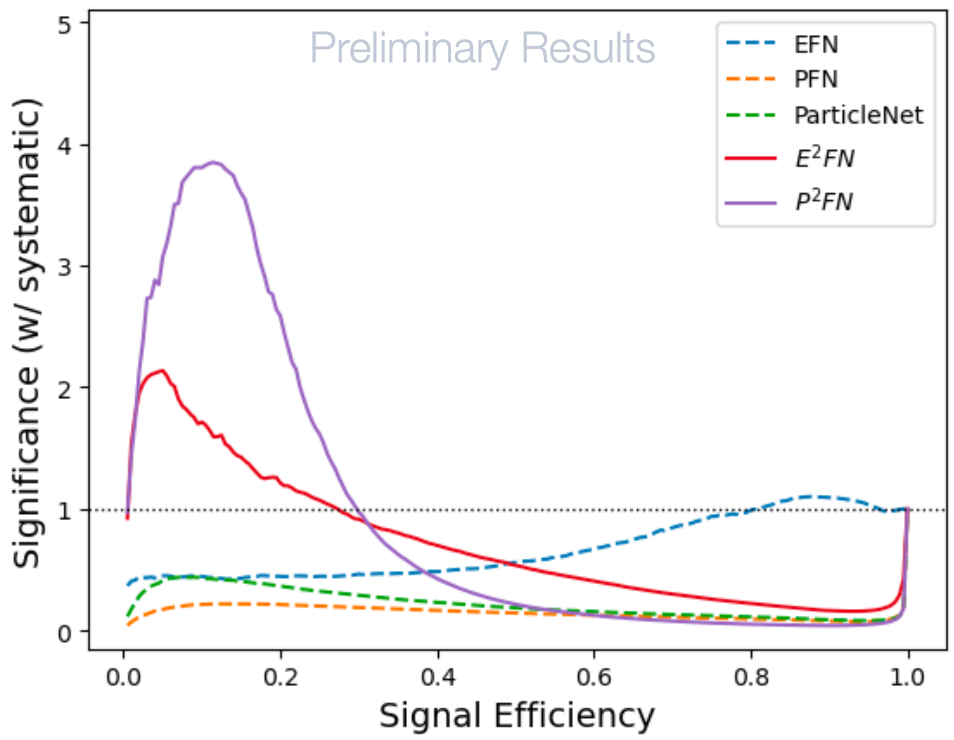
- Background Uncertainties:

$$\frac{1}{10} * [B(\text{hadron}) - B(\text{parton})]$$

⁂ IRC safety can help protect our measurements from theoretical systematics

⁂ There might be other non-IRC safe NNs that might be well-behaved and outperform IRC safe nets!



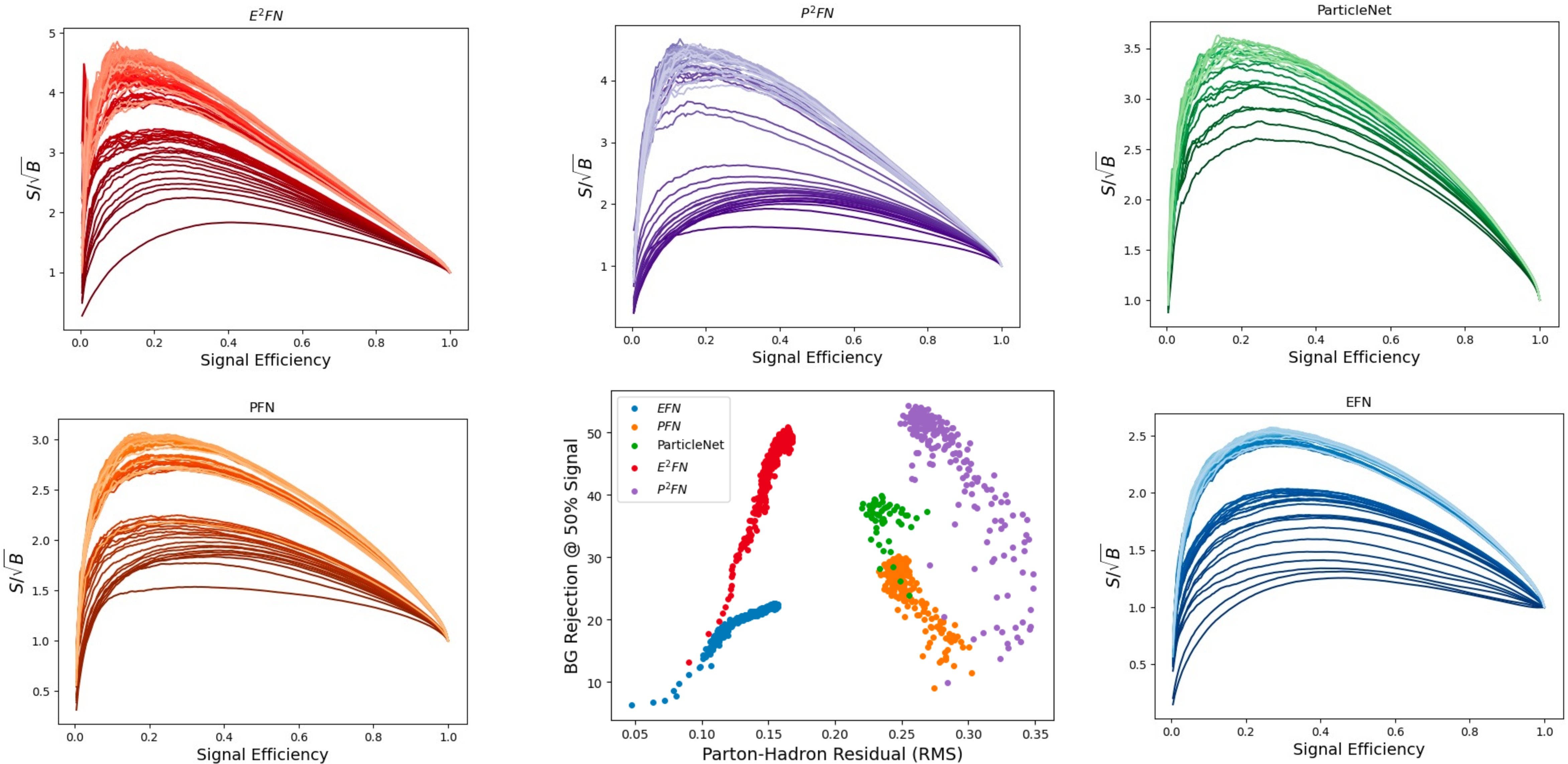


Summary

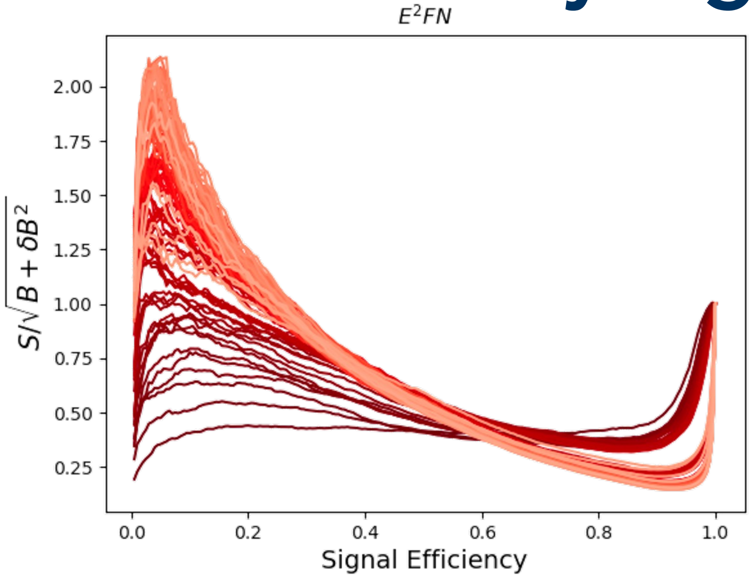
- Extended the notion of EFNs to the more generalized E2FNs which allow for particle correlation and improve the NN performance while ensuring IRC safety
- Understanding the regularization of a NN plays an important role in jet substructure measurements. Especially as we move towards measuring more complicated and high-detail jet observables
- Theoretical uncertainties from non-perturbative processes can have a measurable impact on actual searches and should be further studied

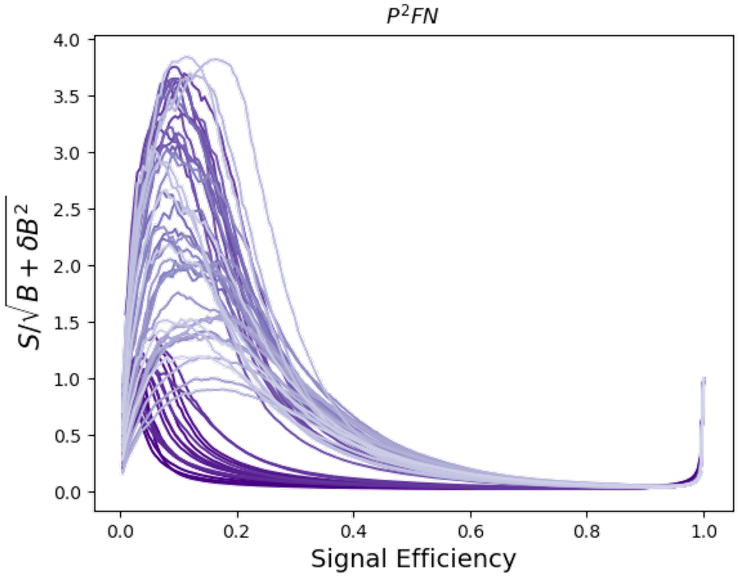

Run: 282712
Event: 474587238
2015-10-21 06:26:57
Thank You!

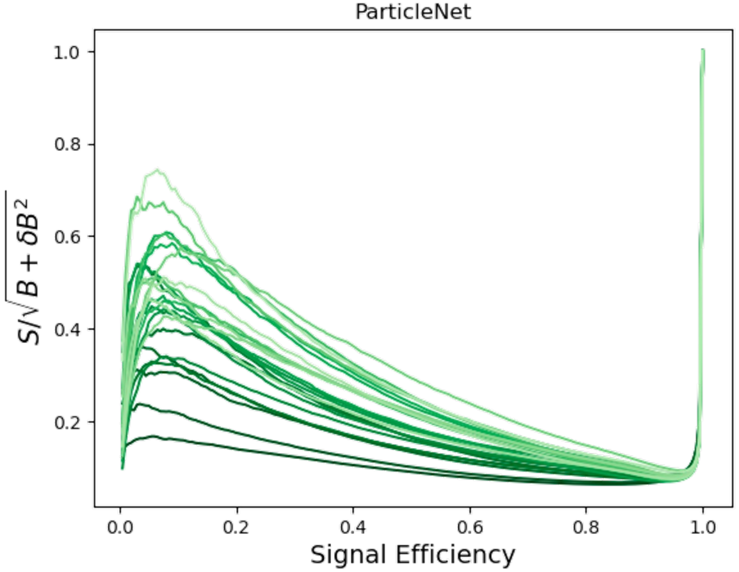
Raw Discovery Significance in Toy Search

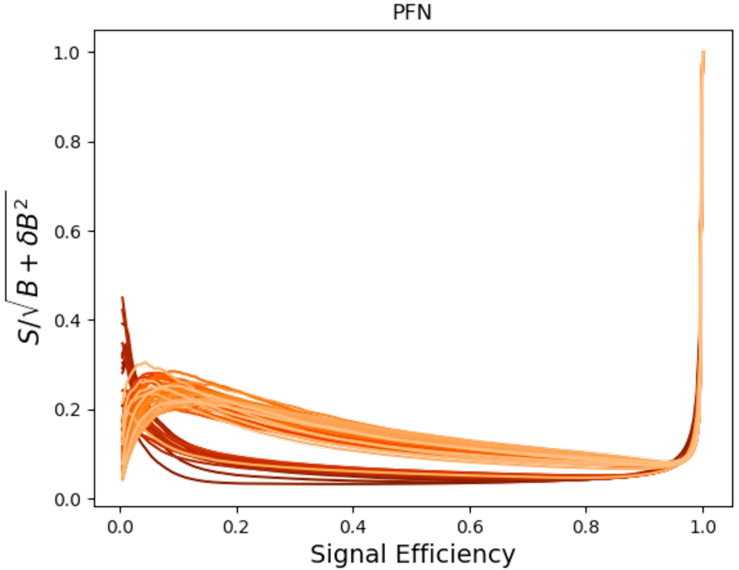


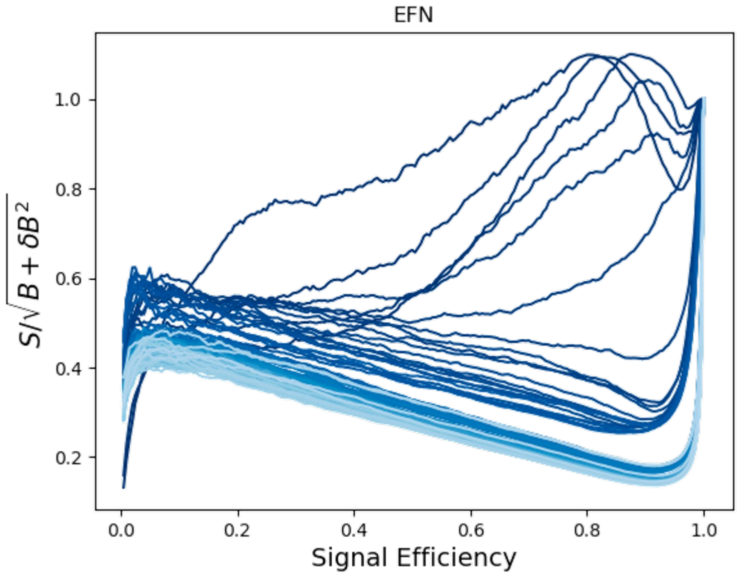
Raw Discovery Significance in Toy Search



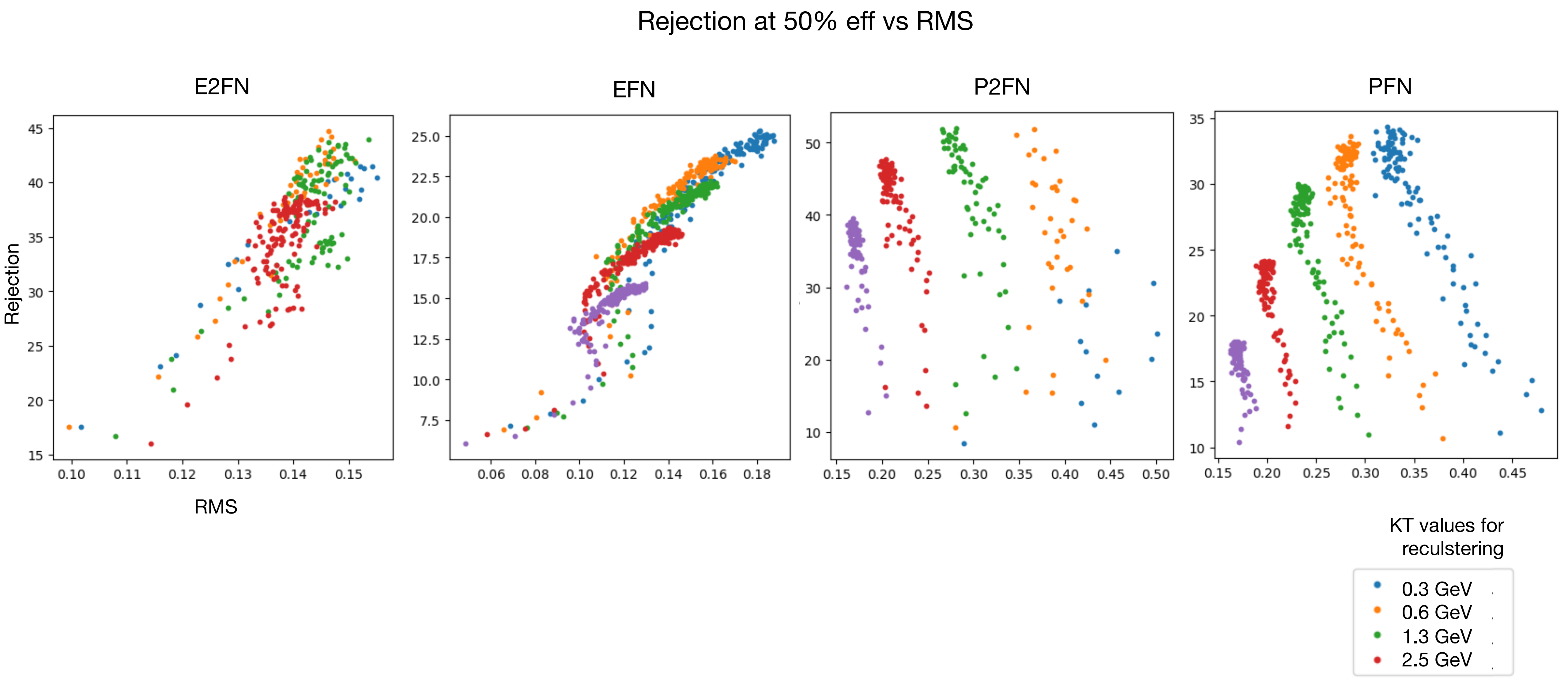




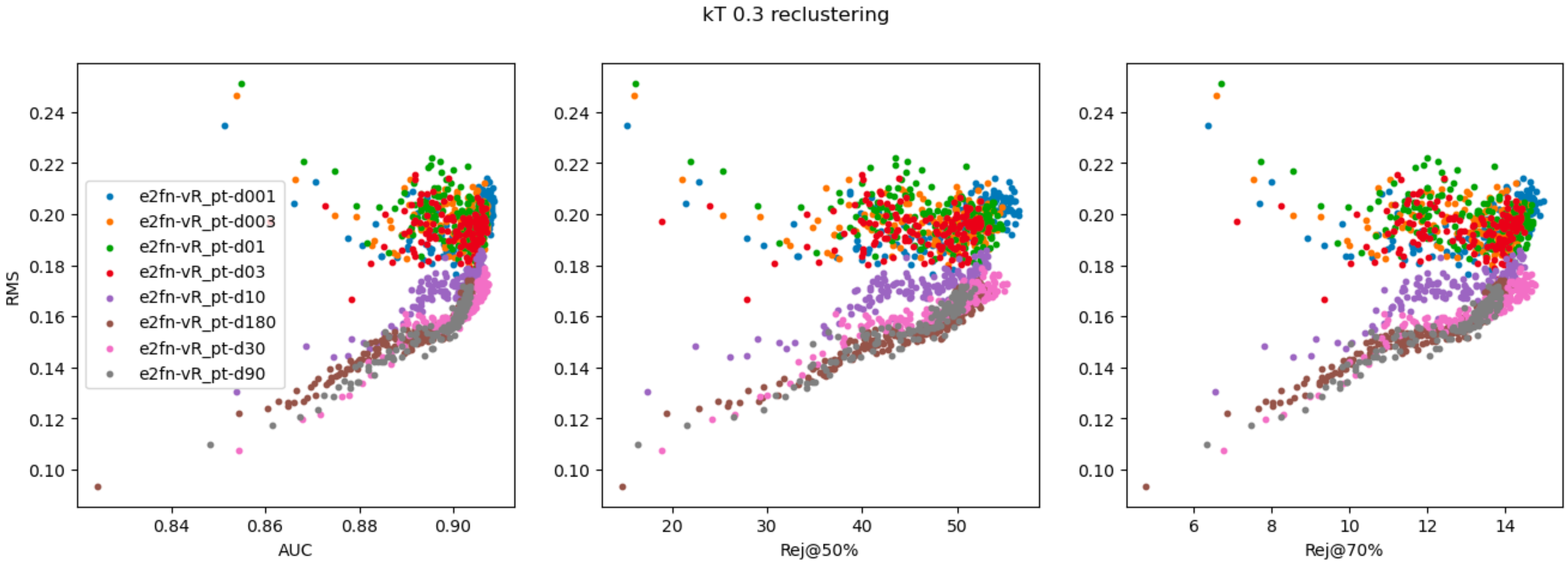




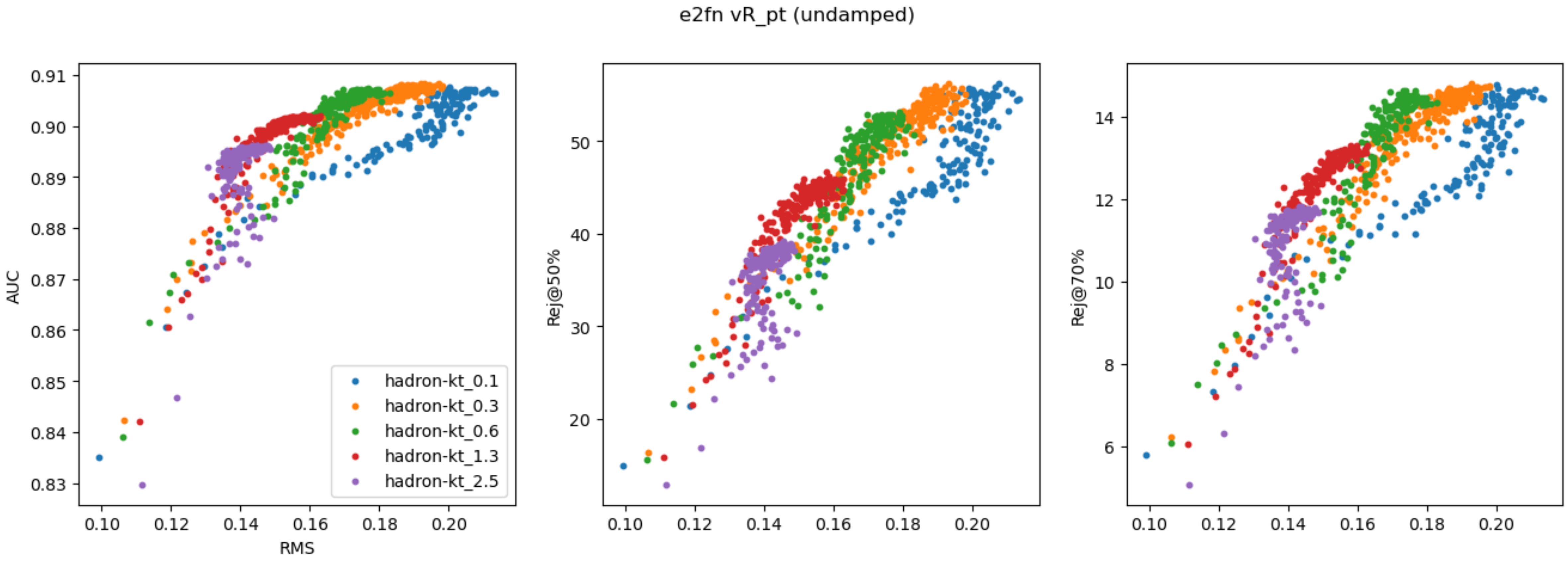
Regularization vs. Performance



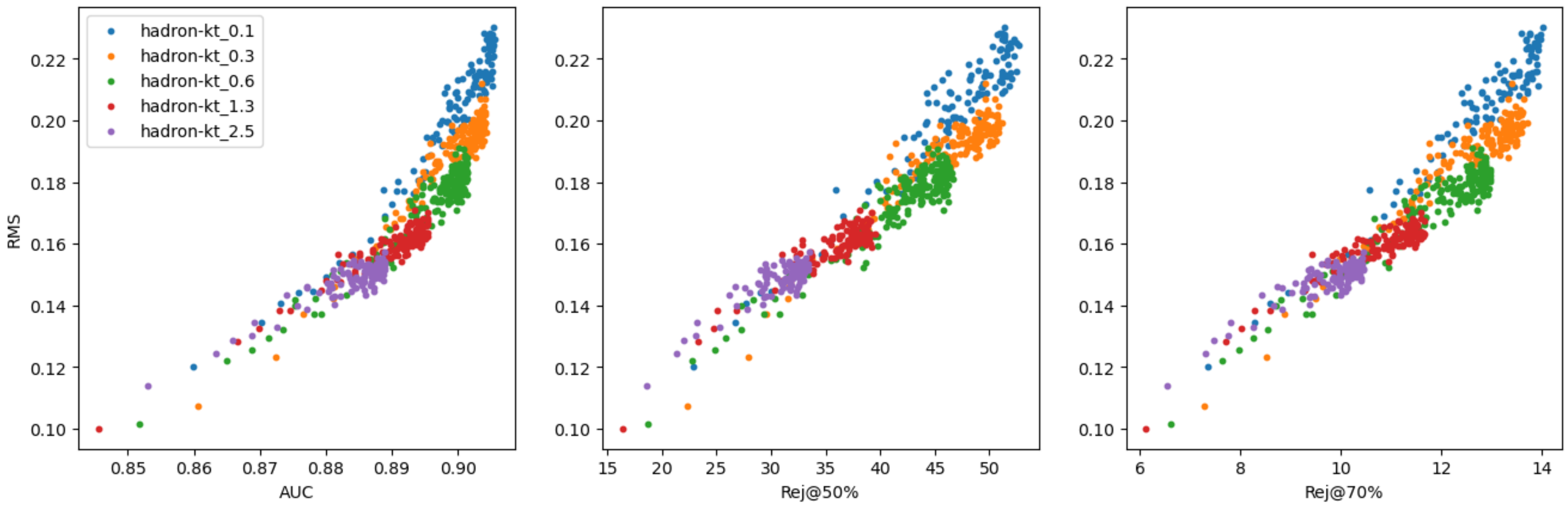
Damping


kT 0.3 reclustering
RMS
0.24
0.22
0.20
0.18
0.16
0.14
0.12
0.10
e2fn-vR_pt-d001
e2fn-vR_pt-d003
e2fn-vR_pt-d01
e2fn-vR_pt-d03
e2fn-vR_pt-d10
e2fn-vR_pt-d180
e2fn-vR_pt-d30
e2fn-vR_pt-d90
0.84
0.86
0.88
0.90
AUC
20
30
40
50
Rej@50%
6
8
10
12
14
Rej@70%

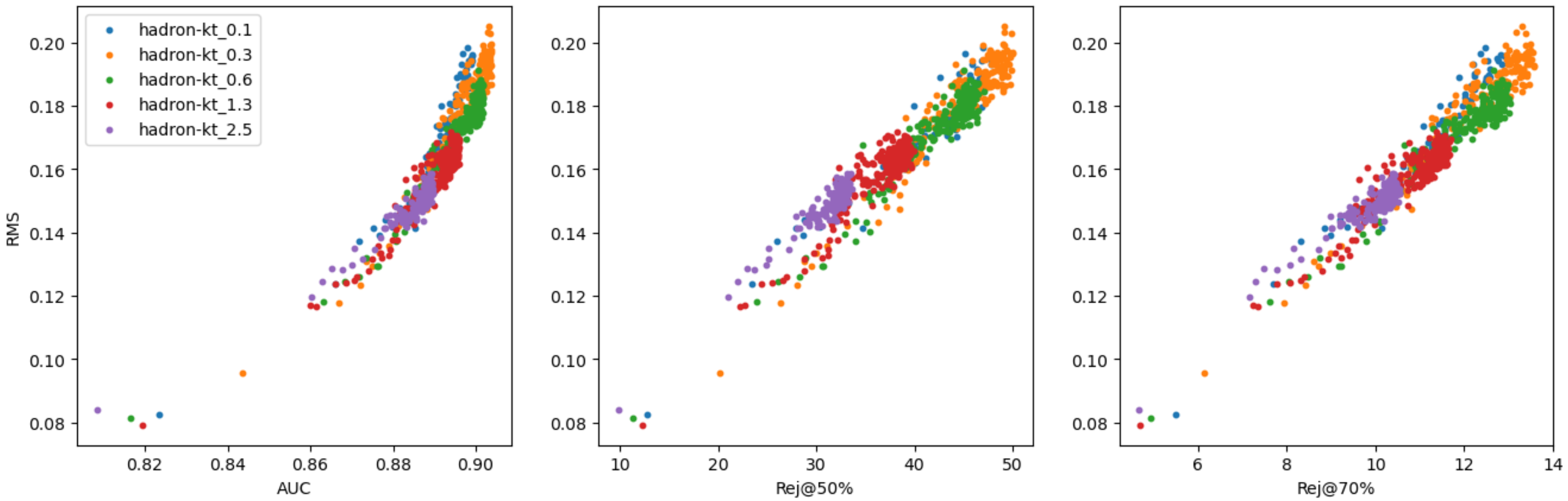
Undamped E2FN RMS vs Performance



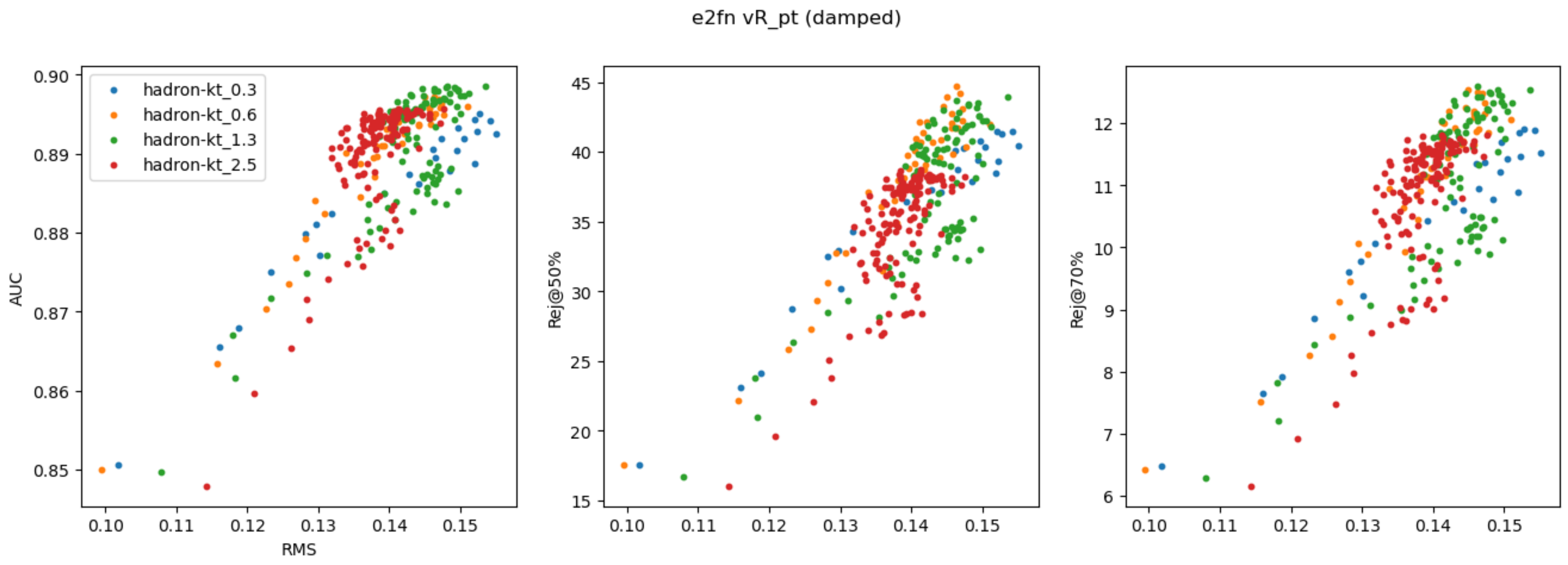
e2fn R_pt



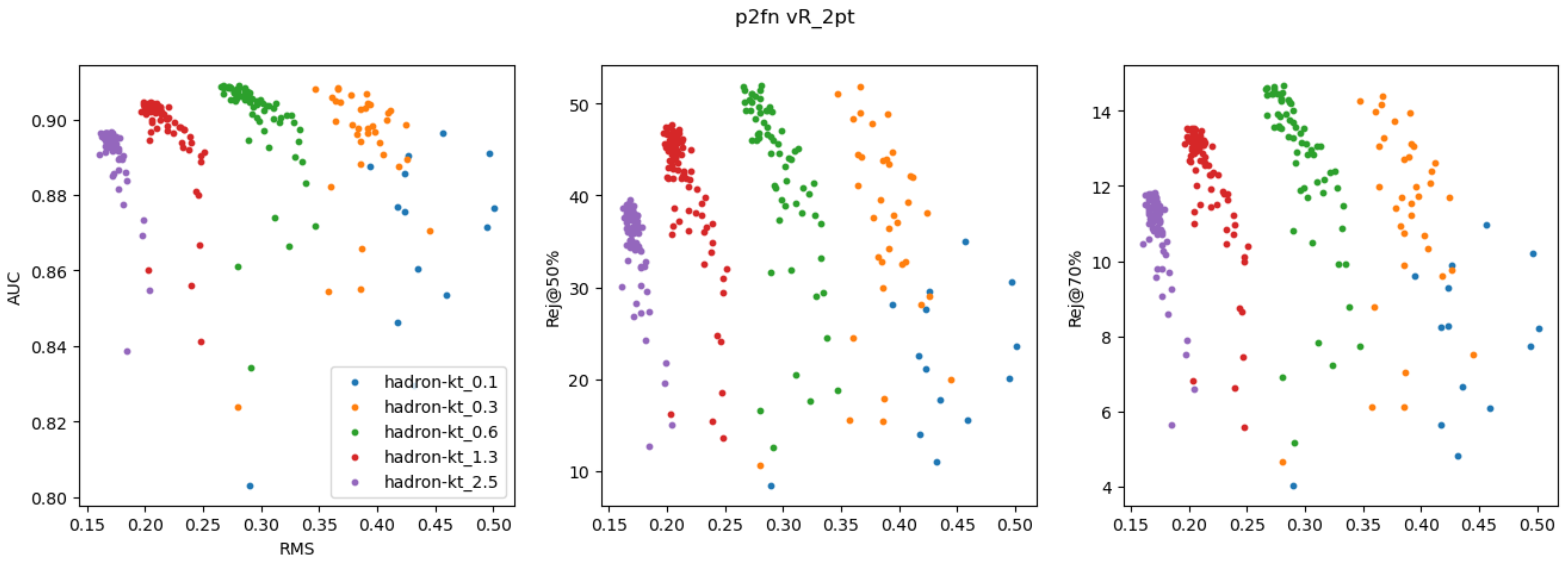
e2fn kt_pt



Damped E2FN RMS vs Performance



P2FN RMS vrs Performance



Energy-Energy Flow Networks (E2FNs)

- Energy-Energy Correlators (EEC)
  - ⇢ Transition between perturbative and non-perturbative physics at a visible angular scale between particles

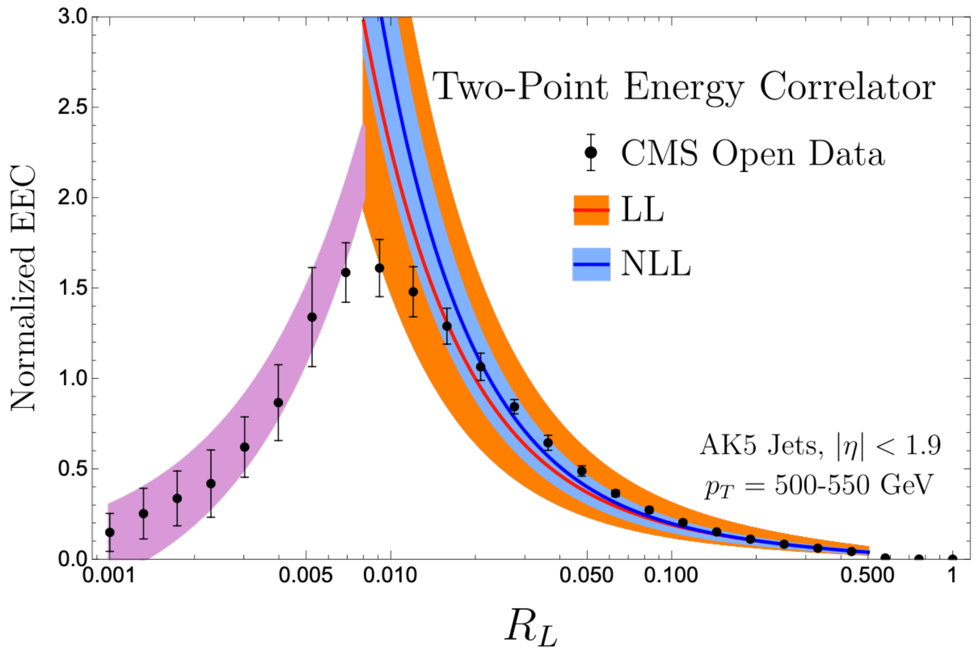


Table 3: Φ inputs for studied E2FN architectures

Input Values	Definition
$\hat{n}_i,\ \hat{n}_j$	$(\eta_i,\ \eta_j,\ \phi_i,\ \phi_j)$
R	$\left(\sqrt{(\eta_i-\eta_j)^2+(\phi_i-\phi_j)^2}\right)$
k_T	$\left(\sqrt{(\eta_i-\eta_j)^2+(\phi_i-\phi_j)^2} * p_{T\,jet}\right)$
Vector R	$(\ (\eta_i-\eta_j)\ ,\ (\phi_i-\phi_j)\ )$
$R,\ p_{T\,jet}$	$\left(\sqrt{(\eta_i-\eta_j)^2+(\phi_i-\phi_j)^2},\ p_{T\,jet}\right)$
$k_T,\ p_{T\,jet}$	$\left(\sqrt{(\eta_i-\eta_j)^2+(\phi_i-\phi_j)^2} * p_{T\,jet},\ p_{T\,jet}\right)$
Vector $R,\ p_{T\,jet}$	$(\ (\eta_i-\eta_j)\ ,\ (\phi_i-\phi_j)\ ,\ p_{T\,jet}\ )$

$$E2FN = F\left(\sum_{i,j}^{N} e_i e_j \Phi\left(\hat{n}_i, \hat{n}_j\right)\right)$$

Hyperparameters & Trainable Parameters

- 2 layers in phi
- 3 layers in F
- Relu activation
- Latent Dim: 32
- Learning rate: 1e-4
- Damping:

$$damp = \frac{\left(k_T/\tau\right)^{\frac{1}{w}}}{\left(k_T/\tau\right)^{\frac{1}{w}} + 1}$$

Table 6: Number of Trainable Parameters

Architecture	Phi	F	Total
PFN	21.2 K	37.5 K	58.7 K
EFN	21.0 K	37.5 K	58.5 K
E2FN	21.2 K	37.5 K	58.7 K
P2FN	21.3 K	37.5 K	58.8 K
ParticleNet	–	–	366 K