Energy Flow Polynomials for Jet Substructure

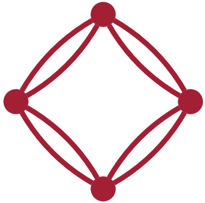
MIT Jet Workshop

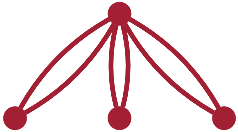
Cambridge, MA – January 11, 2018

Patrick T. Komiske III

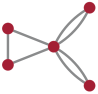
Center for Theoretical Physics, Massachusetts Institute of Technology

PTK, E.M. Metodiev, J. Thaler – 1712.07124





Part I - Introduction to Energy Flow Polynomials



Energy Flow Polynomials (EFPs)

*Part I based on a talk by Eric Metodiev.

EFP Essentials

$$\text{EFP}_G = \underbrace{\sum_{i_1=1}^{M} \cdots \sum_{i_N=1}^{M}}_{\text{Correlator}} \text{ of } \underbrace{z_{i_1} \cdots z_{i_N}}_{\text{Energies}} \text{ and } \underbrace{\prod_{(k,\ell)\in G} \theta_{i_k i_\ell}}_{\text{Angles}}$$

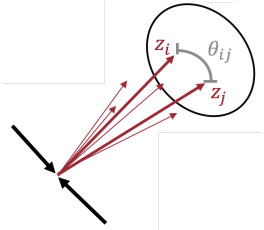


Figure by E.M. Metodiev

EFP Essentials

$$\mathrm{EFP}_G = \underbrace{\sum_{i_1=1}^{M} \cdots \sum_{i_N=1}^{M}}_{\text{Correlator}} \text{ of } \underbrace{z_{i_1} \cdots z_{i_N}}_{\text{Energies}} \text{ and } \underbrace{\prod_{(k,\ell)\in G} \theta_{i_k i_\ell}}_{\text{Angles}}$$

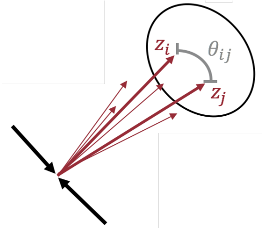


Figure by E.M. Metodiev

e^+e^- measure: $\quad z_i = \frac{E_i}{\sum_k E_k} \qquad \theta_{ij} = \left(\frac{2p_i^\mu p_{j\mu}}{E_i E_j}\right)^{\beta/2}$

Hadronic measure: $\quad z_i = \frac{p_{T,i}}{\sum_k p_{T,k}} \qquad \theta_{ij} = (\Delta y_{ij}^2 + \Delta\phi_{ij}^2)^{\beta/2}$

EFP Essentials

$$\mathrm{EFP}_G = \underbrace{\sum_{i_1=1}^{M} \cdots \sum_{i_N=1}^{M}}_{\text{Correlator}} \text{ of } \underbrace{z_{i_1} \cdots z_{i_N}}_{\text{Energies}} \text{ and } \underbrace{\prod_{(k,\ell)\in G} \theta_{i_k i_\ell}}_{\text{Angles}}$$

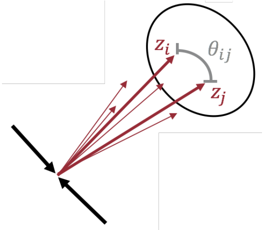
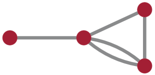


Figure by E.M. Metodiev

e^+e^- measure: $\quad z_i = \dfrac{E_i}{\sum_k E_k} \qquad \theta_{ij} = \left(\dfrac{2p_i^\mu p_{j\mu}}{E_i E_j}\right)^{\beta/2}$

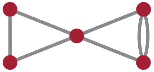
Hadronic measure: $\quad z_i = \dfrac{p_{T,i}}{\sum_k p_{T,k}} \qquad \theta_{ij} = (\Delta y_{ij}^2 + \Delta\phi_{ij}^2)^{\beta/2}$

e.g.



$$= \sum_{i_1=1}^{M}\sum_{i_2=1}^{M}\sum_{i_3=1}^{M}\sum_{i_4=1}^{M} z_{i_1} z_{i_2} z_{i_3} z_{i_4} \theta_{i_1 i_2} \theta_{i_2 i_3} \theta_{i_2 i_4}^2 \theta_{i_3 i_4}$$

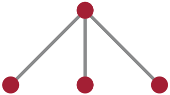
"fly-swatter"



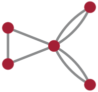
$$= \sum_{i_1=1}^{M}\sum_{i_2=1}^{M}\sum_{i_3=1}^{M}\sum_{i_4=1}^{M}\sum_{i_5=1}^{M} z_{i_1} z_{i_2} z_{i_3} z_{i_4} z_{i_5} \theta_{i_1 i_2} \theta_{i_2 i_3} \theta_{i_1 i_3} \theta_{i_1 i_4} \theta_{i_1 i_5} \theta_{i_4 i_5}^2$$

"bowtie"

Multigraph/EFP Correspondence

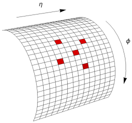
	Multigraph		**EFP**
	 j	$\longleftrightarrow$	z_{i_j}
	k  ℓ	$\longleftrightarrow$	$\theta_{i_k i_\ell}$
		$=$	$\sum_{i_1=1}^{M}\sum_{i_2=1}^{M}\sum_{i_3=1}^{M}\sum_{i_4=1}^{M} z_{i_1}z_{i_2}z_{i_3}z_{i_4}\theta_{i_1i_2}\theta_{i_1i_3}\theta_{i_1i_4}$
N	Number of vertices	$\longleftrightarrow$	N-particle correlator
d	Number of edges	$\longleftrightarrow$	Degree of angular monomial
χ	Treewidth + 1	$\longleftrightarrow$	Optimal VE Complexity
	Connected	$\longleftrightarrow$	Prime
	Disconnected	$\longleftrightarrow$	Composite

Part I - Introduction to Energy Flow Polynomials



Energy Flow Polynomials (EFPs)

*Part I based on a talk by Eric Metodiev.



Energy Flow Basis from IRC safety

EFPs Linearly Span IRC-safe Observables

- Start with an arbitrary IRC-safe observable $\mathcal{S}(p_1^\mu, \ldots, p_M^\mu)$

EFPs Linearly Span IRC-safe Observables

- Start with an arbitrary IRC-safe observable $\mathcal{S}(p_1^\mu, \ldots, p_M^\mu)$
  - Energy expansion*: Approx. $\mathcal{S}$ with polynomials of z_{i_j}
    - IR safety: $\mathcal{S}$ unchanged by addition of infinitesimally soft particles
    - C safety: $\mathcal{S}$ unchanged by collinear splittings of particles
    - Relabeling symmetry: Particle indexing is arbitrary

  See also F. Tkachov hep-ph/9601308, N. Sveshnikov and F. Tkachov hep-ph/9512370

$\Longrightarrow$ Energy correlators linearly span IRC-safe observables

*These expansions generically make use of the Stone-Weierstrass Theorem

EFPs Linearly Span IRC-safe Observables

- Start with an arbitrary IRC-safe observable $\mathcal{S}(p_1^\mu, \ldots, p_M^\mu)$
  - Energy expansion*: Approx. $\mathcal{S}$ with polynomials of z_{i_j}
    - IR safety: $\mathcal{S}$ unchanged by addition of infinitesimally soft particles
    - C safety: $\mathcal{S}$ unchanged by collinear splittings of particles
    - Relabeling symmetry: Particle indexing is arbitrary

See also F. Tkachov hep-ph/9601308, N. Sveshnikov and F. Tkachov hep-ph/9512370

$\Longrightarrow$ Energy correlators linearly span IRC-safe observables

- Angular expansion*: Approx. angular part of $\mathcal{S}$ with polynomials of θ_{ij}
- Simplify: Identify unique analytic structures that emerge as EFPs

Similar emergent multigraphs in M. Hogervorst *et al.* 1409.1581 and B. Henning *et al.* 1706.08520

$\Longrightarrow$ EFPs linearly span IRC-safe observables

$$\boxed{\mathcal{S} \simeq \sum_{G \in \mathcal{G}} s_G \,\mathrm{EFP}_G}$$

*These expansions generically make use of the Stone-Weierstrass Theorem

Organization of the Energy Flow Basis

- Need to select EFP subset $\mathcal{G}$
  - Determine $\mathcal{G}$ by truncating in d
  - Finite # of EFPs at each d
- OEIS A050535:

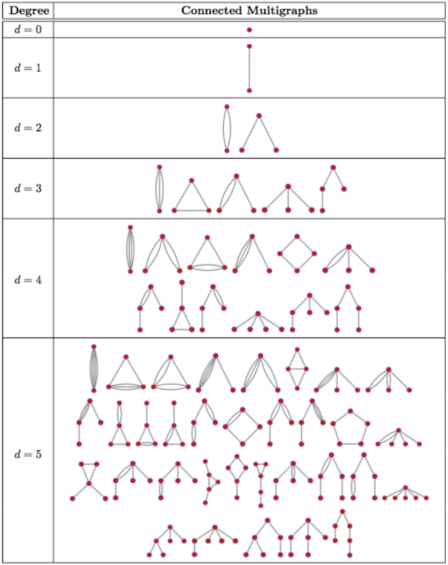
  # of multigraphs with d edges

  # of EFPs of degree d
- OEIS A076864:

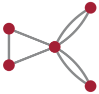
  # of connect. multigraphs with d edges

  # of prime EFPs with degree d

Maximum degree d		0	1	2	3	4	5	6	7	8	9	10
Prime EFPs	A076864	1	1	2	5	12	33	103	333	1 183	4 442	17 576
	Cumul.	1	2	4	9	21	54	157	490	1 673	6 115	23 691
All EFPs	A050535	1	1	3	8	23	66	212	686	2 389	8 682	33 160
	Cumul.	1	2	5	13	36	102	314	1 000	3 389	12 071	45 231

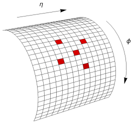


Part I - Introduction to Energy Flow Polynomials

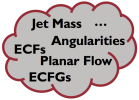


Energy Flow Polynomials (EFPs)

*Part I based on a talk by Eric Metodiev.



Energy Flow Basis from IRC safety



Taming the (IRC-safe) Substructure Zoo

Substructure Observables in the Energy Flow Context

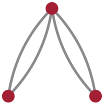
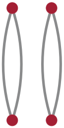
Jet Mass: $$\frac{m_J^2}{p_{TJ}^2} = \sum_{i_1=1}^{M} \sum_{i_2=1}^{M} z_{i_1} z_{i_2} (\cosh(\Delta y_{i_1 i_2}) - \cos(\Delta \phi_{i_1 i_2})) = \frac{1}{2} \times$$  $+ \cdots$

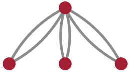
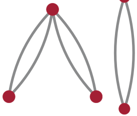
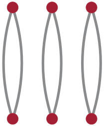
Substructure Observables in the Energy Flow Context

Jet Mass: $$\frac{m_J^2}{p_{TJ}^2} = \sum_{i_1=1}^{M}\sum_{i_2=1}^{M} z_{i_1} z_{i_2} (\cosh(\Delta y_{i_1 i_2}) - \cos(\Delta\phi_{i_1 i_2})) = \frac{1}{2} \times$$  $$+ \cdots$$

Angularities: $$\lambda^{(\alpha)} = \sum_{i=1}^{M} z_i \theta_i^{\alpha}$$ using p_T-centroid axis

C.F. Berger, T. Kucs, and G. Sterman, hep-ph/0303051
S.D. Ellis, *et al.*, 1001.0014
A.J. Larkoski, J. Thaler, and W. Waalewijn, 1408.3122

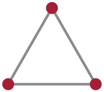
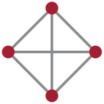
$$\lambda^{(4)} =$$  $$- \frac{3}{4} \times$$ 

$$\lambda^{(6)} =$$  $$- \frac{3}{2} \times$$  $$+ \frac{5}{8} \times$$ 

Substructure Observables in the Energy Flow Context

Energy Correlation Functions:

$$e_N^{(\beta)} = \sum_{i_1}^{M} \cdots \sum_{i_N=1}^{M} z_{i_1} \cdots z_{i_N} \prod_{k<\ell\in\{1,\ldots,N\}} \theta_{i_k i_\ell}^{\beta}$$

$e_2^{(\beta)} =$  , $e_3^{(\beta)} =$  , $e_4^{(\beta)} =$ 

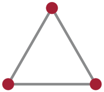
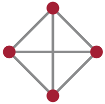
with measure choice β

A.J. Larkoski, G.P. Salam, J. Thaler, 1305.0007

Substructure Observables in the Energy Flow Context

Energy Correlation Functions:

$$e_N^{(\beta)} = \sum_{i_1}^{M} \cdots \sum_{i_N=1}^{M} z_{i_1} \cdots z_{i_N} \prod_{k<\ell\in\{1,\dots,N\}} \theta_{i_k i_\ell}^{\beta}$$

$e_2^{(\beta)} =$  , $e_3^{(\beta)} =$  , $e_4^{(\beta)} =$ 

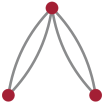
with measure choice β

A.J. Larkoski, G.P. Salam, J. Thaler, 1305.0007

Geometric Moments:

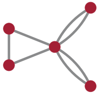
$$\mathbf{C} = \sum_{i=1}^{M} z_i \begin{pmatrix} \Delta y_i^2 & \Delta y_i \Delta\phi_i \\ \Delta y_i \Delta\phi_i & \Delta\phi_i^2 \end{pmatrix}, \qquad \mathrm{Pf} = \frac{4 \det \mathbf{C}}{(\mathrm{Tr}\,\mathbf{C})^2}$$

using p_T-centroid axis

$\mathrm{Tr}\,\mathbf{C} = \frac{1}{2} \times$  , $4 \det \mathbf{C} =$  $- \frac{1}{2} \times$ 

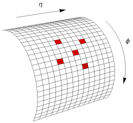
L.G. Almeida, *et al.*, 0807.0234 J. Thaler and Lian-Tao Wang, 0806.0023 J. Gallicchio and M. Schwartz, 1211.7038

Part I - Introduction to Energy Flow Polynomials

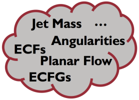


Energy Flow Polynomials (EFPs)

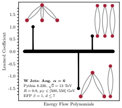
*Part I based on a talk by Eric Metodiev.



Energy Flow Basis from IRC safety



Taming the (IRC-safe) Substructure Zoo



Spanning Substructure with Linear Regression

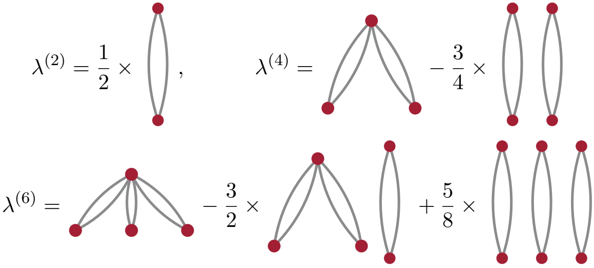
Linear Models

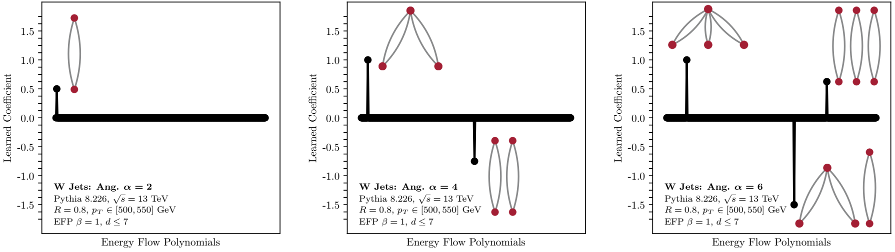
$$\mathcal{S} \simeq \sum_{G\in\mathcal{G}} s_G \mathrm{EFP}_G, \quad \mathcal{S}: \text{ IRC-safe observable}, \quad \mathcal{G}: \text{ set of EFPs}$$

- Machine learn $\{s_G\}$ with a linear model
- Linear models:
  - Convex with few/no hyperparameters to tune
  - Achieve global optimum via closed form solution or convergent iteration
  - Cannot have a simpler model (1 parm./input) sensitive to all inputs
  - Many potential methods to analyze learned model

  See Ch. 3 and 4 of C. Bishop *Pattern Recognition and Machine Learning*

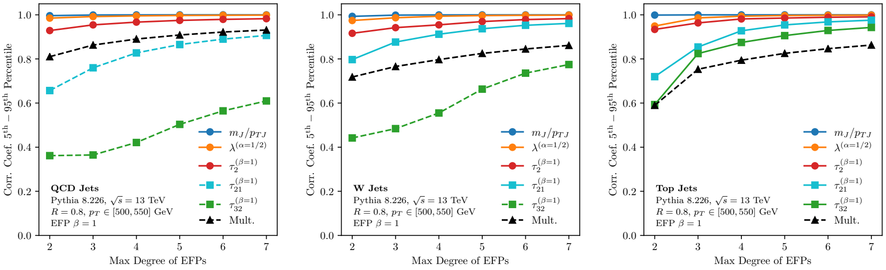
Confirming Analytic Relationships





Linear Regression and IRC-safety

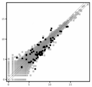
- $m_J/p_{T,J}$: IRC safe - no Taylor expansion due to square root
- $\lambda^{(\alpha=1/2)}$: IRC safe - no simple analytic relationship
- $\tau_2^{(\beta=1)}$: IRC safe - algorithmically defined
- $\tau_{21}^{(\beta=1)}$: Sudakov safe - safe for 2-prong jets and more
- $\tau_{32}^{(\beta=1)}$: Sudakov safe - safe for 3-prong jets and more
- Multiplicity: IRC unsafe



Part I Conclusions

- EFPs:
  - Energy correlators with angular structures indexed by multigraphs
  - Linearly span the space of IRC-safe observables
  - Encompass many existing classes of substructure observables
- Linear regression:
  - Linear models are the easiest and most tractable kind of model
    - Convex with few/no hyperparameters
    - Global optimum via closed form solution or convergent iteration
    - Many potential tools to analyze what's learned
  - Works with EFPs to match onto many IRC-safe observables

Part II - Linear Jet Tagging with EFPs



Linear Classification with EFPs

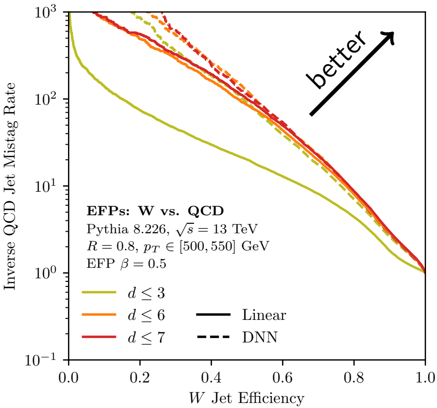
Linear Classification Overview

- Fit a decision plane, determined by a vector $\boldsymbol{w}$
  - Fisher's linear discriminant (LDA): closed-form solution
  - Logistic regression: Convex, iterative solution
- Decision threshold t is determined by distance from the plane
- $\mathcal{G}$ is finite set of graphs corresponding to the inputs
  - Organization by d is natural (equivalent to the order of the expansion)
  - Organization by N or χ also possible, (where is the information?)

$$\text{Classifier} = \begin{cases} \left(t + \sum_{G\in\mathcal{G}} w_G \text{EFP}_G\right) & \geq 0, \quad \text{signal} \\ & < 0, \quad \text{background} \end{cases}$$

Linear Classification with EFPs

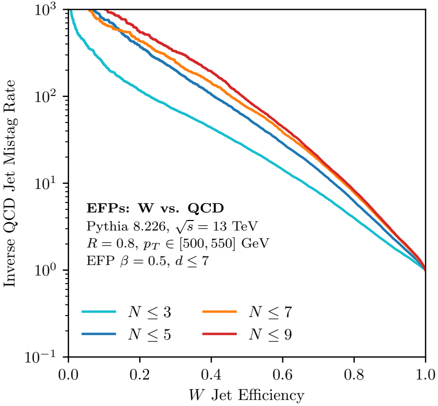
- W vs. QCD jet classification (quark/gluon and top tagging in backup)

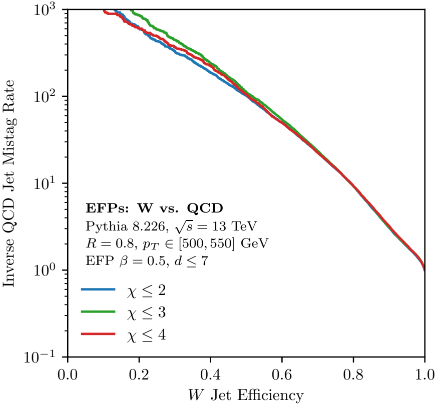


- 300k training samples
- Linear: Fisher's linear discriminant
  - num. params. = num. EFPs < 1000
  - 100k test samples
- DNN: Dense neural net
  - (100 node fully-connected layer) $\times$ 3
  - $\sim$ 120k parameters
  - 50k validation, 50k test samples

Azure

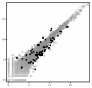
Which EFPs are Important?



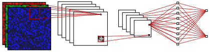


- High-N EFPs are important for classification performance
- Great classification performance with just $\chi = 2$ EFPs!

Part II - Linear Jet Tagging with EFPs

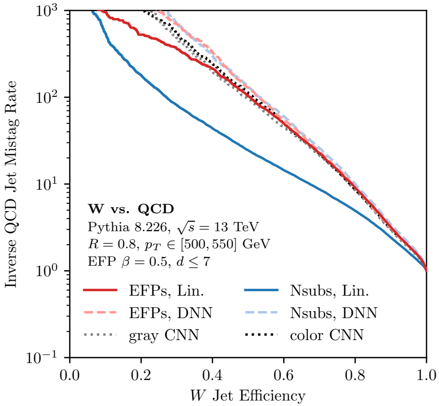


Linear Classification with EFPs



Comparison with Modern Machine Learning

Modern Machine Learning Comparison

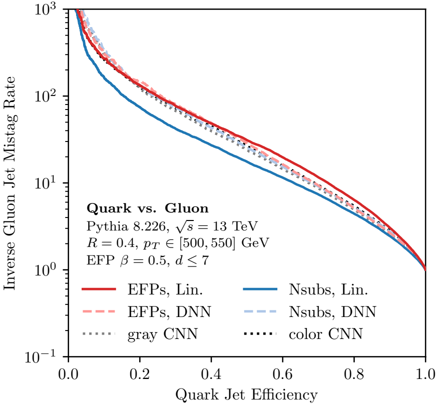


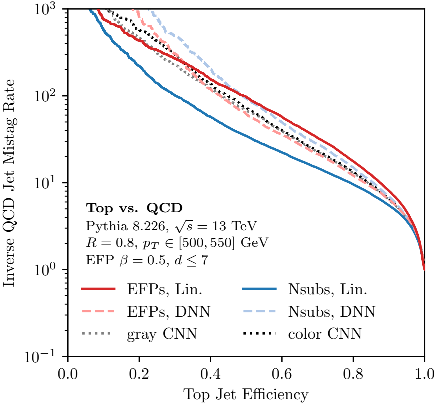
- Linear and DNN same as before
- CNN: Convolutional neural net
  - 33×33 jet images
  - (48 filters) $\times$ 3
  - gray: p_T channel only
  - color: p_T and mult. channels
  - $\sim$ 80k parameters

(Linear classification with EFPs) $\sim$ (MML) for $\varepsilon_s \gtrsim 0.5$!

N-subjetiness: 1011.2268, N-subjetiness basis: 1704.08249, NN Review: 1709.04464

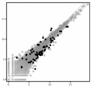
Modern Machine Learning Comparison



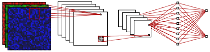


(Linear classification with EFPs) $\gtrsim$ (MML) for $\varepsilon_s \gtrsim 0.5$

Part II - Linear Jet Tagging with EFPs



Linear Classification with EFPs



Comparison with Modern Machine Learning

Fast Computation of EFPs

Computational Complexity of ECF(G)s

$$\sum_{i_1=1}^{M}\cdots\sum_{i_N=1}^{M} E_{i_1}\cdots E_{i_N}\begin{cases}\displaystyle\prod_{i<j\in\{i_1,\ldots,i_N\}}\theta_{ij}^{\beta}, & \mathsf{ECF}_N^{\beta} \quad \text{1305.0007}\\ \displaystyle\prod_{m=1}^{v}\min_{(m)}\{\theta_{ij}^{\beta}\}_{i<j\in\{i_1,\ldots,i_N\}}, & {}_v\mathsf{ECFG}_N^{\beta} \quad \text{1609.07483}\end{cases}$$

- Implementation of ECF(G) formula runs in time $\mathcal{O}(M^N)$
- With $M \sim 100$, ECF(G)$_{N=4}$ $\sim$ one hundred million operations
- $N = 4$ is barely tractable, $N \geq 5$ is essentially inaccessible

Computational Complexity of ECF(G)s

$$\sum_{i_1=1}^{M}\cdots\sum_{i_N=1}^{M} E_{i_1}\cdots E_{i_N}\begin{cases}\displaystyle\prod_{i<j\in\{i_1,\ldots,i_N\}}\theta_{ij}^{\beta}, & \text{ECF}_N^{\beta} \\ \displaystyle\prod_{m=1}^{v}\min_{(m)}\{\theta_{ij}^{\beta}\}_{i<j\in\{i_1,\ldots,i_N\}}, & {}_v\text{ECFG}_N^{\beta}\end{cases}$$

1305.0007

1609.07483

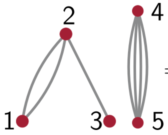
- Implementation of ECF(G) formula runs in time $\mathcal{O}(M^N)$
- With $M \sim 100$, ECF(G)$_{N=4}$ $\sim$ one hundred million operations
- $N = 4$ is barely tractable, $N \geq 5$ is essentially inaccessible

```
// if N > 5, then throw error
if (_N > 5) {
    throw Error("EnergyCorrelator is only hard coded for N = 0,1,2,3,4,5");
}
```

fjcontrib – EnergyCorrelator 1.2.0

Computational Complexity of EFPs

- Like other energy correlators, EFPs are naively $\mathcal{O}(M^N)$
- *Factorability* of summand in EFP formula can speed up computation

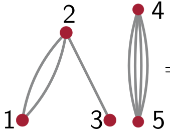


$$= \left(\sum_{i_1=1}^{M} \sum_{i_1=1}^{M} \sum_{i_3=1}^{M} z_{i_1} z_{i_2} z_{i_3} \theta_{i_1 i_2}^2 \theta_{i_2 i_3} \right) \left(\sum_{i_4=1}^{M} \sum_{i_5=1}^{M} z_{i_4} z_{i_5} \theta_{i_4 i_5}^4 \right)$$

Composite EFPs are products of prime EFPs

Computational Complexity of EFPs

- Like other energy correlators, EFPs are naively $\mathcal{O}(M^N)$
- *Factorability* of summand in EFP formula can speed up computation



$$= \left(\sum_{i_1=1}^{M} \sum_{i_1=1}^{M} \sum_{i_3=1}^{M} z_{i_1} z_{i_2} z_{i_3} \theta_{i_1 i_2}^2 \theta_{i_2 i_3} \right) \left(\sum_{i_4=1}^{M} \sum_{i_5=1}^{M} z_{i_4} z_{i_5} \theta_{i_4 i_5}^4 \right)$$

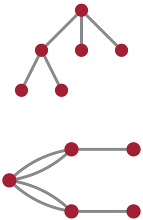
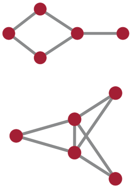
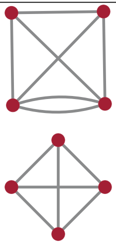
Composite EFPs are products of prime EFPs

$$= \underbrace{\sum_{i_1=1}^{M} \sum_{i_2=1}^{M} \sum_{i_3=1}^{M} \sum_{i_4=1}^{M} \sum_{i_5=1}^{M} \sum_{i_6=1}^{M} \sum_{i_7=1}^{M} \sum_{i_8=1}^{M} z_{i_1} z_{i_2} z_{i_3} z_{i_4} z_{i_5} z_{i_6} z_{i_7} z_{i_8} \theta_{i_1 i_2} \theta_{i_1 i_3} \theta_{i_1 i_4} \theta_{i_1 i_5} \theta_{i_1 i_6} \theta_{i_1 i_7} \theta_{i_1 i_8}}_{\mathcal{O}(M^8)}$$

$$= \underbrace{\sum_{i_1=1}^{M} z_{i_1} \left(\sum_{i_2=1}^{M} z_{i_2} \theta_{i_1 i_2} \right)^7}_{\mathcal{O}(M^2)}$$

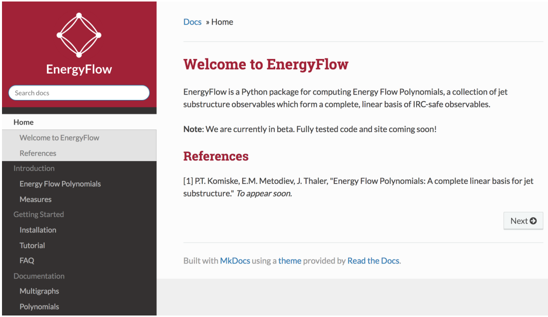
Variable Elimination (VE)

- Algorithm for finding optimal parentheses placement in EFP formula
- Reduces EFP computational complexity to $\mathcal{O}(M^\chi)$:
  - Best case (NP-hard): $\chi = \texttt{treewidth}(G) + 1$
  - Heuristics can be used which work excellently for our small graphs
  - $\chi = 2$ for all tree graphs, $\chi = 3$ for single-cycle graphs, $\chi = N$ for K_N

χ	2	3	4
e.g.			

EnergyFlow Python Package

- A convenient and simple package for efficient implementation of EFPs
- Currently written in pure Python using the NumPy library
  - Need a fast, arbitrary dimension multi-array
  - We're working on a C++ implementation (not simple)




```
pip install energyflow
```

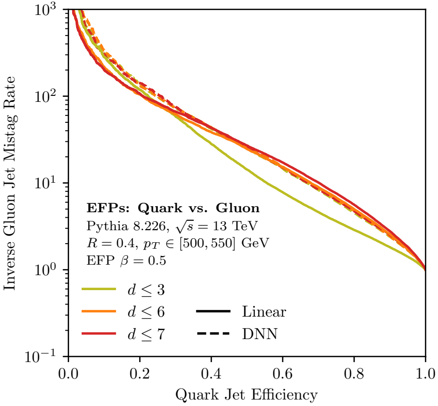
```python
import energyflow as ef
efps = ef.EFPSet('d<=7')
results = efps.compute(event)
```

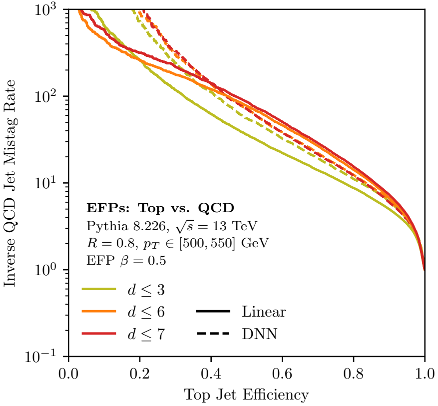
Conclusions

- Linear classification with EFPs very comparable to MML methods
- Linear methods $\implies$ very nice both theoretically and experimentally
  - EFP linear structure potentially allows for theoretical calculation
  - Fully differentiable model, uncertainty/error propagation simple
  - Convex, global minimum is guaranteed
  - No/few hyperparameters
  - Interesting methods made possible by linearity
    - Lasso regression for automatic feature selection
    - PCA, orthogonal subspaces, etc.
- Efficient computation of EFPs has been achieved
  - `EnergyFlow` Python package here, stay tuned for more
- EFPs potentially bridge MML performance & theory understanding

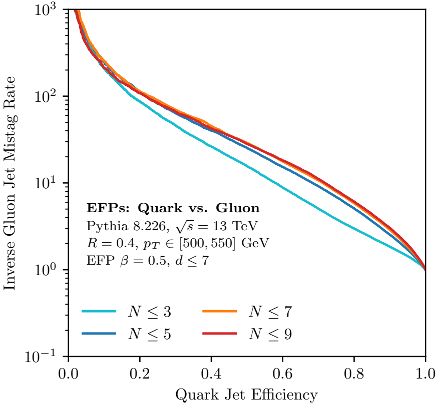
Additional Slides

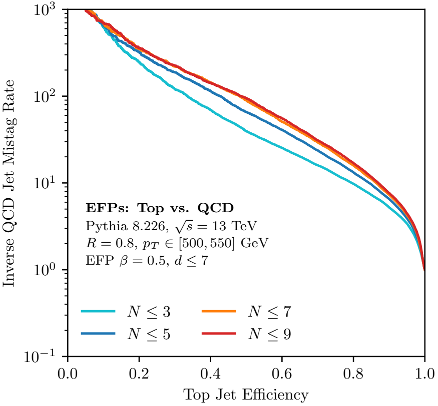
Quark/Gluon, Top Linear Classification with EFPs



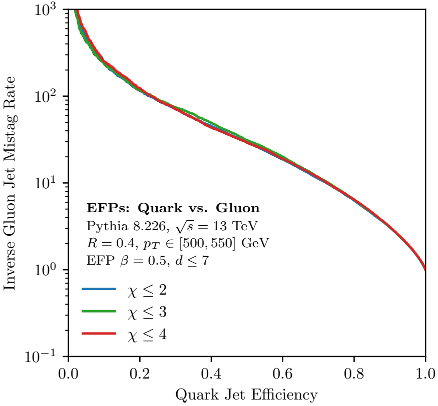


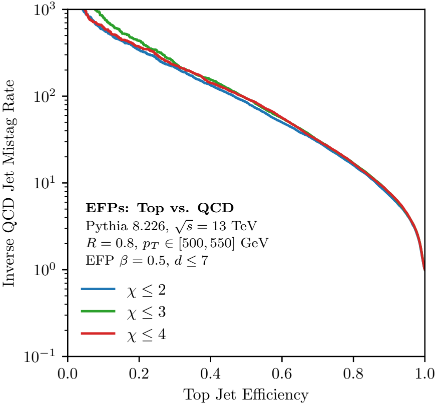
Quark/Gluon and Top Tagging N Sweep



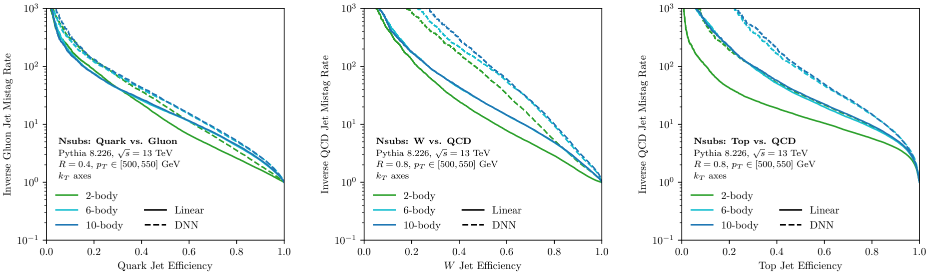


Quark/Gluon and Top Tagging χ Sweep



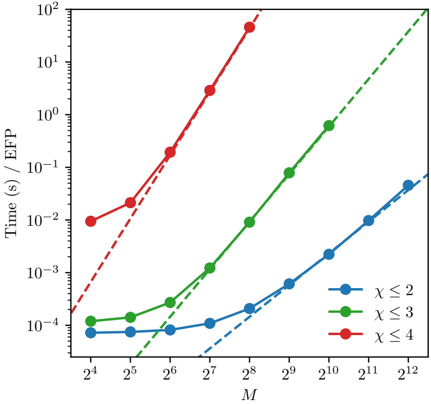


N-Subjettiness Linear/DNN Comparison



VE Timing

- Test our implementation of VE averaged over all EFPs with $d \leq 7$
- This includes prime EFPs up to $N = 8$! (Imagine $N = 8$ ECF, OMG)



N		2	3	4	5	6	7	8
χ	2	7	12	33	50	65	48	23
	3		11	42	82	80	33	
	4			2	1			

Prime EFPs with $d \leq 7$