Automated discovery of jet substructure analyses

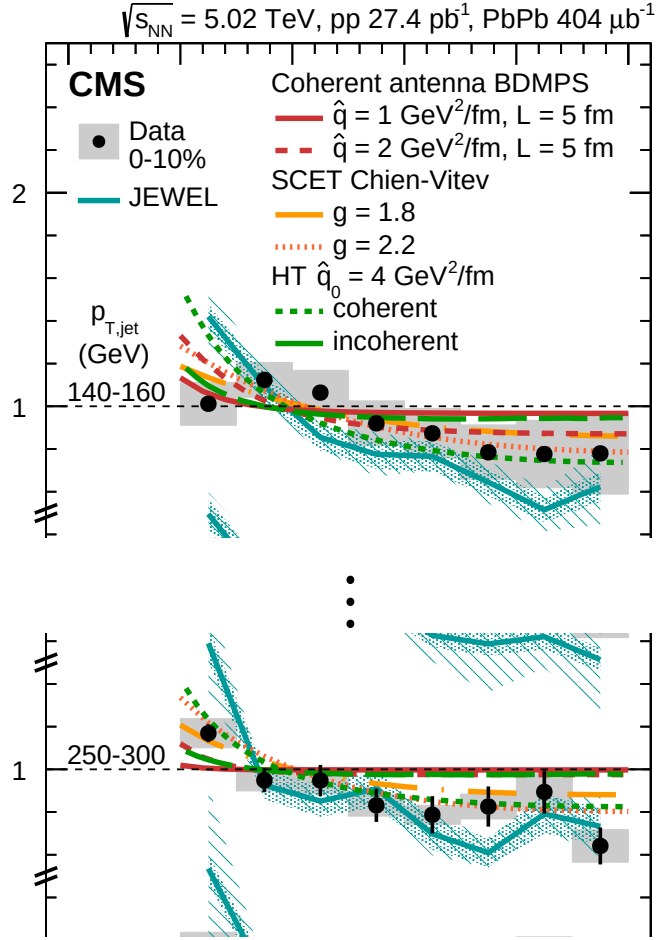
Yue Shi Lai

Lawrence Berkeley National Lab, Division of Nuclear Science

Hard Probes 2018, Aix-les-Bains

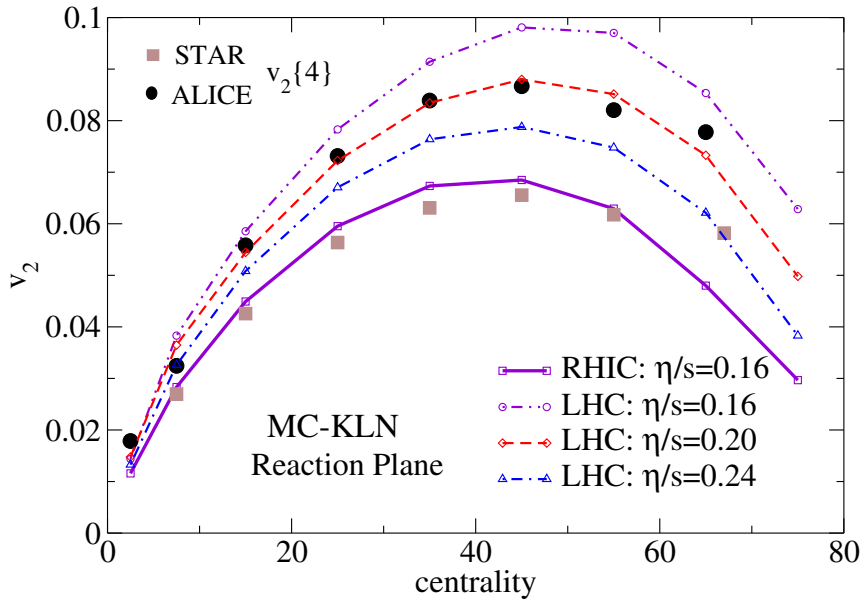
“Why” and “what can you do with it?”

- Jet substructure studied as “on-off switches”, results have not been shown how to probe the medium quantitatively with jet substructure



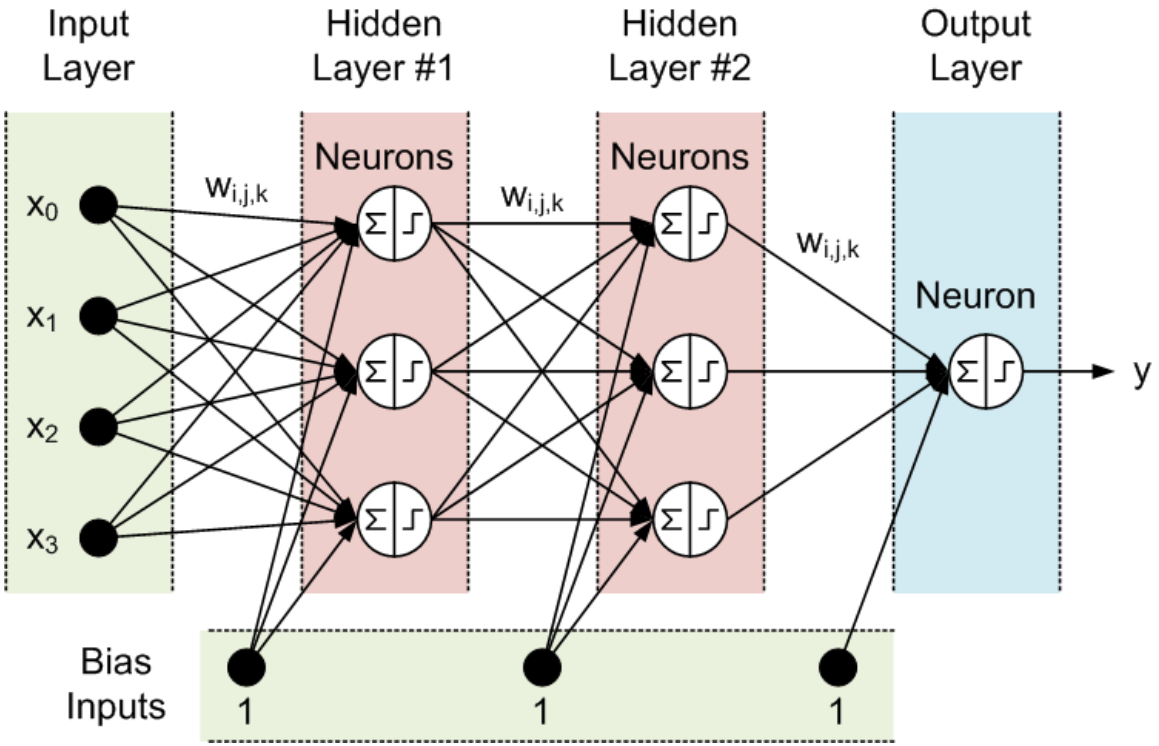
A. M. Sirunyan et al. (CMS), Phys. Rev. Lett. 120, 142302 (2018)

- E.g. CMS z_g: Red-dashed (BDMPS) and orange-dashed (SCET) require 10% accuracy, beyond experimental reach
- State-of-the-art flow analyses look like this:



H.-c. Song, S. A. Bass, U. Heinz, Phys. Rev. C 83, 054912 (2011)

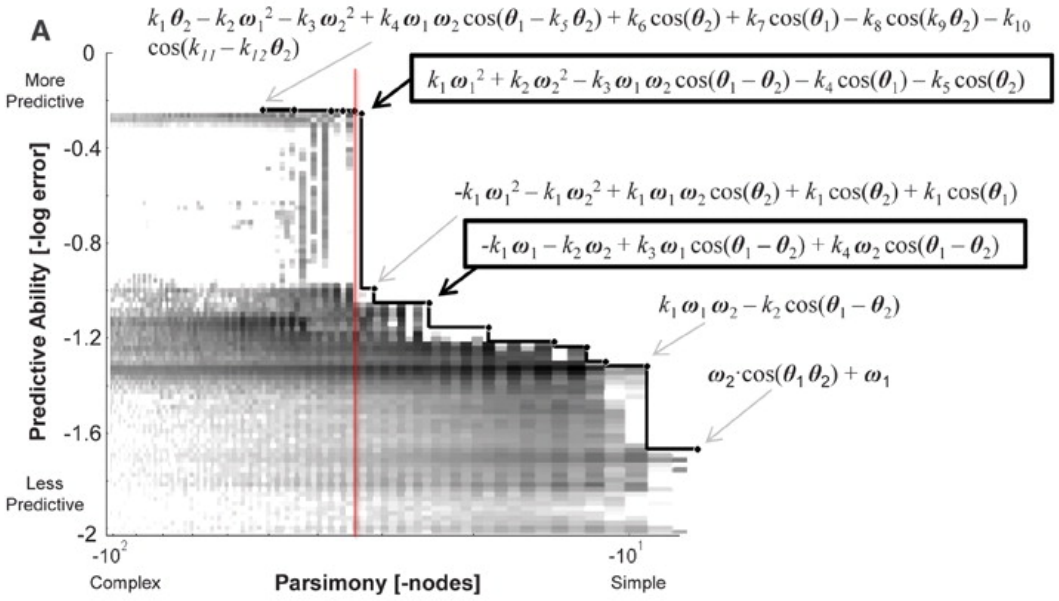
Ingredient: Neural network (NN)

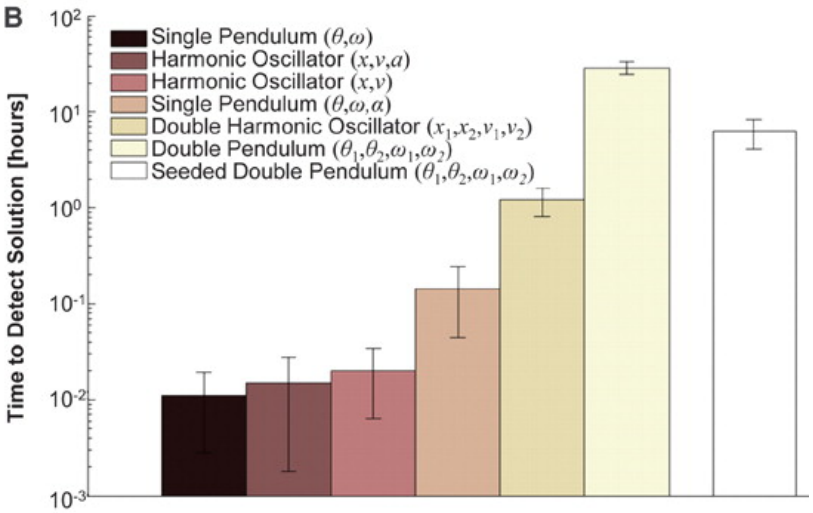


https://www.mql5.com/en/code/9002

- Repetition of: linear transform + nonlinear "activation"
- Universal function approximator
- Scalability/throughput for huge datasets (and many "success stories")
- But learned information is hidden inside 100–10^8 neurons

Ingredient: Symbolic Regression

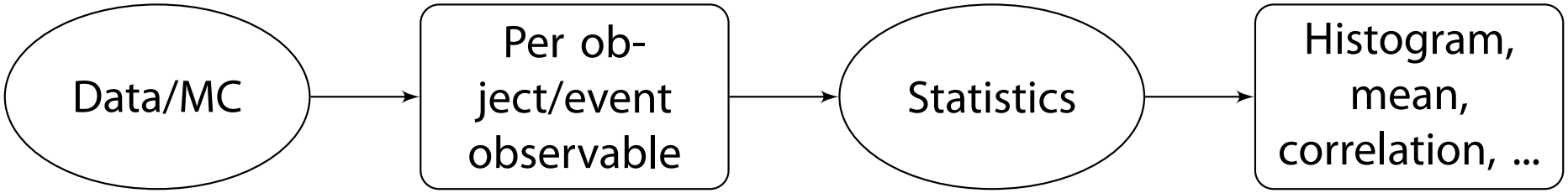




M. Schmidt & H. Lipson, Science. 324(5923), 81 (2009)

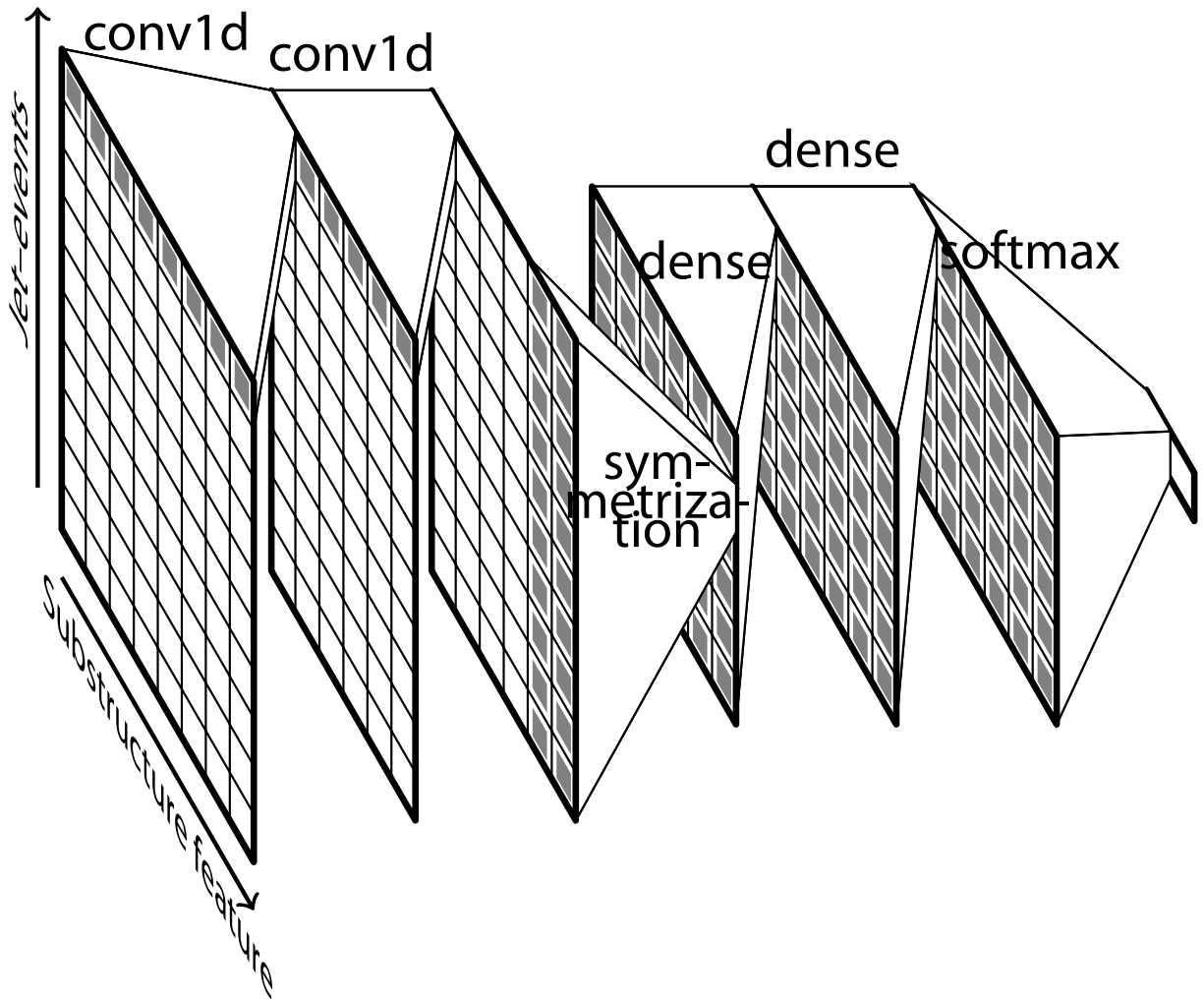
- Genetic algorithm ("mutation" of the equation as a graph) search for the right equation
- Regularize by complexity = number of nodes
- Difficult to apply for high dimension and (HEP/NP-) large data sample
- Schmidt & Lipson criticized for having the exact right phase space for Hamiltonian mechanics as input

Nature of stat. analyses, permutation symmetry



- Analyses are always mathematical functions:
  - Per-event or per-reconstructed object observable
  - Observable is then statistically analyzed (histograms = a series of step-functions, means, ..., correlation, event mixing)
- Key ingredient is permutation symmetry:
  - Regular NN can be made permutation symmetric by symmetric polynomials
  - **New even for NN: previously never attempted large-scale and with more than 1 simultaneous variable**
  - Training works, and each training step corresponds to $N!$ normal NN training $\Rightarrow$ Gigantic speed-up (typically $N \approx 1000$ with a "minimal" analysis)
- See arXiv:1810.00835 for details how to construct them

Layout of the NN that approximates all possible analyses



- conv1d: Per jet/event analysis expressed as 1D convolution
- dense: Fully connected NN layer
- softmax: Transform into probability-like [0, 1]

- A NN that simultaneously builds observables and tests, if a legal (permutation symmetric) analysis can be constructed to use it

Input Monte Carlo

- Jewel and Linearized Boltzmann Transport (LBT) for $\sqrt{s_{NN}} = 5.02$ TeV Pb-Pb 0–10% embedded into HYDJET 1.9 0–10%
- Jet $100 < p_T < 500$ GeV$/c$ selection after UE (UE subtracted) and has the smearing of an actual measurement

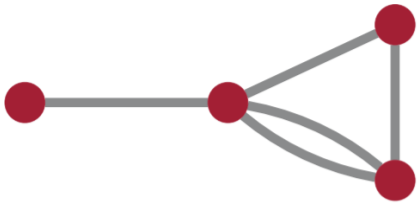
- Jewel with $0.16 \leq T_{\mathrm{i}} \leq 0.76$ GeV$/k$, $0.2 \leq \tau_{\mathrm{i}} \leq 0.8$ fm
- 200k events per T_{i}, τ_{i}, or 3.2M Jewel events total

- LBT with Gubser flow, $0.383 \leq \hat{T} \leq 0.538GeV/k$ (to match Jewel)
- Initial partons + hadronization with PYTHIA 8
- 200k events per $\hat{T}$, or 0.8M LBT events total

- Fully "whitened": jet spectra are reweighted to have no T_{i}, $\hat{T}$ or τ_{i} dependence, Jewel/LBT centrality randomized with HYDJET centrality

Input jet shape/substructure

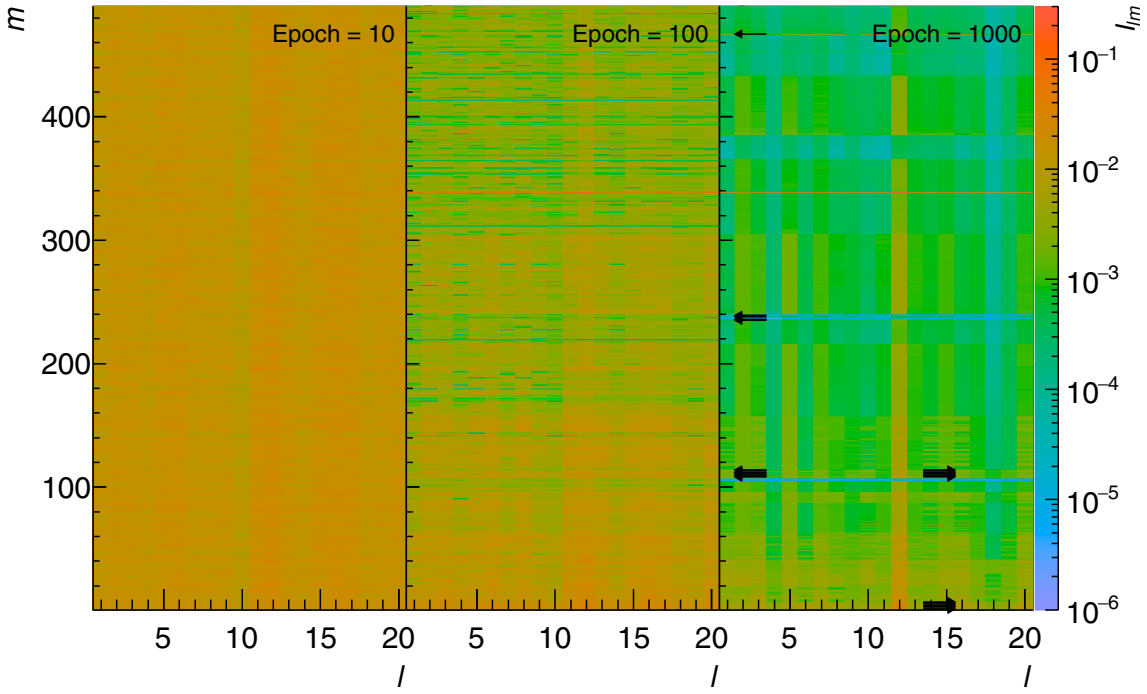
- Energy flow polynomial (EFP) by Komiske, Metodiev, Thaler [J. High Energ. Phys. 04 (2018): 13]
- Each polynomial corresponds to a multigraph $G = (V, E)$
- $\mathrm{EFP}_G = \sum_{j_1} \cdots \sum_{j_{N_V}} \left(z_{j_1} \ldots z_{j_{N_V}} \prod_{(k,l) \in E} \theta_{kl} \right)$
- Set size of E = degree of polynomial
- 490 polynomials up to degree 7, 486 are used (4 remaining takes 6 min per jet to calculate)
- Substructures are not UE subtracted (left for NN)

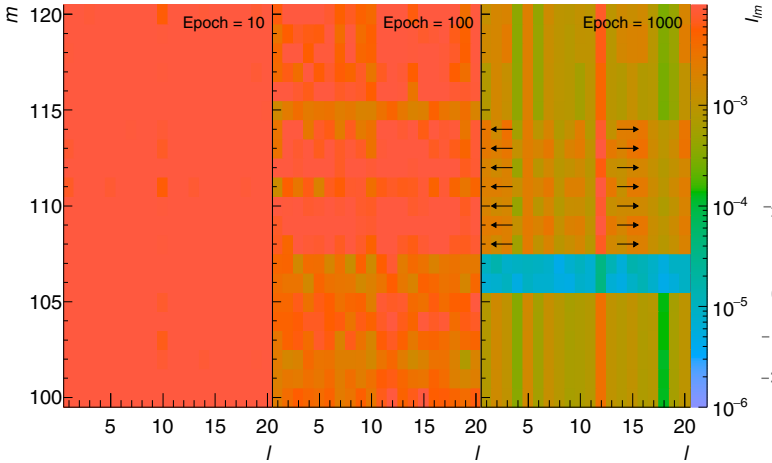


$$= \sum_{i_1=1}^{M} \sum_{i_2=1}^{M} \sum_{i_3=1}^{M} \sum_{i_4=1}^{M} z_{i_1} z_{i_2} z_{i_3} z_{i_4} \theta_{i_1 i_2} \theta_{i_2 i_3} \theta_{i_2 i_4}^2 \theta_{i_3 i_4}$$

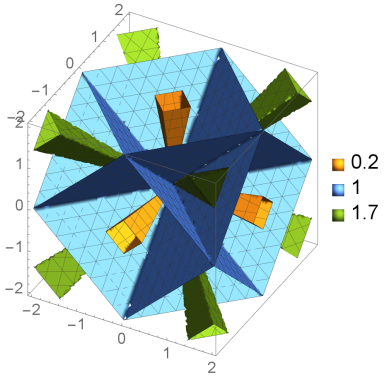
from P. T. Komiske, E. M. Metodiev, J. Thaler, J. High Energ. Phys. 04 (2018): 13

Last ingredient: “Nudge” the NN to simplicity





$R(x_1, x_2, x_3)$ in 3D



See arXiv:1810.00835

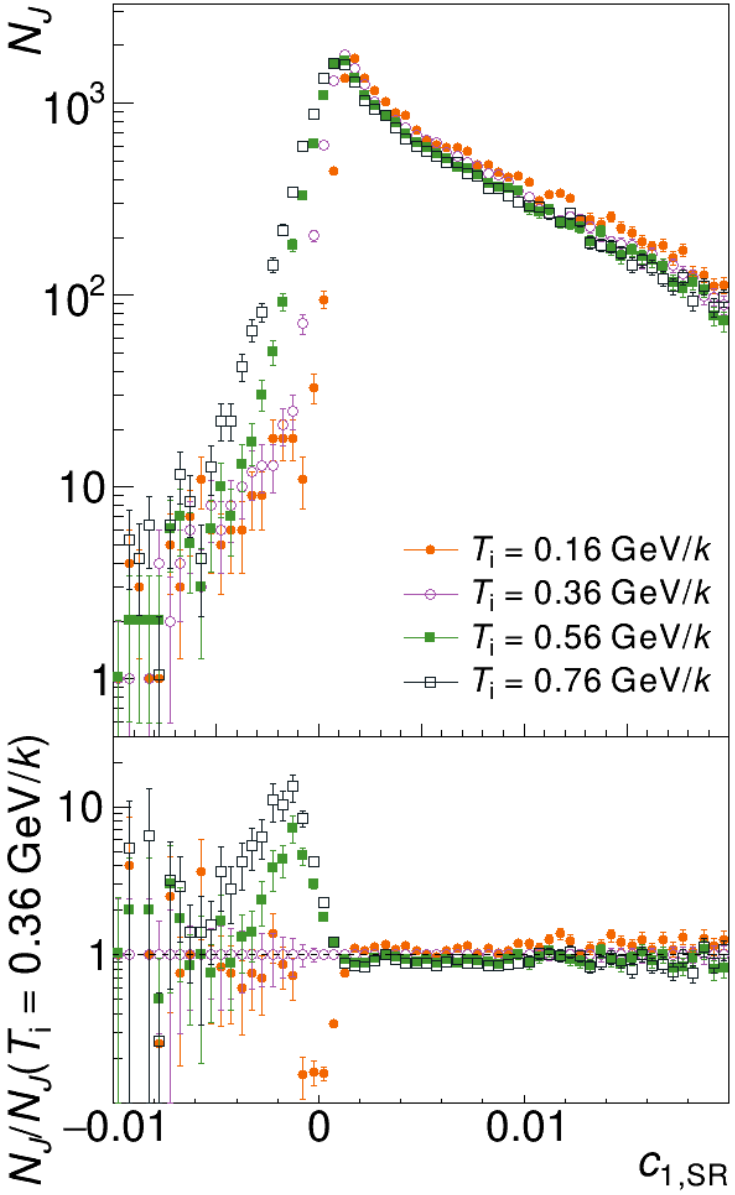
- R takes the bound $l_{lm} = \dfrac{\partial \text{out}_l}{\partial \text{in}_m}$ and counts how many variables are in use
- Adjust the regularization R until the result is simple enough (LASSO)
- **Also new for NN: Interval arithmetic bounds for regularization**
- Regularized NN can then be fed into symbolic regression, with a (weakly) modified FFX [M. Trent, Genetic Programming Theory and Practice IX, Springer (2011)]

Result for Jewel

- $N = 2500$ jets needed to determine T_i
- Overall analysis:
  $d_{1,\mathrm{SR}} = 8.99 - 0.176N\langle c_{16}\rangle - 0.175N\langle c_1\rangle$
- First approximate c_1 (Einstein notation):

$$
\begin{aligned}
c_{1,\mathrm{SR}} = {} & 0.0112 + 45.5\theta_{ab}\theta_{ac}\theta_{bc}\theta_{ad}\theta_{bd}\theta_{ae}\theta_{ce}z_a\cdots z_e + \\
& + 20.9\theta_{ab}\theta_{ac}\theta_{bc}\theta_{ad}\theta_{bd}\theta_{ae}\theta_{be}z_a\cdots z_e + 17.7\theta_{ab}\theta_{ac}\theta_{bd}\theta_{cd}\theta_{ae}\theta_{de}z_a\cdots z_e + \\
& + 8.63\theta_{ab}\theta_{ac}\theta_{bd}\theta_{cd}\theta_{ae}\theta_{de}^2 z_a\cdots z_e - 3.08\theta_{ab}\theta_{ac}\theta_{ad}\theta_{ae}z_a\cdots z_e - \\
& - 1.08\theta_{ab}\theta_{ac}\theta_{ad}^2\theta_{ae}z_a\cdots z_e + 0.769\theta_{ab}\theta_{ac}\theta_{ad}\theta_{ae}\theta_{af}\theta_{ag}\theta_{ah}z_a\cdots z_h - \\
& - 0.233\theta_{ab}^3\theta_{ac}\theta_{ad}\theta_{ae}z_a\cdots z_e - 0.0377\theta_{ab}\theta_{ac}^2\theta_{ad}\theta_{ae}^2 z_a\cdots z_e - \\
& - 0.00483\theta_{ab}\theta_{ac}\theta_{ad}\theta_{ae}^4 z_a\cdots z_e - 0.000508\theta_{ab}\theta_{ac}^3\theta_{ad}^2\theta_{ae}z_a\cdots z_e - \\
& - 4.51\times 10^{-5}\theta_{ab}^2\theta_{ac}\theta_{ad}^2\theta_{ae}^2 z_a\cdots z_e
\end{aligned}
$$

- $c_{1,\mathrm{SR}}$ uses 5th and 7th order ("prongs") correlation
- $c_{1,\mathrm{SR}} < 0$ becomes depopulated with lower temperature
- A feature of Jewel recoil
- The NN never saw any non-recoil events, yet it found out something similar to the groomed jet mass (CMS-HIN-16-024, arXiv:1805.05145)!
- $100 < p_T < 300$ GeV/c plotted

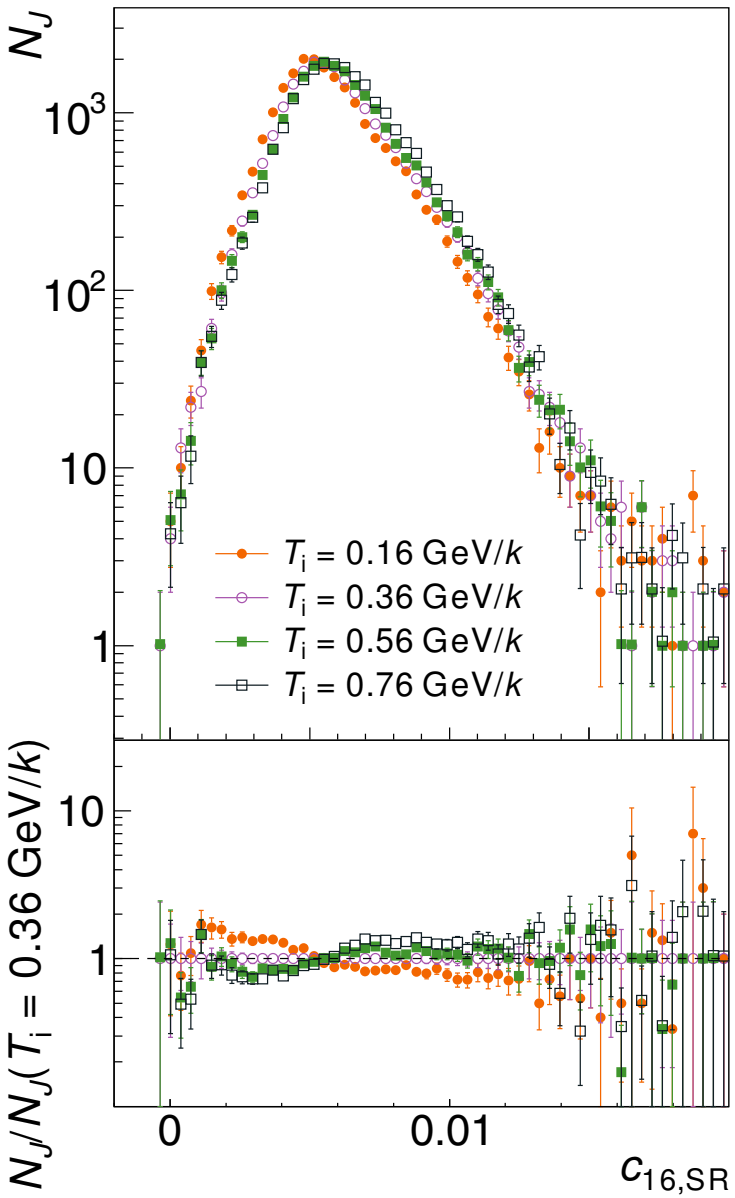


Result for Jewel

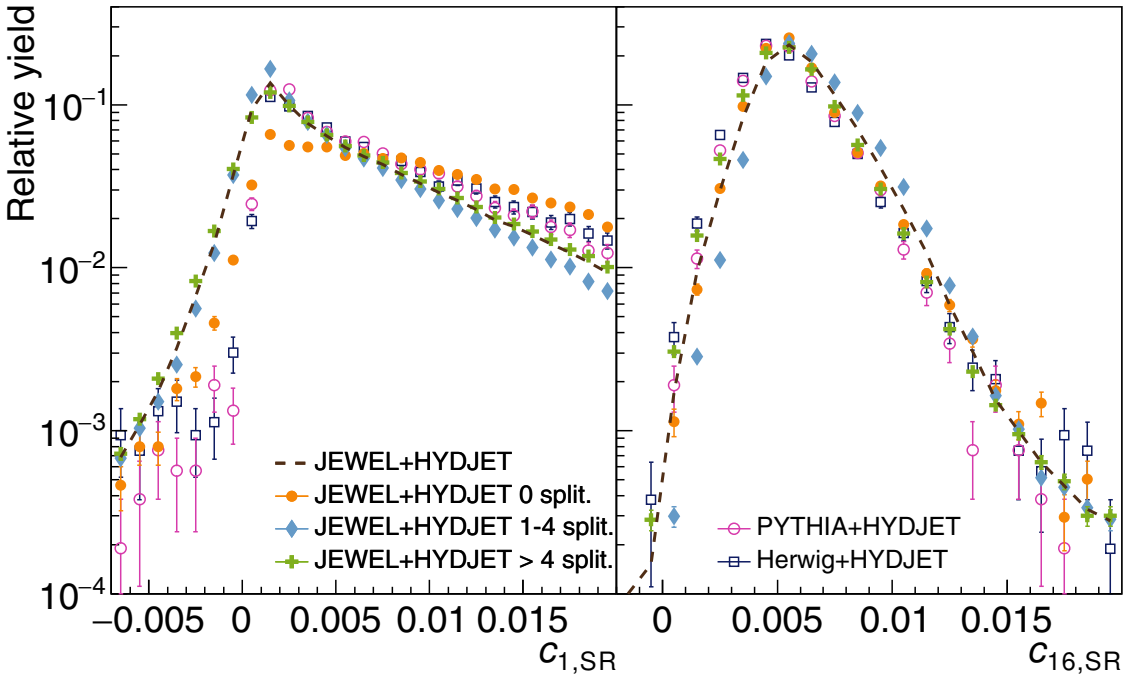
- Approximate c_{16}:

$$
\begin{aligned}
c_{16,\mathrm{SR}} = {} & 0.0362 + 0.594\theta_{ab}^3\theta_{ac}\theta_{ad}\theta_{ae}z_a\cdots z_e + \\
& + 0.575\theta_{ab}\theta_{ac}^2\theta_{ad}\theta_{ae}^2 z_a\cdots z_e + 0.421\theta_{ab}\theta_{ac}\theta_{ad}^2\theta_{ae}z_a\cdots z_e + \\
& + 0.420\theta_{ab}\theta_{ac}\theta_{ad}\theta_{ae}^4 z_a\cdots z_e + 0.246\theta_{ab}\theta_{ac}^3\theta_{ad}^2\theta_{ae}z_a\cdots z_e + \\
& + 0.187\theta_{ab}\theta_{ac}\theta_{ad}\theta_{ae}z_a\cdots z_e + 0.120\theta_{ab}^2\theta_{ac}\theta_{ad}^2\theta_{ae}^2 z_a\cdots z_e - \\
& - 0.0465\theta_{ab}^3 z_a z_b - 0.0453\theta_{ab}^4 z_a z_b - 0.0333\theta_{ab}^5 z_a z_b - 0.0328\theta_{ab}^2 z_a z_b - \\
& - 0.0196\theta_{ab}^6 z_a z_b - 0.0146\theta_{ab} z_a z_b - 0.00963\theta_{ab}^7 z_a z_b
\end{aligned}
$$

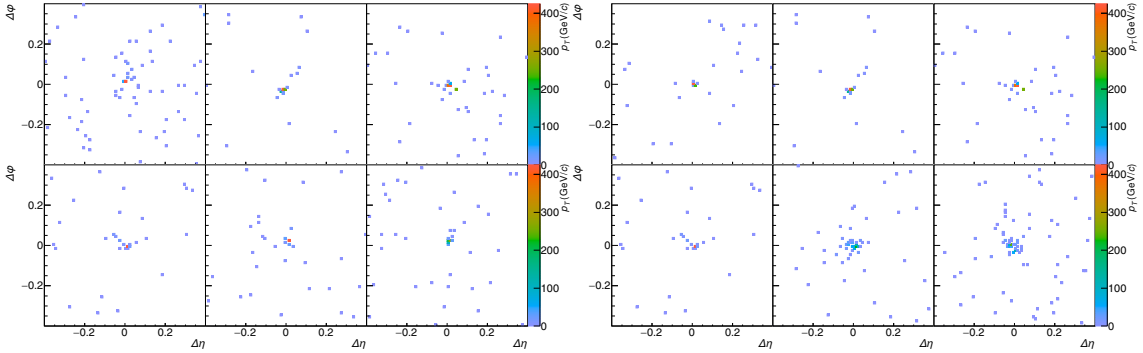
- A gradual, shifting distribution
- Expression contains simple 2-particle correlations, with a nonlinearity $\Rightarrow$ subtraction for average jet expectation
- $100 < p_T < 300$ GeV/c plotted



How does it work?



$c_{1,\mathrm{SR}} < 0.0012$ $\quad$ $0.008 < c_{16,\mathrm{SR}} < 0.012$



Hard event only (for clarity)

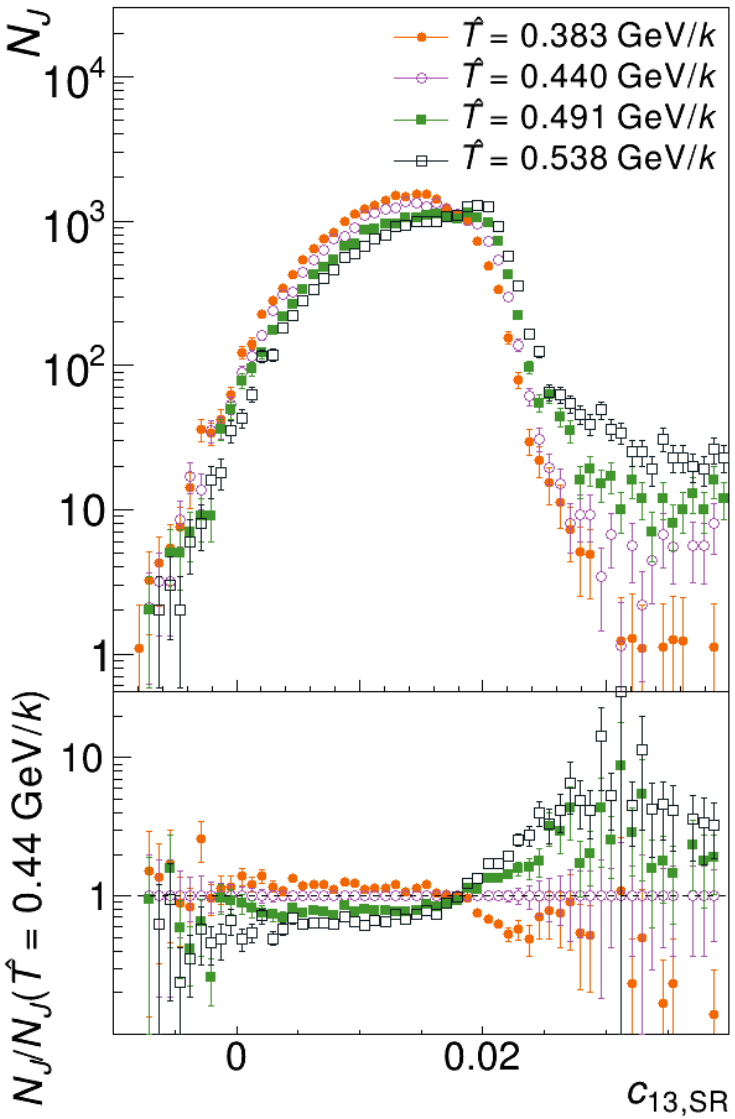
- Jewel was modified to produce (infinitesimal momentum) splitting tags that are clustered into jets
- Count inside each jet the number of splittings in the same area
- $c_{1,\mathrm{SR}}$ and $c_{16,\mathrm{SR}}$ both tagger for no. of internal splittings
- Most of the regions tag few (1–4) splittings
- The $c_{1,\mathrm{SR}} < 0$ is used to tag very high > 4 splittings
- You can also see how $c_{1,\mathrm{SR}} < 0$ tags jets that are rarely produced from PYTHIA 8 (.235, CUETP8M1) and Herwig 7 (.1.1, H7.1-Default)

Result for LBT

- $N = 1200$ jets needed to determine $\hat{T}$
- $d_{1,\mathrm{SR}} = 4.49 - 0.318 N\langle c_{13}\rangle - 0.00653(N\langle c_{13}\rangle)^2$
- Approximate c_{13}:

$$
\begin{aligned}
c_{13,\mathrm{SR}} &= 0.0453 - \\
&\quad - 0.00109 \log_{10}(p_1)(\log_{10}(p_2) + \log_{10}(p_3)) - \\
&\quad - 0.000829 \log_{10}(p_2)\log_{10}(p_3) \\
p_1 &= \theta_{ab}\theta_{ac}\theta_{bc}\theta_{ad}\theta_{bd}\theta_{ae}\theta_{be} z_a \cdots z_e \\
p_2 &= \theta_{ab}\theta_{ac}\theta_{bd}\theta_{cd}\theta_{ae}\theta_{de} z_a \cdots z_e \\
p_3 &= \theta_{ab}\theta_{ac}\theta_{bd}\theta_{cd}\theta_{ae}\theta_{de}^2 z_a \cdots z_e
\end{aligned}
$$

- Demonstrates the system generating non-linear observables
- Possibly log being function of convenience to handle the otherwise long tail (some simplification by hand)
- $100 < p_T < 300$ GeV/c plotted



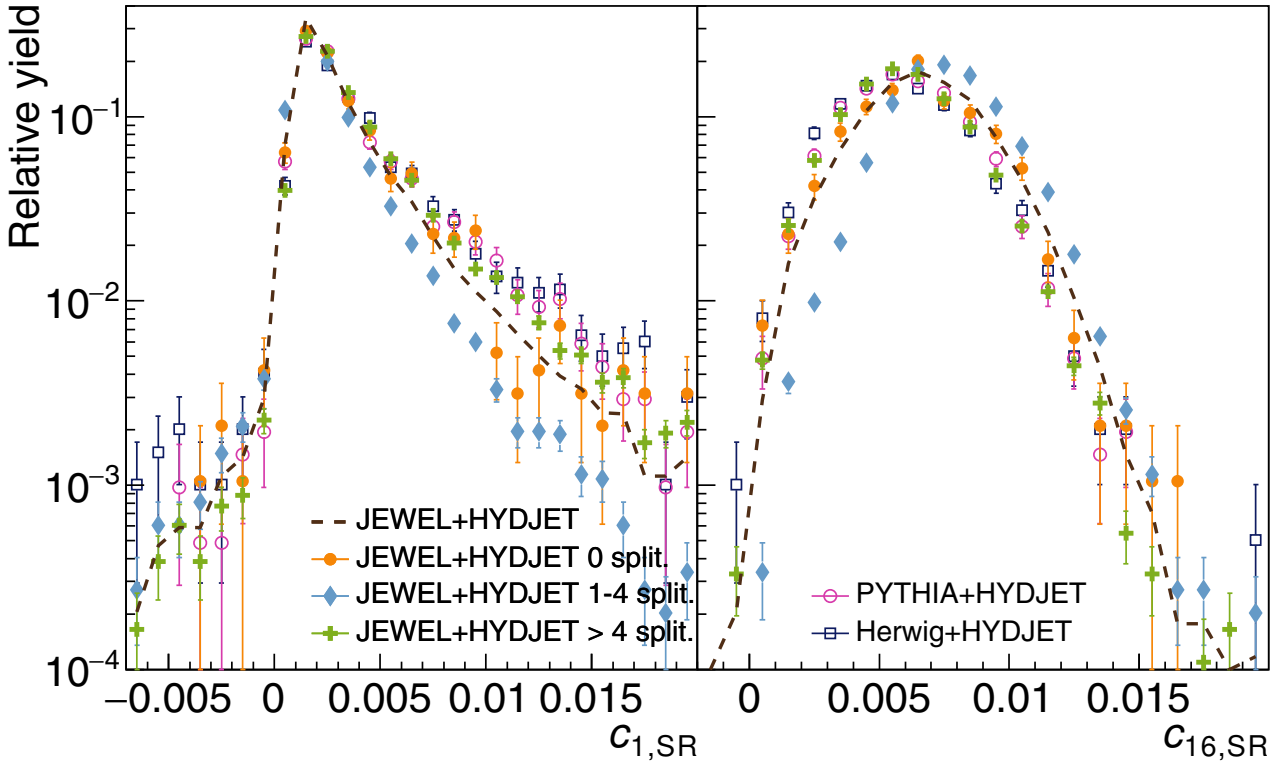
Summary

- First time an end-to-end system is constructed to:
  - Test if jet substructure can be used to independently extract properties of the medium, with realistic no. of events
  - Discover new observables
- NN with properties that are also new for comp. science
- Observables indeed found to extract temperature for Jewel and LBT from purely observing $N = 1200$–2500 jets
- The presented system discovers the effect of Jewel recoil
  $\Rightarrow$ a system probing models at a detail comparable to the current field of human experts
- Many applications beyond the immediate Pb-Pb and jet substructure studies

Part I

Backup

Jewel with no recoil



Permutation symmetry vs. algebra vs. statistics

- Expressibility of permutation symmetry gives the Galois theory of polynomials
- Two well-known types of polynomial:
  - **A** Elementary symmetric polynomials:

$$
\begin{aligned}
e_0(x_1, \ldots, x_n) &= 1 \\
e_1(x_1, \ldots, x_n) &= \sum_{1 \le i \le n} x_i \\
e_2(x_1, \ldots, x_n) &= \sum_{1 \le i < j \le n} x_i x_j \\
&\vdots \\
e_n(x_1, \ldots, x_n) &= \prod_{1 \le i \le n} x_i
\end{aligned}
$$

  - **B** Power sum symmetric polynomials:

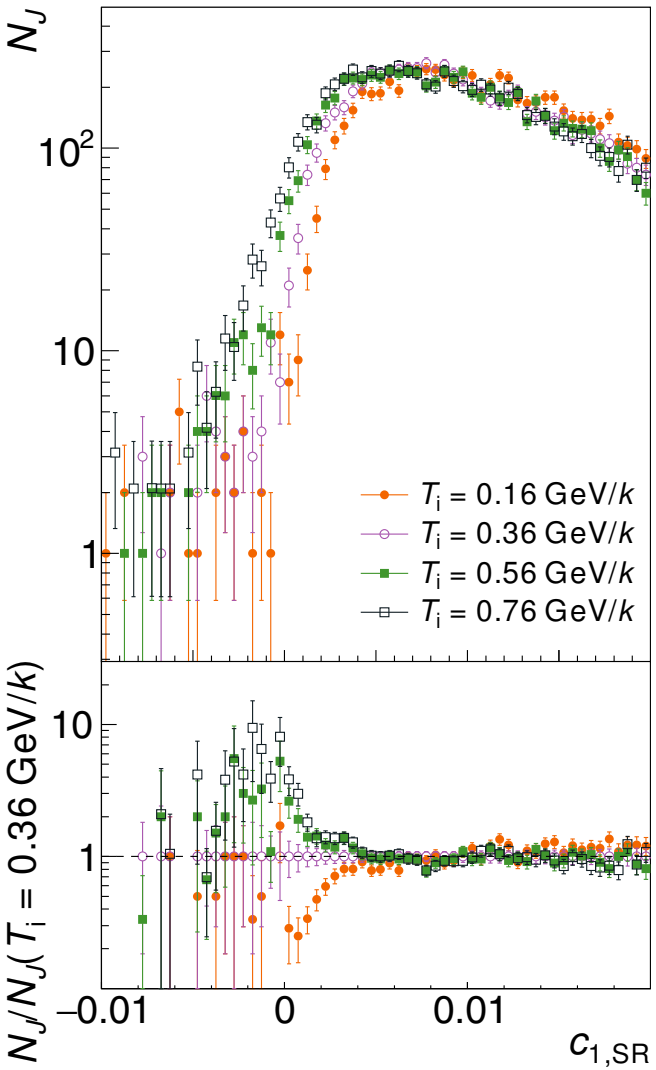
$$
p_k(x_1, \ldots, x_n) = \sum_{1 \le i \le n} x_i^k
$$

- Both are equivalent (Newton's identities), though not equally simple as computational graphs

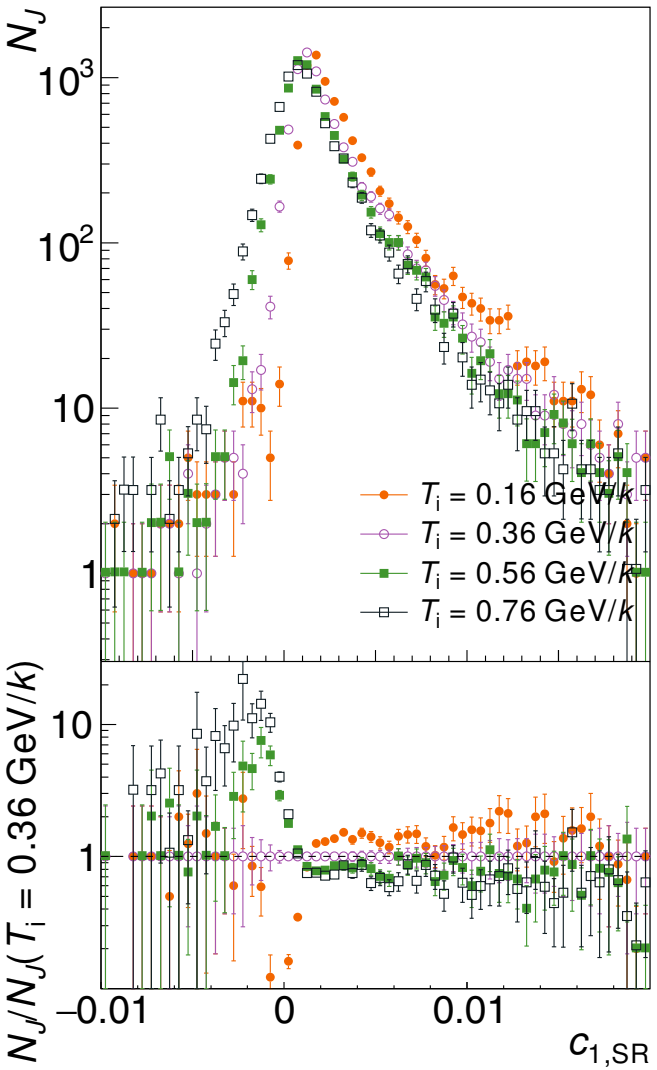
Permutation symmetry vs. algebra vs. statistics

- This still does not help with multivariate functions (where two variables cannot have a mutually different permutation)
- Solution:
  $$s_j(x) = \sum_{k=1}^{N} x_{j\pi(k)}^{m-l} x_{(j+1) \bmod M, \pi(k)}^{l}, \quad l \in \{1, \ldots, m-1\}$$
  for N M-dimensional variables (with a special case for $M = 2$)
- Can check using all M-dimensional symmetric polynomials, that s_j is complete (analogous to Newton's identities)

p_T dependence

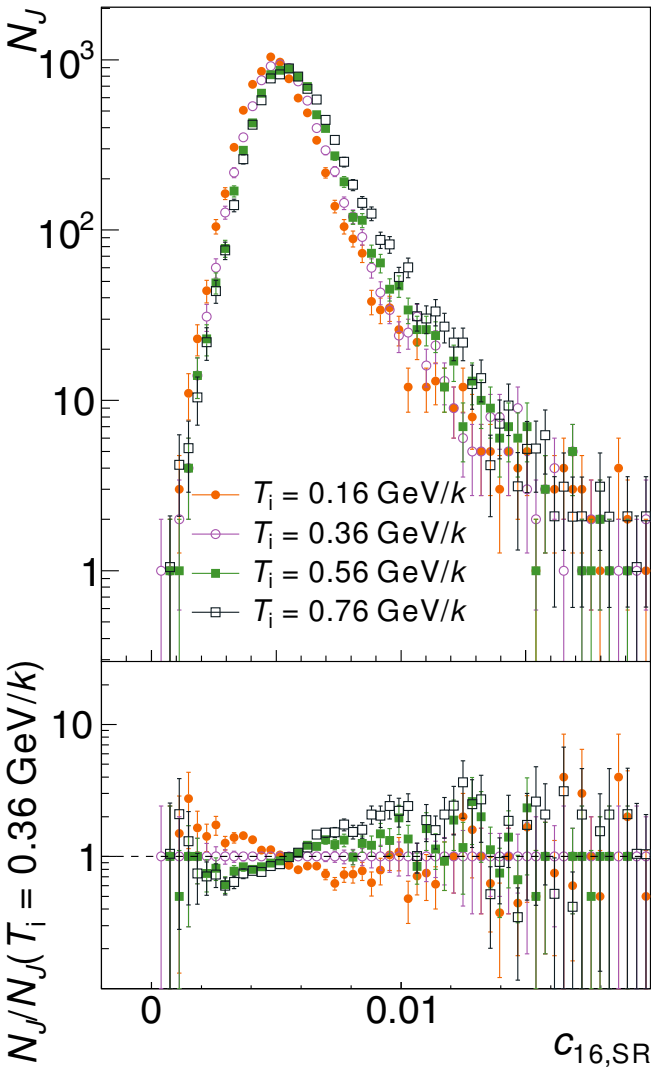


$100 < p_T < 150$ GeV/c

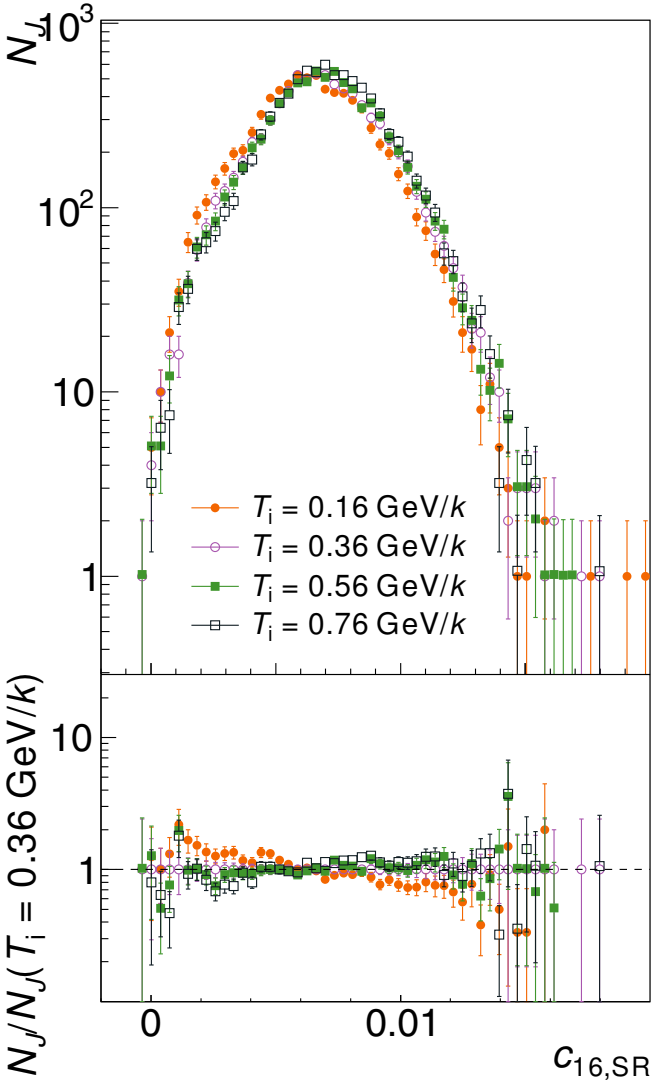


$200 < p_T < 300$ GeV/c

p_T dependence

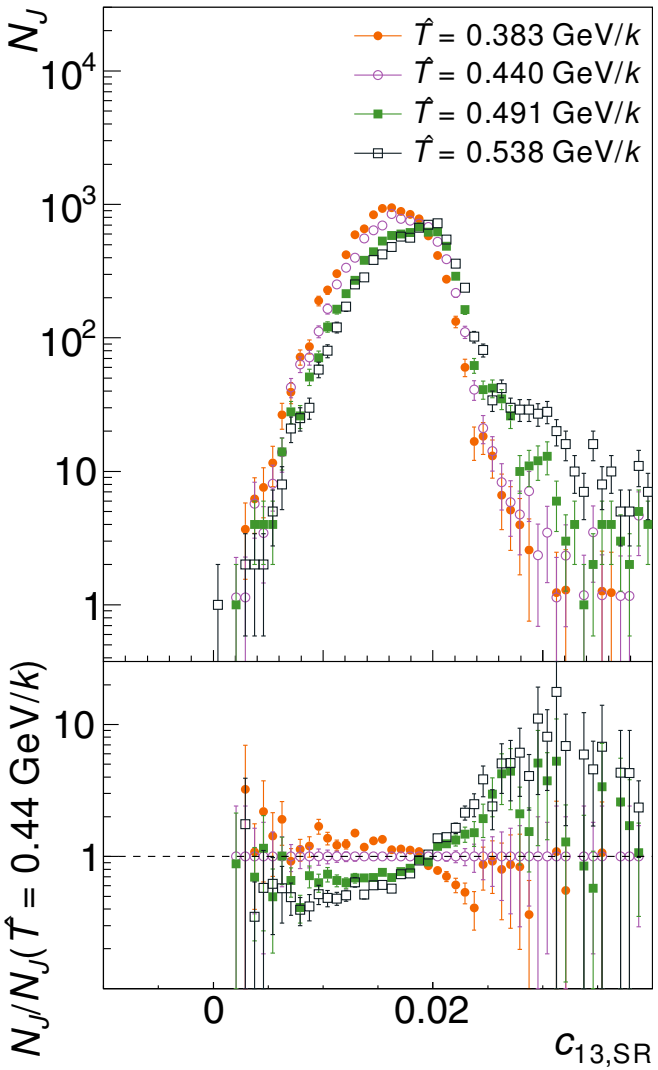


$100 < p_T < 150$ GeV/c

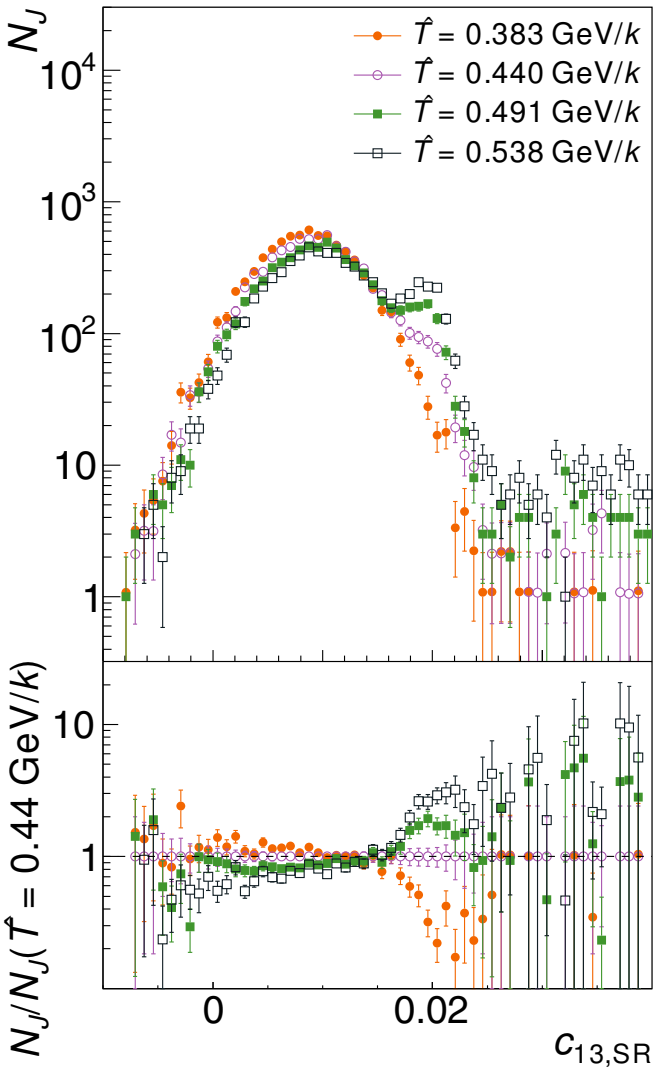


$200 < p_T < 300$ GeV/c

p_T dependence



$100 < p_T < 150$ GeV/c



$200 < p_T < 300$ GeV/c