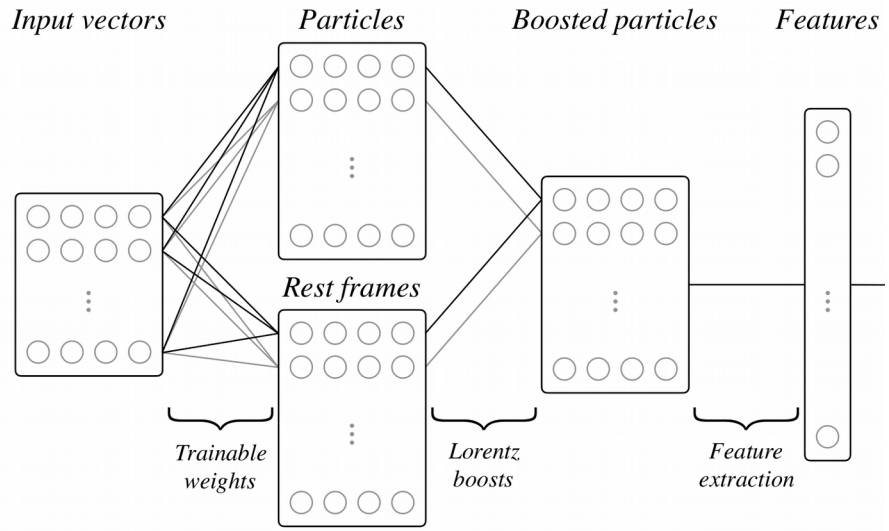
Autonomous Physics-Inspired Feature Engineering with Lorentz Boost Networks

Y. Rath, M. Erdmann, E. Geiser, M. Rieger

III. Physikalisches Institut A, RWTH Aachen University

12.03.19

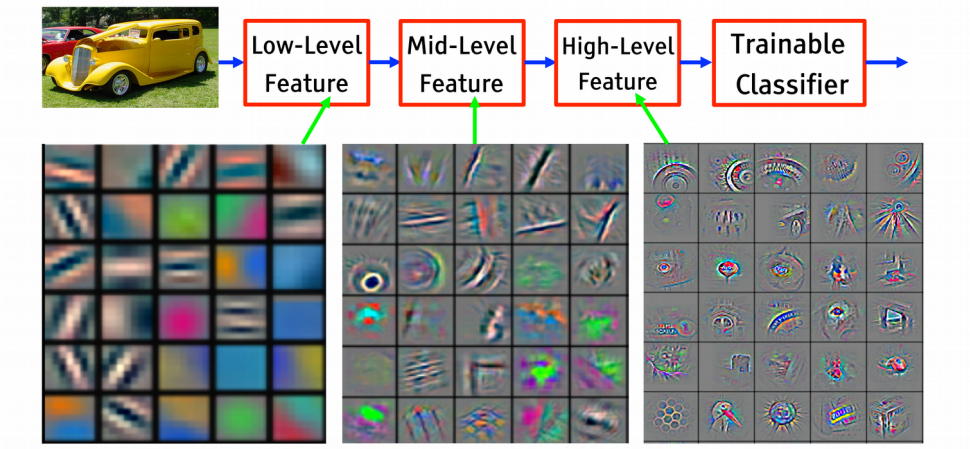


SPONSORED BY THE
Federal Ministry of Education and Research

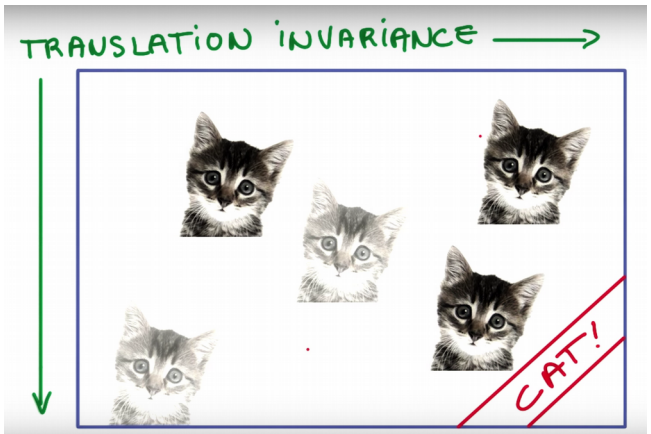
III. Physikalisches Institut

RWTH AACHEN UNIVERSITY

- Wishlist:
  - Model requiring **only low level data** (e.g. particle four momenta)
  - **Interpretable model**
- Computer science: Dedicated architectures exploiting problem structure

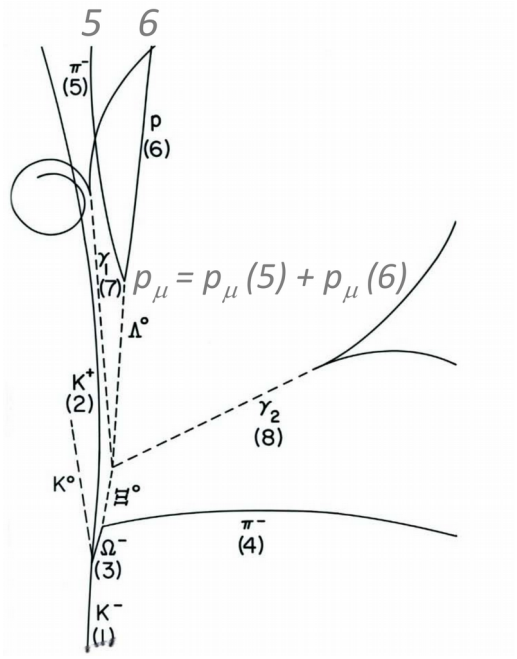


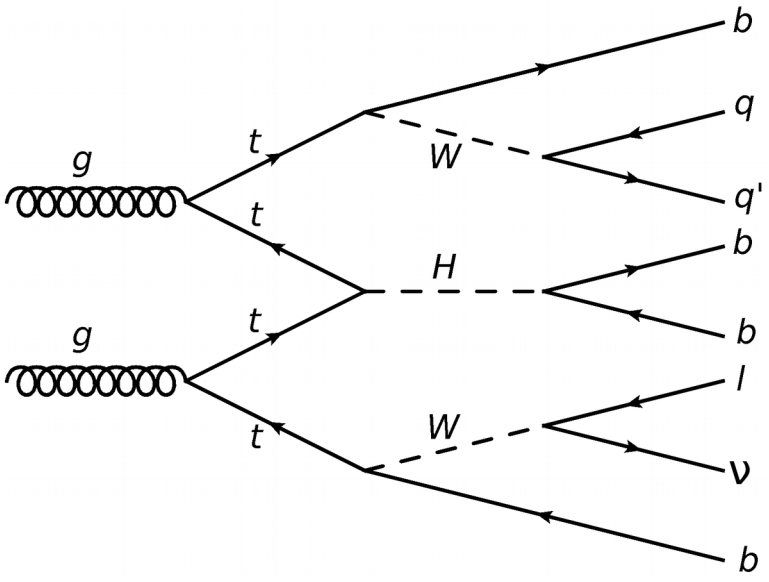
[Zeiler & Fergus 2013], adapted by Yann LeCun



Udacity Course 730, Deep Learning

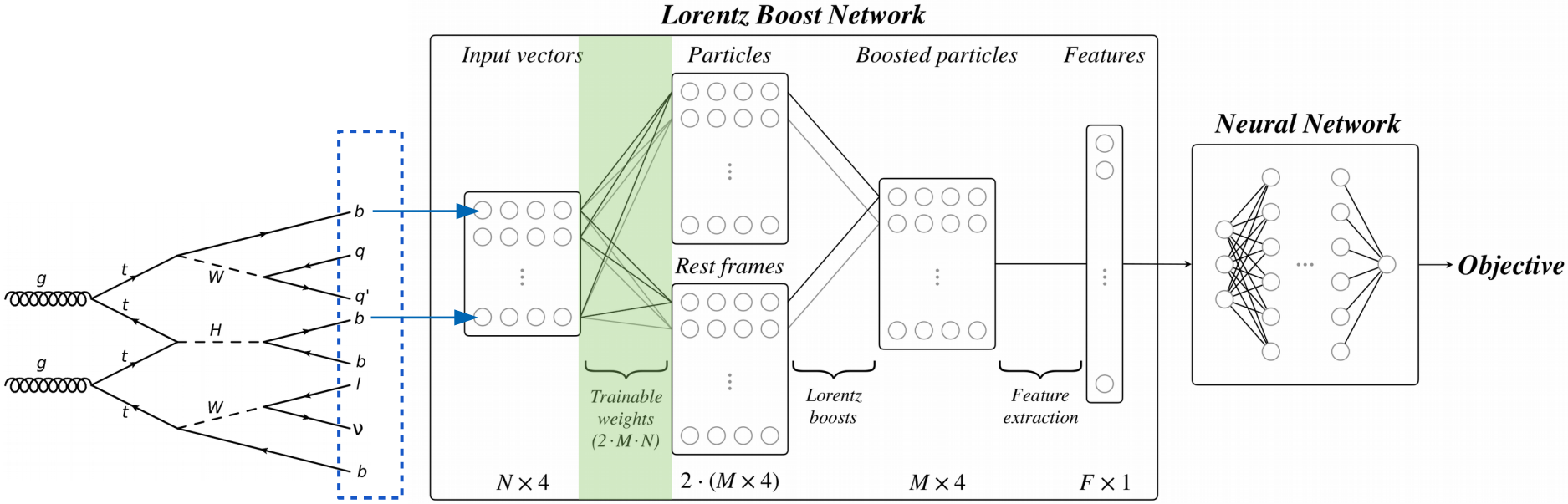
- Recently many similar developments in physics (e.g. LoLa, Recursive NN)
- Our model: Lorentz Boost Network (LBN)





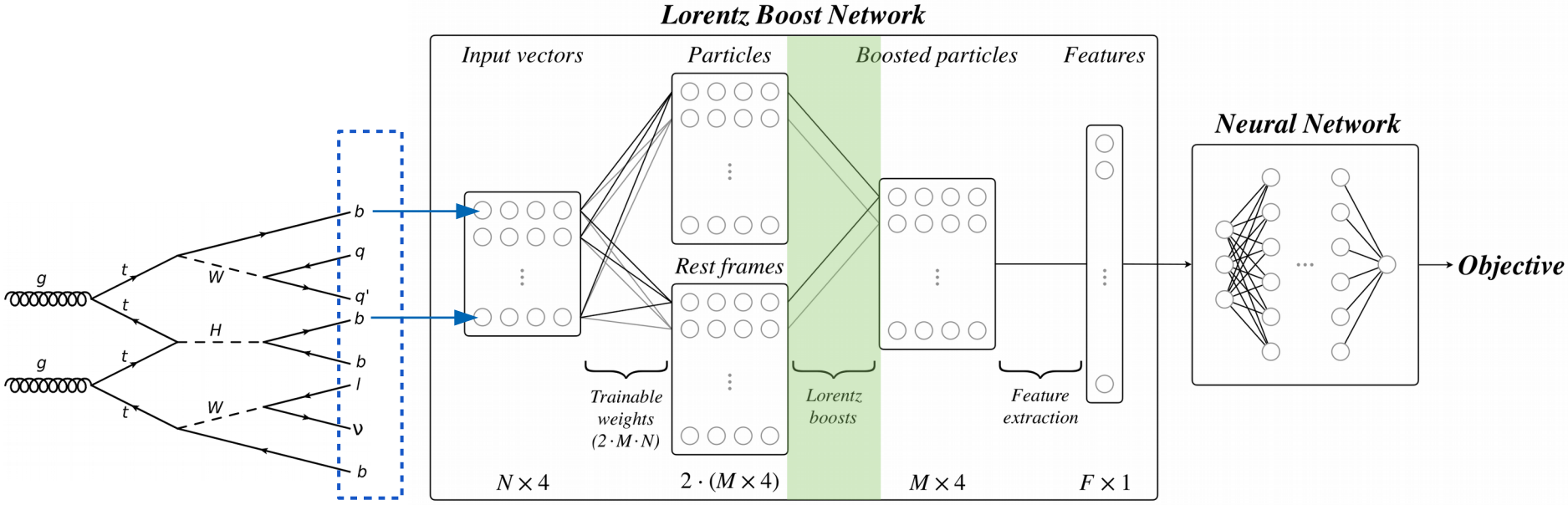
- Physicist: Combine four momenta of final state particles to build intermediate particles
  - Create characteristic variables
    → Analyse decays in rest frame of parent particle
- **Can we incorporate this into a network architecture?**

III. Physikalisches Institut
RWTH AACHEN UNIVERSITY
Yannik Rath – 12.03.19



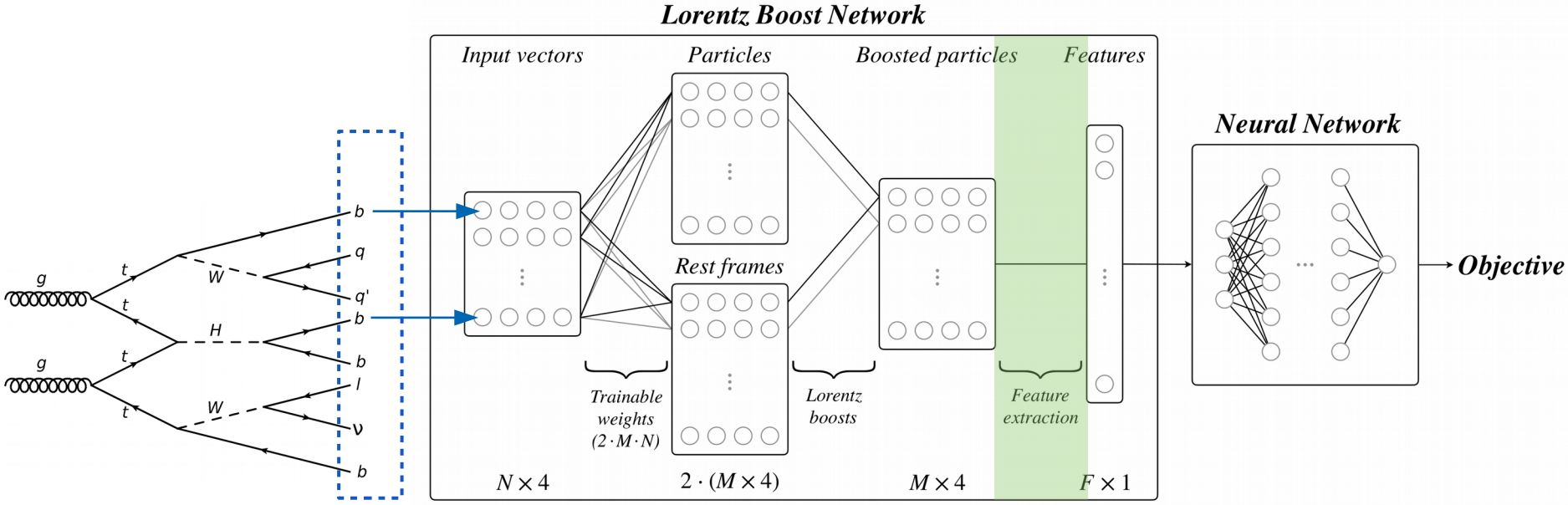
- Create composite particles and rest frames (trainable)

III. Physikalisches Institut
RWTH AACHEN UNIVERSITY
Yannik Rath – 12.03.19



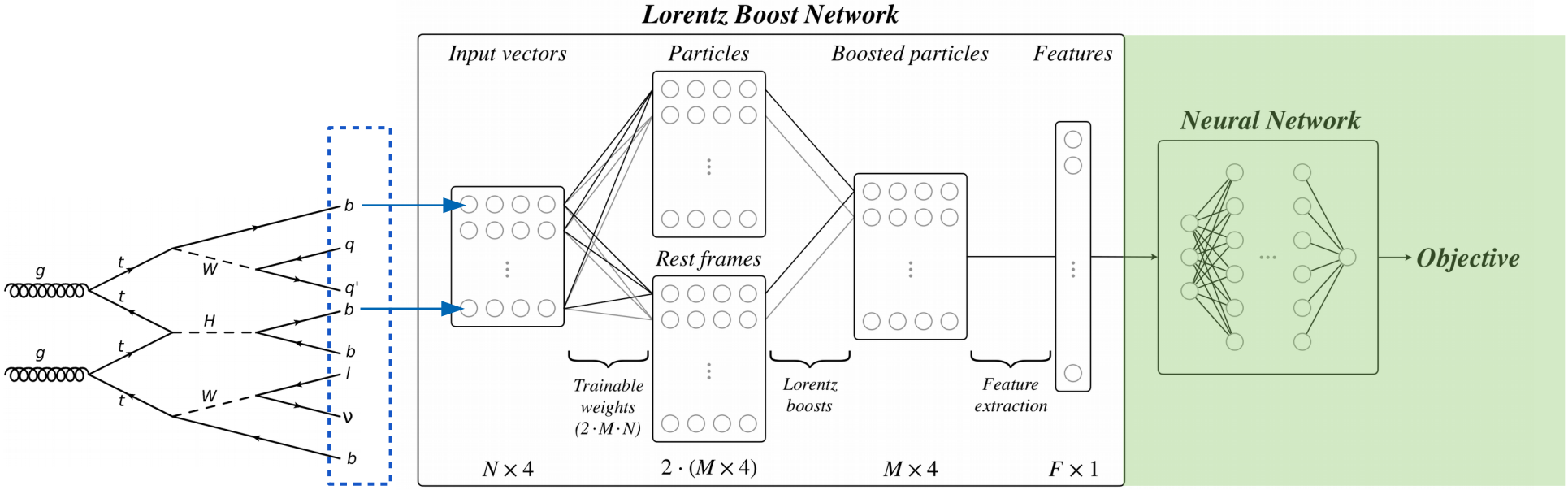
- Create composite particles and rest frames (trainable)
- Perform Lorentz transformation

III. Physikalisches Institut
RWTH AACHEN UNIVERSITY
Yannik Rath – 12.03.19



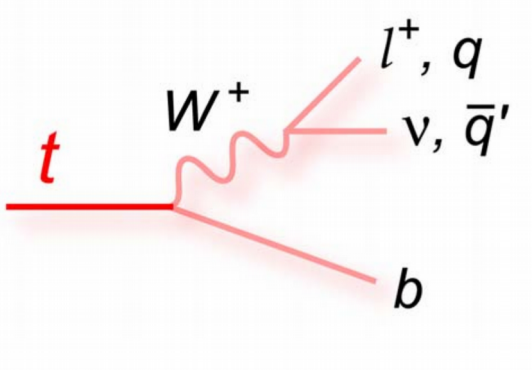
- Create composite particles and rest frames (trainable)
- Perform Lorentz transformation
- Project generic set of features, e.g. energies, pairwise angles

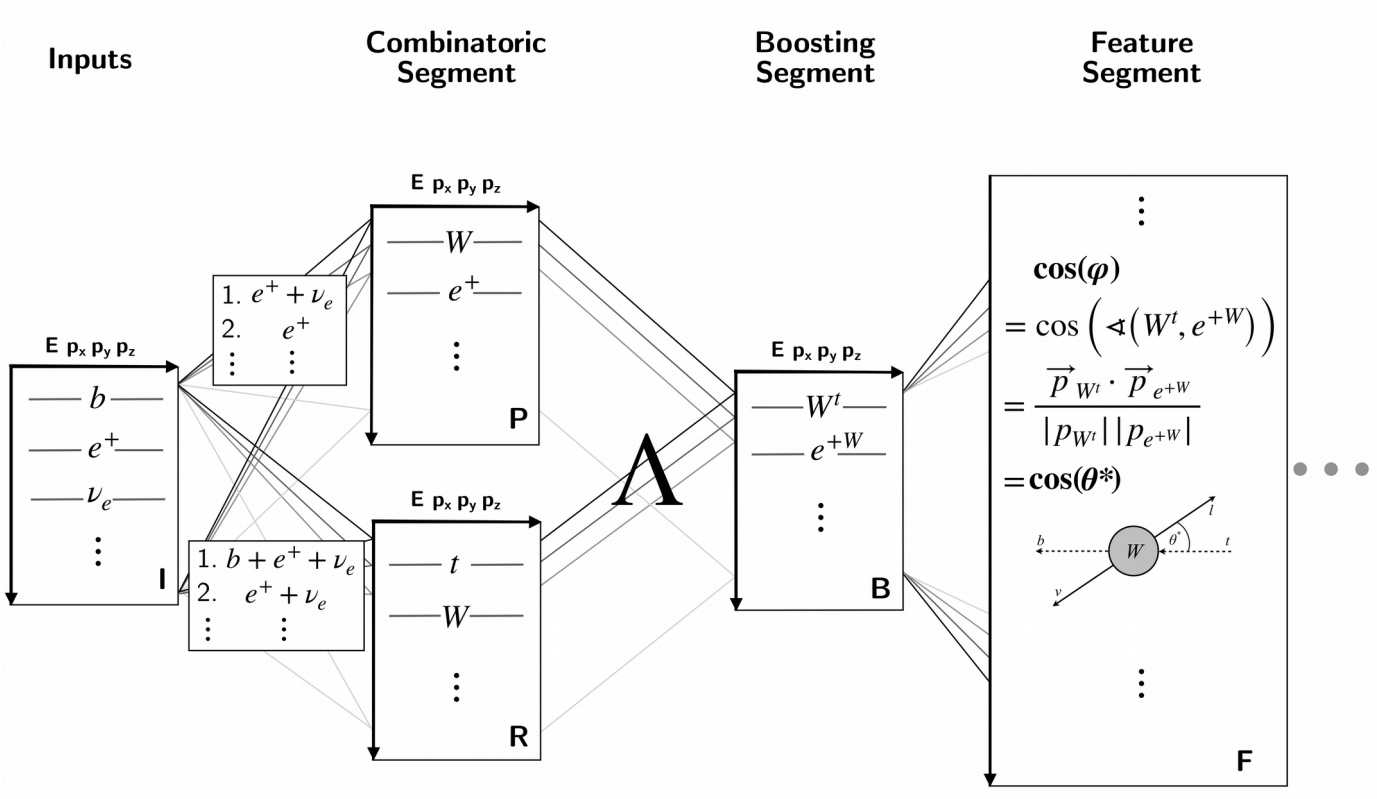
III. Physikalisches Institut
RWTH AACHEN UNIVERSITY
Yannik Rath – 12.03.19



- Create composite particles and rest frames (trainable)
- Perform Lorentz transformation
- Project generic set of features, e.g. energies, pairwise angles
- Use learned features to solve physics problem

- Example of feature construction: helicity angle cos(θ*)



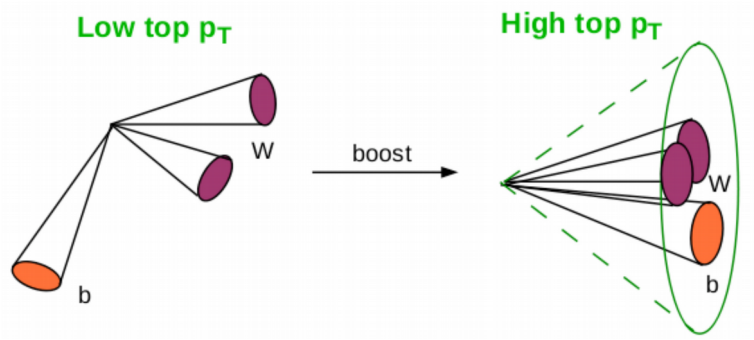


Application

- Currently studying two applications

Top Tagging

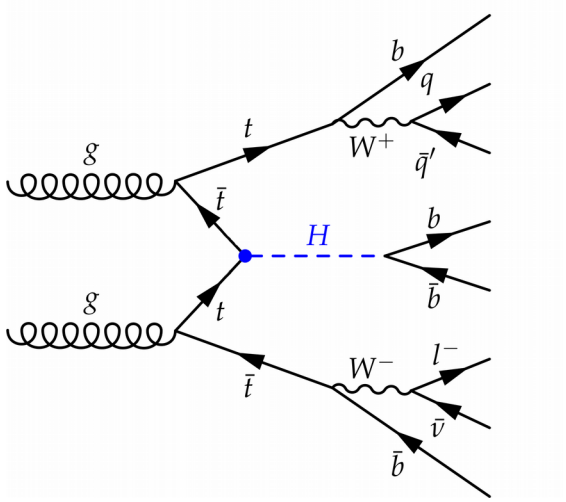
- Distinguish hadronically decaying, highly boosted top quarks from QCD background



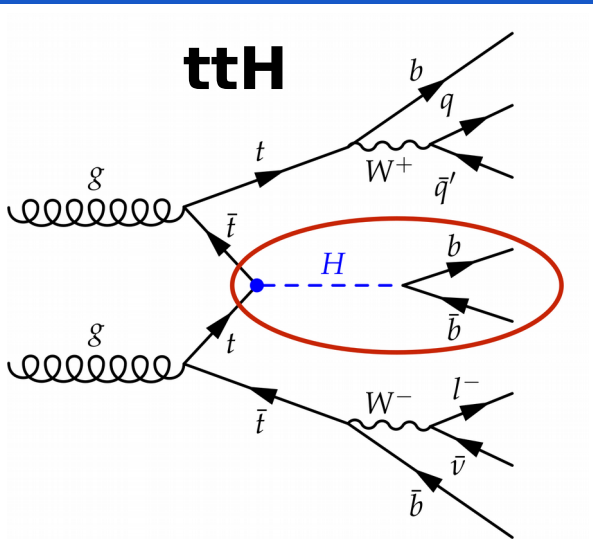
[Method comparison paper: https://arxiv.org/abs/1902.09914]

ttH vs tt+bb

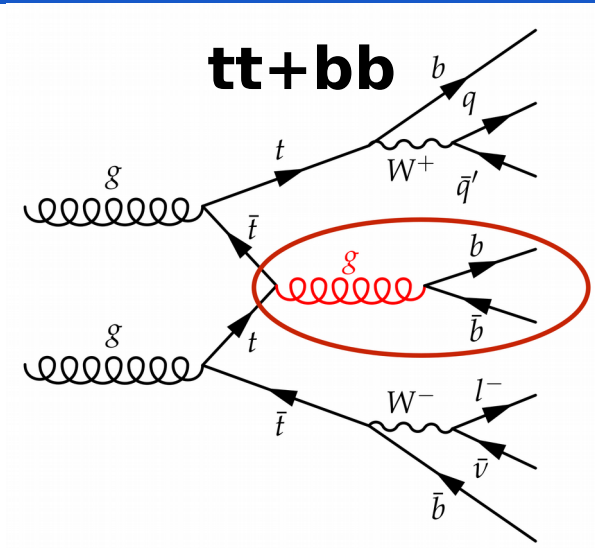
- Separate ttH(H→bb) events from tt+bb background



[https://arxiv.org/abs/1812.09722]



vs.



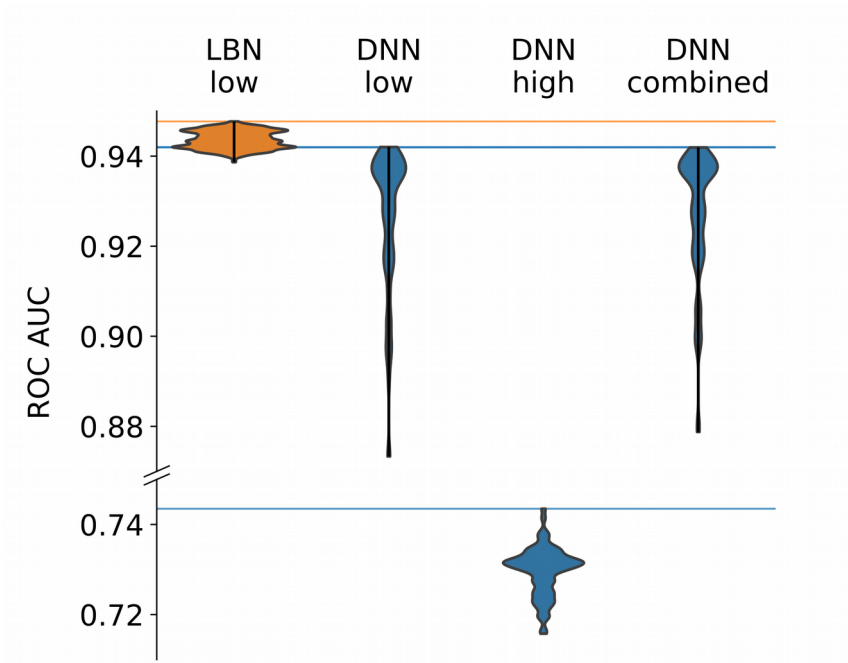
- Final state: 6 jets, 1 charged lepton, 1 neutrino (missing transverse energy)
  → 8 four vectors as input

- Compare LBN against fully connected neural network
  - Trained with same input, with dedicated high-level variables, or both

- Hyperparameter scan with O(100) trainings for each configuration

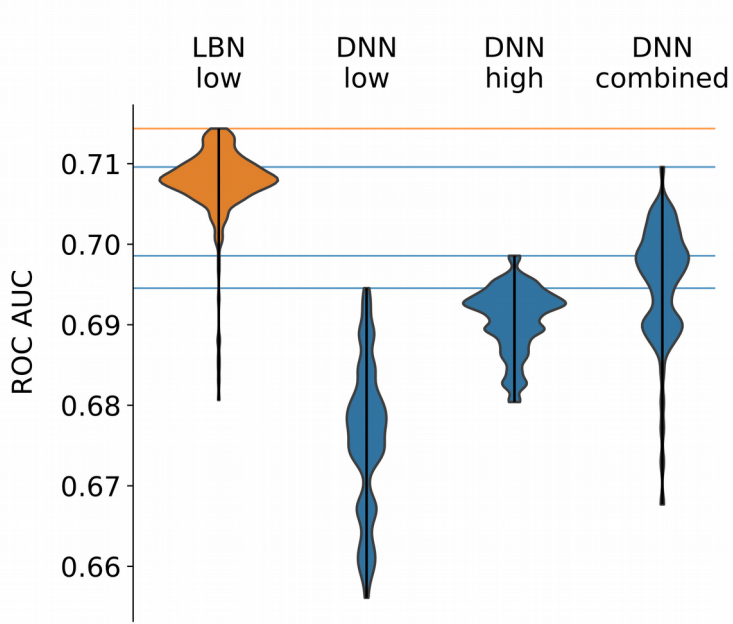
Generator (truth) ordering

→ Isolated LBN performance study



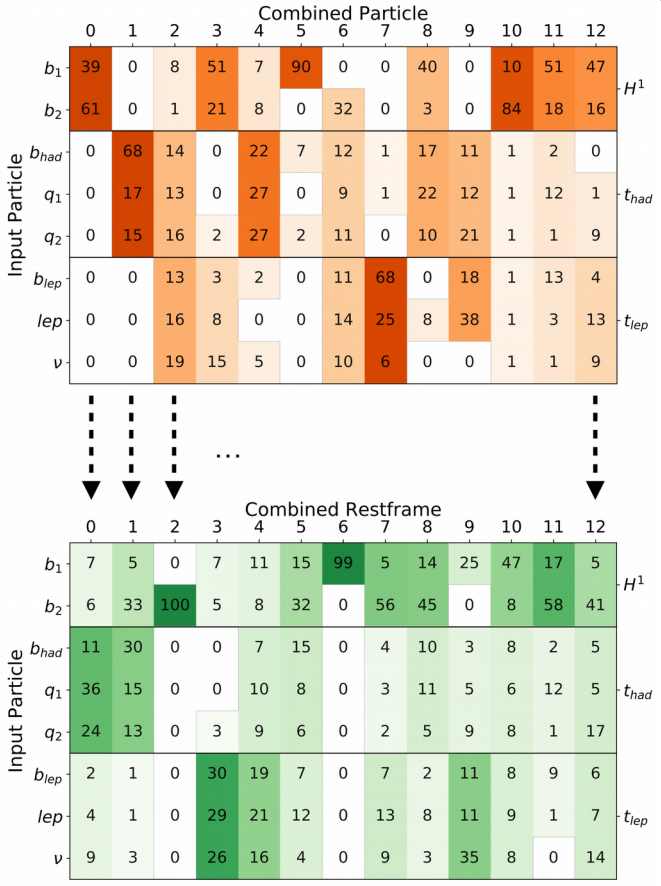
Jet p_T sorting

→ Realistic application



What is learned by the network?

- Study particle combination weights (generator sorting) to gain insights into what the network is learning

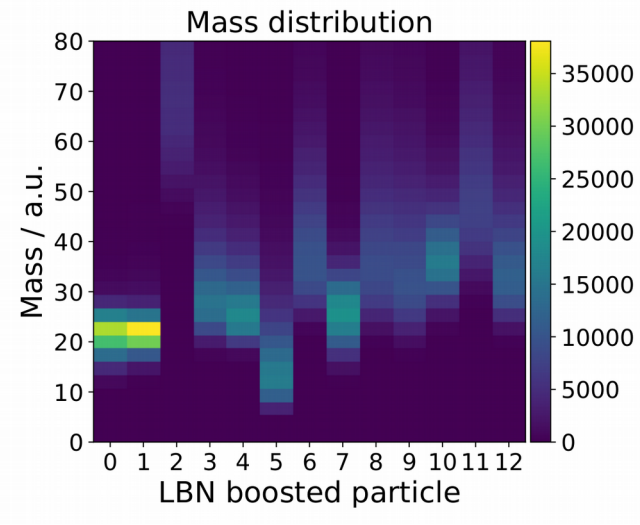


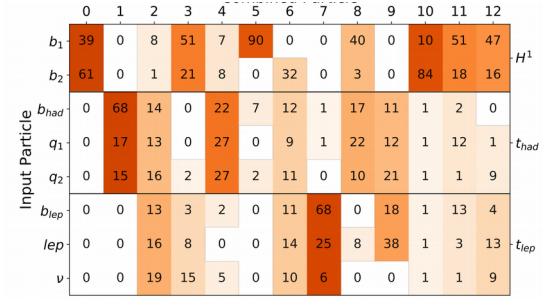
- Approximate construction of intermediate particles by LBN
- Focus on Higgs boson (difference between processes)
- **Which features relevant for the separation?**

Feature Example (Mass)

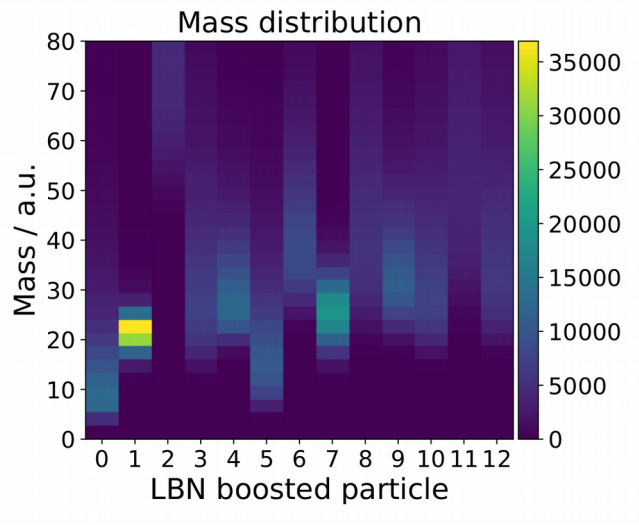
III. Physikalisches Institut
RWTH AACHEN UNIVERSITY
Yannik Rath – 12.03.19

ttH

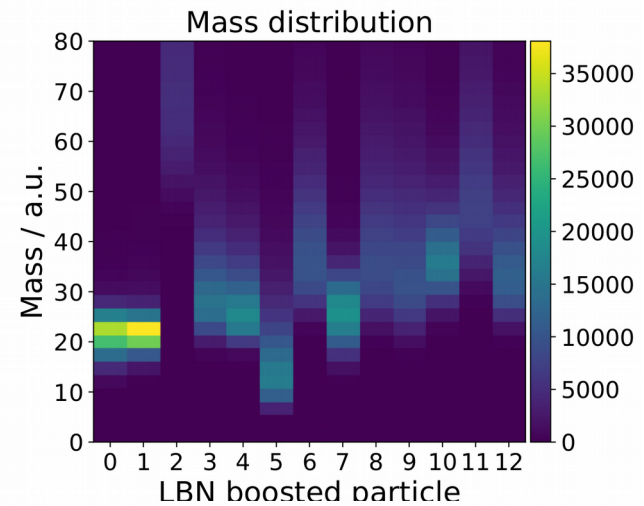


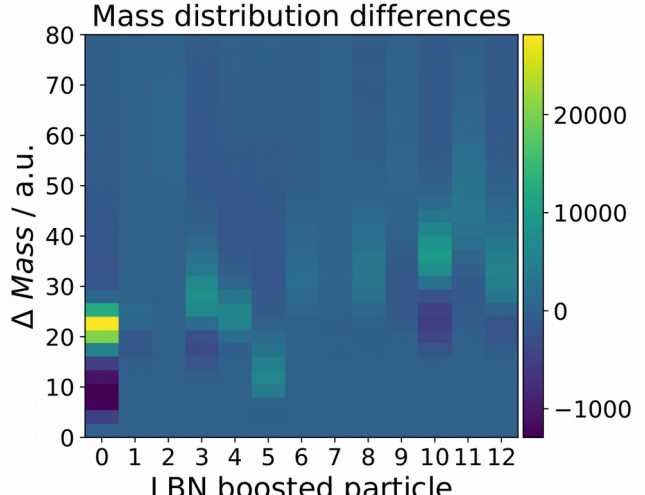


tt+bb

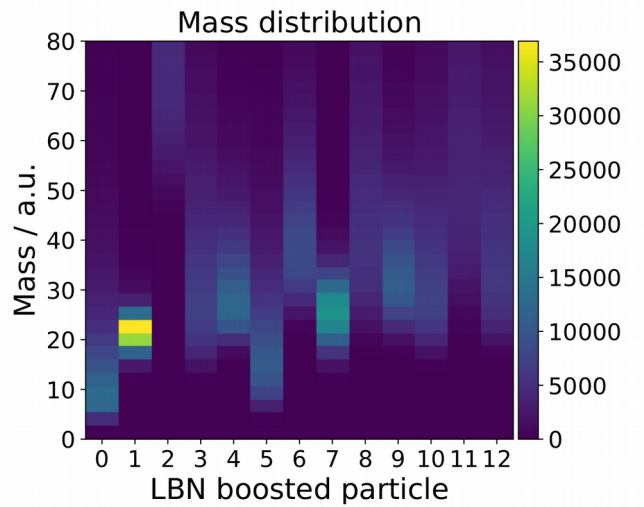


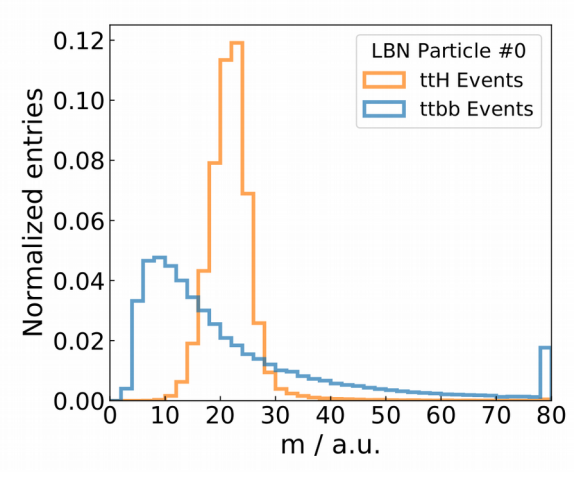
ttH


Mass distribution
Mass / a.u.
LBN boosted particle


Mass distribution differences
Δ Mass / a.u.
LBN boosted particle

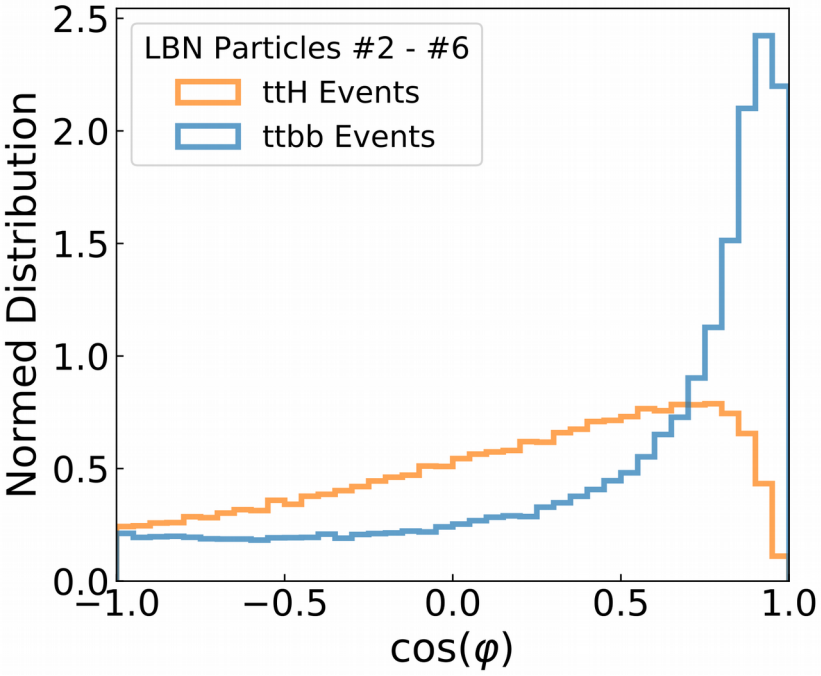
tt+bb

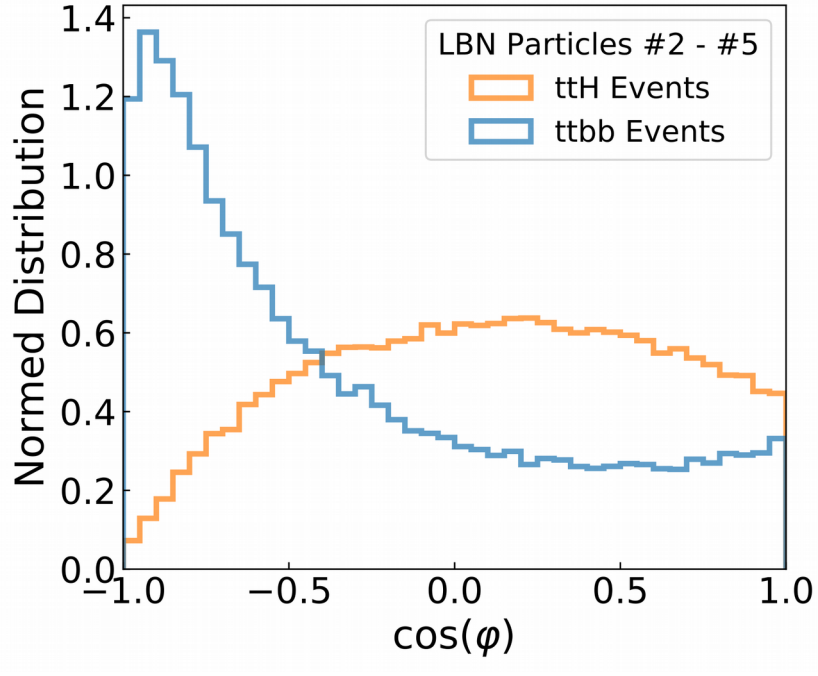

Mass distribution
Mass / a.u.
LBN boosted particle

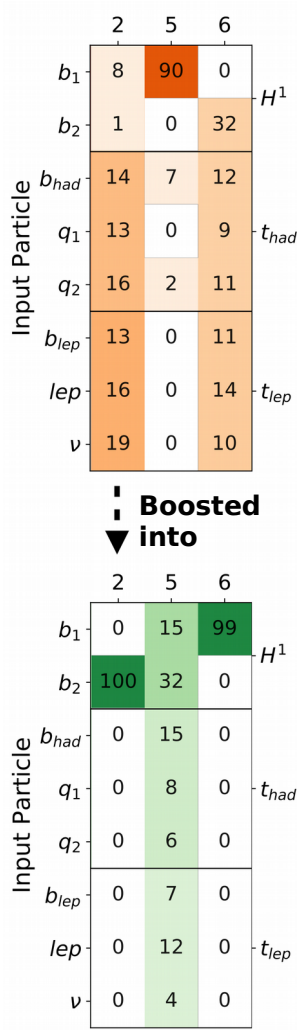

LBN Particle #0
ttH Events
ttbb Events
Normalized entries
m / a.u.

- Mass only dependent on combined particle

  → More complex example: Pairwise angles





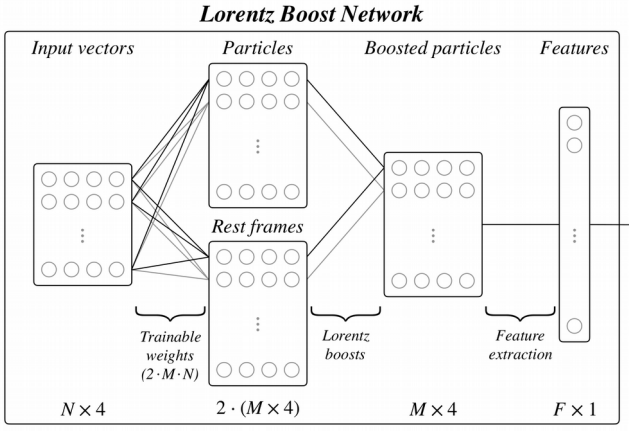


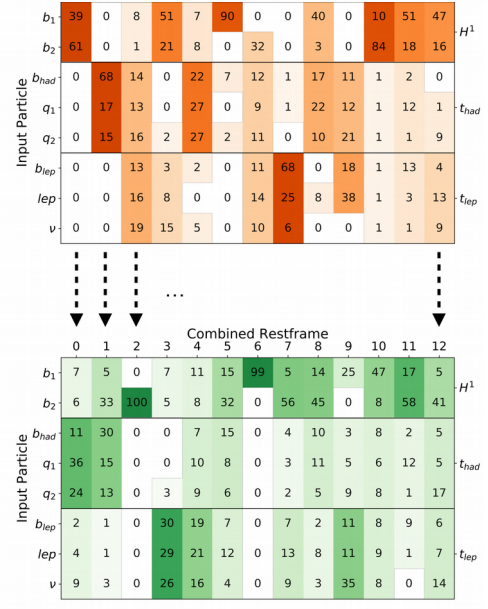
- Lorentz Boost Network: Physics-motivated NN architecture

- Able to autonomously create characteristic features from particle four momenta

- Network components have direct physics interpretation
  - Finds meaningful particle relations, constructs intermediate particles

[https://git.rwth-aachen.de/3pia/lbn]

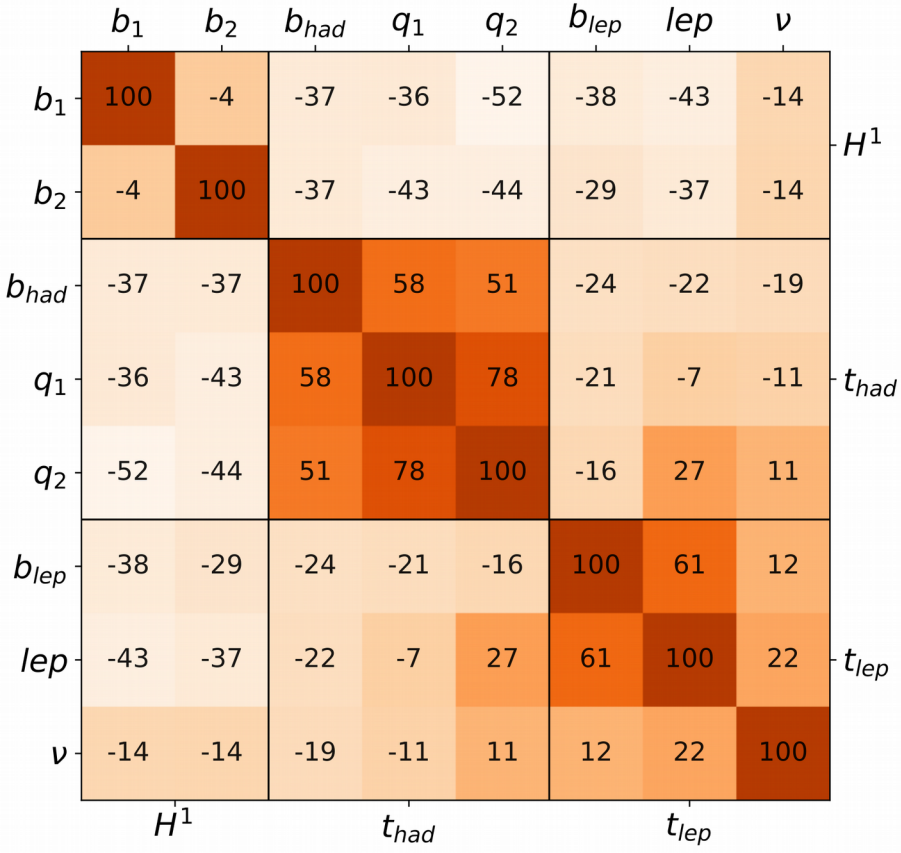




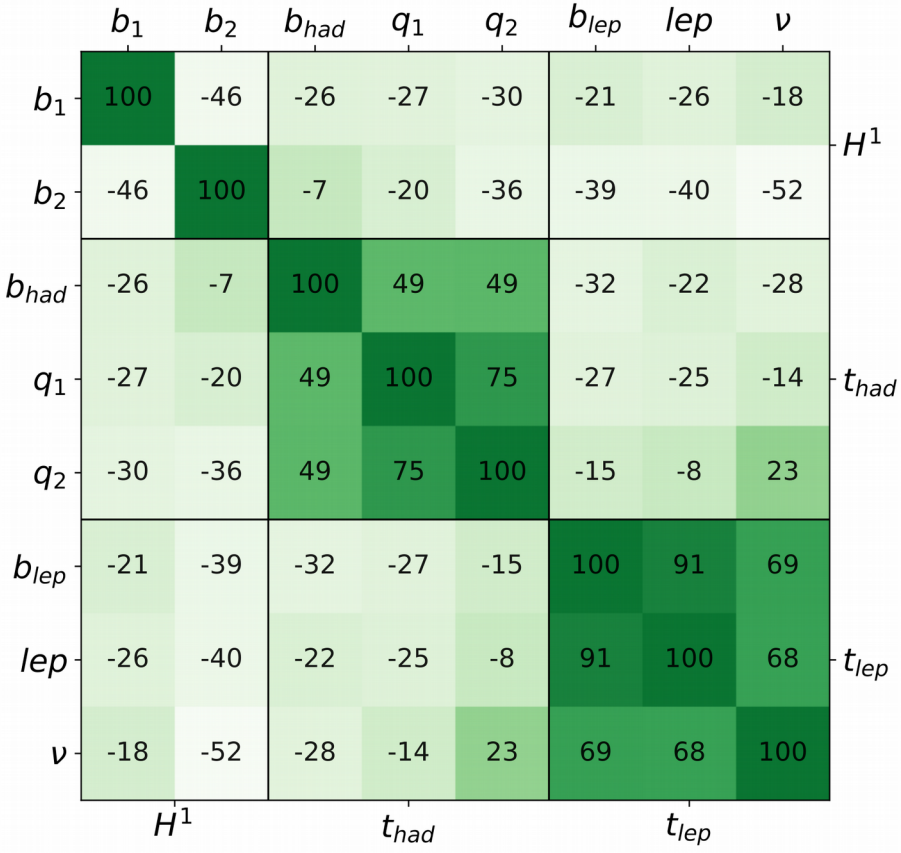
Backup

Correlations

Combined to Particles



Combined to Restframes



Lorentz Boost

$$\Lambda = \begin{bmatrix} \gamma & -\gamma\beta n_x & -\gamma\beta n_y & -\gamma\beta n_z \\ -\gamma\beta n_x & 1+(\gamma-1)n_x^2 & (\gamma-1)n_x n_y & (\gamma-1)n_x n_z \\ -\gamma\beta n_y & (\gamma-1)n_y n_x & 1+(\gamma-1)n_y^2 & (\gamma-1)n_y n_z \\ -\gamma\beta n_z & (\gamma-1)n_z n_x & (\gamma-1)n_z n_y & 1+(\gamma-1)n_z^2 \end{bmatrix}$$

with $\vec{n} = \vec{\beta}/\beta$

- As 4D tensor (batch x particle x 4 x 4):

$$\Lambda = I + (U \oplus \gamma) \odot ((U \oplus 1) \cdot \beta - U) \odot (e \cdot e^T)$$

with $U = \begin{bmatrix} -1^{1\times1} & 0^{1\times3} \\ 0^{3\times1} & -1^{3\times3} \end{bmatrix}$ $e = \begin{bmatrix} 1^{1\times1} \\ -\vec{n}^{3\times1} \end{bmatrix}$

Network Parameters

Network	**LBN+NN**		**DNN**					
Variables	**low-level**		**low-level**		**high-level**		**combined**	
Input Ordering	**gen.**	**p_T**	**gen.**	**p_T**	**gen.**	**p_T**	**gen.**	**p_T**
$M_{\text{part.,rest fr.}}$	13	16	–	–	–	–	–	–
n_{layers}	8	8	8	8	4	4	8	6
n_{nodes}	1024	1024	1024	1024	512	512	1024	1024