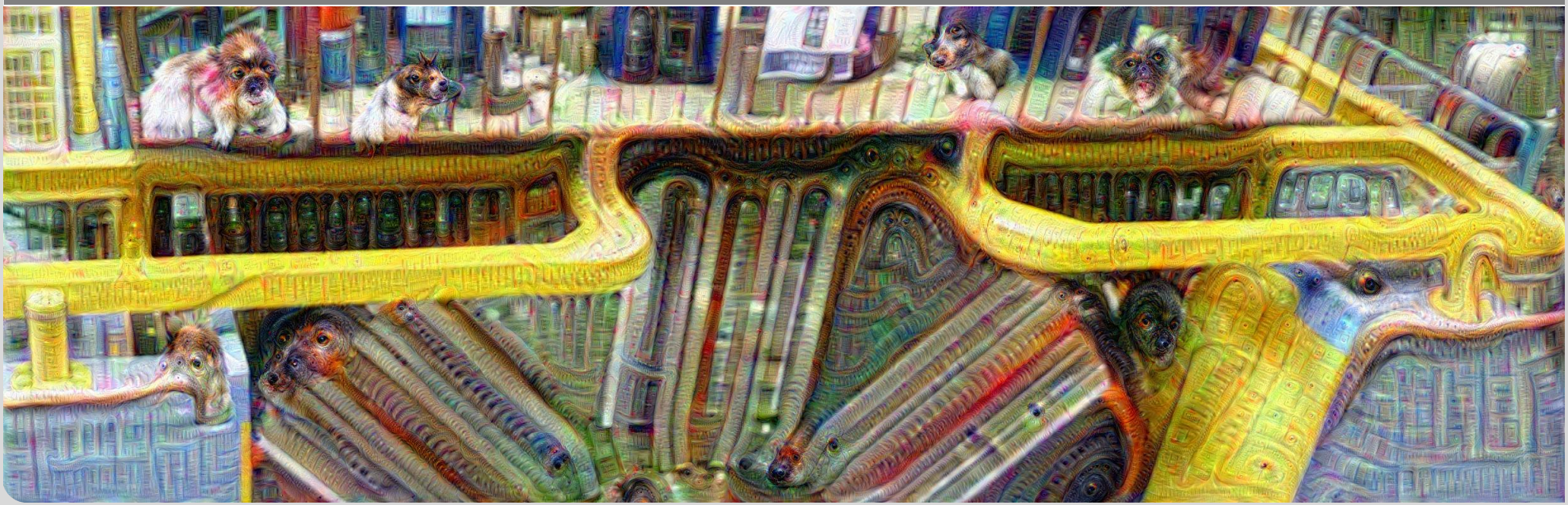
Karlsruher Institut für Technologie

Flavor Tagging using Deep Neural Networks in Belle II

Connecting The Dots 2017

M. Feindt, J. Gemmler, P. Goldenzweig, T. Hauth, M. Heck, T. Keck | 09.03.2017

INSTITUT FÜR EXPERIMENTELLE KERNPHYSIK (IEKP)



Outline

Deep Learning in Physics

ARTICLE

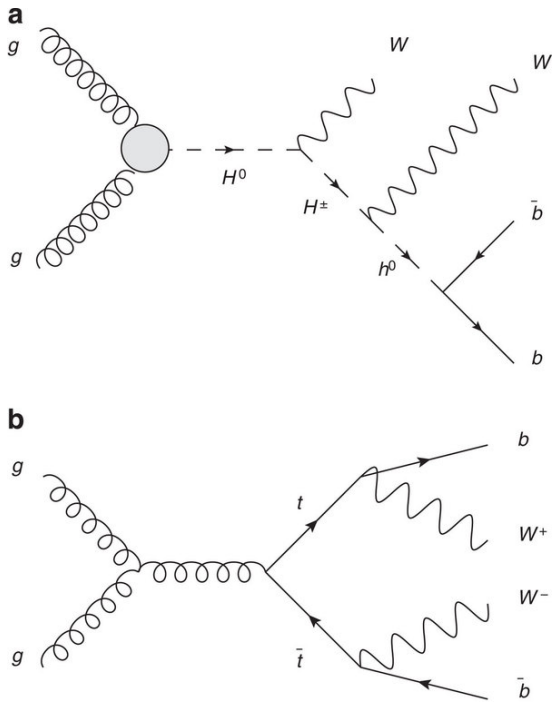
Received 19 Feb 2014 | Accepted 4 Jun 2014 | Published 2 Jul 2014

DOI: 10.1038/ncomms5308

Searching for exotic particles in high-energy physics with deep learning

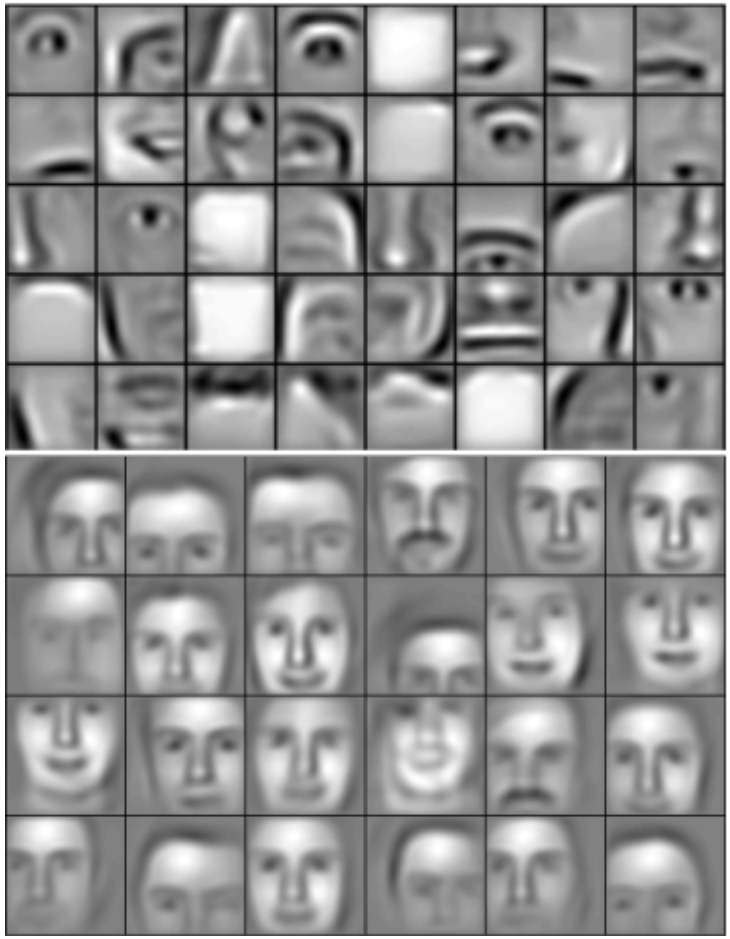
P. Baldi[1], P. Sadowski[1] & D. Whiteson[2]

- 5 layer neural network
- 21 low-level features
  - transverse momentum p_T , pseudorapidity η , ϕ
  - b-tag
  - missing energy ϕ, missing energy magnitude
- compared to 7 high-level features
  - invariant masses of intermediate states



Deep Learning

- from feature engineering to representation learning
- representation attributes[1]
  - abstract and invariant
  - distributed
  - disentangled
- multiple architectures
- supervised and unsupervised
- successfully applied
  - image recognition
  - speech recognition
  - language translation



[1] Representation Learning: A Review and New Perspectives, arXiv:1206.5538v3

[2] Lee, Honglak, et al. "Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations."Proceedings of the 26th annual international conference on machine learning. ACM, 2009.

Deep Learning - Driving Forces

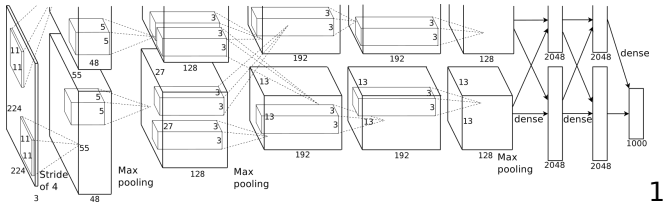
(Big) Data

Architectures/ Algorithms
DNN, CNN, RNN, ...
Dropout, Adam Optimizer, ...

Software Frameworks
Theano, Tensorflow, Torch, ...

Software Libraries
NVIDIA CUDA, OpenCL

Hardware
FPGA, GPU, TPU, ...



[1]

theano

TensorFlow

NVIDIA CUDA

OpenCL

[1] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." Advances in neural information processing systems. 2012.

Deep Learning - Driving Forces

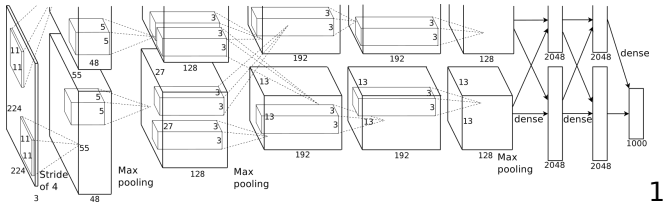
(Big) Data

Architectures/ Algorithms
DNN, CNN, RNN, ...
Dropout, Adam Optimizer, ...

Software Frameworks
Theano, Tensorflow, Torch, ...

Software Libraries
NVIDIA CUDA, OpenCL

Hardware
FPGA, GPU, TPU, ...



[1]

theano

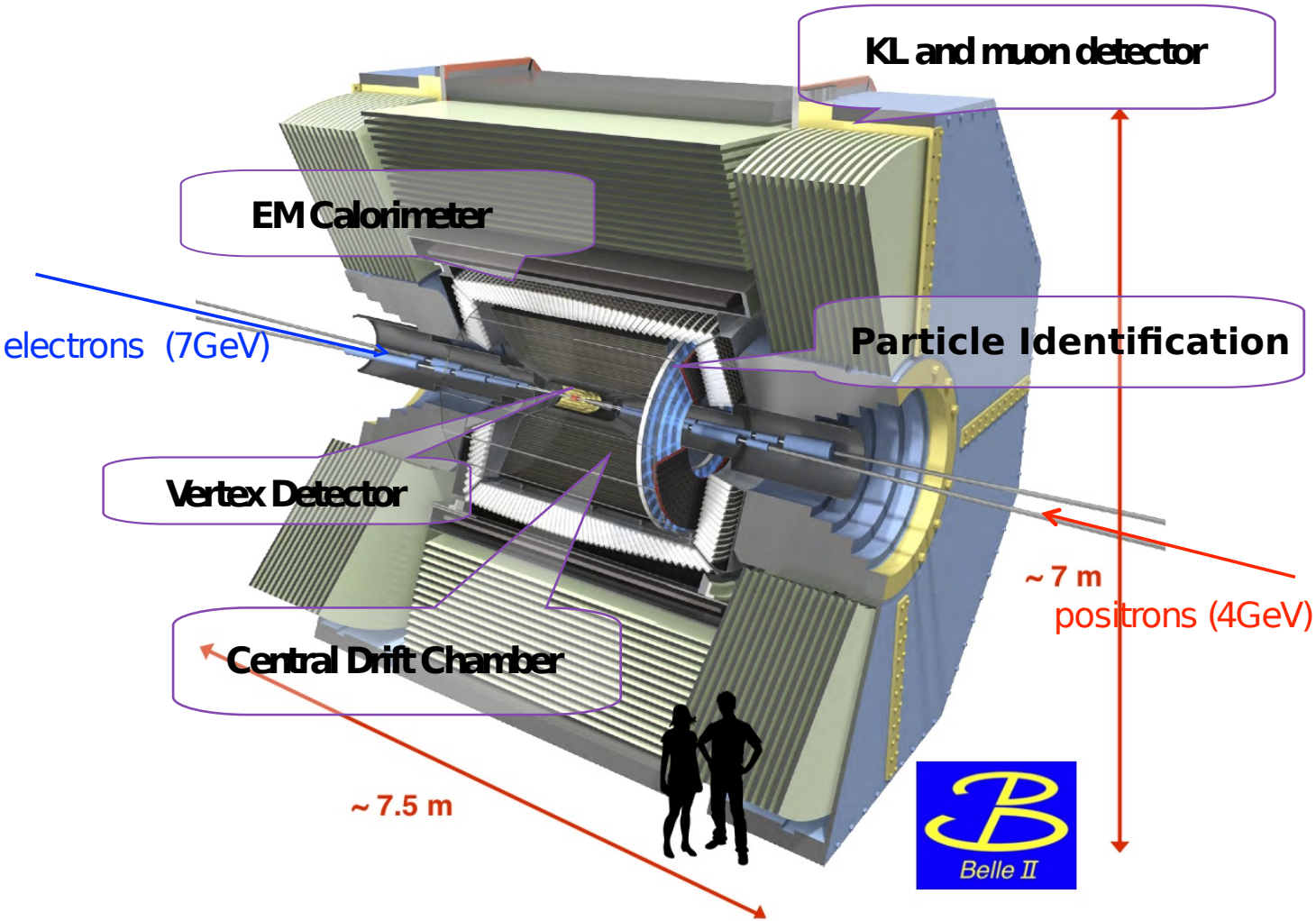
TensorFlow

NVIDIA CUDA

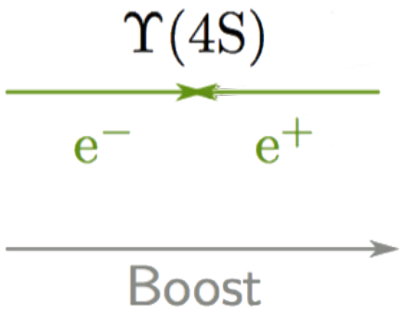
OpenCL

[1] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." Advances in neural information processing systems. 2012.

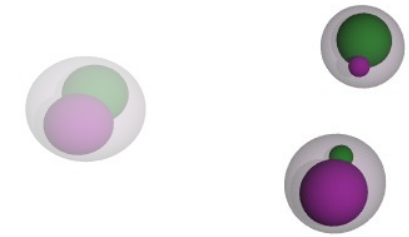
Changing Perspective - The Belle II Experiment

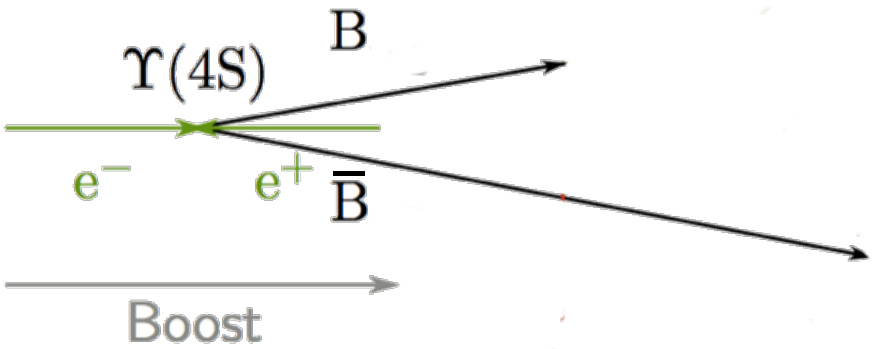


B meson production at Belle II

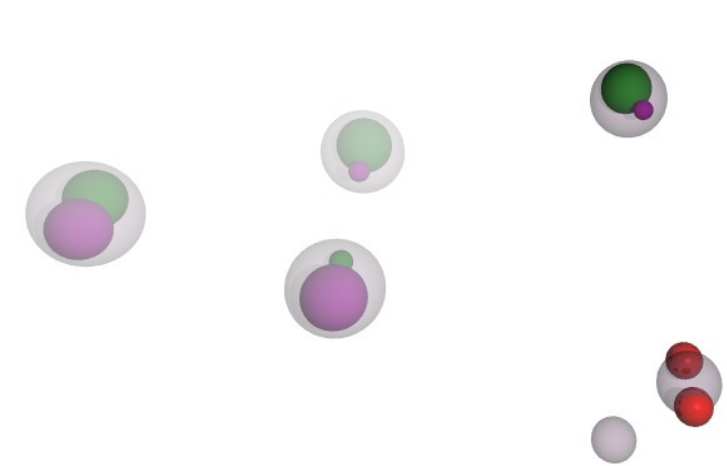


B meson production at Belle II



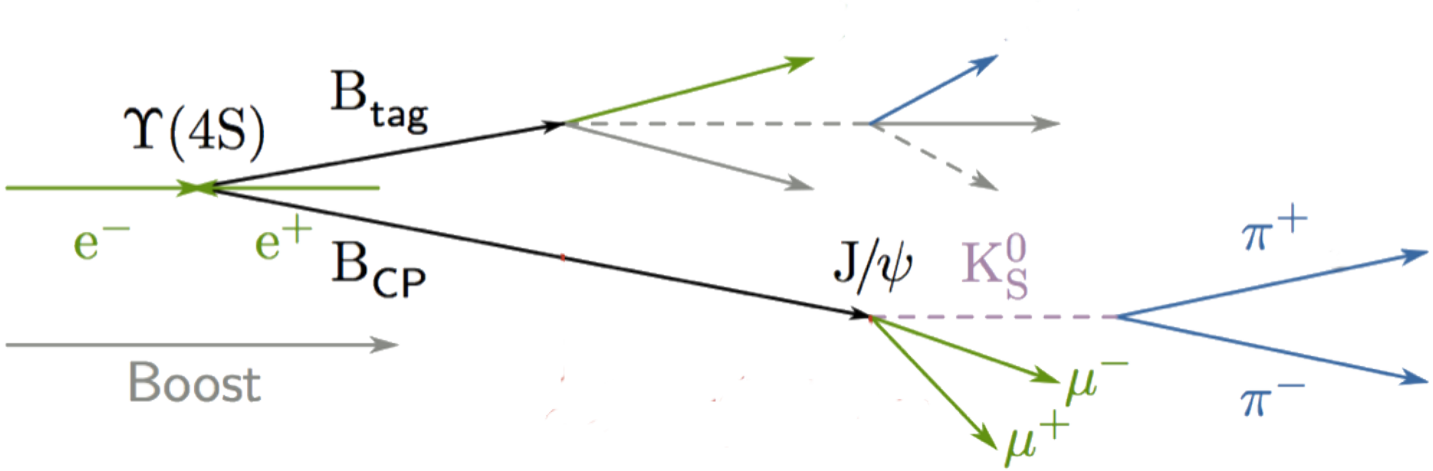


B meson production at Belle II



B-Decay in Flavour-Eigenstate, e.g. $B \rightarrow \boldsymbol{D}\ [\rightarrow \boldsymbol{K}\ \boldsymbol{\pi}]\ \boldsymbol{\pi}$

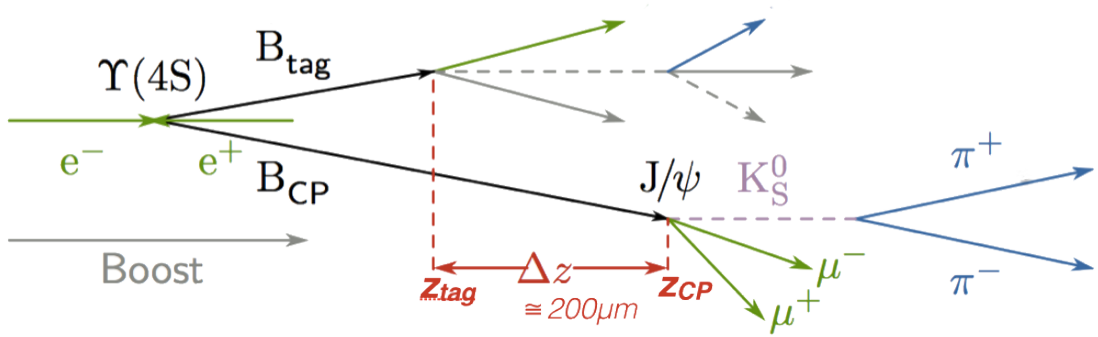
B-Decay in CP-Eigenstate, e.g. $B \rightarrow \boldsymbol{J/\Psi}\ [\rightarrow \boldsymbol{\mu}\ \boldsymbol{\mu}]\ K_s$



Ingredients of CPV-Measurements

$$A(t) = \frac{\Gamma(\bar{B}^0 \to f_{CP}) - \Gamma(B^0 \to f_{CP})}{\Gamma(\bar{B}^0 \to f_{CP}) + \Gamma(B^0 \to f_{CP})} = A^{dec} \cos(\Delta m\, t) + A^{mix-ind} \sin(\Delta m\, t)$$

- $B^0 \to J/\Psi K_s^0$: Measurement of $\sin(2\beta)$
- reconstruct the signal-side
- measure distance between vertices
  $\Delta t = \Delta z / \beta\gamma c$
- important aspect:
  determine the Flavor of the tag-side



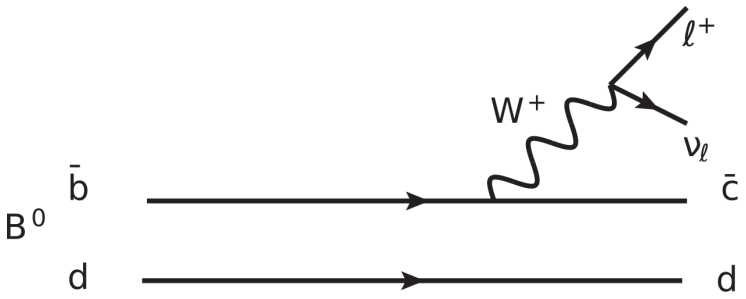
Flavor Tagging Categories

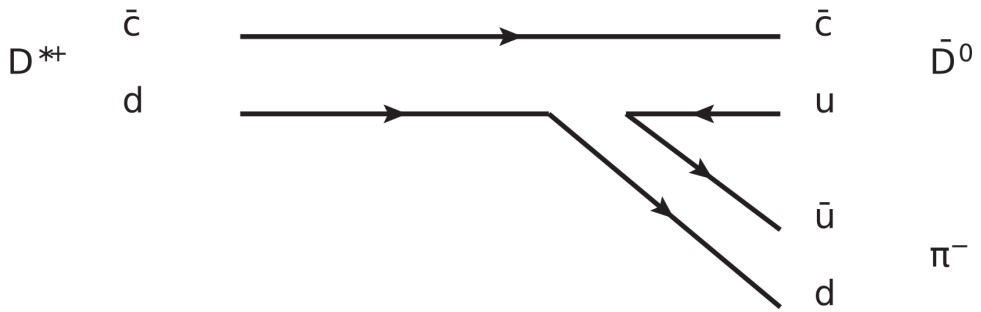
flavor-specific categories

- Primary Leptons
- Secondary Leptons
- Slow Pion
- Fast Strange Particles
- Slow Strange Particles

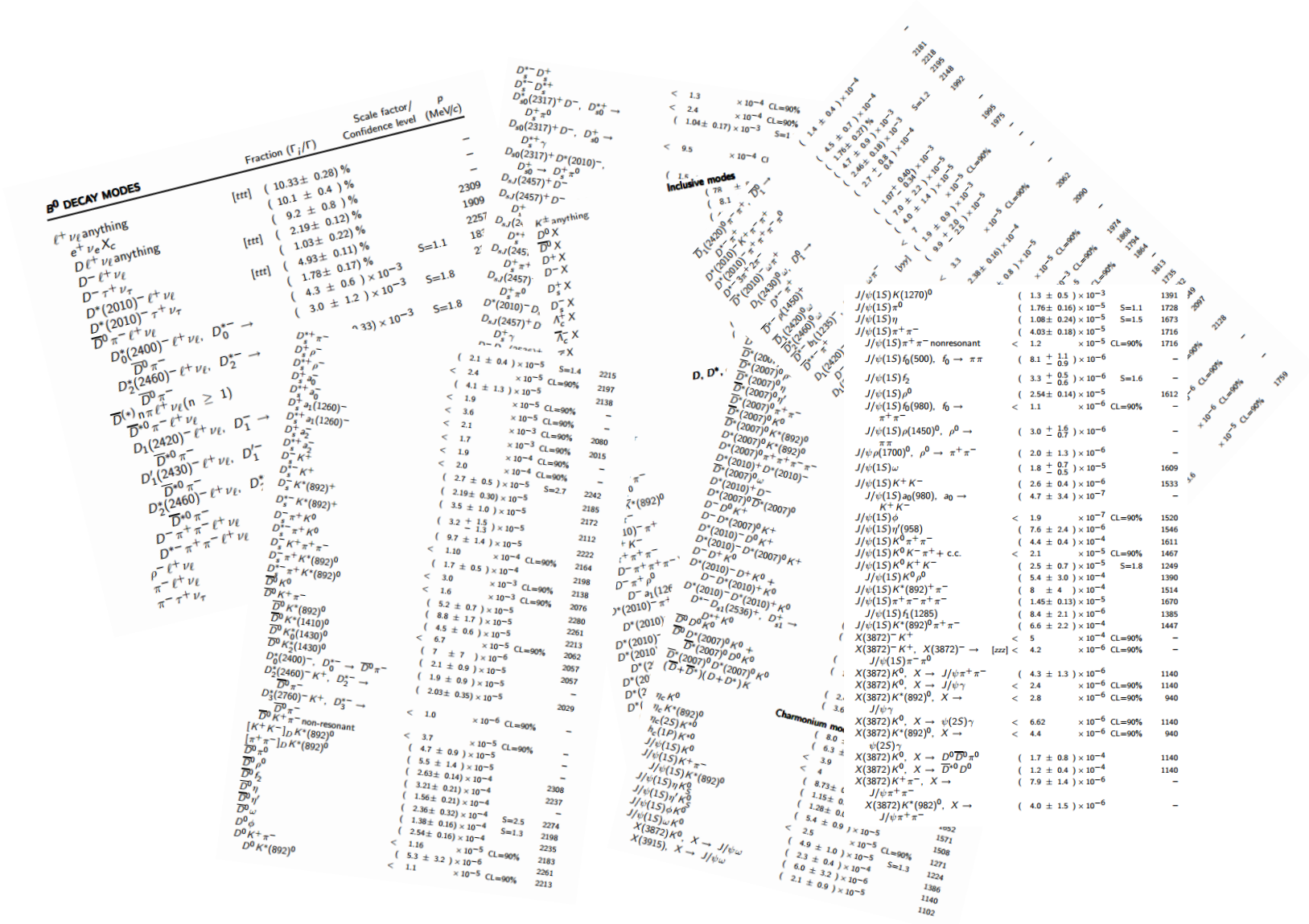
decay processes

- $\bar{b} \to \bar{c}\,\ell^+\nu$
- $\bar{b} \to \bar{c} \to \bar{s}\,\ell^-\bar{\nu}$
- $B^0 \to D^{*-}X, D^{*-} \to \bar{D}^0\pi^-$
- $B^0 \to K^+ X_{c\bar{c}}$
- $\bar{b} \to \bar{c} \to \bar{s}$





Why Flavor Tagging is suitable for Deep Learning

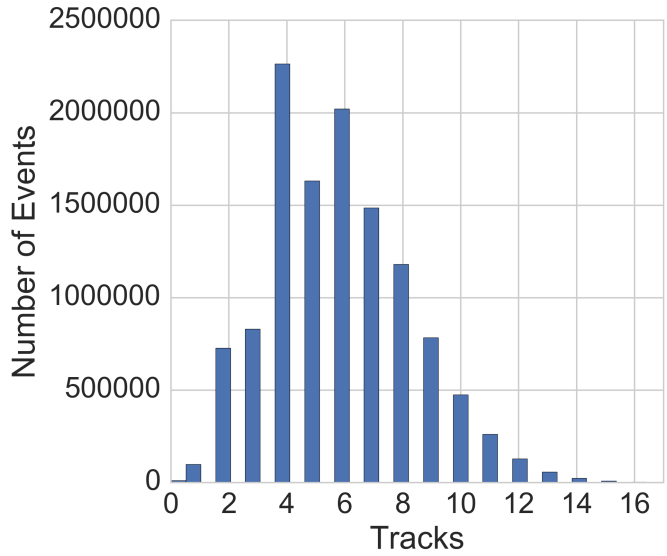


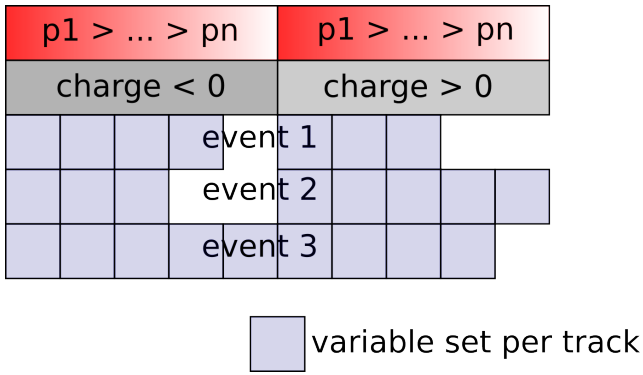
[1] C. Patrignani et al. (Particle Data Group), Chin. Phys. C, 40, 100001 (2016)

Flavor Tagging - Input Features

Features

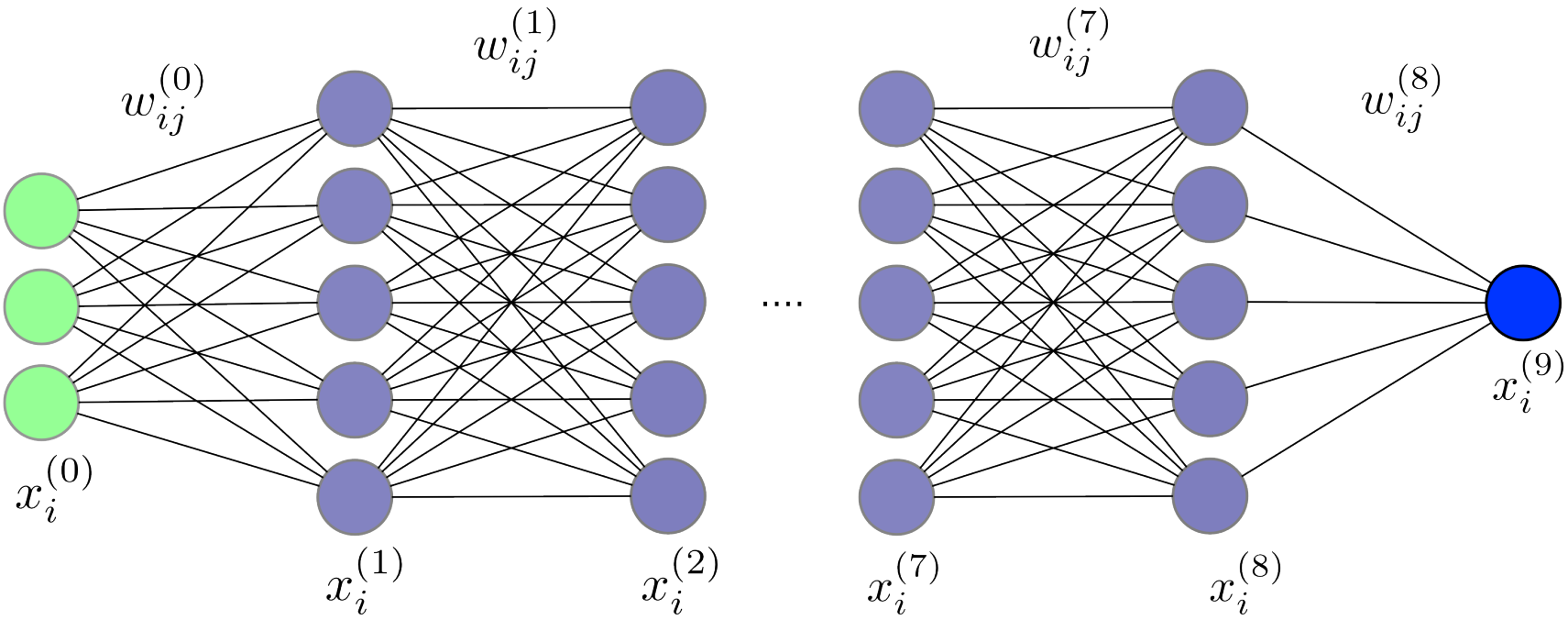
- 10 charged tracks at maximum
- sorted by momentum, grouped by charge
- p_{CMS}, $\cos(\theta_{CMS})$, ϕ_{CMS}
- electron, pion, muon, kaon, proton ID
- hit count of the Belle II-tracking detectors (3)
- track perigee (dr, dz)
- pValue
- overall 140 input features





Flavor Tagging, New Approach

using a 9 layer perceptron



Input Layer Hidden Layers Output Layer

nodes

$$x_i^{(k+1)} = \sigma^{(k+1)}\left(\sum_{j=1} w_{ij}^{(k)} x_j^{(k)} + w_{i0}^{(k)}\right)$$

Training the Neural Network

- binary classification
  $t_n \in \{0, 1\}$

loss function

- compare network output y_n and true information t_n
- cross-entropy
  $E(\mathbf{w}) = -\sum_{n=1}[t_n \ln y_n + (1 - t_n) \ln(1 - y_n)]$

stochastic gradient descent

- weight update
  $\Delta w^i = -\eta \nabla E_n(w^i)$
- weight update with momentum
  $\Delta w^i = -\eta \nabla E_n(w^i) + \mu \Delta w^{i-1}$

Training the Neural Network

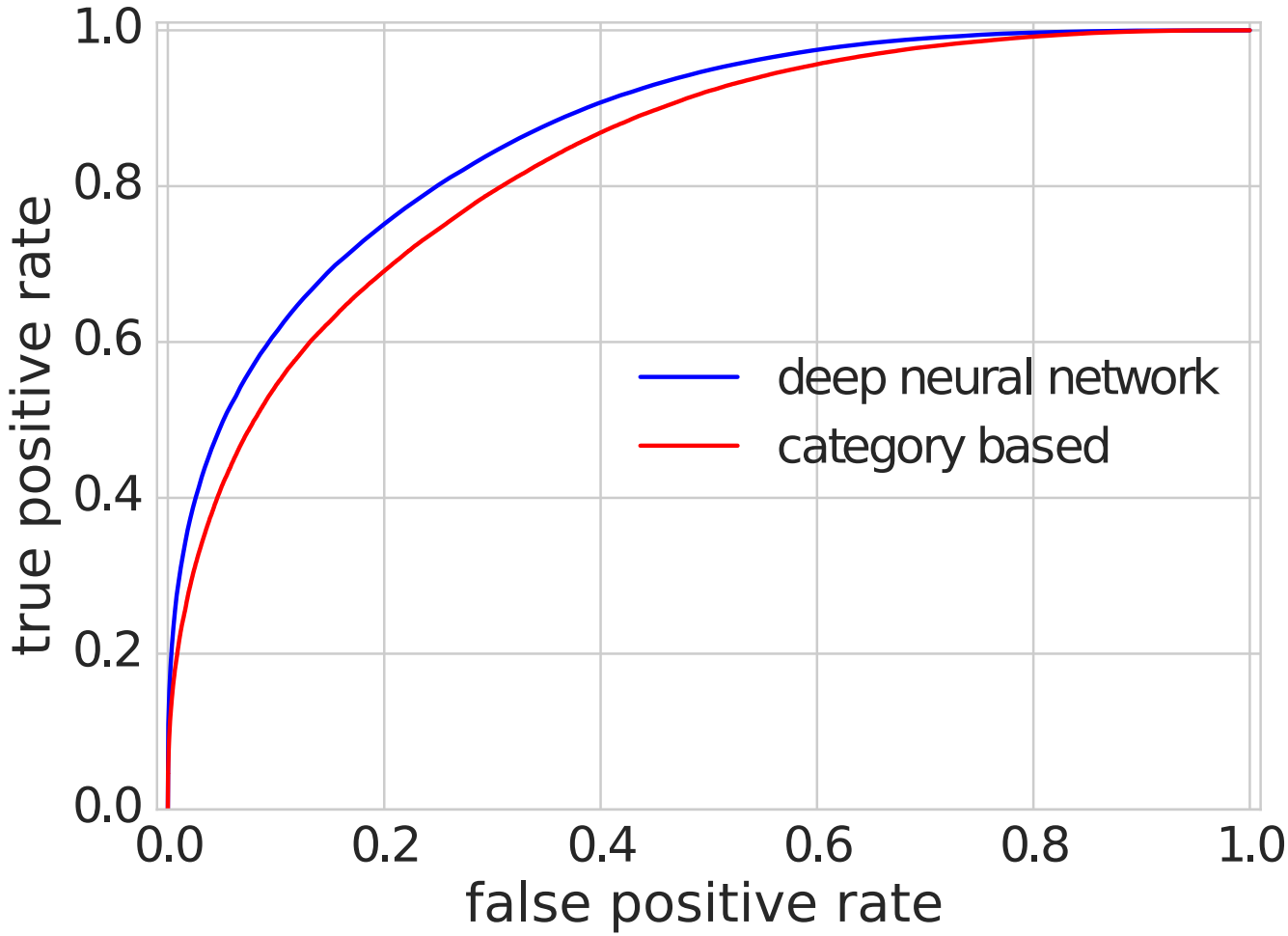
regularization

prevent over-fitting

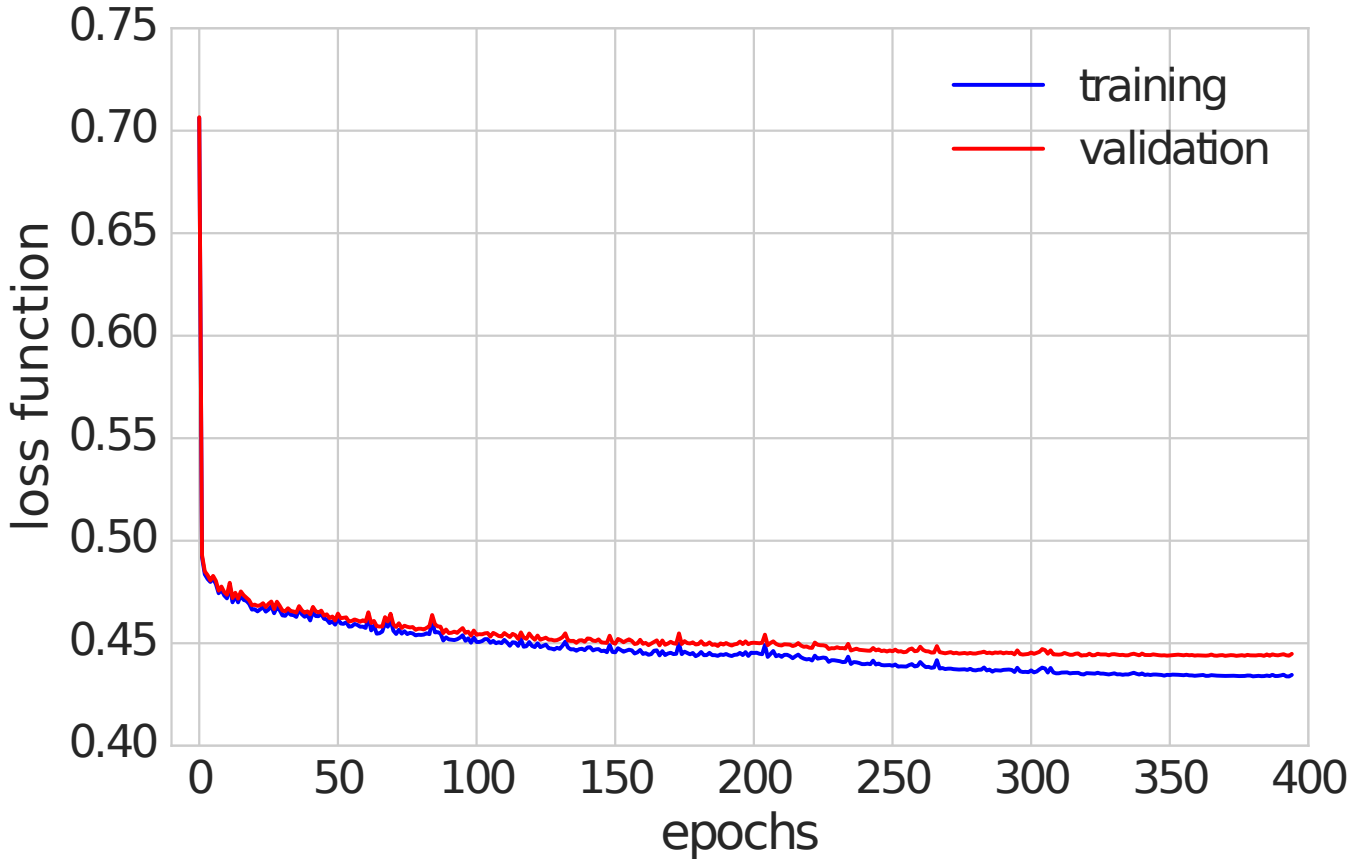
- early stopping
- L2 - weight decay
  $E'_n = E_n + \frac{\alpha}{2} \sum_{i,j} W_{ij}^2$
- dropout

$\rightarrow$ Needs a lot of labeled data:
Accessible in particle physics
Size of the used training data: 12 million events

Receiver Operating Characteristic

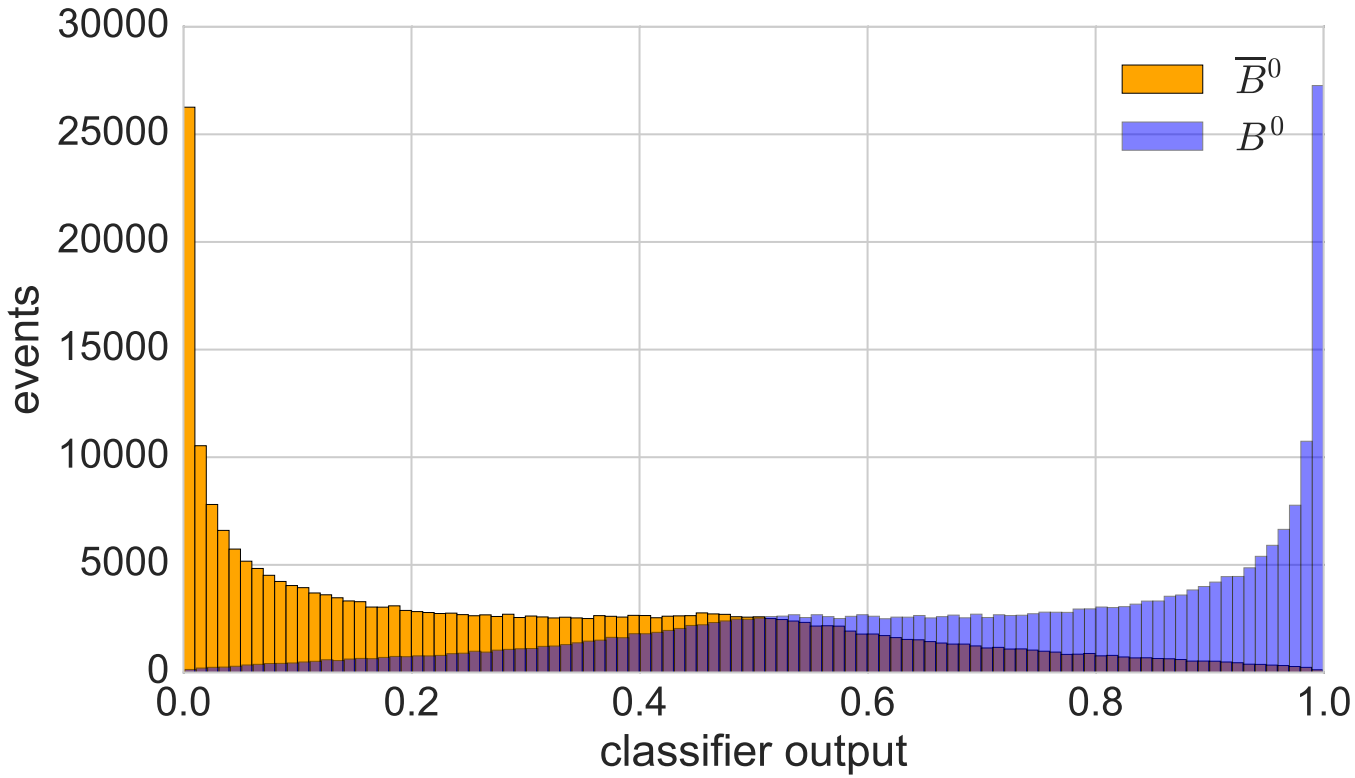


Loss Function



- algorithm converges
- slight over-training on the training dataset

Classifier


30000
25000
20000
15000
10000
5000
0
events
0.0
0.2
0.4
0.6
0.8
1.0
classifier output
$\overline{B}^0$
B^0

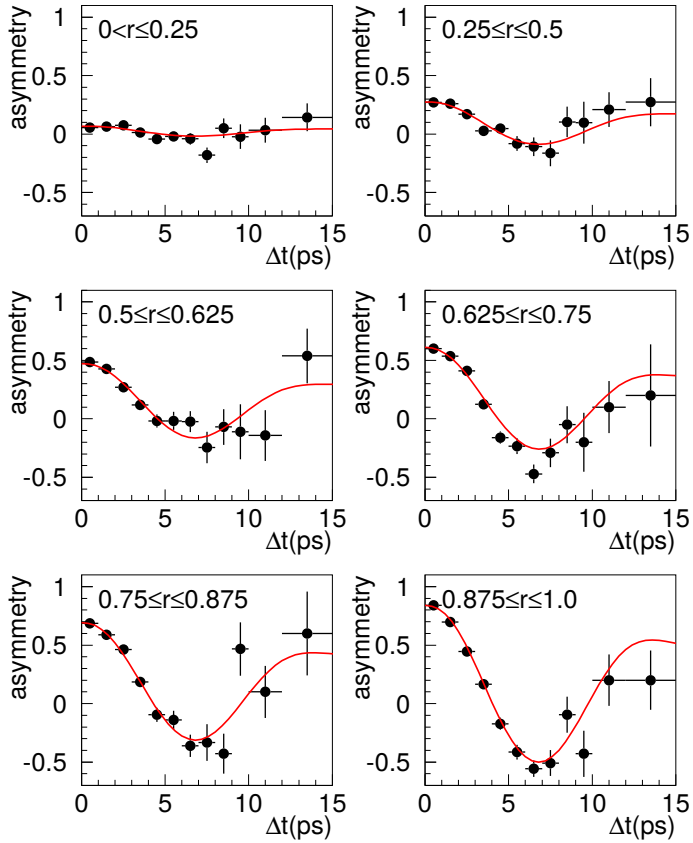
Effective Tagging Efficiency

CP Asymmetry

$$A_{CP}^{obs} = (1 - 2w)A_{CP}^{mes}$$

$$Q = \sum_{i=1}^{6} \epsilon_i (1 - 2w_i)^2$$

$$r_i = 1 - 2w_i$$

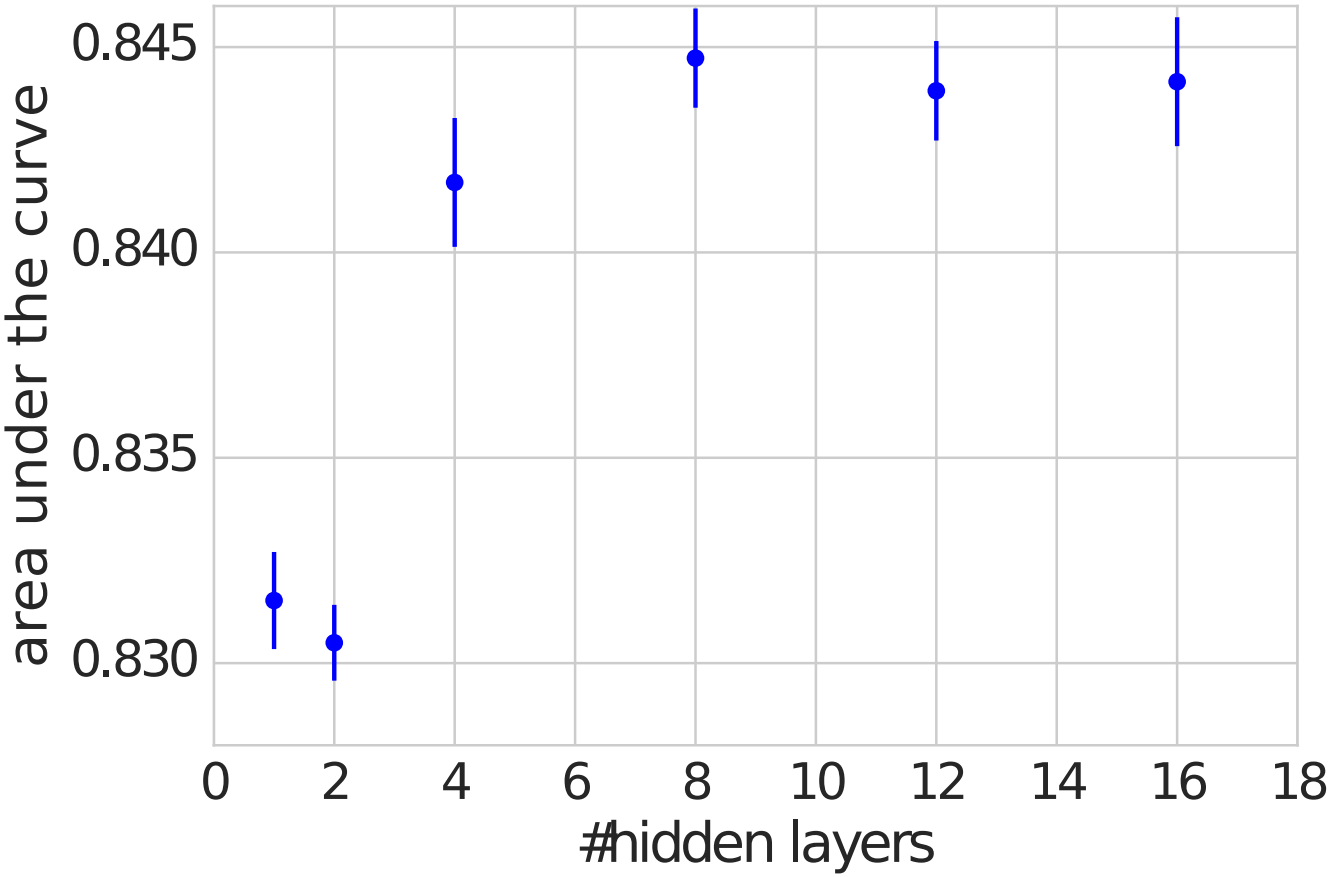


1

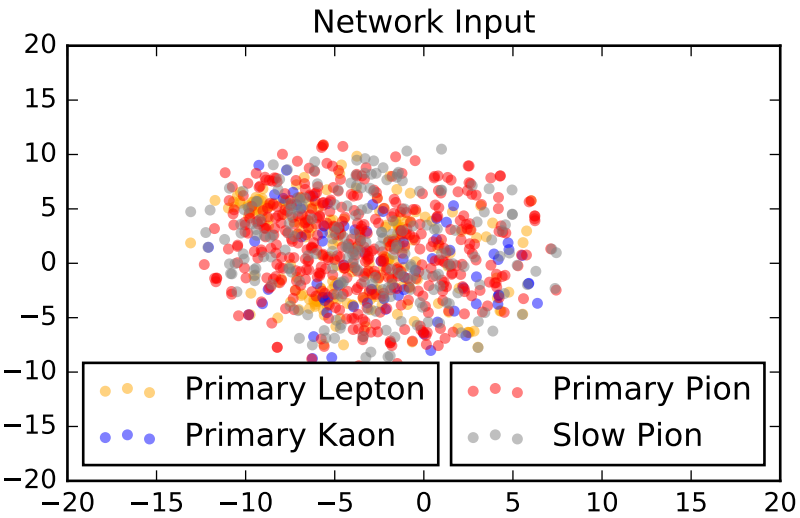
		Category Based	Deep Neural Network
Belle II	$J/\Psi K_S^0$	0.3329±0.0001	0.4069±0.0003
Belle	$J/\Psi K_S^0$	*0.293 ± 0.01* [1]	0.3442±0.0009

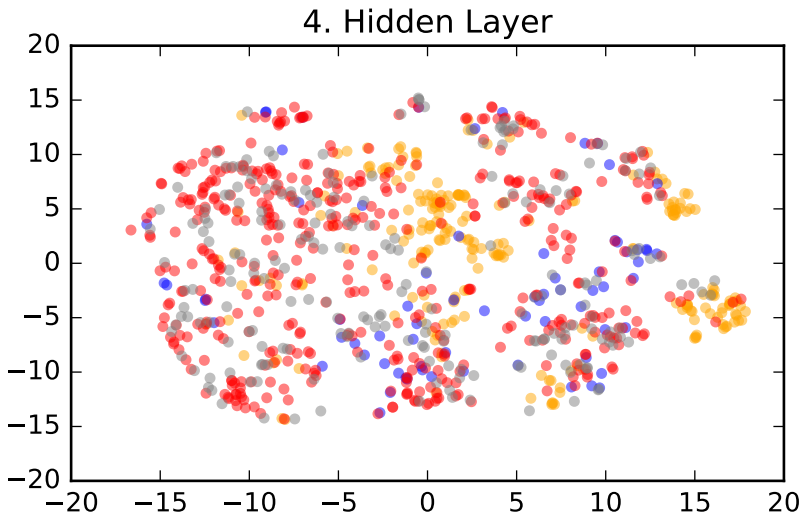
[1] DOI: 10.1016/j.nima.2004.06.159

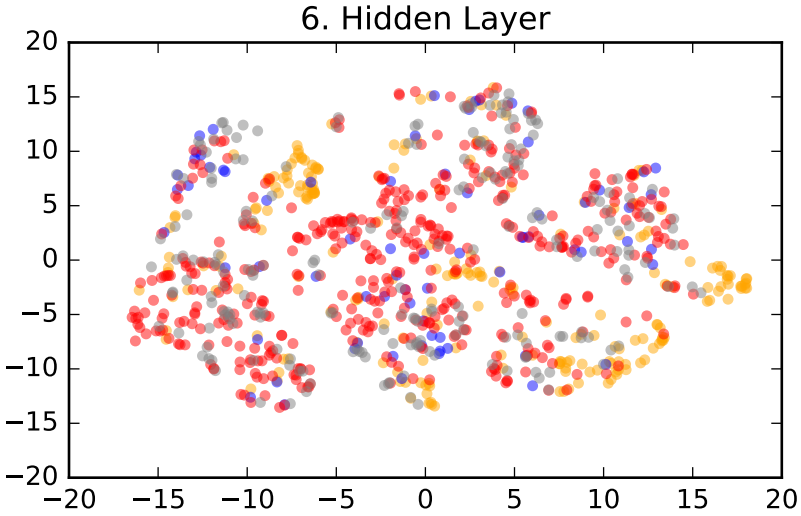
Hyperparameter Influence - Number of Hidden Layers

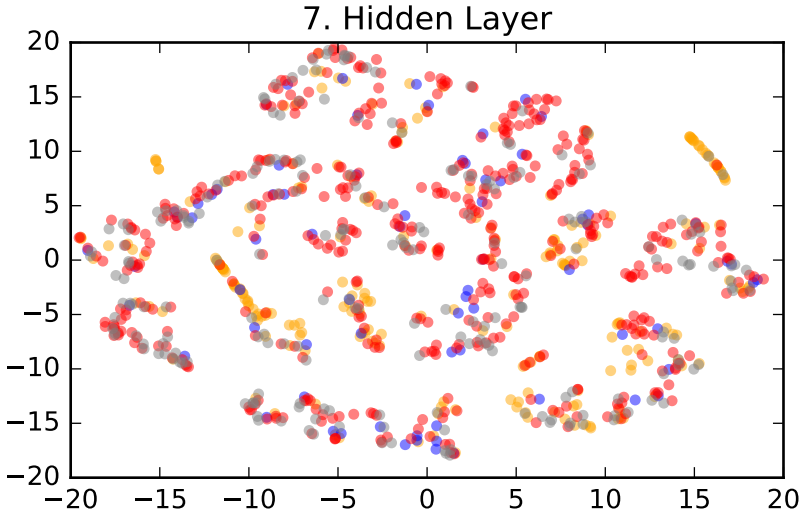


Feature Representation - Hidden Layers



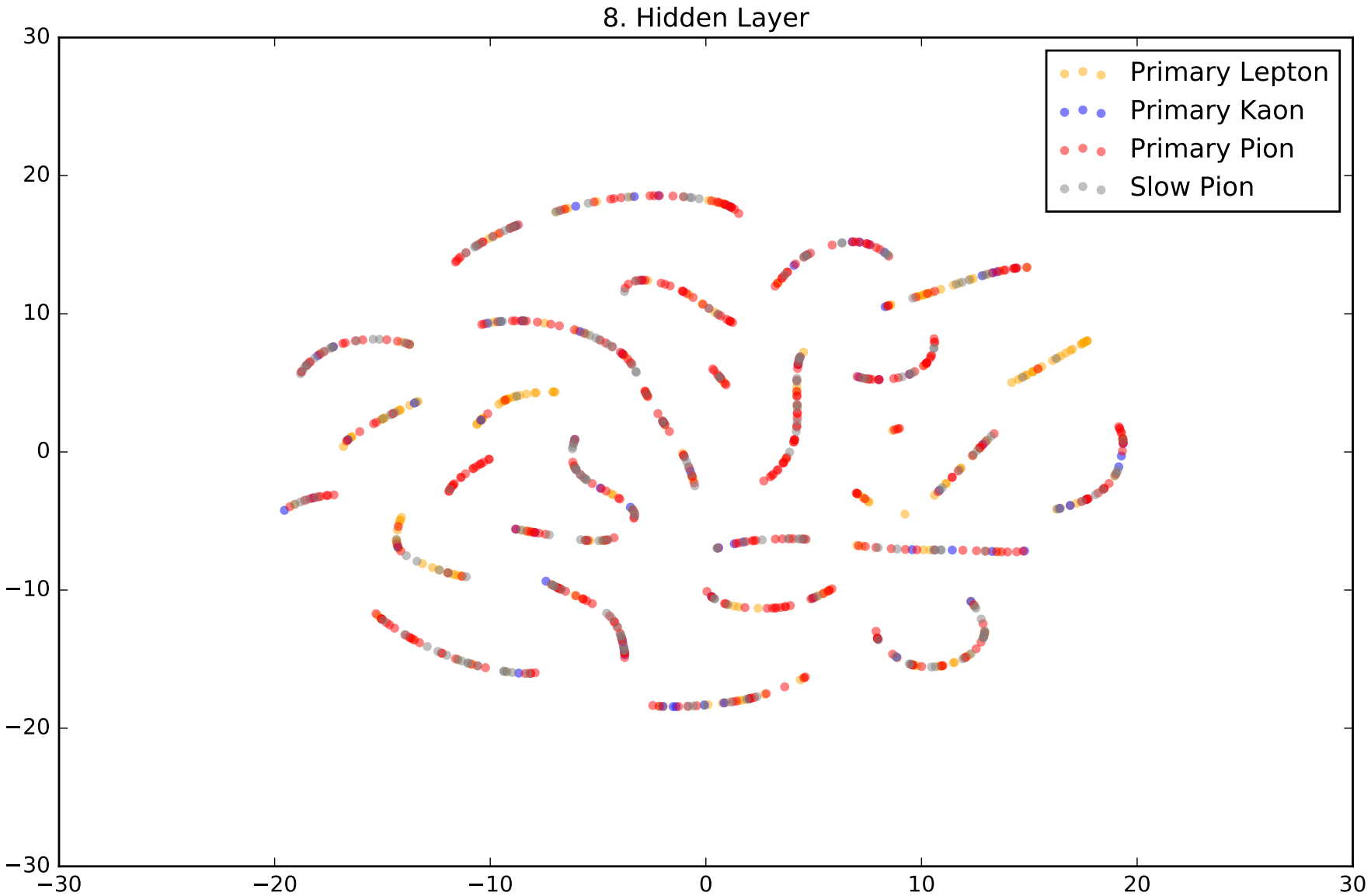




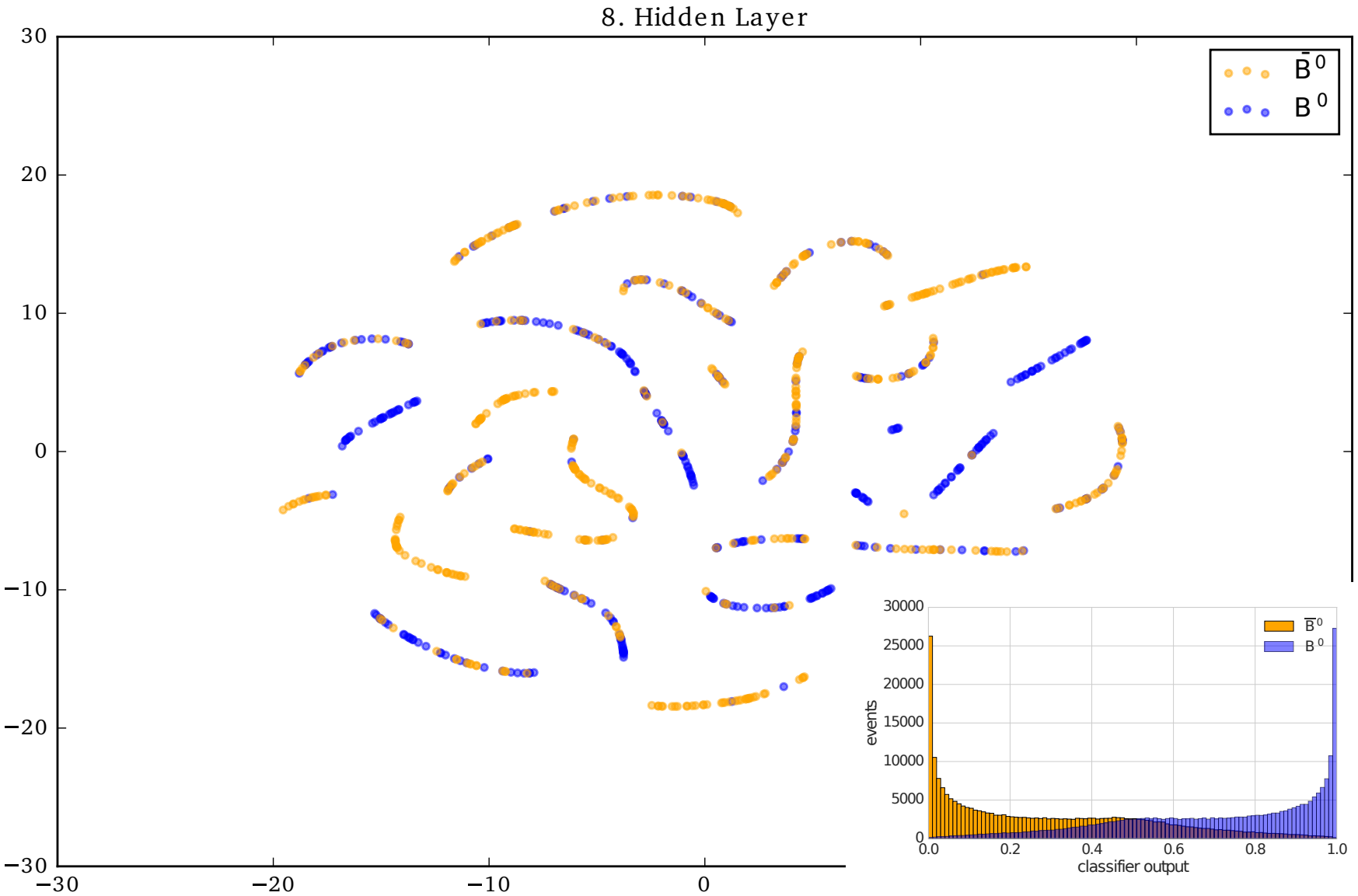


Visualization with T-distributed Stochastic Neighbour Embedding (TSNE)

Feature Representation - Hidden Layer



Feature Representation - Hidden Layer



Summary

- representation learning is advantageous for certain tasks
- new Flavor Tagging method
  - based on Deep Neural Networks
  - relies on low level features
  - shows significant improvement on Monte Carlo compared to the established method
- classifier performance can be evaluated directly on data

Thank you for your attention!

Hyperparameter Influence - Weight decay