Selective background Monte Carlo simulation at Belle II

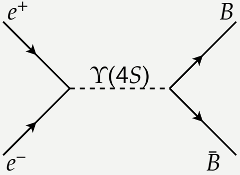
James Kahn
Thomas Kuhr
Martin Ritter

Ludwig Maximilians Universität München
DFG “Excellence Cluster ORIGINS”

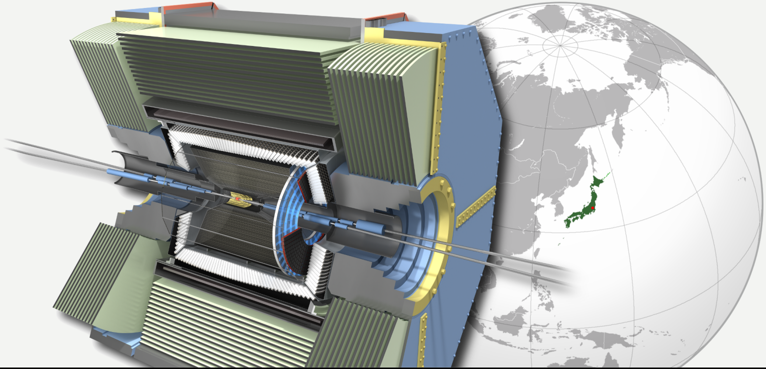
ACAT19, March 2019

Belle II Experiment

Asymmetric e^+e^- experiment mainly at the $\Upsilon(4S)$ resonance (10.58 GeV)

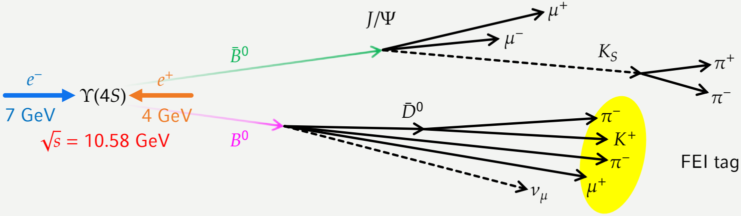


Focus on B, charm and τ physics

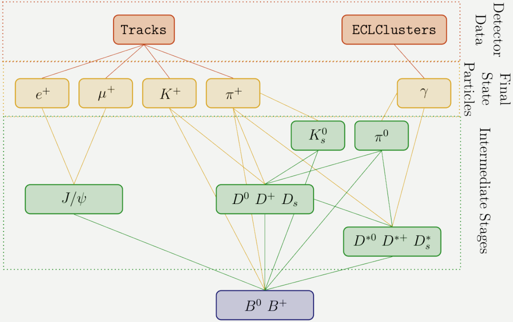


	KEKB/Belle	SuperKEKB/Belle II
Operation	1999–2010	2019–
Peak luminosity	$2.11 \times 10^{34}\ \mathrm{cm^{-2}s^{-1}}$	$8 \times 10^{35}\ \mathrm{cm^{-2}s^{-1}}$
Integrated luminosity	$1\ \mathrm{ab^{-1}}$ (772 million $B\bar{B}$ pairs)	$50\ \mathrm{ab^{-1}}$

Full Event Interpretation

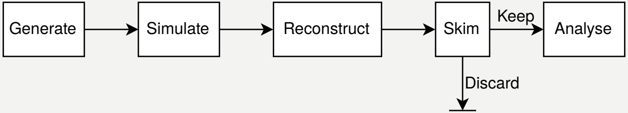


- Hierarchical reconstruction of B mesons
- $O(200)$ decay channels with classifiers trained for each
- Reconstructs $O(10000)$ unique decay chains in six stages
- Talk by W. Sutcliffe on Wednesday, 19:00 (track 2)



Problem

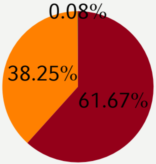
- Approach at Belle → Background MC ≈ 10 × data
- Infeasible at Belle II → still require high statistics

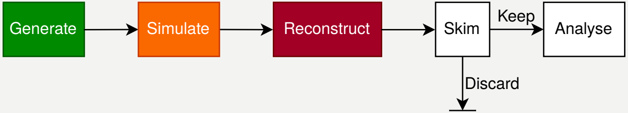


Problem

- Approach at Belle → Background MC ≈ 10 × data
- Infeasible at Belle II → still require high statistics
- Event simulation is time consuming
- Currently for $\Upsilon(4S) \sim 100$ HEPSpec s / event
  To simulate 1 ab^{-1} →∼ 80 GHEPSpec s

Time contribution





Problem

- Approach at Belle → Background MC ≈ 10 × data
- Infeasible at Belle II → still require high statistics
- Event simulation is time consuming
- Currently for $\Upsilon(4S) \sim 100$ HEPSpec s / event
  To simulate 1 ab^{-1} →~ 80 GHEPSpec s
- For given analysis → most background events discarded trivially

Time contribution

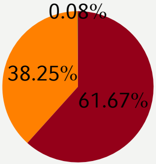
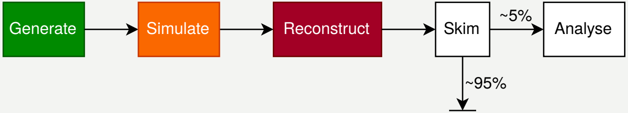


Table: FEI skim retention rates.

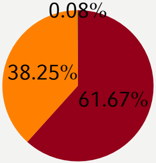
Channel	Had B^+	Had B^0
$B^0\bar{B}^0$	5.62%	4.25%
B^+B^-	8.35%	3.82%
$u\bar{u}$	6.86%	3.78%
$d\bar{d}$	7.20%	3.39%
$c\bar{c}$	12.0%	5.73%
$s\bar{s}$	6.13%	2.95%



Problem

- Approach at Belle → Background MC ≈ 10 × data
- Infeasible at Belle II → still require high statistics
- Event simulation is time consuming
- Currently for $\Upsilon(4S) \sim 100$ HEPSpec s / event
  To simulate 1 ab^{-1} $\rightarrow\sim 80$ GHEPSpec s
- For given analysis → most background events discarded trivially

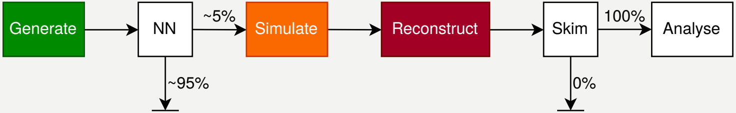
Time contribution



Try to discard events
before simulation

Table: FEI skim retention rates.

Channel	Had B^+	Had B^0
$B^0\bar{B}^0$	5.62%	4.25%
B^+B^-	8.35%	3.82%
$u\bar{u}$	6.86%	3.78%
$d\bar{d}$	7.20%	3.39%
$c\bar{c}$	12.0%	5.73%
$s\bar{s}$	6.13%	2.95%



Dataset

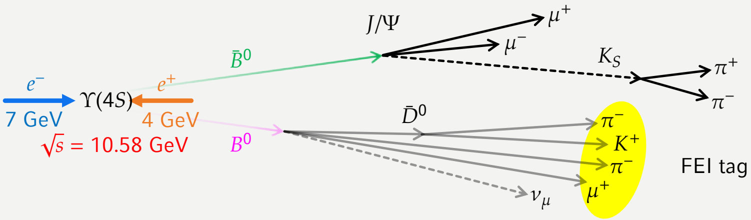
- Train on two FEI skim channels:
  - FEI Had. B^0: 8.5 M events
  - FEI Had. B^+: 4 M events
- 10% reserved for validation
- $100,000$ test events simulated independently
- Implemented in Tensorflow/Keras

Table: FEI skim retention rates.

Channel	Had B^+	Had B^0
$\mathbf{B^0B^0}$	5.62%	**4.25**%
$\mathbf{B^+B^-}$	**8.35**%	3.82%
$u\bar{u}$	6.86%	3.78%
$d\bar{d}$	7.20%	3.39%
$c\bar{c}$	12.0%	5.73%
$s\bar{s}$	6.13%	2.95%

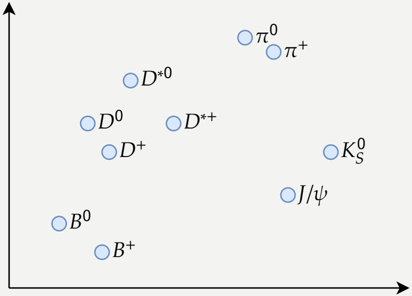
Features

Decay strings

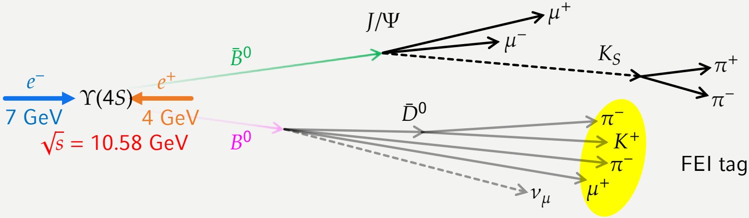


$\bar{B}^0 \to J/\psi(\to \mu^+\mu^-)\, K_S^0(\to \pi^+\pi^-)$
tokenize for embedding:

```
[-511,'(', 443,'(',-13,13,')',310,'(',211,-211,')',')']
```



Decay strings



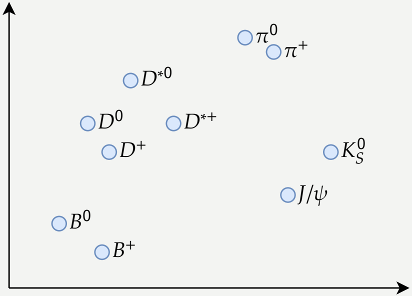
$\bar{B}^0 \to J/\psi(\to \mu^+\mu^-)\, K_S^0(\to \pi^+\pi^-)$
tokenize for embedding:

```
[-511,'(', 443,'(',-13,13,')',310,'(',211,-211,')',')']
```

This study:

- 540 total possible tokens (particles)
- 8 embedding dimensions

To find embedding dims:
Start from 4$^{\text{th}}$ root and work up



Features

MCParticles

1D kernel

$\Upsilon(4S)$ (300553)
$\bar{B}^0$ (-511)
J/ψ (443)
μ^+ (-13)
μ^- (13)
K_S^0 (310)
pi^- (-211)
pi^+ (211)
B^0 (511)
$\bar{D}^0$ (-421)
pi^- (-211)
K^+ (321)
pi^- (-211)
μ^+ (-13)
ν_μ (14)

Feature	Definition
PDG code	Identifier of particle type and charge.
Mother PDG code	Particle parent PDG code.
Mass	Particle mass in GeV/c^2.
Charge	Electric charge of the particle.
Energy	Particle energy in GeV.
Momentum	Three momentum of the particle in Gev/c.
Production time	Production time in ns relative to $\Upsilon(4S)$ production.
Production vertex	Coordinates of particle production vertex.
Status bit	Bitmask representing MC production conditions.

- All simulated particles have same set of features
- Variable number of particles per event $\rightarrow$ pad with zeroes
- Apply 1D convolutions across particles

MCParticles

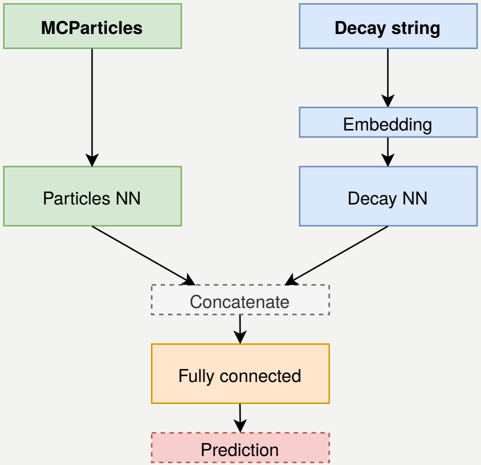
1D kernel

$\Upsilon(4S)$	(300553)	$[E,\ M,\ p_x,\ p_y,\ p_z,\ x,\ y,\ z,\ ...]$
$\bar{B}^0$	(-511)	$[E,\ M,\ p_x,\ p_y,\ p_z,\ x,\ y,\ z,\ ...]$
J/ψ	(443)	$[E,\ M,\ p_x,\ p_y,\ p_z,\ x,\ y,\ z,\ ...]$
μ^+	(-13)	$[E,\ M,\ p_x,\ p_y,\ p_z,\ x,\ y,\ z,\ ...]$
μ^-	(13)	$[E,\ M,\ p_x,\ p_y,\ p_z,\ x,\ y,\ z,\ ...]$
K_S^0	(310)	$[E,\ M,\ p_x,\ p_y,\ p_z,\ x,\ y,\ z,\ ...]$
pi^-	(-211)	$[E,\ M,\ p_x,\ p_y,\ p_z,\ x,\ y,\ z,\ ...]$
pi^+	(211)	$[E,\ M,\ p_x,\ p_y,\ p_z,\ x,\ y,\ z,\ ...]$
B^0	(511)	$[E,\ M,\ p_x,\ p_y,\ p_z,\ x,\ y,\ z,\ ...]$
$\bar{D}^0$	(-421)	$[E,\ M,\ p_x,\ p_y,\ p_z,\ x,\ y,\ z,\ ...]$
pi^-	(-211)	$[E,\ M,\ p_x,\ p_y,\ p_z,\ x,\ y,\ z,\ ...]$
K^+	(321)	$[E,\ M,\ p_x,\ p_y,\ p_z,\ x,\ y,\ z,\ ...]$
pi^-	(-211)	$[E,\ M,\ p_x,\ p_y,\ p_z,\ x,\ y,\ z,\ ...]$
μ^+	(-13)	$[E,\ M,\ p_x,\ p_y,\ p_z,\ x,\ y,\ z,\ ...]$
ν_μ	(14)	$[E,\ M,\ p_x,\ p_y,\ p_z,\ x,\ y,\ z,\ ...]$

Architecture

Investigated components:

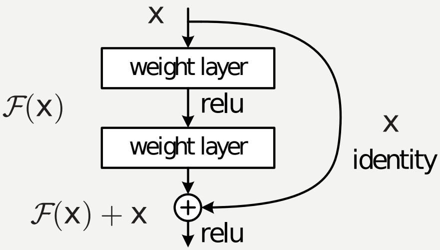
- Fully connected
- Convolutional
- Recurrent (LSTM)

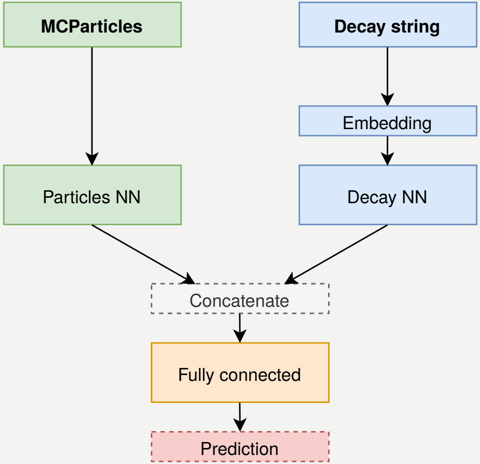


Architecture

Investigated components:

- Fully connected
- Convolutional
- Recurrent (LSTM)
- ResNet (1512.03385) / ResNeXt (1611.05431)

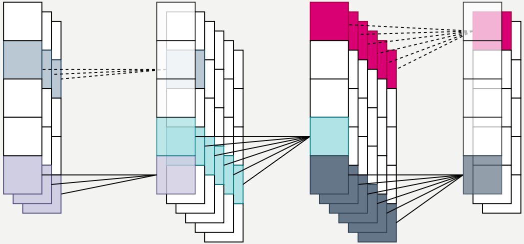


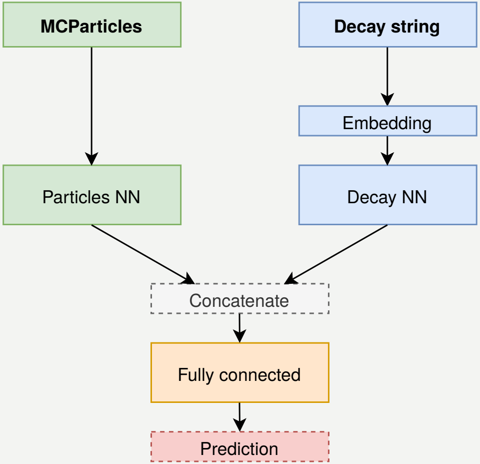


Architecture

Investigated components:

- Fully connected
- Convolutional
- Recurrent (LSTM)
- ResNet (1512.03385) / ResNeXt (1611.05431)
- 1×1 convolutions (1409.4842)

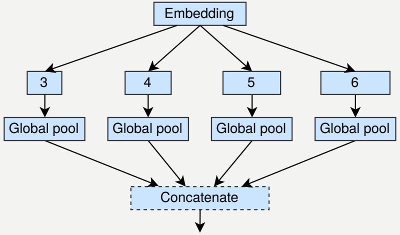


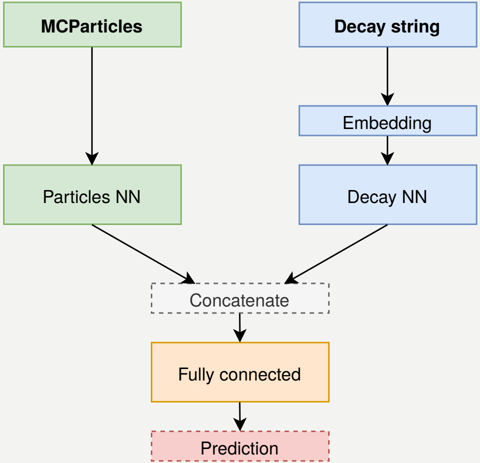


Architecture

Investigated components:

- Fully connected
- Convolutional
- Recurrent (LSTM)
- ResNet (1512.03385) / ResNeXt (1611.05431)
- 1×1 convolutions (1409.4842)
- Wide convolutional (1510.03820)





Training

- Binary classification → binary cross-entropy
- Adam optimiser with AMSGrad
- LeakyReLU activation for intermediate layers
- Sigmoid activation for output layer
- BatchNorm/Dropout used for regularisation
- Trainings performed on consumer grade GPU (Nvidia GTX 1080Ti)

Training

- Binary classification → binary cross-entropy
- Adam optimiser with AMSGrad
- LeakyReLU activation for intermediate layers
- Sigmoid activation for output layer
- BatchNorm/Dropout used for regularisation
- Trainings performed on consumer grade GPU (Nvidia GTX 1080Ti)

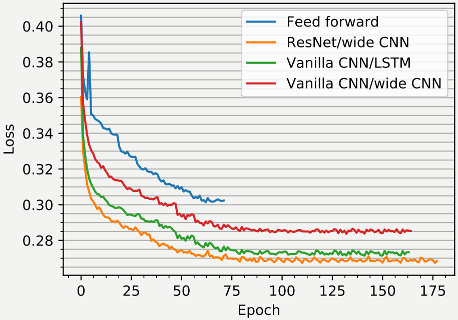


Figure: Example: Charged FEI

Training

- Binary classification → binary cross-entropy
- Adam optimiser with AMSGrad
- LeakyReLU activation for intermediate layers
- Sigmoid activation for output layer
- BatchNorm/Dropout used for regularisation
- Trainings performed on consumer grade GPU (Nvidia GTX 1080Ti)

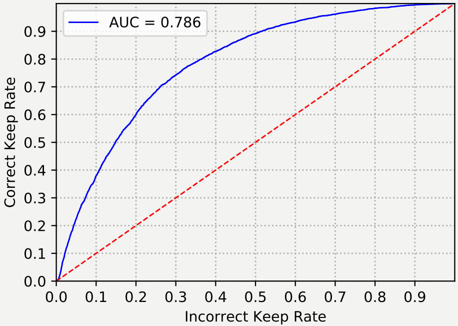
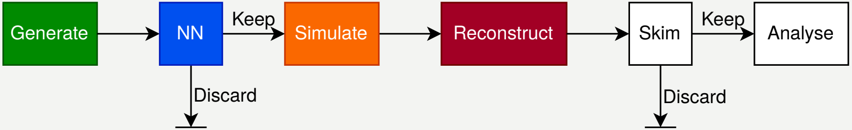
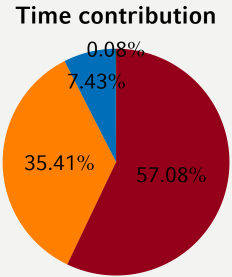


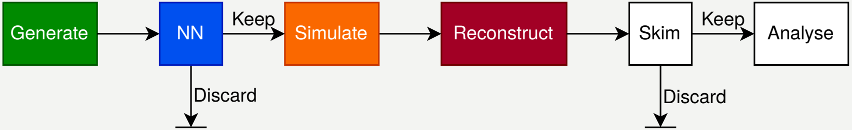
Figure: Example: Charged FEI

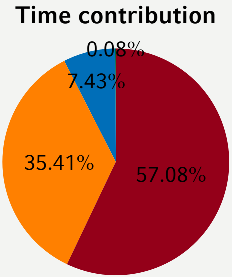
Application / Testing





Application / Testing





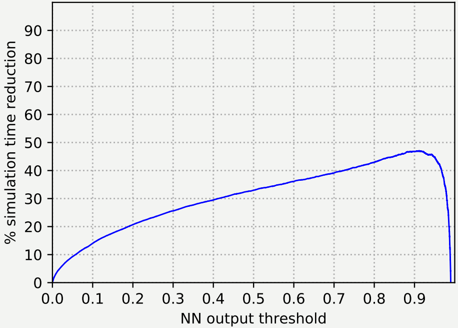
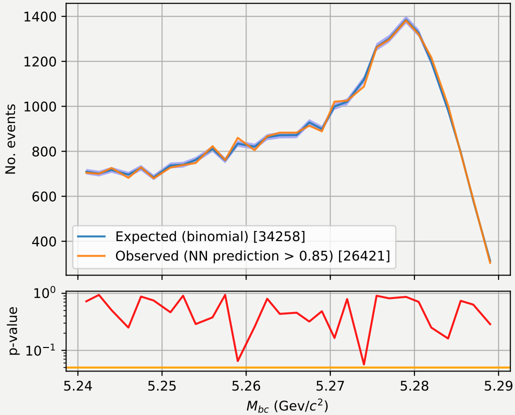
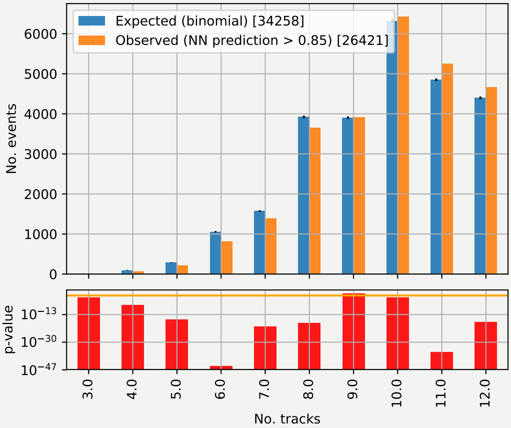


Figure: Example: Charged FEI

Evaluation

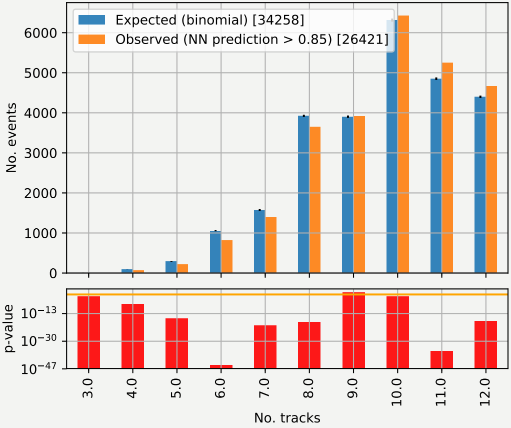
Compare kinematics before/after applying NN threshold:





Evaluation

Compare kinematics before/after applying NN threshold:



Investigating:

- Post-hoc reweighting
- Adversarial network
- Loss penalty (e.g. KL divergence)

Summary

- Belle II experiment presents new challenges in Monte Carlo simulation production
- Shift in focus from *simulate all* to *simulate necessary*
- Possible to preempt event usefulness in early simulation stages
- Requires careful bias consideration → adversarial / loss penalty
- Currently investigating potential for usage in ATLAS simulation