Application of deep learning to the analysis for $B \rightarrow K^*\gamma$ in Belle II

ABHINAV VISHNU, SAM CUNLIFFE, MALACHI SCHRAM, JAN STRUBE, CHARLES SIEGEL
Belle II Overview

- **Goal**: discover new particles and phenomena beyond the Standard Model of particle physics
- Collaboration among **725+ physicists** from 104 institutes in 24 countries
- **50x** the data volume, **40x** rate of collisions relative to previous Belle experiment
- **PNNL led U.S.** (DOE) contribution to Belle II detector construction – now complete
- **Largest** ever U.S. science investment in Japan
  - More Ph.D. physicists (50+) and more institutions (14) than any other country
- **SuperKEKB**: single beam circulation was done **successfully** (phase 1 in 2016)
- Cosmic-ray **data taking is on-going**
- **Physics** run will start in **2018**
Accelerating discovery in HEP/NP with scalable computing solutions

- PNNL leverages virtualization technologies used by PNNL’s team to enable quick solutions to access the LCFs resources at NERSC (Edison/Cori) and ORNL (Titan).
- Currently working on a large scale demonstrator for Distributed Large Scale Data Analytics using MaTEx.
A Belle II physics analysis

- B→K*γ is a Flavor Changing Neutral Current (FCNC) and sensitive to New Physics (NP) contributions to $C_7$
  - Also strong constraint to global NP-Wilson coefficient fits

Standard Model:

- Expect 50–70 events if we get $5fb^{-1}$ of physics Y(4S) next year
- Goal: “re-observation”, validation of detector performance, and benchmarking measurement of $A_{CP}$?
Looking for new physics

- $B \to K^* \gamma$ is a FCNC and sensitive to NP contributions to $C_7$
  - Also strong constraint to global NP-Wilson coefficient fits

Operator product expansion
(~model independent)

$$C_7 = C_7^{SM} + C_7^{NP}$$

Expect 50–70 events if we get 5fb$^{-1}$ of physics $Y(4S)$ next year

- Goal: “re-observation”, validation of detector performance, and benchmarking measurement of $A_{CP}$?
Keras+TensorFlow classifier study

- Keras+TensorFlow are open source 'standard' python ML libraries
  [https://keras.io/][https://www.tensorflow.org/]

- How well do these tools work for our use case? B→K*γ
  - **Pilot** investigation with Belle II simulation where we have well defined benchmarks (TMVA results)

- Same cocktail of variables as in TMVA training
- Only training against B background for now

23rd August 2017
Keras+TensorFlow classifier setup

1. Label samples (in this case: signal or BB background)
   • Can extend categories for continuum (maybe individual qq samples)
2. Split 0.33 test, 0.66 training
3. Normalize variables
4. Define 6 layer 'fully connected' neural network
   • Dropout rate @ 15% .. avoid over-fitting [https://arxiv.org/abs/1207.0580]
5. Use Adam optimiser [https://arxiv.org/abs/1412.6980v8] to minimize loss function = binary cross entropy
   • Minimise misclassification
6. Validation on 0.25 of training sample
   • Continual evaluation through epochs 'sanity check'
7. Train for 10 epochs
   • Can increase as we get more confident: computing resources etc.
Improvements seen with TensorFlow neural networks cf. TMVA

Becoming industry standard, actively maintained / improved

Benchmarking in 'easy' mode: precursor to trying out with more complex analysis (e.g. $B \to K^{(*)}\tau\tau$)
Preliminary Machine learning significance results for $B \to K^*\gamma$

You should compare the **green** LH to **blue** RH

The Kera+TensorFlow neural net has a smoother plateau at optimal

- Less systematics headache due to choice of cut point

- **45%** gain in **significance** = $S/\sqrt{S+B}$
Thoughts on Machine learning for $B \rightarrow K(\ast)\tau\tau$

- Much more complicated final states. Example fully hadronic decay:
  - 3 vertices
  - 8 tracks

- Use “low” level information such as vertices, tracks, track quality, PID, etc.

- Staged classification

23rd August 2017
MaTeX: Scalable Deep Learning Software on HPC

1) Open source software with users in academia, laboratories and industry
2) Supports graphics processing unit (GPU), central processing unit (CPU) clusters/LCFs with high-end systems/interconnects
3) Machine Learning Toolkit for Extreme Scale (MaTeX): [hpc.pnl.gov/matex]
4) 100+ visits/day, ~20 unique visits/day, 3-5 clones/day from github (> 1K clones)
MaTEx computing performance using HEP image like data

- Using up to 32 compute nodes on Cori provides \(~26\times\) speedup
- Conceptually similar to iTOP “ring image” DL study for Belle II PID
- Current evaluation and testing is undergoing on Titan and SummitDev architectures
The Belle II iTOP detector provides PID by comparing between the particle hypothesis and the “measuring” of the Cherenkov angle.

Currently using an analytical solution based on track momentum vector and impact parameters (angle/position).

Previously, minimization solutions were considered computing prohibitive because of the phase space and edge effects.

Scalable DL might provide a solution for a robust PID by training over the phase space using LCFs.
Grid Components for NERSC

**DIRAC**

**Workload Management Agent**
- **SiteDirector**
  - Using modified `GlobusComputingElement`

**Resource Definition**
- **OSG.CORI.us**
  - Defined as `GlobusComputingElement`
  - Mapped to PNNL StorageElement
- **OSG.EDISON.us**
  - Defined as `GlobusComputingElement`
  - Mapped to PNNL StorageElement

**Belle II HPC Docker**
- `/cvmfs/belle.cern.ch`
  - Sync repo to docker
- `pnnlhep/osg-compute MaTEx`

**NERSC**
- Pull and register into Edison/Cori shifter
- Grid submit with docker image and volume host/docker mount points for input/output/repo

**Docker**
- `pnnlhep/osg-compute MaTEx`
Deep learning is used for Belle II analysis

- TensorFlow is available within the Belle 2 software framework
- Kera+TensorFlow shows clear improvements in $B \to K^* \gamma$ analysis
- Investigating more complicated signals such as $B \to K^{(*)} \tau \tau$

Belle II are developing workflows that can use MaTEx for “big” jobs that use GPUs on Titan

- DL application on iTOP images can provide improved and robust PID

Belle II Grid analysis workflow chain are being developed to run on HPC