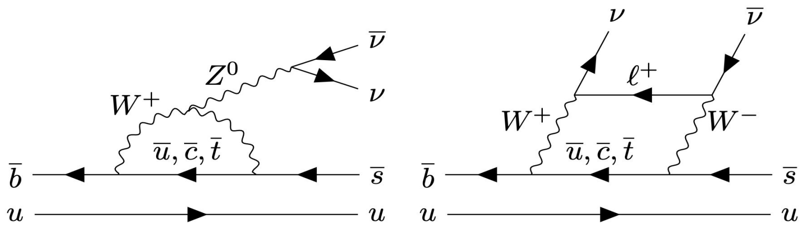
b->svv with inclusive tag

status and updates

Y. FAN, E. GANIEV, S. GLAZOV, Y. HAN, N. ROUT (DESY)
S. STEFKOVA (KIT), C. SCHMITT (LMU)
D. KULAKOV, A. KUCHER (KYIV)
M. LIU (JLU), F. LOZZI (Roma Sapienza/DESY)

Motivation

- b-> svv: a FCNC transition, offers a powerful opportunity to probe the SM
  - Highly suppressed in SM
  - Branching fraction precisely predicted: no charged leptons in the final state
  - Highly sensitive to non-SM contributions
- B-> K(*)vv @Belle II
  - B^+/B^0 decays into a pseudo-scalar (K^+/K_s^0) or a vector meson (K^{*+}/K^{*0})



Example of lowest-order quark-level diagrams for b->svv transition

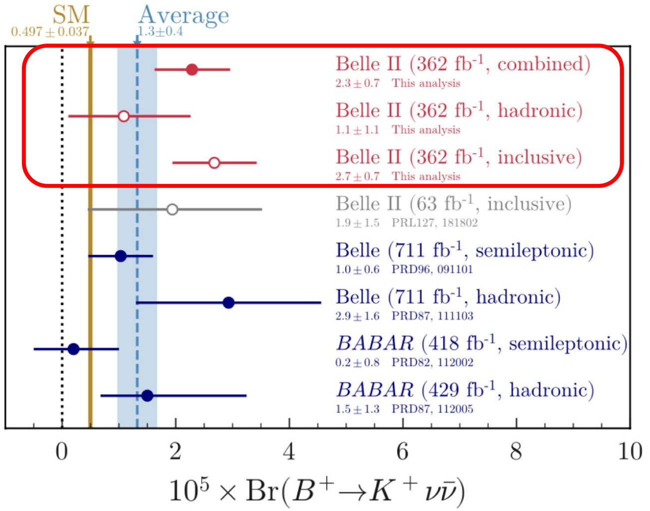
Decay	SM total (x10^{-6})
$B^+ \to K^+\nu\bar{\nu}$	5.06 ± 0.31
$B^0 \to K^0_S\nu\bar{\nu}$	2.05 ± 0.14
$B^+ \to K^{*+}\nu\bar{\nu}$	10.86 ± 1.43
$B^0 \to K^{*0}\nu\bar{\nu}$	9.05 ± 1.37

https://arxiv.org/abs/2301.06990

Where we were

- Previous measurements:
  - B^+-> K^+nunu measurement with Belle II Run 1 data
    - Inclusive tag, Hadronic tag and combined
    - 3.5σ wrt to null hypothesis and 2.7σ wrt SM
    - First evidence of B^+-> K^+nunu

    [PhysRevD.109.112006]



- Extend to K*, K_S^0 channels
  - Use the data sample collected in Run1
  - Challenge: low efficiency of final states with K_S^0s and pi^0s
  - Follow similar strategies with potential improvements
    - Selections, background suppression, new ML tools, etc.
- **Target: moriond 2025**

Workflow and Outline

Recap

The analyses follow a similar workflow as the last iteration:

1) Basic selection:
    - Object selection and event cleanup
    - Signal candidate selection
2) Main background suppression
    - BDT1 for event selection
    - Background suppression: e.g from D
    - Final selection using BDT2
3) Validation with control channels
4) Statistic interpretation

Workflow and Outline

Improvements wrt previous iteration

The analyses follow a similar workflow as the last iteration:

1) Basic selection:
   - Object selection and event cleanup
   - Signal candidate selection
2) Main background suppression
   - BDT1 for event selection
   - Background suppression: e.g from D
   - Final selection using BDT2
3) Validation with control channels
4) Statistic interpretation

Samples:

MC15ri, MC15rd exp18

Gitlab repository:

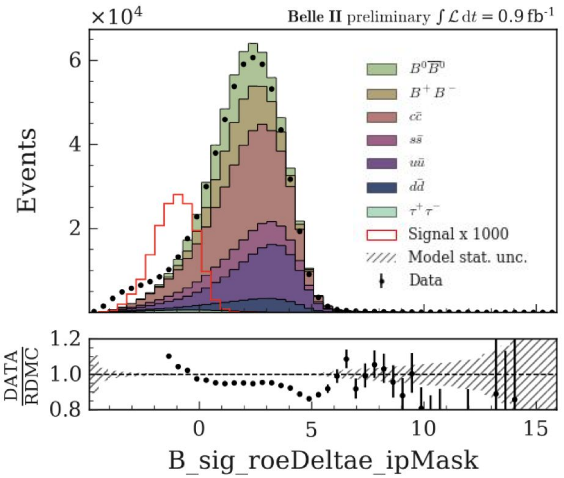
https://gitlab.desy.de/belle2/physics/ewp/b2hnunubar_ITA

1. Candiate selection
2. Optimize D veto
3. Explore new ML tools: DNN LNN
4. Look into important background components

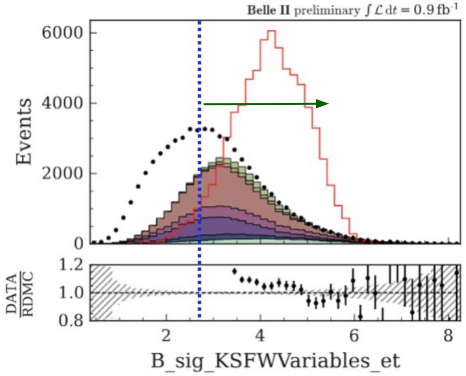
Reject low-multiplicity tracks

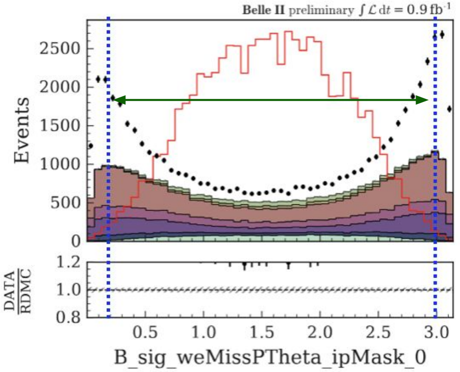
e.g. KS0 channel

- Discrepancy in the low DeltaE range comes from the low-multiplicity tracks: not well modeled in simulation
- Reject using selections on event variables: angle of missing momentum direction and transverse energy

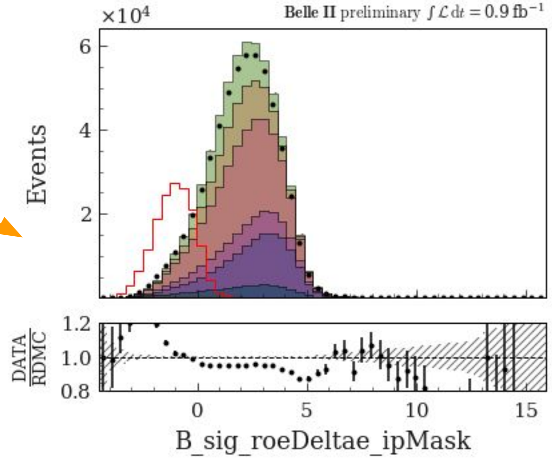


Zoomed in the low-multiplicity abundant region





Efficiency loss: 2.34%



➔ Unify these selections for all channels: study ongoing

Candidates selection strategy

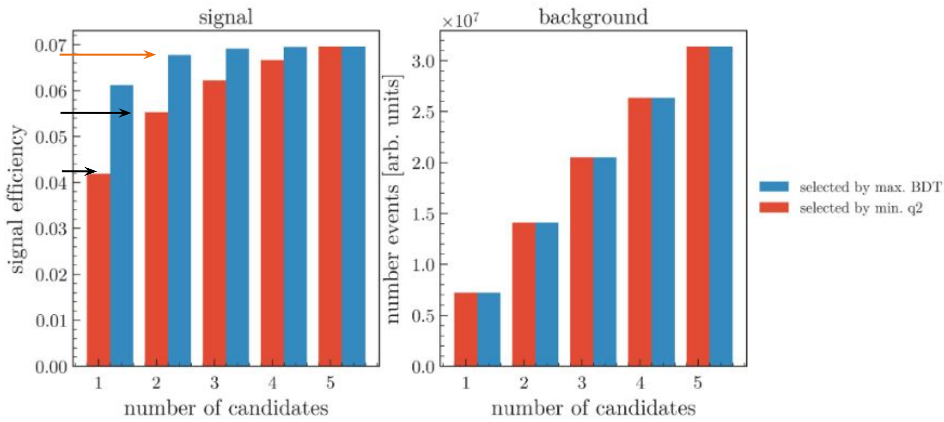
Multi-candidates to improve the signal efficiency:

- KS0/K+:
  - Apply best candidate selection at the basic selection step
    -> BDT1, etc…

- K*0/K*+: postponed best candidate selection to a later stage
  - Two candidates at the basic selection step
    -> BDT1
    -> Background Suppression
    -> Best candidate selection at BDT2

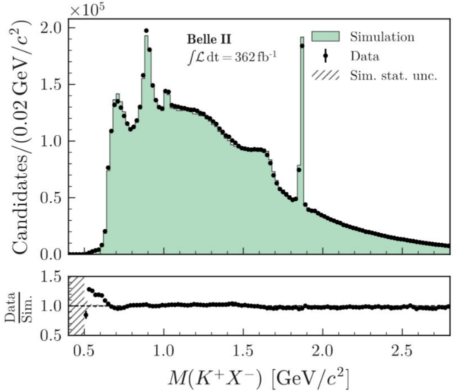
e.g. multi-candidates in the K*+ channel:

- Based on the BDT1 and q^2 info
- Improved signal efficiency while keeping background level under control

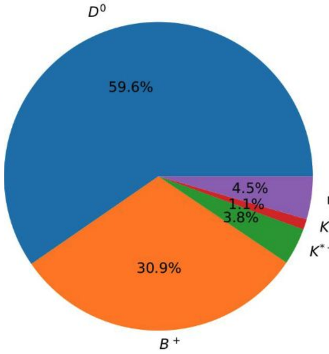


Signal efficiency and background events in K*+ channel as a function of multiple candidates

Improvements related to D veto



Background composition of K+ channel in charged B sample



Dominating background in signal region:

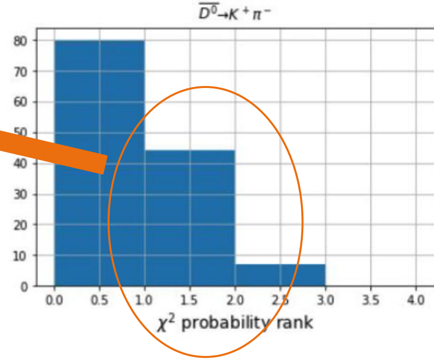
K+,KS0: kaon from D

K*: D →K*X and combinatorial background (kaon from D)

Previously, we suppressed with the reconstructed D0/D+ candidates:

- (KS0) Dp -> pi+Ks, D0-> pi+pi-Ks
- (K*0) Dp -> K*0 e/mu, D0-> K*0 pi+pi- , D0 -> K+ X, Dp-> K*0 X
- (K*+) Dp -> K*+ pi+pi-, D0 -> K*+ pi-, D0 -> K*- X, Dp-> K*+ XX (X: charged tracks)

○ Keep candidate with highest vertex fitting chiProb : can be improved



Ongoing studies and developments for improvement:

➔ Detailed categorization of the background

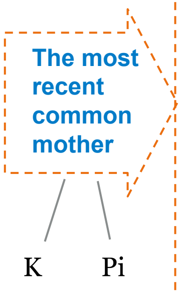
➔ Exploration of new methods and additional variables for improving background suppression

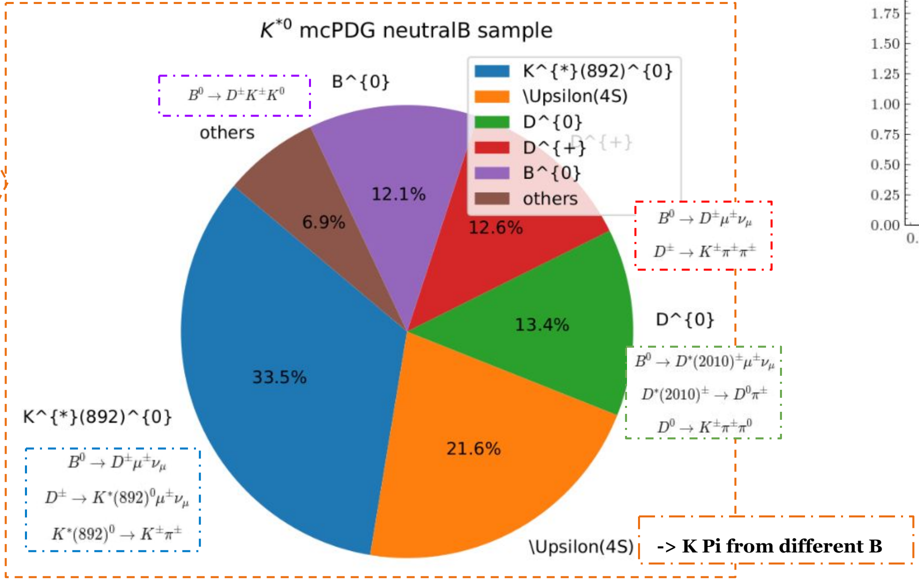
Optimization of the D veto

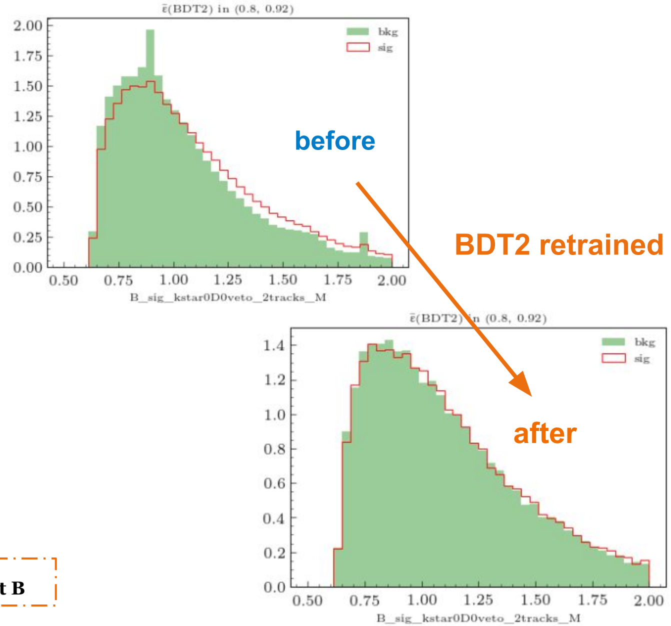
e.g. K*0

1) Background composition -> which particle is mimicking signal K*0
2) Veto strategy:
   a) Reconstruct the particles mimicking K*0
   b) Save and pass the variables to BDT2
3) Check the background composition again

Optimization is ongoing (also for other channels)



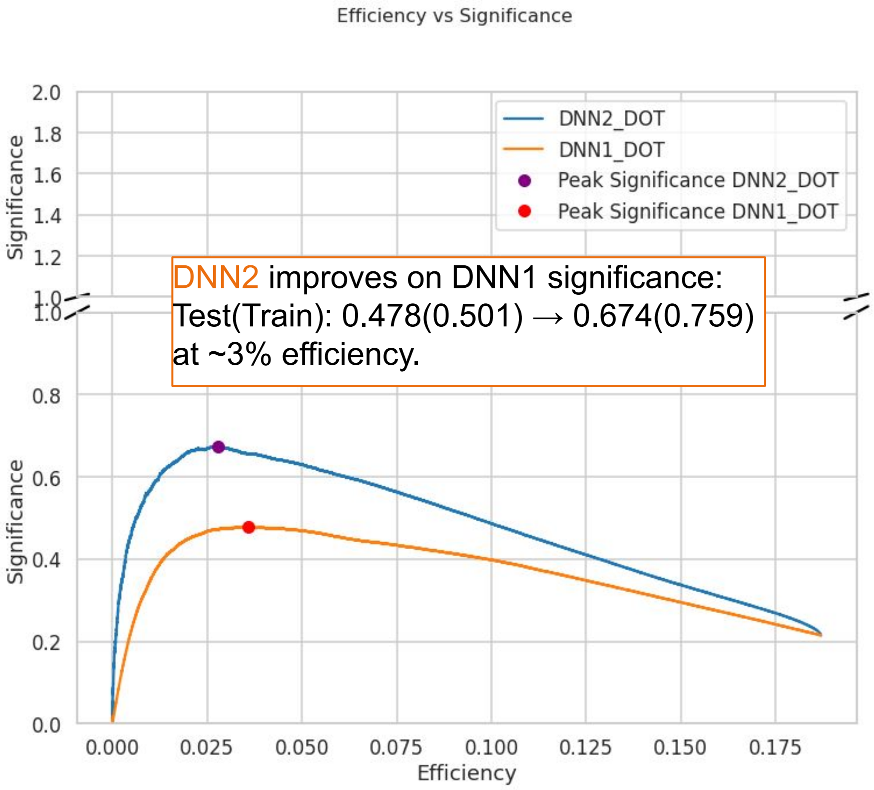




Explore other ML tools: DNN

- Performance is being checked in K*0 channel
    - Train DNN1 and DNN2 as potential alternatives to BDT1/BDT2
    - Good discriminating power

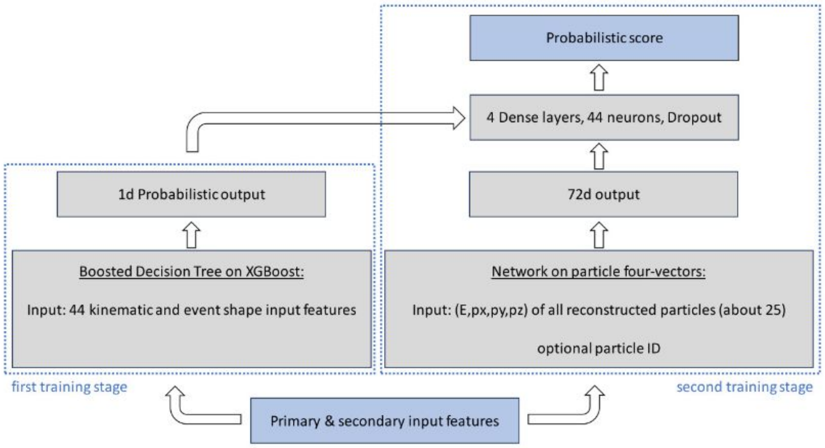
- **Promising starting point, can be further optimised in the future!**

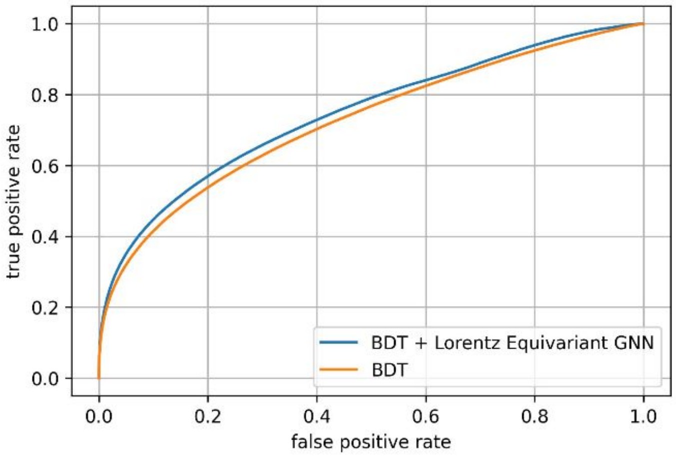


Explore new chance with new architecture

BDT combined with Lorentz equivariant architecture

- Lorentz equivariant Net, similar to a GNN, respects the Lorentz group symmetries of 4-momenta
- Combine BDT1,2 and a LNN on particle 4-vectors for signal-background classification:
    - **Looks promising with initial setup, and further investigation is ongoing …**
    - Talk @EWP



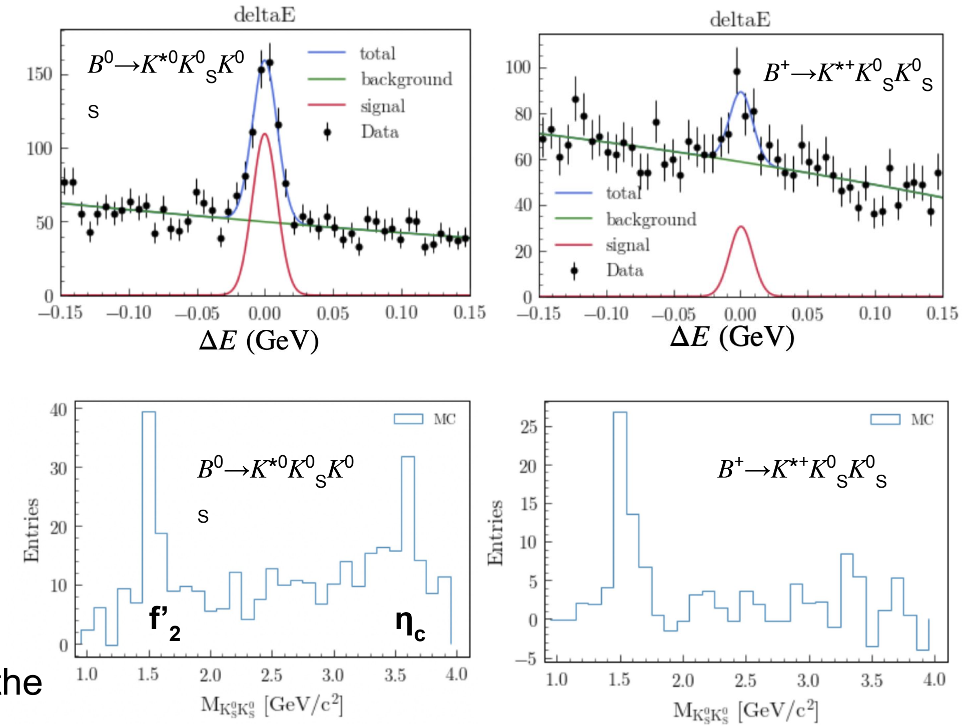


Background from $B \to K^* K^0 \bar{K}^0$

- $B \to K^* K^0 \bar{K}^0$ can mimic our signal
  => Essential to constrain and model them well

- Use $B \to K^* K^0_S K^0_S$ decays to model $B \to K^* K^0_L K^0_L$
- $B \to K^* K^0_S \bar{K}^0_S$ have never been measured before
  => Measure them ourselves

- For now, use generic 400 fb^{-1} MC15ri sample
  - Reconstruct $B \to K^* K^0_S K^0_S$
  - Fit ΔE distribution
  - Apply sPlot technique to obtain signal-only $M(K^0_S K^0_S)$ distributions
    => In simulation, prominent f'$_2$ resonance
    => However, no $B \to K^*$f'$_2$ in PDG
  - Use the obtained model from data to reweight the phase space
    => Now preparing to look in data



Modeling of K0K0K0

Replace the relevant decays before BDT2 training

B0->KS0K0K0 is an important background for KS0 channel and not well modeled in simulation
=> highly overestimated in rel-06
=> double counting of K0K0K0 is fixed in rel-08

Belle2 Decay file (rel6)

```
# PPP
#
0.000006200    K_S0  K_S0  K_S0                PHSP; #Measured, inclusive
0.000006200    K_L0  K_L0  K_L0                PHSP; #Same as KsKsKs
0.000007400    K_S0  K_S0  K_L0                PHSP; #UL
0.000007400    K_S0  K_L0  K_L0                PHSP; #Same as KsKsKL

0.000048000    K0   anti-K0   K0               PHSP; #Assume 8*BF(KsKsKs)~=48 (8 is a guess).
```

BF used to generate signal MC

```
0.000006000   K_S0  K_S0  K_S0        K0K0K0
0.000002460   K_S0  K_S0  K_L0        K0K0K0
0.000002460   K_S0  K_L0  K_L0        K0K0K0
```

(BarBar)
PhysRevD.85.054023

➔ Correction before BDT2 training
- Replace the relevant decays in generic simulation with K0K0K0 (Dalitz) and up-to-date branching fractions
- B->Φ KL/KS is scaled to the latest PDG value
- Improves background modeling

Summary and outlook

The analyses follow a similar workflow as the last iteration:

1) Basic selection:
   - Basic event selection
   - Signal candidate selection
2) Main background suppression
   - BDT1 for event selection
   - Background suppression: e.g from D
   - Final selection using BDT2
3) Validation with the control channel
4) Statistic interpretation

The analysis is progressing efficiently, with efforts aimed at improving various areas.

Target: moriond 2025

Where we were

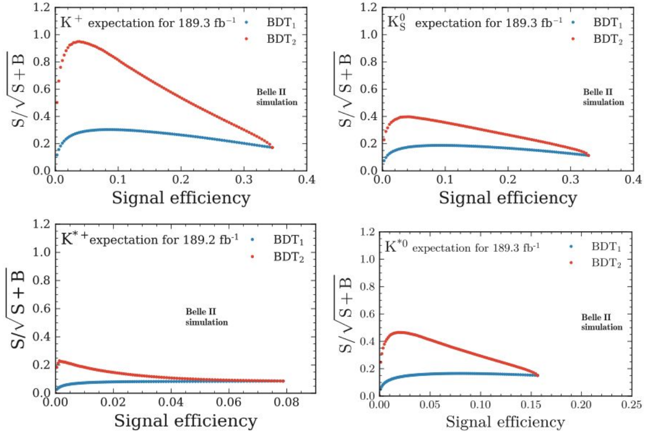
B+ -> K+ nunu

Previously:

- $B^0 \to K^{*0}$, $B^+ \to K^{*+}$ and $B^0 \to$ KS0 were studied pretty detailed for Moriond 2022
- Data sample: (collected by summer 2021) 189.2 fb^{-1} and 17.9 fb^{-1} on- and off-resonance samples
- MC14ri samples
- The measurements were almost ready to be unblinded and code machinery is well developed

Follow similar workflow as last iteration:

1) Basic selection:
    - Signal candidate selection
    - Basic event selection
2) Main background suppression
    - BDT1 for event selection
    - Final selection using BDT2 or BDT2+GraphNN (K*+)
3) Validation with control channel
4) Statistic interpretation



Explore new chance with NN

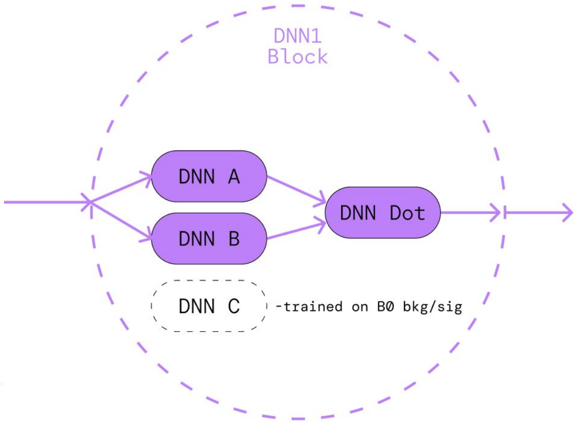
DNN Architecture: 16 Layers, 279 Features.

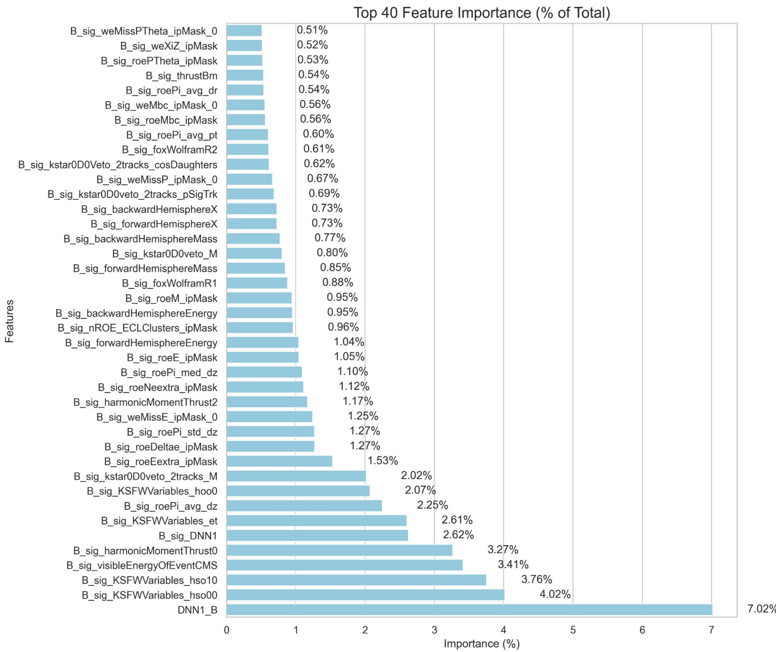
```python
class NeuralNetwork(nn.Module):
    def __init__(
        self,
        input_dim,
        hidden_layers=[1024, 512, 512, 256],
        dropout_rate=0.4,
        alpha=0.01,  # for LeakyReLU
    ):
        super(NeuralNetwork, self).__init__()
        self.input_dim = input_dim
        self.hidden_layers = hidden_layers
        self.dropout_rate = dropout_rate

        layers = []
        prev_dim = input_dim
        for units in hidden_layers:
            layers.extend(
                [
                    nn.Linear(prev_dim, units),
                    nn.LeakyReLU(alpha),
                    nn.BatchNorm1d(units),
                    nn.Dropout(dropout_rate),
                ]
            )
            prev_dim = units
        layers.append(nn.Linear(prev_dim, 1))
        self.model = nn.Sequential(*layers)

    def forward(self, x):
        return self.model(x)
```

- BatchNorm and Dropout layers reduce model overfitting.
- Use appropriate model complexity.
- ReLU family of activation functions fix vanishing gradient problem.
- DNN1 is composed of 3 models dnn1a, dnn1b and dnn1dot which combines their scores.

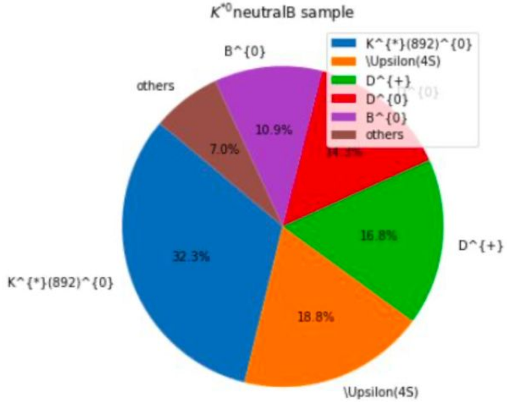




Which particles(decays) mimic signal K*0(decays) in neutral B background

BDT2>0.92

D-veto



	formatted_dec	multiplicity
0	$\Upsilon(4S) \to D^*(2010)^{\pm}\mu^{\pm}\nu_{\mu}$	0.091503
1	$\Upsilon(4S) \to D^*(2010)^{\pm}e^{\pm}\nu_{e}$	0.065359
2	$\Upsilon(4S) \to D^*(2010)^{\pm}\tau^{\pm}\nu_{\tau}$	
3	$\Upsilon(4S) \to D^{\pm}\mu^{\pm}\nu_{\mu}$	
4	$\Upsilon(4S) \to D^*(2010)^{\pm}e^{\pm}\nu_{e}\gamma$	

0	$B^0 \to D^{\pm}K^{\pm}K^0$	0.045198
1	$B^0 \to D^*(2010)^{\pm}e^{\pm}\nu_e$	0.031073
2	$B^0 \to D^*(2010)^{\pm}\tau^{\pm}\nu_{\tau}$	0.028249
3	$B^0 \to D^{\pm}\tau^{\pm}\nu_{\tau}$	0.028249
4	$B^0 \to D^*(2010)^{\pm}K^{\pm}K^0$	0.028249

currently no idea

	formatted_dec	multiplicity
0	$B^0 \to D^{\pm}\mu^{\pm}\nu_{\mu}$ $D^{\pm} \to K^*(892)^0\mu^{\pm}\nu_{\mu}$ $K^*(892)^0 \to K^{\pm}\pi^{\pm}$	0.040913
1	$B^0 \to D^{\pm}\mu^{\pm}\nu_{\mu}$ $D^{\pm} \to K^*(892)^0 e^{\pm}\nu_e$ $K^*(892)^0 \to K^{\pm}\pi^{\pm}$	0.035205
2	$B^0 \to K^0 K^0 K^*(892)^0$ $K^*(892)^0 \to K^{\pm}\pi^{\pm}$	0.030447
3	$B^0 \to D^{\pm}e^{\pm}\nu_e$ $D^{\pm} \to K^*(892)^0 e^{\pm}\nu_e$ $K^*(892)^0 \to K^{\pm}\pi^{\pm}$	0.029496
4	$B^0 \to D^{\pm}\mu^{\pm}\nu_{\mu}$ $D^{\pm} \to K^*(892)^0\pi^{\pm}\pi^0$ $K^*(892)^0 \to K^{\pm}\pi^{\pm}$	0.023787

true K*0 from D->K*0 l nu:
new possible veto:
K*0:sig + ml lepton/pion

	formatted_dec	multiplicity
0	$B^0 \to D^{\pm}\mu^{\pm}\nu_{\mu}$ $D^{\pm} \to K^{\pm}\pi^{\pm}\pi^{\pm}$	0.170330
1	$B^0 \to D^{\pm}e^{\pm}\nu_e$ $D^{\pm} \to K^{\pm}\pi^{\pm}\pi^{\pm}$	0.093407
2	$B^0 \to D^{\pm}e^{\pm}\nu_e\gamma$ $D^{\pm} \to K^{\pm}\pi^{\pm}\pi^{\pm}$	0.049451
3	$B^0 \to D^*(2010)^{\pm}\mu^{\pm}\nu_{\mu}$ $D^*(2010)^{\pm} \to D^{\pm}\pi^0$ $D^{\pm} \to K^{\pm}\pi^{\pm}\pi^{\pm}$	0.045788
4	$B^0 \to D^{\pm}\tau^{\pm}\nu_{\tau}$ $D^{\pm} \to K^{\pm}\pi^{\pm}\pi^{\pm}$	0.042125

new possible veto:
Dp->(sig)K+2(roe)pi

	formatted_dec	multiplicity
0	$B^0 \to D^*(2010)^{\pm}\mu^{\pm}\nu_{\mu}$ $D^*(2010)^{\pm} \to D^0\pi^{\pm}$ $D^0 \to K^{\pm}\pi^{\pm}\pi^0$	0.126882
1	$B^0 \to D^*(2010)^{\pm}e^{\pm}\nu_e$ $D^*(2010)^{\pm} \to D^0\pi^{\pm}$ $D^0 \to K^{\pm}\pi^{\pm}\pi^0$	0.070968
2	$B^0 \to D^*(2010)^{\pm}\mu^{\pm}\nu_{\mu}\gamma$ $D^*(2010)^{\pm} \to D^0\pi^{\pm}$ $D^0 \to K^{\pm}\pi^{\pm}\pi^0$	0.032258
3	$B^0 \to D^*(2010)^{\pm}e^{\pm}\nu_e\gamma$ $D^*(2010)^{\pm} \to D^0\pi^{\pm}$ $D^0 \to K^{\pm}\pi^{\pm}\pi^0$	0.030108
4	$B^0 \to D^*(2010)^{\pm}\tau^{\pm}\nu_{\tau}$ $D^*(2010)^{\pm} \to D^0\pi^{\pm}$ $D^0 \to K^{\pm}\pi^{\pm}\pi^0$	0.030108

new possible veto:
D0->(sig)K (roe)pi pi0

Basic selection

e.g. K*0 channel

- Signal selection efficiency drops a lot for K*0 channels (initial starting point)
    - Good tracks: |dr|<0.5 cm, |dz|<3 cm, pT>0.1 GeV/c, E<5.5 GeV, θ ∈ CDC
    - Cuts on inv.mass of K*0
    - KaonID> 0.9: reduce the multiplicity
    - Number of PXD hits: accurate vertex
    - Best candidate selection

Selection efficiency drops from 45.6% to 23.7%

	efficiency	candidate multiplicity
Track clean up	45.6%	13.9
+ 0.8 < Kstar_M < 1.0	39.7%	5.2
+Kstar_K_kaonID > 0.9	30.2%	1.8
+Kstar_K_nPXDHits>0	28.7%	1.8
+Kstar_Pi_nPXDHits > 0	27.4%	1.7
+Best candidate (lowest q2)	23.7%	1.0

To be improved