A possible TMVA application @ Online Software Trigger

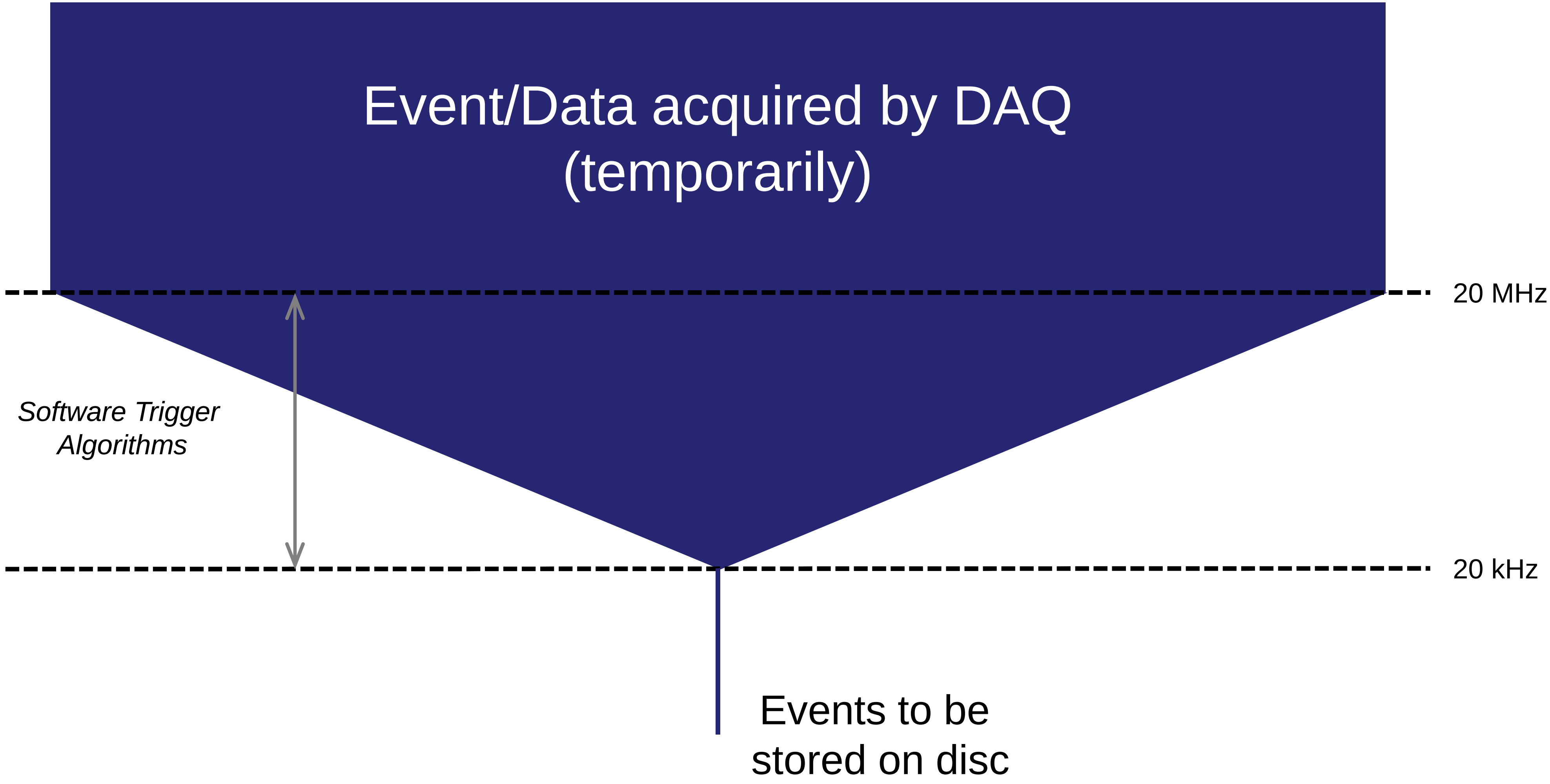
Donghee Kang

Institut für Kernphysik, Universität Mainz

- Study on the event shape with full PANDA simulation

- TMVA application for the event shape variables

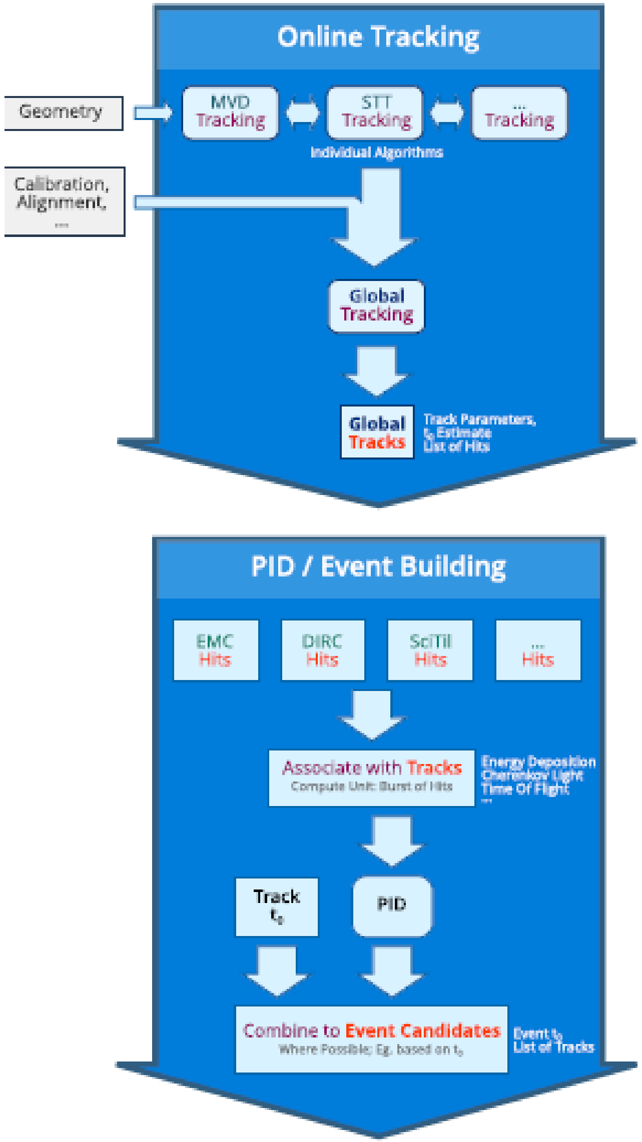
Event rate @ PANDA



Required background reduction rate 1/1000 in total
by means of
online software trigger with high selectivity

Trigger concept

Event Building



Software Trigger

algorithms

$D^0(K\pi)$
$D^0(K\pi\pi^0)$
$D^\pm(K\pi\pi)$
$J/\psi(e^+e^-)$
$J/\psi(\mu^+\mu^-)$
$J/\psi(\pi^+\pi^-\pi^0)$
$D^0(\mu^\pm e^\mp)$
$D^\pm(\pi^\pm e^\mp \mu^\pm)$
$D^0(\mu^+\mu^-)$
e^+e^-
$\phi(K^+K^-)$
$D_s(\phi\pi)$
$\eta_c(\gamma\gamma)$
$\Lambda(p\pi)$
$\Lambda_c^+(pK^-\pi^+)$
$h_c(\gamma\gamma\gamma)$
.
.
.

decision

select event
if one of algorithms
is fulfilled

reject event
if none of algorithms
are satisfied

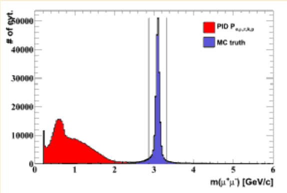
Software Trigger

Event flow

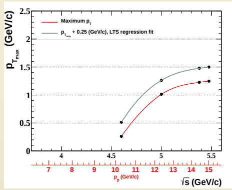
Track/PID candidates

Combinatorial (charged, neutral)

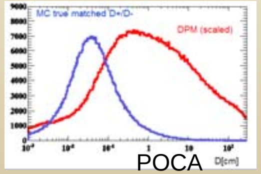
Mass cut (resonances)



Kinematic cut (e.g. p_T)



POCA/Vertex cut



trigger decision (multiple line)

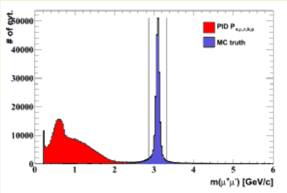
Application of event shape

Event flow

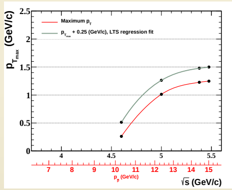
Track/PID candidates

Combinatorial (charged, neutral)

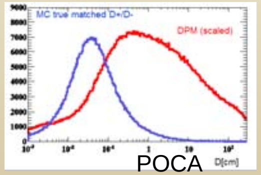
Mass cut (resonances)



Kinematic cut (e.g. p_T)



POCA/Vertex cut



trigger decision (multiple line)

Event shape variables

- Multiplicities of all, charged, neutrals, above a certain momentum threshold, with certain PID quality
- Maximum/Minimum momenta, transverse momenta, cluster energy, transvers energy (cms & lab system)
- Sums of momenta, transverse momenta, energies, transverse energies (cms & lab system)
- Event shape variables like sphericity, aplanarity, planarity, thrust, Fox Wolfram moments (usually cms system)
- Detector information like total size of hits at MUO, STT, MVD and cluster size of EMC (new feature!)

Event shape @ generated MC

With toy MC simulation using cuts on event shape variables
pure reduction of DPM background is found to be a level of few %

Performance of Event Cuts

Energy/channel	Pure eventshape cut		Combined with simultaneous cut			
2.4 GeV	sig eff	dpm eff	eff	eff w/ cut	rel eff	cut
phi (KK) phi	90%	0,6%	95%	90%	94%	nk>1 & sumpc<1.632 & pmax<0.6
Lam(ppi) Lamb	93%	0,2%	96%	91%	95%	npr>0 & npart>3 & npi<4 & fw1>0.132 & fw4>0.2112
DPM			9%	1%	6%	Faktor 16,04
3.77 GeV	sig eff	dpm eff	eff	eff w/ cut	rel eff	cut
phi (KK) phi	93%	0,2%	96%	93%	98%	nk>1 & npart>3 & fw2>0.5456
Lam(ppi) Lamb	92%	0,2%	96%	91%	94%	npr>0 & npart>3 & fw2>0.6248 & fw4>0.3784 & fw5>0.22
J/psi(ll) pipi	91%	0,0%	89%	83%	93%	np10>1 & npi<4 & pmax>1.464 & sumpc>3.424 & fw2<0.8976
D0(Kpi) D0b	91%	2,1%	92%	80%	86%	nk>0 & sumptl>1.38 & pmax>0.768 & pmax<1.056 & fw3<0.1848
D+-(Kpipi) D-+	90%	4,3%	91%	78%	85%	nk>0 & sumptl>1.5 & pmax<0.936
DPM			18%	2%	12%	Faktor 8,52
4.28 GeV	sig eff	dpm eff	eff	eff w/ cut	rel eff	cut
phi (KK) phi	91%	0,1%	96%	90%	94%	nk>1 & sumptl>0.96 & fw2>0.6424
Lam(ppi) Lamb	93%	0,1%	96%	91%	95%	npr>0 & npart>3 & fw2>0.7216 & fw4>0.484 & fw5>0.1408
J/psi(ll) pipi	92%	0,0%	89%	83%	94%	npi<4 & np05>2 & sumpc>3.904 & ptmax>0.612
D0(Kpi) D0b	91%	1,9%	93%	74%	80%	nk>0 & pmax>0.936 & sumptl>2.1
D+-(Kpipi) D-+	90%	5,3%	92%	79%	86%	nk>0 & ptmax>0.516 & sumptl>1.8
Ds+-(KKpi) Ds-+	92%	1,5%	93%	80%	86%	nk>1 & npi<7 & sumptl>1.62
DPM			21%	3%	13%	Faktor 7,74
5.0 GeV	sig eff	dpm eff	eff	eff w/ cut	rel eff	cut
phi (KK) phi	91%	0,1%	96%	90%	94%	nk>1 & sumptl>1.14 & fw2>0.7392
Lam(ppi) Lamb	95%	0,1%	96%	91%	95%	npart>3 & ptmax>0.504 & fw1>-0.0176 & fw2>0.8008
J/psi(ll) pipi	92%	0,0%	89%	84%	94%	npi<4 & sumpc>4.608 & ptmax>0.828
D0(Kpi) D0b	91%	1,4%	93%	75%	81%	nk>0 & pmax>1.152 & sumptl>2.4
D+-(Kpipi) D-+	91%	5,2%	92%	81%	88%	nk>0 & ptmax>0.6 & sumptl>2.04
Ds+-(KKpi) Ds-+	92%	1,6%	94%	82%	87%	nk>1 & npi<7 & sumptl>1.92
Lc (pKpi) Lcb	91%	2,8%	95%	82%	86%	nk>0 & npr>0 & npart>5 & sumpt>1.5 & fw1>-0.0176
DPM			26%	3%	13%	Faktor 7,82

simultaneous tagging with 7 algo.

eff.(bk) = 3% @ 5.0 GeV

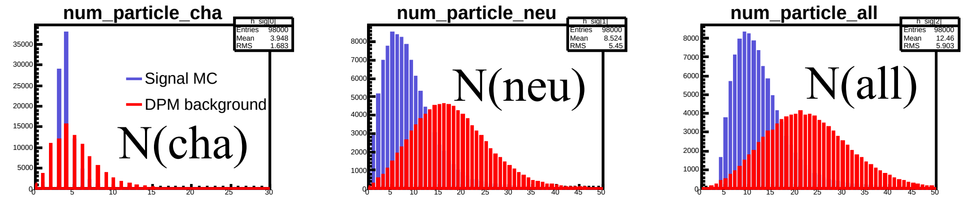
Event shape distribution

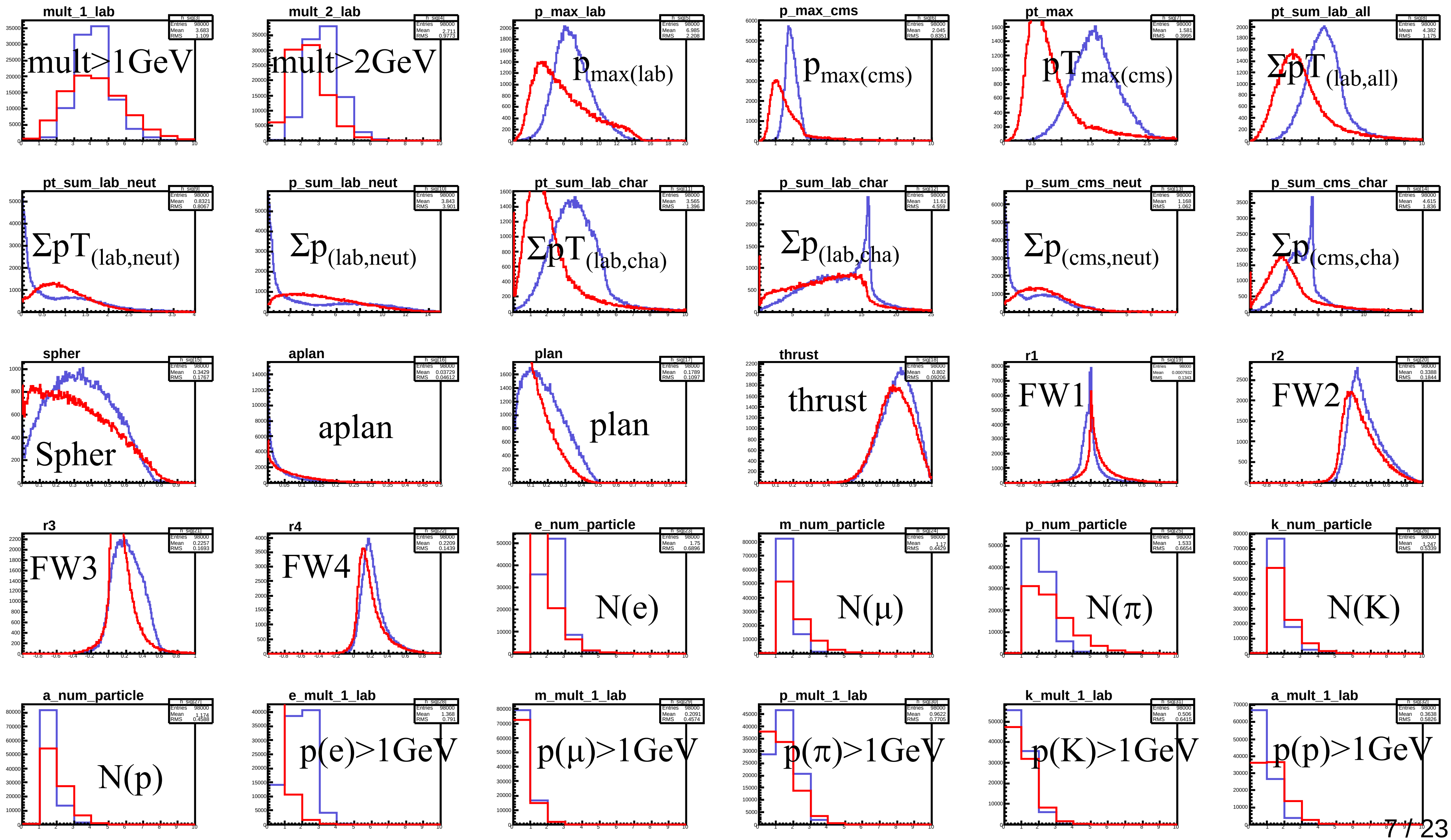
Full MC simulation (Geant3) :

$$\bar{p}p \rightarrow J/\psi\pi^+\pi^- \rightarrow l^+l^-\pi^+\pi^-$$

$$\bar{p} = 15.0\text{GeV/c}\ (\sqrt{s} = 5.474\text{GeV/c})$$

$$\text{Prob.}(e, \mu, \pi, K, p) > 0.25$$ (Loose)





Optimization from distribution

Mode	beam mom. = 15 GeV/c	
Category	Siganl MC	DPM background
$D^0 \to K^-\pi^+$	83.3 %	45.1 %
$D^+ \to K^-\pi^+\pi^+$	84.3 %	49.0 %
$D_s^+ \to \phi\pi^+$	91.9 %	68.3 %
$pp \to e^+e^-$	81.8 %	3.3 x 10^{-5}
$J/\psi\pi^0 \to e^+e^-\gamma\gamma$	81.7 %	1.4 %
$J/\psi\pi^+\pi^- \to e^+e^-\pi^+\pi^-$	82.3 %	10.8 %
$J/\psi\pi^+\pi^- \to \mu^+\mu^-\pi^+\pi^-$	81.3 %	6.8 %
$\Lambda_C \to pK^-\pi^+$	86.2 %	52.8 %
$\Lambda \to p\pi^-$	91.0 %	43.5 %
$\phi \to K^+K^-$	89.3 %	46.8 %
$D^0 \to K^-\pi^+\pi^0$	82.5 %	45.2 %
$D^0 \to K^-\pi^+\pi^+\pi^-$	89.1 %	54.5 %
$D^+ \to K^-\pi^+\pi^+\pi^0$	82.3 %	49.7 %
$D^+ \to K_s\pi^+\pi^+\pi^-$	91.9 %	63.1 %
$D^+ \to K_s\pi^+\pi^0$	88.0 %	58.9 %
$D^+ \to \pi^+e^-\mu^+$	75.5 %	19.1 %
$D_s^+ \to \phi\pi^+\pi^0$	90.2 %	62.2 %

- Find cuts from left/right of event shape distributions : R_{signal} - R_{Back} > 1%
- Background survive more than 68 % (maximum @ D_s^+) by an event shape cut for single category
- Simultaneous tagging with all 17 algorithms
  eff.$_{Background}$ = 95 %

not suitable for tagging

TMVA application

Is it possible to use MVA approach for event shape variables?

- only need a lot of sets of tested/trained data samples according physics channels

- evaluation @ online stream should fast enough because of direct accessing the lookup table via *.xml format

Introduction of TMVA

TMVA (Toolkit for Multivariate Data Analysis with ROOT)

- Evaluation of multi-variated classification by supervised learning algorithms

- make a decision boundary from a set of trained events/data/algorithms

Assessment of MVA properties.

good (**) fair (*), bad(◦)

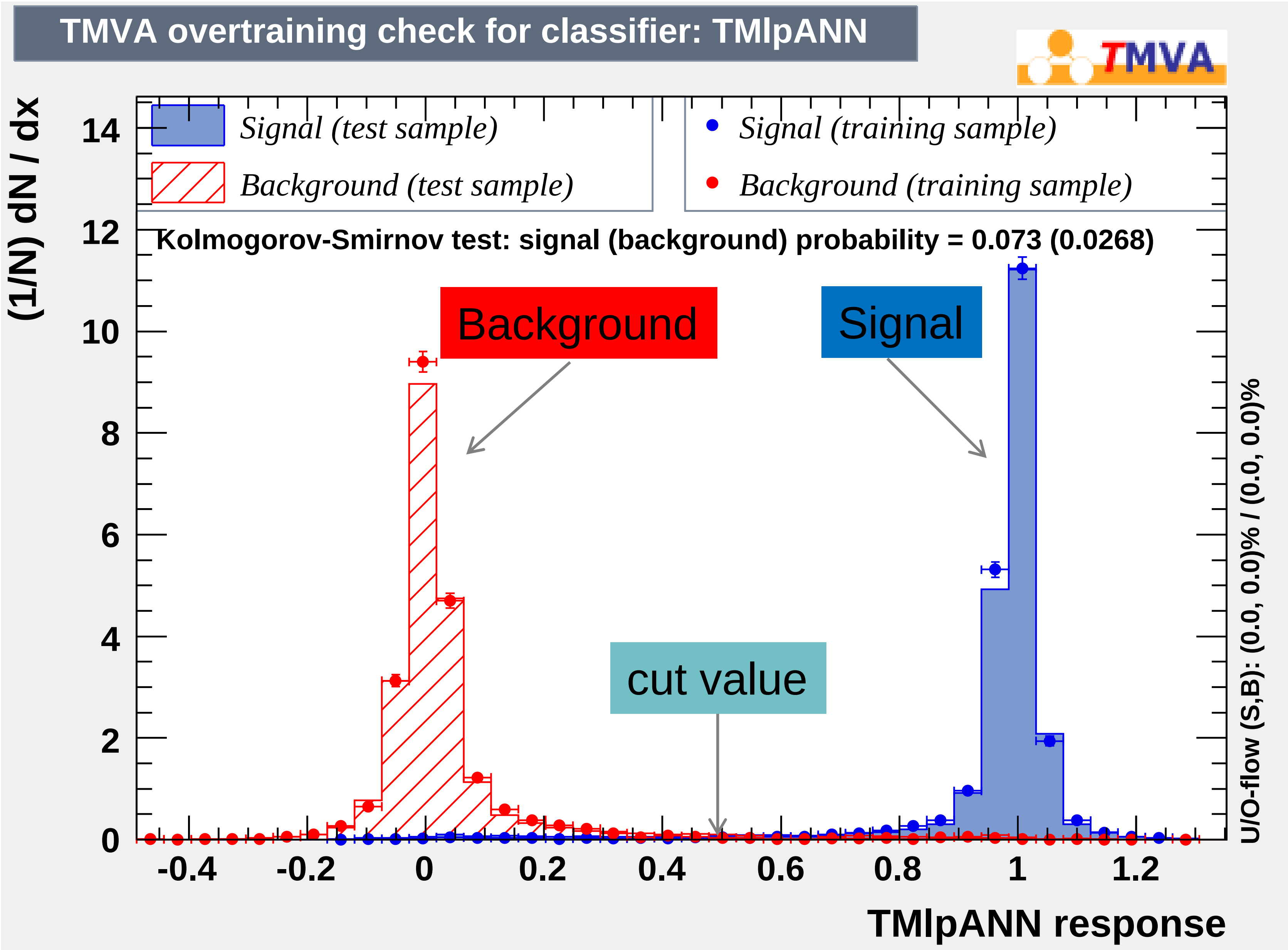
CRITERIA		Cuts	Likelihood	PDE-RS / k-NN	PDE-Foam	H-Matrix	Fisher / LD	MLP	BDT	Rule-Fit	SVM
		MVA METHOD									
Performance	No or linear correlations	⋆	⋆⋆	⋆	⋆	⋆	⋆⋆	⋆⋆	⋆	⋆⋆	⋆
	Nonlinear correlations	◦	◦	⋆⋆	⋆⋆	◦	◦	⋆⋆	⋆⋆	⋆⋆	⋆⋆
Speed	Training	◦	⋆⋆	⋆⋆	⋆⋆	⋆⋆	⋆⋆	⋆	⋆	⋆	◦
	Response	⋆⋆	⋆⋆	◦	⋆	⋆⋆	⋆⋆	⋆⋆	⋆	⋆⋆	⋆
Robustness	Overtraining	⋆⋆	⋆	⋆	⋆	⋆⋆	⋆⋆	⋆	⋆[39]	⋆	⋆⋆
	Weak variables	⋆⋆	⋆	◦	◦	⋆⋆	⋆⋆	⋆	⋆⋆	⋆	⋆
Curse of dimensionality		◦	⋆⋆	◦	◦	⋆⋆	⋆⋆	⋆	⋆	⋆	
Transparency		⋆⋆	⋆⋆	⋆	⋆	⋆⋆	⋆⋆	◦	◦	◦	◦

Working principle

$\bar{p}p \rightarrow J/\psi\pi^+\pi^- \rightarrow l^+l^-\pi^+\pi^-$ @ $\sqrt{s} = 5.474$ GeVc(signal)

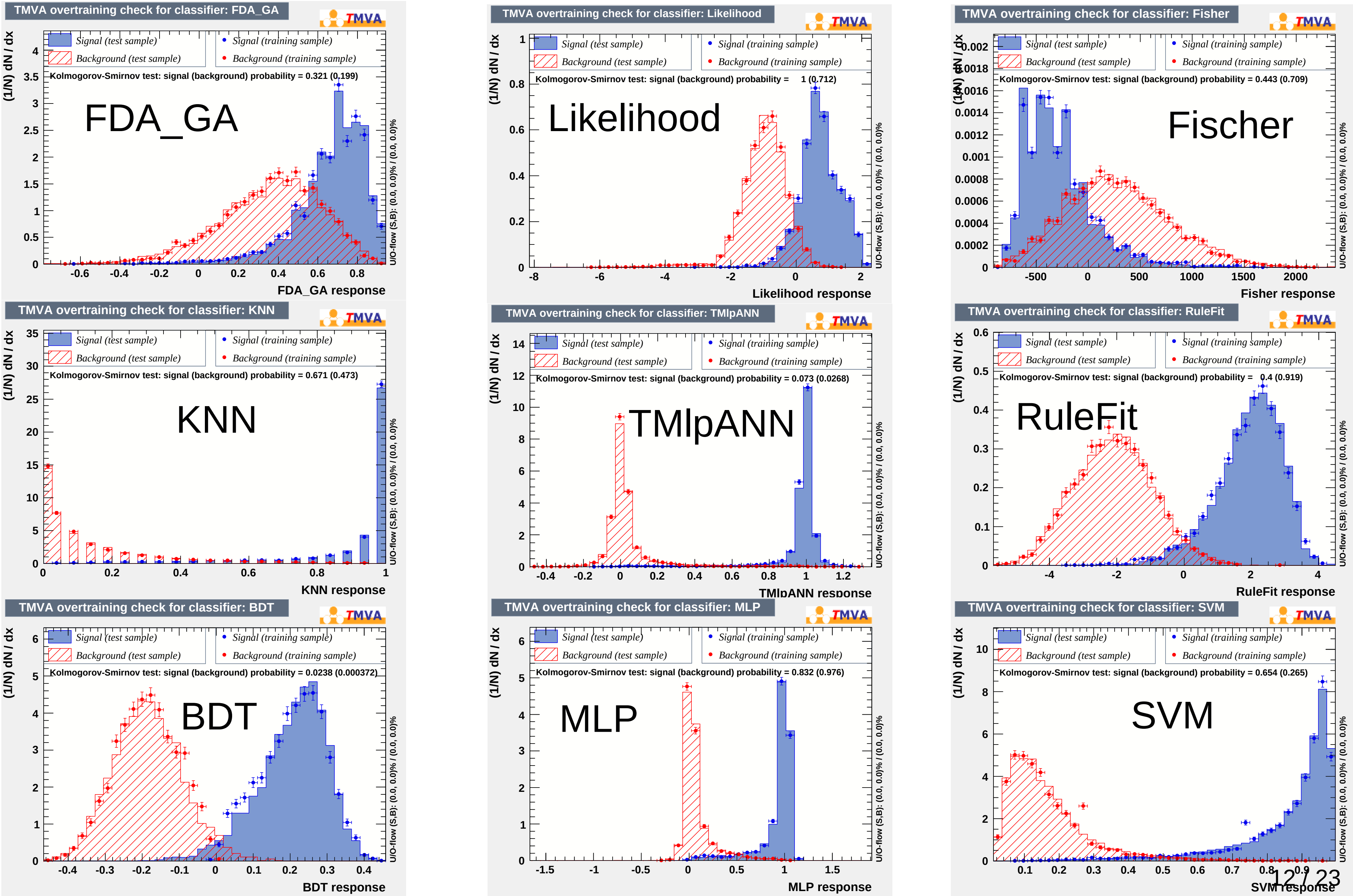
$\bar{p}p \rightarrow$ *generic DPM* @ $\sqrt{s} = 5.474$ GeV/c(background)

Artificial Neural Network



Classification for J/ψ

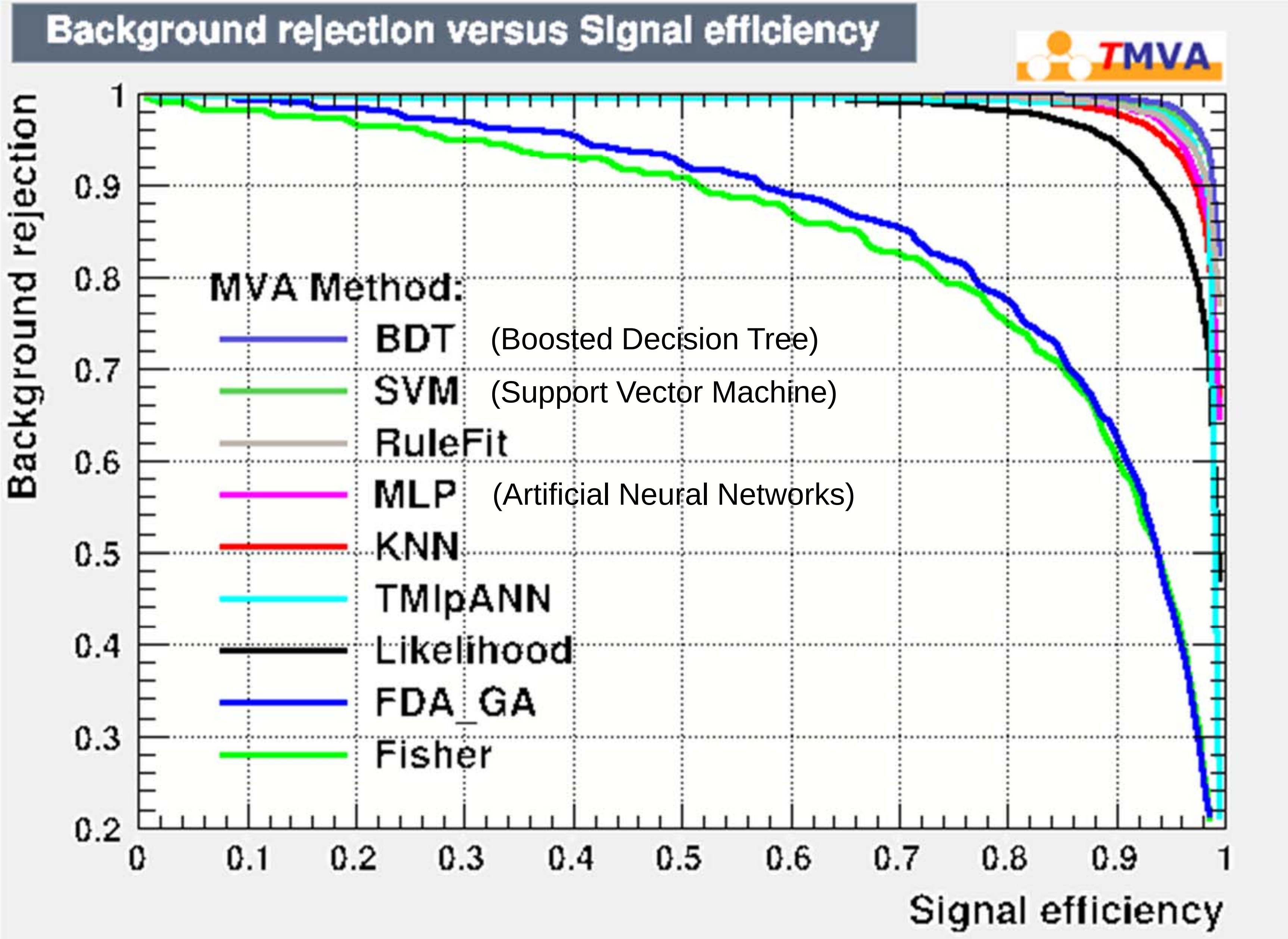
$$\bar{p}p \to J/\psi\pi^+\pi^- \to l^+l^-\pi^+\pi^- \ @\sqrt{s} = 5.474\ \mathrm{GeV/c}$$



ROC of J/ψ

Relative Operation Characteristic

$$\bar{p}p \to J/\psi\pi^+\pi^- \to l^+l^-\pi^+\pi^- \quad @\sqrt{s} = 5.474\,\mathrm{GeV/c}$$

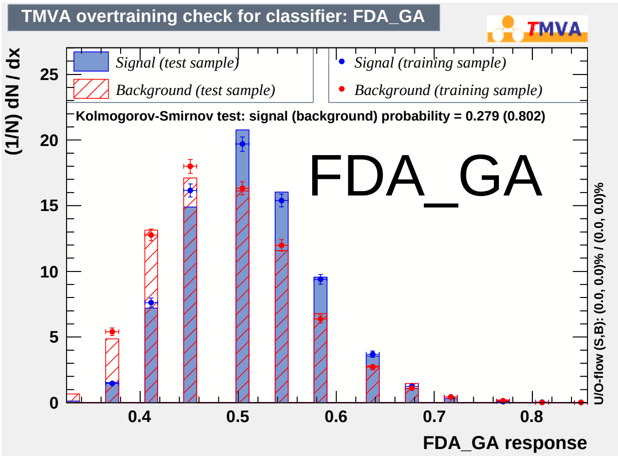


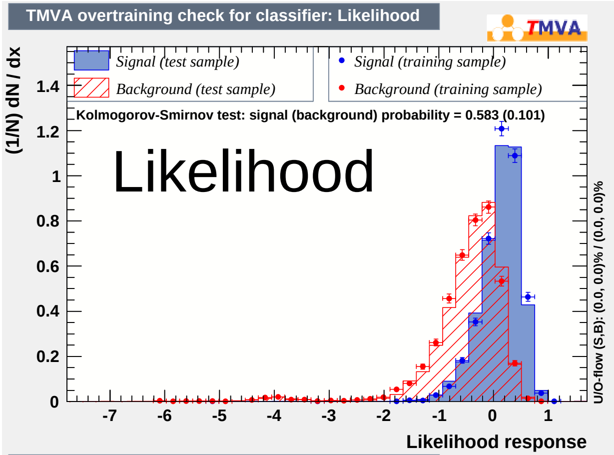
JG|U JOHANNES GUTENBERG UNIVERSITÄT MAINZ

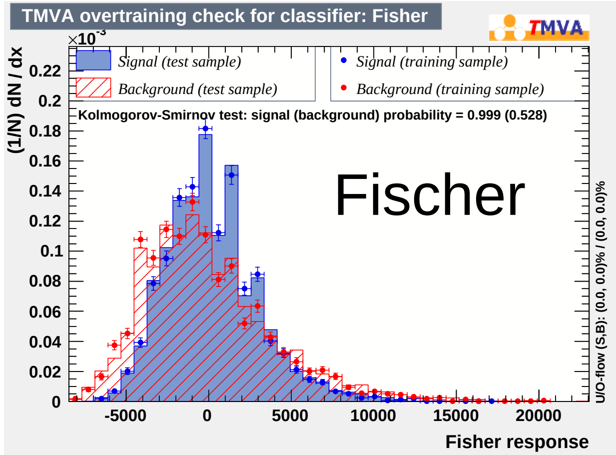
Classification for D$^{\pm}$

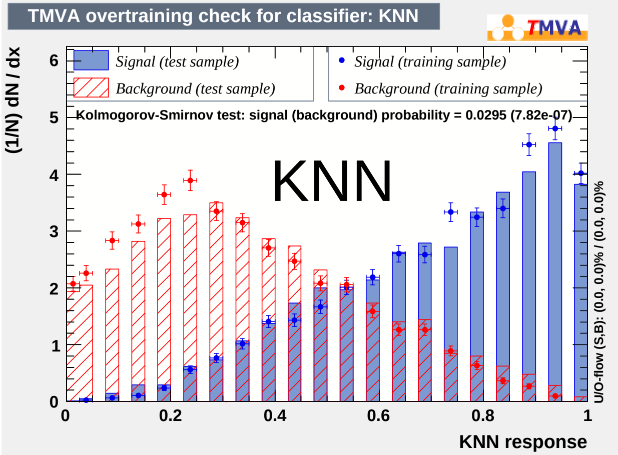
HELMHOLTZ | GEMEINSCHAFT
Helmholtz-Institut Mainz

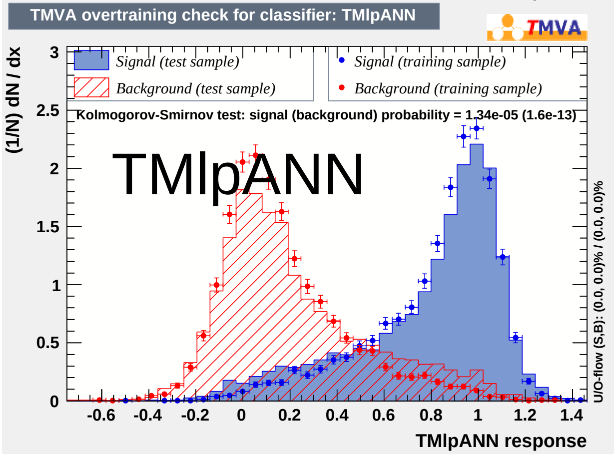
$$\bar{p}p \rightarrow D^{+}D^{-} \rightarrow K^{-}\pi^{+}\pi^{+} + Any \;\; @ \sqrt{s} = 5.474\,\mathrm{GeV/c}$$

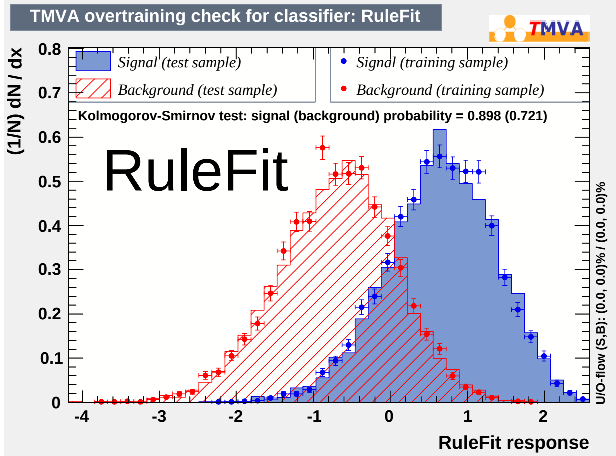


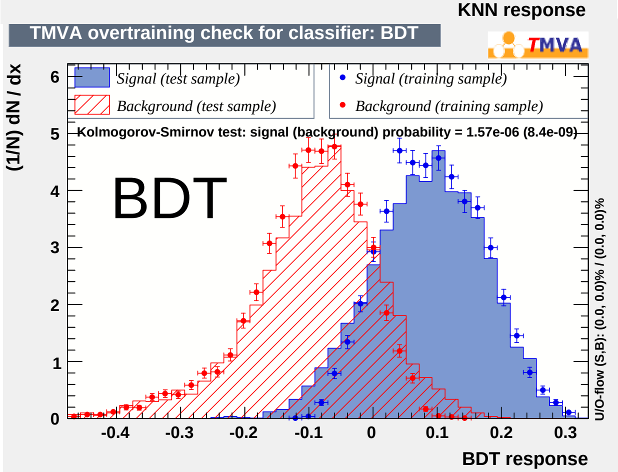


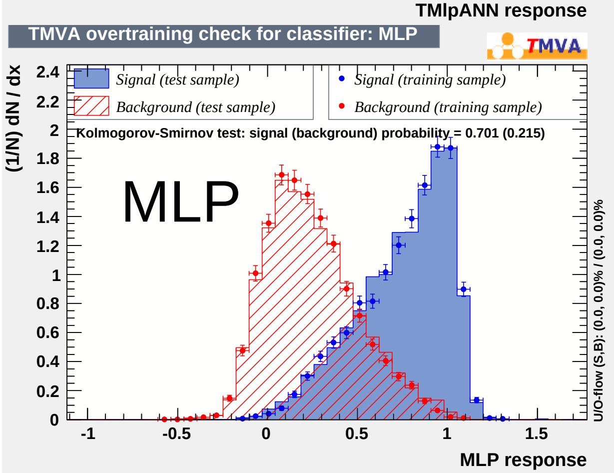


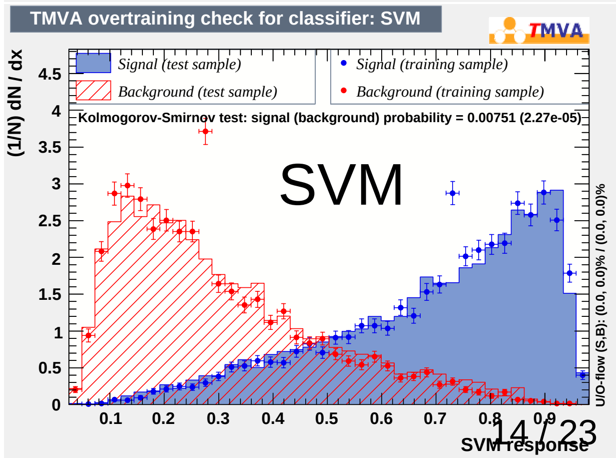








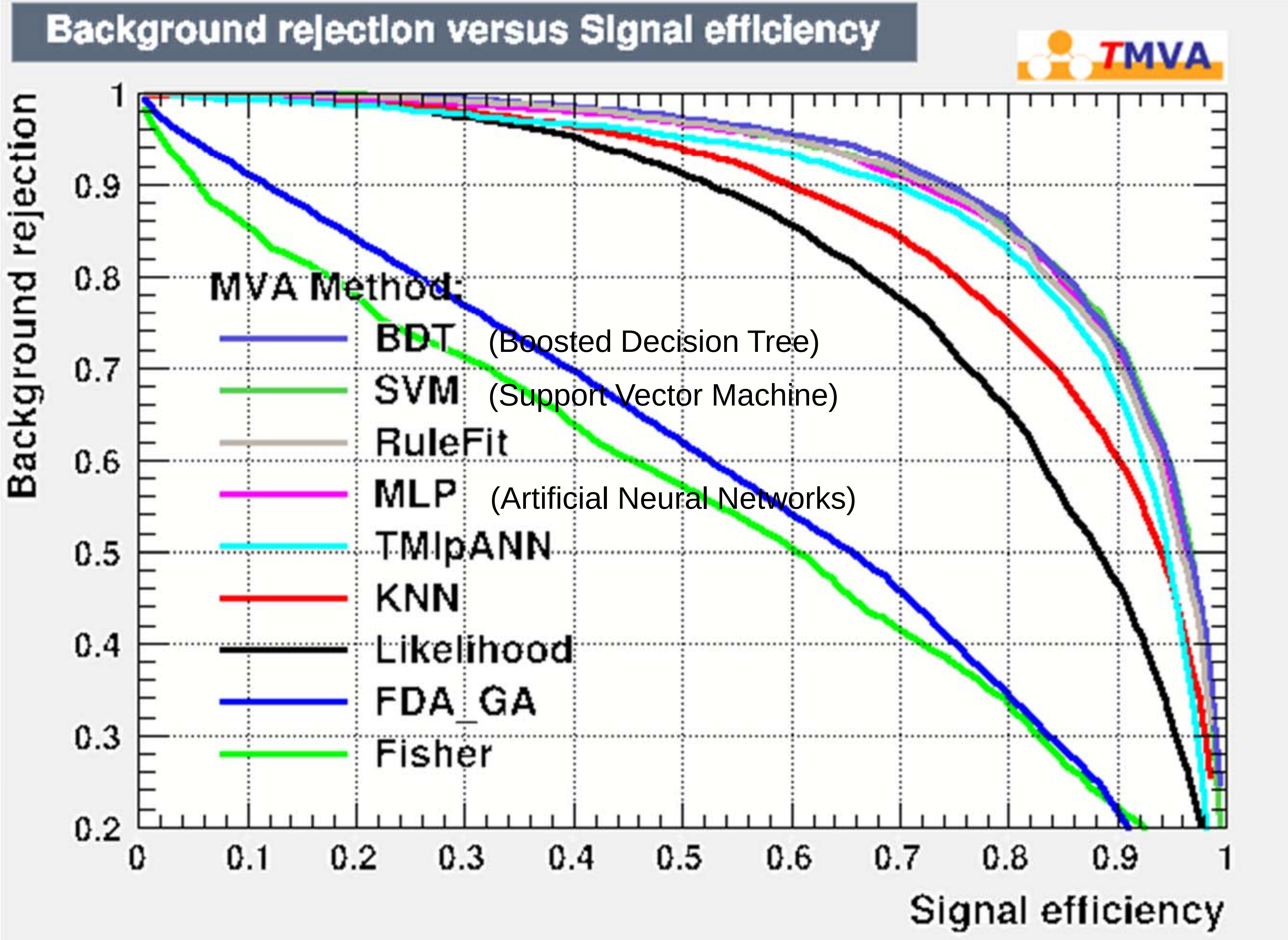




ROC of $D^{\pm}$

Relative Operation Characteristic

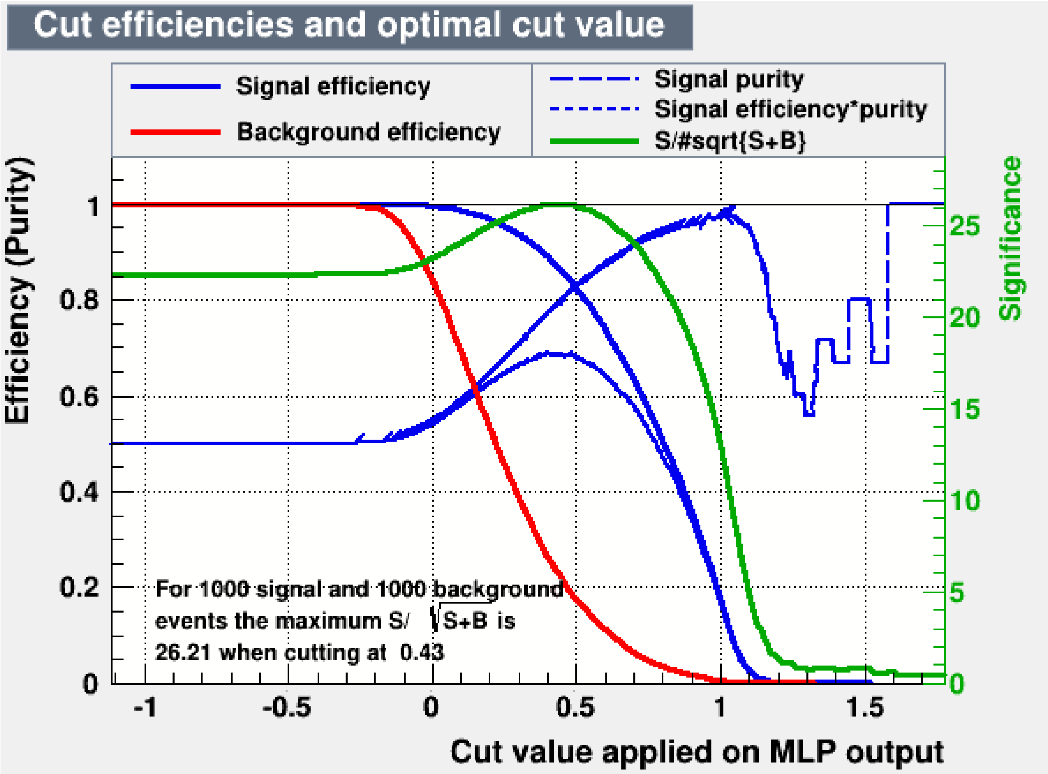
$$\bar{p}p \to D^+D^- \to K^-\pi^+\pi^+ + Any \;\; @ \sqrt{s} = 5.474\,\mathrm{GeV/c}$$



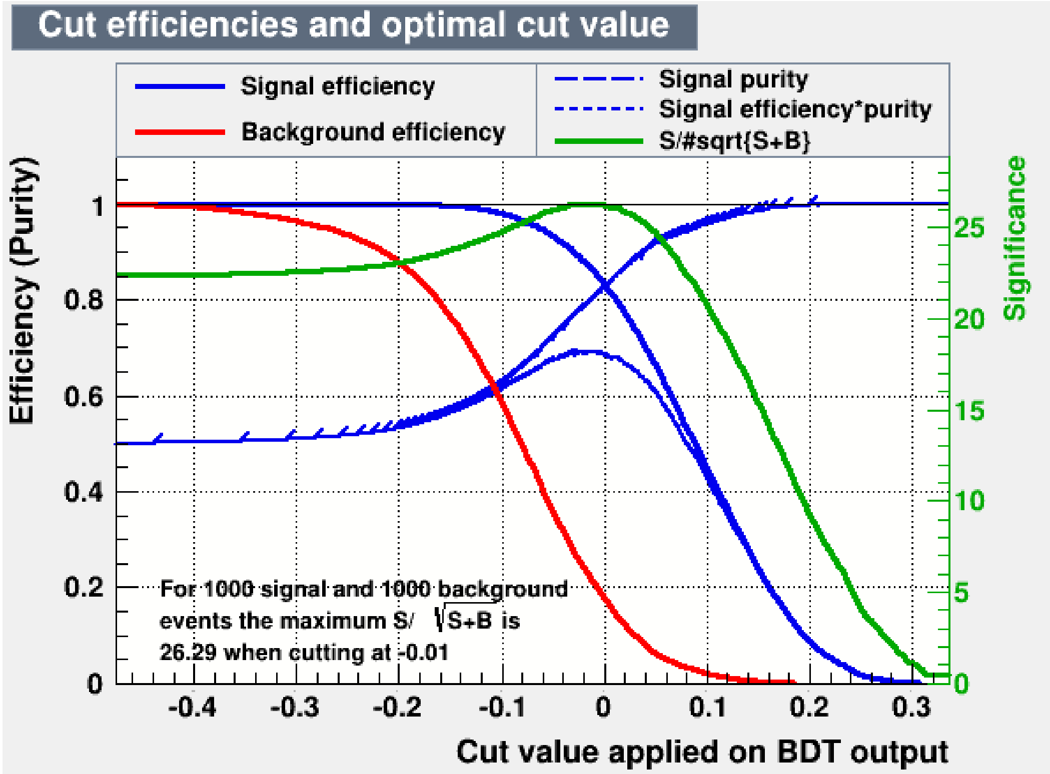
Cut optimization

Optimization of cut value according to significance : $\dfrac{S}{\sqrt{S+B}}$

$$\bar{p}p \rightarrow D^+D^- \rightarrow K^-\pi^+\pi^+ + Any \;\; @ \sqrt{s} = 5.474\,\mathrm{GeV/c}$$



Artifiical Neural Networks (MLP)



Boosted Decision Tree (BDT)

Find optimal cut for different MC data samples / classifiers

Efficiency of TMVA

Determine DPM suppression for simultaneous tagging of all channels

Mode	Beam mom. = 15 GeV/c			
Category(Tagging)	Distr. Optmization	TMVA : BDT[%]	TMVA : MLP[%]	TMVA : SVM[%]
$D^0 \rightarrow K^- \pi^+$	45.1	13.3	15.8	16.1
$D^+ \rightarrow K^- \pi^+ \pi^+$	49.0	25.7	25.9	22.1
$D_s^+ \rightarrow \phi \pi^+$	**68.3**	**32.7**	**27.7**	**22.3**
$pp \rightarrow e^+ e^-$	0.01	1.0	2.4	1.4
$J/\psi \pi^0 \rightarrow e^+ e^- \gamma\gamma$	1.4	0.6	1.8	0.5
$J/\psi \pi^+ \pi^- \rightarrow e^+ e^- \pi^+ \pi^-$	10.8	3.6	5.0	3.1
$J/\psi \pi^+ \pi^- \rightarrow \mu^+ \mu^- \pi^+ \pi^-$	6.8	3.1	6.5	2.9
$\Lambda_C \rightarrow p K^- \pi^+$	**52.8**	**31.1**	**35.3**	**36.8**
$\Lambda \rightarrow p \pi^-$	43.5	7.1	7.7	7.0
$\phi \rightarrow K^+ K^-$	46.8	6.5	8.0	7.5
Total (tagging)	93.0	56.7	56.5	54.7

~ 43 % background reduction (1-εff) by TMVA methods

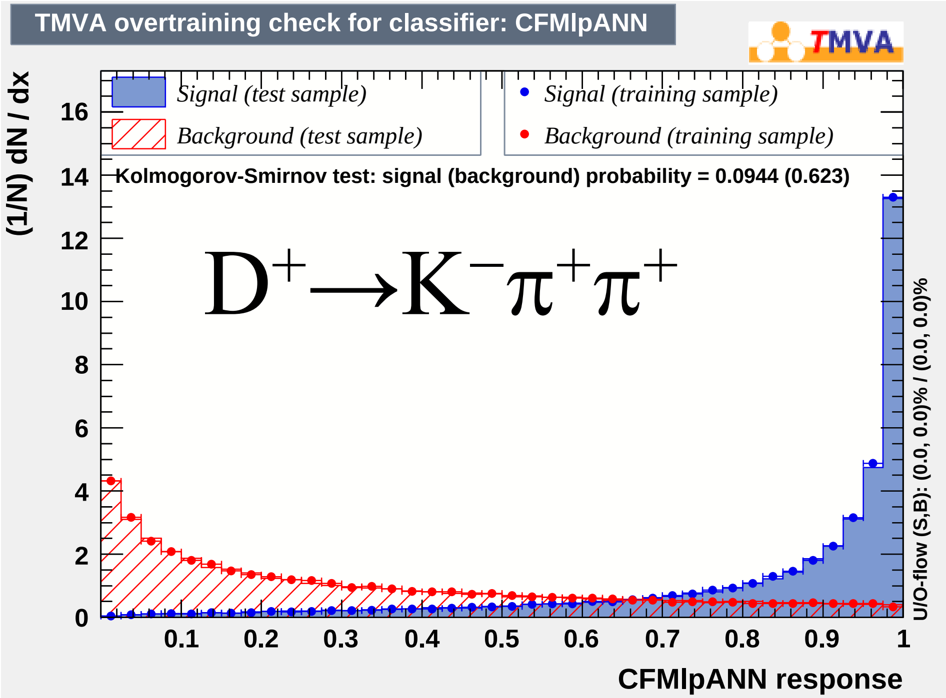
TMVA tuning

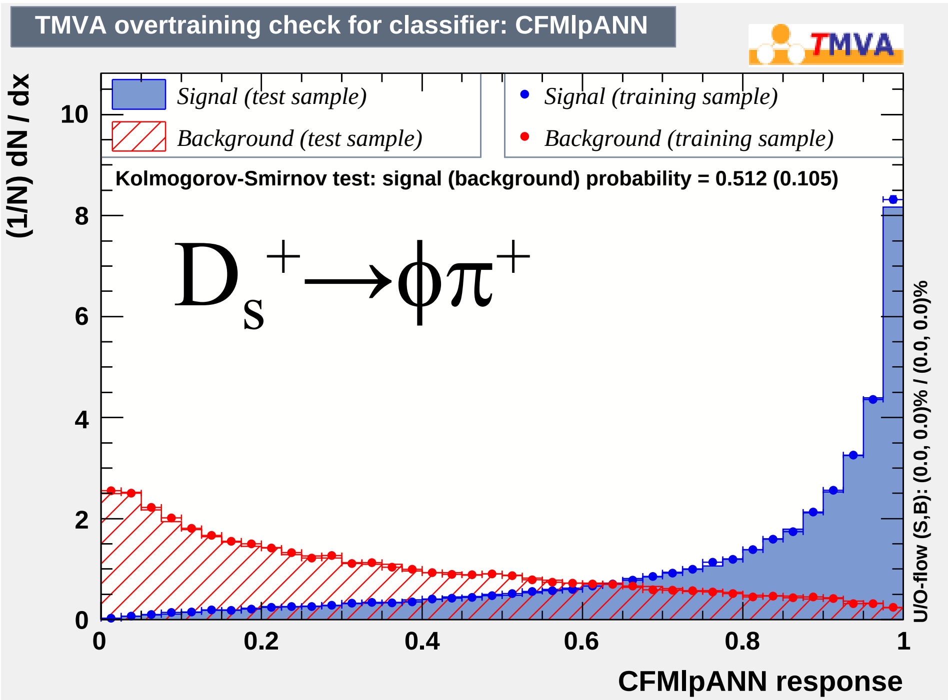
CFMlpANN (Clemont-Ferrant)

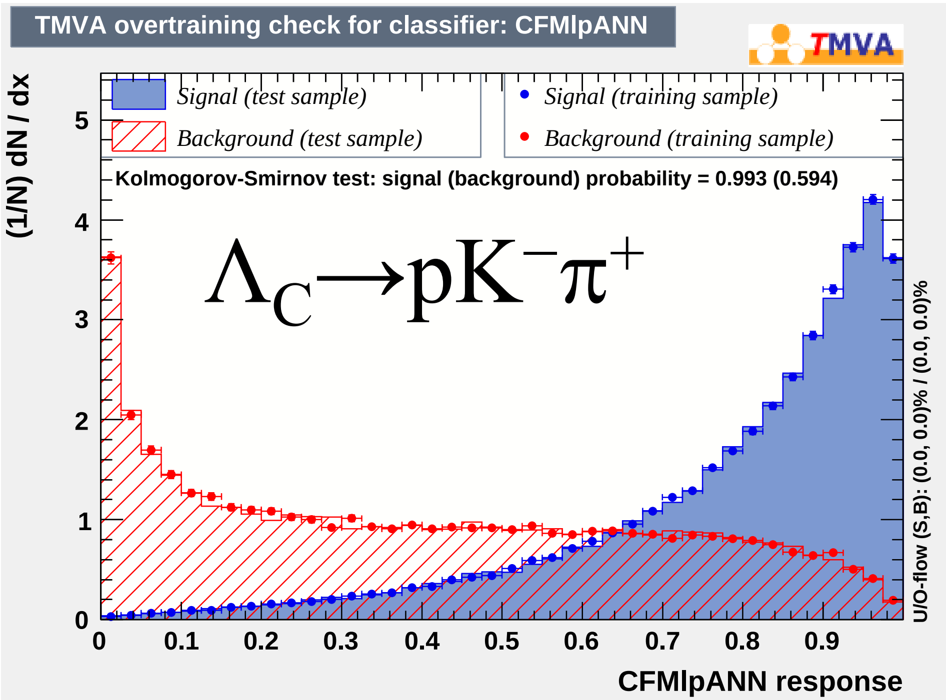
- Artificial Neural Network used in ALEPH and BaBar

- adjust hidden layer structure
  level 2 : N+1,N (level 3 : N+33,N,10)

- good separation for D_s^+







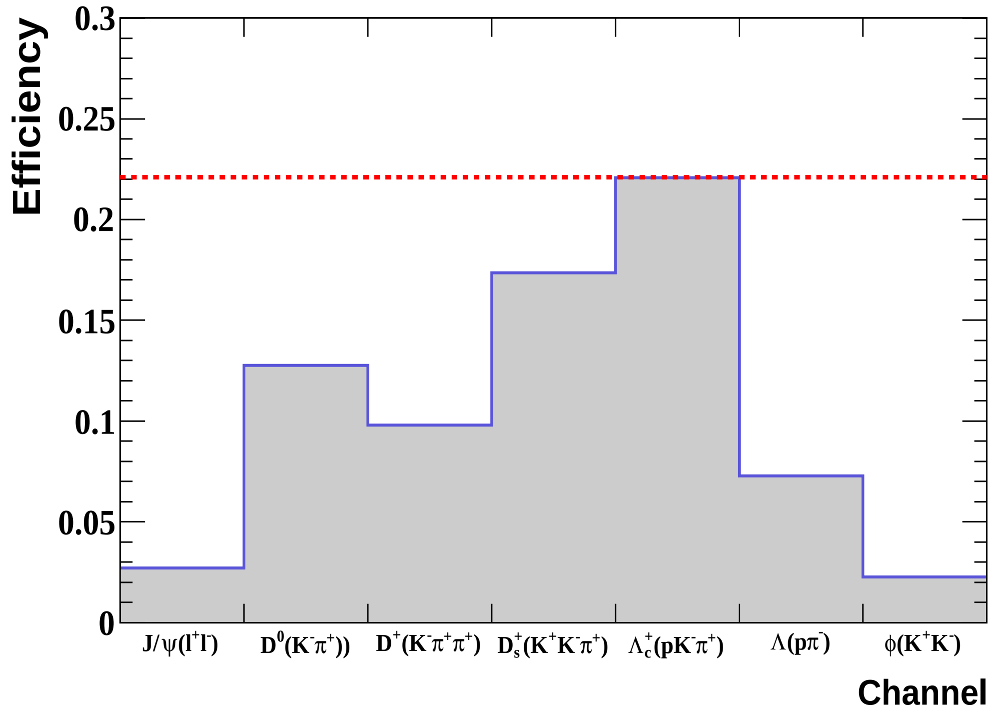
Efficiency of CFMlpANN

CFMlpANN with 33 variables : training 7 different inclusive data samples and DPM

simultaneous tagging

Mode	$\bar{p}$=15 GeV/c
Data	TMVA (level 3) [%]
$D^0 \to K^- \pi^+$	94.8
$D^+ \to K^- \pi^+ \pi^+$	92.8
$D_s^+ \to \phi \pi^+$	89.8
$J/\psi \pi^+ \pi^- \to l^+ l^- \pi^+ \pi^-$	98.5
$\Lambda_C \to p K^- \pi^+$	90.4
$\Lambda \to p \pi^-$	96.2
$\phi \to K^+ K^-$	97.6
DPM	$\boxed{40.3}$

Tagging contribution for DPM background



- 60 % background reduction (1-εff) after tuning of TMVA (20% more gain)

Efficiency of CFMlpANN

CFMlpANN with 33 variables : training 7 different inclusive data samples and DPM

simultaneous tagging

Mode	Beam mom. = 15 GeV/c		
Data	pure TMVA [%]	Mass cut + PID(Loose) [%]	TMVA + Mass cut + PID(Loose) [%]
$D^0 \to K^- \pi^+$	94.8	44.3	43.1
$D^+ \to K^- \pi^+ \pi^+$	92.8	39.7	37.7
$D_s^+ \to \phi \pi^+$	89.8	37.0	34.2
$J/\psi \pi^+ \pi^- \to l^+ l^- \pi^+ \pi^-$	98.5	45.4	45.1
$\Lambda_C \to p K^- \pi^+$	90.4	32.4	30.1
$\Lambda \to p \pi^-$	96.2	-	-
$\phi \to K^+ K^-$	97.6	24.8	24.2
DPM	**40.3**	**20.4**	**9.6**

- 60 % background reduction (1-εff) after tuning of TMVA (20% more gain)
- DPM background contains Λ, cannot use further in efficiency estimation

Efficiency of CFMlpANN

simultaneous tagging

Mode	Beam mom. = 6.569 GeV/c (cms = 3.770 GeV)		
Data	pure TMVA [%]	Mass cut + PID(Loose) [%]	TMVA + Mass cut + PID(Loose) [%]
$D^0 \to K^- \pi^+$	89.1	37.4	35.6
$D^+ \to K^- \pi^+ \pi^+$	88.8	26.2	24.4
$J/\psi \pi^+ \pi^- \to l^+ l^- \pi^+ \pi^-$	98.0	51.2	51.1
$\Lambda \to p \pi^-$	94.1	-	-
$\phi \to K^+ K^-$	95.0	35.5	34.8
DPM	**34.1**	**10.1**	**4.7**

Mode	Beam mom. = 1.7 GeV/c (cms = 2.325 GeV)		
$\Lambda \to p \pi^-$	93.4	-	-
$\phi \to K^+ K^-$	94.8	29.7	29.3
DPM	**30.0**	**3.6**	**0.7**

- 70% background reduction (1-εff) @ low energy mode

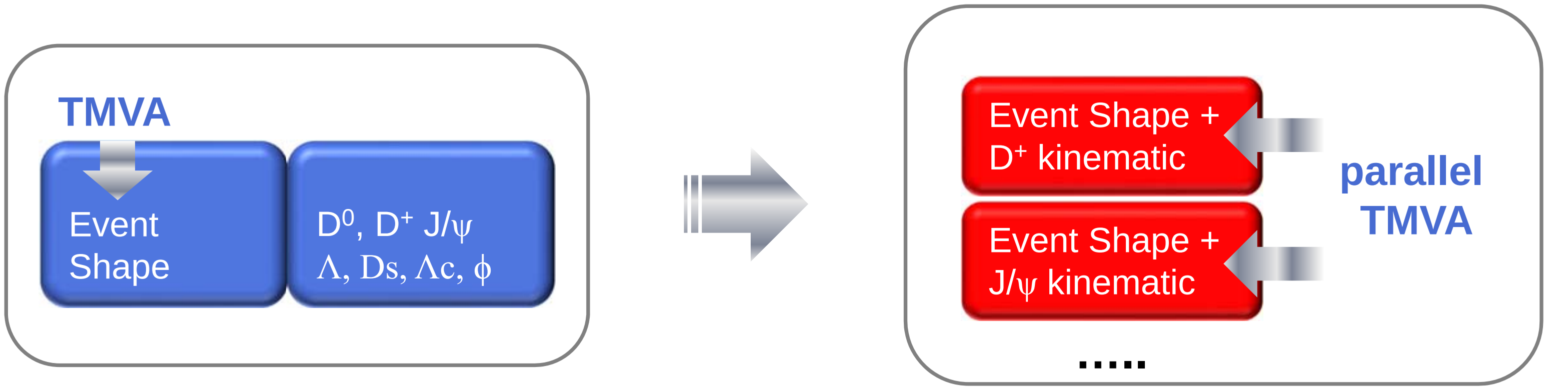
TMVA analysis

Improvement of TMVA for event shape variables

- isolation of large contribution (e.g. $\Lambda_C \rightarrow pK^- \pi^+$) at background suppression
- additional hit information from other detector (e.g. Scitil)

- different combination of hidden layer structure for neural network (e.g. N+10, N, N-1)
- test other classification (e.g. TMlpANN)
- variable transformation by de-correlation in a preprocessing

- combining event shape variables with kinematic variable of resonances



Summary

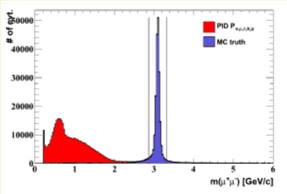
Event flow @ online software trigger

Track/PID candidates

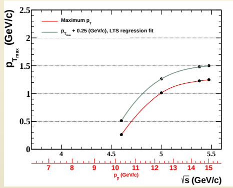
Event Shape cut by MVA

Combinatorial (charged, neutral)

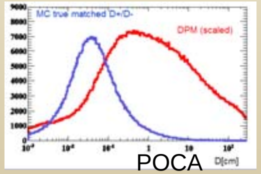
Mass cut (resonances)



Kinematic cut for resonances (e.g. p_T)



POCA / Vertex cut



Trigger decision (multiple line)

20 MHz

start from 40% of background data rate

8 MHz

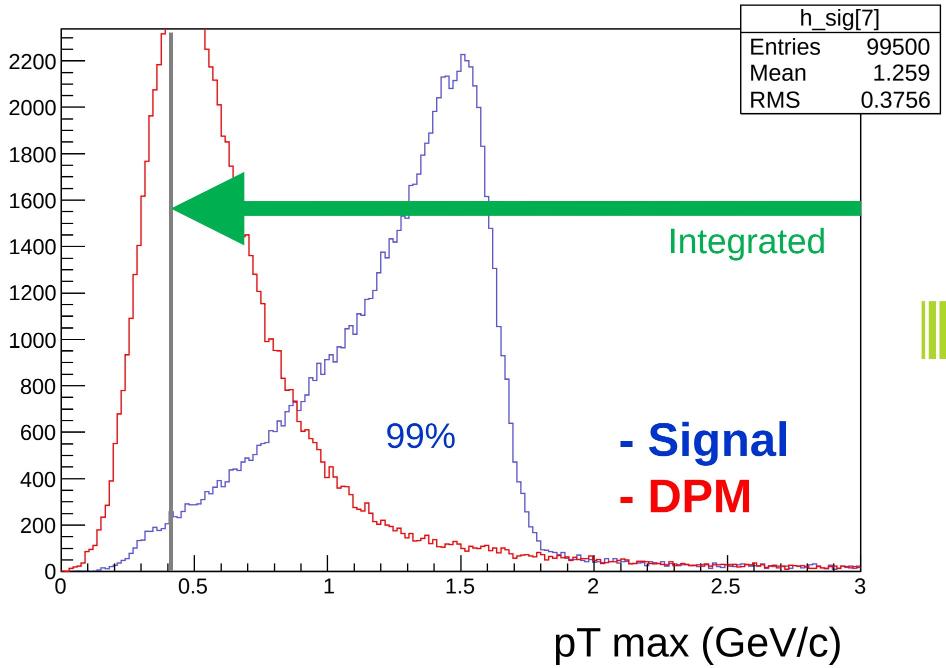
2 MHz

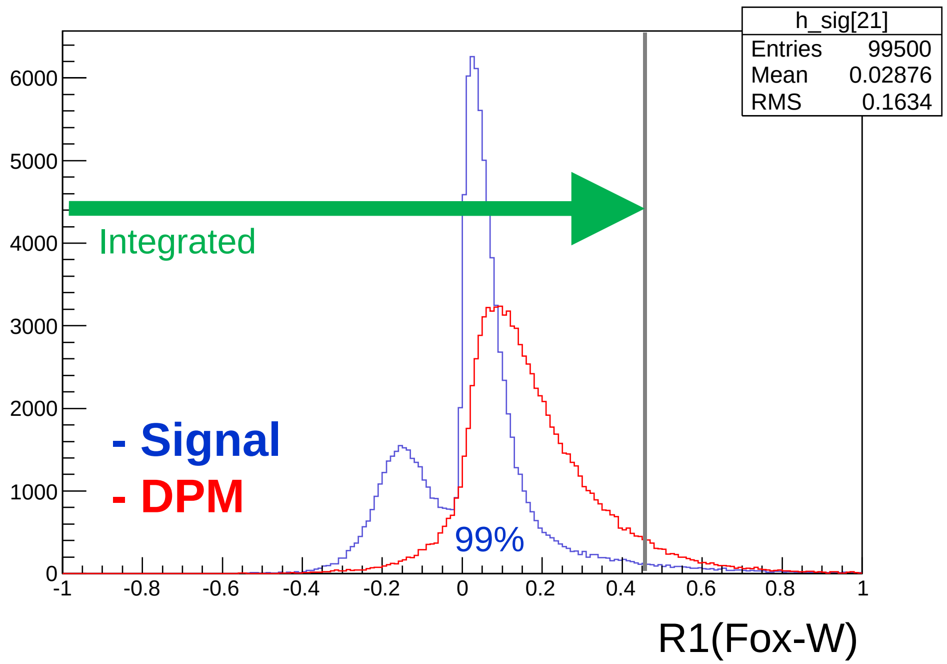
Backup

Distribution optimization approach

For every signal channel and every variable do the following :

- Integrate up to 99% signal area from both sides
- Determine DPM residual and compare with signal residual
- Find cut with $R_{signal} - R_{Background} > 1\%$
- Apply and find next variable

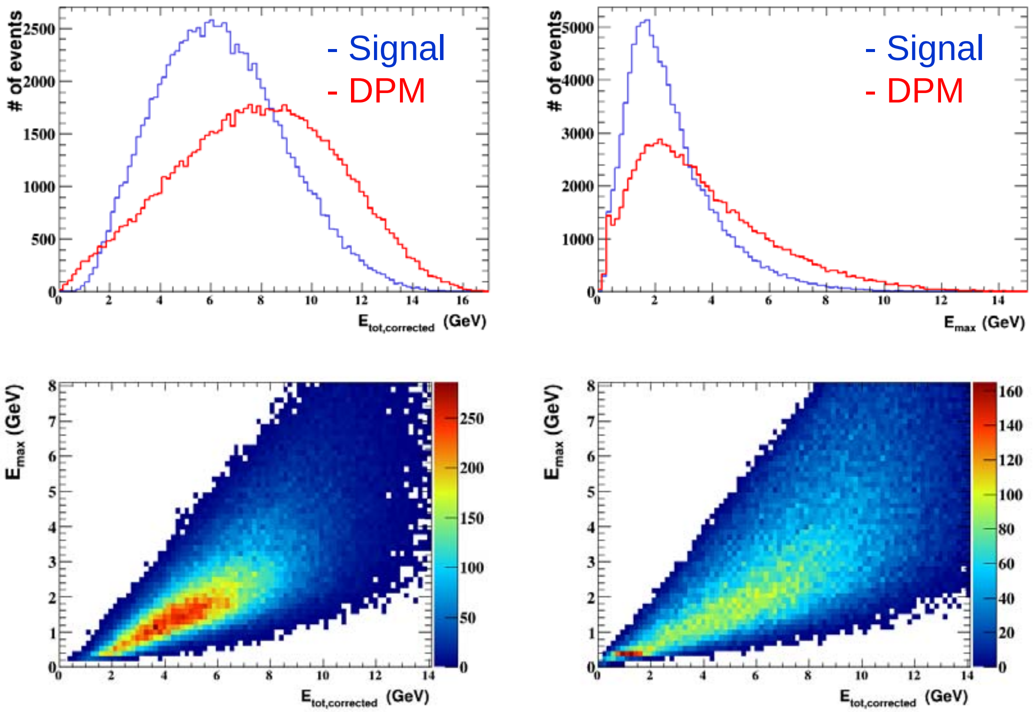




Additional variables

$$\bar{p}p \rightarrow D^+D^- \rightarrow K^-\pi^+\pi^+ + Any \;\; @ \sqrt{s} = 5.474\,\mathrm{GeV/c}$$

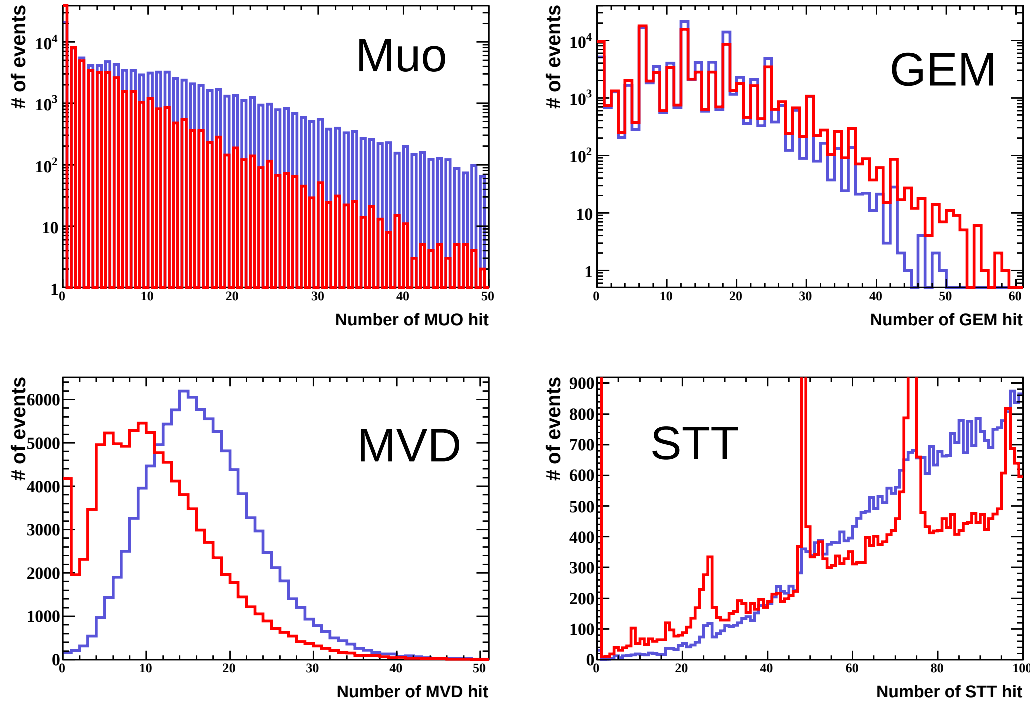
EMC cluster information



Signal Data

DPM background

total number of hit in detector

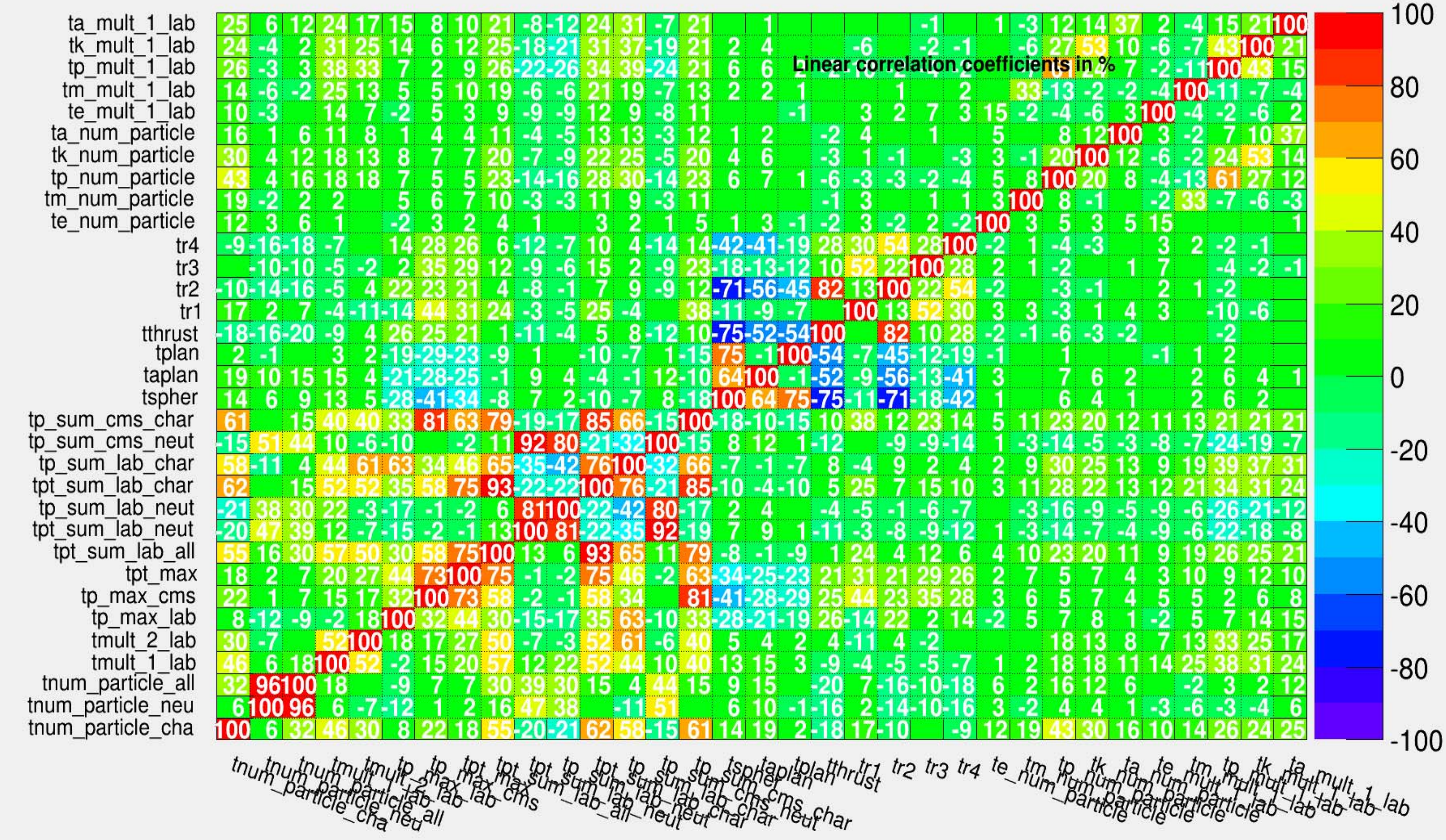


- Signal
- DPM

TMVA tuning

Non-vanishing correlations between input variables would lead to underperforming

33 X 33 variables

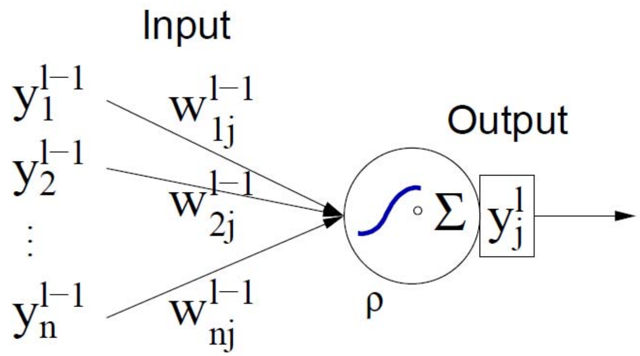


TMVA tuning

Variable ranking in the neural network approach

$$I_i = \overline{x}_i^2 \sum_{j=1}^{n_b} (w_{ij}^{(1)})^2, \qquad i = 1,\ldots..n_{\text{var}},$$

sum of the weights-squared of the connection
between the variable in the input and 1st hidden layer



Determine relevance in the application of TMVA

- search for variable with smallest information loss if removed
- remove variable, calculate information loss again
- start over until no more variable(only improtant var.) left

TMVA tuning

Variable ranking can use to find the importance of variables (for ANN & BDT)