Monte Carlo, fitting, Machine Learning for Tau leptons

Z. Was*, V. Cherepanov, B. Le, E. Richter-Was

*Speaker; Institute of Nuclear Physics, Polish Academy of Sciences Krakow

- **(1)** The τ lepton decays: fascinating laboratory for intermediate energy QCD

- **(2)** In itself, developing models can be very tempting.

NOTE: precision of experimental data is substantially better, than theory predictions.

- **(3)** How to optimize work of in-homogeneous community.

- **(4)** As in the past, I will use TAUOLA, its associated projects and updates as examples.

Few slides prepared by **V. Cherepanov** will serve as two of such examples.

- **(5)** I want to address the question what are the requirements for MC and phenomenology tools, experimental data handling. Especially now in times of Machine Learning $\rightarrow$ multidimensional signatures.

3 talks $\rightarrow$ 15 mins $\rightarrow$ recipe for a disaster?

1. Higgs CP measurement in $H \rightarrow \tau\tau; \tau \rightarrow 2(3)\pi$ channels. My first example of ML usefulness.
2. Tools properties $\rightarrow$ From optimal to expert variables $\rightarrow$ Event space metric
3. Examples of new applications:
   (a) Cracow
   (b) Vladimir
   (c) Literature
4. Comments on rapid evolution of ML algorithms and applications and **what does it mean for phenomenology of τ decays?**
5. Some news on `Tauola Photos TauSpinner` nonetheless.
6. Conclusions and outlook.

The Higgs boson's parity is imprinted in M.E.

- H/A parity information can be extracted from the correlations between τ^+ and τ^- spin components which are further reflected in correlations between the τ decay products in the plane transverse to the $\tau^+\tau^-$ axes.
- The decay probability

$$\Gamma(H/A \to \tau^+\tau^-) \sim 1 - s_{\parallel}^{\tau^+} s_{\parallel}^{\tau^-} \pm s_{\perp}^{\tau^+} s_{\perp}^{\tau^-}$$

is sensitive to the $\tau^{\pm}$ polarization vectors s^{τ^-} and s^{τ^+} (defined in their respective rest frames). The symbols $\parallel$,$\perp$ denote components parallel/transverse to the Higgs boson momentum as seen from the respective $\tau^{\pm}$ rest frames.

- This idea an its practical refinements are universal: 'Higgs spin' is blind on Higgs origin. But it is not true for the background DY processes .

Tau-pair production M.E.

Phenomenology Of Mixed Parity: also from M.E.

- Higgs boson Yukawa coupling expressed with the help of the scalar–pseudo-scalar mixing angle ϕ

$$\bar{\tau} N (\cos\phi + i \sin\phi\gamma_5)\tau$$

- *Decay probability for the mixed scalar–pseudo-scalar case*

$$\Gamma(h_{mix} \to \tau^+\tau^-) \sim 1 - s_{\parallel}^{\tau^+} s_{\parallel}^{\tau^-} + s_{\perp}^{\tau^+} R(2\phi)\, s_{\perp}^{\tau^-}$$

- $R(2\phi) -$ *operator for the rotation by angle* 2ϕ *around the* $\parallel$ *direction.*

$$R_{11} = R_{22} = \cos 2\phi \qquad R_{12} = -R_{21} = \sin 2\phi$$

- *Pure scalar case is reproduced for* $\phi = 0$.
- *For* $\phi = \pi/2$ *we reproduce the pure pseudo-scalar case.*

Tau-decay M.E. and attempted observable.

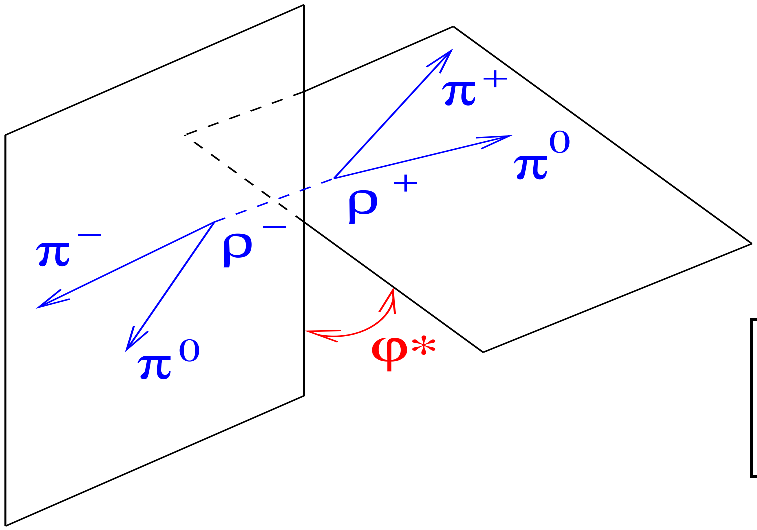
Transverse spin correlations through τ decays

- *Case of $\tau \to \rho\nu_\tau$ decay,* $\mathcal{BR}(\tau \to \rho\nu_\tau) = 25\%$*, also M.E. expressed.*
- *Polarimeter vector h^i is (where q for $\pi^\pm - \pi^0$ and N for ν_τ four momenta.*

$$h^i = \mathcal{N}\Big(2(q \cdot N)q^i - q^2 N^i\Big)$$

$$q \cdot N = (E_{\pi^\pm} - E_{\pi^0})m_\tau$$

- *Acoplanarity of ρ^+ and ρ^- decay prod. (in $\rho^+\rho^-$ r.f.) and events separation.*



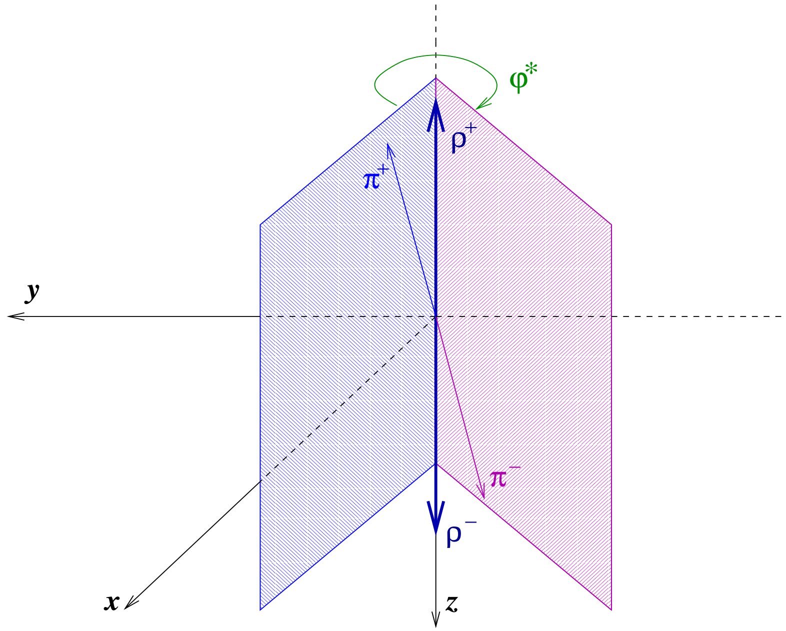
$y_1 y_2 > 0\,;\qquad y_1 y_2 < 0$ *(in $\tau^\pm$ r.f.'s)*

$$y_1 = \frac{E_{\pi^+} - E_{\pi^0}}{E_{\pi^+} + E_{\pi^0}}\,;\qquad y_2 = \frac{E_{\pi^-} - E_{\pi^0}}{E_{\pi^-} + E_{\pi^0}}.$$

Observable of visible products in $H \to \tau^+\tau^- \to \pi^+\pi^0\ \pi^-\pi$

Optimal Observable Mixed Scalar–Pseudoscalar Case

- *For mixing angle ϕ, transverse component of τ^+ spin polarization vector is correlated with the one of τ^- rotated by angle 2ϕ.*
- *Acoplanarity $0 < \varphi^* < 2\pi$ is of physical interest, not just $\arccos \mathbf{n}_- \cdot \mathbf{n}_+$.*
- *Distinguish between the two cases $0 < \varphi^* < \pi$ and $2\pi - \varphi^*$*
- *If no separation made the parity effect would wash itself out.*



Normal to planes: $\mathbf{n}_\pm = \mathbf{p}_{\pi^\pm} \times \mathbf{p}_{\pi^0}$

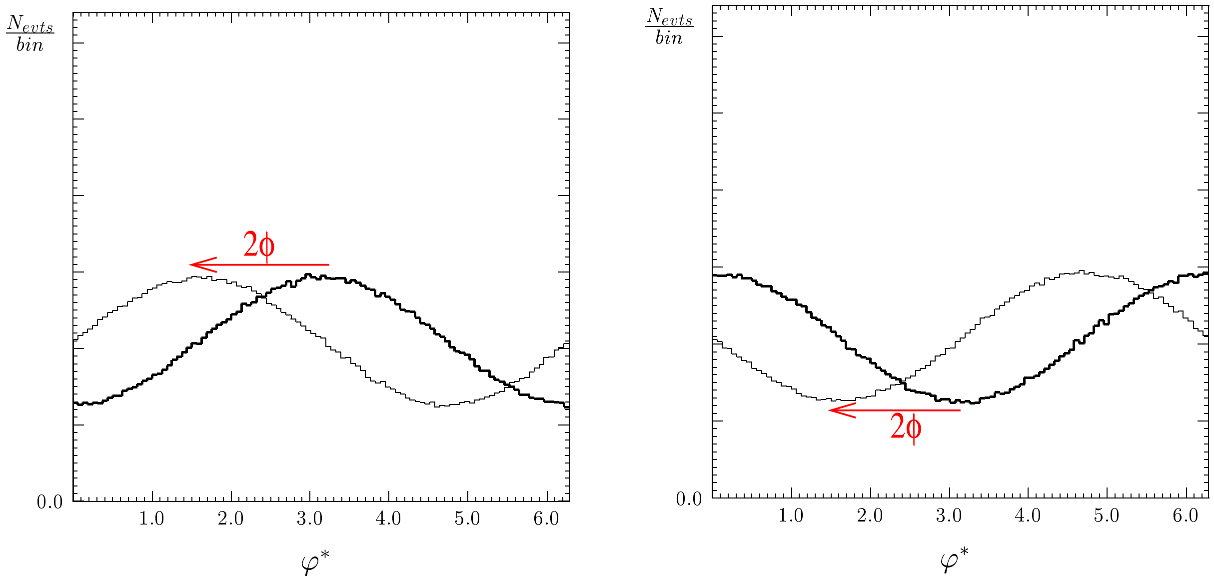
Find the sign of $\mathbf{p}_{\pi^-} \cdot \mathbf{n}_+$

Negative $0 < \varphi^* < \pi$

Otherwise $2\pi - \varphi^*$

Observable of visible products in $H \to \tau^+\tau^- \to \pi^+\pi^0\ \pi^-\pi$

Old attempts, at the end 1-dim plot 'easy' to understand



- *Only events where the signs of y1 and y2 are the same whether calculated using the method without or with the help of the τ impact parameter.*
- *Tesla-like set-up SIMDET used, K. Desch, A. Imhof, ZW,, M. Worek, Phys.Lett. B579 (2004) 157.*
- *The thick line corresponds to a scalar Higgs boson, the thin line to a mixed one.*

Precision on $\phi \sim 6^\circ$, for $1ab^{-1}$ and 350 GeV CMS.

Q: what is ML?

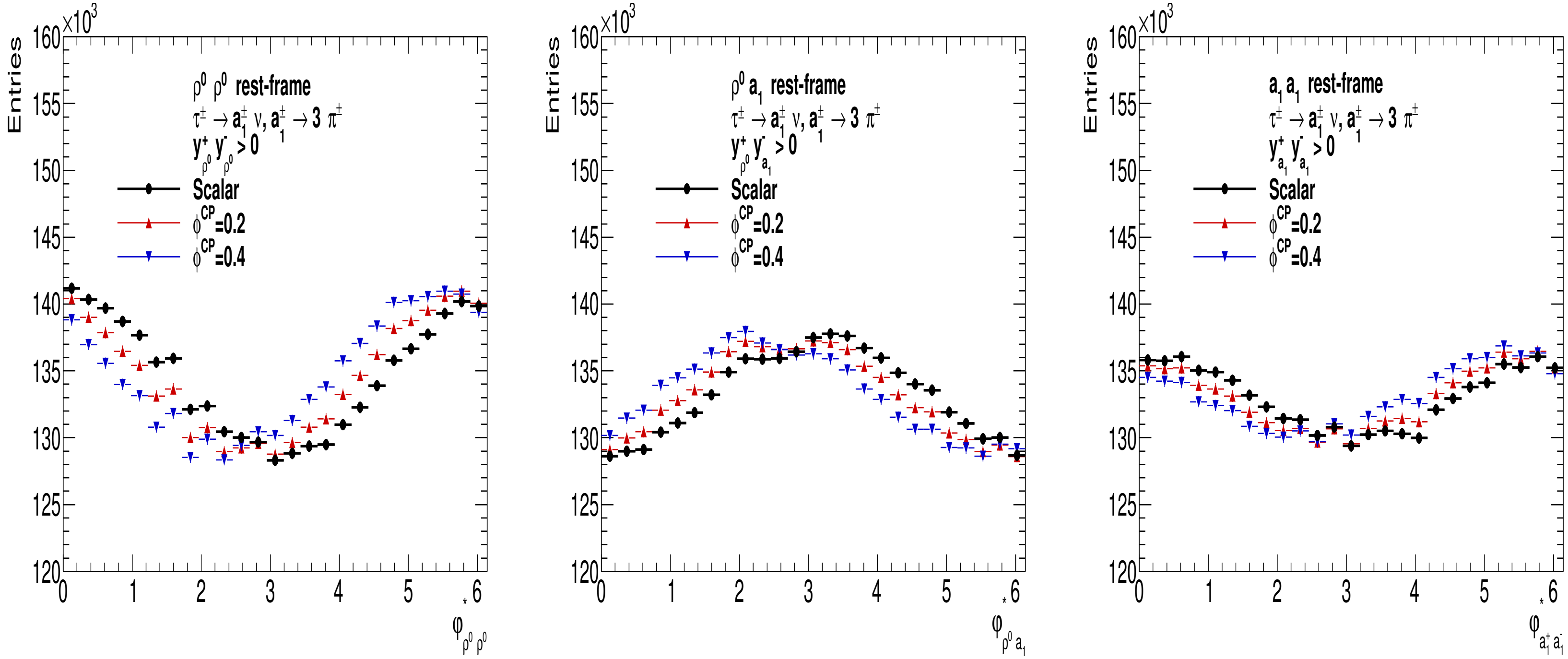
A: I will not cover this topic as well, see some references later.

From my own experience and frustrations:

- $\mathcal{BR}(\tau \to \rho\nu_\tau) = 25\%$, that mean 6$\%$ of $H \to \tau\tau$. Why not use other decay modes? They all have (in principle) the same sensitivity to spin: J. H. Kuhn, Phys. Rev. D52 (1995) 3128, but in practice ν_τ is not observable and:
- from the π^-, π^0 we can define one plane for acoplanarity,
- from the π^-, π^-, π^+ we can define four such planes.
- Each plane bring its own y_i variable to avoid cancellations due to properties of τ decay ME.

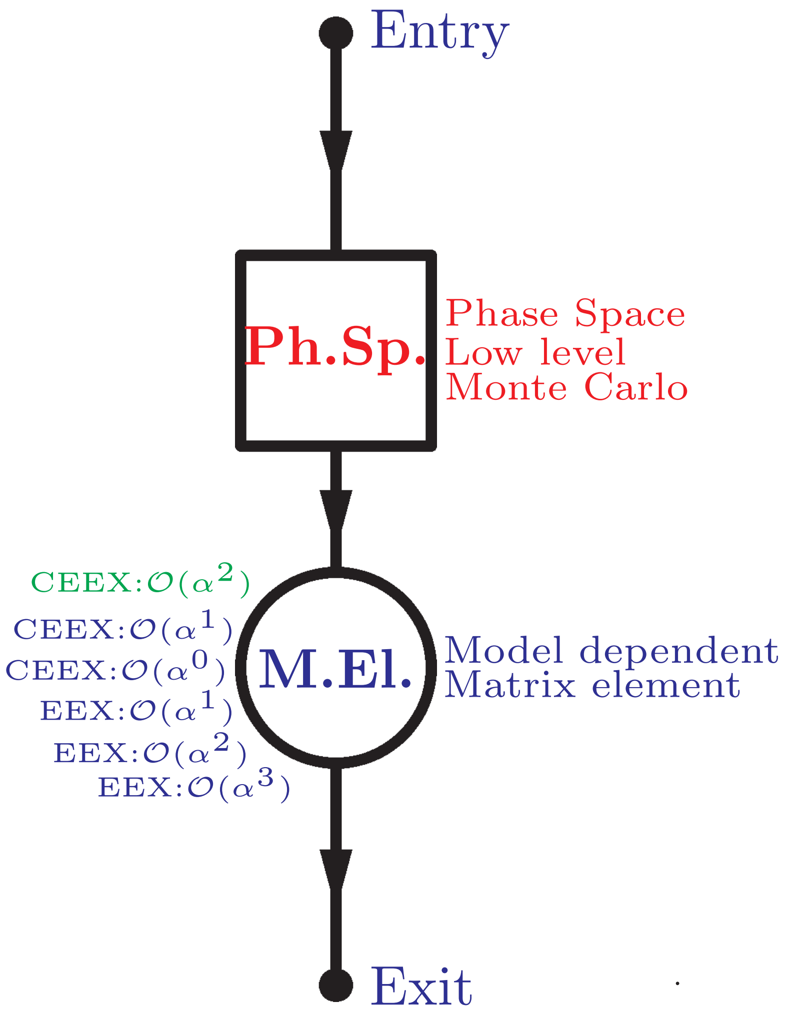
My frustrations: many distributions of little content.

Acoplanarity angles of oriented half decay planes: $\varphi^*_{\rho^0\rho^0}$ (left), $\varphi^*_{a_1\rho^0}$ (middle) and $\varphi^*_{a_1a_1}$ (right), for events grouped by the sign of $y^+_{\rho^0}y^-_{\rho^0}$, $y^+_{a_1}y^-_{\rho^0}$ and $y^+_{a_1}y^-_{a_1}$ respectively. Three CP mixing angles ϕ^{CP} = 0.0 (scalar), 0.2 and 0.4. Note scale, effect on individual plot is so much smaller now. But up to **16 plots like that** have to be measured, correlations understood. Physics model depends on 1 parameter only ϕ^{CP} mixing scalar pseudo-scalar angle, which brings linear shift. I remained frustrated for 15 years, how to digest...



ML: my first solution.

Textbook principle “matrix element × full phase space” useful



- Phase-space Monte Carlo module producing “raw events”.
- Library of models for provides input for “model weight”
- **The scalar from pseudo-scalar distinguished by M.E. weight attributed to each event**
- Ratios define probability that event could be scalar or pseudoscalar Higgs.
- Convenient for ML training sample.

ML: my first solution.

- I can not present ML technology.
- I will flash some results only.
- Essentially probabilities that Network will identify event to be scalar, when it was scalar.
- 0.5 means random choce. 1.0 would mean certainty.
- Anything in-between was something useful.
- To get classification I had to:
- boost events to rest frame of all visible objects combined,
- rotate all to set τ^+ primary decay resonance along z axis.

Features/variables	$\rho^\pm - \rho^\mp$ $\rho^\pm \to \pi^0\,\pi^\pm$	$a_1^\pm - \rho^\mp$ $a_1^\pm \to \rho^0\pi^\mp,\ \rho^0 \to \pi^+\pi^-$ $\rho^\mp \to \pi^0\,\pi^\mp$	$a_1^\pm - a_1^\mp$ $a_1^\pm \to \rho^0\pi^\pm,$ $\rho^0 \to \pi^+\pi^-$
True classification	0.782	0.782	0.782
$\varphi^*_{i,k}$	0.500	0.500	0.500
$\varphi^*_{i,k}$ and y_i, y_k	0.624	0.569	0.536
4-vectors	0.638	0.590	0.557
$\varphi^*_{i,k}$, 4-vectors	0.638	0.594	0.573
$\varphi^*_{i,k}$, y_i, y_k and m_i^2, m_k^2	0.626	0.578	0.548
$\varphi^*_{i,k}$, y_i, y_k, m_i^2, m_k^2 and 4-vectors	0.639	0.596	0.573

Table 1: Average probability p_i that a model predicts correctly event x_i to be of a type A (scalar), with training being performed for separation between type A and B (pseudo-scalar).

$\varphi^*_{i,k}$ and y_i: expert variables In rest frame of all visible, aligned along z. Essential for measure of event distance.

With detector smearing: Phys.Rev. D96 (2017) 073002

Features				Ideal ± (stat)	Smeared ± (stat) ± (syst)
ϕ^*	4-vec	y_i	m_i		
				$a_1 - \rho$ Decays	
✓	✓	✓	✓	0.6035 ± 0.0005	$0.5923 \pm 0.0005 \pm 0.0002$
✓	✓	✓	-	0.5965 ± 0.0005	$0.5889 \pm 0.0005 \pm 0.0002$
✓	✓	-	✓	0.6037 ± 0.0005	$0.5933 \pm 0.0005 \pm 0.0003$
-	✓	-	-	0.5971 ± 0.0005	$0.5892 \pm 0.0005 \pm 0.0002$
✓	✓	-	-	0.5971 ± 0.0005	$0.5893 \pm 0.0005 \pm 0.0002$
✓	-	✓	✓	0.5927 ± 0.0005	$0.5847 \pm 0.0005 \pm 0.0002$
✓	-	✓	-	0.5819 ± 0.0005	$0.5746 \pm 0.0005 \pm 0.0002$
				$a_1 - a_1$ Decays	
✓	✓	✓	✓	0.5669 ± 0.0004	$0.5657 \pm 0.0004 \pm 0.0001$
✓	✓	✓	-	0.5596 ± 0.0004	$0.5599 \pm 0.0004 \pm 0.0001$
✓	✓	-	✓	0.5677 ± 0.0004	$0.5661 \pm 0.0004 \pm 0.0001$
-	✓	-	-	0.5654 ± 0.0004	$0.5641 \pm 0.0004 \pm 0.0001$
✓	✓	-	-	0.5623 ± 0.0004	$0.5615 \pm 0.0004 \pm 0.0001$
✓	-	✓	✓	0.5469 ± 0.0004	$0.5466 \pm 0.0004 \pm 0.0001$
✓	-	✓	-	0.5369 ± 0.0004	$0.5374 \pm 0.0004 \pm 0.0001$

Table 2: AUC for NN to separate scalar and pseudo-scalar hypotheses. Inputs with a ✓used. Results in column "Ideal" - from NNs trained/used with particle-level simulation, in column "Smeared" - from NNs trained/used with smearing. NN trained on smeared samples, for used on exact samples give similar results as "Ideal".

$H \to \tau\tau; \tau \to 2(3)\pi$ channels and ML

1. We have played with input, and we have observed:
2. Our precious expert variables were not necessary from some point
3. But seemingly trivial overall boosts and rotations were indispensable
4. Only some time later we understood why: network required help to separate longitudinal from transverse degrees of freedom.
5. There was not problem that some variables were then systematically big or small. Such properties were easy for NN to understand. Re scaling was in the system.
6. It does not need to be always like that. It will be application domain dependent.
7. My training case was in a sense easy, we could get help from ME. calculation and adjust variable set accordingly.
8. Only some time after finishing work and after some studies of literature I understood this ML contexts.

Already from J. H. Kuhn, Phys. Rev. D52 (1995) 3128 is was clear that having all information on τ decays should give ideal 0.782 classification from all τ decay modes.

Missing neutrinos was a challenge, the following approaches can be listed:

- K. Desch, Z. Was and M. Worek, “Measuring the Higgs boson parity at a linear collider using the tau impact parameter and tau —> rho nu decay,” Eur. Phys. J. C **29** (2003) 491
- A. Rouge, “CP violation in a light Higgs boson decay from tau-spin correlations at a linear collider,” Phys. Lett. B **619**, 43 (2005)
- S. Berge, W. Bernreuther and S. Kirchner, “Prospects of constraining the Higgs boson’s CP nature in the tau decay channel at the LHC,” Phys. Rev. D **92**, 096012 (2015)

Is there anything that ML or new efforts can bring?

Our struggling with Impact parameter 15 years later

- A key for the improvement is to reconstruct ν_τ momenta: 6 variables.
- We have 3 good constraints and two less precise:
  - $m_{\tau^\pm}$
  - m_H
  - $p_{T\ mis.}^{x,y}$
- These constraints are partly correlated.
- We are missing two angles, which need to be obtained from impact parameter measurement.
- On the other hand, attempt to obtain optimal variable may be complicated by all kind of smearing. Kinematical constraints may be non-resolvable.
- ML may be hopefully useful.

Let me show only one technical step for the ML 'expert variable' input construction, or for element in preparation of optimal variable.

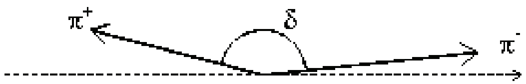
To get neutrinos as good as possible, we need to use approximation.

I can not continue with that. Also, it is still at work: *K. Lasocha et al. Machine learning Classification: case of Higgs boson CP in* $H \to \tau\tau$ *decay at LHC, October 2018.*

I better show , some slides of alternative approach, prepared by Vladimir Cherepanov.

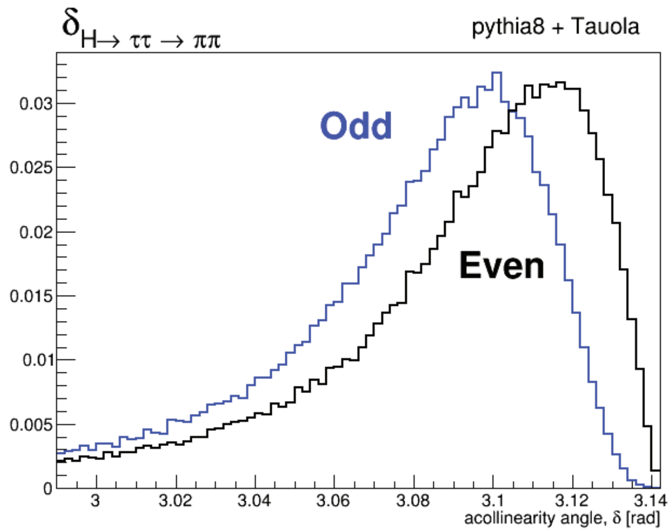
Applications of TauSpinner: Transverse spin correlation in H→ ττ at LHC

- For 1-prong decay acollinearity in Higgs r.f. can be used as analyzer:



M. Kramer, et.al., Z.Phys. C64 (1994) 21-30

Hardly possible at LHC

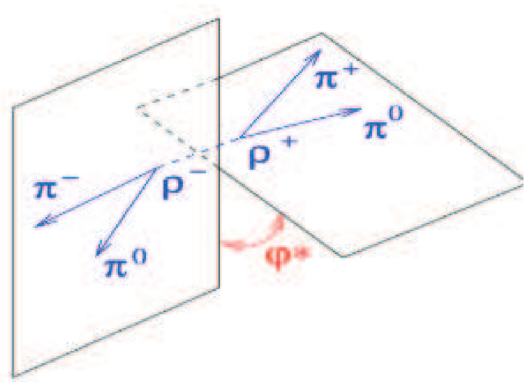


- In several papers it was proposed to build accomplanariy angle using the measured τ decay products (as shown above)

Z. Was, et.al., Phys.Rev. D94 (2016) no.9, 093001

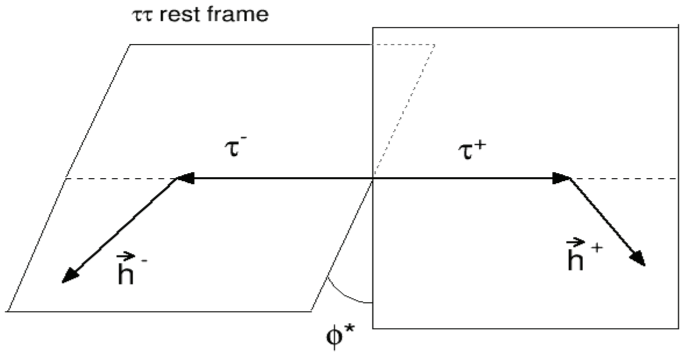
S. Berge, et.al., Nucl.Part.Phys.Proc. 273-275 (2016) 841-845

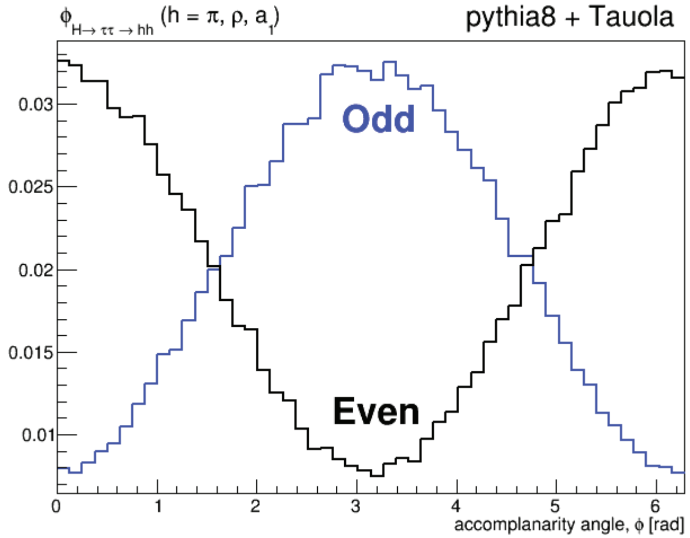
Feasible at LHC but probably not the most optimal approach



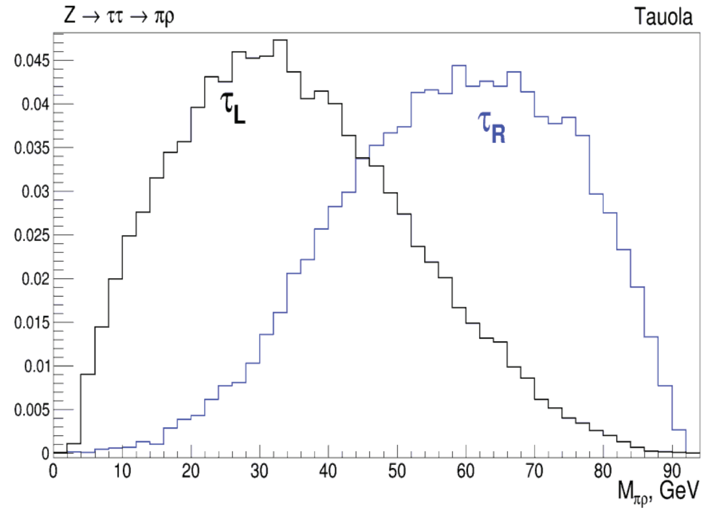
Applications of TauSpinner: Transverse spin correlation in H→ ττ at LHC (Full kinematic analysis)

- Alternatively one may try to estimate the invisible part of the polarimetric vectors of both τ leptons
- The estimation of the Higgs r.f. does not need to be excellent. We need at least some information on neutrinos momenta
- Accomplanarity observable can be build using direction of τ's in H r.f. and polarimetric vectors
- Ideally carries the full analyzing power and irrespective of the τ decay channel





Applications of TauSpinner: Longitudinal τ polarization in Z→ ττ at LHC



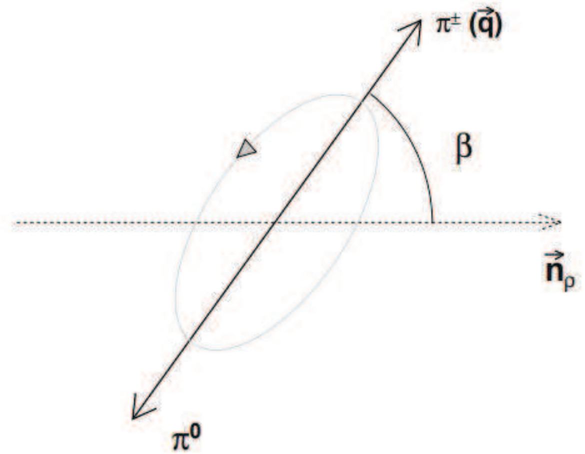
Extensive measurements of tau polarization have been performed at LEP.

At LHC one may apply the visible analysis:

- Acollinearity and energy-energy correlation between decay products of both τ's

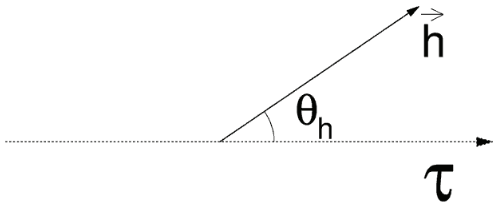
  R. Alemany, et.al., Nucl.Phys. B379 (1992) 3-23

- Angles between measured decay products in τ→ρν and τ→a_1ν decays, for example angle β in τ→ρν decay.



Applications of TauSpinner: Longitudinal τ polarization in Z→ ττ at LHC (Full kinematic analysis)

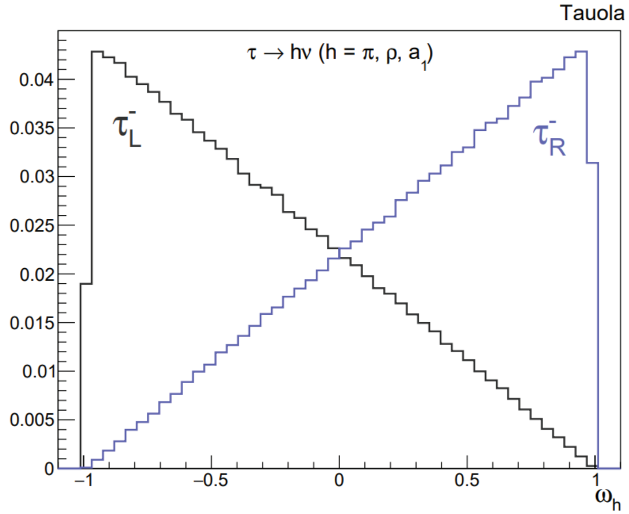
- A competitive precision to LEP can be achieved analysing all τ decays that can be identified at LHC
- Similarly to H→ ττ one may try to maximise the analyzing power reconstructing the event kinematic and the full polarimetric vector
- The observable is an angle between τ direction and the polarimetric vector

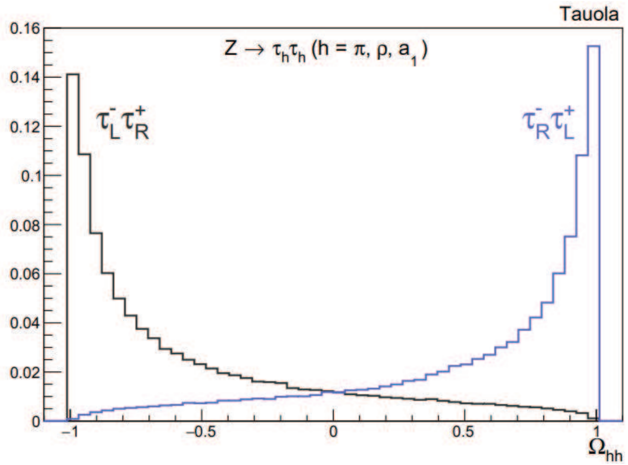


Corresponds to the optimal observable ω = $cos\theta_h$

M. Davier, et.al., Phys.Lett. B306 (1993) 411-417

- ≈100% anti-correlation of τ leptons spins allows to further gain the analysing power





18/09/2018 vladimir.cherepanov@cern.ch 4

- The precision of the longitudinal and transverse spin measurement in H→ $\tau\tau$ and Z→ $\tau\tau$ can be gained considering all possible τ decays
- The full kinematic analysis might help to maximize the analysing power in measurements of τ spin effects in H→ $\tau\tau$ and Z→ $\tau\tau$
- Several tools exist that can be used for estimation the invisible momentum in τ decays (using missing transverse energy, track impact parameters, τ decay vertex etc)
- In practice a reasonably good performance is expected only in channels where robust reconstruction of τ decay point is possible: Z/H→ $\tau\tau$ $\rightarrow a_1 + X$, a_1+a_1 with a_1 decaying to three charged pions.
- The final choice of discriminant in each decay category can be concluded after all detector effects are studied and understood.

CMS work in progress…

Example of papers from huge ML domain

1. Impressive talks https://indico.cern.ch/event/673350/
   https://indico.cern.ch/event/687788/
2. A lot of solutions available from https://root.cern.ch/tmva
3. Examples of applications (papers I read):
   (a) K. Fraser and M. D. Schwartz, arXiv:1803.08066
   (b) P. Baldi, K. Bauer, C. Eng, P. Sadowski and D. Whiteson, "Jet Substructure Classification in High-Energy Physics with Deep Neural Networks," Phys. Rev. D **93** (2016) no.9, 094034
   (c) E. Bothmann and L. Del Debbio, "Reweighting a parton shower using a neural network: the final-state case," arXiv:1808.07802
   (d) D. Guest, K. Cranmer and D. Whiteson, "Deep Learning and its Application to LHC Physics," arXiv:1806.11484
   (e) . Baldi, P. Sadowski and D. Whiteson, Nature Commun. **5**, 4308 (2014)
4. I am only a user interested in how to prepare input for ML solution and in results.
5. I am worried about systematic errors of multi-dimensional distributions.
6. Big interest among students for projects of contact with industrial/other domain techniques is present. They should find the ways. **If** they want to use such tools they **need to understand mathematical foundations**.

News on the programs: PHOTOS, Tauspinner

1. `Photos` `http://photospp.web.cern.ch/photospp/` Monte Carlo for bremsstrahlung in decays was enriched with emission of pairs, tests are published:

   S. Antropov, A. Arbuzov, R. Sadykov and Z. Was, “Extra lepton pair emission corrections to Drell-Yan processes in PHOTOS and SANC,” Acta Phys. Polon. B **48** (2017) 1469

2. `TauSpinner` `http://tauolapp.web.cern.ch/tauolapp/`, algorithm for re-weighting τ production and decays was enriched. This program was used to obtain results presented in earlier parts of my talk. References:

a T. Przedzinski, E. Richter-Was and Z. Was, “Documentation of TauSpinner algorithms – program for simulating spin effects in tau-lepton production at LHC,” arXiv:1802.05459.

b E. Richter-Was and Z. Was, “The TauSpinner approach for electroweak corrections in LHC Z to ll observable,” arXiv:1808.08616

c E. Barberio, B. Le, E. Richter-Was, Z. Was, D. Zanzi and J. Zaremba, “Deep learning approach to the Higgs boson CP measurement in $H \to \tau\tau$ decay and associated systematic,” Phys. Rev. D **96** (2017) no.7, 073002

News on the programs: TAUOLA, τ decay Monte Carlo

1. TAUOLA with new hadronic currents, up to 500 decay channels, which can be manipulated by user is published: M. Chrzaszcz, T. Przedzinski, Z. Was and J. Zaremba, Comput. Phys. Commun. **232**, 220 (2018) . For archivization purposes initialization of hadronic decay channels is compatible with defaults as BaBar was using.

2. Direction for work is essentially set. I have not received much feed-backs, but there were no objections too.

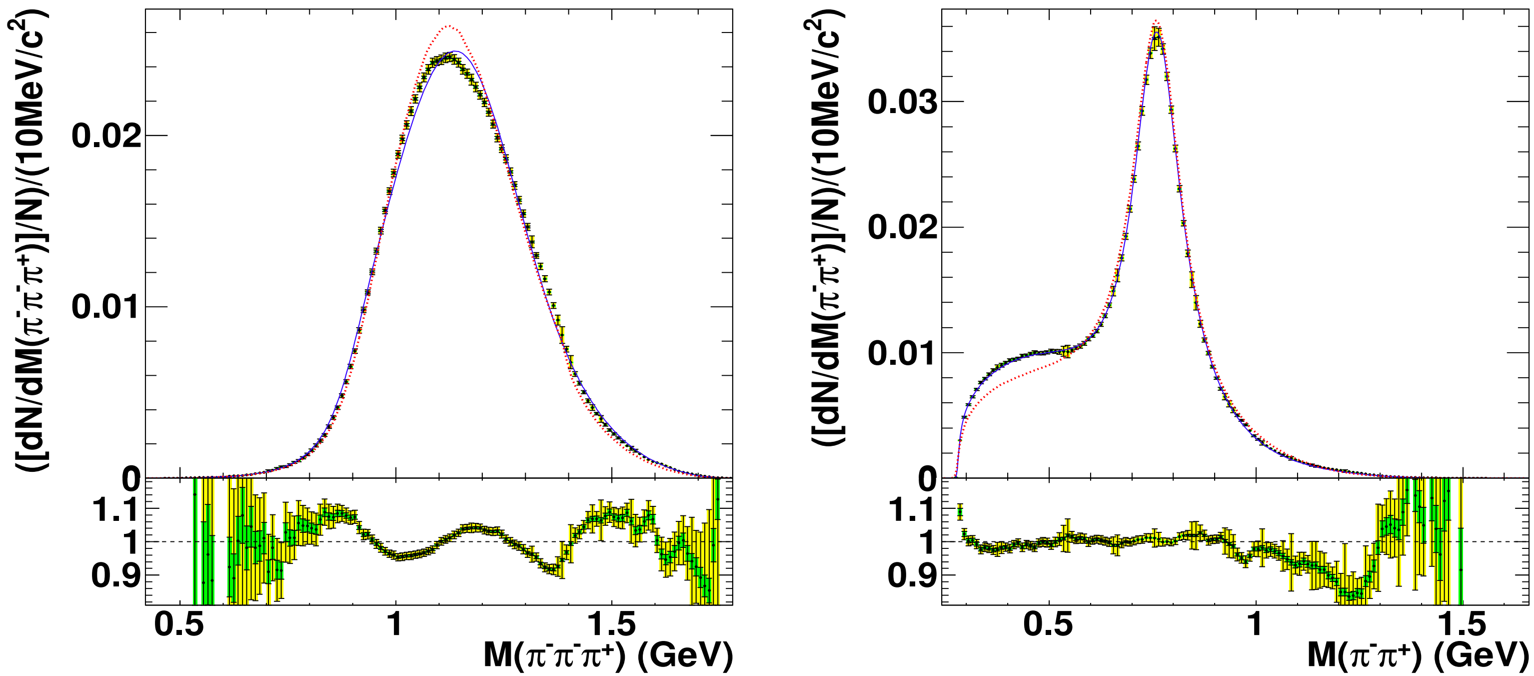
3. Program is prepared to be translated piece after piece into C++, or other language. Whenever a need will arrive.

4. **Theoretical uncertainty of the models can be $\frac{1}{N_C}$, $\frac{1}{N_C^2}$ or $\cdots$,** but experimental precision has to be assumed to be better than 0.001. That is a factor of 100 better.

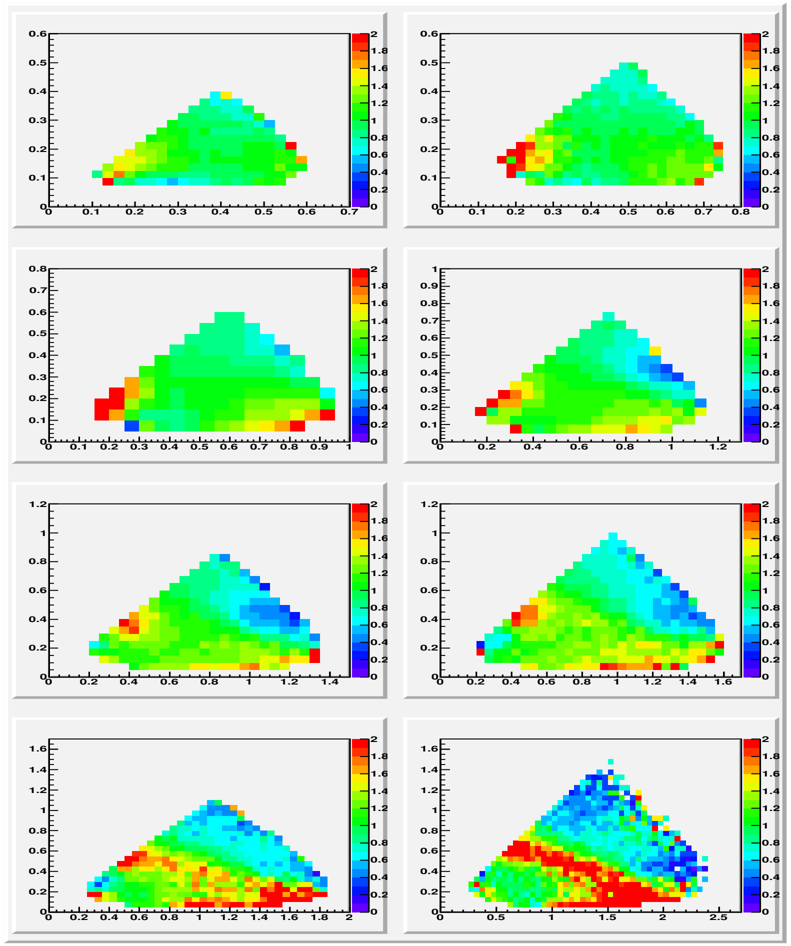
5. Not much to add ...

New currents for $\tau \to 3\pi$ and $\tau \to 2\pi$ decays

Currents from Resonance Chiral Lagrangian approach, fits to BaBar data. Experimental systematic errors considered. From: *Resonance Chiral Lagrangian Currents and Experimental Data for* $\tau^- \to \pi^-\pi^-\pi^+\nu_\tau$, I.M. Nugent, T. Przedzinski, P. Roig, O. Shekhovtsova, Z. Was, Phys. Rev. D 88, 093012 (2013). **Looks like a step of successful strategy. See the next slide for concern.**



Dalitz plot ratios (Red, blue differences ∼50% or more); **TAUOLA RChL** to **TAUOLA CLEO**, $m^2(3\pi)$ in ranges: 0.36-0.81, 0.81-1.0, 1.0-1.21, 1.21-1.44, 1.44-1.69, 1.69-1.96, 1.96-2.25, 2.25-3.24 GeV^2.



Conclusions and outlook

- Some examples of how new techniques for data analysis became useful for multi-dimensional distributions involving $H \to \tau\tau$ decays and Higgs CP-parity measurement.
- Useful for that properties of Monte Carlo simulation programs and properties of ML techniques were underlined.
- Recent developments for `Tauola` `TauSpiunner` and `Photos` programs were listed.
- Important properties of predictions for models used to describe $\tau \to 3\pi\nu$ decays were underlined.
- Fits involving multi-dimensional distributions are highly desirable.
- Essential for future developments will be thus control of systematic errors for such multi-dimensional distributions.
- If experimental data should have background subtracted, or if dominant backgrounds will be fitted simultaneously with the signature is technically less important.
- Question of manpower and training as well as motivation of involved people is very important. Competition for talent from other fields can not be ignored.