LHC Higgs CP sensitive observables in

\[ H \rightarrow \tau^+ \tau^-; \tau^\pm \rightarrow (3\pi)^\pm \nu \]

and Machine Learning benefits

Z. Was

- (1) Goal: observables carrying quality physics messages.
- (2) Often such quantities can be neither measured nor predicted ...
- (3) ... if all necessary details are taken into account.
- (4) Nonetheless it is a goal. However: never break rules.
- (5) Adjust corners: sufficiently well, but not unberably well.
- (6) Challenges: mathematics, physics, algorithms, computing, detector details.
- (7) In my talk Higgs CP measurement will serve as instructive example.
Optimal Variables → Machine Learning → and back

1. Goal: clearly defined quantities easy to measure and interpret.

2. Often it looks like that, later complications arrive:
   (a) experimental side: multi-dimensional, backgrounds, detector structure.
   (b) theory side: multi-dimensional due to jets and/or QED bremsstrahlung. Higher order, loop effects, damage canonical shapes.
   (c) At the end simplicity is gone ...

3. Possible solutions? Is ML good for everything?
   What are the limitations?
**Optimal variables → Machine Learning → and back**

1. 50 years ago (?): measure directly interesting quantities

   
   (a) Look for most sensitive to physics, 1-dimensional distributions, develop optimal variable approach.

   (b) Hide complexity (cut offs, higher order corrections) in shape modifications.

   (c) Use Monte Carlo techniques to prepare templates.

   (d) Prepare functions for fits. Parametrize effects breaking the picture too.
      
      If possible, estimate that they are negligible.

3. What does it mean for multi dimensional signatures and e.g. Machine Learning techniques?

4. I will use my 20 year experience with $H \to \tau\tau$ and CP.
General picture \(\rightarrow\) faulty points \(\rightarrow\) recoveries

1. \(d\sigma = (\sum_{\lambda_1, \lambda_2} |M_{prod}|^2)(\sum_{\lambda_1} |M^{\tau^+}|^2)(\sum_{\lambda_2} |M^{\tau^-}|^2)wt_{spin}d\Omega_{prod} \ d\Omega_{\tau^+} \ d\Omega_{\tau^-}\)

2. Above formula is \textcolor{red}{exact but in practice useless.} We work with:

\[
d\sigma = \sum_{flav.} \int dx_1 dx_2 f(x_1, ...) f(x_2, ...) d\Omega_{part. lev.} \ d\Omega_{\tau^+} \ d\Omega_{\tau^-}
\]

\[
(\sum_{\lambda_1, \lambda_2} |M_{prod \ part. lev.}|^2)(\sum_{\lambda_1} |M^{\tau^+}|^2)(\sum_{\lambda_2} |M^{\tau^-}|^2)wt_{spin}.
\]

3. where \(\sum_{\lambda_1, \lambda_2} |M_{prod \ part. lev.}|^2 \rightarrow d\sigma_{Born}^{q\bar{q}}(\hat{s}, \cos \theta)\) and we sometimes use:

\[
d\sigma_{Born}(x_1, x_2, \hat{s}, \cos \theta) = \sum_{q_f, \bar{q}_f} [f^{q_f}f(x_1, ...) f^{\bar{q}_f}f(x_2, ...) d\sigma_{Born}^{q_f\bar{q}_f}(\hat{s}, \cos \theta)
\]

\[+ f^{\bar{q}_f}f(x_1, ...) f^{q_f}f(x_2, ...) d\sigma_{Born}^{q_f\bar{q}_f}(\hat{s}, - \cos \theta)],
\]

4. \textcolor{red}{Nearly all is faulty.} What can be corrected. What can be avoided.

Z. Was

Moriond, March, 2019
Properties of matrix elements.

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

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

- The decay probability

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

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

- ’Higgs spin’ is blind on Higgs origin. But it is not true for the background DY processes.
Higgs boson Yukawa coupling expressed with the help of the scalar–pseudo-scalar mixing angle $\phi$

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

Decay probability for the mixed scalar–pseudo-scalar case

$$\Gamma(h_{mix} \rightarrow \tau^+ \tau^-) \sim 1 - s^\tau_\parallel s^\tau_\parallel + s^\tau_\perp R(2\phi) s^\tau_\perp$$

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

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

Pure scalar case is reproduced for $\phi = 0$.

For $\phi = \pi/2$ we reproduce the pure pseudo-scalar case.
Note big change: \( s \rightarrow h \) where \( h \) is a function of \( \tau \) decay products.

- Case of \( \tau \rightarrow \pi \nu_\tau \) decay, \( \mathcal{BR}(\tau \rightarrow \pi \nu_\tau) = 10\% \), also M.E. expressed.

- Note that \( h^i \) is defined in \( \tau \) r.f, and always \( |\vec{h}| = 1 \).

\[
\Gamma(h_{mix} \rightarrow \tau^+ \tau^-) \sim 1 - h^\tau_\parallel h^\tau_\parallel + h^\tau_\perp R(2\phi) h^\tau_\perp
\]

\[h^i = p^i\]

- Left: \( \tau \) decay channel independent distribution of polarimetric \( h^i \). Right: Acollinearity of \( \pi^+ \) and \( \pi^- \) is perfect.

All sensitivity is transmitted to acollinearity distribution. Problem: one needs to reconstruct acollinearity of \( \pi^+ \pi^- \) in \( H \) rest-frame. Watch the scale Observable was realistic for Higgs of much lower mass !!!

Quest was set: Z.Phys. C64 (1994) 21

Z. Was

Moriond, March, 2019
Toward realistic observable.

Transverse spin correlations through $\tau$ decays

- **Case of** $\tau \to \rho \nu_\tau$ **decay**, $\mathcal{BR}(\tau \to \rho \nu_\tau) = 25\%$, also M.E. expressed.

- **Polarimeter vector** $\vec{h}^i$ is (where $q$ for $\pi^\pm - \pi^0$ and $N$ for $\nu_\tau$ four momenta.

\[ \vec{h}^i = \mathcal{N} \left( 2(q \cdot N)q^i - q^2 N^i \right) \]

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

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

\[ y_1 y_2 > 0 ; \quad y_1 y_2 < 0 \text{ (in } \tau^\pm \text{ r.f.'s)} \]

\[ y_1 = \frac{E_{\pi^+} - E_{\pi^0}}{E_{\pi^+} + E_{\pi^0}} ; \quad y_2 = \frac{E_{\pi^-} - E_{\pi^0}}{E_{\pi^-} + E_{\pi^0}}. \]
For mixing angle $\phi$, transverse component of $\tau^+$ spin polarization vector is correlated with the one of $\tau^-$ rotated by angle $2\phi$.

Acoplanarity $0 < \varphi^* < 2\pi$ is of physical interest, not just $\arccos n_- \cdot n_+$. 

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: $n_\pm = p_{\pi\pm} \times p_{\pi^0}$

Find the sign of $p_{\pi^-} \cdot n_+$

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

Otherwise $2\pi - \varphi^*$
Toward realistic observable.

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

- Only events where the signs of $y_1$ and $y_2$ are the same whether calculated using the method without or with the help of the $\tau$ impact parameter.


- 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.

Z. Was

Moriond, March, 2019
Toward realistic observable.

Observable relies on 3 variables $\phi^*, y_1, y_2$ even if final plots were 1-dimensional

To improve sample size:

- $\mathcal{BR}(\tau \to \pi \nu_\tau) = 10\%$, that mean 1\% of $H \to \tau\tau$, also a wild challenge to measure acollinearity.
- $\mathcal{BR}(\tau \to \rho \nu_\tau) = 25\%$, 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 $\nu_\tau$ is not observable.
- For the $\rho \to \pi^-, \pi^0$ we can define one plane for acoplanarity definition,
- for the $a_1^- \to \pi^-, \pi^-, \pi^+$ we can define four such planes.
- Each plane bring its own $y_i$ variable. Observable to exploit space of $4 \cdot 4 + 4 \cdot 2 = 24$ dimensions?
Monte Carlo features useful for ML.

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.

Z. Was

Moriond, March, 2019
Feasability → Fundamental ML benchmark.

Acoplanarity angles of oriented half decay planes: $\varphi_{\rho\rho}^*$ (left), $\varphi_{a\rho}^*$ (middle) and $\varphi_{a_a}^*$ (right), for events grouped by the sign of $y_{\rho\rho}^+, y_{\rho\rho}^-$, $y_{a\rho}^+ y_{\rho\rho}^-$ and $y_{a_a}^+ y_{a_a}^-$ respectively. Three CP mixing angles $\phi_{CP}^{CP} = 0.0$ (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 $\phi_{CP}^{CP}$ mixing scalar pseudo-scalar angle, which brings linear shift. I remained frustrated for 15 years, how to digest...

Z. Was

Moriond, March, 2019
Feasability → Fundamental ML benchmark.

- I can not present ML details, I am not good choice for that. Necessary detail:
- Area Under Curve (AUC): essentially probability that Network will identify event to be scalar when it was scalar. There are three other possibilities: (ii) scalar when it was pseudoscalar, (iii) pseudoscalar when it was scalar, (iv) pseudoscalar when it was pseudoscalar, to take into account in AUC.
- 0.5 means random choice. 1.0 would mean certainty.
- Classification in-between is useful, with sufficiently large events sample.
- Results did not came in automatic way. We had to:
  - boost events to rest frame of all visible objects combined,
  - rotate all to set $\tau^+$ primary decay resonance along z axis.
  - But a lot of fancy work, seemed for a while to be irrelevant.
<table>
<thead>
<tr>
<th>Features/variables</th>
<th>$\rho^\pm - \rho^\mp$</th>
<th>$a_1^\pm - \rho^\mp$</th>
<th>$a_1^\pm - a_1^\mp$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\rho^\pm \rightarrow \pi^0 \pi^\pm$</td>
<td>$\rho^0 \rightarrow \pi^+ \pi^-$</td>
<td>$\rho^\mp \rightarrow \pi^0 \pi^\pm$</td>
<td>$\rho^0 \rightarrow \pi^+ \pi^-$</td>
</tr>
<tr>
<td>True classification</td>
<td>0.782</td>
<td>0.782</td>
<td>0.782</td>
</tr>
<tr>
<td>$\varphi_{i,k}^*$</td>
<td>0.500</td>
<td>0.500</td>
<td>0.500</td>
</tr>
<tr>
<td>$\varphi_{i,k}^*$ and $y_i$, $y_k$</td>
<td>0.624</td>
<td>0.569</td>
<td>0.536</td>
</tr>
<tr>
<td>4-vectors</td>
<td>0.638</td>
<td>0.590</td>
<td>0.557</td>
</tr>
<tr>
<td>$\varphi_{i,k}^*$, 4-vectors</td>
<td>0.638</td>
<td>0.594</td>
<td>0.573</td>
</tr>
<tr>
<td>$\varphi_{i,k}^*$, $y_i$, $y_k$ and $m_i^2$, $m_k^2$</td>
<td>0.626</td>
<td>0.578</td>
<td>0.548</td>
</tr>
<tr>
<td>$\varphi_{i,k}^*$, $y_i$, $y_k$, $m_i^2$, $m_k^2$ and 4-vectors</td>
<td>0.639</td>
<td>0.596</td>
<td>0.573</td>
</tr>
</tbody>
</table>

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.
<table>
<thead>
<tr>
<th>Features</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>φ* 4-vec y_i m_i</td>
<td>Ideal ± (stat)</td>
<td>Smeared ± (stat) ± (syst)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>✓ ✓ ✓ ✓</td>
<td>0.6035 ± 0.0005</td>
<td>0.5923 ± 0.0005 ± 0.0002</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>✓ ✓ ✓ -</td>
<td>0.5965 ± 0.0005</td>
<td>0.5889 ± 0.0005 ± 0.0002</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>✓ ✓ - ✓</td>
<td>0.6037 ± 0.0005</td>
<td>0.5933 ± 0.0005 ± 0.0003</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- ✓ - -</td>
<td>0.5971 ± 0.0005</td>
<td>0.5892 ± 0.0005 ± 0.0002</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>✓ ✓ - -</td>
<td>0.5971 ± 0.0005</td>
<td>0.5893 ± 0.0005 ± 0.0002</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>✓ - ✓ ✓</td>
<td>0.5927 ± 0.0005</td>
<td>0.5847 ± 0.0005 ± 0.0002</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>✓ - ✓ -</td>
<td>0.5819 ± 0.0005</td>
<td>0.5746 ± 0.0005 ± 0.0002</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>✓ ✓ ✓ ✓</td>
<td>0.5669 ± 0.0004</td>
<td>0.5657 ± 0.0004 ± 0.0001</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>✓ ✓ ✓ -</td>
<td>0.5596 ± 0.0004</td>
<td>0.5599 ± 0.0004 ± 0.0001</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>✓ ✓ - ✓</td>
<td>0.5677 ± 0.0004</td>
<td>0.5661 ± 0.0004 ± 0.0001</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- ✓ - -</td>
<td>0.5654 ± 0.0004</td>
<td>0.5641 ± 0.0004 ± 0.0001</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>✓ ✓ - -</td>
<td>0.5623 ± 0.0004</td>
<td>0.5615 ± 0.0004 ± 0.0001</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>✓ - ✓ ✓</td>
<td>0.5469 ± 0.0004</td>
<td>0.5466 ± 0.0004 ± 0.0001</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>✓ - ✓ -</td>
<td>0.5369 ± 0.0004</td>
<td>0.5374 ± 0.0004 ± 0.0001</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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”.

Z. Was

Moriond, March, 2019
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 $\tau$ decays should give ideal $0.782$ classification from all $\tau$ decay modes.

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


Is there anything that ML (and present day efforts) can bring?
My struggle with impact parameter, 15 years later

- A key for the improvement is to reconstruct $\nu_{\tau}$ momenta: 6 variables.
- We have 3 good constraints (partly correlated) and 2 less precise:
  - $m_{\tau}^\pm$
  - $m_H$
  - $p_{T,mis.}^{x,y}$
- two missing angles, need to be obtained from impact parameter measurement.

A ML may be helpful? → arXiv:1812.08140

B Attempt to obtain optimal variable for Higgs CP may be complicated because of smearing and difficult kinematical constraints.

C May be use ML to control $h^i$? rather than Higgs decay chain?

D Then at the end we have 1-dim optimal variable, decay mode independent.

Higgs CP is then “easy”: matrix element (weight) very simple, see slide 7.
Applications of TauSpinner: Transverse spin correlation in $H \rightarrow \tau\tau$ at LHC (Full kinematic analysis)

- Alternatively one may try to estimate the invisible part of the polarimetric vectors of both $\tau$ 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 $\tau$'s in H r.f. and polarimetric vectors.

- Ideally carries the full analyzing power and irrespective of the $\tau$ decay channel.
1. We have demonstrated that ML techniques can be useful to distinguish in statistically controllable way between hypotheses of Higgs coupling to tau being CP even, CP-odd or even CP-mix 

*no problem if observables are massively multi-dimensional.*

2. **LESSON:** it is important to separate those degrees of freedom which can be controlled, from those where more effort is still needed.

3. In case of CP from Higgs production and decay, it is easy: Higgs is narrow and its spin is zero, H production is well separated from decay.

4. Problem may come with background. Note that Drell Yan production is by comparison huge, intermediate Z resonance is broad and carries spin.
Conclusions and outlook

- My attempt was to demonstrate how phenomenology of some processes at LHC can be prepared with the help of techniques, where for some sub-set of phase-space dimensions matrix element weights could be defined and used.

- Higgs CP was 'easy' because it is scalar and production does not affect its decay at all.

- Case of Z background was mentioned.

- Our references for the issue of separating good from the badly controlled variables:
Conclusions and outlook

Figure 1: Artificial Neural Networks have spurred remarkable recent progress in image classification and speech recognition. But even though these are very useful tools based on well-known mathematical methods, we actually understand surprisingly little of why certain models work and others don’t.

From http://googleresearch.blogspot.com/2015/06/inceptionism-going-deeper-into-neural.html

Pattern recognition is an active field and deep concern and not only for us.

Z. Was Moriond, March, 2019
Example of papers from huge ML domain

1. Impressive talks https://indico.cern.ch/event/673350/ .../event/687788/


3. A lot of solutions available from https://root.cern.ch/tmva

4. Examples of applications (papers I read):
   (a) K. Fraser and M. D. Schwartz, arXiv:1803.08066
   (e) P. Baldi, P. Sadowski and D. Whiteson, Nature Commun. 5, 4308 (2014)

5. I am only a user interested in how to prepare input for ML solution and in results.

6. **Systematic errors of ML results on multi-dimensional distributions are benchmarked with optimal variables approach.**

Z. Was

Moriond, March, 2019