Karlsruher Institut für Technologie

$t\bar{t}$ + jets/HF model and systematics in CMS

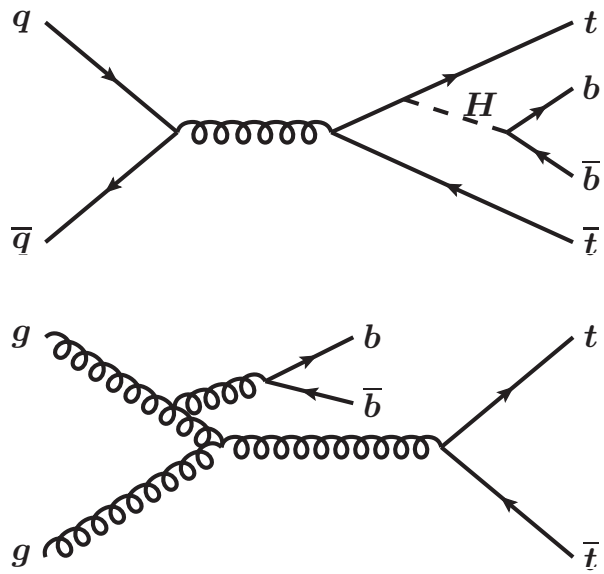
Higgs Toppings Workshop - Probing Top-Higgs Interactions at the LHC (Benasque)

Carmen Diez Pardos (DESY) and **Matthias Schröder** (KIT) | May 28, 2018

INSTITUT FÜR EXPERIMENTELLE TEILCHENPHYSIK (ETP)

$t\bar{t} + b\bar{b}$: Dominant Background to $t\bar{t}H(b\bar{b})$

- Search for $t\bar{t}H(b\bar{b})$ (leptonic)[1]: final states with high jets and b jets multiplicity
- Largest background: $t\bar{t} +$ jets
  - Inclusive $t\bar{t} +$ jets: $\sigma_{t\bar{t}} = 832\,\text{pb}$
  - Irreducible $t\bar{t} + b\bar{b}$ background: $\sigma_{t\bar{t}+b\bar{b}} \approx 4\,\text{pb}$
  - $t\bar{t}H$: $\sigma_{t\bar{t}H} = 0.5\,\text{pb}$
- $t\bar{t} +$ HF modeling very challenging
  - CMS 13 TeV measurement of inclusive $t\bar{t} + b\bar{b}$ cross section[2]: precision $\approx 35\,\%$



Phase space		$\sigma_{t\bar{t}b\bar{b}}$ [pb]
Visible	Measurement	$0.088 \pm 0.012 \pm 0.029$
	SM (POWHEG)	0.070 ± 0.009
Full	Measurement	$4.0 \pm 0.6 \pm 1.3$
	SM (POWHEG)	3.2 ± 0.4

[1] CMS-HIG-17-026, subm. to JHEP

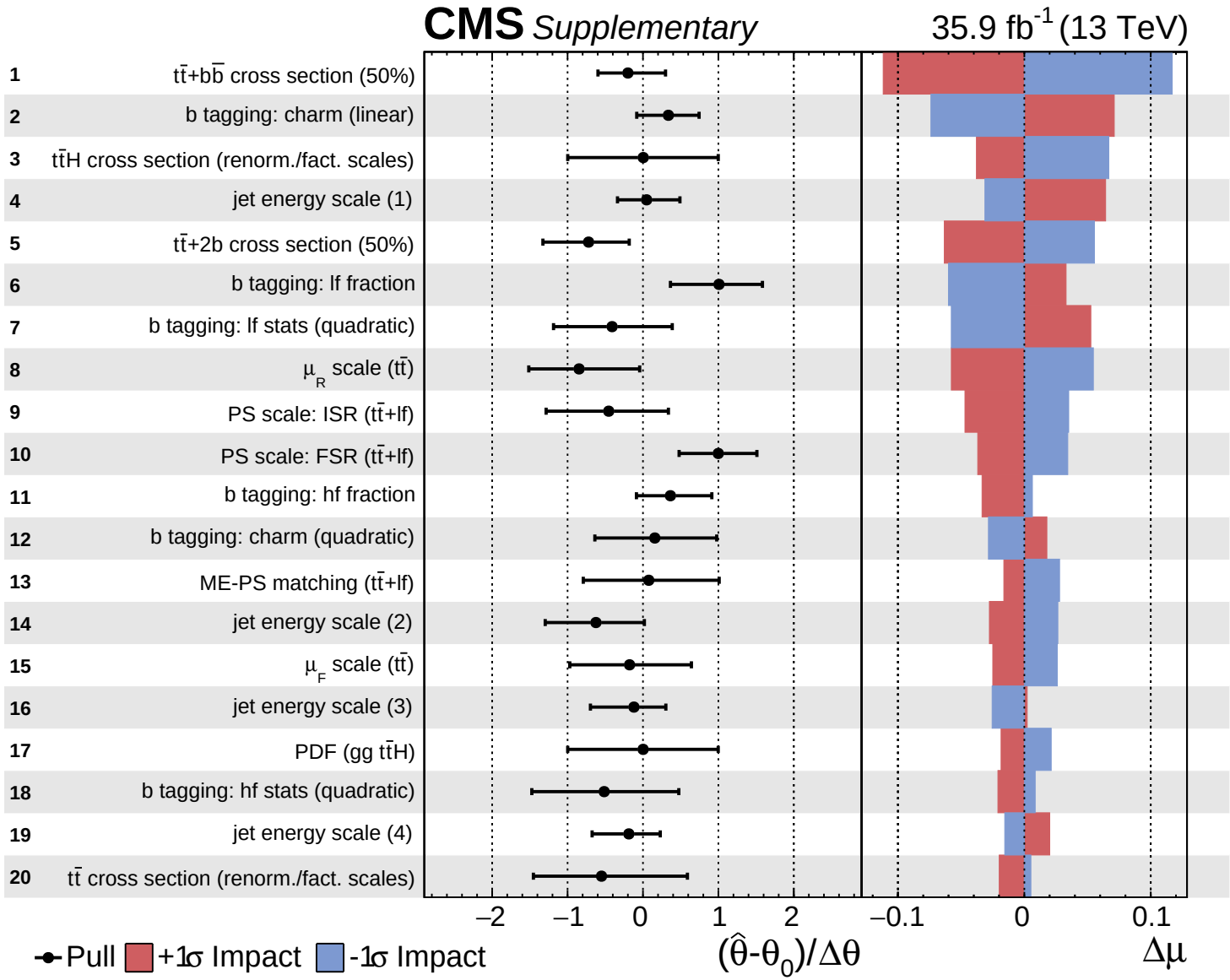
[2] 10.1016/j.physletb.2017.11.043

Impact of Systematic Uncertainties

Uncertainty source	$\pm\Delta\mu$ (observed)	$\pm\Delta\mu$ (expected)
Total experimental	$+0.15/-0.16$	$+0.19/-0.17$
b tagging	$+0.11/-0.14$	$+0.12/-0.11$
jet energy scale and resolution	$+0.06/-0.07$	$+0.13/-0.11$
Total theory	$+0.28/-0.29$	$+0.32/-0.29$
$t\bar{t}$+hf cross section and parton shower	$+0.24/-0.28$	$+0.28/-0.28$
Size of the simulated samples	$+0.14/-0.15$	$+0.16/-0.16$
Total systematic	$+0.38/-0.38$	$+0.45/-0.42$
Statistical	$+0.24/-0.24$	$+0.27/-0.27$
Total	$+0.45/-0.45$	$+0.53/-0.49$

Most important systematic uncertainties related to $t\bar{t}$ + HF modeling

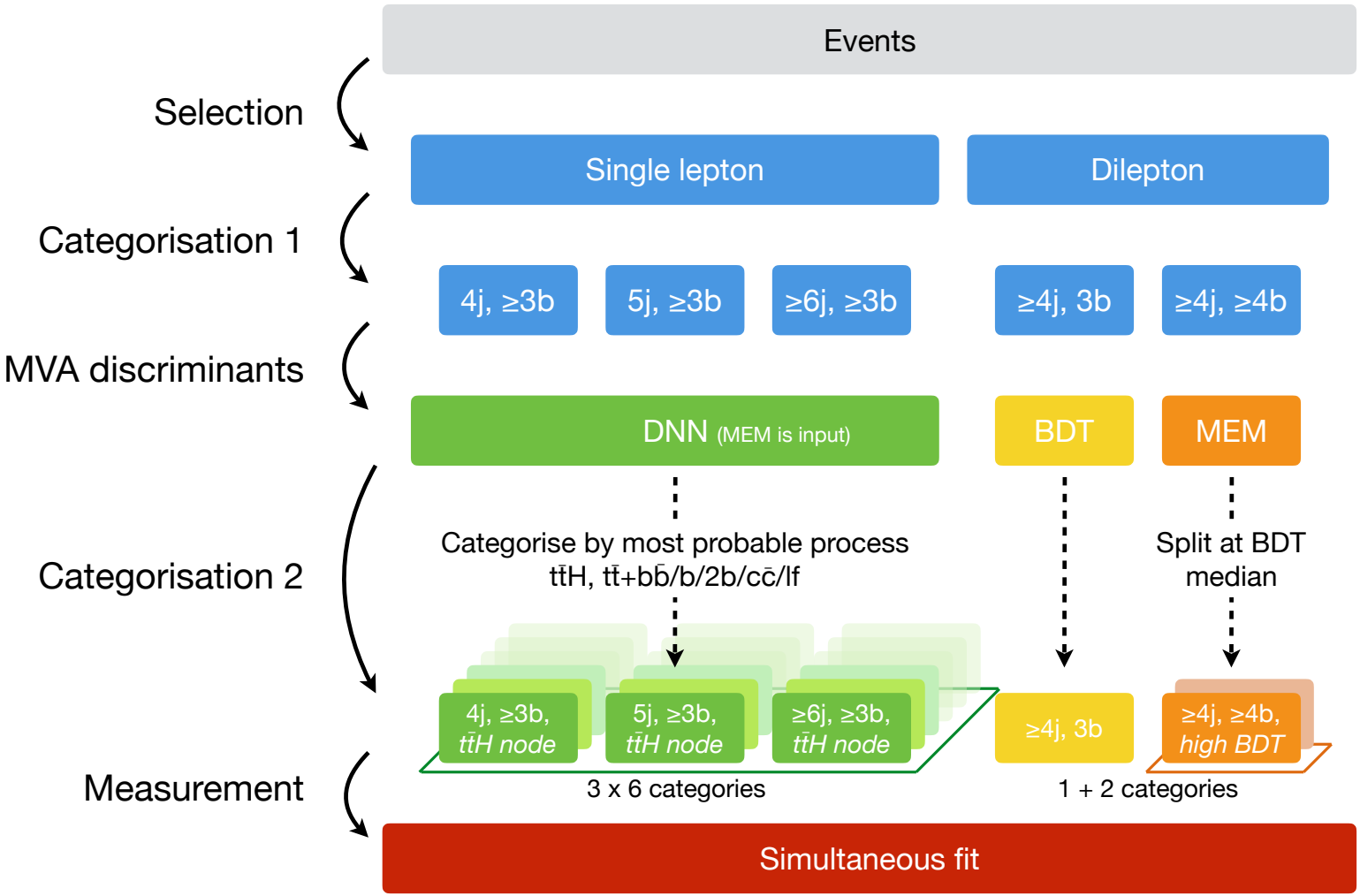
Impact of Systematic Uncertainties



$t\bar{t}$ + HF Definition

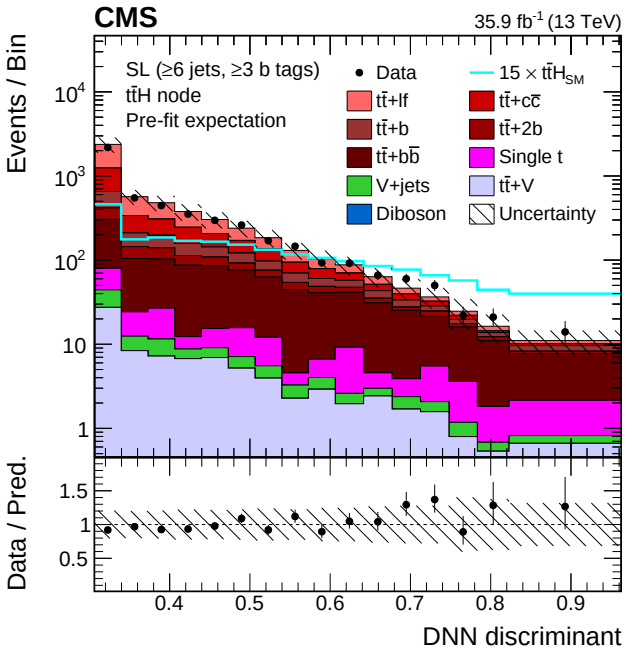
- Events classified based on the **flavour of the additional jets** that do not stem from the decay of a t quark
  - Jets at particle level with $p_T > 15\,\text{GeV}$, $|\eta| < 2.5$
  - HF hadrons associated to jets via ghost-hadron matching
- Define 5 $t\bar{t}$ + X processes:
  - **$t\bar{t}$ + $b\bar{b}$** : $\geq$ 2 add. jets containing $\geq$ 1 b-hadrons each
  - **$t\bar{t}$ + b** : 1 add. jet containing 1 b-hadron
  - **$t\bar{t}$ + 2b** : 1 add. jet containing $\geq$ 2 b-hadrons
  - **$t\bar{t}$ + $c\bar{c}$** : $\geq$ 1 add. jets containing c-hadrons
  - **$t\bar{t}$ + LF** : otherwise

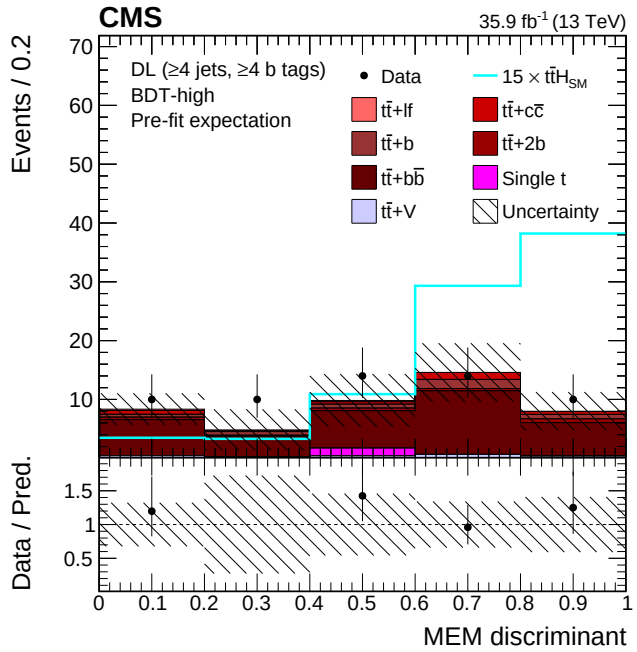
Reminder: Analysis Strategy



$t\bar{t}$ + jets Modelling

- Inclusive POWHEG+PYTHIA8 $t\bar{t}$ NLO simulation
  - CUETP8M2T4 tune, NNPDF3.0
- Normalised to NNLO+NNLL cross section of 832 pb
- Events split into the 5 $t\bar{t}$ + X processes





Shape and normalisation from POWHEG+PYTHIA8

$t\bar{t}$ + jets Uncertainties

- Rate uncertainties on **NNLO $t\bar{t}$ cross section** prediction
  - Includes ME and PDF, correlated among all $t\bar{t}$ + X processes
- Additional **50 % rate uncertainty** per $t\bar{t} + b\bar{b}$, $t\bar{t} + 2b$, $t\bar{t} + b$, and $t\bar{t} + c\bar{c}$ process
  - Conservative prior choice compared to prediction and measurement

$t\bar{t}$ + jets Uncertainties

- Rate uncertainties on **NNLO $t\bar{t}$ cross section** prediction
  - Includes ME and PDF, correlated among all $t\bar{t}$ + X processes
- Additional **50 % rate uncertainty** per $t\bar{t} + b\bar{b}$, $t\bar{t}$ + 2b, $t\bar{t}$ + b, and $t\bar{t} + c\bar{c}$ process
  - Conservative prior choice compared to prediction and measurement
- Shape uncertainty from **NNPDF3.0 PDF** set used at MC generation
  - Envelope of PDF replicas, correlated among all $t\bar{t}$ + X processes and $t\bar{t}H$
- Shape uncertainty from **QCD scales** μ_F and μ_R at ME level
  - By reweighting $t\bar{t}$ sample, correlated among all $t\bar{t}$ subprocesses and $t\bar{t}H$

$t\bar{t}$ + jets Uncertainties

- Rate uncertainties on **NNLO $t\bar{t}$ cross section** prediction
  - Includes ME and PDF, correlated among all $t\bar{t}$ + X processes
- Additional **50 % rate uncertainty** per $t\bar{t} + b\bar{b}$, $t\bar{t}$ + 2b, $t\bar{t}$ + b, and $t\bar{t} + c\bar{c}$ process
  - Conservative prior choice compared to prediction and measurement
- Shape uncertainty from **NNPDF3.0 PDF** set used at MC generation
  - Envelope of PDF replicas, correlated among all $t\bar{t}$ + X processes and $t\bar{t}H$
- Shape uncertainty from **QCD scales** μ_F and μ_R at ME level
  - By reweighting $t\bar{t}$ sample, correlated among all $t\bar{t}$ subprocesses and $t\bar{t}H$
- Uncertainties on **parton shower** (PS), **ME-PS matching**, and **underlying event**
  - From additional $t\bar{t}$ samples generated with varied parameters ($\times$0.5, 2)
  - Due to limited statistical precision, conservatively estimated as jet-multiplicity dependent rate uncertainty
  - Independent for each of the $t\bar{t}$ + X classes

$t\bar{t}$ + jets Uncertainties: Validation

- Sensitivity to $t\bar{t}$ + HF rate changes investigated with toy data
  - **varied prior $t\bar{t}$ + HF uncertainties**
  - **varied $t\bar{t}$ + $b\bar{b}$ rate in toy data**

$t\bar{t}$ + jets Uncertainties: Validation

- Sensitivity to $t\bar{t}$ + HF rate changes investigated with toy data
  - **varied prior $t\bar{t}$ + HF uncertainties**
  - **varied $t\bar{t}$ + $b\bar{b}$ rate in toy data**
- Sensitivity to different $t\bar{t}$ + $b\bar{b}$ shapes as expected from Sherpa+OpenLoops
  - Toy data where $t\bar{t}$ + $\geq$ 1b background template shapes are replaced by **Sherpa+OpenLoops (4F)** prediction
  - Different injected signal strengths: signal recovered within few percent

$t\bar{t}$ + jets Uncertainties: Validation

- Sensitivity to $t\bar{t}$ + HF rate changes investigated with toy data
  - **varied prior $t\bar{t}$ + HF uncertainties**
  - **varied $t\bar{t}$ + $b\bar{b}$ rate in toy data**
- Sensitivity to different $t\bar{t}$ + $b\bar{b}$ shapes as expected from Sherpa+OpenLoops
  - Toy data where $t\bar{t}$ + $\geq$ 1b background template shapes are replaced by **Sherpa+OpenLoops (4F)** prediction
  - Different injected signal strengths: signal recovered within few percent
- Description of **input and control variables verified in data** pre-fit and post-fit
- **Step-wise unblinding** procedure
  - Successively correlating different categories in fit
  - Fitting also control variables

$t\bar{t}$ + jets Uncertainties: Validation

- Sensitivity to $t\bar{t}$ + HF rate changes investigated with toy data
  - **varied prior $t\bar{t}$ + HF uncertainties**
  - **varied $t\bar{t}$ + $b\bar{b}$ rate in toy data**
- Sensitivity to different $t\bar{t}$ + $b\bar{b}$ shapes as expected from Sherpa+OpenLoops
  - Toy data where $t\bar{t}$ + $\geq$ 1b background template shapes are replaced by **Sherpa+OpenLoops (4F)** prediction
  - Different injected signal strengths: signal recovered within few percent
- Description of **input and control variables verified in data** pre-fit and post-fit
- **Step-wise unblinding** procedure
  - Successively correlating different categories in fit
  - Fitting also control variables

Fit model robust against rate and shape changes in data

Questions

- $t\bar{t} + b$ (MPI/FSR) b jets entirely from MPI or FSR gluon splitting: what is the motivation of this process?
- Reweighting of Powheg $t\bar{t} + X$ rates to Sherpa prediction
  - Differences essentially covered by uncertainties of predictions, why reweighting?
  - Comparison/validation with data?
- How well do the alternative samples describe the data?
- What was done to validate the input variables?
  - How are the pre-fit normalisation differences corrected? Freely-floating bkgs. sufficient?
- Technicalities of “2-point” uncertainties
  - How converted to up/down variation?
  - MC statistics for alternative generators? Is a template smoothing applied in all cases and what uncertainties are assigned?

Additional Material

Binary vs. Multi-Classification

Channel	Method	Best-fit μ $\pm\text{tot}\,(\pm\text{stat}\ \pm\text{syst})$
Single-lepton	BDT+MEM	$1.0^{+0.69}_{-0.66}\ \begin{pmatrix}+0.31 & +0.62\\ -0.30 & -0.59\end{pmatrix}$
Single-lepton	DNN	$1.0^{+0.58}_{-0.55}\ \begin{pmatrix}+0.30 & +0.50\\ -0.29 & -0.47\end{pmatrix}$
Dilepton	BDT+MEM	$1.0^{+1.22}_{-1.12}\ \begin{pmatrix}+0.65 & +1.04\\ -0.62 & -0.93\end{pmatrix}$
Dilepton	DNN	$1.0^{+1.38}_{-1.36}\ \begin{pmatrix}+0.71 & +1.18\\ -0.69 & -1.18\end{pmatrix}$
Combined	BDT+MEM	$1.0^{+0.60}_{-0.57}\ \begin{pmatrix}+0.28 & +0.53\\ -0.27 & -0.51\end{pmatrix}$
Combined	DNN	$1.0^{+0.55}_{-0.51}\ \begin{pmatrix}+0.27 & +0.47\\ -0.27 & -0.44\end{pmatrix}$