Accelerating global fits with Machine Learning

Sebastian Liem

GRAPPA
GRavitation AstroParticle Physics Amsterdam

Outline

Part I

Global scan of scalar DM EFT operators.

Real & Complex scalar DM

CMB, Direct detection, and Fermi dSph.

Galactic Centre Excess

Part II

Fast LHC Signal Prediction using Machine Learning

Gaussian Processes

Proof-of-concept: reconstruction

Global analysis of scalar DM EFT operators

Effective field theory of dark matter: a global analysis

Sebastian Liem,[a] **Gianfranco Bertone,**[a] **Francesca Calore,**[a] **Roberto Ruiz de Austri,**[b] **Tim M.P. Tait,**[c] **Roberto Trotta**[d,e] **and Christoph Weniger**[a]

[a] *GRAPPA, University of Amsterdam, Science Park 904, Amsterdam, 1098 XH Netherlands*

[b] *Instituto de Física Corpuscular, IFIC-UV/CSIC, C/ Catedrático José Beltrán 2, Paterna, Valencia, E-46980 Spain*

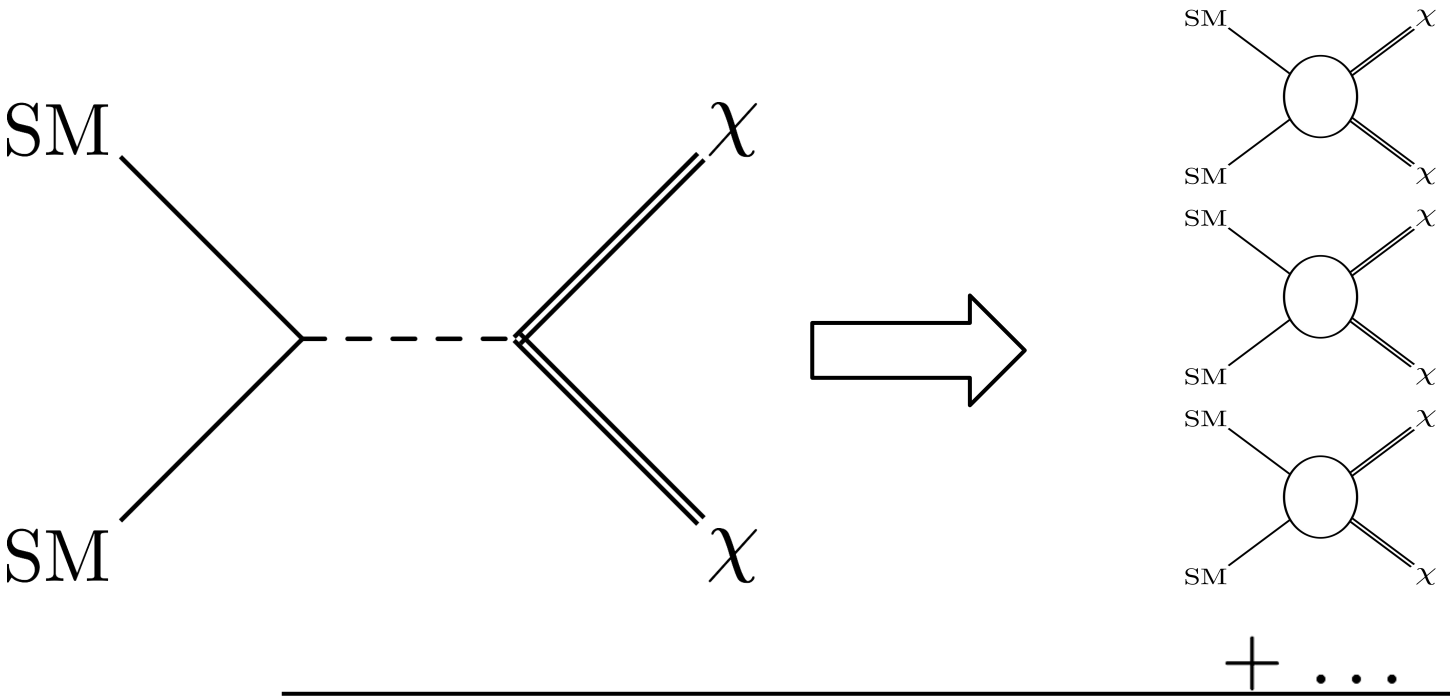
[c] *Department of Physics and Astronomy, University of California, Frederick Reines Hall, Irvine, CA, 92697 U.S.A.*

[d] *Astrophysics Group & Imperial Centre for Inference and Cosmology, Imperial College London, Blackett Laboratory, Prince Consort Road, London, SW7 2AZ United Kingdom*

[e] *Data Science Institute, Imperial College London, William Penney Laboratory, South Kensington Campus, London, SW7 2AZ United Kingdom*

JHEP09(2016)077

Effective field theory



The idea – combine all operators

If the EFT operators span the ‘theory space’ then

scan all EFT op $\Longrightarrow$ scan the ‘theory space’.

$$\mathcal{L} \supset \sum_i \lambda_i \mathcal{O}_i$$

DM-parton EFT operators

Real scalar DM operators				
Label	Coefficient	Operator	σ_{SI}	$\langle\sigma_{ann}v\rangle$
R1	$\lambda_1 \sim \frac{1}{2M^2}$	$m_q\chi^2\bar{q}q$	✓	s-wave
R2	$\lambda_2 \sim \frac{1}{2M^2}$	$im_q\chi^2\bar{q}\gamma^5 q$		s-wave
R3	$\lambda_3 \sim \frac{\alpha_s}{4M^2}$	$\chi^2 G_{\mu\nu}G^{\mu\nu}$	✓	s-wave
R4	$\lambda_4 \sim \frac{\alpha_s}{4M^2}$	$i\chi^2 G_{\mu\nu}\tilde{G}^{\mu\nu}$		s-wave

Complex scalar DM operators				
Label	Coefficient	Operator	σ_{SI}	$\langle\sigma_{ann}v\rangle$
C1	$\lambda_1 \sim \frac{1}{M^2}$	$m_q\chi^\dagger\chi\bar{q}q$	✓	s-wave
C2	$\lambda_2 \sim \frac{1}{M^2}$	$im_q\chi^\dagger\chi\bar{q}\gamma^5 q$		s-wave
C3	$\lambda_3 \sim \frac{1}{M^2}$	$\chi^\dagger\partial_\mu\chi\bar{q}\gamma^\mu q$	✓	p-wave
C4	$\lambda_4 \sim \frac{1}{M^2}$	$\chi^\dagger\partial_\mu\chi\bar{q}\gamma^\mu\gamma^5 q$		p-wave
C5	$\lambda_5 \sim \frac{\alpha_s}{8M^2}$	$\chi^\dagger\chi G_{\mu\nu}G^{\mu\nu}$	✓	s-wave
C6	$\lambda_6 \sim \frac{\alpha_s}{8M^2}$	$i\chi^\dagger\chi G_{\mu\nu}\tilde{G}^{\mu\nu}$		s-wave

Goodman et al.
arXiv:1008.1783

Data and likelihoods

Modified SuperBayeS: MultiNest, FeynRules, micrOMEGAs, PPPC 4 DM ID

$$\ln L = \ln L_{\Omega h^2}$$

Gaussian. Planck determination of **relic density.** (1502.01589)

$$+ \ln L_{\rm LUX}$$

LUXCalc (1502.02667). LUX **SI x-sec** (1310.8214)

$$+ \ln L_{\rm dSph}$$

Official likelihood. Fermi **dSph** Pass 8 (1503.02641)

$$+ \ln L_{\rm CMB}$$

Planck **CMB anisotropies** (1502.01589).
Cline & Scott; Slayter (1301.5908, 1506.03811)

Log prior

$$\log_{10} m_\chi \sim \text{Uniform}(0, 3)$$

$$\log_{10} \lambda_i \sim \text{Uniform}(-20, 0)$$

Dirichlet prior

$$\log_{10} m_\chi \sim \text{Uniform}(0, 3)$$

$$\lambda_i = f_i A$$

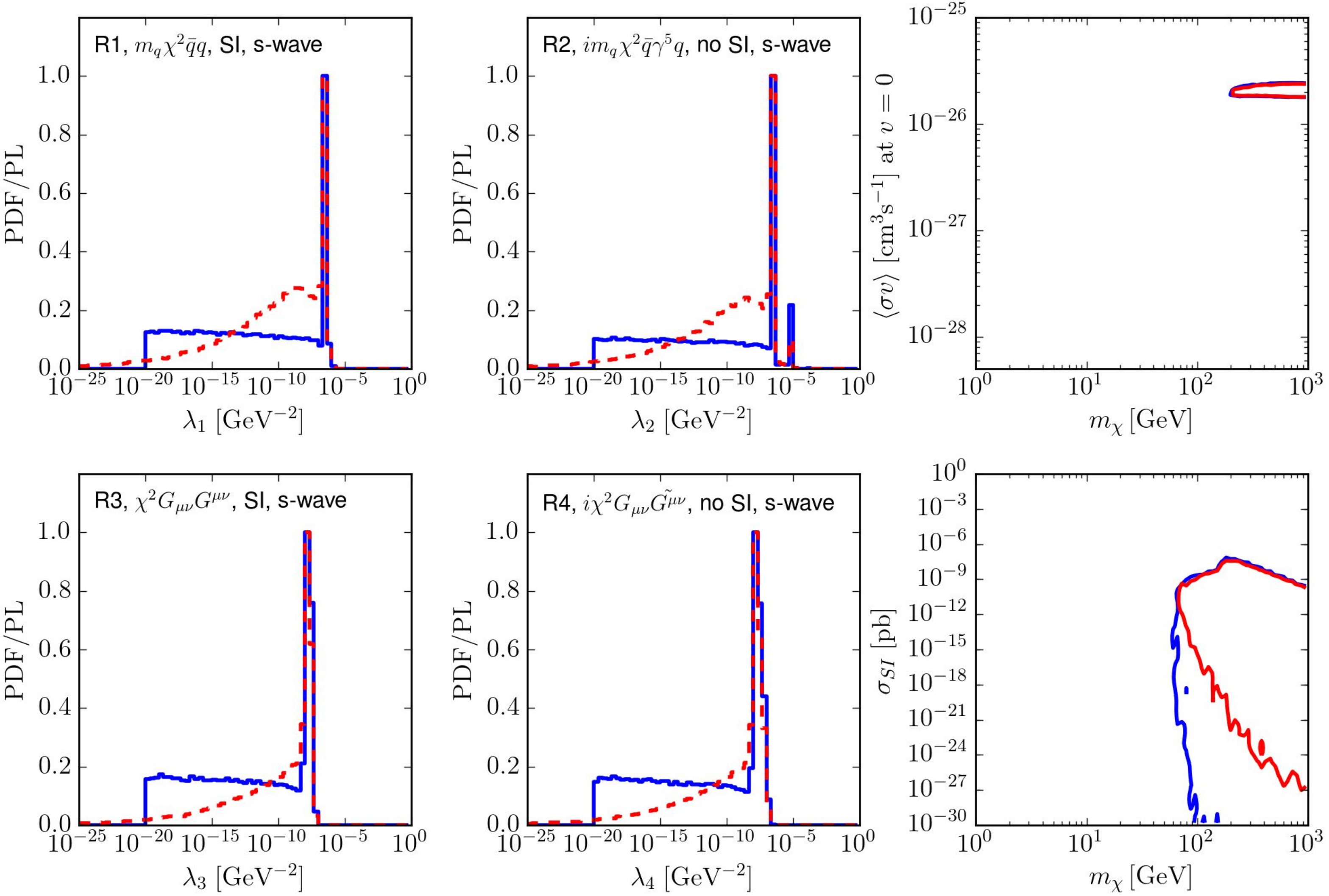
$$\log_{10} A \sim \text{Uniform}(-20, 0)$$

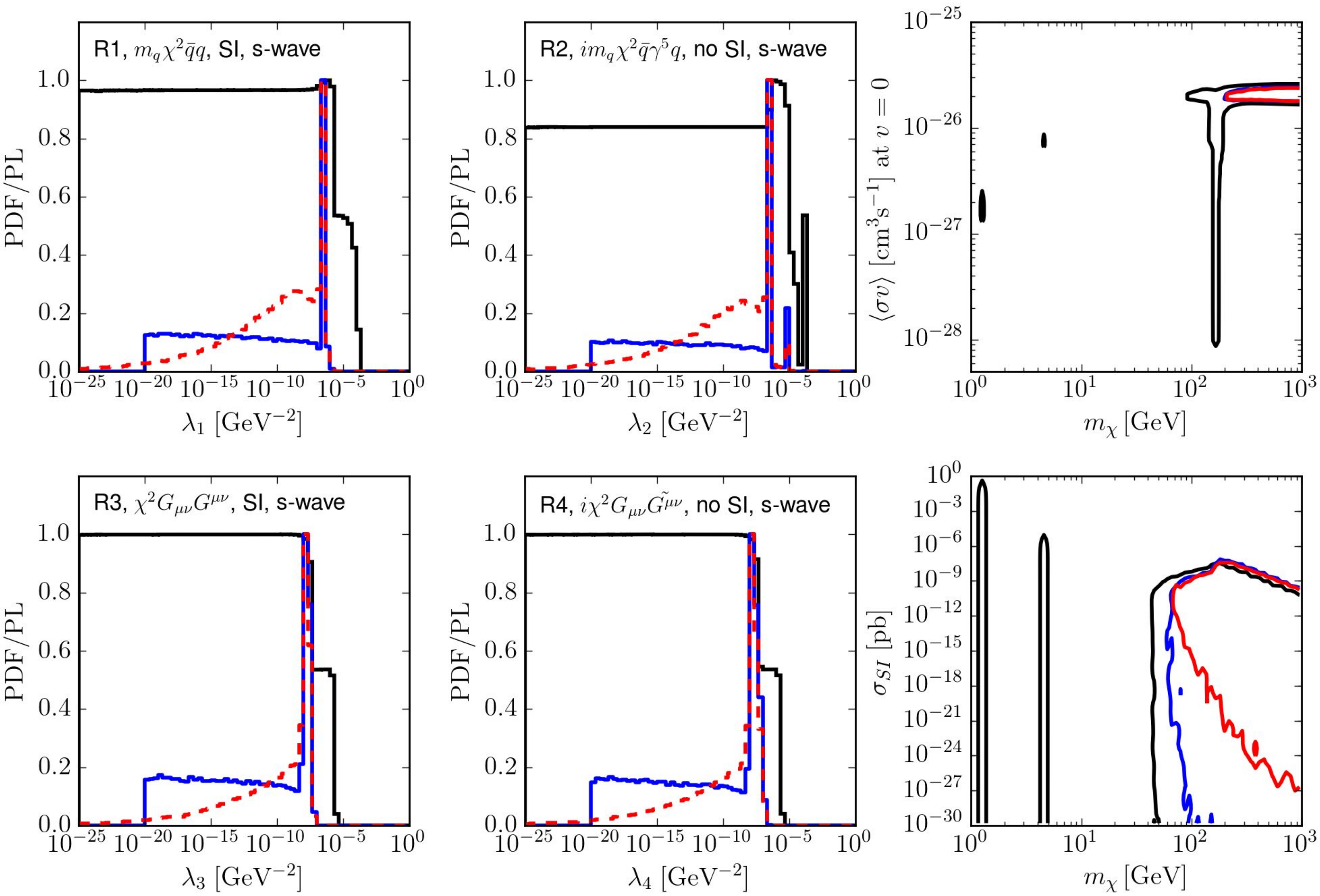
$$\sum f_i = 1$$

$$\bar{f} \sim \text{Dir}(\alpha = 0.1) = \frac{\Gamma(\alpha K)}{\Gamma(K)} \prod_{i=1}^{K} f_i^{\alpha - 1}$$

Two priors

Real scalar DM


R1, $m_q\chi^2\bar{q}q$, SI, s-wave
R2, $im_q\chi^2\bar{q}\gamma^5 q$, no SI, s-wave
R3, $\chi^2 G_{\mu\nu}G^{\mu\nu}$, SI, s-wave
R4, $i\chi^2 G_{\mu\nu}\tilde{G^{\mu\nu}}$, no SI, s-wave
PDF/PL
λ_1 [GeV^{-2}]
λ_2 [GeV^{-2}]
λ_3 [GeV^{-2}]
λ_4 [GeV^{-2}]
$\langle\sigma v\rangle$ [cm^{3}s^{-1}] at $v=0$
σ_{SI} [pb]
m_χ [GeV]

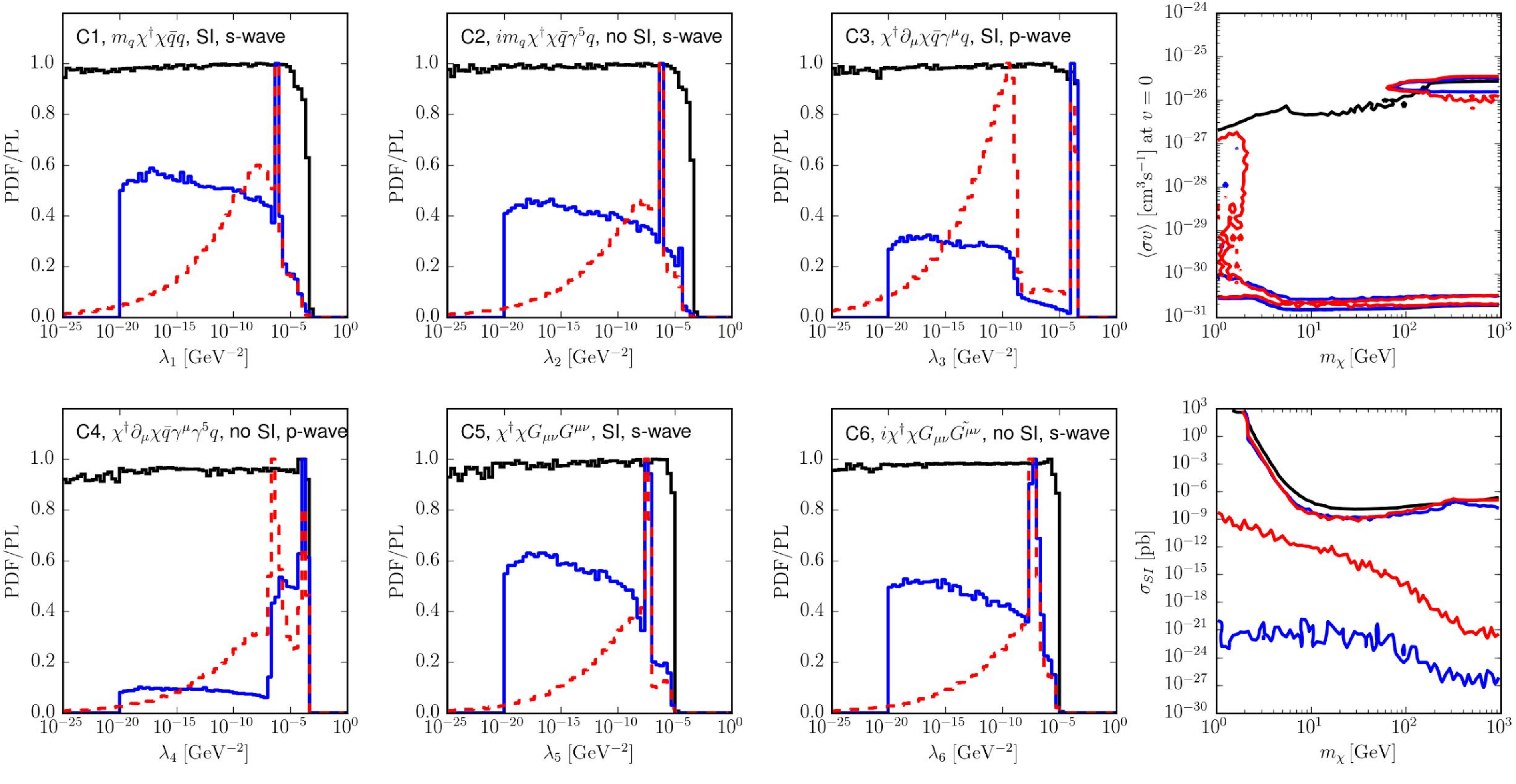


R1, $m_q\chi^2\bar{q}q$, SI, s-wave
R2, $im_q\chi^2\bar{q}\gamma^5 q$, no SI, s-wave
R3, $\chi^2 G_{\mu\nu}G^{\mu\nu}$, SI, s-wave
R4, $i\chi^2 G_{\mu\nu}\tilde{G^{\mu\nu}}$, no SI, s-wave
PDF/PL
λ_1 [GeV^{-2}]
λ_2 [GeV^{-2}]
λ_3 [GeV^{-2}]
λ_4 [GeV^{-2}]
$\langle\sigma v\rangle$ [cm^{3}s^{-1}] at $v=0$
σ_{SI} [pb]
m_χ [GeV]

Complex scalar DM

P-wave operators!

Real scalar DM operators				
Label	Coefficient	Operator	σ_{SI}	$\langle \sigma_{\text{ann}} v \rangle$
R1	$\lambda_1 \sim \frac{1}{2M^2}$	$m_q \chi^2 \bar{q} q$	✓	s-wave
R2	$\lambda_2 \sim \frac{1}{2M^2}$	$i m_q \chi^2 \bar{q} \gamma^5 q$		s-wave
R3	$\lambda_3 \sim \frac{\alpha_s}{4M^2}$	$\chi^2 G_{\mu\nu} G^{\mu\nu}$	✓	s-wave
R4	$\lambda_4 \sim \frac{\alpha_s}{4M^2}$	$i \chi^2 G_{\mu\nu} \tilde{G}^{\mu\nu}$		s-wave
Complex scalar DM operators				
Label	Coefficient	Operator	σ_{SI}	$\langle \sigma_{\text{ann}} v \rangle$
C1	$\lambda_1 \sim \frac{1}{M^2}$	$m_q \chi^\dagger \chi \bar{q} q$	✓	s-wave
C2	$\lambda_2 \sim \frac{1}{M^2}$	$i m_q \chi^\dagger \chi \bar{q} \gamma^5 q$		s-wave
C3	$\lambda_3 \sim \frac{1}{M^2}$	$\chi^\dagger \partial_\mu \chi \bar{q} \gamma^\mu q$	✓	p-wave
C4	$\lambda_4 \sim \frac{1}{M^2}$	$\chi^\dagger \partial_\mu \chi \bar{q} \gamma^\mu \gamma^5 q$		p-wave
C5	$\lambda_5 \sim \frac{\alpha_s}{8M^2}$	$\chi^\dagger \chi G_{\mu\nu} G^{\mu\nu}$	✓	s-wave
C6	$\lambda_6 \sim \frac{\alpha_s}{8M^2}$	$i \chi^\dagger \chi G_{\mu\nu} \tilde{G}^{\mu\nu}$		s-wave

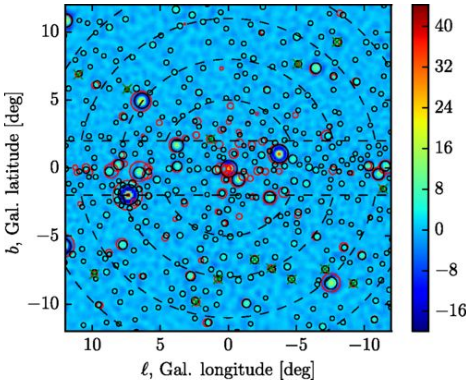


C1, $m_q\chi^\dagger\chi\bar{q}q$, SI, s-wave
C2, $im_q\chi^\dagger\chi\bar{q}\gamma^5 q$, no SI, s-wave
C3, $\chi^\dagger\partial_\mu\chi\bar{q}\gamma^\mu q$, SI, p-wave
C4, $\chi^\dagger\partial_\mu\chi\bar{q}\gamma^\mu\gamma^5 q$, no SI, p-wave
C5, $\chi^\dagger\chi G_{\mu\nu}G^{\mu\nu}$, SI, s-wave
C6, $i\chi^\dagger\chi G_{\mu\nu}\tilde{G^{\mu\nu}}$, no SI, s-wave
PDF/PL
λ_1 [GeV^{-2}]
λ_2 [GeV^{-2}]
λ_3 [GeV^{-2}]
λ_4 [GeV^{-2}]
λ_5 [GeV^{-2}]
λ_6 [GeV^{-2}]
$\langle\sigma v\rangle$ [cm^{3}s^{-1}] at $v=0$
σ_{SI} [pb]
m_χ [GeV]

Galactic centre excess (Galactic Bulge Emission)

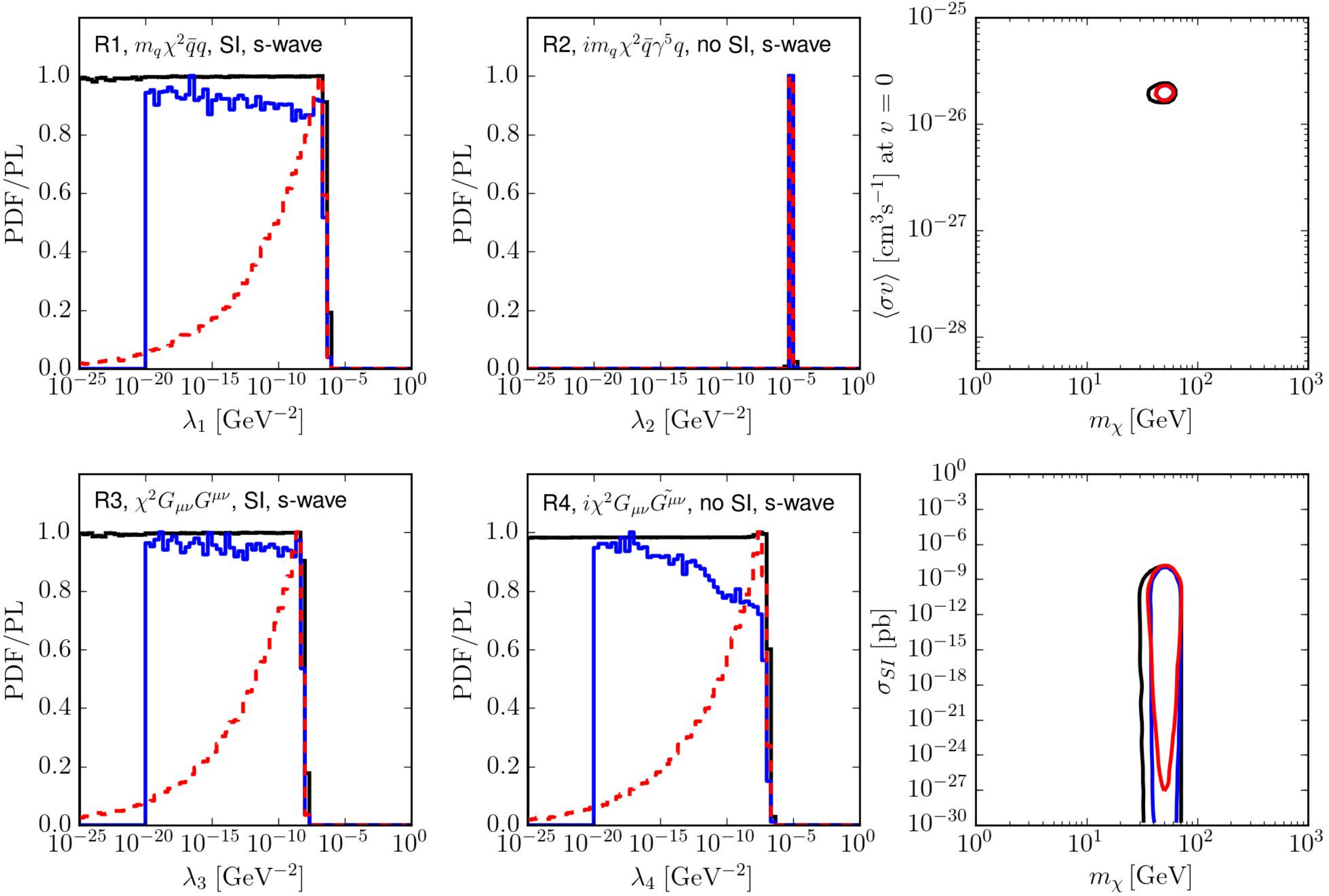
Null results is not (interesting) enough.

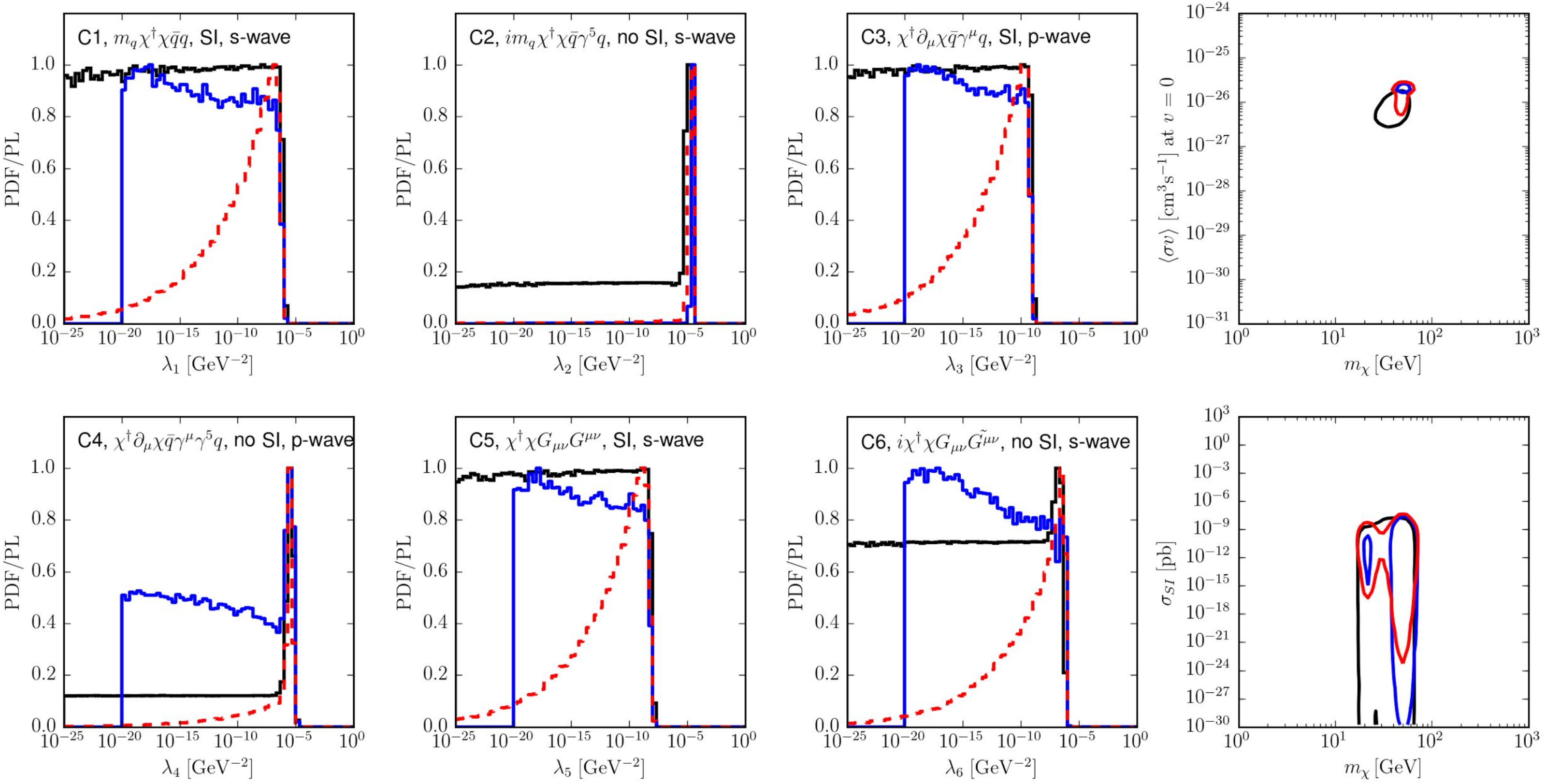
Additional measurements required to identify operators.



Likelihood from Calore et al. arxiv:1409.0042

Bartels et al.
Phys. Rev. Lett. 116, 051102


R1, $m_q\chi^2\bar{q}q$, SI, s-wave
PDF/PL
λ_1 [GeV^{-2}]
R2, $im_q\chi^2\bar{q}\gamma^5 q$, no SI, s-wave
PDF/PL
λ_2 [GeV^{-2}]
$\langle\sigma v\rangle$ [cm^{3}s^{-1}] at $v=0$
m_χ [GeV]
R3, $\chi^2 G_{\mu\nu}G^{\mu\nu}$, SI, s-wave
PDF/PL
λ_3 [GeV^{-2}]
R4, $i\chi^2 G_{\mu\nu}\tilde{G}^{\mu\nu}$, no SI, s-wave
PDF/PL
λ_4 [GeV^{-2}]
σ_{SI} [pb]
m_χ [GeV]



C1, $m_q\chi^\dagger\chi\bar{q}q$, SI, s-wave
C2, $im_q\chi^\dagger\chi\bar{q}\gamma^5 q$, no SI, s-wave
C3, $\chi^\dagger\partial_\mu\chi\bar{q}\gamma^\mu q$, SI, p-wave
C4, $\chi^\dagger\partial_\mu\chi\bar{q}\gamma^\mu\gamma^5 q$, no SI, p-wave
C5, $\chi^\dagger\chi G_{\mu\nu}G^{\mu\nu}$, SI, s-wave
C6, $i\chi^\dagger\chi G_{\mu\nu}\tilde{G^{\mu\nu}}$, no SI, s-wave
PDF/PL
λ_1 [GeV^{-2}]
λ_2 [GeV^{-2}]
λ_3 [GeV^{-2}]
λ_4 [GeV^{-2}]
λ_5 [GeV^{-2}]
λ_6 [GeV^{-2}]
$\langle\sigma v\rangle$ [cm^{3}s^{-1}] at $v=0$
σ_{SI} [pb]
m_χ [GeV]

So far

Simplest possible combined EFT approach using cosmology, direct, and indirect detection.

Measurements are key for minimal theoretical assumption approaches to be interesting in global analysis context.

Alright, cool.
But what about the LHC?

The plan was to include the monojet search.

But the debate on the validity of DM EFT at LHC happened which lead to DM simplified models.

Possible workaround was suggested: Make LHC low-energy experiment by introducing a Mcut = gM and demanding that all events have Ecm < Mcut.

Decent but not satisfying. We want to be all-encompassing.

Considering the cost of simulation and we deemed it not worth it.

Cost is a limitation

Profile likelihood maps of a 15-dimensional MSSM

C. Strege [1], G. Bertone [2], G.J. Besjes [3,4], S. Caron [3,4], R. Ruiz de Austri [5], A. Strubig [3,4], R. Trotta [1]

120M points
400 CPU-years required

Effective field theory of dark matter: a global analysis

Sebastian Liem,[a] Gianfranco Bertone,[a] Francesca Calore,[a] Roberto Ruiz de Austri,[b] Tim M.P. Tait,[c] Roberto Trotta[d,e] and Christoph Weniger[a]

4.2M points

Dark matter interpretations of ATLAS searches for the electroweak production of supersymmetric particles in $\sqrt{s} = 8$ TeV proton–proton collisions

The ATLAS Collaboration

500k points

LHC cannot currently be treated as yet another observation in global fits. We need to be clever.

Part II
Fast LHC Signal Prediction using Machine Learning

Collaborators

Physics:

Gianfranco Bertone

Roberto Ruiz de Austri

Jong Soo Kim

Machine Learning:

Marc Peter Deisenroth

Max Welling

(iDark project - Sascha Caron, Tom Heskes, etc. visualisation, scanning, ML techniques, + actual computer scientists)

Typical setup

Repeated for every point in your sample!

Model parameters

Event generator

Detector simulation

Cross section

Analysis cuts

$$N_s = L\sigma\epsilon$$

$$\mathcal{L}(N_{\mathrm{obs}}|N_s + N_b)$$

Do we really have to do the calculation from scratch for all points in the sample?

Not if we assume that points in parameter space that are close to each other have similar predictions. (We do.)

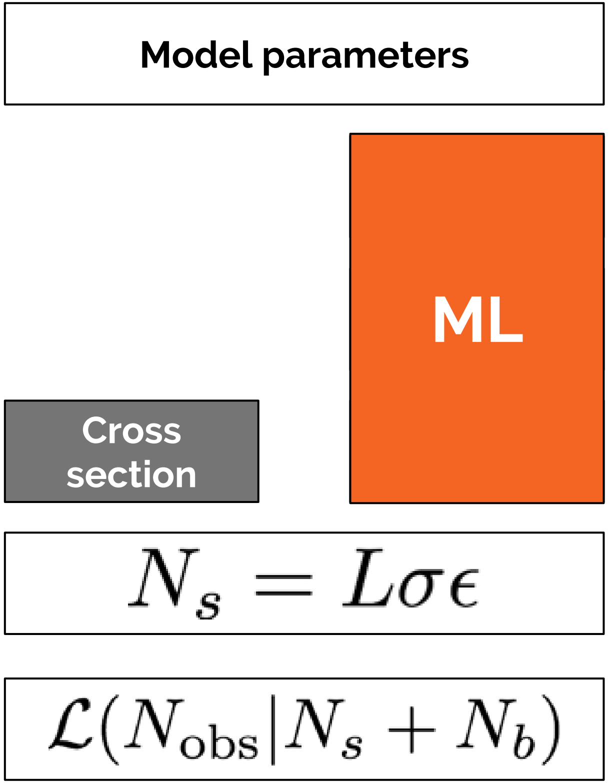
Iffy for cross-sections, probable for SR efficiencies.

Surrogate function

Replace the expensive MC calculation with a **surrogate function.**

$$\epsilon = f(\theta) + \xi$$

In ML language this a supervised regression problem. Many algorithms, we will use Gaussian processes.



Gaussian Processes

Non-parametric: No need to assume a functional form.
Probabilistic: Produces posteriors, i.e. estimates its own error.
Bayesian: Needs a prior on the type of functions. Kernel/covariance function.

$$k_y(x_p, x_q) = \sigma_f^2 \exp\left(-\frac{1}{2\ell^2}(x_p - x_q)^2\right) + \sigma_n^2 \delta_{pq}.$$

Picking hyperparameters is the *learning* task in Gaussian processes.

To deal with large training dataset, we use Distributed Gaussian Processes Deisenroth & Ng, arXiv:1502.02843

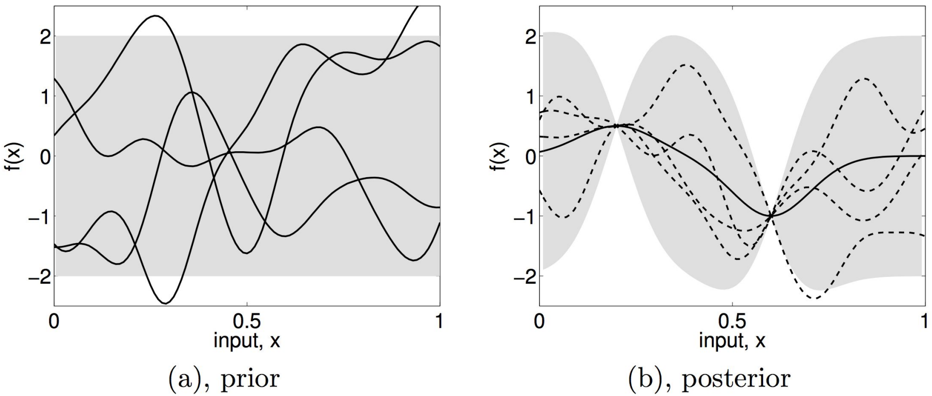


Figure 1.1: Panel (a) shows four samples drawn from the prior distribution. Panel (b) shows the situation after two datapoints have been observed. The mean prediction is shown as the solid line and four samples from the posterior are shown as dashed lines. In both plots the shaded region denotes twice the standard deviation at each input value x.

"Gaussian Processes for Machine Learning"
Rasmussen & Williams, 2006
www.gaussianprocess.org

[Online demo, http://www.tmpl.fi/gp/](http://www.tmpl.fi/gp/)

THE GOAL:
Predict SR efficiencies from theory parameters.

INPUT:
Natural SUSY

Natural because it stabilizes the electroweak scale without fine-tuning.

Only few SUSY states needs to be light.

$$\theta = \{\tan\beta,\ \mu,\ M_3,\ m_{\tilde{Q}_t},\ m_{\tilde{t}_R},\ A_t\}$$

Low-dimensional yet realistic theory.

We already had the training data from previous paper.

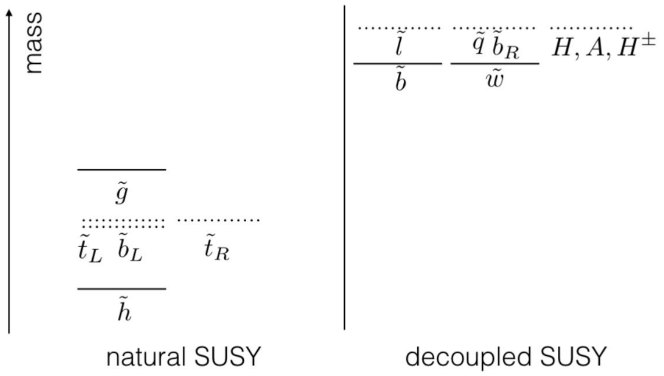
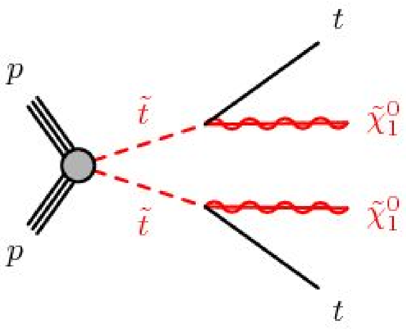


FIG. 1. The minimal natural SUSY mass spectrum on the left while the remaining supersymmetric particles are decoupled on the right.

“Natural SUSY: Now or Never?” Kim et al. arXiv:1606.06738

OUTPUT: Two Signal Regions



Defined in ATLAS-PHYS-PUB-2013-011

Looking direct stop production with HL-LHC, 14 TeV with 3000 fb^{-1}

Stops decay typically to top or b quarks, W/Z or Higgs bosons, and a LSP.
Multiple jets, b-jets, large MET, possibly leptons.

1-lepton

MET > 750 GeV
m_T(lepton, MET) > 550 GeV

Total bkg: 21.1 ± 5.9

0-lepton

MET > 800 GeV
m_T(b-jet, MET) > 400 GeV

Total bkg: 12.2 ± 3.9

Some Machine Learning Lingo

We have a dataset with inputs and outputs. Which is randomly split into

TRAINING DATA which is used to inform the Gaussian process about the functional relationship between our theory parameter and SR efficiency.

TESTING DATA which is **not** use to train the Gaussian process. We use predict the output from the inputs of this data and compare with the true value as defined by the MC.

Training the Gaussian Processes

18647 models split into 16647 models for training and 2000 models for testing.

O(10 min) to train per signal region.

The lunch is not free, just cheaper!

SPheno, Pythia, NLLFAST, CheckMATE, Delphes etc. still needed to generate training data.

Models uniformly sampled from these ranges:

$$\begin{aligned} 0.1\,\text{TeV} \leq \;|\mu|\; &\leq 1.0\,\text{TeV}, \\ 0.1\,\text{TeV} \leq m_{\tilde{Q}_t} &\leq 2.0\,\text{TeV}, \\ 0.1\,\text{TeV} \leq m_{\tilde{t}_R} &\leq 2.0\,\text{TeV}, \\ 0.1\,\text{TeV} \leq |M_3| &\leq 3.0\,\text{TeV}, \\ |A_t| &\leq 3.0\,\text{TeV}, \\ 1 \leq \tan\beta &\leq 20. \end{aligned}$$

All models avoid LEP-II chargino limit, and have reasonable Higgs boson mass.

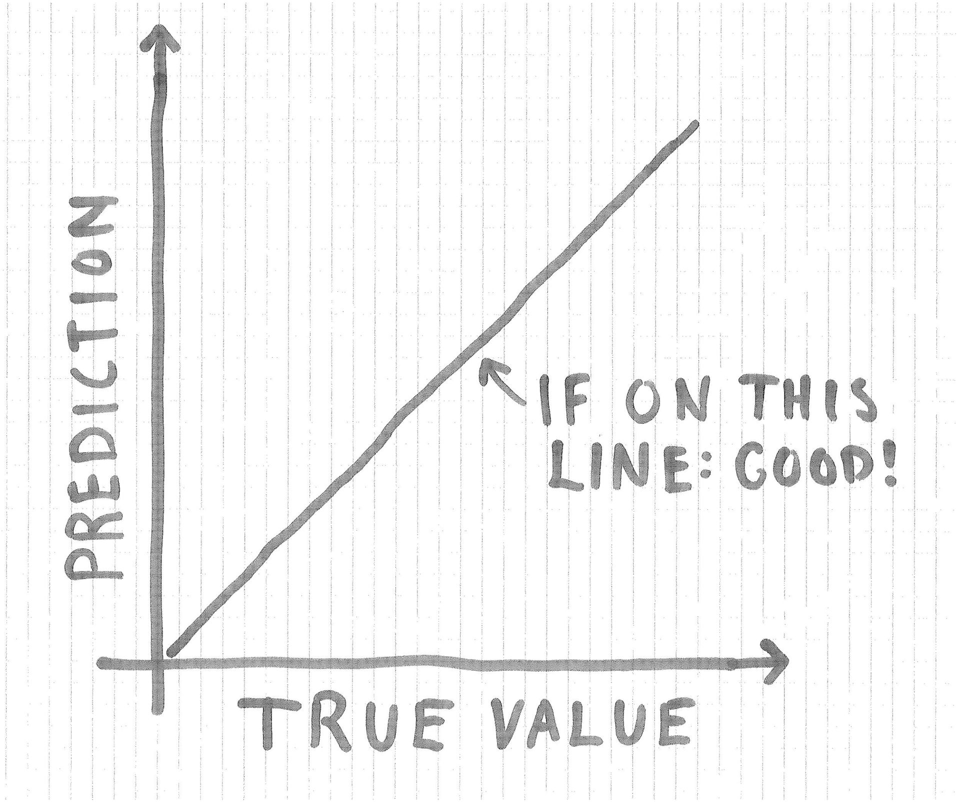
Is it fast?

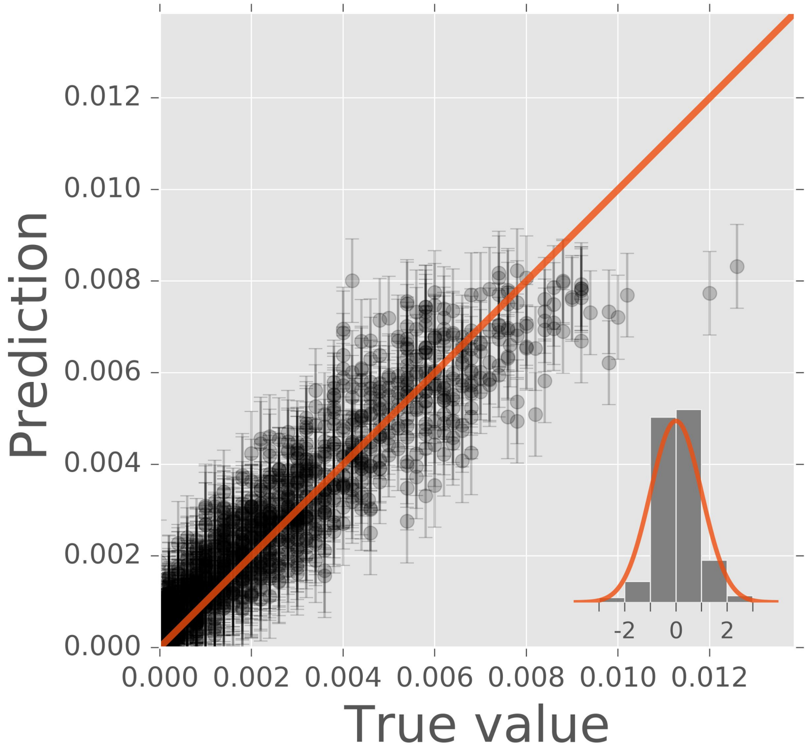
0.06 sec/pred

4 GHz single core, unoptimized Python

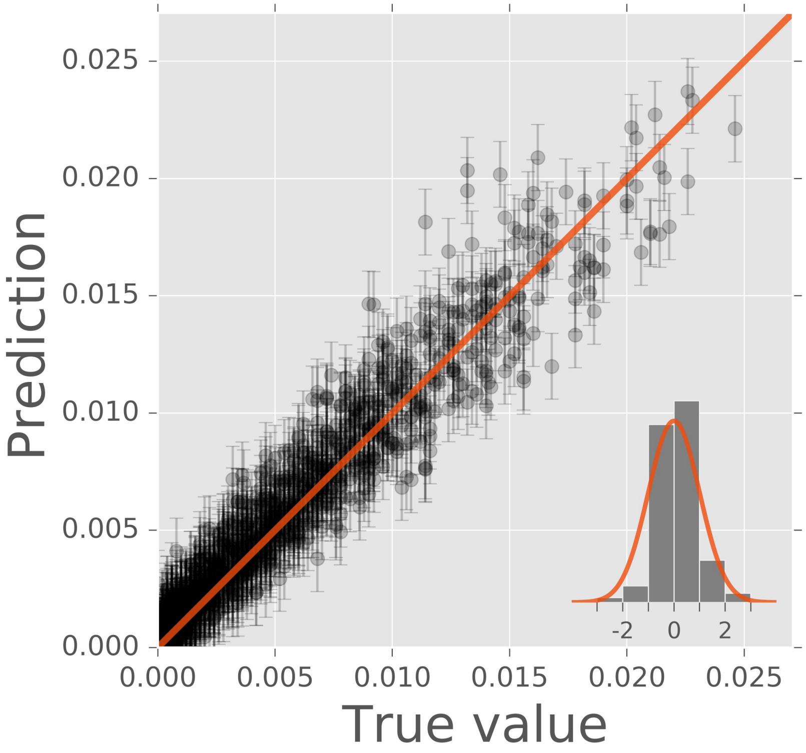
Is it right?

How to evaluate the trained GPs?



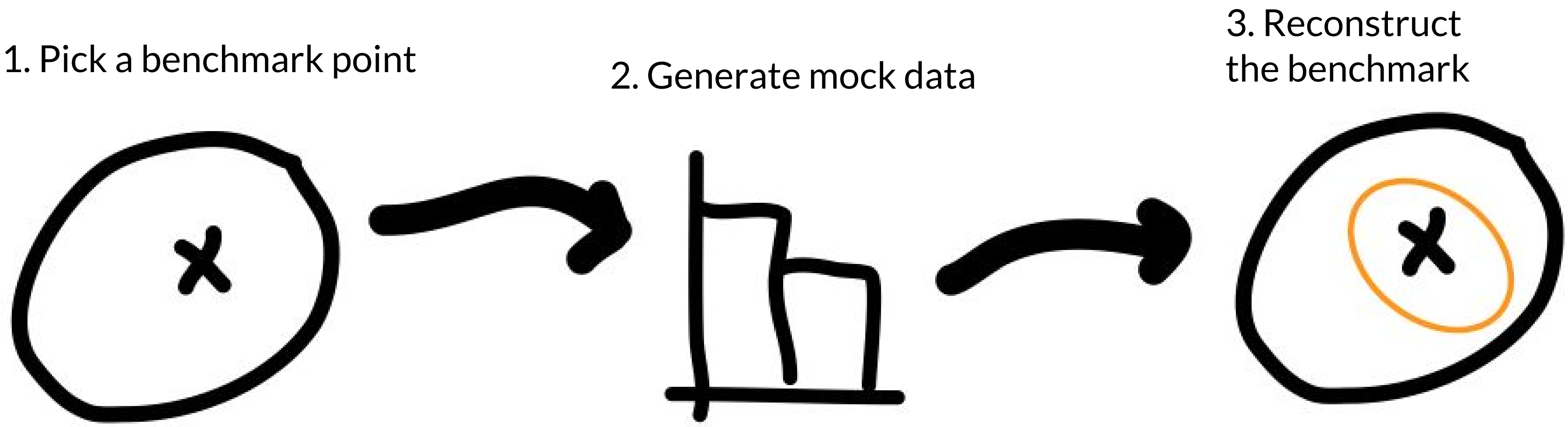


0-lepton signal region



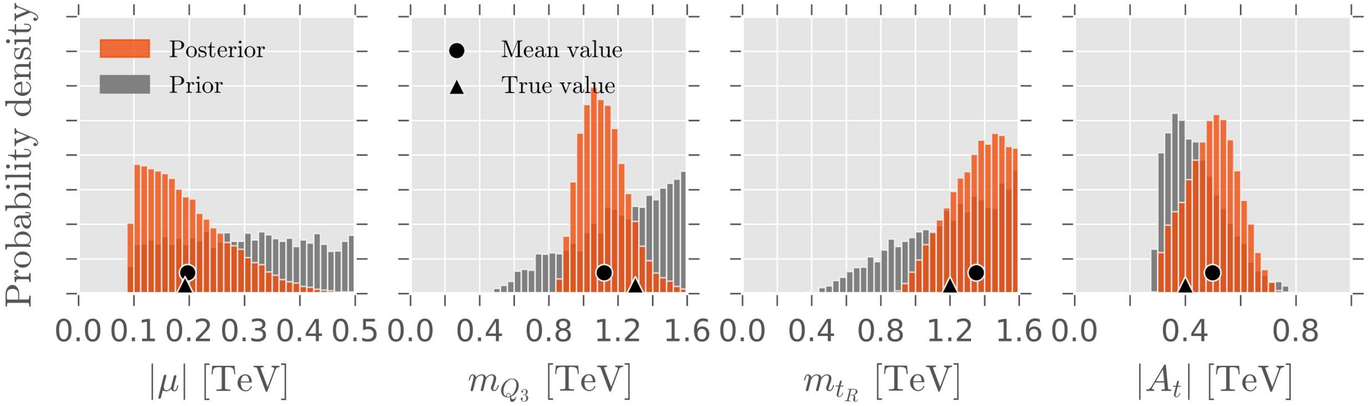
1-lepton signal region

Proof-of-concept: Reconstruction



Schematic view of our reconstruction exercise

PRELIMINARY



No MC performed! No event gen. No detector sim.
~154k likelihood eval
~18h on 10 CPUs

Reconstruction, 4D scan

Outlook

Active learning

Multi-output prediction

Extend to MSSM and other theories

More SRs

Cross-sections

Robust public code library

To conclude...

Using LHC data is a bottleneck.

Gaussian processes makes it really really fast.

Bypasses event generation, detector simulation.

The speed enables a fast reconstruction of theory parameters, and eventually dark matter properties.

The method is completely generic, with many possible applications.

Backup slides

Proof-of-concept Reconstruction

Assume one particular model is true, and measured at the LHC.

Try to reconstruct the parameters using our two signal regions.

Reasonable Higgs mass, LEP chargino limit.

Benchmark model has

63.2 events in 0-lepton SR

161.8 events in 1-lepton SR

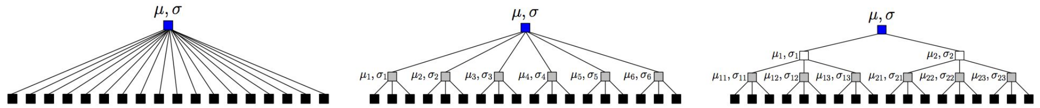
(Bkg: 12.2 ± 3.9 and 21.1 ± 5.9)

—

Distributed Gaussian Processes

The standard Gaussian process scales badly with ***N*** the size of the training dataset. It involves inverting ***N***x***N*** matrices.

We use **distributed** Gaussian processes to avoid this. The data is randomly partitioned and on each partition a Gaussian process is defined. Predictions from each process is then combined.



(a) 1-layer model. (b) 2-layer model. (c) 3-layer model.

Figure 1. Computational graphs of hierarchical PoE models. Main computations are at the leaf nodes (GP experts, black). All other nodes recombine computations from their direct children. The top node (blue) computes the overall prediction.

“Distributed Gaussian Processes”
Deisenroth & Ng, arXiv:1502.02843

—

Benchmark

$$\mu = -192.5$$
$$M_3 = 2100$$
$$m_{\tilde{Q}_3} = 1300$$
$$m_{\tilde{t}_R} = 1200$$
$$\tan\beta = 8.43$$
$$A_t = 2000$$