Dark matter at colliders

10th IDPASC School

Benjamin Fuks, Luca Panizzi

13/09/2021

This exercise will guide you through a simplified version of a phenomenological analysis involving the analysis of a final state compatible with production of dark matter candidates at the LHC.

The main steps of the exercise are:

1. simulate the events corresponding to different signal hypotheses;
2. identify and simulate the most relevant SM background;
3. compare signal and background to obtain exclusion and discovery significances at different luminosities;
4. extract the main features of the signal, *i.e.* how the cross-section and kinematical distributions depend on the parameters of the model;
5. apply kinematical cuts to optimise the reach of the analysis.

Disclaimer: in this exercise we are only simulating the hard scattering events at reconstruction level, which means that we will include parton shower and hadronisation but not, at the level of software, the simulation of the detector. The reason is twofold: on one hand detector simulation with DELPHES requires to install ROOT, which might be a long and not so smooth process depending on the system, and on the other because the time for the exercise is limited.

We will go through the analysis of a mono-jet processes in the first hour and in the second hour you will be asked to repeat the steps for a 2j+MET process and evaluate if it is more or less relevant for determining the exclusion and discovery reaches of the considered scenario at the LHC. If simulations take long and/or issue error messages, the relevant files for each dataset discussed in the text and for those of the 2j+MET process are available in the indico page.

Contents

1 Introduction: the DM model

We will consider a simplified model where dark matter (DM) candidates interact with the Standard Model (SM) particles via a new particle, usually called **mediator**.

The new states share one key property: **both mediator and DM are odd under a $\mathbb{Z}_2$ symmetry**, whereas **all SM states are even**. The Lagrangian is assumed to be invariant under such $\mathbb{Z}_2$ symmetry. This automatically ensures that every interaction involving new particles must contain an even number of the new states, and, as a consequence, that the lightest new state is stable. A stable new state must be electrically and colour neutral to be compatible with cosmological observations, and therefore, if it is massive, it is a DM candidate.

In this model the mediator carries colour charge and belongs to the fundamental representation of $SU(3)$: as a consequence, the only possible SM particles which can interact with the DM are quarks. We will make one working assumption: **both DM and mediator interact only with the SM right-handed up quark**. This assumption is by no means justified by specific observations, the DM could interact with any other quark. It is just for the scope of the exercise. We will finally assume the spin assignments of new particles: **the DM is a real scalar (S) and the mediator is a fermion ψ.** The corresponding interaction Lagrangian is therefore:

$$\mathcal{L}_{\text{int}} = \left[\lambda_\psi \bar{\psi} u_R S + \text{h.c.}\right] \ , \tag{1}$$

where λ_ψ is the coupling strength. The syntax of the UFO model for particle names and their interaction is given in table 1.

Model name	DM		Mediator		Coupling
	ID	name	ID	name	
`DMSimpt_NLO_v1_2_UFO-F3S_uR`	`51`	`xs`	`5920002`	`yf3u1 yf3u1~`	`lamf3u1x1`

Table 1: UFO model syntax for the scenario of the exercise.

The ID of the particles are unique numbers which are assigned to each particle of the model. For the SM particles (including all hadrons and a few hypothetical particle) these ID numbers have been assigned by the Particle Data Group (PDG) (`http://pdg.lbl.gov/index.html` and are hardcoded in each software for simulations. For new particles one needs to use numbers which do not already exist. The list of assigned PDG codes is available at `http://pdg.lbl.gov/2019/reviews/rpp2019-rev-monte-carlo-numbering.pdf`.

The model can be imported in MG5_aMC via the command `import DMSimpt_NLO_v1_2_UFO-F3S_uR`.

2 Generation of topologies

Disclaimer: we will be working at **leading order** in both QCD and EW, which means we will not consider any loop topology. This is potentially a severe approximation: we are working with a model containing coloured new states, and QCD corrections can strongly affect the results. However, the analysis strategies would not change, and therefore for the sake of the exercise we will ignore loop corrections.

Since we are excluding loop corrections, any signal of new physics involving a DM candidate **must** contain the DM candidate in the final state (Why?). In our context, we have three possible production:

1. $pp \to SS$;
2. $pp \to \psi S \to u + SS$ and $pp \to \bar{\psi} S \to \bar{u} + SS$;
3. $pp \to \psi\bar{\psi} \to u\bar{u} + SS$.

The first case would lead to a completely invisible state, but jets are emitted by the initial state and by the mediator, producing and unbalance in the transverse energy measured in the detector. Representative topologies corresponding to the three processes are shown in table 2.

The final state for the first two topologies is, in both cases, one jet and missing transverse energy (MET), but the kinematics will be completely different, due to the different topologies at the origin of it. Furthermore, part of the topologies interfere (look at the first and third Feynman diagrams in table 2) complicating the picture. These consideration play a very relevant role when considering QCD corrections and justify treating the three processes separately (with some care, see Benjamin's slides), but *since we are working at leading order*, we can

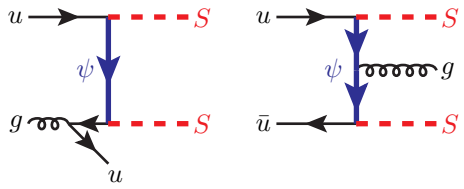
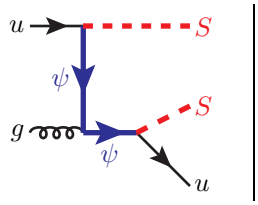
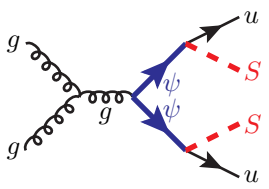
$pp \to SS + j$	$pp \to \psi S \to u + SS$ $pp \to \bar{\psi} S \to \bar{u} + SS$	$pp \to \psi\bar{\psi} \to u\bar{u} + SS$
		

Table 2: Topologies for DM production at LHC with real scalar DM and fermionic mediator.

simply combine the mono-jet topologies and study them together. The topology corresponding to di-jet plus missing transverse energy can be studied separately.

The syntax to simulate the mono-jet and di-jet + MET processes and output the simulation files in different folders is the following:

1j+MET

```
generate p p > xs xs j
output ./DMcollider/F3S_1j
```

2j+MET

```
define yy = yf3u1 yf3u1~
define uu = u u~
generate p p > yy yy, yy > xs uu
output DMcollider/F3S_2j
```

3 Setting the signal parameters

All the parameters of the model can be set in a file called `param_card.dat` inside the folder `Cards`. In this file the couplings and the masses and decay width (in GeV) of all particles of the model are stored.

We don't know the masses of the DM candidate and of the mediator, so in a complete analysis we should perform a scan over different values. We also do not know the strength of the coupling between DM, mediator and SM states, and again a scan should be performed. Here the coupling is fixed to a specific value, $\lambda_\psi = 2$. In this exercise, therefore, we limit ourselves to four representative benchmark points (BPs), summarised in table 3.

	λ_ψ	m_ψ (GeV)	m_S (GeV)	description
BP1	2	500	100	light mediator, large mass gap
BP2	2	500	450	light mediator, small mass gap
BP3	2	1000	100	heavy mediator, large mass gap
BP4	2	1000	950	heavy mediator, small mass gap

Table 3: Signal benchmark points.

A fixed coupling means that, depending on the mass configurations, the *decay width* of the mediator, defined as:[1]

$$\Gamma_\psi = \frac{\lambda_\psi^2}{32\pi m_\psi^3}(m_\psi^2 - m_S^2)^2 \tag{2}$$

is not the same for each benchmark. We are of course assuming here that ψ does not have any further decay channel, as we are working with a simplified model. The Γ_ψ/m_ψ ratio of the four BPs is given in table 4. All our

	BP1	BP2	BP3	BP4
Γ_ψ/m_ψ (%)	3.7	0.14	3.9	0.04

Table 4: Ratios between width and mass of the mediator for the BPs of the exercise.

benchmarks correspond to mediators with a small width with respect to their mass. This check is potentially very important in case the simulation requires the factorisation of resonant production and decay of the mediator (such in the case of the 2j+MET process), which is consistent only in the narrow-width approximation (NWA).

[1] The analytical expressions of decay widths for each particle of the UFO model can be found in the file `decays.py`.

In our case we are allowing the mediator to propagate either on-shell or off-shell, so our results will be consistent regardless of its width. Widths can also be computed numerically in MG5_aMC: in the `madevent` environment it is enough to type `compute_widths yf3u1` and set the masses and couplings in the `param_card.dat` file inside the folder `Cards`; this file will be updated and at the end of it all the decay widths of the particle are shown, together with the BR. For example, for BP3 the end of the file will show the following lines:

```
#      PDG        Width
DECAY  5920002   3.899694e+01
#  BR             NDA  ID1    ID2   ...
   1.000000e+00   2    51  2 # 38.99693993109166
```

4 Generation of signal events

Navigate to the `DMcollider/F3S_2j` folder and start the simulation environment by typing either `./bin/madevent` or `python2.X ./bin/madevent` or `python3.X ./bin/madevent` with `X>6` depending on your system.

We will be simulating the four BPs one at a time. There are also automatic ways to perform scans in MG5_aMC, and one can write customised scripts to do the job. Since in our case the number of signal points is limited, we will not exploit these features.

In the `madevent` environment launch the simulation for BP1:

```
panizzi@panizzi-Precision-3530:~/Research/MadGraph/MG5_aMC_v3_1_0/DMcollider/F3S_1j$ python3.7 ./bin/madevent
launch in debug mode
************************************************************
*                                                          *
*                    W E L C O M E to                      *
*           M A D G R A P H 5 _ a M C @ N L O              *
*                    M A D E V E N T                       *
*                                                          *
*                *                       *                 *
*                  *       * *       *                   *
*                    * * * * 5 * * * *                     *
*                  *       * *       *                   *
*                *                       *                 *
*                                                          *
*         VERSION 3.1.0                20xx-xx-xx          *
*                                                          *
*    The MadGraph5_aMC@NLO Development Team - Find us at   *
*    https://server06.fynu.ucl.ac.be/projects/madgraph     *
*                                                          *
*              Type 'help' for in-line help.               *
*                                                          *
************************************************************
INFO: load configuration from /home/panizzi/Research/MadGraph/MG5_aMC_v3_1_0/DMcollider/F3S_1j/Cards/me5_configuration.txt
INFO: load configuration from /home/panizzi/Research/MadGraph/MG5_aMC_v3_1_0/input/mg5_configuration.txt
INFO: load configuration from /home/panizzi/Research/MadGraph/MG5_aMC_v3_1_0/DMcollider/F3S_1j/Cards/me5_configuration.txt
Using default text editor "vi". Set another one in ./input/mg5_configuration.txt
Using default eps viewer "gv". Set another one in ./input/mg5_configuration.txt
Using default web browser "firefox". Set another one in ./input/mg5_configuration.txt
F3S_1j>launch monojet_signal_BP1
```

and switch on Pythia8.2 [1] and MadAnalysis 5 [2,3]:

```
/================== Description ==================|================= values ==================|===== other options ======\
| 1. Choose the shower/hadronization program       |        shower = Pythia8                   |     OFF                  |
| 2. Choose the detector simulation program        |      detector = OFF                       |     Delphes              |
| 3. Choose an analysis package (plot/convert)     |      analysis = MadAnalysis5              |     OFF|MadAnalysis4     |
| 4. Decay onshell particles                       |       madspin = OFF                       |     ON|onshell|full     |
| 5. Add weights to events for new hypp.           |      reweight = OFF                       |     ON                   |
\=================================================|===========================================|==========================/
```

At this point we can choose how to set the parameters of the model and of the simulation:

```
Do you want to edit a card (press enter to bypass editing)?
/-----------------------------------------------------------\
|  1. param              : param_card.dat                   |
|  2. run                : run_card.dat                     |
|  3. pythia8            : pythia8_card.dat                 |
|  4. madanalysis5_parton : madanalysis5_parton_card.dat    |
|  5. madanalysis5_hadron : madanalysis5_hadron_card.dat    |
\-----------------------------------------------------------/
 you can also
   - enter the path to a valid card or banner.
   - use the 'set' command to modify a parameter directly.
     The set option works only for param_card and run_card.
     Type 'help set' for more information on this command.
   - call an external program (ASperGE/MadWidth/...).
     Type 'help' for the list of available command
 [0, done, 1, param, 2, run, 3, pythia8, 4, enter path, ... ][90s to answer]
>
```

As already mentioned, all the model parameters are stored in the `param_card.dat`. All parton-level simulation parameters are stored in the `run_card.dat`, while the parameters for Pythia8.2 and MadAnalysis 5 are stored in the corresponding files, listed above.

We can either work from command line using the `set` followed by the name of the parameter and its value, or type a number of the list and open the corresponding files in a terminal editor. Alternatively one can set the parameters in an editor beforehand and then run the simulation pressing enter at this stage. Here we work via command line and type the following sequence:

```
set mass 5920002 500
set width 5920002 auto
set mass 51 100
set lamf3u1x1 2
set use_syst False
```

At each step MG5_aMC will confirm your choice and allow you to select the next parameter until the setting is finished. Setting the mediator width to `auto` tells MG5_aMC to compute the width depending on the input parameters, instead of fixing it to a given value. The last setting removes the calculation of systematic uncertainties. It is done **exclusively to spare time** in the exercise, but in real anayses it is always crucial to estimate systematics, as they are always the dominant theoretical uncertainty. In fact, statistical uncertainties can be made as small as we need by increasing the number of MC events, but systematics depend on the order of the simulation (going to NLO reduces the uncertainties but it is more challenging to do, see Benjamin's slides) and on the choice of the PDFs, which are determined by different collaborations using different techniques (data driven, neural networks...) with different degrees of accuracy.

We are now all set and the simulation can start by pressing `enter`. The details about how the Monte Carlo (MC) simulation is performed are out of the scope of the exercise; in a nutshell, events are not distributed evenly in the allowed phase space, but are concentrated in the regions which are most relevant to determine the cross-section.

With the default settings, the simulation will be done with 10^4 events, which correspond to a Poissonian statistical uncertainty of 1%. The default LHC energy is 13 TeV, the parton distribution functions (PDFs) are NNPDF2.3 at LO [4] (there are more updated ones, but it is not important here) and the renormalisation and factorisation scales are computed on an event-by-event basis to the transverse energy of the system (see the file `SubProcesses/setscales.f` in the simulation folder and `https://cp3.irmp.ucl.ac.be/projects/madgraph/wiki/FAQ-General-13` for more details. Minimal kinematic cuts are also imposed on the jets: $p_T(j) > 20$ GeV and $|\eta(j)| < 5$.

Hadronisation and parton showering are performed by Pythia8.2, and the subsequent reconstruction of the final state is done by MadAnalysis 5 via FastJet using as default settings the anti-k_T algorithm with $p_T^{\min} = 5$ GeV and cone radius $R = 0.4$ (see the FastJet manual [5] for details about the algorithm and related quantities). All the settings related to the reconstruction can be modified in the `madanalysis5_hadron_card.dat`, which can be edited before running the simulation when MG5_aMC asks to set the simulation parameters.

The simulation will take a few minutes depending on the system...[2]

After it is done, let's repeat the same procedure with the other benchmarks, being careful to give a different name to the run each time to avoid overwriting previous results.[3] **Remember to re-set the mediator width to `auto` at each step, otherwise MG5_aMC will take the value of the width which was set in the previous simulation**.

After all simulations are done, we can exit the `madevent` environment and proceed to the analysis of the results.

5 Reading the signal events

Here a description of how to read the events is provided, both at parton level for the reconstructed final states. This is done in practice by sofwtare tools (MadAnalysis 5 already did it during the simulation), but it is very instructive to be able to understand how these tools extract the information.

[2] MG5_aMC might issue warnings telling that some particles have a too small width; ignore the warnings, as these particles do not participate into the process.

[3] During the tutorial only the simulation for BP1 is performed, while pre-simulated data (available in the indico page) are used for the others, to spare time.

5.1 The parton-level results

In the folder `DMcollider/F3S_1j/Events/monojet_signal_BP1/` (and in the material in the indico page) you will find a file called `unweighted_events.lhe.gz`.[4] It is a compressed file but it can be uncompressed using `gunzip -k unweighted_events.lhe.gz` or `tar -xzvf unweighted_events.lhe.gz`. The resulting `.lhe` file is a text file containing all the events of the simulation. It is rather lengthy, but we only need to analyse the beginning of it.

The standard references for a description of the variables are [6, 7], and we strongly suggest you to refer to these papers for any doubt about the meaning of the numbers in the events.

The header of the file (between the strings `<header>` and `</header>`) is composed of the following parts:

- The first ∼750 lines, until the string `</slha>`, contain information about the parameters of the simulation. Here you will find the syntax you used for the generation of the diagrams, and a copy of the `param_card.dat` and the `run_card.dat`.
- If systematics uncertainties are evaluated, the subsequent lines, until the string `</initrwgt>` contain information about the reweighting and the members of the PDF set we used. These lines will not appear in our simulations, as we switched off the evaluation of systematics.
- Subsequently there are information about the cross-section.

```
<MGGenerationInfo>
#  Number of Events        :       10000
#  Integrated weight (pb)  :       15.071898
</MGGenerationInfo>
```

Information about the collider and the MC events are then displayed.

```
<init>
2212 2212 6.500000e+03 6.500000e+03 0 0 247000 247000 -4 1
1.507190e+01 4.480162e-02 1.507190e+01 1
<generator name='MadGraph5_aMC@NLO' version='3.1.0'>please cite 1405.0301 </generator>
</init>
```

You can notice how the first pair of numbers correspond to the PDG code of the proton, the second pair of numbers corresponds to the beam energy of this simulation, then some information about the PDFs and in the second line you can recognise the cross-section with uncertainty. Notice that the value of the cross-section might be different in your case, as it fluctuates for each simulation. This is a MC simulation and it will depend on a (pseudo-)random number generator, the seed of which can be fixed to obtain a perfect reproducibility, but it is not necessary for this exercise (the corresponding parameter is in the `run_card.dat` for those who are interested). However, the numbers you obtain should be compatible with those above within the uncertainties.

The interesting part for us is the following, which corresponds to the first event (it might contain different numbers in your case). Each simulated event is contained between the strings `<event>` and `</event>`, which repeat 10000 times in the file.

The event displayed below is initiated by a gluon and an up quark, it produces the mediator ψ as intermediate state and a DM candidate S in the final state; the mediator then decays to another DM particle and to an up quark. Let's see how we can get this information.

```
<event>
 6      1 +1.5071898e+01 6.06060900e+02 7.81860800e-03 1.00544000e-01
      21 -1    0    0  501  502 +0.0000000000e+00 +0.0000000000e+00 +9.5081898896e+01 9.5081898896e+01 0.0000000000e+00 0.0000e+00 -1.0000e+00
       2 -1    0    0  502    0 -0.0000000000e+00 -0.0000000000e+00 -2.4741784187e+03 2.4741784187e+03 0.0000000000e+00 0.0000e+00 1.0000e+00
 5920002  2    1    2  501    0 -8.9179143462e+01 +3.3769728572e+02 -1.4403821407e+03 1.5626933492e+03 4.9529528148e+02 0.0000e+00 0.0000e+00
      51  1    3    3    0    0 -2.0507008140e+02 +2.8095628725e+02 -5.2144048081e+02 6.3473644031e+02 1.0000000000e+02 0.0000e+00 0.0000e+00
      51  1    1    2    0    0 +8.9179143462e+01 -3.3769728572e+02 -9.3871437911e+02 1.0065669684e+03 1.0000000000e+02 0.0000e+00 0.0000e+00
       2  1    3    3  501    0 +1.1589093794e+02 +5.6740998475e+01 -9.1894165991e+02 9.2795690887e+02 0.0000000000e+00 0.0000e+00 1.0000e+00
</event>
```

The first line contains information about the number of particles in the event (and indeed there are 6 lines after the first one), a label which is not important for us, the weight of the event in pb (all the same for us, corresponding to setting `event_norm=average` in the `run_card.dat`, such that $\sigma = \sum_i^{\#\text{events}} \text{weight}_i/\#\text{events}$), the factorisation scale of the PDFs in GeV (which also corresponds to the renormalisation scale of α_S), and the running values of α_{em} and α_S (respectively).

All the other lines correspond to each particle of the event, and contain analogous information, therefore describing one of them is enough to understand all the others. In the example above let's take the line corresponding to the up quark in the final state (the last one). The numbers have the following meaning.

[4] LHE stands for Les Houches Events, and it is a standardized form for storing MC events in collider simulations.

2	**PDG number of the particle**
1	**State of the particle**: -1 = initial state, 2 = intermediate resonance, 1 = final state
	Mothers
3	Mother 1: from which particle it comes from (in this case from the third particle of the event, *i.e. the mediator* ψ*).*
3	Mother 2: from which other particle it comes from. In this case it is the same, but if you look at the lines corresponding to the mediator or the second DM, you can see that they are generated by the initial state, as they are not products of the decay of a resonance.
	Colour flow
501	Colour of the particle. The actual numerical values here do not have a meaning.
0	Anti-colour of the particle. Notice that this is a colour triplet, so it has only one colour. The gluons have both numbers.
The important thing is that the colour flow is conserved in the event. In this case the initial state has colours 501 and 502 and anticolour 502, and since 502 contracts away, the net colour and anticolour of the initial state are 501 and 0 respectively. If you look at the final state particles and contract all the indices you will see that the colour is preserved.	
Four-momentum in the lab frame and generated mass of the particle (all quantities in GeV)	
+1.1589093794e+02	p_x
+5.6740998475e+01	p_y
-9.1894165991e+02	p_z
9.2795690887e+02	E
0.0000000000e+00	M such that $M^2 = E^2 - \|\vec{p}\|^2$. The up quark is massless, but this quantity is not zero for the mediator and the DM. Notice that for the mediator, which is an intermediate state, this does not correspond exactly to the pole mass: event by event, this quantity will be smeared around the pole mass according to the widths of the particles.
	Other information
0.0000e+00	invariant lifetime (in mm)
1.0000e+00	spin/helicity

5.2 The reconstructed events

After the inclusion of parton showering and hadronisation via Pythia8.2and a reconstruction of the final state (without including the detector simulation), MadAnalysis 5 produces another file in `lhe` format, where all the events have the very same structure as in the parton-level example above, but where all jets in the final state (excluding b-jets) are represented by the PDG label 21, and all invisible particle (contributing to the missing transverse energy) are associated to the PDG label 12. Here the labels do not correspond to gluon and electron-neutrino respectively, they are just meant to identify final state objects. Information about the colour flow is also not propagated, as it has been already processed by Pythia8.2.

The reconstructed event corresponding to the example in the previous section, which can be found in DMcollider/F3S_1j/Events/monojet_signal_BP1/tag_1_pythia8_BasicReco.lhe.gz, reads:

```
<event>
21 9999  1.5071899E-03 -1.0000000E+00 -1.0000000E+00 -1.0000000E+00
       21   -1    0    0    0    0  0.00000000000E+00  0.00000000000E+00  0.95081901550E+02  0.95081901550E+02  0.00000000000E+00 0.00000 0.00000
        2   -1    0    0    0    0  0.00000000000E+00  0.00000000000E+00 -0.24741784668E+04  0.24741784668E+04  0.00000000000E+00 0.00000 0.00000
  5920002    3    1    2    0    0 -0.89179145813E+02  0.33769729614E+03 -0.14403820801E+04  0.15626933594E+04  0.49529528809E+03 0.00000 0.00000
       51    3    1    2    0    0  0.89179145813E+02 -0.33769729614E+03 -0.93871435547E+03  0.10065669556E+04  1.00000000000E+02 0.00000 0.00000
       51    3    0    0    0    0 -0.19193368530E+03  0.27872113037E+03 -0.50221389771E+03  0.61379376221E+03  1.00000000000E+02 0.00000 0.00000
        2    3    0    0    0    0  0.13515042114E+03  0.53464000702E+02 -0.89075323486E+03  0.90253277588E+03 -0.00002412626E+00 0.00000 0.00000
       21    1    0    0    0    0  0.13002043962E+01  0.56275610924E+01  0.35472080231E+02  0.35986225128E+02  0.18384406567E+01 0.00000 0.00000
       21    1    0    0    0    0 -0.45636730194E+01  0.22123250961E+01 -0.88516677856E+02  0.88671020508E+02  0.12754055262E+01 0.00000 0.00000
       21    1    0    0    0    0  0.62020902634E+01 -0.33388130665E+01  0.25538522720E+02  0.26544168472E+02  0.16622546911E+01 0.00000 0.00000
       21    1    0    0    0    0  0.49047603607E+01  0.39266762733E+01 -0.75958206177E+02  0.76299285889E+02  0.35293636322E+01 0.00000 0.00000
       21    1    0    0    0    0 -0.58412227631E+01 -0.22551760674E+01 -0.12785321045E+03  0.12801364136E+03  0.13579218388E+01 0.00000 0.00000
       21    1    0    0    0    0 -0.10233658791E+02 -0.55511226654E+01  0.96387863159E+01  0.15456365585E+02  0.32326951027E+01 0.00000 0.00000
       21    1    0    0    0    0  0.17238742113E+01 -0.12144783020E+02  0.63934040070E+01  0.14929090500E+02  0.56155705452E+01 0.00000 0.00000
       21    1    0    0    0    0 -0.15499397516E+01  0.79484233856E+01 -0.19885877609E+02  0.21670598984E+02  0.29303524494E+01 0.00000 0.00000
       21    1    0    0    0    0  0.79085421562E+01 -0.55437464714E+01  0.18413026810E+02  0.21196403503E+02  0.41194400787E+01 0.00000 0.00000
       21    1    0    0    0    0 -0.11111829758E+02 -0.16524000168E+02  0.13016361237E+02  0.24907608032E+02  0.73788805008E+01 0.00000 0.00000
       21    1    0    0    0    0 -0.23837696075E+02  0.16388479233E+02  0.30531206131E+02  0.42729549408E+02  0.75393490791E+01 0.00000 0.00000
       21    1    0    0    0    0  0.28916299820E+02 -0.31888859272E+01  0.29265570068E+03  0.29414956665E+03  0.55034775734E+01 0.00000 0.00000
       21    1    0    0    0    0 -0.42811523437E+02  0.41915535927E+01  0.58977008820E+02  0.73747299194E+02  0.10487195969E+02 0.00000 0.00000
       21    1    0    0    0    0  0.13036148071E+03  0.52272178650E+02 -0.84723815918E+03  0.85899810791E+03  0.18404953003E+02 0.00000 0.00000
       12    1    0    0    0    0 -0.84294715881E+02 -0.59450878143E+02  0.00000000000E+00  0.10315040588E+03  0.00000000000E+00 0.00000 0.00000
</event>
```

This is the file we will be using to extract the kinematical information of our signal.

5.3 The cross-sections

The first piece of information we need to perform the analysis is the cross-section of the signal. This will tell us how many signal events we can expect at a given luminosity, and it is crucial to assess if the signal is potentially observable or not.

The cross-sections for the BPs of the exercise are shown in table 5 (as usual, numbers may fluctuate for different simulations, but they should be consistent within uncertainties).

	BP1	BP2	BP3	BP4
m_ψ (GeV)	500	500	1000	1000
m_S (GeV)	100	450	100	950
σ (pb)	15.072 ± 0.045	3.971 ± 0.011	0.5849 ± 0.0019	0.076003 ± 0.00024

Table 5: Cross-sections for the BPs of the exercise. Uncertainties here are **only** statistical to spare time in the simulations. In real analyses, systematics must be included as well.

The numbers can be read in different ways: they are shown (together with many more information about the runs) in the page `crossx.html` which can be accessed with any browser, they appear in the terminal at the end of each simulation, they are reported in the `lhe` file described in section 5.1 or in the banner of the simulation (a file with extension `txt` which can be found in the `Events/<name_of_run>/` folder, which reports the header of the `lhe` file).

The number of signal events can be computed for an integrated luminosity L as:

$$s = L\sigma\epsilon \,, \tag{3}$$

where ϵ represents the efficiency of the analysis, which includes the detector acceptances and any selection or kinematical cut introduced to reduce the SM background.

In the ideal detector scenario (100% of the events are detected), and before any cut, for the nominal luminosity at the end of Run 3 of the LHC, $L = 300$ fb^{-1}, the above cross-sections correspond to $s(\text{BP1}) \simeq 4.5 \times 10^6$, $s(\text{BP2}) \simeq 1.2 \times 10^6$, $s(\text{BP3}) \simeq 1.75 \times 10^5$, $s(\text{BP4}) \simeq 22800$. Are these numbers big or small? To understand it we need to evaluate the number of background events.

6 Generation of the SM background

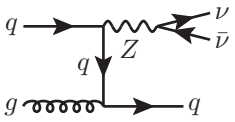
Determining which SM processes constitute a background for our signal is crucial to assess if the signal is observable, or above which luminosity it will become detectable.

Disclaimer: The determination of the background is more often than not a very complex procedure, which is heavily affected by how the detector is sensitive to the kinematical properties of the final state objects we are planning to use for our analysis. For this exercise we will limit ourselves to a basic category of background, the **irreducible background**, which is the set of processes which produce the very same final state of the signal. In our case, processes which produce one jet and missing transverse energy. Many other backgrounds, involving more particles in the final state, can be very relevant, as the detector might miss jets, confuse other particles such as a tau or a photon or an electron with a jet, and very often the determination of the backgrounds does not rely on simulations, but is data-driven.

The only invisible particles of the SM at the LHC are neutrinos, and the irreducible background processes will therefore involve any number of neutrino pairs (of same flavour to conserve lepton flavour number) and one jet. These processes can be simulated by typing the following in the MG5_aMCenvironment:

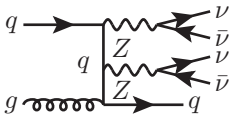
$1j{+}2\nu$

```
generate p p > j vl vl~
output DMcollider/SM_background/1j2nu
```



$1j{+}4\nu$

```
generate p p > j vl vl~ vl vl~
output DMcollider/SM_background/1j4nu
```



$1j + 6\nu$ ...

After performing the simulation with the same parameters as for the signal (the corresponding `lhe` files are provided in the indico page), we can extract the cross-sections of the background processes, which are shown in table 6.

	$1j + 2\nu$	$1j + 4\nu$
σ (pb)	2487.8 ± 7.97	0.18658 ± 0.000574

Table 6: Cross-sections of background processes.

Clearly, the background is heavily dominated by the process with three particles in the final state, and this is due to the much smaller phase space suppression. In case of ideal detector, at the end of Run 3 the background would give around 746×10^6 events. This is way more than the number of signal events, even for BP1, which has the largest yield, but to understand if the signal is statistically large enough, we need to estimate the significance.

7 Exclusion and discovery: computing significances

Before computing the significance of the signal we need to define the null hypothesis. The possible null hypotheses are: 1) only background contributes to the observed data; 2) both signal and background are present in the data. We assume that both signal and background are governed by Poisson statistics. In the first case we can calculate the **exclusion significance** of our signal by simulating a large number of pseudo-experiments assuming that signal and background are present and compute the p-values with respect to the background-only null-hypothesis. In the second case we can calculate the **discovery significance** of the signal by performing simulations and computing the p-values reversing the role of the hypotheses.

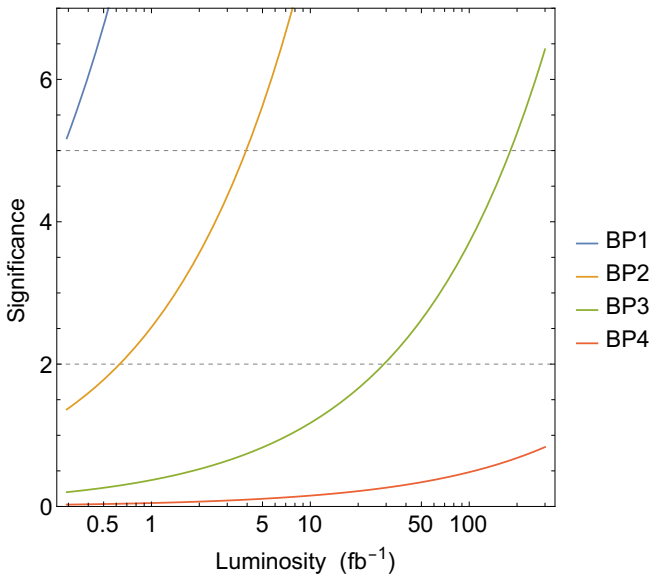
There are various techniques to estimate the significance, involving different statistical considerations, and also considering the uncertainties of signal and background. We refer to the literature for a treatment of these aspects [8–11]. For this exercise, considering that both signal and background are large, we can use here the asymptotic limit of both exclusion and discovery significances, which reads:

$$z = \frac{s}{\sqrt{s+b}} \,. \tag{4}$$

It is commonly assumed that $z \approx 2$ corresponds to the 2σ, or 95%CL exclusion limit, while $z \approx 5$ is the 5σ discovery reach.

This is a rough approximation, but it serves our purposes here. Considering an ideal detector, let's compute the significances of our BPs as function of the luminosity, up to the nominal Run 3 luminosity:

$$z(BP, L) = \sqrt{L}\frac{\sigma_s(BP)}{\sqrt{\sigma_s + \sigma_b}} \,, \tag{5}$$



This is telling us that:

- The cross-section of BP1 is so large that even without imposing any cut, it would generate so many events that it would have been already discovered by now;
- BP2 has also a very large cross-section, which would require such a low luminosity to be discovered, that it would have been observed already;
- BP3 is already excluded at 95% CL, but its discovery reach is close to the luminosity achieved at the end of Run 2 (around 150 fb^{-1}), so there is still some (small) room for discovery during Run 3.
- BP4 seems impossible to either discover or even exclude during Run 3.

However, remember that these are results for an ideal detector and that we are not considering any uncertainty. Assuming that both signal and background yields are reduced by the same amount due to detector acceptances would reduce the significances and make the results weaker.

In the following let's see if we can design kinematical cuts to optimise the significance for BP4 and be able at least to exclude it during Run 3.

8 Kinematical distributions and imposing cuts

Disclaimer: due to time limitations, we will use MadAnalysis 5 to produce plots. It is a powerful tool, and since it is already interfaced with MG5_aMC, it can be easily used to produce plots of your simulations. All references for using the tool can be found in the MadAnalysis 5 webpage `https://launchpad.net/madanalysis5`. However, if used "blindly" it is a black-box, as no information is displayed about how the events shown in the previous section are read, which information is used and so on. For this reason we suggest you to try *at least once* to analyse the events by writing your code or codes for reading the LHE file, extracting the physical observables from the events, and generateing the corresponding histograms. This will give you a feeling of how to perform independent analyses, also because in some cases one needs to consider non-standard quantities which might not be implemented by default in public tools.

The number of background events is so large that the signal corresponding to BP4 would be statistically not significant enough to be observed. However, if the kinematical properties of the final states of background and signal are different enough, we could exclude the region where most of background events fall and optimise the significance for the surviving ones. To do this let's see the properties of some key distributions.

We will run MadAnalysis 5 using a set of instructions which import the reconstructed events for both background and BP4 signal, and superimpose their distributions.

For this we shall write a text file containing the instructions for MadAnalysis 5 (called for example `<MG path>/DMcollider/madanalysis5_hadron_card_BP_comparison.dat`) with the following commands (this file is also included in the indico page, but paths should be suitably modified):

```
import <MG path>/DMcollider/F3S_1j/Events/monojet_signal_BP4/tag_1_pythia8_BasicReco.lhe.gz as BP4
import <MG path>/DMcollider/SM_background/1j2nu/Events/1j2nu/tag_1_pythia8_BasicReco.lhe.gz as SM

set BP4.xsection=0.076003
set SM.xsection=2487.8

set BP4.type = signal
set SM.type = background

set main.lumi=300

set main.stacking_method = normalize2one
set BP4.linecolor = red+2
set SM.linecolor = black

# Basic plots
plot MET 40 0 1000
plot MET 40 0 1000 [logY]
plot THT 40 0 500
plot THT 40 0 1000 [logY]
# basic properties of the non-b-tagged jets
plot PT(j[1]) 40 0 1000 [logY]
plot ETA(j[1]) 40 -10 10 [logY]
# Angular distance distributions
plot DELTAR(j[1],invisible) 40 0 10 [logY]

select MET > 100
select MET > 200
select MET > 300
select MET > 400
select MET > 500
select MET > 600

submit <MG path>/DMcollider/ma5_BPcomparison_F3S_1j
Y
```

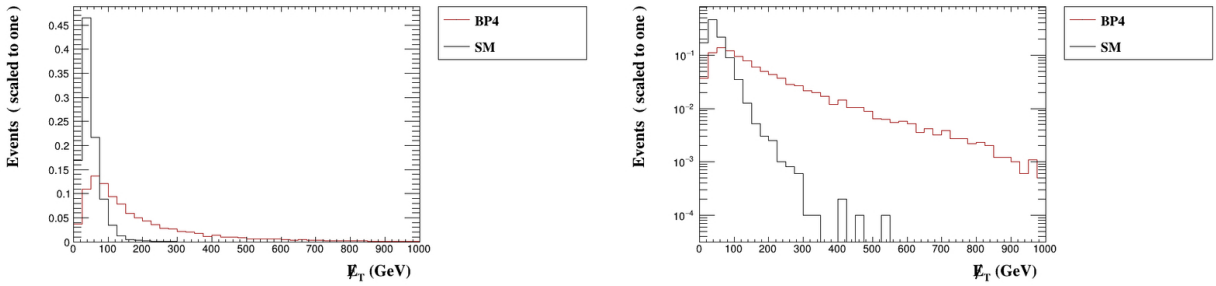
The file does the following:

- Import the samples for BP4 and SM background
- Associate the corresponding cross-sections (in pb)
- Define which is signal and which is background
- Define the luminosity at which the analysis is performed (in fb^{-1})
- Set some graphical parameters and plot some quantities: missing transverse energy, the scalar sum of transverse momenta of visible objects H_T, the transverse momentum and pseudo-rapidity of the leading jet, and the distance between the leading jet and the vector of the missing transverse momentum. Plots are either in linear or logarithmic scale: this might help to better visualize differences.
- A sequence of increasing cuts is imposed on the missing transverse energy
- The whole sequence is submitted to MadAnalysis 5and the output is sent to a folder, for example `DMcollider/ma5_BPcomparison_F3S_1j`
- The last line simply confirms to overwrite the folder if it already exists

To run the analysis navigate to `<MG path>/HEPTools/madanalysis5/madanalysis5/bin` and run the following command:

```
python3.7 ma5 < <MG path>/DMcollider/madanalysis5_hadron_card_BP_comparison.dat
```

Let's see how the distribution of missing transverse energy looks like. All the results of the analysis can be found in a PDF file contained in `DMcollider/ma5_BPcomparison_F3S_1j/Output/PDF/MadAnalysis5job_0/` (and provided in the indico page). In this file we can see the MET distribution, which looks like:



We can see that while the background events are mostly concentrated in the low MET region, the signal events are spreading also to much higher values. By selecting only events with a large MET we should thus be able to remove most of the background and, if the number of surviving signal events is sufficient, observe them.

The sequence of cuts at the end of the input file shows how the significance changes as the cut on MET increases:

Cuts	Signal (S)	Background (B)	S vs B
Initial (no cut)	22800	746340000	0.835
SEL: MET > 100.0	13575.7 +/- 74.1	45526729 +/- 6538	2.012 +/- 0.011
SEL: MET > 200.0	7173.2 +/- 70.1	4104870 +/- 2020	3.5374 +/- 0.0346
SEL: MET > 300.0	4099.6 +/- 58.0	447804 +/- 668	6.098 +/- 0.086
SEL: MET > 400.0	2487.6 +/- 47.1	298536 +/- 546	4.5339 +/- 0.0855
SEL: MET > 500.0	1479.8 +/- 37.2	74634 +/- 273	5.364 +/- 0.134
SEL: MET > 600.0	930.3 +/- 29.9	0.0 +/- 0.0	30.50 +/- 0.49

As we can see, by applying a simple cut we can increase the significance to above the discovery threshold. However, care must be taken in not hitting a region where the background has a very limited statistics. If the cut is above 300 GeV, the number of background events fluctuates significantly, and the significance follows this behaviour.

References

[1] T. Sjöstrand, S. Ask, J. R. Christiansen, R. Corke, N. Desai, P. Ilten, S. Mrenna, S. Prestel, C. O. Rasmussen, and P. Z. Skands, *An introduction to PYTHIA 8.2*, *Comput. Phys. Commun.* **191** (2015) 159–177, [arXiv:1410.3012].

[2] E. Conte, B. Fuks, and G. Serret, *MadAnalysis 5, A User-Friendly Framework for Collider Phenomenology*, *Comput. Phys. Commun.* **184** (2013) 222–256, [arXiv:1206.1599].

[3] E. Conte, B. Dumont, B. Fuks, and C. Wymant, *Designing and recasting LHC analyses with MadAnalysis 5*, *Eur. Phys. J. C* **74** (2014), no. 10 3103, [arXiv:1405.3982].

[4] R. D. Ball et al., *Parton distributions with LHC data*, *Nucl. Phys. B* **867** (2013) 244–289, [arXiv:1207.1303].

[5] M. Cacciari, G. P. Salam, and G. Soyez, *FastJet User Manual*, *Eur. Phys. J. C* **72** (2012) 1896, [arXiv:1111.6097].

[6] E. Boos et al., *Generic User Process Interface for Event Generators*, in *2nd Les Houches Workshop on Physics at TeV Colliders*, 9, 2001. hep-ph/0109068.

[7] J. Alwall et al., *A Standard format for Les Houches event files*, *Comput. Phys. Commun.* **176** (2007) 300–304, [hep-ph/0609017].

[8] A. L. Read, *Modified frequentist analysis of search results (The CL(s) method)*, in *Workshop on Confidence Limits*, 8, 2000.

[9] A. L. Read, *Presentation of search results: The CL(s) technique*, *J. Phys. G* **28** (2002) 2693–2704.

[10] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, *Asymptotic formulae for likelihood-based tests of new physics*, *Eur. Phys. J. C* **71** (2011) 1554, [arXiv:1007.1727]. [Erratum: Eur.Phys.J.C 73, 2501 (2013)].

[11] P. N. Bhattiprolu, S. P. Martin, and J. D. Wells, *Criteria for projected discovery and exclusion sensitivities of counting experiments*, *Eur. Phys. J. C* **81** (2021), no. 2 123, [arXiv:2009.07249].