Introduction to Data Analysis in High Energy Physics

Matheus Pereira Coelho* and Pedro Galli Mercadante†
Centro de Ciências Naturais e Humanas - UFABC
(Dated: April 20, 2023)

The research uses event generators, as MadGraph and Pythia, to simulate the production of new physics signals in reactions involving particle accelerators, particularly in the CERN CMS experiment. The search for physics beyond standard model (SM) goes through the refined understanding about the generation process performed by these frameworks with SM reactions. This work presents particular interest in the study of dark matter (DM) production described by Freeze-in scenarios and, for that reason, models which predict the existence of Feebly Interacting Massive Particles (FIMPs) were adopted. The chosen model predict the existence of a scalar and stable DM candidate, invariant under $SU(3)_C \times SU(2)_L \times U(1)_Y$ transformations and also the existence of a heavy charged parent, which is a $SU(2)$ singlet, and therefore, could be lepton-like or quark-like. The properties of the new model were included into the softwares and the signals of the new particles were discussed.

I. INTRODUCTION

Throughout the twentieth century, as a result of a wide range of cosmological and astronomical observations, dark matter (DM) has become one of the most important topics in modern physics. Comparisons between baryonic mass and gravitational effects in structures with vastly different scales, ranging from small galaxies to galaxy clusters, strongly suggest the existence of an exotic form of matter [1], [2]. Along with that, some remarkable examples comes from gravitational lenses, galaxies rotation curves and stellar velocities, where the last one is represented on figure 1.

With the assumption that such matter can be described as particles, some models try to explain possible mechanisms for DM production in the early universe in order to match the actual relic density estimated by cosmological probes, such as cosmic microwave background anisotropies [3]. The Freeze-out and Freeze-in scenarios furnish two different pictures of DM synthesis. The Freeze-out assumes an initial production in which thermal equilibrium is reached, meaning that DM particles were created and annihilated at the same rate. With the expansion of the universe and temperatures decreasing, DM production became less likely, while annihilation keep occurring until became highly improbable, leaving a constant density. In contrast, the Freeze-in scenario assumes that thermal equilibrium was not reached during the initial production of DM particles and initially they have not density enough, turning the annihilations unlikely in this early stage. Unlike the freeze-out scenario, the DM density increased as the universe expanded and cooled down, eventually reaching a constant density [4], [5]. While the Freeze-out propose the existence of the Weakly Interactin Massive Particles (WIMP's), Freeze-in propose the existence of the Feebly Interacting Massive Particles (FIMP's), both neutral, but the second one with

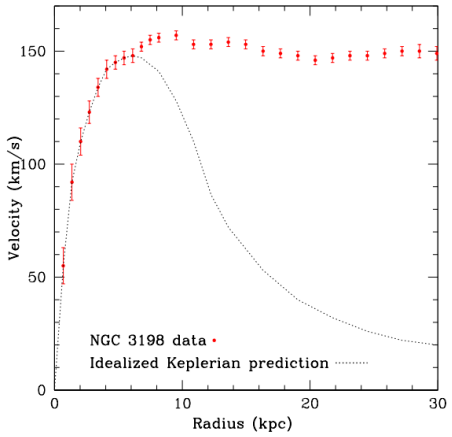

FIG. 1. Comparison between keplerian prediction for star velocities and data [6].

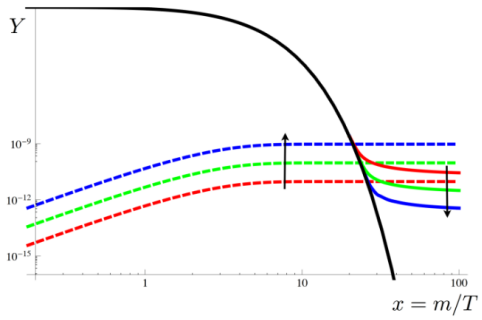

FIG. 2. Freeze-out and Freeze-in are described by solid and dashed lines, respectively. The plot illustrates the evolution of DM density over time, where the X-axis represents the ratio of mass to temperature [7].

even small couplings to standard model (SM) particles. Figure 2 put in contrast the two mechanisms.

The proposition of DM candidates that can explain the observations of these gravitational and cosmological phenomena essentially passes through SM extensions. An extension to the SM as well as strategies for experimen-

* matheus.coelho@aluno.ufabc.edu.br
† pedro.mercadante@ufabc.edu.br

tal searches, particularly in the context of the Compact Muon Solenoid (CMS) experiment at CERN [8], are discussed. The simulation of new reactions using Monte Carlo generators along with real data is a powerful tool to test new models, thus, events were generated to investigate possible signals of DM in CMS detectors. Additionally, a deep understanding of SM reactions that can act as background for new models is required, making it a critical focus of this work.

II. STANDARD MODEL

Based on fundamental pillars of the modern physics, the SM is constructed as a quantum field theory. One of the key concepts of quantum mechanics and special relativity, thus of the quantum field theory itself, is the notion of symmetry. Such concept appears very naturally even in classical theories, which makes it an essential ingredient to search for new physics [9]. SM is represented on figure 3.

The SM is composed of fermions and bosons and is proposed to describe all particles and their interactions. Fermions are composed of quarks and leptons, and one of the fundamental differences between them is that quarks interact via both the electroweak and strong forces, while leptons only interact via the electroweak force. These interactions are mediated by vector bosons.

The symmetry of the SM is described by the group $SU(3) \otimes SU(2) \otimes U(1)$. It is well known that members of the $SU(N)$ algebra are both hermitian and traceless, which implies the existence of $N^2 - 1$ generators. The $SU(2) \otimes U(1)$ group is generated by a linear combination of $SU(2)$ and $U(1)$ generators, and since the electroweak bosons obey the same Lie algebra, $W^{\pm}$, Z and the photon are intrinsically associated with them. On the other hand, the eight gluons of strong interactions are also related to $SU(3)$ generators [10].

Understanding these groups is essential for proposing extensions, as they assume the existence of new fields. Therefore, defining the way in which these new fields relate to the SM group and other symmetries is a crucial point to new theories.

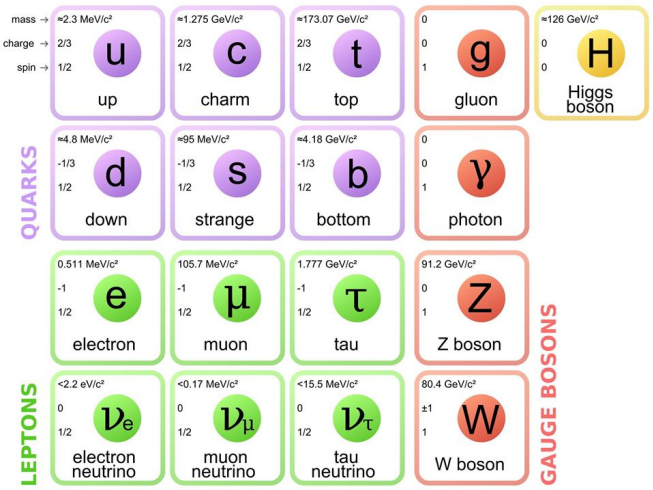

FIG. 3. Representation of the Standard Model.

III. ADOPTED MODEL

A Freeze-in model is adopted [11]. The idea is to make a minimal extension of the SM which can explain the actual dark matter density in universe and the current absence of evidence in particle colliders.

Two new fields (F, s) and a new symmetry described by the discrete group Z_2 is introduced. The new fields are denoted as Z_2-odd while the SM particles are Z_2-even and, since it is desirable a stable candidate, the conservation of the new quantum number is required.

The field denoted by s is the dark matter candidate and, besides of being stable, it is also a real scalar (spinless) and invariant under $SU(3) \otimes SU(2) \otimes U(1)$. On the other hand, F is taken as a heavy charged particle and a $SU(2)$ singlet. Thus, it is interesting to see that it can be taken as lepton or quark-like depending on the way that the field transforms under $SU(3) \otimes U(1)$. However, here only the lepton-like scenario is investigated. The general lagrangian of the model is written as

$$\mathcal{L} = \mathcal{L}_{SM} + \partial_\mu s \partial^\mu s - \frac{\mu_s^2}{2} s^2 + \frac{\lambda_s}{4} s^4 + \lambda_{sh} s^2 \left(H^\dagger H\right) + \bar{F}(i\not{D})F - m_F \bar{F} F - \sum_f y_s^f \left(s\bar{F} \left(\frac{1+\gamma^5}{2} \right) f + \text{h.c} \right). \quad (1)$$

The set of parameters $\{m_s, m_F, \lambda_s, \lambda_{sh}, \{y_s^f\}\}$ in equation (1) takes into account the masses of the new fields, the self-coupling and Higgs coupling for s and the coupling with fermions. The first four extensions terms in lagrangian are, respectively, the Klein-Gordon lagrangian, the self-interacting and the Higgs coupling term for the massive field s. The last three terms are the free lagrangian for the massive spinorial field F and the interaction term between s, F and SM fermions denoted by f.

The extension terms obey fundamental rules for a consistent theory, such as unitarity and Lorentz and gauge invariance. For the present analysis, the self-coupling and Higgs coupling parameters could be set as $\lambda_s = \lambda_{sh} = 0$

[12] . Also for F as a lepton-like field, the lagrangians turns into

$$\mathcal{L} = \mathcal{L}_{SM} + \partial_\mu s \partial^\mu s - \frac{m_s^2}{2} s^2 + \bar{F}(i\not{\partial})F + \\ - m_F \bar{F} F - \sum_l y_s^l \left(s\bar{F} \left(\frac{1+\gamma^5}{2} \right) l + \text{h.c} \right). \quad (2)$$

The conservation of the new quantum number implies that the decay of F has the form of $F \to s + l$, where l is a SM charged lepton.

IV. THE CMS DETECTOR AND EVENT SIMULATION

A. Detector Apparatus

Some properties of the new fields are taken into account according to the characteristics of the CMS detector, otherwise any conclusions could be inferred about the new model. Due to this, a brief description about the detector is necessary [13], [14].

The CMS detector has two major detection schemes composed by a tracker system and two calorimeters. The matching between the information given by both systems allows the description of important physical properties.

The track system is composed of semi-conductor shells and is imbued in a strong magnetic field. The first part, namely silicon pixels, is the closest detector to the collision beam and is fundamental to the detection of charged short-lived particles. Particles with sufficient lifetimes can pass to the next part, the silicon strips. The last one is made up of silicon barrels and reach approximately 1.2m from beam. The key point of the track system functionality lies in the physics of the semi-conductors. Essentially, in the presence of a particle with ionizing energy coming from the beam, these components produce electrical signals which make it possible to reconstruct the particle's path in a non-destructive way. From the curved trajectory due to the magnetic field, it is possible to infer the charge and momentum of these particles.

As mentioned before, the CMS has two calorimeters: the electromagnetic (ECAL) and the hadronic (HCAL), both composed of dense materials and scintillators. They are designed to reconstruct the properties of particles in a destructive way, which means that it is expected that the particles will deposit all their energy in these detectors. The energy deposition of particles can be described by the Bethe-Block formula, Bremsstrahlung effect, pair production, and other phenomena. As a consequence, in general, charged particles produce electromagnetic showers. On the other hand, quantum chromodynamics effects produce many hadrons that can be observed in HCAL as jets, which are generally determined by a bunch of particles within a certain radius on the calorimeter plane.

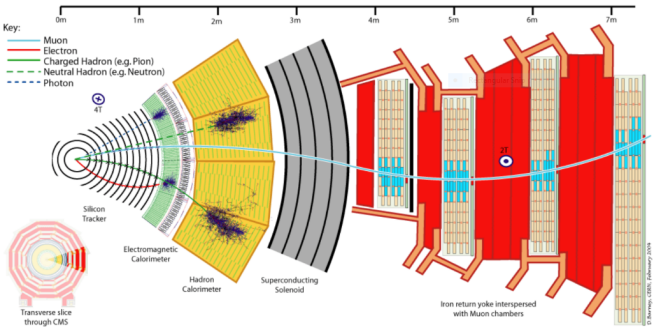

FIG. 4. CMS detector representation.

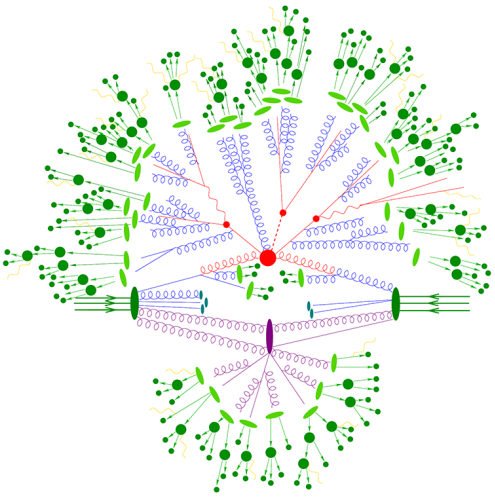

FIG. 5. Representation of different stages of event generation. The core reaction is simulated by MadGraph and hadronized by Pythia, reproducing what comes to detectors.

As said earlier, a global reconstruction with information of these different detection strategies is fundamental to the measurement of physical quantities. Given the cylindrical geometry of the experiment, with the beam in z direction, the azimuthal angle of the transverse plane is Lorentz invariant. Quantities such as missing transverse energy (MET) and transverse momentum (p_T) are interesting due to this invariance property. A simple sketch about the CMS detector apparatus is shown on figure 4.

B. Production and analysis software

The data production were performed by MadGraph5 and Pythia [15], [16]. The first one deals with matrix elements of the initial and finals states of the reactions, the fundamental stage of the generation process, while the second simulates secondary effects such as hadronization, and initial and final state radiation (ISR, FSR). These Monte Carlo generators produce a .ROOT file with all physical information present in events, such as four-

momentum and energy of the particles, jets coordinates in calorimeter plane and other information that makes possible a lot of analysis. From this output, the data analysis is done with ROOT, an oriented object framework, fundamental for the data interpretation [17]. Finally, with algorithms in C++ the information in events can be accessed and used to produce graphs. The event generation is represented in figure 5.

C. New Model Implementation

In order to generate reactions with the new fields, we implemented the new model into MadGraph. The software needs some inputs based on new fields properties, such as the decay width, mass and coupling constants. As it will be explained in next sections, one of the main focus of the analysis is the investigation of disappearing tracks, which set boundaries for the distances traveled by these particles in the detector, $d = \beta\gamma c\tau$. Taking $c\tau = 1\text{m}$, the associated decay width can be obtained:

$$\begin{aligned} \Gamma_{F\to sl} &= \frac{\hbar}{\tau} \\ \Gamma_{F\to sl} &= 1.9733 \times 10^{-16}\ \text{GeV}. \end{aligned} \tag{3}$$

On the other hand, cosmological considerations sets a lower bound for the s mass ($m_s \geq 12$ KeV) [11], and since s is the lighter Z_2-odd field, $m_F > m_s$. Among many possible scenarios, these masses can be chosen to be closer to each other and here the analysis is constrained to $m_F = 200$ GeV and $m_s = 199$ GeV. Furthermore, the partial decay widths are, in fact, very similar in a way that $\Gamma_{F\to es} \approx \Gamma_{F\to\mu s} \approx \frac{\Gamma_{F\to sl}}{2}$. The following equation establishes the connection between these quantities and makes possible the coupling constants calculation [18]. We have,

$$\Gamma_{F\to sl} = \frac{\left(y_s^l\right)^2}{32\pi} m_F \left(1 - \frac{m_s^2}{m_F^2} - \frac{m_l^2}{m_F^2}\right)^2. \tag{4}$$

These values, along with the partial decay widths returned by MadGraph, are shown in the table bellow.

Electron Channel
$y_s^e = 7.0599 \times 10^{-7}$
$\Gamma_{F\to es} = 9.8663 \times 10^{-17}$GeV
Muon Channel
$y_s^\mu = 7.0601 \times 10^{-7}$
$\Gamma_{F\to\mu s} = 9.8659 \times 10^{-17}$GeV

It is important to note that only muon and electron decay channels are possible, since the difference of mass between s and F is less than the tau mass.

Therefore, all free parameters in the effective lagrangian (2), $\{m_s, m_F, \{y_s^l\}\}$, are described and reactions with the new fields can be generated.

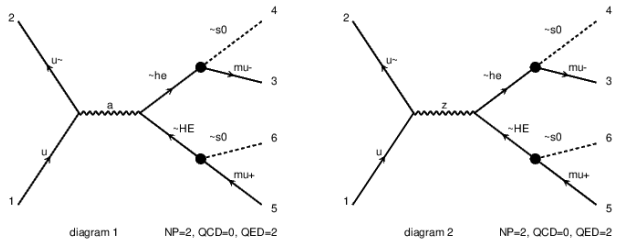

FIG. 6. Feynman diagrams of the reactions with muons in decay product. Here we identify $F \to he$ and $s \to s0$.

V. RESULTS AND ANALYSIS

As described in previous sections, the analysis is focused in a leptonic F. The nature of this particle is fundamental for the experimental search, since a hadronic origin could lead to different signals in the detector. However, there are some typical signatures for particles which its decay products do not interacts with detectors, such as the presence of remarkable missing energy in the events. A few comments about possible signals and strategies to search for these particles are discussed.

A. Leptonic decay

Given the model specifications, one of the reactions of interest is the following:

$$p\,p \to F\,\bar{F},\ (F \to l_-\ s)\,,\ \left(\bar{F} \to l_+\ s\right).$$

Its Feynman diagram is represented in figure 6.

The chosen m_s, m_F and $c\tau$ implies that the charged parent will decay inside the CMS tracker leaving a track that is discontinued, and therefore disappears. For the case in which the masses of F and s are close to each other, we can assume that s carry all momenta of F. Such assumption requires that the lepton has no substantial deposits in calorimeter since they are produced with low momentum. In most cases, these leptons fail to be reconstructed or do not pass in criteria selection, especially because they cannot leave the tracker due to the strong magnetic field. This not only emphasizes the need to search for disappearing tracks but also for displaced vertices, especially for $c\tau < 1\text{m}$ [11]. Such signals are typical for long lived particles due to the small couplings, which is a property of FIMP candidates. In contrast, WIMP candidates tends to decay quite instantly, demanding different search strategies.

The transverse momentum (p_T) distribution can bring some important information about the reactions. We generate 25k events at MadGraph and Pythia with the specified reaction above. The results are displaced in figure 7.

Comparing the two histograms on the left, given that MadGraph generates the $F\bar{F}$ pair with zero momentum,

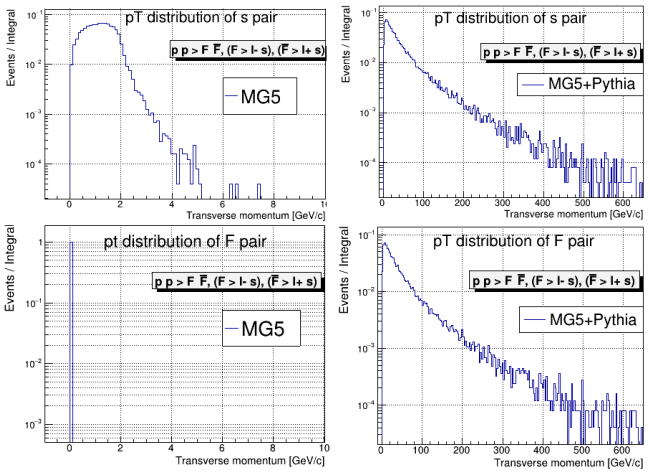

FIG. 7. Normalized histograms with hard process level events on the left (MadGraph) and events that have passed through hadronization, ISR and FSR effects on the right (MadGraph+Pythia). The upper and lower side histograms show the reconstructed p_T of the F pair and the s pair, respectively.

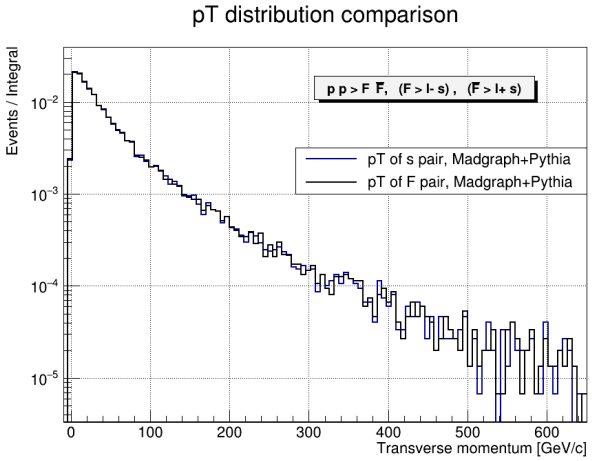

FIG. 8. Superposition of the two histograms at the right of figure 7. Both distributions seems very similar, indicating that considering the leptons or not does not produce any significant difference.

the downside histogram peaked at zero is expected. In contrast, considering only the s pair (i.e., disregarding the leptons) leads to a residual distribution with typical low momenta, as seen in the upside histogram. In fact these small differences at generator level does not change the distributions after hadronization and ISR/FSR effects, which is the reason for both distributions on the right being so similar. Moreover, it is essential to highlight that the histograms on the right represent what would actually come to the detectors and, since that there is no huge difference whether we consider the leptons or not, these distributions enforces the fact that these leptons are not important in this particular analysis. As an

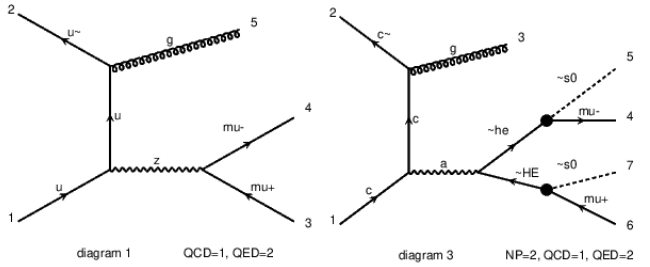

FIG. 9. Feynman diagrams of $p\ p \to Z + jet,\ (Z \to \mu^+\mu^-)$ and $p\ p \to F\ \bar{F} + jet,\ (F \to l_-\ s)\,,\ \left(\bar{F} \to l_+\ s\right)$.

important consequence, the relevant physical information is not measured by the detectors, meaning that the signal is expressed in terms of a substantial missing momentum. Figure 8 shows the superposition of the two histograms with Pythia, making these conclusions clearer.

There are other interesting physical processes that can also generate a final state with the dark matter candidate and SM leptons, especially those which can produce signals that are directly measurable in the detector. One remarkable example is given by:

$$p\ p \to F\ \bar{F} + jet,\ (F \to l_-\ s)\,,\ \left(\bar{F} \to l_+\ s\right).$$

This reaction demonstrates that the initial quarks can emit radiation in the form of quarks and gluons, resulting in a jet and a pair of F. Due to momentum conservation, a high energy jet is expected to counterbalance the momentum of the F particles, generating an important signal for this type of reaction. It should be noted that in real events all processes occur, however, in this context, only those involving jets in the final state are capable of producing detectable signals in the calorimeters.

At this point, it is instructive to look back at SM reactions and investigate processes that also includes jet production along with Z boson. For example:

$$p\ p \to Z + jet,\ (Z \to \nu\bar{\nu}).$$

There are two main reasons why understanding Z processes are really important in this context. Firstly, just like dark matter candidates do not interact with the detector, the Z boson also has decay channels with invisible particles (neutrinos). Therefore, some considerations can be extended to the new particles. Secondly, this reaction is background for a lot of beyond SM models, including this one. It is fundamental to comprehend such processes to make a correct background modeling and discrimination. The figure 9 put in contrast the Feynmann diagrams for $Z + jet$ and $F\ \bar{F} + jet$.

As the discussion goes through how these reactions are generated, the next subsection is focused on the Z boson, in order to compare the differences between $p\ p \to Z + jet\ , (Z \to \mu^+\mu^-)$ and $p\ p \to Z, (Z \to \mu^+\mu^-)$ at the Monte Carlo generators.

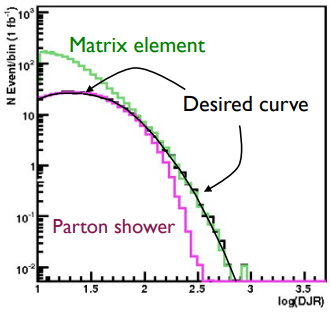

FIG. 10. Number of events in terms of DJR [19]. The curve above the desired curve shows an overestimated number of jets.

B. Matching

A few questions emerges when jets are produced in MadGraph. It is important to remember that in reaction $p\ p \rightarrow Z, (Z \rightarrow \mu^+\mu^-)$ the jets are produced only by Pythia ISR and FSR effects, with tipical low momenta, especially when compared to those produced directly by MadGraph. Moreover, a double counting problem arises because whether or not jets are considered in the hard process, Pythia will hadronize these events and produce more jets, meaning that an overestimated number of jets will be produced. This problem is solved with a matching procedure.

The matching consists in avoid this double counting by setting cuts in the transverse momentum at the generator ($XQCUT$) and clustering ($QCUT$) levels. These parameters are important to the guarantee that the events has been generated in a correct way. In fact, this is evidenced by looking to the number of events or cross section in terms of the differential jet rate (DJR), a variable related to the cuts in p_T [20]. The goal is to choose these cuts in a way that the DJR remains continuous, in order to have the correct number of jets per event and work with MadGraph and Pythia in the best of their regimes. The desired smooth curve is represented in the figure 10.

However, the analysis here is focused on observing how the p_T distribution of the Z boson changes with these cuts, especially $XQCUT$. A first look at the figures 11, 12, and 13 shows that the momentum distribution is quite similar for different cuts. This means that regardless the jets are produced on MadGraph or Pythia, the Z boson is produced with the same properties, and there will be no remarkable differences in its momentum distribution. In fact, these distributions seem to be insensitive not only to the cuts applied but also to the number of jets in the events. The ratio between the distributions below each histogram supports these conclusions.

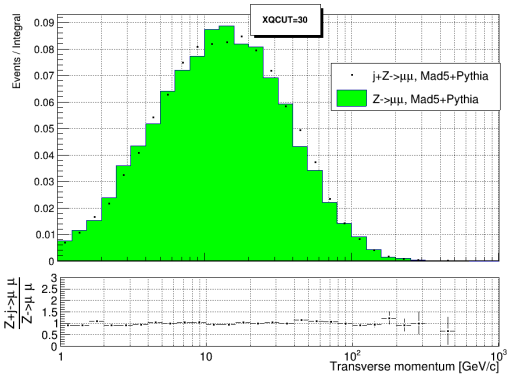

FIG. 11. Comparison between the p_T distribution of the Z boson for $Z \rightarrow \mu\mu$ and $Z + j$, $(Z \rightarrow \mu\mu)$ processes with $XQCUT = 30$ GeV/c.

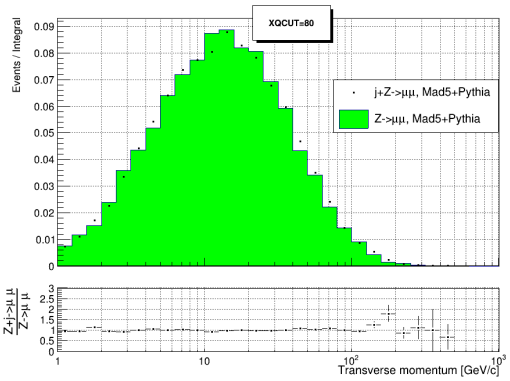

FIG. 12. Comparison between the p_T distribution of the Z boson for $Z \rightarrow \mu\mu$ and $Z + j$, $(Z \rightarrow \mu\mu)$ processes with $XQCUT = 80$ GeV/c.

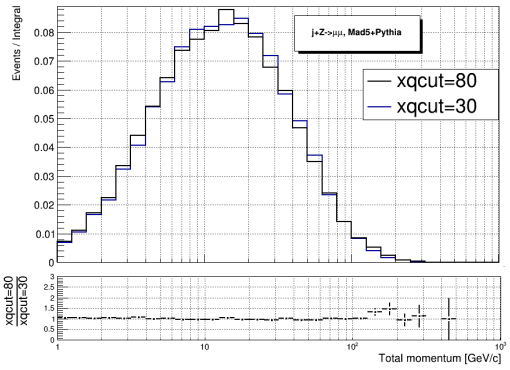

FIG. 13. Comparison between the p_T distribution of the Z boson for $Z+j$, $(Z \rightarrow \mu\mu)$ process with $XQCUT = 30$ GeV/c and $XQCUT = 80$ GeV/c.

VI. CONCLUSIONS

Based on these results, a few conclusions can be drawn about how to search for these specific dark matter candidates. With the scenarios covered here, the search for

events with at least one high p_T jet and missing energy, along with investigation of disappearing tracks, defines the main possible signals produced by the candidates of the worked model. Due to the introductory character of the analysis presented here, parameters such as the masses of the new fields and $c\tau$ were constrained for specific values. Other interesting results, as exclusion curves, can be obtained by the variations of these values. Furthermore, the investigation of a F field with quark-like properties needs also a very cautious investigation.

The matching analysis shows that it is perfectly possible to generate reactions including quarks/gluons at the matrix elements generation level with coherent results. The advantage is the possibility to understand better reactions that are background for beyond standard model processes, as well as the signals left by both.

Therefore, it was possible to discuss important topics about the search for dark matter candidates in the LHC-CMS experiment. The existence of different models based on different kinds of dark matter makes the experimental search rich and diverse. The study presented here is only an initial search among the vast quest for dark matter.

[1] J. D. Simon, *The Faintest Dwarf Galaxies* (arXiv:1901.05465 [astro-ph.GA], 2019).

[2] S. W. Allen, A. E. Evrard, and A. B. Mantz, *Cosmological Parameters from Observations of Galaxy Clusters* (arXiv:1103.4829 [astro-ph.CO], 2011).

[3] Planck Collaboration, *Planck 2018 results. VI. Cosmological parameters* (arXiv:1807.06209 [astro-ph.CO], 2018).

[4] A. Boveia and C. Doglioni, *Dark Matter Searches at Colliders* (arXiv:1810.12238v1 [hep-ex], 2018).

[5] A. Goudelis, *Freeze-in: making sense of very weakly coupled dark matter candidates* (GGI Workshop "Collider Physics and the Cosmos" Florence, Italy, 2017).

[6] K. G. Begeman, *HI rotation curves of spiral galaxies. I - NGC 3198* (stron. and Astrophys. 223, 47-60, 1989).

[7] A. Green, *Dark Matter Cookbook: Freeze-In* (https://www.particlebites.com/?p=7132, 2020).

[8] CMS Collaboration, *The CMS experiment at the CERN LHC* (doi:10.1088/1748-0221/3/08/S08004, 2008).

[9] S. F. Novaes, *Standard Model: An Introduction* (arXiv:hep-ph/0001283, 1982).

[10] A. Zee, *Group Theory in a Nutshell for Physicists* (Princeton University Press, 2016).

[11] G. Bélanger, N. Desai, A. Goudelis, J. Harz, A. Lessa, J. No, A. Pukhov, S. Sekmen, D. Sengupta, B. Zaldivar, and J. Zurita, *LHC-friendly minimal freeze-in models* (arXiv:1811.05478 [hep-ph], 2018).

[12] N. Bernal, X. Chu, C. Garcia-Cely, T. Hambye, and B. Zaldivar, *Production Regimes for Self-Interacting Dark Matter* (arXiv:1510.08063 [hep-ph], 2016).

[13] CMS Collaboration, *CMS detector description* (https://cms.cern/index.php/detector).

[14] T. Tomei, *Aspects of High Energy Phyics* (III Escola Jayme Tiomno de Física Teórica, 2021).

[15] J. Alwall et al, *The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations* (arXiv:1405.0301 [hep-ph], 2014).

[16] 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* (arXiv:1410.3012 [hep-ph], 2014).

[17] R. Brun and F. Rademakers, *ROOT - An Object Oriented Data Analysis Framework* (Nucl.Instrum.Meth.A 389 81-86, 1997).

[18] F. O. de Aguiar, *Busca Por Sinais de Desaparecimento de Traços no Interior do CMS* (Tese de mestrado, Universidade Federal do ABC, 2022).

[19] A. George, *What the Hell is Matching* (University of California).

[20] F. Krauss, A. Schälicke, S. Schumann, and G. Soff, *Simulating W/Z+jets production at the Tevatron* (arXiv:hep-ph/0409106, 2004).