Using the modified matrix element method to constrain $L_\mu - L_\tau$ interactions

Fatemeh Elahi and Adam Martin

Department of Physics, 225 Nieuwland Science Hall, University of Notre Dame,
Notre Dame, Indiana 46556, USA

In this paper, we explore the discriminatory power of the matrix element method (MEM) in constraining the $L_\mu - L_\tau$ model at the LHC. The Z' boson associated with the spontaneously broken $U(1)_{L_\mu - L_\tau}$ symmetry only interacts with the second and third generation of leptons at tree level, and is thus difficult to produce at the LHC. We argue that the best channels for discovering this Z' are in $Z \to 4\mu$ and $2\mu + \not{E}_T$. Both these channels have a large number of kinematic observables, which strongly motivates the usage of a multivariate technique. The MEM is a multivariate analysis that uses the squared matrix element $|\mathcal{M}|^2$ to quantify the likelihood of the testing hypotheses. As the computation of the $|\mathcal{M}|^2$ requires knowing the initial and final state momenta and the model parameters, it is not commonly used in new physics searches. Conventionally, new parameters are estimated by maximizing the likelihood of the signal with respect to the background, and we outline scenarios in which this procedure is (in)effective. We illustrate that the new parameters can also be estimated by studying the $|\mathcal{M}|^2$ distributions, and, even if our parameter estimation is off, we can gain better sensitivity than cut-and-count methods. Additionally, unlike the conventional MEM, where one integrates over all unknown momenta in processes with $\not{E}_T$, we show an example scenario where these momenta can be estimated using the process topology. This procedure, which we refer to as the "modified squared matrix element," is computationally much faster than the canonical matrix element method and maintains signal-background discrimination. Bringing the MEM and the aforementioned modifications to bear on the $L_\mu - L_\tau$ model, we find that with 300 fb^{-1} of integrated luminosity, we are sensitive to the couplings of $g_{Z'} \gtrsim 0.002 g_1$ and $M_{Z'} < 20$ GeV, and $g_{Z'} \gtrsim 0.005 g_1$ and 20 GeV $< M_{Z'} < 40$ GeV, which is about an order of magnitude improvement over the cut-and-count method for the same amount of data.

I. INTRODUCTION

The highest priority of LHC run II is finding physics beyond the standard model (BSM). In the pursuit of optimal sensitivity to BSM physics, it is important to develop a diverse set of techniques that improve our sensitivity to the BSM signal against the SM background. The most straightforward technique is the cut-and-count method, where one determines the suitable region of phase space by studying the kinematic distributions of the simulated events (e.g. the invariant mass of two detected particles, or the angle difference between them) and looks for an excess over the SM expectation. While simple, this method can miss correlations among kinematic variables. To improve our discriminatory power, one can select events based on more complicated variables that take into account correlations. These complicated variables can be a linear or a nonlinear combination of the kinematic variables used in the cut-and-count method. The techniques that consider combinations of kinematic variables are generally called "multivariate analysis" (MVA) [1]. Some of the main methods developed in this category are "neural network" [2–6] and "boosted decision tree" [7,8]. However, these MVA methods usually require a phase of computer training and techniques that are not transparent to theorists. Another

example of MVA that is calculated based on the theoretical assumptions for a given process is the matrix element method (MEM) [9–15].

The MEM was originally developed in the Tevatron experiment and was successfully applied to measurement of the top quark mass and single top electroweak production [9,13,16–29]. In the MEM, given a reconstructed event and a theoretical hypothesis, the probability that the event matches the hypothesis is quantified based on the value of the partonic matrix element for the hypothesis processes evaluated using the final state four momenta. Repeating this procedure using multiple hypothesis processes, e.g. a background process and a signal process, and comparing the results, one can quantify how signal-like or backgroundlike a given ensemble of events is. By definition, the MEM contains all of the kinematic information of the hypothesis process so it captures all correlations. Moreover, the MEM has a clear physical meaning and there is a transparent link between the theoretical assumptions and event reconstruction.

Despite the successes of the MEM in the SM measurements, this method has not been applied extensively in BSM searches. Some of the main difficulties of the MEM are the following [9,14]:

(i) The squared matrix element depends on all of the momenta—both initial and final—in the event. In circumstances where one or more momenta is not determined uniquely, we must integrate over all possible values. The more integrations required for each event, the more time and computational resources required. Events with missing energy, a hallmark of many new physics scenarios, automatically fall into this category of events with unconstrained momenta.

(ii) BSM hypotheses necessarily introduce new parameters (α = masses, couplings of new particles), which need to be determined before we can calculate the matrix element. Without a separate experiment telling us what α to use, our best option is to choose the α such that they maximize our reach for detecting the signal given the background. However, finding the best-suited values for these parameters can be difficult, as shown in Sec. IV A.

Ideally, we would like to develop techniques to overcome the aforementioned challenges for *any* BSM hypothesis. However, as a first step, we focus on dealing with some of these challenging for a specific model. The BSM hypothesis we consider in this paper is one of the simplest extensions of the SM, the $L_\mu - L_\tau$ model [30]. In the $L_\mu - L_\tau$ model, the difference between muon number and tau number—an accidental symmetry of the SM—is gauged and spontaneously broken, generating a massive Z' that interacts only with the second and third generation of leptons at tree level [30]. As the $L_\mu - L_\tau Z'$ is hadrophobic and does not interact with electrons at tree level, it is difficult to produce at conventional experiments and therefore is challenging to constrain [31]. In fact, tree-level Z' production at the LHC always involves four second/third generation lepton (either charged or neutral) final states: the initial partons create a pair of leptons via $W^\pm/Z/\gamma$ exchange, one of which radiates a Z' that subsequently decays to a second lepton pair. At the LHC, identifying muons is much easier and more accurate than identifying taus; thus in this work, we only consider combinations of muons and $\not{E}_T$ (due to ν_μ or ν_τ) as potential final states. In particular, we study the two processes of (i) four muons (4μ), and (ii) two muons and missing energy ($2\mu + \not{E}_T$). Of course, as all of the final state particles are SM particles, there is interference between the events produced via Z' and SM events. We put this interference to use in Secs. IV A and IV B.

The four muon final state is fully reconstructable, and it has 12 independent observables. The existence of this many kinematic variables begs for the use of a MVA. The MEM is a suitable choice because it optimally [9] uses all of the available kinematic variables to distinguish signal from background. Applying MEM to the process $2\mu + \not{E}_T$ with six observables is also more lucrative; however, the existence of $\not{E}_T$ makes its computation more challenging.

Fortunately, the main SM background to this process ($pp \to \tau\tau \to 2\mu\not{E}_T$) has a very specific topology, and thus the missing momenta for this process can be estimated from the momenta of muons. In Sec. IV B, we show that by using the squared matrix element of the $\tau^+\tau^-|_{\text{dimuon}}$ background as a discriminatory variable (as opposed to the conventional likelihood function) with the guessed missing momenta derived from its topology (i.e. without having to integrate over missing momenta), we are able to sufficiently differentiate our signal from the remaining SM backgrounds as well. Due to the deviation from the canonical MEM procedure, we refer to this approach as the "modified MEM."

The organization of the rest of the paper is as follows. In the next section, we explain the matrix element method and how it is used to discriminate the signal from the background. In Sec. III, we introduce the model and discuss some of the constraints on its parameters from the LHC and other experiments. In Sec. IV, we explore how the bounds can be improved at the LHC. Finally, a discussion about the results and concluding remarks are made in Sec. V.

II. MATRIX ELEMENT METHOD

At the LHC, we are searching for BSM physics against the SM background. Although we could look for generic departure from the SM, our sensitivity is greater if we look for a particular BSM physics hypothesis. Therefore, we are usually dealing with two hypotheses: a specific new physics model (signal) and the null hypothesis (the SM background). To statistically analyze these two test hypotheses, the MEM uses their likelihood ratio, where the likelihoods are calculated based on the squared matrix element ($|\mathcal{M}|^2$) of a process at tree level,[1] and evaluated using the final state momenta of individual events. Since all the kinematic information of a process is contained in the matrix element, $|\mathcal{M}|^2$ is a powerful variable to discriminate between the two test hypotheses. If the empirical LHC events are inconsistent with either of the conjectures, $|\mathcal{M}|^2$ should also favor neither of them.

The function $|\mathcal{M}|^2$ depends on the momentum of initial and final state partons (p^{par}). The squared matrix element of the signal $|\mathcal{M}|^2_{\text{signal}}$ additionally depends on the new model parameters (α): $|\mathcal{M}|^2_{\text{signal}} = |\mathcal{M}(p^{\text{par}}; \alpha)|^2$, and $|\mathcal{M}|^2_{\text{SM}} = |\mathcal{M}(p^{\text{par}})|^2$. In the special case where the final states of the signal exactly match the SM background (e.g., no BSM particles in the final state), $|\mathcal{M}|^2_{\text{signal}}$ must include the interference term with the SM as well.

If the LHC could detect the p^{par} of all of the final state particles, the likelihood (P) that an observed event is due to a particular hypothesis would be defined as

---

[1]Some papers have studied the MEM at NLO as well [32–35].

$$P(p^{\text{par}}|\alpha) = \frac{1}{\sigma} \int dx_1 dx_2 \frac{f(x_1)f(x_2)}{2sx_1x_2} |\mathcal{M}(p^{\text{par}};\alpha)|^2$$
$$\times \delta^4(p^{\text{par}}_{\text{initial}} - p^{\text{par}}_{\text{final}}), \quad (1)$$

where x_i and $p^{\text{par}}_{\text{initial}}$ are intimately related: $p^{\text{par}}_{\text{initial,i}} \equiv \frac{\sqrt{s}}{2}(x_i, 0, 0, \pm x_i)$. The factors $f(x_i)$ are the parton distribution functions (PDF) of the initial states, the collider center of mass energy of collision is represented by $\sqrt{s}$, and σ is the total cross section with which $P(p^{\text{par}}|\alpha)$ is normalized to ensure $\int P(p^{\text{par}}|\alpha)dp^{\text{par}} = 1$. The factor $\delta^4(p^{\text{par}}_{\text{initial}} - p^{\text{par}}_{\text{final}})$ ensures the conservation of energy and momentum in the process. If we know the final state four momenta, we can use this delta function to infer information about the initial state momenta. More specifically, the delta function involving the $p^{\text{par}}_{\text{initial}}$ can be translated to a delta function on x_i's, uniquely defining the x_i's and collapsing the integrals,

$$\delta^4(p^{\text{par}}_{\text{initial}} - p^{\text{par}}_{\text{final}}) = \delta\left(\frac{\sqrt{s}}{2}(x_1 + x_2) - p^{\text{par,Energy}}_{\text{final}}\right)$$
$$\times \delta^2(p^{\text{par,transverse}}_{\text{final}})$$
$$\times \delta\left(\frac{\sqrt{s}}{2}(x_1 - x_2) - p^{\text{par,longitudinal}}_{\text{final}}\right). \quad (2)$$

The events at the LHC, however, are defined according to reconstructed momenta (p^{rec}) at the detector which may not equal the p^{par}. Hence, we must modify the likelihood (P) to be a function of detector level momenta,

$$P(p^{\text{rec}}|\alpha) = \frac{1}{\sigma} \int d\Phi(p^{\text{par}}_{\text{final}})dx_1 dx_2 \frac{f(x_1)f(x_2)}{2sx_1x_2}$$
$$\times |\mathcal{M}(p^{\text{par}};\alpha)|^2 \delta^4(p^{\text{par}}_{\text{initial}} - p^{\text{par}}_{\text{final}})W(p^{\text{rec}}, p^{\text{par}}), \quad (3)$$

with an integration over the possible values of partonic momenta given the reconstructed momenta [represented by $\int d\Phi(p^{\text{par}}_{\text{final}})dx_1 dx_2$]. When there is missing energy (e.g., neutrino) in the event, we integrate over all unconstrained momenta. However, if the particle is detected, we use the transfer functions $W(p^{\text{rec}}, p^{\text{par}})$ to translate between the partonic momenta and reconstructed momenta. These functions are usually Gaussian (or bi-Gaussian), where the arguments are estimated according to Monte Carlo (MC) simulations[2] [10–12]. If the detected particle is a lepton or a photon, the transfer function is well

---

[2]In the matrix element technique, the theoretical assumptions (the $|\mathcal{M}|^2$) and the assumptions about the experiment and detectors (the transfer functions) factorize and are independent of each other. Therefore, the improvement in any of these assumptions can easily be implemented in the matrix element method.

approximated by a delta function [$\delta(p^{\text{par}} - p^{\text{rec}})$], while for colored objects, the reconstructed momenta may be significantly different from the partonic momenta.

After calculating the likelihood function (P) of the signal and the SM, we need to determine whether a given event is more likely to be due to the signal hypothesis or the background. Therefore, we study the likelihood ratio,

$$\mathcal{L}(p^{\text{rec}}_i;\alpha) = \frac{P(p^{\text{rec}}_i|\text{new phyiscs}(\alpha))}{P(p^{\text{rec}}_i|\text{null hypothesis})}, \quad (4)$$

where the i subscript refers to the ith event. If the value of α were known, we could plot the distribution of $\mathcal{L}$ for given events, just like any usual kinematic distribution that uses the reconstructed momenta. In the $\mathcal{L}$ distribution, larger values would indicate that the signal hypothesis is favored and lower values mean the data are more consistent with the null hypothesis. Hence, with a cut on the larger values of $\mathcal{L}$, we could find the phase space that increases the signal fraction.

Having said that, in the case of new physics, we do not know the value of α. Since we want to optimize our reach for the signal hypothesis, we want to choose α that maximizes $\mathcal{L}(p^{\text{rec}}_i;\alpha)$ (or equivalently $\log[\mathcal{L}(\alpha)]$). Therefore, we plot $\log[\mathcal{L}(\alpha)]$ with respect to α and look for maxima in the plot. This process can be done for each event. For multiple events, we simply sum over i,

$$\sum_i \log[\mathcal{L}(p^{\text{rec}}_i;\alpha)] = \sum_i \log\left[\frac{P(p^{\text{rec}}_i|\text{new phyiscs}(\alpha))}{P(p^{\text{rec}}_i|\text{null hypothesis})}\right]. \quad (5)$$

We denote α^* for the value of α that maximizes Eq. (5). As it will be important later on, we emphasize that Eq. (5) only yields one number for an entire set of events. Previous studies have shown that if α is the mass of a particle, the α^* returned by maximizing Eq. (5) is actually the same as the true value of α; in fact, the current most precise measurement of the top quark mass is obtained with this process of maximizing the likelihood ratio with respect to top mass [9,13]. This procedure is used only to determine α^*, and it gives no information on the prospect of discovering the signal. Once the optimal value α^* is determined, the sensitivity to the signal is determined by selecting a region of the $\mathcal{L}(\alpha^*)$ [with $\alpha = \alpha^*$ in Eq. (4)] distribution that optimizes the signal over background ratio.

Despite the power of the MEM in discriminating signal against the null hypothesis, due to its computational difficulties, it is not commonly used in BSM searches. One reason may be the number of phase space integrations in the calculation of $P(p^{\text{rec}}|\alpha)$. If the process of interest has multiple sources of missing energy or contains colored objects, the number of integrals can be high and may overconsume computational resources. Another reason is that the maximization of likelihood ratio with respect to α

can be challenging depending on the nature of parameters. As we see in Sec. IIIA, the maximization of log likelihood works if α is the mass of a new particle, but this approach breaks down if α is a coupling. That is because $\log[\mathcal{L}(\alpha)]$ only increases with respect to α and has no local maxima. Although one naively might expect that larger values of α result in greater sensitivity to the BSM hypothesis, in Sec. IV A, we show that is not correct.

In this paper, we study a simple model to give a working example of how some of the challenges in MEM for BSM can be overcome. The example we consider is the gauged $U(1)_{L_\mu - L_\tau}$ symmetry, summarized in the next section. This model contains a massive Z' gauge boson that only interacts with the second and third generation of leptons. We show that the LHC sensitivity can be improved by an order of magnitude compared with the cut-and-count method if we apply the MEM.

III. $L_\mu - L_\tau$ MODEL

The difference between muon number and tau number $L_\mu - L_\tau$ is one of the accidental global symmetries present in the SM. This $U(1)_{L_\mu - L_\tau}$ symmetry is anomaly free and therefore can be gauged. However, from the oscillation of tau or muon neutrinos to electron neutrinos, we know this symmetry is not respected in nature and needs to be broken. The consequence of the breaking is a neutral, color singlet, massive Z' that couples only to muon number and tau number at tree level. The interactions of Z' are described by the Lagrangian below,

$$\mathcal{L} \ni -\frac{1}{4}(Z')_{\alpha\beta}(Z')^{\alpha\beta} + \frac{1}{2}M_{Z'}^2 Z'^\alpha Z'_\alpha$$
$$- \epsilon g_1 Z'_\alpha(\bar{\ell}_2\gamma^\alpha\ell_2 + \bar{\mu}\gamma^\alpha\mu - \bar{\ell}_3\gamma^\alpha\ell_3 - \bar{\tau}\gamma^\alpha\tau), \quad (6)$$

where $Z'_{\alpha\beta} = \partial_\alpha Z'_\beta - \partial_\beta Z'_\alpha$ is the field strength tensor, and $\ell_2 = (\nu_\mu, \mu)^T$, $\ell_3 = (\nu_\tau, \tau)^T$. As shown in the Lagrangian, the Z' has the same coupling to the left-handed and right-handed muon (tau), with a relative minus sign between the coupling of muons and taus [30]. The new parameters in the model are $\alpha = (M_{Z'}, \epsilon)$, where ϵg_1 is the coupling of Z' to muons and taus.

Studying this model is important because some region of its parameter space can explain the long persisting discrepancy in the SM prediction and experimental measurement of the muon anomalous magnetic moment $(g-2)_\mu$ [36–44]. Some anomalies observed in B physics and flavor changing Higgs coupling [45–47] can also be explained by gauged $L_\mu - L_\tau$, which further motivates studying this model. In particular, the anomalies recently observed in $R_K = \mathrm{Br}(B \to K\mu^+\mu^-)/\mathrm{Br}(B \to Ke^+e^-) \simeq 0.745$ with 2.6σ discrepancy between theoretical expectations, and in $R_{K^*} = \mathrm{Br}(B \to K^*\mu^+\mu^-)/\mathrm{Br}(B \to K^*e^+e^-) \simeq 0.7$ with $\sim 2.5\sigma$, by LHCb [48] can be explained by the $L_\mu - L_\tau$ model, assuming $M_{Z'} \sim O(\mathrm{TeV})$ and $g_{Z'} \sim O(1)$ [45,46,49].

Because the $L_\mu - L_\tau Z'$ only interacts with second and third generation leptons at tree level, it is not very constrained (see Ref. [31] for the quantitative estimation of the constraints on such leptophilic Z'). One constraint comes from $Z - Z'$ mixing that arises from loops of muons and taus (and their respective neutrinos), inducing a coupling of $O(10^{-3}\epsilon)$ between the Z' and all fermions [50–52]. The factor of 10^{-3} is a rough estimate based on the loop suppression and the couplings of Z' and Z with muon and tau. Precision measurements of Z-electron coupling at the *BABAR* and Belle II experiment [53–58] require the coupling of Z' to electrons to be $\lesssim 10^{-3}$ for $10\,\mathrm{MeV} < M_{Z'} < 10\,\mathrm{GeV}$, which translates to $\epsilon \lesssim 1$ in the $L_\mu - L_\tau$ model.

The strongest current bound on $L_\mu - L_\tau$ for $M_{Z'} < 10$ GeV is from fixed target neutrino beam experiments. In particular, CHARMII [59] and CCFR [60,61] tightly constrain $L_\mu - L_\tau Z'$ via the trident process: $N + \nu_\mu \to N + \nu_\mu + \mu\mu$ [62]. This process occurs through the exchange of off-shell $W^\pm/Z$ bosons in the SM. However, in the $L_\mu - L_\tau$ model, the exchange of Z' can significantly contribute to the rate of the process, especially if the Z' is produced on shell [59,60,62,63]. The neutrino trident experiment excludes a $L_\mu - L_\tau Z'$ with $\epsilon \gtrsim 0.005$ and $M_{Z'} \lesssim 1$ GeV. The bounds loosen for heavier Z' to $\epsilon \sim 0.05$ for $M_{Z'} = 20$ GeV.

LHC bounds on this model come from recasting the $pp \to Z \to 4\mu$ searches by CMS and ATLAS [64,65]. This bound surpasses the trident bound for 10 GeV $< M_{Z'} \lesssim 45$ GeV, excluding $\epsilon \gtrsim 0.04$ for $M_{Z'} = 10$ GeV, and a looser bound of $\epsilon \gtrsim 0.5$ for $M_{Z'} \sim 45$ GeV [62]. In Ref. [66], we discussed the potential LHC reach using a dedicated cut-and-count Z' analysis in the $pp \to Z \to 4\mu$ (for $M_{Z'} > 2m_\mu$) and $pp \to \mu^+\mu^- \not{E}_T$ (for $M_{Z'} < 2m_\mu$) channels. These channels were proposed for their cleanliness. Additionally, since the contribution of Z' is greatest when it is produced on shell, the mass ranges of the channels were chosen such that an on-shell Z' can decay to muons ($pp \to 4\mu$) or neutrinos ($pp \to 2\mu\not{E}_T$). Although in [66] a large region of parameter space could be uncovered with the cut-and-count method after the full 3 ab^{-1} of HL-LHC run, in this paper we show that our sensitivity can be enhanced further if we use the matrix element technique.

IV. LHC BOUNDS ON THE $L_\mu - L_\tau$ MODEL USING THE MATRIX ELEMENT METHOD

A. Looking for Z' with mass range $2m_\mu < M_{Z'} < M_Z$ in $pp \to Z \to 4\mu$

The rare process $Z \to 4\mu$ occurs through Z boson decay into two muons, one of which radiates a neutral boson V that subsequently splits into a second pair of muons. In the SM, V can be an off-shell Z or photon, while in the $L_\mu - L_\tau$ model, the on-shell/off-shell Z' also contribute (see Fig. 1).
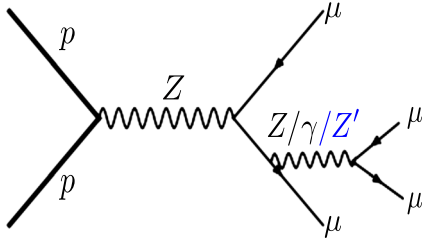
FIG. 1.   The Feynman diagram for $Z \to 4\mu$ in the SM as well as the Z' contribution.

Four muons reconstructing an on-shell Z boson is a clean process that has been studied extensively. Therefore, it is an ideal channel for constraining the $L_\mu - L_\tau$ model. Moreover, the related channels $pp \to Z \to 4e$ and $pp \to Z \to 2e2\mu$ can be used as a background control sample to suppress the experimental uncertainties of this channel. Within the SM the cross section of $Z \to 4\mu$ is the same as the $Z \to 4e$ and $Z \to 2e2\mu$ processes up to $O(m_e^2/m_\mu^2) \sim 10^{-4}$; hence by measuring $Z \to 4e$ and $Z \to 2e2\mu$, we can obtain a precise prediction of $(pp \to Z \to 4\mu)_{\rm SM}$. As such, in the following, we assume that the systematic uncertainties of $(pp \to Z \to 4\mu)_{\rm SM}$ are subpercent.

To study the $L_\mu - L_\tau$ Z', we generated a Universal FeynRules Output model [67] using FeynRules [68]. We then fed the model to MadGraph [69,70] to generate our events at leading order. The background sample only contains the SM contributions to the four muon production $(pp \to 4\mu)_{\rm SM}$, while the signal event sample includes both the Z' as an intermediate state (where its width is calculated using MadGraph) *and* the SM gauge bosons $(pp \to 4\mu)_{\rm SM+Z'}$ to capture the interference among processes. The signal events were generated for various values of $\alpha^{\rm gen} = (M_{Z'}^{\rm gen}, \epsilon^{\rm gen})$ between $2m_\mu \leq M_{Z'}^{\rm gen} \leq 40$ GeV and $0.001 \leq \epsilon^{\rm gen} \leq 0.1$.

Before applying the MEM, we impose some preliminary cuts to ensure the events have been triggered upon. Specifically, we impose a dilepton trigger used in LHC-13 [71] that selects events with $p_T(\mu_1) > 17$ GeV and $p_T(\mu_2) > 8$ GeV, where μ_1 is the leading muon and μ_2 is the subleading one. We also require $p_T > 4$ GeV for all muons and the separation $\Delta R_{\mu\mu} > 0.05$ [64,65,71]. To ensure $\sqrt{\hat{s}} \sim M_Z$, we impose 76 GeV $< m_{4\mu} <$ 106 GeV and veto extra jets in the event, since they would hurt the cleanliness of the process.

As lepton momenta are accurately measured at the detector, we can be confident that the detected momenta very closely represent the partonic level momenta; stated in terms of transfer functions introduced in Eq. (3), we assume $W(p^{\rm rec}, p^{\rm par}) = \delta(p^{\rm rec} - p^{\rm par})$. Furthermore, because the final states are fully reconstructable and we have vetoed jets,[3] we can find the initial state energies [or equivalently

---

[3]We have done our analysis at parton level, but given $m_{4\mu} \sim M_Z$, it is reasonable to assume jet contamination is negligible.

the x_is in Eq. (1)] from conservation of energy and momentum, as shown in Eq. (2). As a result, all of the integrations collapse due to delta functions, and the likelihood ratio becomes

$$\mathcal{L}(p^{\rm rec}; M_{Z'}, \epsilon) = \frac{\frac{1}{\sigma}\frac{f(x_1)f(x_2)}{2sx_1x_2}|\mathcal{M}(p^{\rm rec}; M_{Z'}, \epsilon)_{q(x_1)q(x_2)\to 4\mu}|^2}{\frac{1}{\sigma}\frac{f(x_1)f(x_2)}{2sx_1x_2}|\mathcal{M}(p^{\rm rec}; {\rm SM})_{q(x_1)q(x_2)\to 4\mu}|^2}.$$

Because the initial state quarks (with energies parametrized by x_1 and x_2) are the same for both the signal and the background at leading order, the PDF and x_i factors cancel in the ratio. Hence, the likelihood ratio for a single event can be simplified to

$$\mathcal{L}(p^{\rm rec}; M_{Z'}, \epsilon) = \frac{|\mathcal{M}(p^{\rm rec}; M_{Z'}, \epsilon)|^2}{|\mathcal{M}(p^{\rm rec}; {\rm SM}|^2}. \tag{7}$$

The analytical calculation of $|\mathcal{M}|^2$ can be simplified if we just consider the process $Z \to 4\mu$, where Z is the vector sum of the four muons. Although the events could also be due to the processes $qq \to \gamma^* \to 4\mu$ or the interference of γ^* and Z mediators, requiring 76 GeV $< m_{4\ell} <$ 106 GeV assures us that $|\mathcal{M}_{Z\to 4\mu}|^2$ is a reasonably accurate description of the events. For the analytic evaluation of $|\mathcal{M}|^2$, we use the form given in MCFM [72–74], which is easily modified to include Z' intermediate states (for the case of the signal).

Even when all final state particles are reconstructed, there are complications in evaluating $|\mathcal{M}|^2$. Specifically, at the detector level, we do not know the right pair of muons that reconstruct the $V = \gamma/Z/Z'$, leading to a combinatorics problem. The most naive way we can account for this problem is by summing over all four possible $|\mathcal{M}|^2$ with different combinations of muon pairs originating from V. With a better algorithm, the sensitivity to signal may be further improved, but we do not attempt that here.

The likelihood ratio, as shown in Eq. (7), can be calculated for each event as a function of $M_{Z'}$, and ϵ. Before proceeding, we need to differentiate between the analysis $\alpha = (M_{Z'}, \epsilon)$ values used in the signal matrix element hypothesis $[\mathcal{M}(p^{\rm rec}, M_{Z'}, \epsilon)]$ and the "truth" values—the values of $M_{Z'}$ and ϵ used to generate the signal events (which, in actual data, would be unknown). To avoid confusion, we use $\alpha^a = (M_{Z'}^a, \epsilon^a)$ for the analysis values and $\alpha^{\rm gen} = (M_{Z'}^{\rm gen}, \epsilon^{\rm gen})$, which were used for event generation.

Our goal is to determine the values of $M_{Z'}^a$ and ϵ^a that maximize our sensitivity to the signal hypothesis. Keeping this task in mind, it is worth looking at how the α^a parameters enter into the matrix element,

$$\mathcal{M}(p^{\rm rec}; M_{Z'}^a, \epsilon^a) = \mathcal{M}_{\rm SM} + \frac{(\epsilon^a)^2}{M_{\mu\mu}^2 - (M_{Z'}^a)^2 - i\Gamma_{Z'}^a M_{Z'}^a}(\cdots),$$
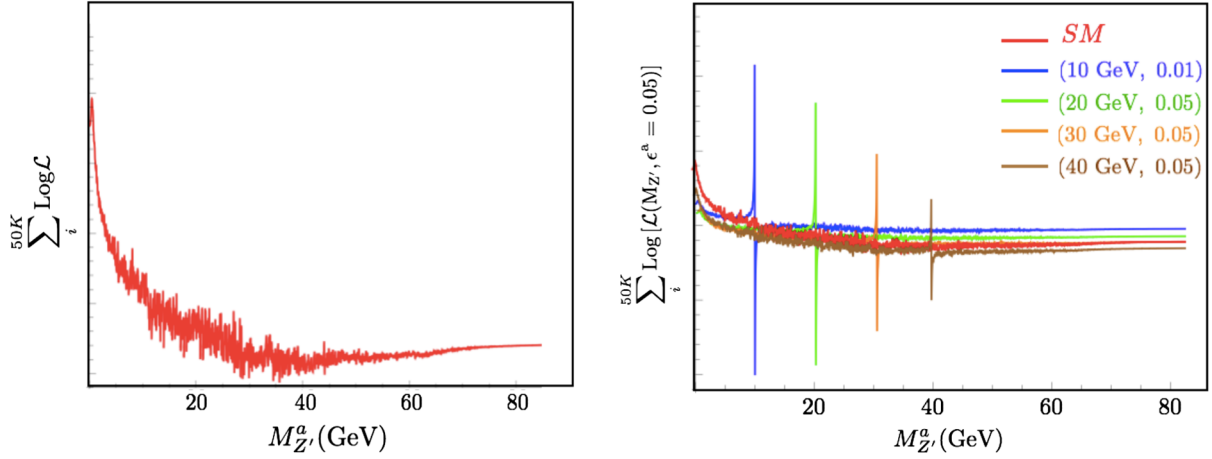
$$\tag{8}$$

FIG. 2. The plot of log likelihood ratio summed over 50000 events is shown above. The left plot is the distribution of SM events, where there is a spike at every reconstructed $M_{\mu\mu}$. The peak at $M_{Z'}^a \sim 0$ GeV is due to the photon propagator. For larger values of $M_{Z'}^a$, there are fewer events with $M_{\mu\mu} = M_{Z'}^a$, and the function looks smoother, which is the result of our basic cuts. In the right plot, the distribution of the signal events compared to the simulated background events is shown. In the signal events, which include both SM and Z' contributions, there is a peak at $M_{Z'}^a = M_{Z'}^{\mathrm{gen}}$ because of the contribution from the on-shell Z'. It is important to mention that the optimal value of $M_{Z'}^a$ is obtained by comparing the relative shapes of the signal plot with the background one. The scale on the y axis is irrelevant because it is highly sensitive to the number of events in our sample, and by increasing the sample size the numerical values on the y axis of the plots become more comparable. For our analysis, we manually brought the plots to similar y-axis values, to compare their shapes.

where $\Gamma_{Z'}^a = \Gamma_{Z'}(M_{Z'}^a, \epsilon^a)$ and the ellipses represent the part of the matrix element that is independent of $M_{Z'}^a, \epsilon^a$. Knowing the dependence of $\mathcal{M}$ on α^a can help us better understand the behavior of the likelihood ratio for various α^a values. For example, the appearance of $M_{Z'}^a$ only in the denominator suggests that the likelihood ratio is very sensitive to the value of $M_{Z'}^a$ (e.g., the possibility of resonance). With ϵ^a in the numerator, we suspect the likelihood ratio to not be as sensitive to a particular value of ϵ^a, because regardless of which testing hypotheses the events belongs to, their likelihood ratio will increase by increasing ϵ^a. Although $\Gamma_{Z'}^a$ in the denominator also depends on ϵ^a, its contribution is suspected to be much smaller than the ϵ^a in the numerator.

To examine these conjectures and hunt for the α^a that maximize the likelihood ratio, we proceed by plotting $\sum_i \log[\mathcal{L}(p_i, \alpha^a)]$ defined in Eq. (5) as a function of α^a. Thereby, for each α^a we can compare the relative sizes of the summed log likelihood ratio of the signal sample events with that of background sample, and look for the values of α^a that maximize the likelihood of the signal with respect to the background. As we have two new parameters $(\alpha^a = M_{Z'}^a, \epsilon^a)$, for simplicity we fix one of the parameters and plot Eq. (5) as a function of only one parameter.

First, we assume a nonzero value for ϵ^a and plot Eq. (5) for a range of $M_{Z'}^a$ values. The value of ϵ^a can be any arbitrary nonzero value, and varying ϵ^a does not alter our results. For every event, the reconstructed invariant mass of two muons $(M_{\mu\mu})$ will produce a spike at $M_{Z'}^a = M_{\mu\mu}$, which is not necessarily equal to the $M_{Z'}^{\mathrm{gen}}$. In fact, because we are summing over all four possible combinations of muon pairs,

we have four spikes in $M_{Z'}^a$ for each event. Therefore, summing over all events, the net distribution (in our case, 50000 events) has spikes at all values of $M_{Z'}^a \lesssim M_Z$ as shown in the left plot of Fig. 2. In the signal sample, due to the higher number of events with $M_{\mu\mu} = M_{Z'}^{\mathrm{gen}}$, the spike at $M_{Z'}^{\mathrm{gen}}$ is more noticeable (see Fig. 2, right plots). The maximum likelihood ratio of the signal events is consistently at $M_{Z'}^a = M_{Z'}^{\mathrm{gen}}$ for any arbitrary value of $\epsilon^a \neq 0$, while the pure background events show no interesting behavior at $M_{Z'}^a = M_{Z'}^{\mathrm{gen}}$.

The height of the peak at $M_{Z'}^a = M_{Z'}^{\mathrm{gen}}$ relative to the other spikes depends on ϵ^{gen} (regardless of ϵ^a). The peak is more visible for larger values of ϵ^{gen}, but becomes less distinguishable from other spikes for smaller values of ϵ^{gen}. As one example, for $\epsilon^{\mathrm{gen}} \lesssim 0.01$ for $M_{Z'}^{\mathrm{gen}} \gtrsim 10$ GeV, we could not distinguish the peak at $M_{Z'}^a = M_{Z'}^{\mathrm{gen}}$ from other spikes). Given that our analysis highly depends on whether we are able to find $M_{Z'}^{\mathrm{gen}}$ by plotting $\sum_i \log[\mathcal{L}(p_i, M_{Z'}^a, \epsilon_{\mathrm{arbitrary}}^a \neq 0)]$ as a function of $M_{Z'}^a$, our strategy needs to bifurcate depending on the value of ϵ^{gen}.

  (i) Large ϵ^{gen}: we determine the value of $M_{Z'}^a$ by observing the peak at $M_{Z'}^a = M_{Z'}^{\mathrm{gen}}$.

  (ii) Small ϵ^{gen}: we need to chose an arbitrary value of $M_{Z'}^a$.

However, as the choice of the strategies depends on whether we are able to *observe* the peak at $M_{Z'}^a = M_{Z'}^{\mathrm{gen}}$, rather than trying to quantify what it takes to observe a peak, we study both strategies for all values of ϵ^{gen}. More specifically, we first assume that we can find $M_{Z'}^a = M_{Z'}^{\mathrm{gen}}$ even for small values of ϵ^{gen}; then we fix an arbitrary value
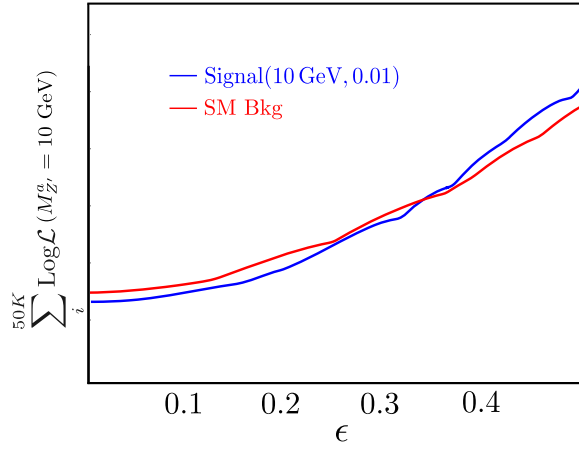
FIG. 3. The distribution of the log likelihood ratio for the signal and background events summed over 50000 sample events. We have fixed $M_{Z'}^a = M_{Z'}^{\text{gen}} = 10$ GeV while $\epsilon^{\text{gen}} = 0.01$. The signal and the background MC sample events increase as ϵ^a increases, providing us with no special ϵ^a value that optimizes the signal likelihood ratio. The numerical value on the y axis of the two plots as explained earlier is irrelevant and thus not shown.

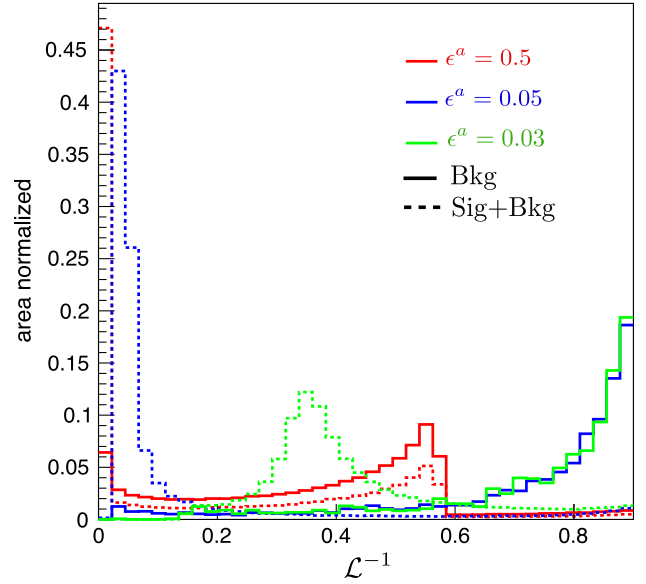FIG. 4. The distribution of $\mathcal{L}^{-1} = |\mathcal{M}_{\text{SM}}|^2/|\mathcal{M}_{Z'+\text{SM}}|^2$ for $M_{Z'}^{\text{gen}} = 10$ GeV, and $\epsilon^{\text{gen}} = 0.05$. This plot shows that for $M_{Z'}^a = 10$ GeV, if we have $\epsilon^a = 0.03$, the resonance region is not very well separated from the background region. If $\epsilon^a = 0.5$, although the signal is far from 1, the background also spreads more. The middle value $\epsilon^a = 0.05$ is the optimal one.

for $M_{Z'}^a$ and study the signal assuming we cannot find $M_{Z'}^{\text{gen}}$ even for large values of ϵ^{gen}. The first method represents the best we can do with the MEM, while the second represents a more realistic reach for $\epsilon \lesssim 0.01$.

Now that we have discussed how to find the optimal value of $M_{Z'}^a$, we follow the same procedure to determine ϵ^a. For simplicity, we fix $M_{Z'}^a = M_{Z'}^{\text{gen}}$ and let ϵ^a be a free parameter. The $\sum_i \log[\mathcal{L}(p_i, M_{Z'}^a = M_{Z'}^{\text{gen}}, \epsilon^a)]$, where the sum is over 50000 MC generated events for each sample, is plotted in Fig. 3 below as a function of ϵ^a. Both the signal and background MC events increase as a function of ϵ^a. Such behavior is anticipated, because as we can see in Eq. (8), larger ϵ results in larger $|\mathcal{M}(p^{\text{rec}}; M_{Z'}, \epsilon)|^2$ and thus a larger likelihood ratio for any event. Stated another way, fixing $M_{Z'}^a$ and studying the likelihood ratio with respect to ϵ^a is not useful because the simulated signal and background events behave the same way as a function of ϵ^a. Thus, for the remainder of this section, we explore an alternative different technique for estimating the optimal α^a.

The problem with the approach of $\mathcal{L}$ maximization is that, for a given α^a, it compares only two numbers: the summed log likelihood ratio of the simulated signal events vs that of background events. With this approach, we are throwing away a lot of information about each event. We can get a better handle on the signal vs the background, if we look at their *distributions* for a given α^a and how those distributions evolve as we vary α^a. The distribution illustrates to us the behavior of events for a given α^a, rather than just their sum. We can use this distribution to find α^a that results in best discrimination of the signal from the background, even if the sum over events is similar. The

signal usually resides in high values of the likelihood ratio, as expected according to the definition of the likelihood ratio [Eq. (4)]. So, we distribute the events in the inverse likelihood ratio ($\mathcal{L}^{-1}$) to be able to see the excess of the signal spread over a narrower window [0, 1].[4]

As an example of the α^a distribution method, in Fig. 4 below we show the distribution of the MC generated events for the signal and the background for different choices of ϵ^a, assuming $M_{Z'}^a = M_{Z'}^{\text{gen}}$. We can see that as we increase ϵ^a, the resonance region, defined as events with $M_{\mu\mu} \sim M_{Z'}^{\text{gen}}$, is more separated from the background. However for very large ϵ^a (for example $\epsilon^a = 0.5$ in Fig. 4), the background distribution also becomes more spread out. This is again consistent with Eq. (8), because no matter whether the events are at resonance or not, increasing ϵ^a increases their likelihood ratio.

The optimal value of ϵ^a depends on the distribution of the signal with respect to the background. For a fixed sample size of signal and background simulated events, optimal is defined as maximization of

$$\frac{\text{number of signal events}}{\sqrt{\text{number of background events}}}.$$

---

[4]Not all of the background will fall in this [0, 1] region; however, we only care about how the background is distributed in the region where the signal resides.
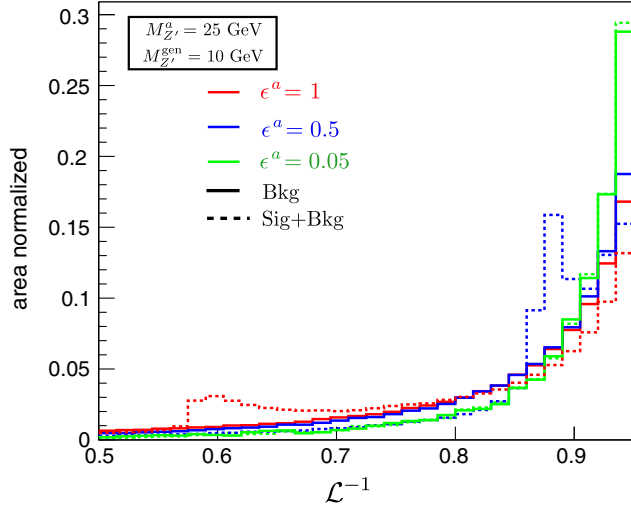
FIG. 5. The distribution of $\mathcal{L}^{-1} = |\mathcal{M}_{SM}|^2/|\mathcal{M}_{Z'+SM}|^2$ for $M_{Z'}^{gen} = 10$ GeV, and $\epsilon^{gen} = 0.05$ is shown, where we are assuming $M_Z^a = 25$ GeV, and we vary ϵ^a. Here, a relatively large ϵ^a is needed to separate the signal from background. However, with increasing ϵ^a, we also get broader peaks and therefore our sensitivity is not good as when $M_{Z'}^a = M_{Z'}^{gen}$. In this particular example $\epsilon^a = 0.5$ is the optimal choice.

After examining a few values[5] of ϵ^a, we find $\epsilon^a = 0.05$ to give the best discrimination, regardless of the ϵ^{gen}. This result is consistent for all values of $M_{Z'}^{gen}$ that we studied.

Thus far, we have seen that with the assumption that we know $M_{Z'}^{gen}$, the optimal values for α^a are $M_{Z'}^a = M_{Z'}^{gen}$ and $\epsilon^a = 0.05$. However, if we cannot find the value of $M_{Z'}^{gen}$ by maximizing $\sum_i \log[\mathcal{L}(p_i, \alpha^a)]$ with respect to $M_{Z'}^a$, we must explore how our sensitivity changes if $M_{Z'}^a$ is fixed to an arbitrary value $\neq M_{Z'}^{gen}$. Because the Z' mass for the generated events is no longer the same as the Z' mass used in the analysis matrix element, the events with $M_{\mu\mu} \sim M_{Z'}^{gen}$ are no longer at resonance in the matrix element $\mathcal{M}(p^{rec}; M_{Z'}, \epsilon)$ [Eq. (8)]. Therefore, when we go to search for the optimal ϵ^a using distributions, the signal and background are less separated than in the previous case. To increase the separation, we need to increase ϵ^a. Consequently, the optimal ϵ^a is no longer fixed at 0.05 and depends on the difference between $M_{Z'}^a$ and $M_{Z'}^{gen}$. For example, in Fig. 5, the distribution of the signal and the background MC generated events for $M_{Z'}^a = 25$ GeV and $M_{Z'}^{gen} = 10$ GeV, with $\epsilon^a = 0.05, 0.5, 1$, are shown. In this example, we see that $\epsilon^a = 0.05$ does not give a good discrimination of the signal from the background, and we have to use larger values of ϵ^a. Furthermore, as the plot illustrates, although large ϵ^a increases the separation between the signal region and the background, it broadens

---

[5]The benchmark points ϵ^a we studied are $\epsilon^a \in [0.01, 0.5]$ with increments of 0.01. With a larger sample of ϵ^a, the best value of ϵ^a may slightly vary.

TABLE I. Determining $\alpha^a = (M_{Z'}^a, \epsilon^a)$ using a combination of likelihood maximization and the $\mathcal{L}^{-1}$ distribution.

$M_{Z'}^a = M_{Z'}^{gen}$	$\epsilon^a = 0.05$
Using likelihood maximization	
$M_{Z'}^a$ arbitrary	$\epsilon^a > 0.05$
Likelihood maximization fails, because of small e^{gen}	

the signal region. Because we are interested in distinguishing the region where the signal to background ratio is maximized, having a broad signal region is not ideal. Hence, we expect our reach for $M_{Z'}^a \neq M_{Z'}^{gen}$ to be more limited that the previous case, even with the most optimal ϵ^a.

Table I recapitulates the value ϵ^a depending on whether or not plotting $\sum_i \log[\mathcal{L}(p_i, \alpha^a)]$ as a function of $M_{Z'}^a$ can tell us M_Z^{gen}.

To quantify the reach of the strategy proposed here, let us define the following variables:

$$S = \text{Luminosity} \times (\sigma(pp \to 4\mu)_{SM+Z'} - \sigma(pp \to 4\mu)_{SM}),$$

$$B = \text{Luminosity} \times (\sigma(pp \to 4\mu)_{SM}).$$

The cuts on $\mathcal{L}^{-1}$ are imposed such that we get the maximum $S/\sqrt{B}$ with integrated luminosity of 300 fb^{-1}. We also require $S > 10$ to avoid confusion of signal events for statistical fluctuations. We find that, by using the MEM approach we are able to get $S/\sqrt{B} \geq 3$ up to $\epsilon^{gen} \sim 0.002$ and $M_{Z'} \lesssim 20$ GeV, and up to $\epsilon^{gen} \sim 0.005$ for 20 GeV $< M_{Z'}^{gen} \lesssim 40$ GeV, provided that we can determine the true value of $M_{Z'}^{gen}$ by conventional MEM means. This best case scenario is indicated with the dashed red line in Fig. 6. The dashed-dotted brown line in Fig. 6 shows our reach assuming $M_{Z'}^a = 2$ GeV while optimizing ϵ^a. Similarly, the dashed brown line is for $M_{Z'}^a = 10$ GeV, and the dotted green line is for $M_{Z'}^a = 25$ GeV. As expected, the lines with arbitrary $M_{Z'}^a$ touch the dashed red line for $M_{Z'}^a = M_{Z'}^{gen}$, and have a relatively good sensitivity when $M_{Z'}^a \sim M_{Z'}^{gen}$, but their sensitivity declines as $M_{Z'}^a$ moves away from $M_{Z'}^{gen}$. The current constraints from *BABAR* [75], CCFR [60,61] experiments and LHC-8 are also shown in Fig. 6 in black and purple solid lines, respectively. The expected exclusion bound (3σ) from LHC-14, with 300 fb^{-1} integrated luminosity, using the cut-and-count method that is explored in Ref. [66] is the dashed blue line in Fig. 6. With the assumption that we know $M_{Z'}^a = M_{Z'}^{gen}$, our reach with MEM is about a factor of 10 greater than the canonical cut-and-count method, and even with $M_{Z'}^a$ fixed to an arbitrary value, we have an improved sensitivity for some range of $M_{Z'}^{gen}$ compared with the cut-and-count method. By trying several values of $M_{Z'}^a$ and choosing
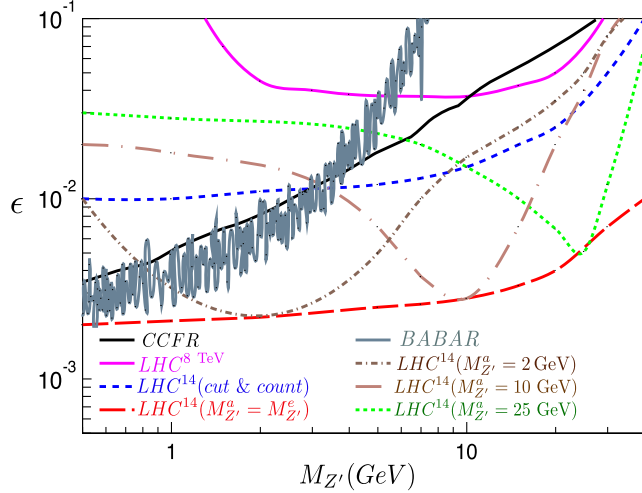
FIG. 6. The bound on $L_\mu - L_\tau$ from different experiments. The purple line is the LHC-8 bounds, the black line is the CCFR bound, and the solid gray line is the bound from the *BABAR* experiment [75]. The dashed blue line is the bound from LHC-14 with luminosity 300 fb^{-1} up to 3σ with the cut-and-count method [66]. The red dashed line in the figure shows the reach with MEM, but assuming we know the mass of Z', and with $\epsilon^a = 0.05$. This is the best sensitivity we could get using MEM. The dotted green line is assuming $M_{Z'}^a = 25$ GeV, the dashed brown line is $M_{Z'}^a = 10$ GeV, and the dotted-dashed dark brown line is $M_{Z'}^a = 2$ GeV while choosing the optimal value for ϵ^a. The MEM bounds are also with luminosity 300 fb^{-1} up to 3σ.

$\epsilon^a \sim 0.05$, one can achieve a sensitivity near the red curve shown in Fig. 6 at the LHC.

B. Looking for Z' with mass range $M_{Z'} < 2m_\mu$ in $pp \to \mu^+\mu^- \not{E}_T$

If $M_{Z'} < 2m_\mu$, the decay of on-shell Z' to muons is kinematically forbidden. In fact, for this mass range, on-shell $L_\mu - L_\tau Z'$ can only go to neutrinos. While we could hunt for off-shell Z' in this mass range using the 4μ final state (as in the previous section), off-shell Z' production is suppressed relative to on-shell production by two additional powers of ϵ. Therefore, as proposed in [66], we give up the benefits of a completely visible final state in favor of a larger rate and hunt for Z' in $pp \to \mu^+\mu^- \not{E}_T$. The signal contribution to $pp \to \mu^+\mu^- + \not{E}_T$ comes from the production of a pair of muons, one of which radiates an on-shell Z' that decays to a pair of neutrinos[6]: $pp \to \mu^+\mu^- Z' \to \mu^+\mu^- \nu_\ell\bar{\nu}_\ell$.

---

[6]The signal also captures the diagram where the neutrinos and muons are produced in opposite order: a pair of neutrinos is first produced, and then one of the neutrinos emits a Z' which splits to two muons. In this topology, the Z' must be off shell, so the cross section is suppressed by more powers of ϵ and is, therefore, negligible in the parameter space of interest.
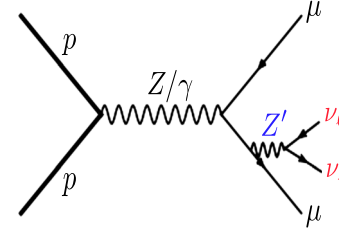

FIG. 7. The main Feynman diagram that contributes to the signal. The sources of missing energy are shown in red.

Due to the presence of missing energy, we can no longer impose an invariant mass cut on the final state. As a consequence, some new, important backgrounds emerge,

$$pp \to \tau^+\tau^-|_{\text{dimuon decay}},$$
$$pp \to VV = \begin{cases} W^+W^-|_{\text{dimuon decay}}, \\ Z(Z/\gamma), \end{cases}$$
$$pp \to \mu^+\mu^- + \text{jets}. \tag{9}$$

The last background arises as a result of jet mismeasurements and pileup, and so peaks at low values of $\not{E}_T$. The Feynman diagrams of the rest of the processes are shown in Fig. 8. Ditau production is the largest irreducible background, followed by diboson production. Technically, we include both resonant and nonresonant contributions in this category, as the latter can be non-negligible. Thus, the VV background is more accurately described as $(pp \to \mu^+\mu^-\nu_\ell\bar{\nu}_\ell)_{\text{SM}}$. Similarly, the signal in this section is defined as $pp \to \mu^+\mu^-\nu_\ell\bar{\nu}_\ell$ including Z' as a possible intermediate state: $(pp \to \mu^+\mu^-\nu_\ell\bar{\nu}_\ell)_{\text{SM}+Z'}$. As in the previous section, the signal is defined including SM contributions to incorporate interference.

To study this channel in more detail, we generated events for $pp \to \tau^+\tau^- \to \mu^+\mu^- + \not{E}_T$, $(pp \to \mu^+\mu^-\nu_\ell\bar{\nu}_\ell)_{\text{SM}}$, and $(pp \to \mu^+\mu^-\nu_\ell\bar{\nu}_\ell)_{\text{SM}+Z'}$ via the MC chain `MadGraph5-aMC@NLO` [70] plus `Pythia 6.4` [76], where the latter step is used here to decay the taus.[7] Before any MEM analyses, we require all events to pass the same dilepton trigger requirement used in Sec. IV A [$p_T(\mu_1) > 17$ GeV and $p_T(\mu_2) > 8$ GeV, where μ_1 is the leading muon and μ_2], veto any jets with $p_T > 20$ GeV, $\eta < 2.5$, and impose a minimum missing energy cut of $\not{E}_T > 20$ GeV. The last cut is imposed to suppress the $pp \to \mu^+\mu^- + \text{jets}$ background.

After the initial set of cuts, muonic tau production $pp \to \tau^+\tau^- \to \mu^+\mu^- \not{E}_T$ is our main irreducible background. The cross section of the $\tau^+\tau^-|_{\text{dimuon}}$ background is roughly 2 orders of magnitude larger than $\sigma(pp \to \mu^+\mu^-\nu_\ell\bar{\nu}_\ell)_{\text{SM}}$. Therefore, if we want to have any chance to be sensitive to a

---

[7]We restricted the possible τ decays to leptonic channels only within `Pythia 6.4` to make event generation more efficient.
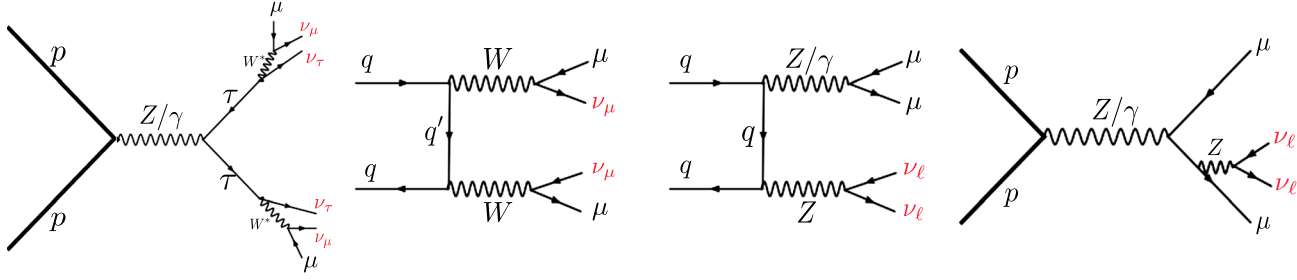
FIG. 8.   The main irreducible backgrounds to the dimuon + MET process. The sources of missing energy are shown in red. The left-most diagram is the $\tau^+\tau^-|_{\mathrm{dimuon}}$, which is the main background to our signal. If the neutrinos are the same (e.g., $\nu_\ell = \nu_\mu$), the right three diagrams interfere, and thus we must use $(pp \to \mu^+\mu^-\nu_\ell\bar{\nu}_\ell)_{\mathrm{SM}}$ to describe them. The diagram with Z' (shown in Fig. 7) may also interfere with the right three diagrams and thereby, the signal is described by $(pp \to \mu^+\mu^-\nu_\ell\bar{\nu}_\ell)_{\mathrm{SM}+Z'}$.

Z' signal, we need to first make the $\tau^+\tau^-|_{\mathrm{dimuon}}$ background more manageable. So, instead of trying to discriminate Z' signal against background, we focus on distinguishing $\tau^+\tau^-|_{\mathrm{dimuon}}$ from other processes.

The most efficient way to eliminate the $\tau^+\tau^-|_{\mathrm{dimuon}}$ background is by using a variable that is most faithful to $\tau^+\tau^-|_{\mathrm{dimuon}}$ background and thus localizes its simulated events to a small region. Motivated by the benefits of the MEM discussed previously, we use the $|\mathcal{M}_{\tau\tau}|^2$, the squared matrix element of the muonic ditau production, as the discriminating variable. For each event we calculate $|\mathcal{M}_{\tau\tau}|^2$ using the observed final state momenta, then search for and select out regions (using MC) of $|\mathcal{M}_{\tau\tau}|^2$ that $pp \to \tau^+\tau^-$ does not populate. In doing this, we are not following the traditional MEM in this section, because we are not using $|\mathcal{M}|^2$ of the signal to distinguish that from other processes. Rather, we are only using the $|\mathcal{M}|^2$ of (part of) the background, which makes our approach independent of the signal (Z') model.

As a further deviation from the traditional MEM, we weight each event by only one squared matrix element, $|\mathcal{M}_{\tau\tau}|^2$, rather than two (a signal hypothesis and a background hypothesis). We may lose some discriminating power by not calculating the likelihood ratio as described in Sec. II, but our approach is more time efficient as we do not have to deal with other squared matrix elements that contain missing energy.

Even after reducing our discriminant to the evaluation of a single $|\mathcal{M}|^2$ for each event, evaluating $|\mathcal{M}_{\tau\tau}|^2$ is still a difficult task. In $\tau^+\tau^-|_{\mathrm{dimuon}}$ production, there are four sources of missing energy, which translates to 12 unknown momenta. Moreover, because we do not know the energies of the initial state, we have in total 14 unknowns. Energy-momentum conservation $\delta^4(p_{\mathrm{initial}} - p_{\mathrm{final}})$ reduces the number of unknowns to 10, but ten integrations for each event is still extremely computationally cumbersome. Thankfully, the specific topology of the $\tau^+\tau^-|_{\mathrm{dimuon}}$ production can help us approximate the unknowns. The list of our assumptions to approximate the unknowns is as follows.

(I)   We assume that the τs were produced on shell. We know that the invariant mass of an on-shell tau decay product is the tau mass. Therefore, we can determine two unknowns from this assumption since there are two τs in each process.

(II)   Instead of calculating $\tau \to \nu_\tau\bar{\nu}_\mu\mu$, we replace the two neutrinos with one massive scalar[8] neutrino (νs for the notation), and we calculate $\tau \to \mu\nu s$.
Three body decay has different kinematic distributions compared to two body decays, but in this approach we can reduce our number of unknowns by 4, leaving four remaining. Because we are no longer dealing with the actual $|\mathcal{M}|^2$ and we are calculating $pp \to \tau^+\tau^- \to \mu^+\mu^-\nu s\bar{\nu}s$, we refer to the matrix element we calculate as the "modified" $|\mathcal{M}_{\tau\tau}|^2$, or $|\mathcal{M}_{\tau\tau}^{\mathrm{mod}}|^2$.

(III)   The tau pair can be produced from either a photon or a Z boson. However, as a result of our basic cuts (dilepton trigger and $\not{E}_T > 20$ GeV), we can be confident that the production of $\tau^+\tau^-$ is dominated by Z exchange. Therefore, we assume that the taus are produced from an on-shell Z, and so $\sqrt{\hat{s}} = M_Z$. This assumption eliminates another unknown.

(IV)   Based on the previous assumptions, we expect the tau decay products to be nearly collinear. Consequently, the η and ϕ of the νs_i should be close to η and ϕ of the corresponding μ_i, with the subscript i defined as follows: $\tau_i \to \mu_i\nu s_i$. For our analysis, we assume $\eta(\mu_i) = \eta(\nu s_i)$, and $\Delta\phi(\mu_i, \nu s_i) = \epsilon_\phi^i$ where $\epsilon_\phi \ll 1$. Therefore, we only work to first order in ϵ_ϕ. These assumptions specify two more unknowns and results in the relationship $p_T^{\tau_i} = p_T^{\nu s_i} + p_T^{\mu_i}$.

(V)   We can also assume $p_T^{\tau_1} = p_T^{\tau_2}$. This is the same as assuming there is no initial or final state radiation, which is reasonable given that $\sqrt{\hat{s}} \lesssim M_Z$, and we

---

[8]Technically we should also consider vector massive neutrinos, though we do not expect that this choice affects our results. For $m_{\nu_s} = 0$ the vector results are identical to the scalar case, while there is a small shift in the matrix element if $m_{\nu_s} \neq 0$.

have vetoed jets in our events. This assumption leads to specification of one more unknown.

Making the above approximations,[9] we can determine all of the unknown kinematic parameters and therefore calculate $|\mathcal{M}_{\tau\tau}^{\mathrm{mod}}|^2$ with no integrations, *significantly* reducing the time and computational power needed to do the analysis. The analytical expression of $|\mathcal{M}_{\tau\tau}^{\mathrm{mod}}|^2$ is given in Appendix A.

We emphasize that these assumptions are only reasonable in the context of the $\tau^+\tau^-|_{\mathrm{dimuon}}$ production. For the rest of the processes $[(pp \to \mu^+\mu^-\nu_\ell\bar{\nu}_\ell)_{\mathrm{SM}}$ background and $(pp \to \mu^+\mu^-\nu_\ell\bar{\nu}_\ell)_{\mathrm{SM}+Z'}$ signal], the approximations I–V are not faithful to the kinematics and we might get unphysical results, i.e. $|\mathcal{M}_{\tau\tau}^{\mathrm{mod}}|^2 < 0$. To get an idea of how negative matrix element squared can arise, let us look at the conclusion of assumption IV: $p_T^\tau = p_T^{\nu s_i} + p_T^{\mu_i}$, which means $p_T^{\nu s}$ is calculated based on p_T^τ. We can determine p_T^τ using conservation of energy and momentum and a combination of assumptions. We get

$$(p_T^\tau)^{\mathrm{approx}} = \sqrt{\frac{M_Z^2}{2(1 + \cosh(\Delta\eta(\mu^+, \mu^-)))} - m_\tau^2}. \quad (10)$$

Consequently, the p_T of the vector sum of the two neutrinos coming from a tau can be deduced from these approximations: $p_T^{\nu s_i} = (p_T^\tau)^{\mathrm{approx}} - p_T^{\mu_i}$. These approximations were reasonable in the framework of $\tau^+\tau^-|_{\mathrm{dimuon}}$. However, for $(pp \to \mu^+\mu^-\nu_\ell\bar{\nu}_\ell)$ we can have $(p_T^\tau)^{\mathrm{approx}} < p_T^{\mu_i}$ and therefore a negative (unphysical) magnitude for the transverse momentum, which may lead to negative $|\mathcal{M}_{\tau\tau}^{\mathrm{mod}}|^2$. The weights (area normalized) of $|\mathcal{M}_{\tau\tau}^{\mathrm{mod}}|^2$ for MC generaed $(pp \to \tau^+\tau^- \to \mu^+\mu^-\not{E}_T)$, $(pp \to \mu^+\mu^-\nu_\ell\bar{\nu}_\ell)_{\mathrm{SM}}$ and $(pp \to \mu^+\mu^-\nu_\ell\bar{\nu}_\ell)_{\mathrm{SM}+Z'}$ are shown below in Fig. 9. As expected, the $\tau^+\tau^-|_{\mathrm{dimuon}}$ distribution is more localized and all of its events have $|\mathcal{M}_{\tau\tau}^{\mathrm{mod}}|^2 > 0$. Therefore, if we restrict ourselves to events with $|\mathcal{M}_{\tau\tau}^{\mathrm{mod}}|^2 < 0$, we can safely assume that $\tau\tau$ background is negligible. We have generated 10 million $\tau^+\tau^-|_{\mathrm{dimuon}}$ events, and 1 million events for each of the $(pp \to \mu^+\mu^-\nu_\ell\bar{\nu}_\ell)_{\mathrm{SM}}$ and $(pp \to \mu^+\mu^-\nu_\ell\bar{\nu}_\ell)_{\mathrm{SM}+Z'}$ processes, to make sure we have captured the tail of the $\tau^+\tau^-|_{\mathrm{dimuon}}$ distribution correctly.

Inspecting Fig. 9, we can see that the signal $(pp \to \mu^+\mu^-\nu_\ell\bar{\nu}_\ell)_{\mathrm{SM}+Z'}$ and background $(pp \to \mu^+\mu^-\nu_\ell\bar{\nu}_\ell)_{\mathrm{SM}}$ have slightly different weights according to $|\mathcal{M}_{\tau\tau}^{\mathrm{mod}}|^2$. This is not completely surprising because there are contributions from more diagrams in the signal events. The difference is most significant at large, negative $|\mathcal{M}_{\tau\tau}^{\mathrm{mod}}|^2$; in particular, $(pp \to \mu^+\mu^-\nu_\ell\bar{\nu}_\ell)$ including Z' intermediate states populates $|\mathcal{M}_{\tau\tau}^{\mathrm{mod}}|^2 < -4$ more than when the Z' is
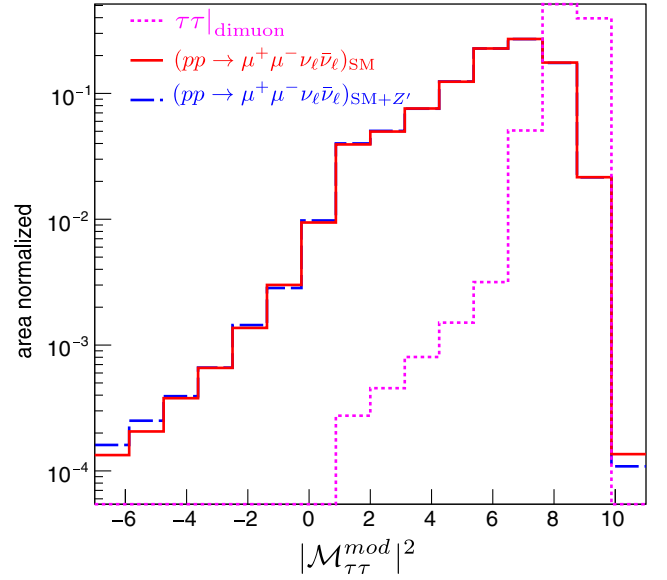

FIG. 9. The weights $|\mathcal{M}_{\tau\tau}^{\mathrm{mod}}|^2$ for various MC sample events. The dotted purple line shows the weights for MC generated $\tau^+\tau^-|_{\mathrm{dimuon}}$ (10 million generated events). The solid red line shows the MC $(pp \to \mu^+\mu^-\nu_\ell\bar{\nu}_\ell)$ background (1 million events), and the dashed blue line is MC $(pp \to \mu^+\mu^-\nu_\ell\bar{\nu}_\ell)$ including Z' (1 million events). Because the assumptions were chosen based on $\tau^+\tau^-|_{\mathrm{dimuon}}$ topology and were not reasonable in other processes, we have $|\mathcal{M}_{\tau\tau}^{\mathrm{mod}}|^2 < 0$ for part of the distribution of other processes.

excluded. This $|\mathcal{M}_{\tau\tau}^{\mathrm{mod}}|^2$ region is shown in greater detail in Fig. 10.

To comprehend why $(pp \to \mu^+\mu^-\nu_\ell\bar{\nu}_\ell)_{\mathrm{SM}+Z'}$ prefers to be in the region $|\mathcal{M}_{\tau\tau}^{\mathrm{mod}}|^2 < -4$, let us look once again at the squared matrix element of the signal

$$
\begin{aligned}
|\mathcal{M}_{\mathrm{signal}}|^2 &= |\mathcal{M}_{\mathrm{SM}} + \mathcal{M}_{Z'}|^2 \\
&= |\mathcal{M}_{Z'}|^2 + 2|\mathcal{M}_{\mathrm{SM}}^*\mathcal{M}_{Z'}| + |\mathcal{M}_{\mathrm{SM}}|^2 \\
&\supset |\mathcal{M}_{Z',\mathrm{on-shell}}|^2 + 2|\mathcal{M}_{WW}^*\mathcal{M}_{Z'}|,
\end{aligned}
$$

where $\mathcal{M}_{WW}$ refers to the portion of the SM matrix element that involves the WW contribution. Obviously, the departure of the signal from the SM background is most ideal for discrimination when Z' is on shell.[10] That is because (1) the $|\mathcal{M}|^2$ is suppressed by only two powers of ϵ, only one power of ϵ at the production vertex of Z' (amplitude level) and no ϵ suppression at the decay vertex, and (2) the

---

[9]Although we have defined modified specifically for the second approximation, we generalize its definition to encompass *all* of the aforementioned approximations.

[10]The greatest contribution to the signal (only considering the Z' contribution) is when on-shell Z' comes from an on-shell Z. On the other hand, because both $\tau^+\tau^-|_{\mathrm{dimuon}}$ background and signal have $\sqrt{\hat{s}} \lesssim M_Z$, a significant fraction of the signal (only the Z' contribution) is removed when removing $\tau\tau$ background. This is inevitable, and the same challenge was faced with the cut-and-count method described in [66], when $M_T(\mu\mu, \not{E}_T) < 100$ GeV was imposed to remove $\tau^+\tau^-|_{\mathrm{dimuon}}$. Therefore, it is really important to use the interference to look for our signal.
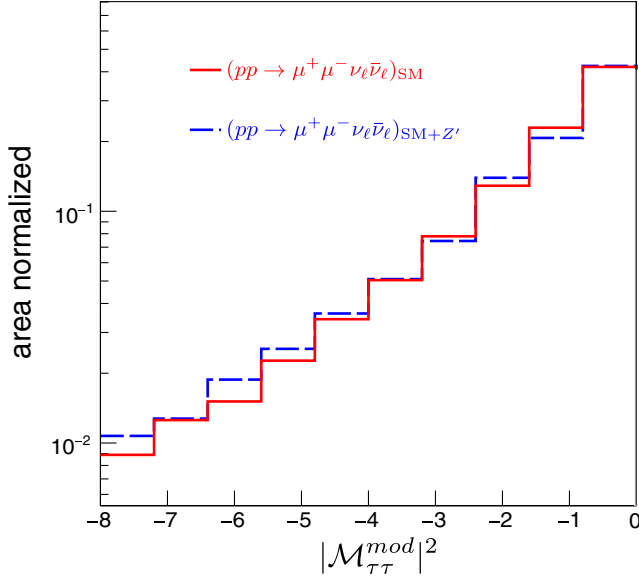
FIG. 10. The weights $|\mathcal{M}_{\tau\tau}^{\text{mod}}|^2$ for $(pp \to \mu^+\mu^-\nu_\ell\bar{\nu}_\ell)$ background (SM only, solid line) and signal (SM $+ Z'$, dashed line) in the negative $|\mathcal{M}_{\tau\tau}^{\text{mod}}|^2$ region.

topology of process with an on-shell Z' mediator is different from the SM background, and so with some careful cuts we can make the SM background small. The next most important contribution of the signal is in the $Z' -$ SM interference, also suppressed by only two powers of ϵ. The interference term is significant when either the SM piece is sizable or when the portion with Z' contribution is big. Each option requires different kinematics; large Z' contribution means the invariant mass of the neutrinos is small (or equivalently angular separation between the neutrinos is small), while large contribution of the SM could be when some of the intermediate states are produced on resonance. The only SM background that can have on-shell resonances

and yet have other kinematics consistent with a (nearly) on-shell Z' is the t-channel W^+W^- background. We suspect the excess in the signal in the region of $|\mathcal{M}_{\tau\tau}^{\text{mod}}|^2 < -4$ is due to the interference of Z' piece with the WW contribution.

To check this intuition, we study the distributions of the events in the invariant mass of a muon and the associated neutrino (i.e. $M_{\nu_\ell\mu^+}$ or $M_{\bar{\nu}_\ell\mu^-}$), and the separation between the two neutrinos $[\Delta R(\nu_\ell, \bar{\nu}_\ell)]$ after requiring $|\mathcal{M}_{\tau\tau}^{\text{mod}}|^2 < -4$ in Fig. 11. These are not kinematic variables that we could actually measure, as neutrinos are not observed at the detector. However, studying them can help us understand the behavior of the simulated events for different regions of $|\mathcal{M}_{\tau\tau}^{\text{mod}}|^2$.

We can see from Fig. 11 that in the region of $|\mathcal{M}_{\tau\tau}^{\text{mod}}|^2 < -4$, all of the signal events have $M_{\mu\nu} \sim M_W$, and have a small separation between the two neutrinos. This is perfectly consistent with what we expected from the interference term; the effect of the interference is enhanced when W s are near resonance, and other kinematic distributions are more consistent with the Z' signal. Thereby, we can be confident that the excess at $|\mathcal{M}_{\tau\tau}^{\text{mod}}|^2 < -4$ is due to the interference between the signal and the W^+W^- background.

With the basic cuts, $\not{E}_T > 20$ GeV, and modified $|\mathcal{M}_{\tau\tau}|^2 < -4$, we get the cross section of the background $\sigma(pp \to \mu^+\mu^-\nu_\ell\bar{\nu}_\ell)_{\text{SM}}$ of 80 ± 3 ab, while $\sigma(pp \to \mu^+\mu^-\nu_\ell\bar{\nu}_\ell)_{\text{SM}+Z'}$ is 133 ± 4 ab. The uncertainties on the cross sections are derived based on the number of events in our simulation that survive the imposed cuts. Let us define

$$S \equiv \text{luminosity} \times (\sigma(pp \to \mu^+\mu^-\nu_\ell\bar{\nu}_\ell)_{\text{SM}+Z'}$$
$$- \sigma(pp \to \mu^+\mu^-\nu_\ell\bar{\nu}_\ell)_{\text{SM}})$$
$$B \equiv \text{luminosity} \times (\sigma(pp \to \mu^+\mu^-\nu_\ell\bar{\nu}_\ell)_{\text{SM}}$$
$$+ \sigma(pp \to \tau^+\tau^- \to \mu^+\mu^-\nu_\mu\bar{\nu}_\mu\nu_\tau\bar{\nu}_\tau)_{\text{SM}}).$$
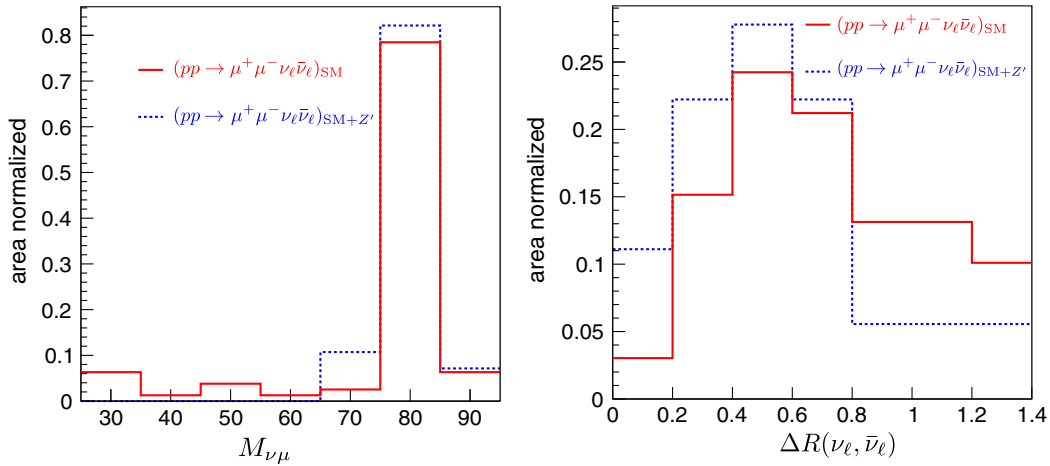

FIG. 11. The left plot represents the invariant mass of a muon and the associated neutrino M_{ν_μ}, and the right plot is the separation between the two neutrinos $\Delta R(\nu_\ell, \bar{\nu}_\ell)$. These plots show the distribution of the events after requiring $|\mathcal{M}_{\tau\tau}^{\text{mod}}|^2 < -4$, and demonstrate that the signal events belong to on-shell W production with Z' (near) on shell as well.
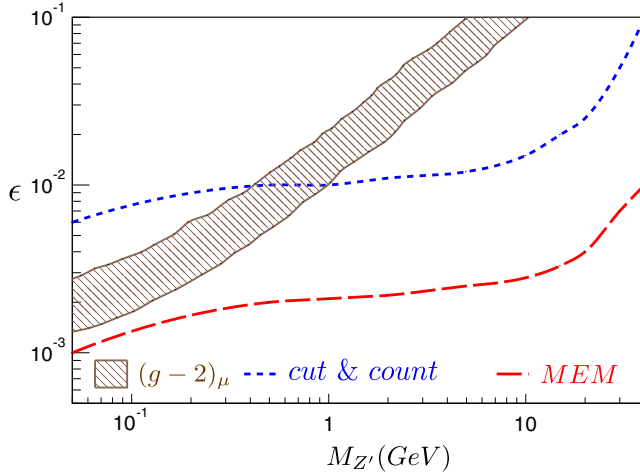
FIG. 12. The new bounds according to our study using MEM compared with the cut-and-count method [66]. The $(g-2)_\mu$ band is also shown in brown. For $M_{Z'} > 2m_\mu$, the MEM bound depends on whether we know $M_{Z'}$ or not. The red line shown here is our best bound. For $M_{Z'} < 2m_\mu$, our analysis is independent of $M_{Z'}$. The bounds from the MEM and cut-and-count method are with luminosity of 300 fb^{-1} and are up to 3σ.

If we use the significance $S/\sqrt{B}$ as a test statistic and assume that $pp \rightarrow \mu^+\mu^- + \text{jets}$ (which was our reducible background) is 0^{11} after imposing the $\not{E}_T$ cut and $|\mathcal{M}^{\text{mod}}_{\tau\tau}|^2 < -4$, we find that after 300 fb^{-1} luminosity, we get $S/\sqrt{B} \geq 3$.

Our reach in the region of the parameter space using MEM in contrast with the cut-and-count method and the $(g-2)_\mu$ band is shown in Fig. 12. The MEM and cut-and-count method bounds are based on the benchmark point of $M_{Z'}^{\text{gen}} = 0.05$ GeV and $\epsilon^{\text{gen}} = 0.001$. We see that with MEM, our sensitivity improves by a factor of 5–10 compared to the cut-and-count method, and we can explore a greater region of parameter space including the $(g-2)_\mu$ band.

V. DISCUSSSION

In this paper, we applied the MEM to the $L_\mu - L_\tau$ model and concluded that our sensitivity improves by up to an order of magnitude compared with the cut-and-count method. The MEM uses the matrix element of a process to distinguish signal from background, and it has proven to be a powerful tool in several SM measurements (e.g., top quark mass) and searches (e.g., Higgs to four lepton and

electroweak single top production). However, it has not been extensively applied to BSM searches. The two main hurdles are that the MEM needs the physical parameters of the model and the four momenta of all initial and final states as inputs. In the quest to find BSM signatures, we do not know the value of the new model parameters and the processes often contain unknown momenta in the form of missing energy.

To investigate how we can combat these difficulties, as a first small step, we applied the MEM to the $L_\mu - L_\tau$ model. This model, being one of the simplest extensions of the SM, is already very well motivated because it can explain some of the current observational anomalies in the $(g-2)_\mu$ and B decays. In this model, there is a Z' that couples to only second and third generation leptons at tree level. As a result, any tree-level process at the LHC involving Z' has to include four leptons of second or third generation. We considered two cases, 4μ and $2\mu + \not{E}_T$. Both of these processes have a large number of kinematic observables, making them ideal test grounds for the MEM approach. The new parameters introduced by the $L_\mu - L_\tau$ model are the Z' mass $M_{Z'}$ and coupling (parametrized by ϵ: $g_{Z'} = \epsilon g'$).

Although the favorite mass region to explain the B decay anomalies is $O(\text{TeV})$, in this paper, we only focused on light Z': $M_{Z'} < M_Z$. The extension of our analysis to $M_{Z'} \sim O(\text{TeV})$ is nontrivial and so we leave it to future works. In the mass range $2m_\mu < M_{Z'} < M_Z$, we looked at the process $pp \rightarrow Z \rightarrow 4\mu$. This channel is clean and well understood, and the presence of $Z \rightarrow 4e$ and $Z \rightarrow 2e2\mu$ control samples can be used to mitigate systematic uncertainties. Therefore, we can be sensitive to percent-level deviations. As the matrix element of the signal depends on $M_{Z'}$ and ϵ, we first discussed how we can find values of these parameters that best separate signal from background. The optimal value of $M_{Z'}$ can be found by maximizing likelihood ratio with respect to $M_{Z'}$, depending on the strength of the Z' coupling. However, the likelihood ratio function increases monotonically as a function of ϵ regardless of whether events belong to the signal sample or the background sample. Hence, we had to deviate from the conventional MEM and look for the most optimal analysis value of ϵ by studying the distribution of the signal and background MC generated events as a function of likelihood ratio for various fixed values of ϵ. In the best case scenario, the $L_\mu - L_\tau$ model can be explored up to 3σ for $\epsilon \gtrsim 0.002$ for $2m_\mu < M_{Z'} < 20$ GeV and $\epsilon \gtrsim 0.005$ for $20 < M_{Z'} < 40$ GeV with luminosity of 300 fb^{-1} at the LHC, which is about an order of magnitude improvement compared with the cut-and-count method.

For lighter Z', we studied the process $pp \rightarrow 2\mu\not{E}_T$. Due to the presence of missing energy, this channel is not as clean as the all muonic final state and is afflicted by several backgrounds. One significant background is $pp \rightarrow \tau^+\tau^-$ with the taus decaying to muons, which has a cross section

---

[11]This assumption is backed up by a MC study of 500 K $pp \rightarrow \mu^+\mu^-$ events generated at the detector level (PGS [77]) with the default smearing algorithm. Requiring events pass the dilepton trigger and contain no jets with $p_T^j > 20$ GeV and photons with $p_T^\gamma > 10$ GeV, the cross section was 111 pb. After imposing $\not{E}_T = p_T(\text{dimuon}) > 20$ GeV, the cross section drops to 0.22 pb. With the further requirement of $|\mathcal{M}^{\text{mod}}_{\tau\tau}|^2 < -4$, we find that the $pp \rightarrow \mu^+\mu^-$ background can be removed completely.

that is orders of magnitude greater than signal. To be sensitive to the signal, we first work towards eliminating the $\tau^+\tau^-|_{\text{dimuon}}$. We use the squared matrix element of $\tau^+\tau^-|_{\text{dimuon}}$ for this task. This is a departure from the canonical MEM, as we weight the events by the squared matrix element of only one process, whereas in MEM we usually weight the events by the ratio of the squared matrix element of the signal processes over the background ones. This alternative approach has two main benefits: (1) given that we have missing energy in the process, calculating the squared matrix elements is challenging; hence, focusing our attention to only one can save us time and computational power. (2) Furthermore, this approach is independent of the model parameters and can be used for any BSM physics with this signature at the LHC.

Due to the presence of four sources of missing energy in the $\tau^+\tau^-|_{\text{dimuon}}$ process, we have ten unconstrained momenta in this channel. Traditionally, one would proceed by integrating over the unknown momenta. In this paper, we instead showed how the unknown momenta could be estimated by exploiting the topology of the $\tau^+\tau^-|_{\text{dimuon}}$. Without the need for any integrations, we calculate the (modified) squared matrix element of $\tau^+\tau^-|_{\text{dimuon}}$ ($|\mathcal{M}_{\tau\tau}^{\text{mod}}|^2$) and use the resulting weight as a discriminant. With a judicious cut on $|\mathcal{M}_{\tau\tau}^{\text{mod}}|^2$, we find that the $\tau^+\tau^-|_{\text{dimuon}}$ background can be completely eliminated. We then observe that the signal reacts differently to the $|\mathcal{M}_{\tau\tau}^{\text{mod}}|^2$ compared to other (non-*tau*) SM backgrounds, and trace the difference to interference between Z' contributions to the amplitude and contributions containing two on-shell W's. As a result, we can differentiate the signal from *all* SM backgrounds using $|\mathcal{M}_{\tau\tau}^{\text{mod}}|^2$ alone. With this method, we find we can reach to 3σ up to $\epsilon \sim 0.001$ for $M_{Z'} < 2m_\mu$ assuming an integrated luminosity of 300 fb^{-1}, covering the $(g-2)_\mu$ band. This result may be improved if we relax some of the kinematic assumptions and instead integrate over a subset of the unconstrained momentum, something which may be worth investigating in the future.

Because we did not use the squared matrix element of the signal in our analysis of $pp \to 2\mu\slashed{E}_T$, our procedure can be applied to any BSM searches with leptons and missing energy in the final states. Scenarios with leptons and missing energy are particularly well motivated in many dark matter and dark photon searches at the LHC [78–86]. More generally, we argue that processes with $\slashed{E}_T$ that have a specific topology can benefit from the MEM, while not suffering from its computational challenges.

In conclusion, in this paper we provide a working example where, after approximating unknown momenta using the topology of the process, the MEM yields a superior sensitivity compared to the cut-and-count method without having to integrate. Consequently, even if applying the canonical MEM to BSM searches has obstacles and

appears to be computationally challenging, we may be able to *modify* the MEM with reasonable assumptions to ease the computational difficulty and yet gain a better sensitivity than the cut-and-count method.

ACKNOWLEDGMENTS

APPENDIX: THE MODIFIED SQUARED MATRIX ELEMENT OF THE $\tau^+\tau^-|_{\text{dimuon}}$ PROCESS ($|\mathcal{M}_{\tau\tau}^{\text{mod}}|^2$)

In this appendix, we detail the calculation of $|\mathcal{M}_{\tau\tau}^{\text{mod}}|^2$. The first step is to define the four-vector of νs (the vector sum of the two neutrinos coming from each tau) using the momenta of muons in the framework of assumptions discussed in Sec. IV B, keeping in mind the νs are not massless,

$$p^{\nu s_i} = \left(\sqrt{(p_T^{\nu s_i})^2 + m_{\nu s_i}^2} \cosh\eta_{\nu s_i}, p_T^{\nu s_i}\cos\phi_{\nu s_i}, p_T^{\nu s_i}\sin\phi_{\nu s_i}, \right.$$
$$\left. \times \sqrt{(p_T^{\nu s_i})^2 + m_{\nu s_i}^2}\sinh\eta_{\nu s_i} \right),$$

where the subscript i is defined such that $\tau_i \to \mu_i\nu s_i$. In the following, we define $p_T^{\nu s_i}$, $m_{\nu s_i}$, $\eta_{\nu s_i}$, and $\phi_{\nu s_i}$ in terms of known or measurable parameters,

$$p_T^{\nu s_i} = \sqrt{\frac{M_Z^2}{2(1+\cosh(\Delta\eta(\mu^+,\mu^-)))} - m_\tau^2 - p_T^{\mu_i}}$$

$$m_{\nu s_i} = \sqrt{m_\tau^2 + 2(p_T^{\mu_i})^2 - 2\sqrt{(p_T^{\mu_i})^2(m_\tau^2 + (p_T^{\mu_i})^2 + (p_T^{\nu s_i})^2)}}$$

$$\eta_{\nu s_i} = \eta_{\mu_i}$$

$$\phi_{\nu s_i} = \phi_{\mu_i} + \epsilon_\phi^i, \quad \text{where } \epsilon_\phi^i \ll 1,$$

where we have ignored the muon mass ($m_\mu = 0$), and ϵ_ϕs are calculated from the conservation of momenta in the transverse plane.

Furthermore, one of the assumptions (III) in Sec. IV B is that the taus are produced from an on-shell Z. Therefore, in calculating $|\mathcal{M}_{\tau\tau}^{\text{mod}}|^2$, we also assume the process is $Z \to \tau^+\tau^- \to \mu_1\mu_2\nu s_1\nu s_2$, where Z is simply the vector sum of the final state products, shown by p^Z. The modified squared matrix element of $\tau^+\tau^-|_{\text{dimuon}}$ is

$$
\begin{aligned}
|\mathcal{M}_{\tau\tau}^{\mathrm{mod}}|^2 =\ & \frac{m_\tau^2}{(m_\tau^2 - m_{\nu_{S_1}}^2)^2 (m_\tau^2 - m_{\nu_{S_2}}^2)^2} \Bigg(\frac{8 m_{\nu_{S_1}} m_{\nu_{S_2}} (p^Z \cdot p^{\mu_1})(p^Z \cdot p^{\mu_2})}{m_\tau^2 M_Z^2} + \frac{4 m_{\nu_{S_1}} m_{\nu_{S_2}} (p^{\mu_1} \cdot p^{\mu_2})}{m_\tau^2} \\
& - \frac{8 m_{\nu_{S_1}} [(p^Z \cdot p^{\mu_1})(p^Z \cdot p^{\mu_2}) + 2(p^Z \cdot p^{\nu_{S_1}})(p^Z \cdot p^{\mu_2})]}{m_\tau M_Z^2} - \frac{2 m_{\nu_{S_1}} [(p^{\mu_1} \cdot p^{\mu_2}) + 2(p^{\nu_{S_1}} \cdot p^{\mu_2})]}{m_\tau} \\
& - \frac{4 m_{\nu_{S_2}} [(p^Z \cdot p^{\mu_1})(p^Z \cdot p^{\mu_2}) + 2(p^Z \cdot p^{\mu_1})(p^Z \cdot p^{\nu_{S_2}})]}{m_\tau M_Z^2} - \frac{2 m_{\nu_{S_2}} [(p^{\mu_1} \cdot p^{\mu_2}) + 2(p^{\mu_1} \cdot p^{\nu_{S_2}})]}{m_\tau} \\
& + \frac{1}{M_Z^2} [4(p^Z \cdot p^{\mu_1})(p^Z \cdot p^{\nu_{S_2}}) + 2(p^Z \cdot p^{\mu_1})(p^Z \cdot p^{\mu_2}) + 8(p^Z \cdot p^{\nu_{S_1}})(p^Z \cdot p^{\nu_{S_2}})] \\
& \times \frac{4}{M_Z^2} (p^Z \cdot p^{\nu_{S_1}})(p^Z \cdot p^{\mu_2}) + 2(p^{\mu_1} \cdot p^{\nu_{S_2}}) + (p^{\mu_1} \cdot p^{\mu_2}) + 4(p^{\nu_{S_1}} \cdot p^{\nu_{S_2}}) + 2(p^{\nu_{S_1}} \cdot p^{\mu_2}) \Bigg).
\end{aligned}
$$

The numerical coefficient in front of $|\mathcal{M}_{\tau\tau}^{\mathrm{mod}}|^2$ is irrelevant and thus can be ignored. This is because it does not matter whether the plots are with respect to $|\mathcal{M}_{\tau\tau}^{\mathrm{mod}}|^2$ or $16\pi |\mathcal{M}_{\tau\tau}^{\mathrm{mod}}|^2$.

---

[1] A. Hocker *et al.*, TMVA—Toolkit for Multivariate Data Analysis, *Proc. Sci.*, ACAT (2007) 040 [arXiv:0703039].

[2] B. H. Denby, Neural networks and cellular automata in experimental high-energy physics, Comput. Phys. Commun. **49**, 429 (1988).

[3] L. Bellantoni, J. S. Conway, J. E. Jacobsen, Y. B. Pan, and S. L. Wu, Using neural networks with jet shapes to identify b jets in e + e− interactions, Nucl. Instrum. Methods Phys. Res., Sect. A **310**, 618 (1991).

[4] C. Bortoletto, A. De Angelis, and L. Lanceri, Tagging the decays of the Z0 boson into b quark pairs with a neural network classifier, Nucl. Instrum. Methods Phys. Res., Sect. A **306**, 459 (1991).

[5] W. S. Babbage and L. F. Thompson, The use of neural networks in gamma—pi0 discrimination, Nucl. Instrum. Methods Phys. Res., Sect. A **330**, 482 (1993).

[6] K. H. Becks, F. Block, J. Drees, P. Langefeld, and F. Seidel, B quark tagging using neural networks and multivariate statistical methods: A comparison of both techniques, Nucl. Instrum. Methods Phys. Res., Sect. A **329**, 501 (1993).

[7] B. P. Roe, H.-J. Yang, J. Zhu, Y. Liu, I. Stancu, and G. McGregor, Boosted decision trees, an alternative to artificial neural networks, Nucl. Instrum. Methods Phys. Res., Sect. A **543**, 577 (2005).

[8] H.-J. Yang, B. P. Roe, and J. Zhu, Studies of boosted decision trees for MiniBooNE particle identification, Nucl. Instrum. Methods Phys. Res., Sect. A **555**, 370 (2005).

[9] F. Fiedler, A. Grohsjean, P. Haefner, and P. Schieferdecker, The matrix element method and its application in measurements of the top quark mass, Nucl. Instrum. Methods Phys. Res., Sect. A **624**, 203 (2010).

[10] P. Artoisenet and O. Mattelaer, MadWeight: Automatic event reweighting with matrix elements, *Proc. Sci.*, CHARGED2008 (2008) 025.

[11] O. Mattelaer, PhD thesis, Louvain University, 2011.

[12] A. Mertens, The automated matrix-element reweighting and its applications at the LHC, J. Phys. Conf. Ser. **523**, 012028 (2014).

[13] L. Bianchini, B. Calpas, J. Conway, A. Fowlie, L. Marzola, and C. Veelken, Reconstruction of the Higgs mass in events with Higgs bosons decaying into a pair of tau leptons using matrix element techniques, Nucl. Instrum. Methods Phys. Res., Sect. A **862**, 54 (2017).

[14] I. Volobouev, Matrix element method in HEP: Transfer functions, efficiencies, and likelihood Normalization, arXiv:1101.2259.

[15] B. Biedermann, A. Denner, S. Dittmaier, L. Hofer, and B. Jager, Next-to-leading-order electroweak corrections to the production of four charged leptons at the LHC, J. High Energy Phys. 01 (2017) 033.

[16] V. M. Abazov *et al.* (D0 Collaboration), A precision measurement of the mass of the top quark, Nature (London) **429**, 638 (2004).

[17] V. M. Abazov *et al.* (D0 Collaboration), Search for single top quark production in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$-TeV, Phys. Lett. B **622**, 265 (2005).

[18] K. Kondo, Dynamical likelihood method for reconstruction of events with missing momentum. 2: Mass spectra for $2 \to 2$ processes, J. Phys. Soc. Jpn. **60**, 836 (1991).

[19] R. H. Dalitz and G. R. Goldstein, The decay and polarization properties of the top quark, Phys. Rev. D **45**, 1531 (1992).

[20] A. Abulencia *et al.* (CDF Collaboration), Precise Measurement of the Top Quark Mass in the Lepton + Jets Topology at CDF II, Phys. Rev. Lett. **99**, 182002 (2007).

[21] T. Aaltonen *et al.* (CDF Collaboration), Measurement of the Top Quark Mass with Dilepton Events Selected using Neuroevolution at CDF, Phys. Rev. Lett. **102**, 152001 (2009).

[22] V. M. Abazov *et al.* (D0 Collaboration), Precise Measurement of the Top Quark Mass from Lepton + Jets Events at D0, Phys. Rev. Lett. **101**, 182001 (2008).

[23] V. M. Abazov *et al.* (D0 Collaboration), Observation of Single Top Quark Production, Phys. Rev. Lett. **103**, 092001 (2009).

[24] T. Aaltonen *et al.* (CDF Collaboration), Observation of single top quark production and measurement of $|V_{tb}|$ with CDF, Phys. Rev. D **82**, 112005 (2010).

[25] V. M. Abazov *et al.* (D0 Collaboration), Measurement of the top quark mass using the matrix element technique in dilepton final states, Phys. Rev. D **94**, 032004 (2016).

[26] G. Cortiana, Top quark mass measurements: Review and perspectives, Rev. Phys. **1**, 60 (2016).

[27] D. Schouten, A. DeAbreu, and B. Stelzer, Accelerated matrix element method with parallel computing, Comput. Phys. Commun. **192**, 54 (2015).

[28] N. A. Graf, M. E. Peskin, and J. L. Rosner, in *Proceedings, 2013 Community Summer Study on the Future of U.S. Particle Physics: Snowmass on the Mississippi (CSS2013)* (2013), http://www.slac.stanford.edu/econf/C1307292.

[29] V. M. Abazov *et al.* (D0 Collaboration), Precise Measurement of the Top Quark Mass in the Dilepton Channel at D0, Phys. Rev. Lett. **107**, 082004 (2011).

[30] X.-G. He, G. C. Joshi, H. Lew, and R. Volkas, Simplest Z-prime model, Phys. Rev. D **44**, 2118 (1991).

[31] F. del Aguila, M. Chala, J. Santiago, and Y. Yamamoto, Collider limits on leptophilic interactions, J. High Energy Phys. 03 (2015) 059.

[32] J. M. Campbell, W. T. Giele, and C. Williams, The matrix element method at next-to-leading order, J. High Energy Phys. 11 (2012) 043.

[33] R. Baumeister and S. Weinzierl, The matrix element method at next-to-leading order for arbitrary jet algorithms, Phys. Rev. D **95**, 036019 (2017).

[34] T. Martini and P. Uwer, The matrix element method at next-to-leading order accuracy, Acta Phys. Pol. B **46**, 2143 (2015).

[35] T. Martini and P. Uwer, The matrix element method at next-to-leading order accuracy, Acta Phys. Polon. **B46**, 2143 (2015).

[36] J. Mott, The muon g-2 experiment at fermilab, Nucl. Part. Phys. Proc. **287–288**, 65 (2017).

[37] S. Baek, N. G. Deshpande, X. G. He, and P. Ko, Muon anomalous g-2 and gauged L(muon)—L(tau) models, Phys. Rev. D **64**, 055006 (2001).

[38] E. Ma, D. P. Roy, and S. Roy, Gauged L(mu)—L(tau) with large muon anomalous magnetic moment and the bimaximal mixing of neutrinos, Phys. Lett. B **525**, 101 (2002).

[39] S. N. Gninenko and N. V. Krasnikov, The muon anomalous magnetic moment and a new light gauge boson, Phys. Lett. B **513**, 119 (2001).

[40] M. Pospelov, Secluded U(1) below the weak scale, Phys. Rev. D **80**, 095002 (2009).

[41] J. Heeck and W. Rodejohann, Gauged L_mu-L_tau symmetry at the electroweak scale, Phys. Rev. D **84**, 075007 (2011).

[42] K. Harigaya, T. Igari, M. M. Nojiri, M. Takeuchi, and K. Tobe, Muon g-2 and LHC phenomenology in the $L_\mu - L_\tau$ gauge symmetric model, J. High Energy Phys. 03 (2014) 105.

[43] C.-H. Chen and T. Nomura, $L_\mu - L_\tau$ gauge-boson production from LFV τ decays at Belle II, arXiv:1704.04407.

[44] W. Altmannshofer, C.-Y. Chen, P. S. Bhupal Dev, and A. Soni, Lepton flavor violating Z? Explanation of the muon anomalous magnetic moment, Phys. Lett. B **762**, 389 (2016).

[45] A. Crivellin, G. D. Ambrosio, and J. Heeck, Addressing the LHC flavor anomalies with horizontal gauge symmetries, Phys. Rev. D **91**, 075006 (2015).

[46] A. Crivellin, G. D. Ambrosio, and J. Heeck, Explaining $h \to \mu^\pm \tau^\mp$, $B \to K^* \mu^+ \mu^-$, and $B \to K \mu^+ \mu^- / B \to K e^+ e^-$ in a Two-Higgs-Doublet Model with Gauged $L_\mu - L_\tau$, Phys. Rev. Lett. **114**, 151801 (2015).

[47] J. Heeck, M. Holthausen, W. Rodejohann, and Y. Shimizu, Higgs in Abelian and non-Abelian flavor symmetry models, Nucl. Phys. **B896**, 281 (2015).

[48] S. Bifani, Status of New Physics searches with $b \to s\ell^+\ell^-$ transitions @ LHCb, in 52nd Rencontres de Moriond on EW Interactions and Unified Theories (Moriond EW 2017) La Thuile, Italy, 2017 (2017), http://inspirehep.net/record/1598464/files/arXiv:1705.02693.pdf.

[49] R. Alonso, P. Cox, C. Han, and T. T. Yanagida, Anomaly-free local horizontal symmetry and anomaly-full rare B decays, arXiv:1704.08158.

[50] K. S. Babu, C. F. Kolda, and J. March-Russell, Implications of generalized Z—Z-prime mixing, Phys. Rev. D **57**, 6788 (1998).

[51] A. A. Andrianov, P. Osland, A. A. Pankov, N. V. Romanenko, and J. Sirkka, On the phenomenology of a Z' coupling only to third family fermions, Phys. Rev. D **58**, 075001 (1998).

[52] J. Erler, P. Langacker, and T.-j. Li, The Z—Z' mass hierarchy in a supersymmetric model with a secluded U(1)-prime breaking sector, Phys. Rev. D **66**, 015002 (2002).

[53] G. Eigen (*BABAR* Collaboration), Direct searches for new physics particles at *BABAR*, J. Phys. Conf. Ser. **631**, 012034 (2015).

[54] D. Curtin, R. Essig, S. Gori, and J. Shelton, Illuminating dark photons with high-energy colliders, J. High Energy Phys. 02 (2015) 157.

[55] G. Eigen (*BABAR* Collaboration), Recent *BABAR* results, J. Phys. Conf. Ser. **631**, 012033 (2015).

[56] R. Essig, J. Mardon, M. Papucci, T. Volansky, and Y.-M. Zhong, Constraining light dark matter with low-energy e^+e^- colliders, J. High Energy Phys. 11 (2013) 167.

[57] B. Wang, Searches for New Physics at the Belle II Experiment, in *Proceedings, Meeting of the APS Division of Particles and Fields (DPF 2015): Ann Arbor, Michigan, USA, 2015* (2015), https://inspirehep.net/record/1402351/files/arXiv:1511.00373.pdf.

[58] T. Araki, S. Hoshino, T. Ota, J. Sato, and T. Shimomura, Cornering the $L_\mu - L_\tau$ gauge boson at Belle-II, Phys. Rev. D **95**, 055006 (2017).

[59] D. Geiregat *et al.* (CHARM-II Collaboration), First observation of neutrino trident production, Phys. Lett. B **245** (1990) 271–275.

[60] S. R. Mishra *et al.* (CCFR Collaboration), Recent electroweak results from the CCFR Collaboration: Neutrino tridents and $W - Z$ interference and the Lorentz structure of the weak current, in *Proceedings, Leptonic Session of the 26th Rencontres de Moriond, Les Arcs, France, 1991* (1991), pp. 87–96, http://lss.fnal.gov/archive/preprint/fermilab-conf-91-374-ae.shtml.

[61] S. R. Mishra *et al.* (CCFR Collaboration), Neutrino Tridents and W Z Interference, Phys. Rev. Lett. **66,** 3117 (1991).

[62] W. Altmannshofer, S. Gori, M. Pospelov, and I. Yavin, Quark flavor transitions in $L_\mu - L_\tau$ models, Phys. Rev. D **89,** 095033 (2014).

[63] Y. Kaneta and T. Shimomura, On the possibility of search for $L_\mu - L_\tau$ gauge boson at Belle-II and neutrino beam experiments, arXiv:1701.00156.

[64] S. Chatrchyan *et al.* (CMS Collaboration), Observation of Z decays to four leptons with the CMS detector at the LHC, J. High Energy Phys. 12 (2012) 034.

[65] G. Aad *et al.* (ATLAS Collaboration), Measurements of Four-Lepton Production at the Z Resonance in pp Collisions at $\sqrt{s} = 7$ and 8 TeV with ATLAS, Phys. Rev. Lett. **112,** 231806 (2014).

[66] F. Elahi and A. Martin, Constraints on $L_\mu - L_\tau$ interactions at the LHC and beyond, Phys. Rev. D **93,** 015022 (2016).

[67] C. Degrande, C. Duhr, B. Fuks, D. Grellscheid, O. Mattelaer, and T. Reiter, UFO—The Universal FeynRules Output, Comput. Phys. Commun. **183,** 1201 (2012).

[68] A. Alloul, N. D. Christensen, C. Degrande, C. Duhr, and B. Fuks, FeynRules 2.0—A complete toolbox for tree-level phenomenology, Comput. Phys. Commun. **185,** 2250 (2014).

[69] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H. S. Shao, T. Stelzer, P. Torrielli, and M. Zaro, The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations, J. High Energy Phys. 07 (2014) 079.

[70] J. Alwall, M. Herquet, F. Maltoni, O. Mattelaer, and T. Stelzer, MadGraph 5: Going beyond, J. High Energy Phys. 06 (2011) 128.

[71] V. Khachatryan *et al.* (CMS Collaboration), Search for new physics in final states with two opposite-sign, same-flavor leptons, jets, and missing transverse momentum in pp collisions at $\sqrt{s} = 13$ TeV, J. High Energy Phys. 12 (2016) 013.

[72] J. M. Campbell and R. K. Ellis, An update on vector boson pair production at hadron colliders, Phys. Rev. D **60,** 113006 (1999).

[73] J. M. Campbell, R. K. Ellis, and C. Williams, Vector boson pair production at the LHC, J. High Energy Phys. 07 (2011) 018.

[74] R. Boughezal, J. M. Campbell, R. K. Ellis, C. Focke, W. Giele, X. Liu, F. Petriello, and C. Williams, Color singlet production at NNLO in MCFM, Eur. Phys. J. C **77,** 7 (2017).

[75] J. P. Lees *et al.* (*BABAR* Collaboration), Search for a muonic dark force at *BABAR*, Phys. Rev. D **94,** 011102 (2016).

[76] T. Sjostrand, S. Mrenna, and P. Z. Skands, PYTHIA 6.4 physics and manual, J. High Energy Phys. 05 (2006) 026.

[77] "Pgs4," http://conway.physics.ucdavis.edu/research/software/pgs/pgs4-general.htm.

[78] E. Gramstad, PhD thesis, Oslo University, 2013.

[79] G. Aad *et al.* (ATLAS Collaboration), Search for direct slepton and gaugino production in final states with two leptons and missing transverse momentum with the ATLAS detector in pp collisions at $\sqrt{s} = 7$ TeV, Phys. Lett. B **718,** 879 (2013).

[80] G. Aad *et al.* (ATLAS Collaboration), Search for direct production of charginos, neutralinos, and sleptons in final states with two leptons and missing transverse momentum in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector, J. High Energy Phys. 05 (2014) 071.

[81] S. Baek, H. Okada, and T. Toma, Two loop neutrino model and dark matter particles with global B-L symmetry, J. Cosmol. Astropart. Phys. 06 (2014) 027.

[82] T. Kruker (ATLAS Collaboration), Searches for supersymmetric gaugino production in final states with leptons with the ATLAS detector, *Proc. Sci.*, ICHEP2012 (2013) 113.

[83] J. Wittkowski (ATLAS Collaboration), Search for electroweak production of SUSY particles in events with leptons in the final state with the ATLAS Detector at the LHC, Acta Phys. Pol. B **44,** 1657 (2013).

[84] T. Cheng and T. Li, Electroweak supersymmetry (EWSUSY) in the NMSSM, Phys. Rev. D **88,** 015031 (2013).

[85] M. Teroerde, PhD thesis, Aachen, Tech. Hochsch., 2016.

[86] W. Altmannshofer, S. Gori, S. Profumo, and F. S. Queiroz, Explaining dark matter and B decay anomalies with an $L_\mu - L_\tau$ model, J. High Energy Phys. 12 (2016) 106.