Modern Physics Letters A
Vol. 25, No. 17 (2010) 1–19

SEARCHES FOR NEW PHYSICS WITH JETS-PLUS-$\not\!\!E_T$ SIGNATURES USING DATA-DRIVEN BACKGROUND ESTIMATES

PIERRE-HUGUES BEAUCHEMIN

Particle Physics Sub-Dept., University of Oxford, Denys Wilkinson Building, Keble Road, Oxford, OX1 3RH, UK
h.beauchemin@physics.ox.ac.uk

PIERRE SAVARD

Department of Physics, University of Toronto, 60 St. George St. Toronto, Ontario M5S 1A7, Canada
savard@physics.utoronto.ca

Received 12 April 2010

Events featuring high energy jets and a large amount of missing transverse energy constitute a key signature for a wide spectrum of new physics models. In this review, the results of two searches with such signatures are presented. The benefits of performing these searches in a model-independent way are discussed and data-driven techniques used to estimate Standard Model backgrounds are described in detail. These data-driven techniques will be an important part of searches for new physics at the LHC, especially in the early data-taking period.

Keywords: Jets-plus-$\not\!\!E_T$; data-driven; model-independent.

1. Introduction

The Standard Model of particle physics provides an accurate description of the data collected so far by a wide variety of particle physics experiments. The extraordinary precision of many experimental results, such as electroweak parameter measurements, the anomalous magnetic moment of the electron (g-2) or the running of the strong interaction coupling constant (α_S)[1,2] are a few examples that have confirmed Standard Model predictions at the level of small higher order quantum corrections. Despite this success, the Standard Model is considered by many high energy physicists as no more than an effective theory valid only below the TeV energy scale. The reason for this is that many fundamental questions about nature are not answered within the context of the Standard Model. For example, the SM does not explain the origin of the matter–anti-mater asymmetry of the Universe, it does not describe the gravitational interaction, and does not include a viable dark matter particle. It also does not explain why the strong CP terms, *a priori* allowed in the Standard Model Lagrangian, are so suppressed (strong CP problem)[3] and there is no natural mechanism that explains what keeps the Higgs boson mass at the electroweak mass scale in the presence of very large higher order corrections.

In the past few decades, many different theories have been developed to address some of the questions left unanswered within the Standard Model. The construction of these new physics theories are now severely constrained by existing measurements in particle physics, astrophysics, and cosmology. They typically predict the existence of new particles or interactions that would profoundly impact particle physics phenomenology at the energy scale of electroweak symmetry breaking and above. The study of this energy range could help us understand why this symmetry is broken, and solve the naturalness problem of the Standard Model. Moreover, if dark matter is made of Weakly Interacting Massive Particles (WIMPs), the mass of these particles would have to be close to this energy range.[4] The Tevatron Collider and the Large Hadron Collider are able to probe this energy scale and their experiments are in a position to reveal new phenomena should they be close to the electroweak breaking scale.

Discrepancies with SM expectations could manifest themselves in a variety of ways including production cross sections, branching ratios, or more complex relationships between different measurable quantities (e.g. unitarity of the CKM triangle). A key advantage of the high energy hadron colliders mentioned above is their potential to directly produce these new particles which is why they are often referred to as “discovery machines”. This direct observation of new particles is to be contrasted with their indirect observation which would manifest itself in discrepancies in some SM parameter values. The simplest case of the observation of a new particle is a resonance in the invariant mass distribution obtained from the reconstruction of all of its decay products. A more challenging case involves final states with undetectable particles. Here, the missing information becomes “missing transverse energy” or $\not\!\!E_T$, and results in an incomplete kinematic reconstruction of the event preventing the use of the invariant mass as a discriminating variable. Also, the missing momentum potentially opens the door to additional instrumental backgrounds that require dedicated efforts to understand and control. It is this challenge of finding new physics in final states with jets and missing momentum that will be addressed in this review.

After a brief survey of some new physics models that can be used as benchmarks to interpret the results of searches in jets-plus-$\not\!\!E_T$ final states, we will present data-driven techniques that can be used to estimate the Standard Model background to jets-plus-$\not\!\!E_T$ events. Example of such searches for new physics conducted at the Tevatron will be used to illustrate the use of data-driven techniques to estimate the major Standard Model backgrounds.[5–7] The result of these searches and their interpretation will then be presented in Sec. 5.

2. Jets-Plus-$\not\!\!E_T$ Searches

Final-state signatures featuring energetic jets and a large amount of missing transverse energy have a great potential for new physics discoveries. A large variety of new physics scenarios, including supersymmetry,[8,9] extra dimensions[10–12] and leptoquarks,[13,14] predict enhanced contributions of jets-plus-$\not\!\!E_T$ events compared to SM predictions. In general, any model predicting associated production of partons and weakly interacting particles or pair production of unstable particles whose decay products are a single parton and a non-interacting particle could be observed as an excess of events above the Standard Model expectation in the jets-plus-$\not\!\!E_T$ channels.

Large amounts of missing transverse energy in these new physics signals could be produced by Standard Model particles like neutrinos. For example, leptoquarks decaying to a quark and a neutrino would produce jets-plus-$\not\!\!E_T$ events.[13,14] The missing momentum could also come from particles not included in the Standard Model like gravitons which could escape into extra dimensions.[10–12] If the non-interacting particle is protected against eventual decay by a discrete symmetry like the R-parity conservation in supersymmetry,[15,16] or the KK-number conservation in Universal Extra Dimension scenarios[17,18] or T-parity conservation in smallest Higgs models,[19] it would be stable and therefore a candidate for a dark matter particle.

The production of particles in the new physics models described above generally involves the strong interaction which implies high production rates. The fact that new physics events with jets-plus-$\not\!\!E_T$ are expected to be produced at a high rate at hadron colliders and that these events could feature dark matter particles justify the importance assigned to these searches at hadron colliders.

2.1. *Signature-based searches*

As discussed above, many final state signatures can be used to search for a wide range of new physics models. However, the sensitivity associated with the choice of the kinematic phase space will be model-dependent: the maximisation of the sensitivity to a specific new physics model can lead to very different choices of kinematic selections depending on the model considered. This optimisation approach is not necessarily the one best suited to find new physics beyond the Standard Model. In general, we are more interested in ruling out the Standard Model than we are interested in ruling out any given particular new physics scenario. There are so many models that predict significant contributions to a jets-plus-$\not\!\!E_T$ final state, that optimising selections on any one of them could reduce our chances to find new physics at all. There are generally no good reasons to favour one model over others, especially since the models that are often available in Monte Carlo based calculations are essentially toy models that are not realistic, even if the underlying theory is potentially correct.

The idea of signature-based searches is to keep the highest possible independence with respect to theoretical models by choosing event selections based mainly on experimental criteria rather than *a priori* optimisation for a given new physics model. The choice of selection cuts for this approach will be limited more by experimental considerations rather than any particular theoretical considerations. This approach will not yield a sensitivity to new physics that is as high as a dedicated search for a given model. However, such a generic search is less likely to miss a new physics signal. Furthermore, should an actual excess be observed, dedicated searches would typically not be in a position to claim evidence for the particular model that was searched for without ruling out other possible alternatives for the excess and without looking at many other final states. Therefore, searches for new physics performed in a model-independent way in many exclusive final states represent a better strategy for understanding the nature of potential non-SM contributions.

In the end, if no significant excess over Standard Model expectations is observed, results can be used to set constraints on a wide variety of new physics models. In general the limits can be set on a simple benchmark model as it provides a good reference with which similar measurements could be compared to.

3. Data-driven Background Calculations

The sensitivity to potential new physics depends not only on the number of new physics and background events but also on the precision of the background estimate. Therefore, a key element to control in order to maximise our sensitivity to the numerous new physics scenarios is the precision with which the background is estimated. Many systematic uncertainties, such as those coming from the jet energy scale and resolution effects, the detector acceptance and efficiency estimates, and the modelling of underlying event and parton distribution functions, will be relatively large in the early data-taking period of the LHC experiments. Analyses designed to minimise the impact of these uncertainties will in general be more robust. Such robust analyses can be achieved by using data-driven estimates of the Standard Model contributions to the signature under study. These techniques should therefore play an important role in searches for new physics, both at the Tevatron and the LHC.

3.1. *Motivation for data-driven estimate of jets-plus-$\not{E}_T$ backgrounds*

Typically, Standard Model predictions for final states involving jets and $\not{E}_T$ suffer from large systematic uncertainties due to effects that are hard to model properly in simulation. These effects can be divided into two categories: non-calculable contributions to the prediction and detector modelling.

The production of jets, due to the confined nature of the strong interaction, involves non-perturbative processes which cannot be predicted by the usual techniques used in quantum field theory. The general procedure is to start from the well-understood perturbative calculation performed in the centre-of-mass of the parton (quark and gluons) system and then to add effects that account for the parton density function of the colliding hadrons (protons or anti-protons), the potential gluon emissions of the initial state or final state partons, the hadronisation of the final partons to create jets of colourless hadrons, and the other potential partonic collisions that can occur within the same colliding hadrons. Most of these effects cannot be computed within the quantum chromodynamic theory and require approximate models for which a large number of parameters need to be tuned to experimental data. Despite this tuning, the uncertainty due to the approximate modelling of these effects can have a large impact on the predicted production of jets. Moreover, such tuning is performed in a jet kinematic phase space region which is different from the one that is probed for the new physics searches motivated above. Using data events containing jets with similar kinematics to predict the Standard Model contribution will reduce significantly the impact of uncertainties associated with the modelling of such non-perturbative effects.

The second important factor which contributes to the uncertainty on the predicted background is the imperfect modelling of the calorimeter response to jets. Although the simulation will in general reproduce well the energy scale and resolution on average, the inadequate modelling of calorimeter cracks, dead regions, particle punch-through, and rare fragmentation effects will lead to large tails in the $\not\!\!E_T$ distribution.

For all the above reasons, the exclusive use of Monte Carlo simulation to predict the Standard Model contribution to jets-plus-$\not\!\!E_T$ events can result in large systematic uncertainties. However, using data itself for a Standard Model prediction can greatly reduce many of the systematic uncertainties mentioned above. For example, if the calibration of the energy of jets was off by 10%, both the prediction and the observation would be off by the same amount and the difference between the Standard Model prediction and the observation would remain unaffected. Calibrating a prediction on data is therefore a robust way to minimise resulting jet energy scale and resolution uncertainties.

From these considerations, we can see that the objective of data-driven estimates of Standard Model backgrounds is to significantly reduce the systematic uncertainty on the predictions by using some well-identified data sample to model the background rather than relying on Monte Carlo modelling of the above effects. In particular, the idea here is to find a data sample orthogonal to the jets-plus-$\not\!\!E_T$ final state (control region), but similar enough to be able to provide a good model of the shape of the jet energy or $\not\!\!E_T$ distributions. The sources of uncertainty on such estimates would therefore essentially be reduced to the statistics of the control regions data. If such background estimate techniques are an asset for mature experiments like CDF and D0, they are even more desirable for newly experiments like ATLAS and CMS for which Monte Carlo simulation have not yet been tuned to collider data.

The use of such data-driven estimates will be advantageous provided that the choice of the control region does not introduce significant biases on kinematic distributions, and that sufficiently large statistical samples can be found such that the normalisation can be estimated precisely. The following sections will illustrate how this can be done with examples of data-driven estimates of the electroweak and QCD backgrounds for searches for new physics performed at CDF.

4. Jets-Plus-$\not\!\!E_T$ Searches with CDF

Two examples of analyses performed at the Tevatron will be presented in the following section. Although the details of the selection cuts will be different from one experiment to the next, the overall strategy and techniques are applicable in general to hadron collider experiments.

4.1. *Event selections and non-collision backgrounds*

The cases of one jet (monojet) and two jets plus $\not\!\!E_T$ (dijet+$\not\!\!E_T$) analyses are discussed below. Although both measurements feature different final states and are sensitive to different new physics processes, most of the event selections can be applied to both analyses. For example, in both measurements, jets are reconstructed with a cone-based algorithm algorithm using a cone size of $\Delta R = \sqrt{\Delta(\eta)^2 + \Delta(\phi)^2} = 0.7$.[20,a] Only jets with a E_T threshold of 20 GeV after the application of jet energy scale corrections[21] were considered for these analyses. The monojet analysis requires exactly one of those jets while the dijet+$\not\!\!E_T$ analysis requires exactly two jets. Events for which the $\not\!\!E_T$ points in the direction of a jet in ϕ are rejected to reduce the QCD background. The requirement is that the difference in the azimuthal angle ϕ between the $\not\!\!E_T$ and each jet satisfies $\Delta(\phi_{\not E_T} - \phi_{\mathrm{jet}}) \geq 0.5$ rad. In order to reduce the W+jets background, a lepton veto is applied. This veto rejects events in which an isolated track of $P_T \geq 10$ GeV is reconstructed. The isolation requirement is defined by the fraction of the energy contained in a cone of 0.4 centred on the track which has to be less than 10% of the transverse momentum of that track. To further reduce the electron contribution, none of the selected jets must have more than 90% of their energy contained in the electromagnetic calorimeter.

Cosmic rays, beam-gas collisions and beam halo events can contribute to the monojet and dijet+$\not\!\!E_T$ signatures. They are the largest source of background to the monojet final state without the "cleanup" cuts described below. It is sufficient for a cosmic ray muon or a beam halo muon produced upstream of the detector to cross the calorimeter and emit large bremsstrahlung radiation to obtain a large deposition

[a] We use a coordinate system where θ is the polar angle to the proton beam, ϕ is the azimuthal angle about this beam axis, and η is the pseudorapidity defined as $-\ln\tan(\theta/2)$. The missing transverse energy, $\not\!\!E_T$, is then defined as the magnitude of $-\sum_i E_T^i \hat{n}_i$ where $\hat{n}_i$ is a unit vector in the azimuthal plane that points from the beamline to the ith calorimeter tower and E_T^i is the transverse component of measured energy in the tower, defined as $E^i \cdot \sin(\theta)$.

of energy in the calorimeter which will yield a jet signature, and by construction, a correlated equal amount of $\not\!\!E_T$. This kind of background is relatively hard to predict, even from Monte Carlo simulation, and therefore requires an extra set of selections, to maximally reduce its contribution. This non-collision background gets almost completely eliminated by requiring that:

- a reconstructed primary vertex with at least five tracks pointing to it and within 60 cm of the centre of the detector;
- the sum of the transverse momentum of the tracks matched to each jet fiducial to the tracker be at least 10% of the jet transverse energy;
- the leading jet is central ($|\eta| < 1.0$);
- there is at least 10% of the total energy deposited in the full calorimeter which is contained in the electromagnetic component of the calorimeter.

Timing information available from the hadronic calorimeter has been used to estimate the small, residual non-collision background. The number of jets-plus-$\not\!\!E_T$ events for which the time of a significant energy deposition in the calorimeter is not synchronised with the nearest collision is compared with the number of such events in W/Z+jets events to set an upper limits in the residual non-collision background. The cosmic ray background is therefore estimated by a data-driven method. After the cleanup cuts are applied, physics backgrounds (electroweak and QCD multijets) need to be tackled next.

4.2. *QCD multijets background*

QCD multijet events can become a background to jets-plus-$\not\!\!E_T$ searches if one or more jets are badly mis-measured. This background would dominate the electroweak backgrounds without the requirement that the $\not\!\!E_T$ vector does not point in the direction of any jet. After this selection the QCD background is not expected to be large based on the MC simulation. For the MC simulation to be used, however, we must trust its ability to accurately model detector cracks and uninstrumented regions, and to model the fragmentation that leads to jets that have very few particles which can more easily generate missing transverse energy. In the analyses described here, data-driven techniques have been chosen to estimate this source of background.

After the selection above, a QCD multijet event will become a background if at least one of the jets in the event is completely lost, i.e. loses enough energy to fall below the jet counting threshold. We can therefore classify the QCD background in two orthogonal categories:

- Category A: events where one jet is dominantly responsible for the observed $\not\!\!E_T$. In those events, the $\not\!\!E_T$ would point in the direction of the jet, if it was not lost.
- Category B: events where at least two jets significantly contribute to the $\not\!\!E_T$. In those events, the $\not\!\!E_T$ does not point in the general direction of any jet.

According to the Monte Carlo simulation, category A dominates for low multiplicity jets-plus-$\not\!\!E_T$ final states (less than four jets). For this reason, the general strategy is to use the data to estimate the absolute contribution of the dominant category A events and use the MC simulation to calculate the small residual category B contribution.

In order to estimate the category A background with n-jets+$\not\!\!E_T$ events, events are selected that have $(n+1)$-jets+$\not\!\!E_T$. For these events, one of the jets points in the direction of the $\not\!\!E_T$. These events are orthogonal and unbiased with respect to the signal events, and can therefore be used for the data-driven estimate of the QCD multijet background. By studying the transverse energy distribution of that jet as more and more of the jet transverse energy is lost, we can estimate how many events will have jets below the jet counting threshold. This can be done by fitting the E_T distribution of this jet and doing an extrapolation below the jet counting threshold using the fit. This is the region populated with QCD background events of category A. To avoid double counting of the electroweak background, the W/Z+jets contribution to the fitted data must be removed. Such a correction generally amounts to 15–20% of the overall prediction for the QCD background of category A. Once the correction for electroweak contamination is applied, the integral of the extrapolated function in the signal region provides the data-driven estimate for this background. An example of a fitted distribution for the corrected data, taken from the CDF monojet search for new physics, is shown in Fig. 1. In this example, the jet counting threshold was set to 20 GeV, although the method has been redone with a threshold of 15 GeV, as can be seen on the figure, to test that the method works well. The predicted number of QCD background events of category A, extracted from the extrapolation of Fig. 1, is 591 ± 87 events. The quoted uncertainty of 15% accounts for the statistics of the control region, the systematic uncertainty on the fit and its extrapolation, and the systematic uncertainty on the electroweak contamination correction. The contribution of category B events still needs to be evaluated. Similar data-driven methods to estimate the QCD background to jets-plus-$\not\!\!E_T$ events could not be applied on the QCD background of category B. This is because there is no way to build a control region which would not get important contributions from the signal region. At the LHC, attempts to use the category A control region to construct a $\not\!\!E_T$ transfer function that could be applied to estimate the background of category B is under study. In the analyses described here, the Monte Carlo simulation was used to estimate the relative contribution of category B events over category A events. The use of the ratio reduces the expected systematic uncertainty but because of the low MC statistics left after all cuts and the small impact of the category B contribution on the overall jets-plus-$\not\!\!E_T$ measurements, a conservative 100% uncertainty on the relative contribution of the category B background was used. For example, in the CDF monojet analysis, the relative contribution of category B events was estimated to $20 \pm 20\%$, for an overall prediction of 708 ± 146 QCD background events.

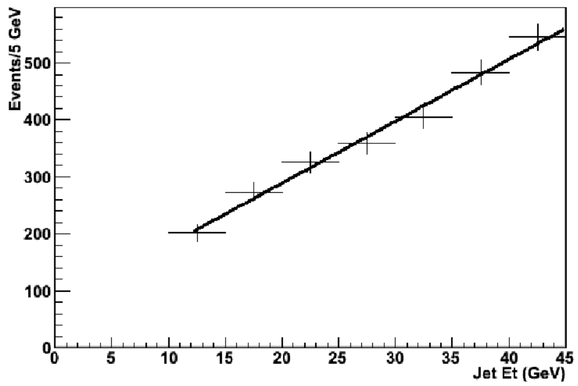

Fig. 1. Transverse energy of the second most energetic jet of multijet data events. There is no lepton in these events and the large $\not\!\!E_T$ points in the direction of the sub-leading jet. The extrapolation of the fitted distribution below 20 GeV is used for the QCD background estimate to 80 GeV monojet events.

4.3. *Electroweak backgrounds*

After the application of cleanup selections and the substantial reduction in the QCD background thanks to the $\Delta\phi$ cut, the remaining background will be from electroweak processes. In particular, the largest source of Standard Model background to jets-plus-$\not\!\!E_T$ events is predicted to come from Z+jets events, where the Z-boson decays to a pair of neutrinos and W+jets events where the W decays to a lepton (e, μ or τ) that is not observed in the detector, i.e. escaping our lepton veto requirements. To estimate these backgrounds, we can use Z+jets and W+jets events for which the leptons (e or μ) are well reconstructed and identified rather than being vetoed, and on which the "+jets" requirement corresponds to the exactly same set of selections applied in the jets-plus-$\not\!\!E_T$ searches. It is because the jet selections are the same for the signal (jets-plus-$\not\!\!E_T$) and control regions and that they are applied on the same type of processes (W/Z+jets) that the data-driven estimates are less affected by large systematic uncertainties which would result from the modelling of jet quantities. These control region events therefore satisfy important criteria for data-driven background estimates. First, they use an orthogonal data sample to the signal sample. Because of the lepton selections, they do not contain any contamination from the jets-plus-$\not\!\!E_T$ events selected in the signal region. Second, the shape of the jets and $\not\!\!E_T$ distributions are not biased by the choice of the control sample. This has been checked in the simulation and in the data by varying kinematic selections. In order to get the $\not\!\!E_T$ distribution of Z($\rightarrow \nu\nu$)+jets events from Z($\rightarrow \ell\ell$)+jets events, the energy of the measured charged leptons has to be removed from the event. Since the impact of the lepton energy scale and resolution uncertainties on the $\not\!\!E_T$ measurement is much smaller than the same uncertainties for jets, the lepton removal procedure does not induce noticeable biases on the modelling of the Z($\rightarrow \nu\nu$)+jets $\not\!\!E_T$ distribution.

In order to be able to use these distributions to estimate the electroweak background to jets-plus-$\not\!\!E_T$ events, they must be properly normalised. These normalisation factors will vary depending on whether the distributions from the controlled regions are used to estimate the Z($\rightarrow \nu\nu$)+jets events or if they are used to estimate W($\rightarrow \ell\nu$)+jets events where the charged lepton is lost. In both cases, a correction to the distribution normalisation must be made for the lepton acceptance (A_{lept}) as well as for the lepton trigger, reconstruction and identification efficiencies (ϵ_{lept}). Of course, these efficiencies are different for W and Z events, because they involve a different number of charged leptons.

The second important contribution to the normalisation factor that needs to be computed is the estimate of the Standard Model background to the W/Z+jets events selected in the control regions. For example, QCD multijet events for which one of the jets fully mimics the signature of an electron will contribute to the W $\rightarrow e\nu$+jets distributions used to estimate the jets-plus-$\not\!\!E_T$ background, without being a background to jets-plus-$\not\!\!E_T$ events itself. It has been verified that these backgrounds to W $\rightarrow \ell\nu$+jets or Z $\rightarrow \ell\ell$+jets events do not noticeably distort the shape of the jet energy or $\not\!\!E_T$ distributions extracted from the W/Z+jets event candidates. This is because of the way they are selected and the smallness of their contribution ($\sim$ 20% of the total W($\rightarrow e\nu$)+jets events, and $\sim$ 5% of the Z($\rightarrow ee$)+jets events for example). Their impact on the final prediction is therefore essentially just a change in the normalisation. To limit the effect of these backgrounds on the overall systematic uncertainty of the measurement, data-driven techniques have been used to estimate the contribution of the QCD background to W/Z+jets events.

Once the above normalisation corrections are applied, we obtain a measurement of W $\rightarrow \ell\nu$+jets and Z $\rightarrow \ell\ell$+jets cross sections for the same jet selections to those that are probed in the jets-plus-$\not\!\!E_T$ searches. An estimate of the Z $\rightarrow \nu\nu$+jets background to jets-plus-$\not\!\!E_T$ events can thus be obtained from the Z $\rightarrow \ell\ell$+jets cross section measurements by multiplying this measured cross section by the ratio of Z $\rightarrow \nu\nu$ to Z $\rightarrow \ell\ell$ branching ratios as measured at LEP: $\frac{\mathrm{Br}(Z\rightarrow\nu\nu)}{\mathrm{Br}(Z\rightarrow ee)} = 5.942 \pm 0.018$.[1] However, before obtaining the final Z $\rightarrow \nu\nu$+jets background prediction, there is another correction factor that needs to be included in the normalisation calculation to account for small differences in the geometrical jet acceptance of the Z $\rightarrow \ell\ell$+jets cross section measurement, compared to the geometrical volume available to jets in Z $\rightarrow \nu\nu$ events. This is because the removal of charged leptons from the event, which is done in order to calculate the $\not\!\!E_T$, reduces the geometrical volume available for jets. The jet acceptance therefore differs slightly and a correction ($A^{\mathrm{corr}}_{\mathrm{jets}}$) must be added, in order to get a correct prediction for the Z $\rightarrow \nu\nu$+jets background. The following equation summarises how the Z $\rightarrow \nu\nu$+jets background prediction is obtained from Z $\rightarrow \ell\ell$+jets events passing the same jet selections as those applied in the jets-plus-$\not\!\!E_T$ searches ($N_{\mathrm{Z}(\rightarrow\ell\ell)+\mathrm{jets}}$):

$$N(\mathrm{Z} \rightarrow \nu\nu + \mathrm{jet}) = 5.942 \times \frac{N_{\mathrm{Z}(\rightarrow\ell\ell)+\mathrm{jets}} - N_{\mathrm{bkg}}}{A_{\mathrm{lept}} \times \epsilon_{\mathrm{lept}}} \times A^{\mathrm{corr}}_{\mathrm{jets}} \times \epsilon_{\mathrm{trig}} \times \frac{L_{\mathrm{trig}}}{L_{\mathrm{lept}}}\,, \quad (1)$$

Table 1. Normalisation factors estimated for $Z \to \ell\ell$+1-jets cross section measurements ($\ell = e$ or μ) when the jet transverse energy and the $\not\!\!E_T$ are above 80 GeV. The numbers are for 1.1 fb^{-1} of CDF run II data. Note that here, the acceptance includes a Monte Carlo-based efficiency correction. To reflect the data–driven estimate of the lepton selection efficiencies, we also quote the scale factor ($\epsilon^{\rm SF}_{\rm lept} = \frac{\epsilon^{\rm Dat}_{\rm lept}}{\epsilon^{\rm MC}_{\rm lept}}$) that need to be applied to the quoted acceptances.

Normalisation factors	$Z \to ee$+1-jets	$Z \to \mu\mu$+1-jets
Raw data events ($N_{Z(\to\ell\ell)+\rm jets}$)	112	100
Acceptance ($A_{\rm lept} \times \epsilon^{\rm MC}_{\rm lept}$)	0.156 ± 0.006	0.240 ± 0.007
Efficiency SF ($\epsilon^{\rm SF}_{\rm lept}$)	1.017 ± 0.010	0.923 ± 0.023
Background ($N_{\rm bkg}$)	3 ± 3	negligible
Jets acceptance corr. ($A^{\rm corr}_{\rm jets}$)	0.957 ± 0.027	1.120 ± 0.025
$\not\!\!E_T$ trigger efficiency ($\epsilon_{\rm trig}$)	0.993	0.993
Luminosity correction ($\frac{L_{\rm trig}}{L_{\rm lept}}$)	0.887	0.930
$Z \to \nu\nu$+1-jets estimate	3470 ± 377	2798 ± 296

where $\epsilon_{\rm trig}$ is the efficiency of the trigger (jet or $\not\!\!E_T$ triggers) used to gather the jets-plus-$\not\!\!E_T$ data sample, $L_{\rm trig}$ is the integrated luminosity of this sample, and $L_{\rm lept}$ is the integrated luminosity of the sample collected with the lepton triggers. Table 1 gives an example of these normalisation corrections in the context of a $Z \to \nu\nu$+jets background estimate for monojet events where both the leading jet transverse energy and the $\not\!\!E_T$ of the events (after lepton removal) are above 80 GeV. As can be seen in the table, the total uncertainty on the number of $Z \to \nu\nu$+jets events predicted is higher than 10%, while the total uncertainty from each of the normalization factors adds up to less than 6%. The difference is due to the limited $Z \to \ell\ell$+jets statistics available for the high kinematic region.

There is an order of magnitude more $W \to \ell\nu$+jets than $Z \to \ell\ell$+jets events. If the $W \to \ell\nu$+jets events could be exploited to obtain a prediction for $Z \to \nu\nu$+jets events, the total uncertainty on the prediction would be significantly reduced. To do this, the ratio $R_{\rm jets}$ of $W \to \ell\nu$+jets to $Z \to \ell\ell$+jets cross sections is used. This ratio is calculated at Next-to-Leading order in perturbation theory and the resulting systematic uncertainties are relatively small. It should be noted that the use of the ratio cancels some of the uncertainties of the individual cross section calculations. By computing this ratio for the phase space regions under study, a $Z \to \nu\nu$+jets prediction can be obtained from $W \to \ell\nu$+jets events. For example, in the W/Z+1-jet case where the jet transverse energy and the $\not\!\!E_T$ are above 80 GeV, theoretical calculations performed with the MCFM program[22] yield $R_{\rm jets} = 9.0 \pm 0.3$. Applying the normalisation correction procedure outlined above on $W \to \ell\nu$+jets events gives a total of four statistically independent estimates for the $Z \to \nu\nu$+jets background to jets-plus-$\not\!\!E_T$ events. The details of the estimate from W+jets events for the CDF

Table 2. Normalisation factors estimated for W $\to \ell\nu$+1-jets cross section measurements ($\ell = e$ or μ) when the jet transverse energy and the $\not\!\!E_T$ are above 80 GeV. The numbers are for 1.1 fb^{-1} of CDF run II data. Note that here, the acceptance includes a Monte Carlo-based efficiency correction. To reflect the data-driven estimate of the lepton selection efficiencies, we also quote the scale factor ($\epsilon^{\rm SF}_{\rm lept} = \frac{\epsilon^{\rm Dat}_{\rm lept}}{\epsilon^{\rm MC}_{\rm lept}}$) that need to be applied to the quoted acceptances.

Normalisation factors	W $\to e\nu$+1-jets	W $\to \mu\nu$+1-jets
Raw data events ($N_{{\rm W}(\to\ell\nu)+{\rm jets}}$)	1610	1299
Acceptance ($A_{\rm lept} \times \epsilon^{\rm MC}_{\rm lept}$)	0.223 ± 0.006	0.244 ± 0.007
Efficiency SF ($\epsilon^{\rm SF}_{\rm lept}$)	0.965 ± 0.008	0.868 ± 0.005
Background ($N_{\rm bkg}$)	312 ± 47	232 ± 24
Jets acceptance corr. ($A^{\rm corr}_{\rm jets}$)	0.973 ± 0.026	1.055 ± 0.026
$\not\!\!E_T$ trigger efficiency ($\epsilon_{\rm trig}$)	0.993	0.993
Luminosity correction ($\frac{L_{\rm trig}}{L_{\rm lept}}$)	0.887	0.930
Z $\to \nu\nu$+1-jets estimate	3440 ± 231	3267 ± 179

monojet search are given as an example in Table 2. As can be seen in Tables 1 and 2, the four predictions are consistent within uncertainties. By combining them we obtain a final prediction of 3207 ± 138 Z $\to \nu\nu$+jets background events. The relative uncertainty on this combined prediction using both W+jets and Z+jets events is two times smaller than what would be obtained from Z $\to \ell\ell$+jets events alone, and about five time smaller than what would be obtained from Monte Carlo-based predictions.

Starting from the measurement of W $\to \ell\nu$+jets cross sections, we can estimate the contribution of the other W/Z+jets backgrounds to the jets-plus-$\not\!\!E_T$ candidate sample. To this end, one simply needs to estimate the fraction of events in which the charged lepton is not reconstructed, is outside of the detector acceptance or otherwise fails to be rejected by the lepton veto requirements. Such probability to "lose" the charged leptons (P($\ell \to \not\ell$)) must be estimated from Monte Carlo. However, by exploiting the relevant ratios of Monte Carlo events to obtain such estimate, systematic uncertainties due to non-perturbative QCD effects (e.g. hadronisation, PDF, etc.) or due to the modelling of detector effects does not significantly increase the overall uncertainty on the final prediction. Table 3 displays these probabilities that the lepton survives veto cuts for the various lepton types in the context of a CDF search for new physics in monojet events. As can be seen in the table, electrons are less likely to be lost than muons because of the wider coverage of the CDF calorimeter compared to its tracker. The fact that the tau background is the largest is a consequence of events where a tau fakes a jet object, enhancing their contribution to jets-plus-$\not\!\!E_T$ events. Although the probability to lose the two charged leptons from Z $\to \ell\ell$+jets events is small, their contribution to jets-plus-$\not\!\!E_T$

Table 3. Probability that all the charged leptons of a $W \to \ell\nu$+jets or a $Z \to \ell\ell$+jets event get lost, i.e. survive the lepton veto requirements, when the jet transverse energy and the $\not\!\!E_T$ are above 80 GeV.

Processes	P($e \to \not e$)	P($\mu \to \not\mu$)	P($\tau \to \not\tau$)
$W \to \ell\nu$+1-jets	$20.8 \pm 0.3\%$	$33.0 \pm 1.0\%$	$54.6 \pm 0.8\%$
$Z \to \ell\ell$+1-jets	0.0%	$11.9 \pm 0.2\%$	$7.9 \pm 0.2\%$

events have been estimated in the same way as for the W+jets contribution. This completes the data-driven estimate of the electroweak backgrounds to jets-plus-$\not\!\!E_T$ events which amounts to $\sim 90\%$ of the total background.

5. CDF Results

The data-driven ideas and techniques to estimate Standard Model backgrounds to jets-plus-$\not\!\!E_T$ events have been used in CDF for monojet and dijet+$\not\!\!E_T$ signature-based searches for new physics. In order to illustrate how these methods perform in actual measurements, we report the results of these two searches. Examples of interpretations of the result obtained from the comparison of the predictions with the observations are also provided.

5.1. *Monojet search*

A background prediction for three different kinematic regions is obtained for the monojet analysis using the data-driven techniques presented above. This increases the sensitivity to different new physics scenarios, while keeping a model-independent approach. As mentioned before, two of those kinematic regions are determined from the trigger requirement that the $\not\!\!E_T$ or jet triggers are fully efficient. As a consequence, a requirement of 80 GeV on the jet E_T and on the $\not\!\!E_T$ is imposed on the monojet sample collected from the $\not\!\!E_T$ trigger, while a jet of $E_T > 150$ GeV and 120 GeV of $\not\!\!E_T$ are required for monojet events collected from the jet trigger. A third region is probed based on the criteria that there are enough W/Z+jets events with a well-reconstructed lepton to perform a data-driven background estimate with 1 fb^{-1} of data. A jet of 180 GeV and an $\not\!\!E_T$ of 150 GeV are then required to define this third kinematic region probed for new physics. As can be seen in Table 4, the observations in the three kinematic regions probed for new physics are consistent with Standard Model expectations obtained with the methods outlined above. Although the search performed consists in a counting experiment, a comparison between the observed and predicted $\not\!\!E_T$ distributions for monojet events is provided in Fig. 2. It shows that the agreement between Standard Model predictions and the monojet data is good over the entire distribution. From these results, no evidence for new physics in 1 fb^{-1} was found in the CDF monojet data.

Although no new physics has been found with these monojet analyses, results can be used to constrain the parameters of different new physics scenarios which

Table 4. Estimated SM backgrounds and the number of observed monojet data events for 80/80 ($E_T > 80$ GeV, $\not\!\!E_T > 80$ GeV), 150/120 ($E_T > 150$ GeV, $\not\!\!E_T > 120$ GeV) and 180/150 ($E_T > 180$ GeV, $\not\!\!E_T > 150$ GeV) candidate samples.

Background	80/80	150/120	180/150
$Z \to \nu\nu$+jets	3207 ± 138	338 ± 30	139 ± 17
$W \to e\nu$+jets	1959 ± 67	187 ± 14	58 ± 5
$W \to \mu\nu$+jets	1530 ± 53	117 ± 9	35 ± 3
$W \to \tau\nu$+jets	808 ± 28	58 ± 4	18 ± 2
$Z \to \ell\ell$+jets	96 ± 4	8 ± 1	2 ± 0
QCD multi-jet	708 ± 146	23 ± 20	12 ± 12
γ+jets	209 ± 41	17 ± 5	8 ± 3
Non-collision	52 ± 52	10 ± 10	3 ± 3
Total expected	8564 ± 331	808 ± 62	275 ± 30
Data observed	8449	809	319

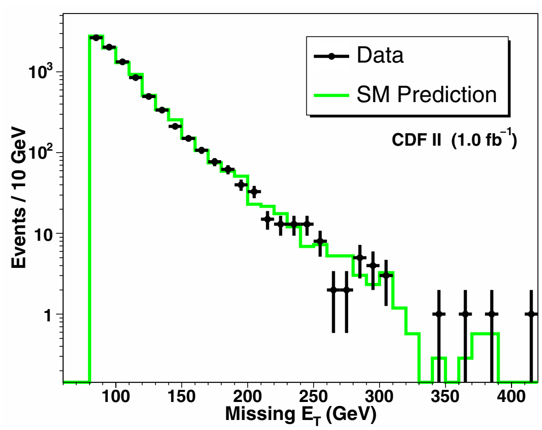

Fig. 2. Comparison of the event $\not\!\!E_T$ for the 8449 events in the candidate sample used for the 80/80 monojet search for new physics with the predicted Standard Model distribution. The prediction does not include a systematic uncertainty on the shape of the Standard Model background distribution which is directly taken from Monte Carlo.

predict an enhanced monojet signal with respect to Standard Model predictions. From the number of observed events, the predicted number of Standard Model background events, and the uncertainty on this background prediction, an upper limit on the number of new physics events contributing to our sample events can be obtained. For example, in the 150/120 region, the measurement is consistent at 95% confidence level with a maximum of 125 new physics events. This number is essentially model-independent and can be used to set limits on the parameters of any new physics model of interest.

Table 5. 95% C.L. lower limits on the fundamental Planck scale M_D in TeV obtained from the 150/120 monojet analysis as a function of $n = 2$–6 extra dimensions. Comparison with LEP results are also provided.

	M_D (TeV)	
Extra dimensions	CDF monojet	LEP combined
$n = 2$	1.33	1.60
$n = 3$	1.10	1.20
$n = 4$	0.99	0.94
$n = 5$	0.92	0.77
$n = 6$	0.89	0.66

Graviton production in the context of Large Extra Dimension (LED) models have been chosen as a benchmark model that can be constrained by the monojet measurement. After an *a priori* comparison, the 150/120 measurement was determined to be the most sensitive to LEDs and therefore was used for the calculation of the limits on the fundamental Planck scale in $4 + n$ dimensions. To that end, the maximum number of new physics events obtained above can be converted into an upper limit on the monojet production cross section from LED models[23] after calculation of the signal acceptance. Monte Carlo simulations are used for this estimate. For example, in the case of two large extra dimensions, the signal acceptance is estimated to $9.9 \pm 1.3\%$. The uncertainty includes effects from the modelling of jet energy scale and resolution (8%), partons density function (6%), initial and final state radiation (3%) and integrated luminosity (6%). This systematic uncertainty is twice as large as the total uncertainty on the background prediction for the 150/120 measurement, which is essentially all statistical. This illustrates the gain that can be obtained from using data-driven estimates of the Standard Model background. Note that this gain is larger for lower kinematic regions where the statistics of the control sample are larger. For example, in the region where both the jet E_T and the $\not\!\!E_T$ are required to be above 8 GeV, the total uncertainty on the Standard Model background prediction is less than 4%, which is approximately a factor of four lower than what would have been obtained from Monte Carlo based predictions. This shows that the systematic uncertainty on background prediction is well controlled when this prediction is extracted directly from data. The upper limit on the graviton production cross section in the ADD scenario can then be used to constrain the fundamental Planck scale in $4 + n$ dimensions M_D. Table 5 presents the lower limit on M_D as a function of the number of large extra dimensions n, and is compared with LEP limits on M_D obtained from photon-plus-$\not\!\!E_T$ events. LEP constraints are better than CDF limits for $n = 2$ and $n = 3$. However, these two cases are significantly disfavoured by astronomical considerations.[24–27] The cases $n > 3$ are not disfavoured by astrophysics, and the CDF monojet results set the best constraints on the fundamental Planck scale M_D.

Table 6. Estimate SM of backgrounds and the number of observed dijet+$\not\!\!E_T$ event for loose ($H_T > 125$ GeV, $\not\!\!E_T > 80$ GeV) and ($H_T > 225$ GeV, $\not\!\!E_T > 100$ GeV) tight candidate samples.

Background	Loose Sample	Tight Sample
Z $\to \nu\nu$+jets	888 ± 54	86 ± 13
W $\to e\nu$+jets	669 ± 42	50 ± 8
W $\to \mu\nu$+jets	399 ± 25	33 ± 5
W $\to \tau\nu$+jets	256 ± 16	14 ± 2
Z $\to ee$+jets	0 ± 0	0 ± 0
Z $\to \mu\mu$+jets	13 ± 2	1 ± 0
Z $\to \tau\tau$+jets	16 ± 2	1 ± 0
Top quark production	74 ± 9	11 ± 2
QCD multi-jet	49 ± 30	9 ± 9
γ+jets	75 ± 11	5 ± 1
Non-collision	4 ± 4	1 ± 1
Total expected	2443 ± 145	211 ± 30
Data observed	2506	186

5.2. *Dijet+$\not\!\!E_T$ search*

As in the case of the monojet analysis, different kinematic regions have been probed in the dijet+$\not\!\!E_T$ analysis. These regions have been defined in terms of the $\not\!\!E_T$ and the scalar sum of the transverse energy of each of the two jets contained in the event ($H_T = E_T(j_1) + E_T(j_2)$). From the trigger requirement ($\not\!\!E_T$ trigger), the lowest kinematic region is defined as $\not\!\!E_T > 80$ GeV and $H_T > 125$ GeV. We refer to this region as the loose sample. By requiring that there are enough W/Z+jets events to perform a data-driven estimate of the Standard Model contribution to dijet+$\not\!\!E_T$ events with 2 fb^{-1} of CDF data, a second kinematic region has been defined as $\not\!\!E_T > 100$ GeV and $H_T > 225$ GeV (tight sample). Once again, predictions have been independently obtained for these two regions. Results of the predictions and a comparison with the observed number of dijet+$\not\!\!E_T$ events are presented in Table 6. The agreement between the observations and the Standard Model predictions are excellent, showing no evidence for new physics. As was done in the monojet analysis, differential distributions were compared to the observed data and the agreement between with Standard Model predictions was again good. Results for the $\not\!\!E_T$ distribution of the loose sample are displayed in Fig. 3.

The results above can be used to constrain any new physics model which predicts a dijet+$\not\!\!E_T$ signature. A 95% C.L. limit on the number of new physics events in each of the candidate samples is obtained from the number of observed events, the number of predicted Standard Model events and its uncertainty. For the loose region, this upper limit consists of 359 events, while for the tight sample this limit is

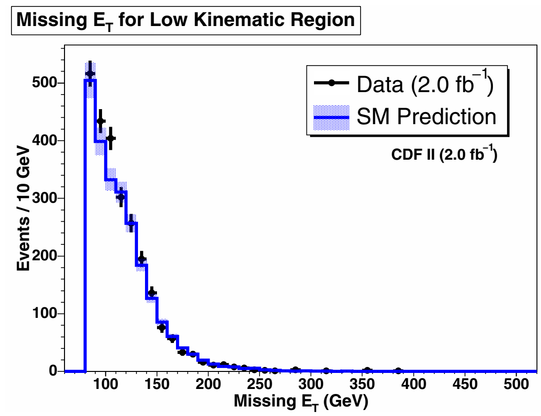

Fig. 3. Comparison of the event $\not\!\!E_T$ for the 2,443 events in the candidate sample used for the loose dijet+$\not\!\!E_T$ search for new physics with the predicted Standard Model distribution. The prediction does not include a systematic uncertainty on the shape of the Standard Model background distribution which is directly taken from Monte Carlo.

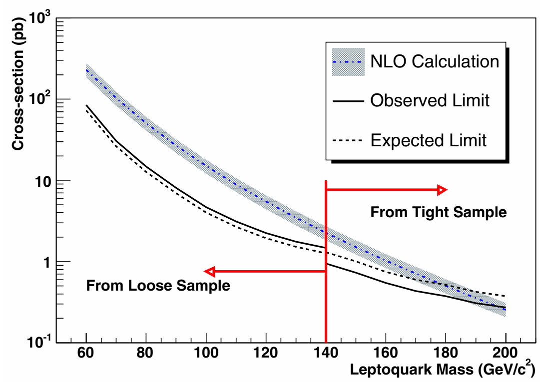

Fig. 4. 95% C.L. cross section limits on first and second generations $q\nu$ scalar leptoquark pair production (q being u, d, s or c) as a function of leptoquark masses M_{LQ}). The lower limit on the leptoquark mass is obtained from the intersection between the observed limit (straight black line) and the NLO calculation (blue dotted line).

51 events. After the signal acceptance to dijet+$\not\!\!E_T$ selection is calculated, the upper limits on new physics model cross sections can be obtained. Leptoquark models for which a pair of leptoquarks is produced each decaying to a quark and a neutrino was chosen for its simplicity: it depends only on two free parameters (the leptoquark mass and branching ratio), facilitating comparisons with other experimental results. For each tested leptoquark mass, an *a priori* choice of the kinematic region to be used for the constraints was made. Up to 140 GeV, constraints have been set from the loose sample prediction. Results are shown in Fig. 4. As can be seen in the figure, the mass point for which the 95% C.L. distribution from data crosses the leptoquark cross section distribution marks the upper limit set by the dijet+$\not\!\!E_T$

measurement on the leptoquark mass. This limit of 190 GeV/c^2 corresponds to an upper limit of 0.31 pb on the leptoquark pair production cross section.

6. Conclusion

Jets-plus-$\not\!\!E_T$ signatures have a lot of potential for discoveries of new physics beyond the Standard Model. These signatures are predicted in many new physics scenarios and in some of those models, Dark Matter particles are responsible for the observed $\not\!\!E_T$.

In order to maximize the sensitivity and robustness of these searches, data-driven methods were developed and used in measurements performed at the Tevatron with the CDF detector. Two examples of such searches were discussed: a monojet and a dijet+$\not\!\!E_T$ analysis. The advantage of data-driven background estimates applied in this context in reducing the overall uncertainty and in increasing the search sensitivity was demonstrated. The CDF searches found no evidence of physics beyond the Standard Model, and constraints on parameters of new physics models were obtained. In particular, the monojet analysis set limits on the fundamental Planck scale of a large extra dimension model while the dijet+$\not\!\!E_T$ analysis was used to set limits on the mass of leptoquarks.

The strategies and methods presented in this review and tested at the Tevatron will hopefully play an important role in future discoveries at the LHC.

Acknowledgments

The authors would like to thank Dr. Eric James and Dr. Dan MacQueen for their help in developing the analyses and techniques described in this review. We also thank our CDF collaborators, the Fermilab staff and the technical staffs of the participating institutions for their vital contributions.

References

1. W.-M. Yao *et al.*, *J. Phys. G* **33**, 1 (2006), also see http://pdg.lbl.gov
2. For example of review, see *TASI Lectures on Precision Electroweak Physics*, K. Matchev, UFIFT-HEP-03-11, CLNS-04/1862.
3. For example of review, see *TASI Lectures on The Strong CP Problem*, M. Dine, SCIPP-00/30.
4. J. L. Feng, *J.Phys. G* **32**, R1 (2006).
5. CDF Collab. (A. Abulencia *et al.*), *Phys. Rev. Lett.* **97**, 171802 (2006).
6. CDF Collab. (T. Aaltonen *et al.*), *Phys. Rev. Lett.* **101**, 181602 (2008).
7. CDF Collab. (T. Aaltonen *et al.*), submitted to *Phys. Rev. Lett.*, arXiv:0912.4691.
8. G. L. Kane and J. P. Leville, *Phys. Lett. B* **112**, 227 (1982).
9. P. R. Harrison and C. H. Llewellyn-Smith, *Nucl. Phys. B* **213**, 223 (1983).
10. I. Antoniadis, N. Arkani-Hamed, S. Dimopoulos and G. Dvali, *Phys. Lett. B* **436**, 257 (1998).
11. G. Azuelos, P.-H. Beauchemin and C. P. Burgess, *J. Phys. G: Nucl. Part.* **31**, 1 (2005).
12. D. Atwood, C. P. Burgess, E. Filotas, F. Leblond, D. London and I. Maksymyk, *Phys. Rev. D* **63**, 025007 (2001).

13. J. L. Hewett and T. G. Rizzo, *Phys. Rev. D* **56**, 5709 (1997).
14. M. Krämer, T. Plehn, M. Spira and P. M. Zerwas, *Phys. Rev. Lett.* **79**, 341 (1997).
15. H. Goldberg, *Phys. Rev. Lett.* **50**, 1419 (1983).
16. J. R. Ellis, J. S. Hagelin, D. V. Nanopoulos, K. A. Olive and M. Srednicki, *Nucl. Phys. B* **238**, 453 (1984).
17. G. Servant and T. M. P. Tait, *Nucl. Phys. B* **650**, 391 (2003).
18. H. C. Cheng, J. L. Feng and K. T. Matchev, *Phys. Rev. Lett.* **89**, 211301 (2002).
19. H. C. Cheng and I. Low, *JHEP* **0309**, 051 (2003).
20. CDF Collab. (F. Abe *et al.*), *Phys. Rev. D* **45**, 1448 (1992).
21. A. Bhatti *et al.*, *Nucl. Instrum. Meth. A* **566**, 375 (2006).
22. J. Campbell and K. Ellis, *Monte carlo for FeMtobarn processes at hadron Colliders*, see http://mcfm.fnal.gov/.
23. G.F. Giudice, R. Rattazzi and J. D. Wells, *Nucl. Phys. B* **544**, 3 (1999).
24. S. Hannestad and G. Raffelt, *Phys. Rev. Lett.* **87**, 051301 (2001).
25. S. Hannestad and G. Raffelt, *Phys. Rev. Lett.* **88**, 071301 (2002).
26. S. Hannestad and G. Raffelt, *Phys. Rev. D* **67**, 125008 (2003) [Erratum-*ibid.* **69**, 029901 (2004)].
27. S. Hannestad, *Phys. Rev. D* **64**, 023515 (2001).