A Deep Neural Network for Missing Transverse Momentum Reconstruction in ATLAS

Matthew Leigh

University of Cape Town, South Africa

E-mail: `matthew.leigh@cern.ch`

Abstract. The missing transverse momentum ($\mathbf{E}_\mathrm{T}^\mathrm{miss}$) of a proton-proton ($pp$) collision is an important observable as it serves as an experimental proxy for the net transverse momentum carried away by undetected particles. Precisely measuring $\mathbf{E}_\mathrm{T}^\mathrm{miss}$ is critical for the understanding many physical processes which take place at the Large Hadron Collider. The ATLAS experiment currently utilises several algorithms to reconstruct the $\mathbf{E}_\mathrm{T}^\mathrm{miss}$. These proceedings describe the development of a novel algorithm based on a deep neural network to estimate $\mathbf{E}_\mathrm{T}^\mathrm{miss}$. The network reconstruction had resolution of 22.7 GeV when tested in simulated $t\bar{t}$ events which was better than the next best performing algorithm which had a resolution of 27.3 GeV. The new estimate was more robust to an increase in pileup and outperformed all other methods across the full pileup range. The network was also demonstrated to generalise well in real Z events, improving the resolution from 15.53 to 11.29 GeV.

1. Introduction

The ATLAS detector [1] is an example of a hermetic detector. It was thus designed to observe nearly all possible decay products produced in the high-energy pp collisions provided by the Large Hadron Collider (LHC) at CERN. Standard Model neutrinos and many theorised particles in beyond the SM physics are weakly interacting and are able travel through the detector without leaving any measurable signal. However, since the total linear momentum of the colliding protons in the transverse plane is negligible, the presence of these undetectable particles can be inferred by the resultant transverse momentum imbalance of all observable decay products emerging from the collision. The negative vectoral sum of the visible momenta in the transverse plane, $\mathbf{E}_\mathrm{T}^\mathrm{miss}$, therefore serves as an experimental proxy for the net transverse momentum of undetected particles. This is an essential part of many ongoing physics analyses at the LHC [2,3], including the discovery of the Higgs boson in 2012 [4]. Reconstructing $\mathbf{E}_\mathrm{T}^\mathrm{miss}$ requires the output of all detector subsystems, as well as the most complete representation of the hard scatter (defined in ATLAS as the vertex with the highest $\sum(p_\mathrm{T}^\mathrm{track})^2$) which is obscured by the presence of additional "pileup" interactions from other pp scatters in the same bunch crossing. Thus $\mathbf{E}_\mathrm{T}^\mathrm{miss}$ resolution is expected to get worse as the luminosity of the LHC, and thus the number of pileup interactions, increases over the next few years. Monte-Carlo (MC) simulations are used in most analyses to provide a baseline from which one can compare real data captured by ATLAS. To produce this simulated data, the underlying parton interaction is generated, soft radiation and hadronization modelling is applied, before the passage of the resultant stable particles through a detector simulation, to estimate the signals they would produce [5]. The event reconstruction process applied to MC is identical to the one applied to real data. In this project we extracted

and stored the true transverse momentum of all non-interacting particles in each MC event before the detector simulation, called $\mathbf{E}_{\mathrm{T}}^{\mathrm{miss,\ true}}$. A deep neural network [6] is then trained to learn a mapping from a set of reconstructed observables to $\mathbf{E}_{\mathrm{T}}^{\mathrm{miss,\ true}}$. This mapping is then applied and evaluated on real and MC datasets, all of which are orthogonal from those used to train the network.

2. The ATLAS Detector

The ATLAS detector at the LHC is a general purpose particle detector with nearly 4π solid angle coverage and a nominal forward-backward symmetry. It consists of an inner tracking detector (ID) surrounded by a superconducting solenoid magnet which provides a 2 T magnetic field. Encompassing the ID are electromagnetic (EM) and hadronic calorimeters, and a muon-spectrometer (MS). The ID consists of a pixel detector, a semiconductor tracker and a transition radiation tracking detector. It provides tracking information in a pseudo rapidity range of $\eta < |2.5|$ [1]. The EM calorimeter provides high granularity energy measurements over a range of $\eta < |3.2|$. Steel-scintillating hadronic calorimeters provides central coverage within $\eta < |1.7|$. End-cap regions of the detector contain additional calorimeters up to $\eta < |4.9|$. The MS is the outermost layer of the ATLAS detector and features three large air-core toroidal superconducting magnet systems with eight coils each. It contains precision tracking chambers covering $\eta < |2.7|$. ATLAS data-taking utilises a two-level trigger system. The first is a hardware-based trigger which reduces the event rate to around 100 kHz. The second is a software-based high level trigger, which further reduces the rate to approximately 1 kHz.

3. Missing Transverse Momentum Reconstruction at ATLAS

Over the past few years, different algorithms have been developed to reconstruct $\mathbf{E}_{\mathrm{T}}^{\mathrm{miss}}$ at ATLAS [7]. ATLAS currently offers several variants for physics analyses with different requirements, several of which are used in this project [8]. Most definitions are object based and are characterised by two contributions. The first one is from *hard-event* signals from identified and calibrated physics objects associated with the hard-scatter. These are fully reconstructed jets, photons, electrons, and muons. Dedicated rejection procedures are carried out to ensure that all contributing objects were reconstructed from mutually exclusive detector signals. The second contribution to $\mathbf{E}_{\mathrm{T}}^{\mathrm{miss}}$ is from *soft-event* signals. These are detector signals that are not used to reconstruct the objects mentioned above. Most algorithms use a Track Soft Term (TST), which is created purely from unused particle tracks in the ID which are associated with the hard scatter. The tracks are required to have passed high-quality reconstruction requirements and have $\eta < |2.5|$ and $p_T > 0.4$ GeV. An example of this algorithm is shown in Equation 1, which show the negative vectoral summation of the terms.

$$\mathbf{E}_{\mathrm{T}}^{\mathrm{miss}} = - \underbrace{\sum_{\substack{\text{selected} \\ \text{electrons}}} \mathbf{p}_{\mathrm{T}}^{e}}_{\mathbf{E}_{\mathrm{T}}^{\mathrm{miss},e}} - \underbrace{\sum_{\substack{\text{selected} \\ \text{muons}}} \mathbf{p}_{\mathrm{T}}^{\mu}}_{\mathbf{E}_{\mathrm{T}}^{\mathrm{miss},\mu}} - \underbrace{\sum_{\substack{\text{selected} \\ \text{photons}}} \mathbf{p}_{\mathrm{T}}^{\gamma}}_{\mathbf{E}_{\mathrm{T}}^{\mathrm{miss},\gamma}} - \underbrace{\sum_{\substack{\text{selected} \\ \text{jets}}} \mathbf{p}_{\mathrm{T}}^{\mathrm{jet}}}_{\mathbf{E}_{\mathrm{T}}^{\mathrm{miss,\ jet}}} - \underbrace{\sum_{\substack{\text{unused} \\ \text{tracks}}} \mathbf{p}_{\mathrm{T}}^{\mathrm{track}}}_{\mathbf{E}_{\mathrm{T}}^{\mathrm{miss,\ soft}}} \tag{1}$$

$$\underbrace{\phantom{\mathbf{E}_{\mathrm{T}}^{\mathrm{miss},e} \quad \mathbf{E}_{\mathrm{T}}^{\mathrm{miss},\mu} \quad \mathbf{E}_{\mathrm{T}}^{\mathrm{miss},\gamma} \quad \mathbf{E}_{\mathrm{T}}^{\mathrm{miss,\ jet}}}}_{\text{hard terms}} \quad \underbrace{\phantom{\mathbf{E}_{\mathrm{T}}^{\mathrm{miss,\ soft}}}}_{\text{soft term}}$$

The total transverse energy in the detector, ΣE_{T}, is a scalar which quantifies the total event activity. This is a crucial variable for understanding the scale and resolution of $\mathbf{E}_{\mathrm{T}}^{\mathrm{miss}}$. It uses

---

[1] ATLAS uses a right-handed coordinate system with its origin at the nominal interaction point in the center of the detector. The z-axis lies along the beam pipe, the x-axis points towards the center of the LHC ring, and the y-axis points vertically upwards. Cylindrical coordinates are used in the transverse plane where ϕ is the azimuthal angle around the z-axis. The pseudorapidity is defined in terms of the polar angle as $\eta \equiv -\log \tan \frac{\theta}{2}$, where θ is the polar angle off the z-axis.

the same terms, and follows the same overlap removal procedures, but is created using the scalar sum of the various transverse momenta. The magnitude of the missing transverse momentum ($E_\mathrm{T}^\mathrm{miss}$), must be non-negative by definition, and in a realistic environment where not all relevant $\mathbf{p}_\mathrm{T}$ terms are accurately represented in Equation 1, an observation bias towards non-vanishing values for $E_\mathrm{T}^\mathrm{miss}$ is introduced. This is most noticeable in final states without genuine missing transverse momentum, meaning that due to measurements $E_\mathrm{T}^\mathrm{miss}$ is often non-zero even when there are no undetected particles.

3.1. Hard Terms

Muons are identified by matching an ID track with an MS track or segment, and are required to have $p_T > 10$ GeV to be included in the muon term for $\mathbf{E}_\mathrm{T}^\mathrm{miss}$. Electrons are reconstructed from clusters in the EM calorimeter associated with an ID track and must also have $p_T > 10$ GeV to contribute to $\mathbf{E}_\mathrm{T}^\mathrm{miss}$ or ΣE_T. Photons are identified from their very distinctive electromagnetic showers in the calorimeters. In this project, photon candidates are required to satisfy a set of tight criteria to reduce backgrounds and must have $p_T > 25$ GeV. Jets are reconstructed from clusters of topologically connected calorimeter cells using the anti-kt algorithm [9] with a radius parameter of $R = 0.4$. All jets are first required to have $p_T > 20$ GeV after calibration and $\eta < |4.5|$. Jets are further decorated using a tagging algorithm to select jets likely from emerging from the hard-scatter, known as jet-vertex-tagging (JVT) [10]. The JVT value attributed to a jet ranges from 0 (likely from pile-up) to 1 (hard-scatter). All central jets ($\eta < |2.5|$) are required to have either a JVT value greater than 0.59 or $p_T > 60$ GeV.

3.2. Alternate Definitions of $\mathbf{E}_T^{miss}$

Five different algorithms are used and compared in this work. The first three are constructed as mentioned above and only differ in their treatment of forward jets. *Loose* $\mathbf{E}_\mathrm{T}^\mathrm{miss}$ includes contributions from any forward jet with $p_T > 20$ GeV. *Tight* $\mathbf{E}_\mathrm{T}^\mathrm{miss}$ removes all forward jets with $p_T < 30$ GeV. *FJVT* $\mathbf{E}_\mathrm{T}^\mathrm{miss}$ once again uses the same terms as Loose, but all forward jets with $20 > pT > 50$ GeV are required to pass the Loose FJVT criteria [11]. *CST* $\mathbf{E}_\mathrm{T}^\mathrm{miss}$ uses the same jet selection as the Tight, but replaces the TST with a more inclusive soft term created from unused calorimeter clusters [12]. Although noise suppression is applied, no additional pileup suppression techniques are used. The fifth and final algorithm used in this work is somewhat distinct from the others. *Track* $\mathbf{E}_\mathrm{T}^\mathrm{miss}$ is not object based and only employs ID tracks. It is therefore inherently more immune to pileup but ignores all neutral particles.

4. The Deep Neural Network

The main goal of this project was to produce a new definition, *Network* $\mathbf{E}_\mathrm{T}^\mathrm{miss}$, that would, on an event by event basis, choose the unique non-linear combination of all other algorithms to produce a more accurate estimate. The model chosen was a dense feed-forward artificial neural network (ANN) [6]. The network did not receive any raw detector signals or even individual particles as inputs. This was due to the desired scope of the project, which was to provide a final level correction to the $\mathbf{E}_\mathrm{T}^\mathrm{miss}$ of an event only, and not to redefine object selection or identification. 65 variables were chosen to be inputs for the neural network. These included the outputs of the 5 mentioned algorithms, their unique hard and soft terms and additional variables which might indicate the accuracy of these definitions. These additional variables included the number of reconstructed vertices, which indicates the amount of pileup, the amount of forward jet activity and object based $\mathbf{E}_\mathrm{T}^\mathrm{miss}$ significance [13].

4.1. Network Training

The network was trained on MC simulated datasets where the target variable on an event by event basis was $\mathbf{E}_\mathrm{T}^{\mathrm{miss,\ true}}$. The dataset was made up of 8 million $t\bar{t}$, Wt, $W\bar{t}$ and diboson (WW,

WZ, ZZ) processes. Underlying events were generated using Powheg-Box [14] and subsequent parton showers and hadronisation were handled by Pythia [15].

A grid search to was performed to find the optimal parameter features. Over the course of this work, more than 1000 different networks were tested. Features that were varied were the network depth, width, optimiser, learning rate, activation functions, and regularisation technique. The final network had five hidden layers, each with 1000 neurons featuring a Swish activation function [16]. After each non-linearity a dropout layer with $p = 0.2$ was applied. Gradient descent was performed by the ADAM optimiser [17] based on Huber-Loss [18].

5. Performance on $Z \rightarrow \mu\mu$ and $t\bar{t}$ events

5.1. Event Selection

Events involving $t\bar{t}$ production provide a good topology to measure the reconstruction performance, especially in interactions with large jet multiplicities. Studies in this region involved MC simulation only, thus the absence of background allows for very loose selection criteria, so only standard ATLAS event quality checks were performed [19].

The $Z \rightarrow \mu\mu$ final state is ideal for the evaluation of $\mathbf{E}_T^{\mathrm{miss}}$ performance in real events. Z boson production is relatively abundant at the LHC and can be selected with high signal to background ratios, and these events have well understood topologies. In this channel neutrinos are produced only through very rare heavy-flavour decays in the hadronic recoil. Therefore the distribution of $\mathbf{E}_T^{\mathrm{miss}}$ is strongly peaked close to zero with a finite width that can be largely attributed to miss-measurements of the inputs to the $\mathbf{E}_T^{\mathrm{miss}}$ algorithm. Events were selected if they fired a single lepton trigger and contained exactly two oppositely charged muons with $p_T > 25$ GeV, which have an invariant mass between 76 GeV and 106 GeV. Data used in this channel was recorded by the ATLAS experiment in 2017 at $\sqrt{s} = 13$ TeV with a total integrated luminosity of 44.3 fb^{-1}.

All events in both channels had to pass standard detector quality assessment criteria. Monte-Carlo samples of $Z \rightarrow \mu\mu$ were generated using Sherpa [20]. $t\bar{t}$, single-t and diboson events were generated using Powheg, with Pythia handling the parton shower.

5.2. Results

The resolution is determined by the width of the combined distribution of the differences between the measured $E_{x(y)}^{\mathrm{miss}}$ and the components of the true $\mathbf{E}_T^{\mathrm{miss, true}}$. The width is taken from the of the root mean square error (RMSE) of the distributions,

$$\mathrm{RMSE} = \begin{cases} RMS(E_{x(y)}^{\mathrm{miss}} - E_{x(y)}^{\mathrm{miss, true}}) & t\bar{t} \text{ sample } (E_T^{\mathrm{miss, true}} > 0) \\ RMS(E_{x(y)}^{\mathrm{miss}}) & Z \rightarrow \mu\mu \text{ sample } (E_T^{\mathrm{miss, true}} = 0) \end{cases} \quad (2)$$

The performance of the different algorithms was compared to each other in simulated $t\bar{t}$ events. In Figure 1a, the resolution is plotted with respect to the ΣE_T, which can be taken as a measurement of the hardness of the interaction, providing a useful scale for the evaluation of the resolution. The network outperforms all other estimates and shows less degradation with increased event activity. Furthermore, from the plots in Figure 1b, the network shows an excellent resilience to pileup, based on the consistently greater resolution with respect to the number of reconstructed vertices.

Data vs MC simulation comparisons for the Tight and the Network $\mathbf{E}_T^{\mathrm{miss}}$ algorithms in the $Z \rightarrow \mu\mu$ channel are shown in Figure 2. The error on the ratio points show the uncertainty of the MC samples, accounting for uncertainty in luminosity, cross section and statistics, but not for the Jet Energy Scale systematic uncertainty (which would be the largest factor in most of the bins but has yet to be added to this work). Good agreement in the bulk and the same overall shape, showing that the network, which was trained on MC datasets, is still able to generalise well to

real data. Over 98% of events in the MC $Z \to \mu\mu$ contribution, which significantly dominates all other processes, have $|\mathbf{E}_{\mathrm{T}}^{\mathrm{miss,\ true}}|$ less than 5 GeV and while the aforementioned observation bias is still visible in both plots, the network more accurately reflects this distribution and leads to a greater distinction between the signal and background processes. The average resolution using all algorithms applied to both datasets are shown in Table 1.
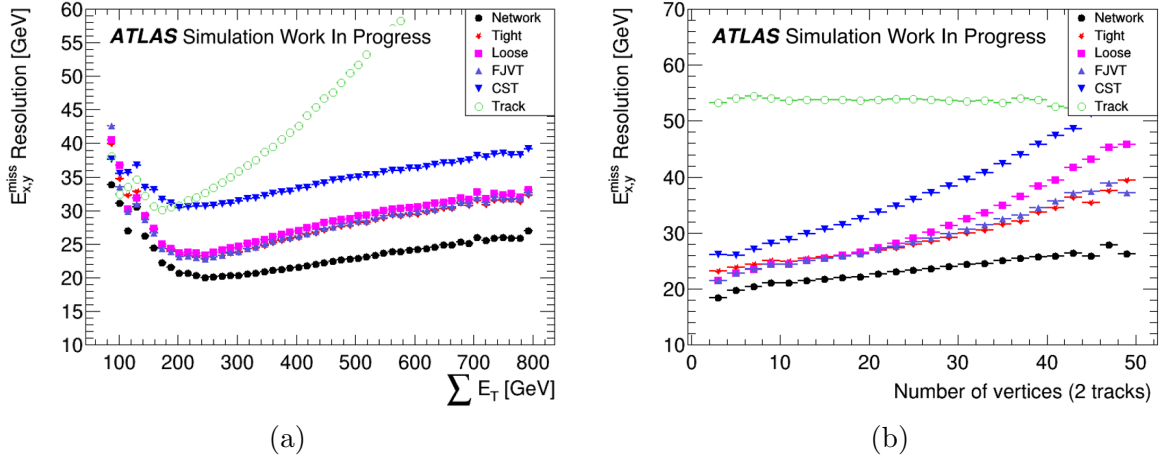
Figure 1. Comparisons of the $\mathbf{E}_{\mathrm{T}}^{\mathrm{miss}}$ resolution using different algorithms with respect to the ΣE_{T} (a) of the event and the number of reconstructed vertices (b).
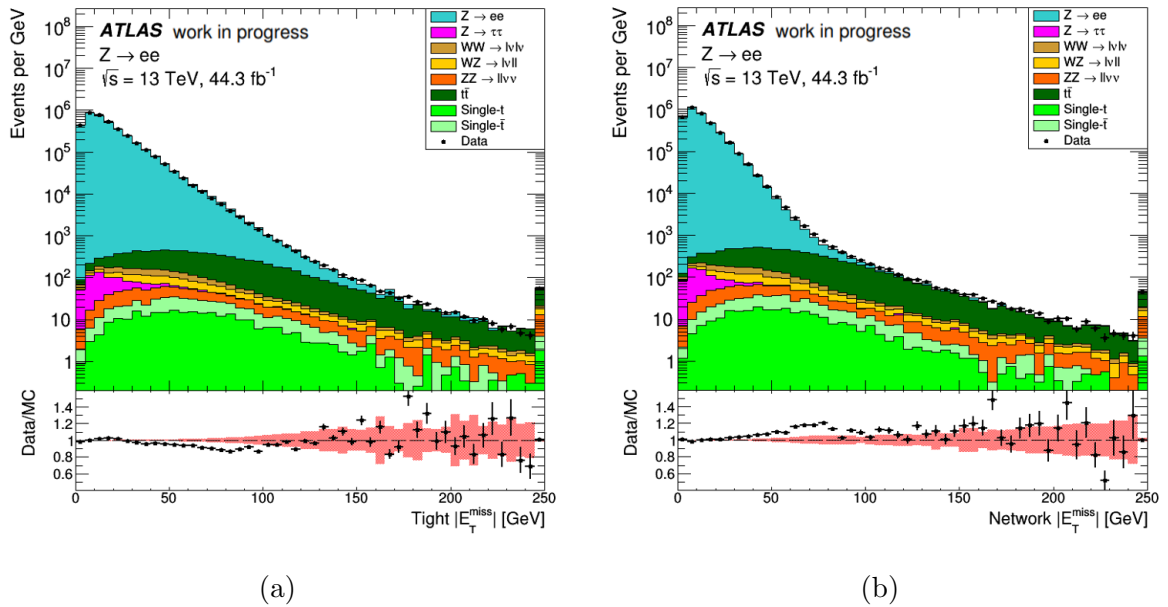
Figure 2. Data to Monte-Carlo comparisons of the magnitude of the $\mathbf{E}_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction using the Tight (a) and the Network (b) definitions in the $Z \to \mu\mu$ channel.

	$t\bar{t}$	$Z \to \mu\mu$
Tight	27.30	15.53
Loose	28.20	18.06
FJVT	27.48	16.55
CST	34.31	24.24
Track	53.96	18.48
Network	**22.70**	**11.29**

Table 1. The $\mathbf{E}_\mathrm{T}^\mathrm{miss}$ resolution, measured in GeV, of the different algorithms measured on the the simulated $t\bar{t}$ and real $Z \to \mu\mu$ datasets.

6. Conclusions

The measurement of $\mathbf{E}_\mathrm{T}^\mathrm{miss}$ is an important contribution to many different physics analyses. The high luminosities at the LHC however mean that there are large amounts of pileup interactions which degrade its reconstruction quality. Therefore there are a number of existing algorithms to reconstruct and estimate $\mathbf{E}_\mathrm{T}^\mathrm{miss}$ depending on the topology of the event. A neural network was trained to combine these different definitions and produced a reconstruction method which lead to a greater $\mathbf{E}_\mathrm{T}^\mathrm{miss}$ resolution in both data and MC, as well as an excellent stability with pileup. While more quantitative analysis still needs to be performed on the resolution improvements, this project reflects the ongoing work to improve reconstruction performance at ATLAS.

7. References

[1] The ATLAS Collaboration 2008 *Journal of Instrumentation* **3** S08003
[2] The ATLAS Collaboration 2019 *Phys. Rev. D* **100**(1) 012006
[3] The ATLAS Collaboration 2018 *Phys. Rev. D* **98**(9) 092002
[4] The ATLAS Collaboration 2012 *Physics Letters B* **716** 1 – 29 ISSN 0370-2693
[5] The ATLAS Collaboration 2010 *The European Physical Journal C* **70** 823–874 ISSN 1434-6052
[6] Nielson M A 2015 *Neural Networks and Deep Learning* (Determination Press)
[7] The ATLAS Collaboration 2018 *The European Physical Journal C* **78** 903 ISSN 1434-6052
[8] The ATLAS Collaboration 2018 $E_\mathrm{T}^\mathrm{miss}$ performance in the ATLAS detector using 2015-2016 LHC p-p collisions Tech. Rep. ATLAS-CONF-2018-023 CERN Geneva URL https://cds.cern.ch/record/2625233
[9] Cacciari M, Salam G P and Soyez G 2008 *Journal of High Energy Physics* **2008** 063–063
[10] The ATLAS Collaboration 2016 *The European Physical Journal C* **76** 581 ISSN 1434-6052
[11] The ATLAS Collaboration 2017 *The European Physical Journal C* **77** 580 ISSN 1434-6052
[12] The ATLAS Collaboration 2017 *The European Physical Journal C* **77** 241 ISSN 1434-6052
[13] 2018 Object-based missing transverse momentum significance in the ATLAS detector Tech. Rep. ATLAS-CONF-2018-038 CERN
[14] Nason P 2004 *Journal of High Energy Physics* **2004** 040–040
[15] Sjstrand T, Mrenna S and Skands P 2006 *Journal of High Energy Physics* **2006** 026–026
[16] Ramachandran P 2017 *CoRR* **abs/1710.05941**
[17] Kingma D P and Ba J 2014 *arXiv preprint arXiv:1412.6980*
[18] Huber P J 1964 *Ann. Math. Statist.* **35** 73–101
[19] 2010 Data-Quality Requirements and Event Cleaning for Jets and Missing Transverse Energy Reconstruction with the ATLAS Detector in Proton-Proton Collisions at a Center-of-Mass Energy of $\sqrt{s} = 7$ TeV Tech. Rep. ATLAS-CONF-2010-038 CERN
[20] Gleisberg T, Hche S, Krauss F, Schnherr M, Schumann S, Siegert F and Winter J 2009 *Journal of High Energy Physics* **2009** 007–007