# Cern Summer Student Report 2017
## "FOAM Approximation of $t\bar{t}$ Phase-Space Distributions for Monte-Carlo Generation"

Graham Van Goffrier
advised by Alexander Held and Prof. Tancredi Carli
ATLAS EP-ADE-CA

Aug 29, 2017

# 1 Introduction

The $t\bar{t}H$ analysis group, as a part of the ATLAS Experiment, seeks to measure the properties of Higgs boson production in association with top-antitop quark pairs. Although this process has a cross-section two orders of magnitude below the leading Higgs decay mode (at LHC center-of-mass energies), direct measurement of the top Yukawa-coupling [1], which can help to further constrain Beyond the Standard Model (BSM) physics, is only possible in this production mode.

The rarity of $t\bar{t}H$ compared to $t\bar{t} + b\bar{b}$ background processes, as well as the irreducibility of these background processes, makes many preferred analysis methods ineffective. In this situation a technique known as the Matrix Element Method (MEM) [3] is applied, which provides a strong discriminant between the $t\bar{t}H$ signal and the background, allowing physics parameters to ultimately be extracted from measurements. The power of this technique comes from its thorough use of all available theoretical information about a process; however, it has heavy computational requirements for the same reason. Very many Monte-Carlo (MC) events must be generated using the matrix-element distributions in phase-space for each parameter set, and for complicated distributions, this event generation is challenging to perform with precision.

## 1.1 What is FOAM, and how does it work?

FOAM is an algorithm designed to generate an approximation of any function, made up of weighted polyhedral "cells" which cover the same domain as the original function. In relation to MC methods, FOAM is able to not only store an approximation of a probability distribution in an efficient manner, but also very efficiently generate MC events in a manner consistent with the true distribution.

The FOAM build-up process begins by initiating a single cell which covers the entire domain; each cell is then recursively split into two daughter cells. The dimension and position of splitting are chosen so as to maximize the decrease of a chosen error variable (as compared with the source distribution) between the parent cell and the sum of the two daughter cells. These are determined by generation of a small quantity of MC events within the cell and subsequent evaluation of the distribution integral over the cell; this process is described in total detail in [2].

Cell generation is stopped when some total quantity of cells (as specified by the user) is attained. The coordinates and weights of all cells may then be stored in a memory-efficient format using the ROOT libraries. The overall FOAM build-up process must sample the given distribution on the order of millions of times, and so has a runtime that grows as the number of cells desired; but this build-up is an investment that pays off, as it need only be performed once in order for any number of MC events to be produced from the approximate distribution.

## 1.2 Project outline

This summer student project generally sought to improve the use of MEM to classify $t\bar{t}H$ events, using established multivariate data methods to supplement MEM. The use of FOAM for this purpose rapidly became the primary focus the project. Although a long-term goal of this work is to successfully generate a FOAM which represents the entire $t\bar{t}H$ process diagram, many small steps of progress are required first. During this summer, FOAMs for three processes have been primarily explored, namely gluon fusion to $t\bar{t}$ where zero, one, or two of the outgoing top quarks are allowed to decay to a bottom quark and a W boson. Promising results have been obtained in the cases of zero and one decays. Perhaps an equally important result of this work is the body of code that has been developed (including expansion of existing code) for

interfacing FOAMs to MC event generators, generating FOAMs for phase-space distributions with different parameterizations or transformations of variables, and evaluating the performance of FOAMs. Graphical tools for visualizing FOAM performance have been an especially helpful outcome of this project, and have been designed in a general enough manner so as to be applicable to higher-dimensional FOAMs in the future. Specifically, they will play an important role in identifying and correcting the difficulties that currently exist in applying FOAM to the gluon to $t\bar{t}$ process with both top quarks allowed to decay.

# 2 Challenge: FOAM complexity-dependence

Theoretical considerations of the FOAM algorithm state that in the limit of infinitely many generated cells, the FOAM distribution should converge perfectly to its reference distribution. A key aspect of improving the usefulness of FOAM is to minimize the number of cells which are necessary to achieve agreement with the reference within a given error bound. This required quantity of cells is known to increase sharply as the number of dimensions and peak steepness of a desired distribution are increased; a more analytical solution for this quantity was beyond the scope of this project.

An existing FOAM interface setup for the gluon to $t\bar{t}$ process, with no subsequent decays, was employed to confirm functionality of all relevant software components. At this stage, experience was gained with the histogram and plotting capabilities of ROOT, producing images such as Figure 1 to evaluate FOAM performance.
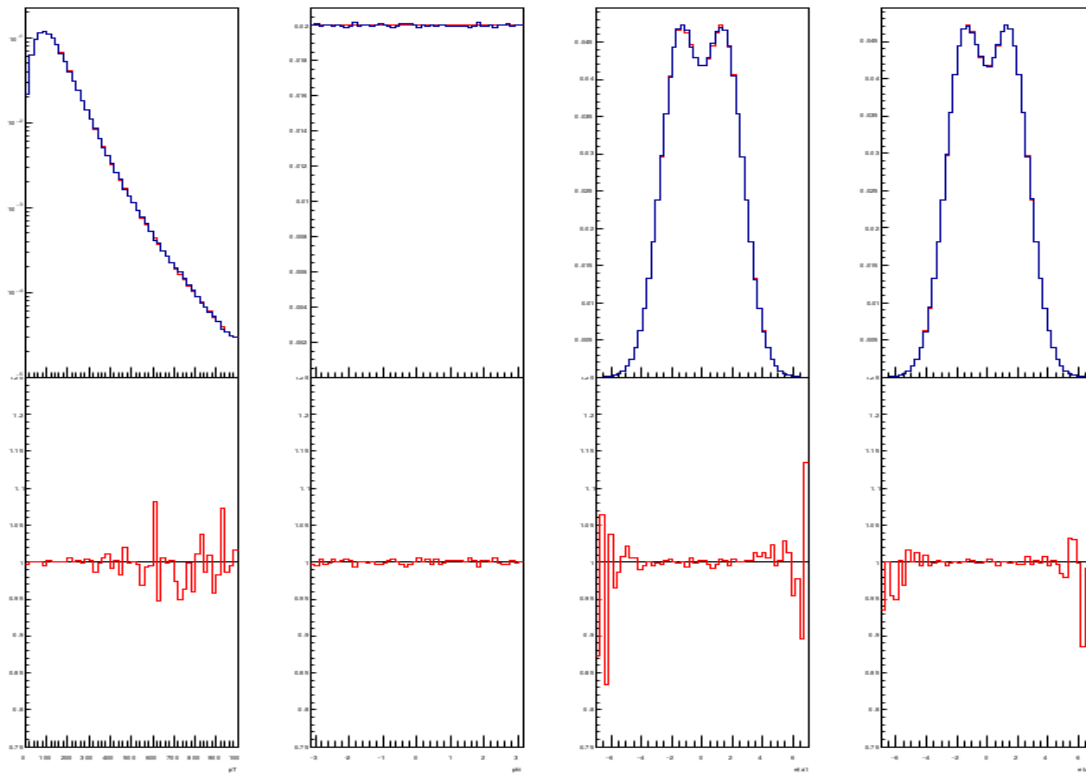


Figure 1: g g > t $\bar{t}$: FOAM vs. reference in angular coordinates

The four physical variables plotted in the above figure ($p_T$, $\phi$, $\eta_t$, and $\eta_{\bar{t}}$) correspond to the four degrees of freedom of the process after momentum conservation constraints have been applied. The plots in the top sector show the FOAM cells projected into histograms in each physical dimension, superimposed on reference histograms generated with MadGraph. The red plots in the bottom sector give the ratio between these two histograms, and show overall good agreement.

Modified interface software allowed for the building of FOAMS for $t\bar{t}$ processes with one and two top decays, corresponding to seven- and ten-dimensional systems respectively. Even for large quantities of cells, these FOAMs showed little agreement with reference distributions in most physical variables. This situation was a clear example of the impact of high-dimensional, strongly-peaked distributions on FOAM performance, and a motivator for the remainder of summer work.

## 3  Progress: g g > (t > b w+) $\bar{t}$

The $t\bar{t}$ process with one decay included was the lowest-dimensional FOAM that displayed significant inaccuracy. Trial-and-error of different parameterizations and single-variable transformations played an important role in improving this accuracy. One important factor in the selection of a phase-space parameterization was mathematical convenience, as the Jacobian of the variable change would need to be calculated for each generated MC event. Simpler algebraic forms in this expression would simplify debugging of relevant code, and also decrease computation time.

Ultimately, a particular angular parameterization was chosen for the FOAM that gave strong agreement: this included all three angular coordinates ($p_T$, $\phi$, and $\eta$) for the $b$ and $W^+$ particles, and only the pseudorapidity $\eta$ for the $\bar{t}$ quark. Although the many single-variable transformations studied did not prove to be useful for this particular use case, it is expected that they will become more necessary in even higher-dimensional processes. Figure 2 shows one of the best-performing dimensions of this FOAM, the pseudorapidity of the b quark, with the distributions plotted as normalized to unity.
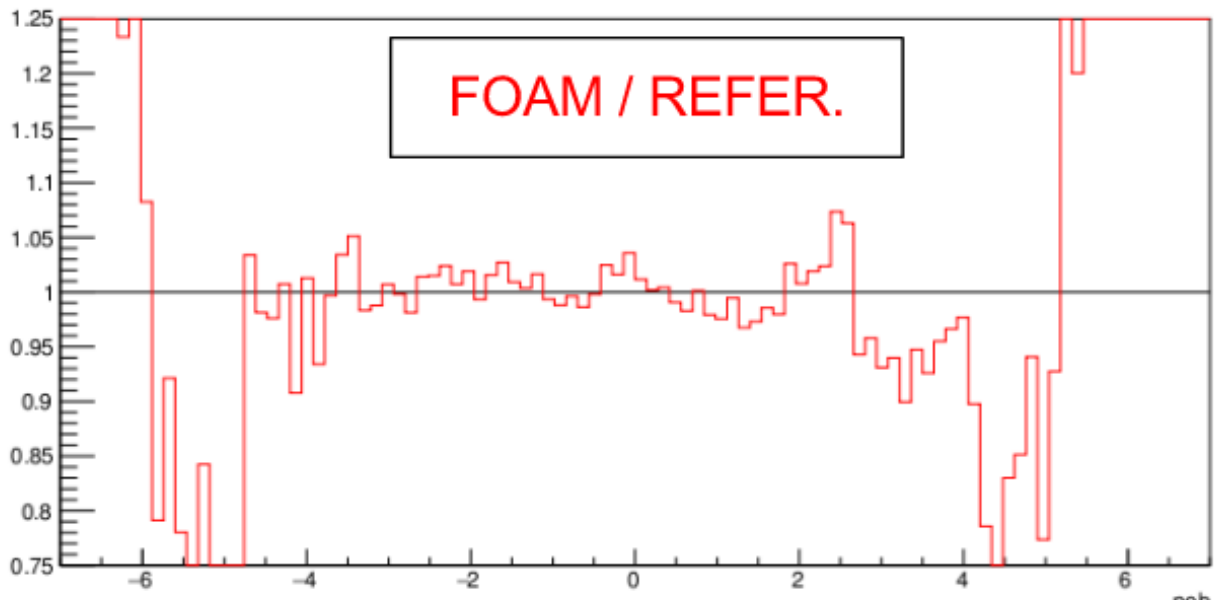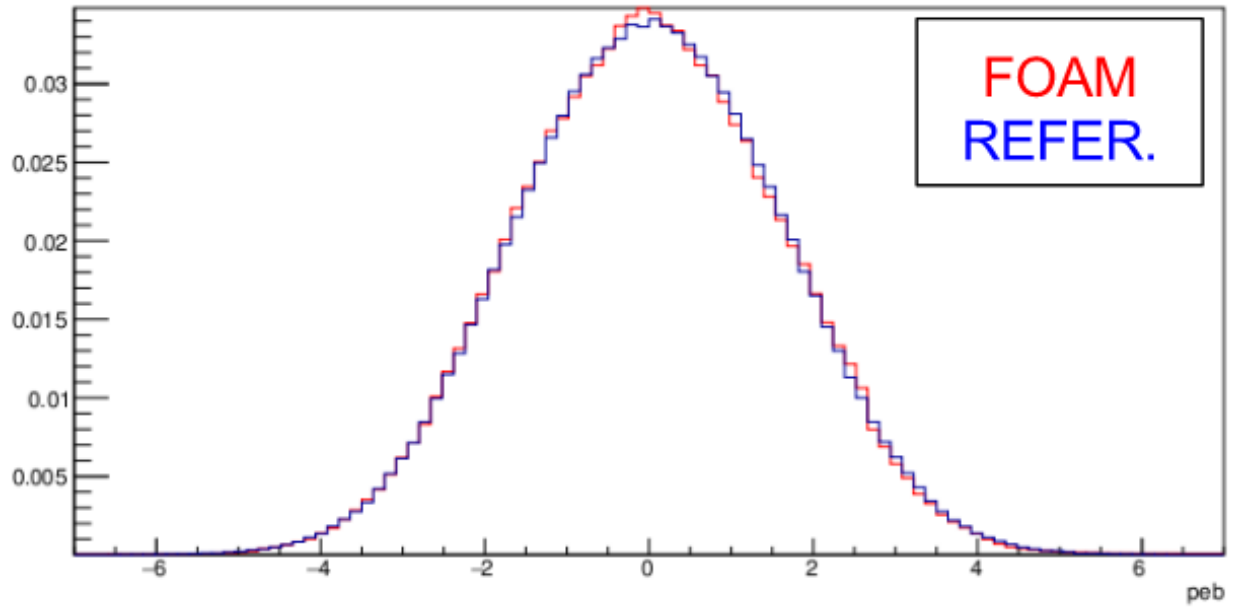
Figure 2: g g > $t\bar{t}$ with one decay: $\eta_b$ comparison

Although agreement was good overall, some lingering asymmetries appeared in the transverse momentum ($p_T$) of the b quark and the pseudorapidity ($\eta$) of the antitop quark, as shown in Figure 3.



(a) $p_{T_b}$ distribution


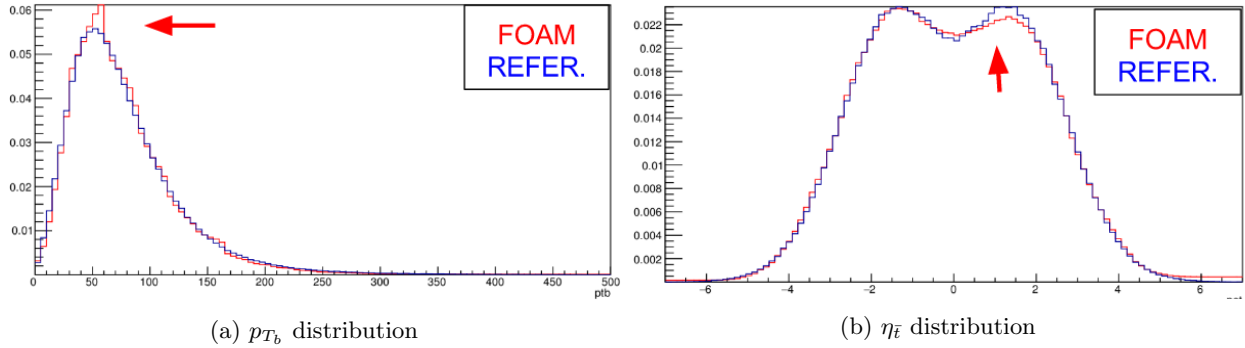
(b) $\eta_{\bar{t}}$ distribution

Figure 3: Asymmetries in gluon to $t\bar{t}$ single-decay process

The task of uncovering the cause of these disagreements inspired several key visualization tools which are likely to play a role in future work. One example is a cell-split mapping tool, which recursively searches the binary tree formed by cells in a FOAM and sorts cell-split locations by dimension into separate histograms. This analysis is possible because the FOAM generation algorithm only ever splits cells in one dimension at a time, leaving cell coordinates untouched in all other orthogonal directions. The resultant plots (e.g. in Figure 4a) aid in the identification of regions where the FOAM fails to devote sufficient splitting resources, either by some fault in the algorithm or by some flaw in the parameterization that might be corrected. A function for calculating the top mass only from the physical parameters stored in the FOAM was also implemented; the resultant plot for this process, shown below in Figure 4b, peaked appropriately near the value of 172.5 GeV which was used in the FOAM generation, and so gave some confirmation that major physics/calculation errors were not the source of the asymmetries. This distribution is spread out due to the non-infinitesimal size of all cells in the FOAM, and with more cells the shape should approach a Breit-Wigner function.
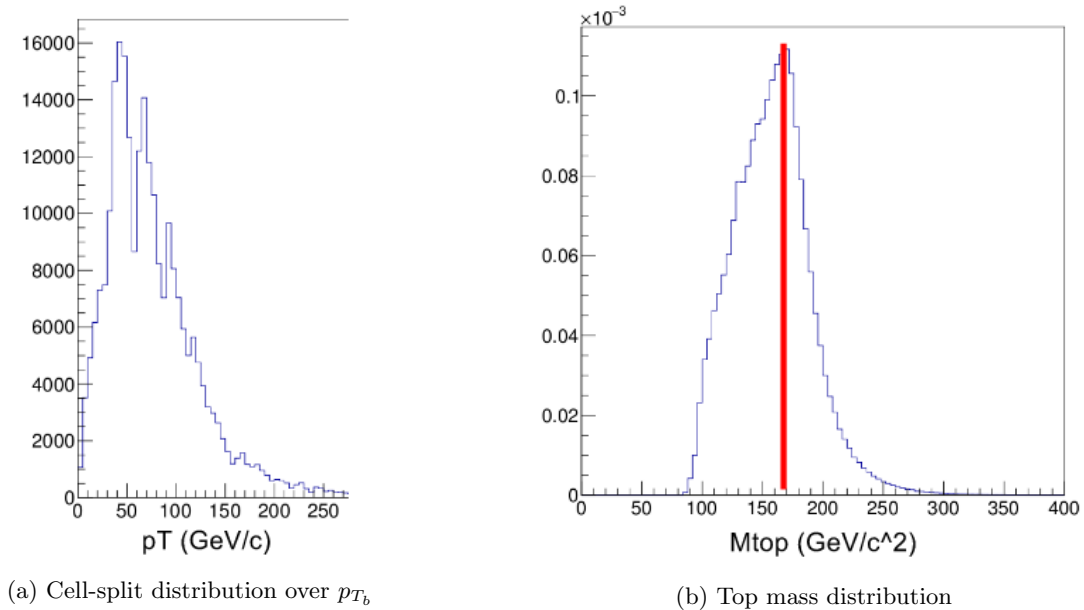


(a) Cell-split distribution over $p_{T_b}$



(b) Top mass distribution

Figure 4: Visualization tools for FOAMs

5

Most recently, two-dimensional projections of the FOAM were considered as a means of exploring hidden features of the distribution. Although one-dimensional projections such as those shown in Figure 2 and 3 contain substantial information, they cannot give a total description of the multidimensional distribution; the only exception is a distribution which is exactly factorizable. Two-dimensional projections are capable of displaying richer detail about the FOAM and the distribution, while still remaining intuitively readable to humans. Two such examples are shown in Figure 5.



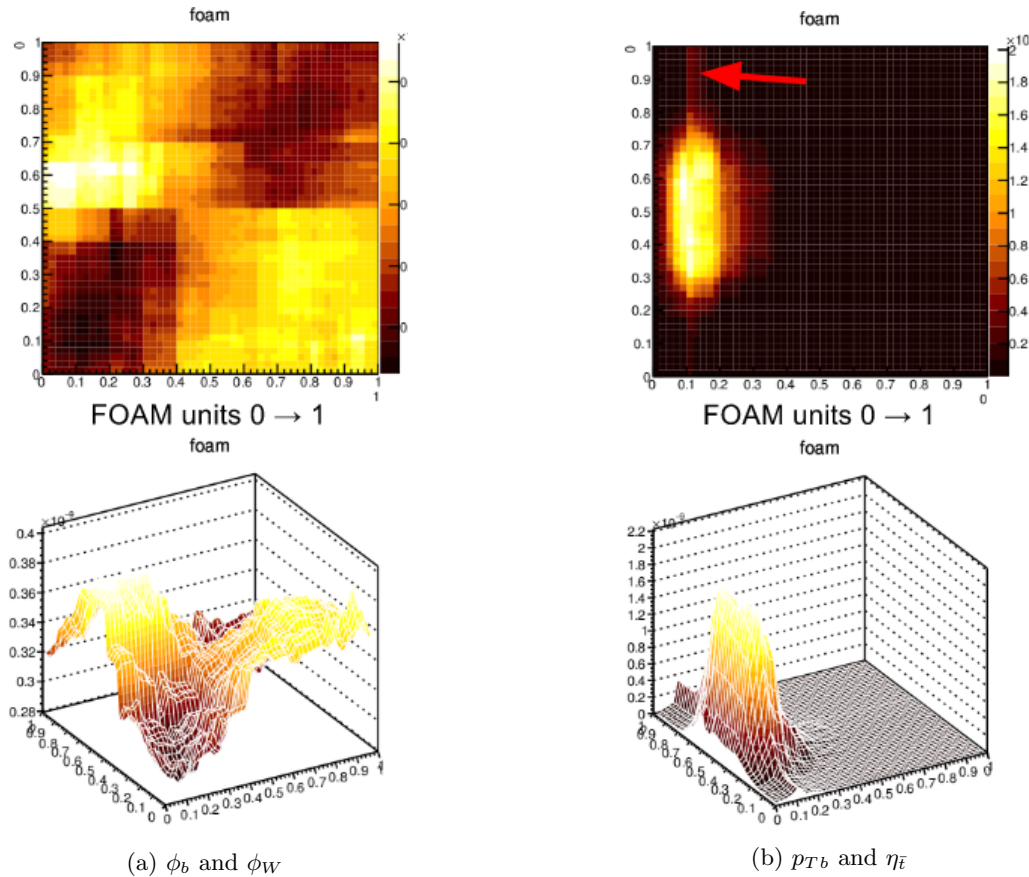(a) $\phi_b$ and $\phi_W$

(b) $p_{Tb}$ and $\eta_{\bar{t}}$

Figure 5: Two-dimensional projection histograms

Figure 5a is given as a pedagogical example of this functionality; the top and bottom plots are different camera views of the same histogram, for convenience, where brighter shading of a cell corresponds to a larger weight value. The single-dimensional projections of phi distributions for the b and W particles are both individually quite flat; however, when plotted together, a clear anticorrelation appears which matches with physical intuition about the opposing directions in which decay products from the same source should generally travel.

Figure 5b shows a more meaningful application of this tool, combining the two error-ridden variables given in Figures 4a and 4b into a single visual representation. The essential feature is the very large volume, mid-weight cell marked here by an arrow; by projecting into the two component dimensions, the cause of both physical asymmetries was revealed as this single FOAM feature. How to best treat this issue remains an open question, though some solutions involving stricter cuts have been proposed. It seems that additional splitting of the largest-volume cells might correct the observed disagreement, but gaining understanding of why these cells were not automatically split during FOAM generation remains an important goal for

generally improving accuracy. Perhaps the most important result from this example is reinforcement of the effectiveness of the visualization tools developed, giving hope that they will maintain their effectiveness to higher dimensions.

# 4    Ongoing work

The next goal of this continuing project will be to extend the concepts applied to the single-decay $t\bar{t}$ process to the double-decay version. In principle, the two share enough physical properties that this extension should not be unreasonable to achieve; in the event of unforeseen complications, the visualization tools developed (and perhaps new ones) will be applied to investigate each problem.

Another important short-term goal will be the collection of all implemented functions and classes (in both Python and C++) into a single bash-based package for easier use. Any researcher, experienced with MC methods but perhaps not specifically with FOAM, should be able to make use of these tools to experiment with FOAM and adapt it to any desired application.

# 5    References

[1] Bezrukov, F., and M. Shaposhnikov. \Why Should We Care about the Top Quark Yukawa Coupling?" Journal of Experimental and Theoretical Physics, vol. 120, no. 3, 2015, pp. 335{343., doi:10.1134/s1063776115030152.

[2] Jadach, S. \Foam:  A General-Purpose Cellular Monte Carlo Event Generator", 2002. Comput.Phys.Commun. 152 (2003) 55-100. Web, arXiv:physics/0203033.

[3] Nackenhorst, Olaf. "Search for the Standard Model Higgs boson produced in association with tt and decaying into bb at 8 TeV with the ATLAS detector using the Matrix Element Method", 2015. CERN-THESIS-2015-186. Web, cds.cern.ch.