The Island Algorithm - A Photon Clusterizer at sPHENIX

Brandon McKinzie, Bi Ran, and Gunther Roland

Laboratory for Nuclear Science, Massachusetts Institute of Technology

Abstract

The sPHENIX detector is a proposed major upgrade to the PHENIX experiment at the Relativistic Heavy Ion Collider at Brookhaven National Laboratory [1]. The sPHENIX detector will have an electromagnetic calorimeter (EMCal) to detect the energies of charged particles such as electrons, positrons, and photons. In order to help identify which region of towers in the EMCal absorbed the energy deposited by a given particle, this research implemented and analyzed a clustering algorithm known as the island algorithm. Its performance can be analyzed by running collision simulations and comparing the algorithm's outputs with the known energies of the simulated particles. For simulations of single-particle events, the results indicate close agreement between the known generated energies and the reconstructed energies from the island algorithm.

Introduction

The collisions at sPHENIX involve protons and/or heavy nuclei traveling near the speed of light. When particles collide at these high energies, hundreds or even thousands of particles can be produced. The electromagnetic calorimeter (EMCal) is one of many specialized sub-detectors required to accurately reconstruct the event information. The EMCal is responsible for collecting the energy and position information of electrically charged particles that pass through the detector. The EMCal can be seen in the two figures below as the innermost dark gray ring.

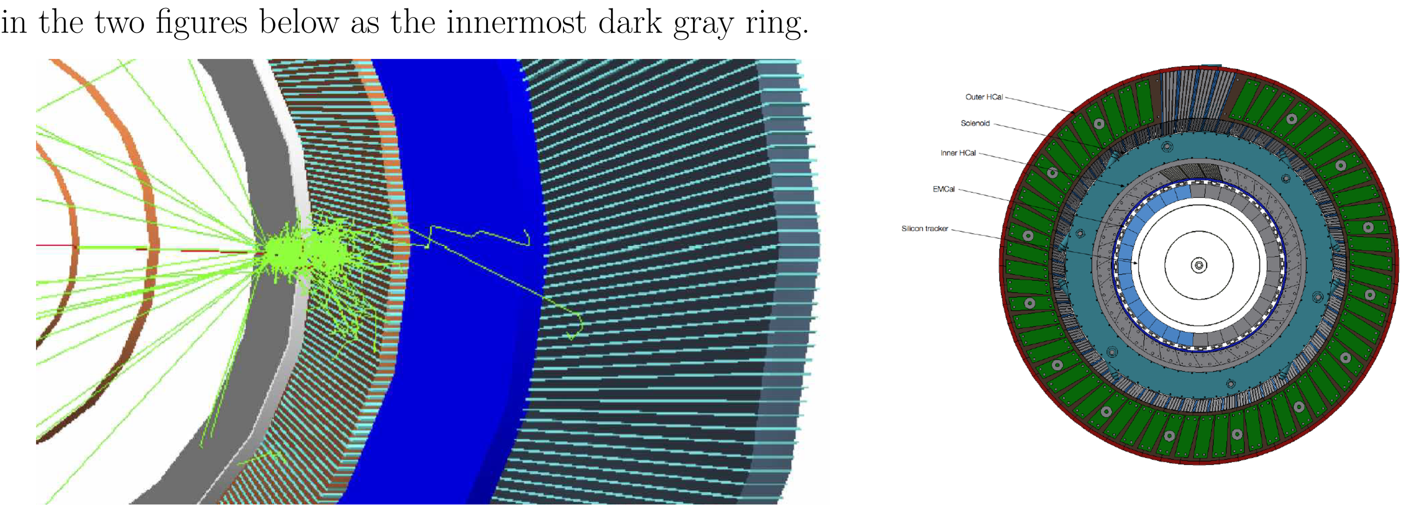

Figure 1: Left: Transverse view of a 10 GeV/c electron in sPHENIX, showing it showering mainly in the EMCal. The particle's energy is spread out over a group of towers. The purpose of a clusterizer is to determine which group of towers correspond to a given single particle. Right: Cross section of the sPHENIX detector indicating the locations of the main sub-detectors. Image source (both): ref. [1].

This research was focused on implementing an algorithm that, given raw data from the EMCal, could translate that data into the energies of the particles produced in the collision. What follows is an outline of the main steps in the algorithm, visualizations of example clusters built by the algorithm, and an analysis of how well the algorithm reconstructed the energies of certain particles in comparison with a more primitive clusterizer implementation.

The Island Algorithm

1. Begin by preparing a list of seeds, which are defined as towers with a transverse energy above a certain threshold.
2. Order the seed list in decreasing energy, then remove seeds that are adjacent to higher energy ones.
3. Starting from the most energetic seed, conditionally assign surrounding towers to its cluster:
   1. Search both directions in φ, collecting towers until a rise in energy or a hole is encountered.
   2. Move one step forward in η. If the tower energy is larger than the previous, or if it is a hole, discontinue η steps in this direction and move to next step. Otherwise, apply logic in step 3.1 again for this position in η.
   3. When forward η steps are complete, do the same in the opposite η direction.

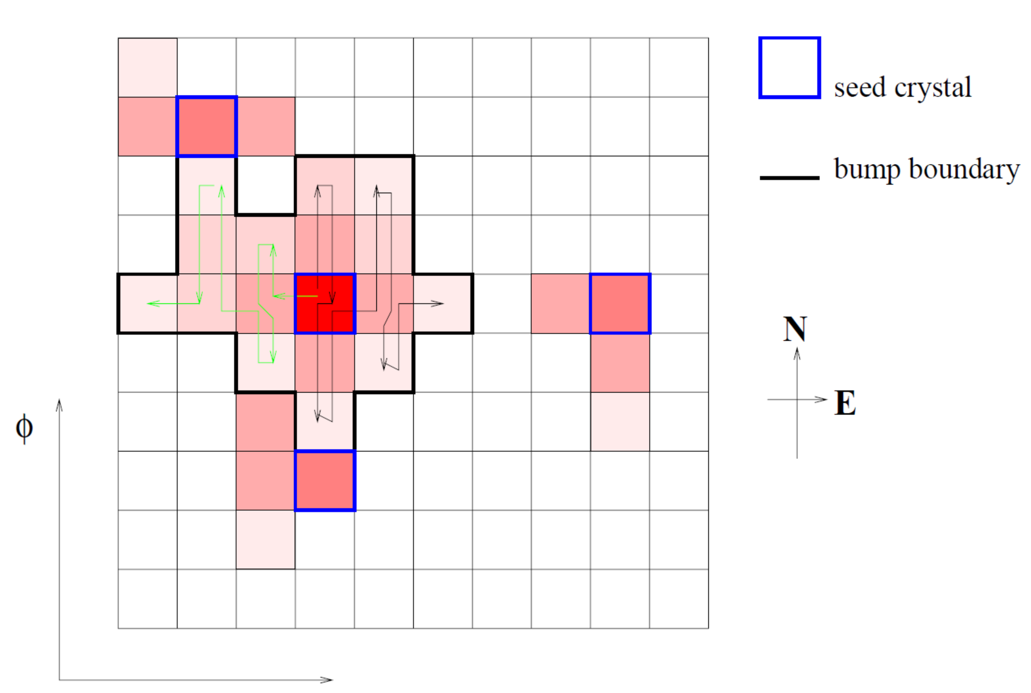

Figure 2: Illustration of the island algorithm. Given the seed tower near the center (dark red square), apply the steps enumerated above to collect the surrounding towers that will define the associated island cluster. Image source: ref. [2].

Cluster Visualizations

Shown below are visual representations of clusters that were built both by the island algorithm (upper) as well as a simple 5x5 clusterizer (lower). As its name implies, the 5x5 clusterizer simply, for each seed tower, defines the associated cluster as all towers (with non-zero energy) in the 5x5 region centered on the seed. For both clustering algorithms, three example clusters are plotted, one for each simulated set of (left to right) electron, photon, and pion events. Visually inspecting these output clusters helps one develop a stronger intuition and understanding of each algorithm's behavior.

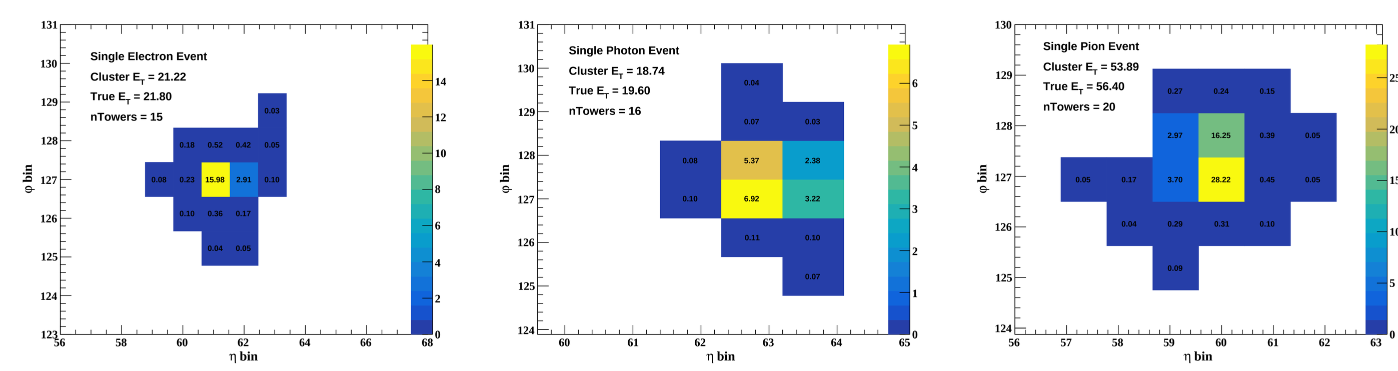

Figure 3: Examples of clusters built with the **island algorithm**. From left to right, the clusters correspond to an electron, photon, and neutral pion. Notice how electrons appear to deposit nearly all their energy in a single tower, while pions, which are known to decay into two photons, are seen to have the bulk of their energy spread over two or more towers. These visualizations show how the algorithm got its name; the bright yellow squares, the seed towers, appear as an island surrounded by the residual (lower-energy) towers.

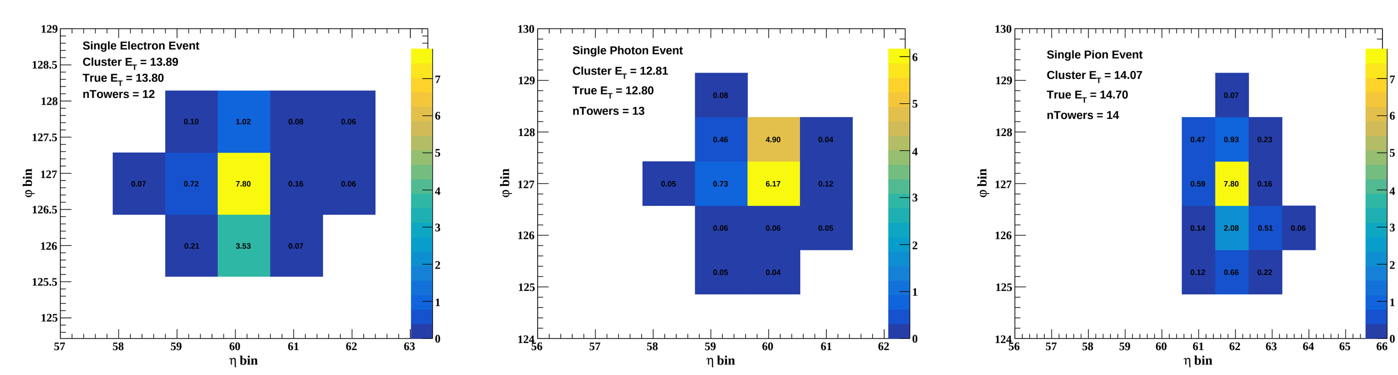

Figure 4: Examples of clusters built with the **simple 5x5 clusterizer**. From left to right, the clusters correspond to an electron, photon, and neutral pion. As the amount of background processes increase, this method will become more prone to overestimating a given particle's energy, since it will collect all towers in the 5x5 region. It will also exhibit cluster shapes that would be impossible for the island algorithm, such as the rightmost (pion) cluster. Since the island algorithm first searches left/right from the seed to determine when to stop, it would never encounter the tower in the lower-right region with $E_T = 0.06 GeV$.

Applying Cuts

Like all clustering algorithms, the island algorithm has its limitations. A common remedy for this is to apply "cuts," which are simply conditions required for data that determine whether or not it should be rejected. For example, since the island algorithm begins by assigning towers with energy above a predefined threshold, it can erroneously cluster towers that contain nothing but noise (energy from unwanted background).

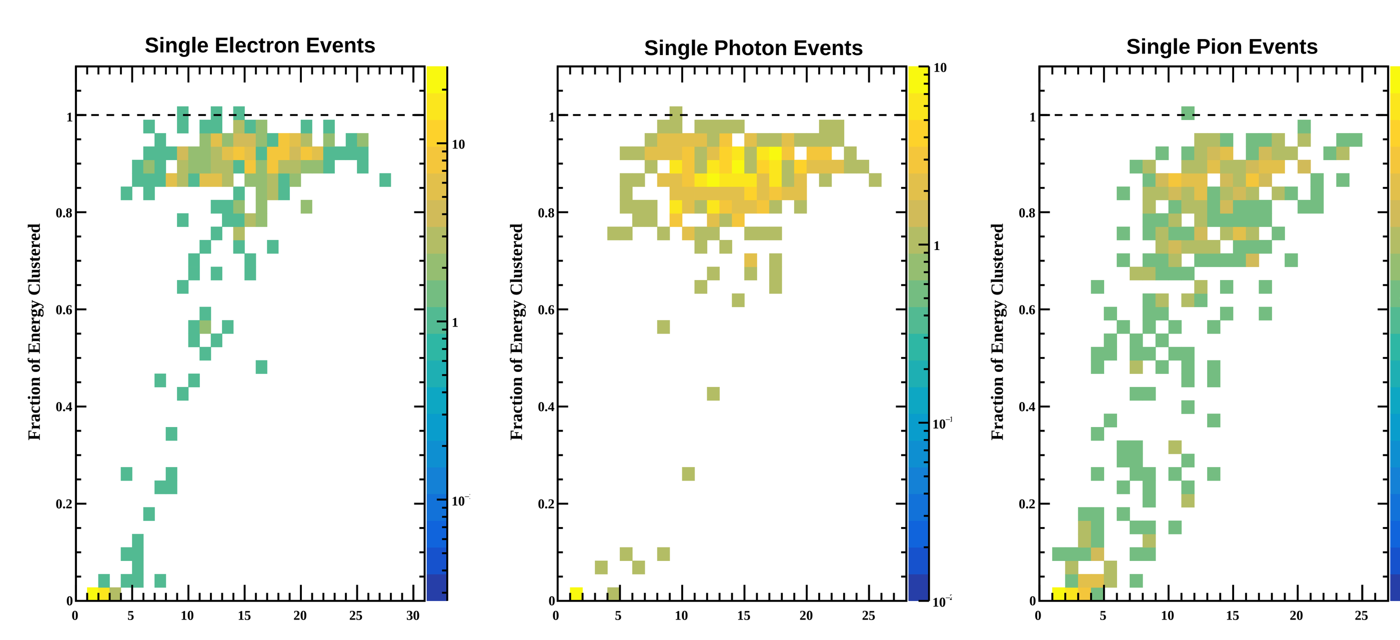

Figure 5: The fraction of the single-particle energy reconstructed in clusters built by the island algorithm, shown as a function of the number of towers in the cluster. The simulated particles shown are (from left to right) electrons, photons, and neutral pions. The majority of clusters that severely underestimate the particle's energy tend to contain less EMCal towers. To improve performance, then, we apply a cut that rejects any clusters containing fewer than six towers.

Results

Since the goal of both the island algorithm (upper row) and the simple 5x5 clusterizer (lower row) is to accurately reconstruct the energies of particles that strike the detector, one straightforward evaluation method is to plot the reconstructed (cluster) energies against the known particle energies. On all plots, the dotted line of unit slope indicates where an ideal clusterizer would be, corresponding to an exact match between all true and reconstructed energies. For reasons explored in the previous section, a cut is applied requiring all clusters to contain at least six towers.

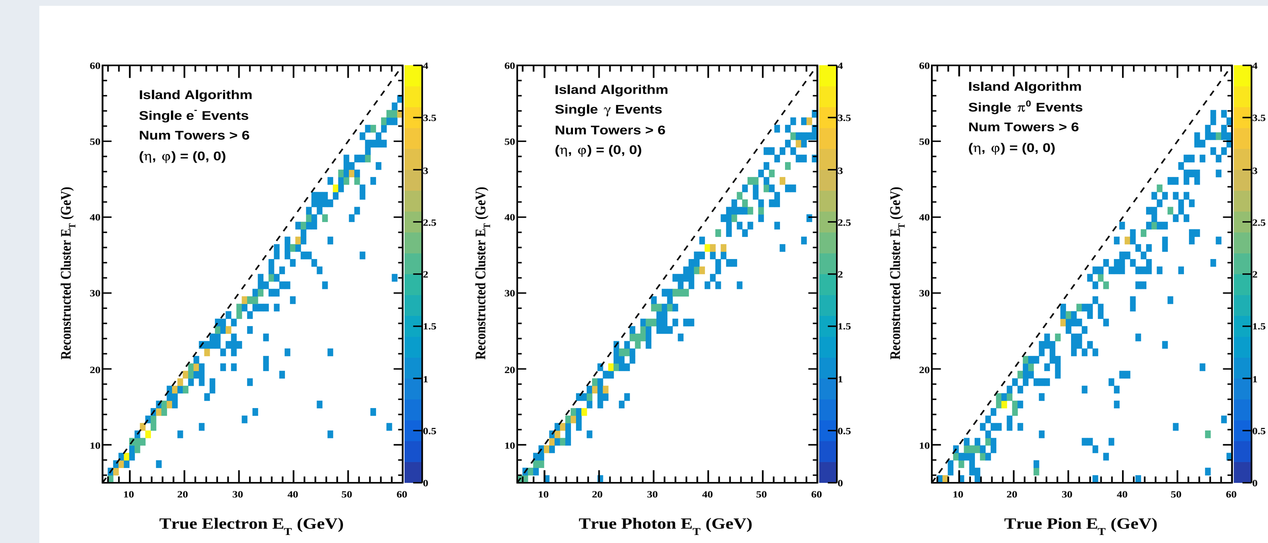

Figure 6: Reconstructed particle energies determined by the **island algorithm**, shown as a function of the known particle energies. As anticipated, electron and photon events are closely aligned with the ideal dotted line, since these particles deposit most of their energy in a small region of towers. Neutral pions, however, decay into two photons before reaching the EMCal. If the splitting angle between the two photons is large enough the island algorithm may cluster the two photons separately, resulting in two clusters with half the energy.

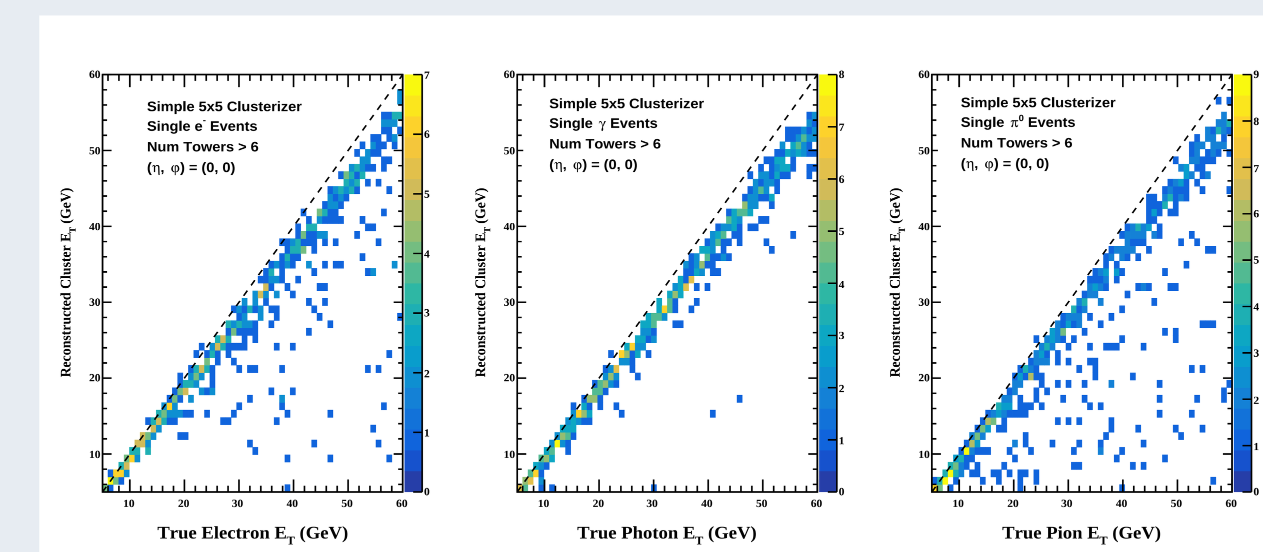

Figure 7: Reconstructed particle energies determined by the **simple 5x5 clusterizer**, shown as a function of the known particle energies. Although this clustering method conforms reasonably well to the dotted line for these single-particle events, it can be seen to overestimate the particle energy far more often than exhibited by the island algorithm. This is due to the 5x5 clusterizer blindly accepting all towers in a fixed region around the seed, and the number of overestimated energies will increase as the simulated events become more complex and with increasing background.

Conclusions

The island algorithm is shown to be a reliable clusterizer for single electron, photon, and neutral pion events. The number of false positives, regions of noise that were erroneously clustered, can be reduced by cutting on, for example, the minimum number of towers per cluster. A more comprehensive analysis, one that tests the algorithm using PYTHIA and HIJING samples, is required before this implementation can be fully integrated within the sPHENIX software framework. However, this research project succeeded as a proof-of-concept that the island algorithm, ported from the CMS collaboration, could be a useful contribution to the sPHENIX software framework.

Acknowledgments

I'd like to thank all members of the MIT Heavy Ion Group, in particular Gunther Roland, Bi Ran, Alex Barbieri, and George Stephans for always being available when I had questions. In addition, this project would not have been possible without the tremendous help from sPHENIX collaboration members Jin Huang and Dennis Perepelitsa. Finally, thank you to all members of the MSRP staff for putting this program together and mentoring the students on how to best prepare for a graduate education.

References

[1] A. Adare et al., "An Upgrade Proposal from the PHENIX Collaboration," (2015). arXiv:1501.06197 [nucl-ex]

[2] E. Meschi et al., "Electron Reconstruction in the CMS Electromagnetic Calorimeter," (2001). CMS-NOTE-2001-034