Note: blue font means that this is text and figures which will be replaced/added to the paper

Reviewer #2

1) What is the typical energy, spatial and timing distributions of the gamma rays emitting from the target? If this question is too general, maybe you can use the case of the most dominant type of events encountered at R3B, or the most interesting one.

The energy goes up to 10 MeV- really depending on the excited states which get occupied in the reaction process. However when forward boosted energies up to 20 and more MeV are possible. There is no typical energy, it is experiment dependent (qfs,fission,coulomb excitation). The forward boost determines also the spatial distribution in forward direction.

For example in the commissioning experiment S444 in 2020 we had a ^{12}C beam at 400/550/650 and 800 AMeV impinging on a carbon (or plastic) target. In the qfs reaction, where a proton from the 12C projectile gets knocked out by a proton from the target and you end up with a 11B fragment. This can be in the ground state or excited. If excited it immediately de-excites and emits a (boosted) gamma (2.1MeV in the 11B frame) which we then detect in CALIFA.

2) What does the 4µs window correspond to, in term of the experiment? Is it around a beam signal, or a trigger signal?

At R3B we usually set our event window at 4us for time stitching in the data stream. This window is opened by the first signal from one of the detectors in the whole R3B setup, which is mostly the fast signal from the Start detector at the entrance of our experimental hall Cave C. All signals within 4us are then grouped together an form one physics event.
As mentioned in the paper for this cluster reconstruction analysis the time within the stitched event is normalized between 0 and 1.

3) The input gamma rays are sampled from a uniform energy distribution between 0.3 MeV and 10 MeV, what is the angular distribution sampled, in particular the polar angle, also uniform?

Both polar as azimuthal angular distributions are uniformly distributed. This is the experimental case for calibration runs with a radioactive source (e.g. 22Na, 60Co).
In experiments with relativistic beams the gammas are boosted in forward direction (no uniform distribution in polar angle); the azimuthal angular distribution remains also in this case uniformly distributed.
Here the distribution of the input features (energy, theta, phi, normalized time).

This picture with description will be added in the “supplementary material” part of the paper:

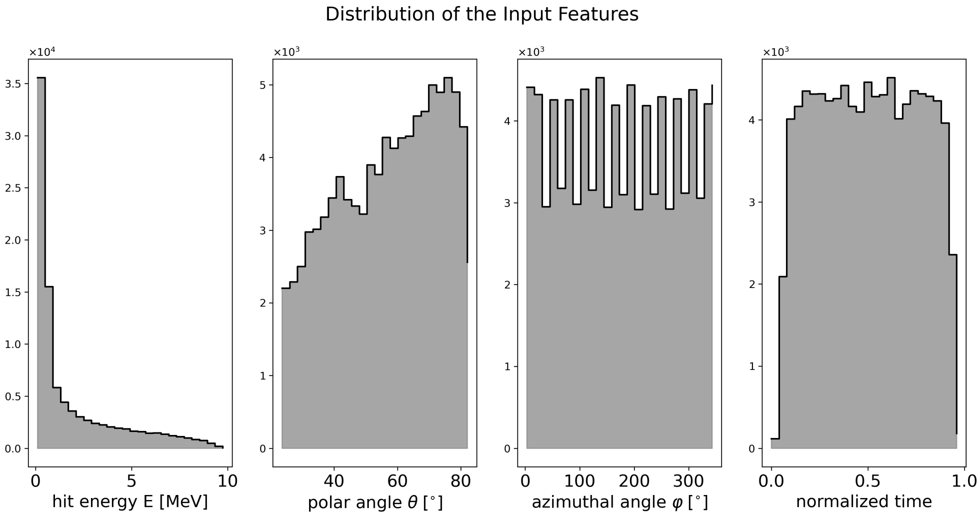


Fig. X.: Feature distributions of the simulated hits in CALIFA obtained with Geant4 from incident gammas. Although the primary gammas were generated with uniform distributions in energy, polar and azimuthal angle, as well as normalized time, the reconstructed hit features show detector- and physics-driven effects. The hit energy spectrum is dominated by low-energy deposits, reflecting the increased probability of pair production and Compton scattering at higher gamma energies, while the contribution of the photoelectric effect decreases (cf. Fig. 1). The polar-angle distribution exhibits a reduced response at small angles. This effect arises from two sources: (i) histogram binning artifacts, since the polar angles of the crystals are not uniformly distributed owing to variations in crystal geometry, and (ii) detector-specific effects, as the CALIFA forward region employs shorter crystals (15 cm) while the backward region uses longer crystals (20 cm) with slightly different shapes.
The azimuthal and normalized time distributions are approximately flat, as expected from the isotropic emission and uniform generation in time.

4) Does the MC simulation include the detector response and electronics readout effect?

The MC simulation includes the detector response thanks to the Geant4 library which simulates the interaction of the emitted gammas with the detector material CsI(Tl).
The electronics readout effect/response is not simulated with Geant4. By smearing out the times of the hits belonging to one cluster (true information) with a Gaussian with sigma of 200us I try to roughly include the electronic response with respect to the time.

5) Can you show the distributions of 12 inputs for the Edge model training? Is it possible to compare these variables between MC simulation and real data with some kind of calibration runs?
The four input feature distributions, as shown in answer 3) will be added to the text.
In principle it is possible to compare these variables between MC simulation and real data - more precisely with calibration runs (since for experiments the gammas are boosted in forward direction and therefore not uniformly distributed for the polar angle).

6) Can you compare the output of Edge model output between MC and real data from

Calibration runs?

Yes, this should be in principle possible. However I do not have any data available where the position of the source is same as for the simulation. And as we do usually have only sources such as 22Na/60Co with peaks at 1.275 and 1.3 MeV accordingly, we do not expect to see much of the complex hit pattern due to the low probability of pair production at gamma energies ~ 1MeV, see also Fig. 1 in text.

7) Figure 2: why do you select the case of 3 mono-energetic gammas at 2.1 MeV? If it is to represent a mixture of signal and background, shouldn't the gamma rays be of different energy?

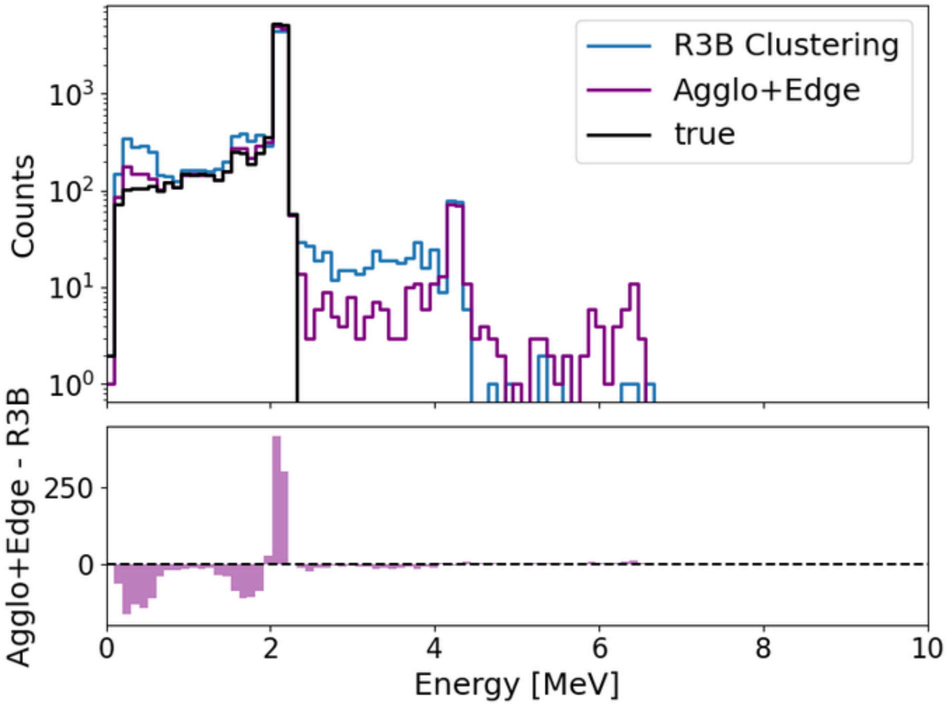
You are correct that from a pure physics standpoint, it is unexpected to have exactly three gammas with same energy for each event. This was a rudimentary observation based on specific experimental campaigns, such as (p2p) experiments where we observed around 10 low-energy hits, which should be assigned to 1-2 background clusters and around one cluster from particle de-excitation in the reaction.

For other physics cases, like fission, we might expect to see more clusters in CALIFA, while others may produce fewer.

The main goal of our algorithm is not to focus on the exact number of gammas or on their energies. Instead, we want the algorithm to learn the correlation between hits so that it can accurately group hits together. The choice to use a fixed number of gammas per event with same energy was an initial step and makes also easier to interpret reconstructed cluster energy distributions, as in Fig2. We can see there the single escape peak at ~1.6MeV and the corresponding 511keV peak. Using uniformly distributed gamma energies would smear out the distributions and would it make also more difficult to see just by looking at the plot, if the applied model improves reconstruction.

8) Figure 2: can you add the truth distribution: i.e. where the fake rates are 0 and TP=1?

This is now implemented in Fig. 2.



9) Figure 2: It would also be interesting to have the same plot but comparing cluster energy, rather than event energy, between different models. I guess this will show up as peaks centered at 2.1 MeV with different height (efficiency) and width (resolution).

Figure 2 actually shows the cluster energy, not the event energy. As you can see there I compare this cluster energy for the different models.
I rewrote the caption of the figure to make this clear:
Reconstructed gamma cluster energy spectrum from simulated events, each consisting of three 2.1 MeV gamma photons emitted from the target point. The upper histogram shows the reconstructed cluster energy distribution using the geometrical R3B clustering (blue) and the Agglo+Edge(pink) method accordingly. In black the true energy cluster distribution. The lower panel displays the bin-by-bin count difference between the two approaches. The Agglo+Edge model demonstrates a significant improvement by successfully reattaching escaped hits, notably in cases where sparse energy deposits around 1.6 MeV and 0.5 MeV result from pair production and subsequent annihilation processes of the original gamma photons. This clean-up step leads to a marked reduction in false negatives (i.e. reduction of bin counts at 0.5 and 1.6 MeV) compared to the geometrical R3B clustering and an enhancement of 2.1 MeV peak.

10) What is the significance of various fake rates and WR with respect to physics measurements of experiment? This study focuses only on the correct association of hits to clusters. Is this the most important requirement of the calorimeter reconstruction? For example, I can imagine in some applications where noise hits are non-negatable missing a few low energy hits may be less important than a precise localization or energy measurement of the cluster. Is there other "Performance Metrics" worth exploring for the CALIFA calorimeter reconstruction?

As you mention, when handling with high energy hits, e.g. protons, missing few low energy hits

would not be that important.
However in this analyis we focus on gamma reconstruction. A reduction in the energy reconstruction is not negligible, even for low energy hits missing. Especially when considering that for experimental runs the Doppler correction makes the energy reconstruction even worse, when missing hits.
As you mention, a precise localization – the incident (polar) angle of the gamma – is of huge importance, since gammas cannot be tracked by any pixel tracking detector. And again – misreconstruction in the polar angle have non negligible effects when applying Doppler correction. Which means, that for gammas we have to focus on both precise localization and energy measurement.
The topic of precise localization is out of the scope of this work. However this may should be pointed out in the outlook and should be further investigated.

11) The numbers in table 1 are measured with the uniform energy distribution between 0.3 MeV and 10 MeV. Is there any energy dependence of these numbers?

This picture with description will be added in the “supplementary material” part of the paper:

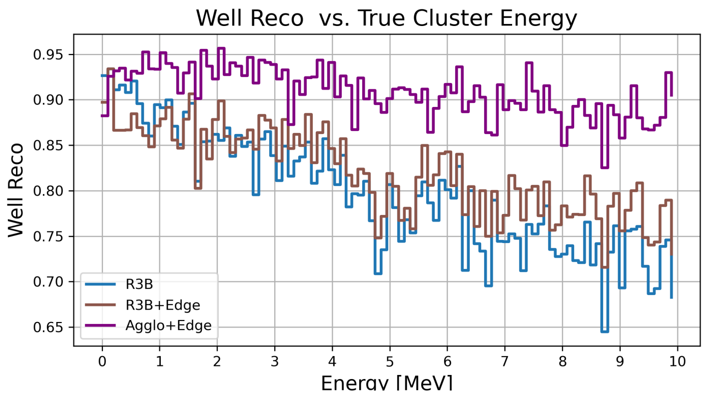


Fig. X: Ratio of well-reconstructed clusters as a function of the true cluster energy. For the geometrical R3B clustering algorithm, the well-reconstruction ratio decreases with increasing cluster energy. The R3B+Edge approach partially compensates this degradation, while the Agglo+Edge method maintains a consistently high well-reconstruction ratio over the full energy range.

12) Can you discuss in the "outlook" section how this model will be integrated into the R3B data analysis to produce improved measurements, and cite some examples where the improvements will bring significant impact on the science output of R3B?

As mentioned in the “outlook” section, the current model is computational expensive and not suited for online analysis.
For offline analysis this model could be implemented as the time complexity does not play a critical role.
It has to first be analyzed if the model can be shrunk so that it can be used for online analysis or if it should only be used for offline analysis.

TODO: maybe add more in my outlook part about this...

13) Publication details are missing for Reference [5].

Done, thanks for the hint!

14) It would be useful to define "A GeV".

I decided to use GeV per nucleon instead of defining AGeV.