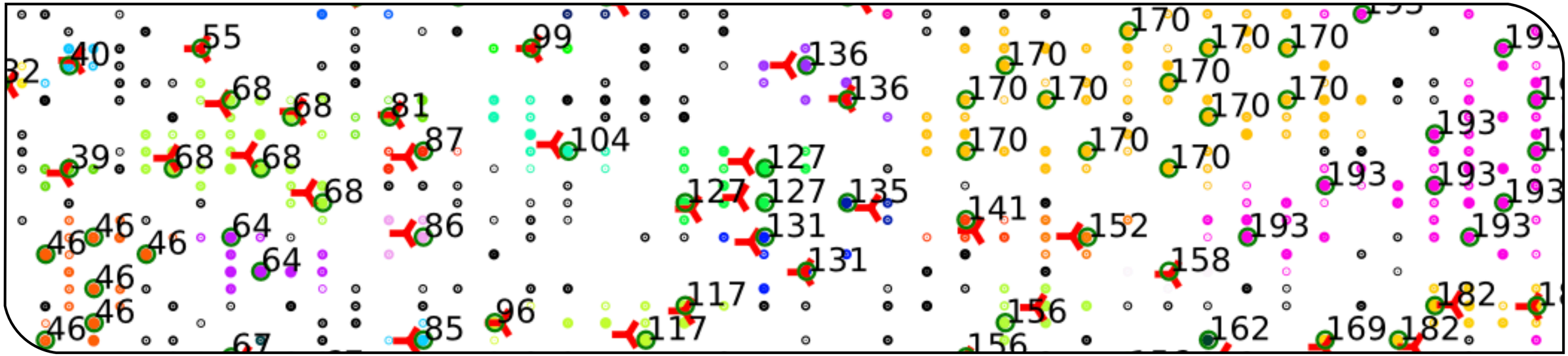
Karlsruhe Institute of Technology

ECL Clustering at very high beam backgrounds

FSP Belle II Germany, 01.10.2024

Torben FERBER (torben.ferber@kit.edu), Isabel HAIDE
Institute of Experimental Particle Physics (ETP)



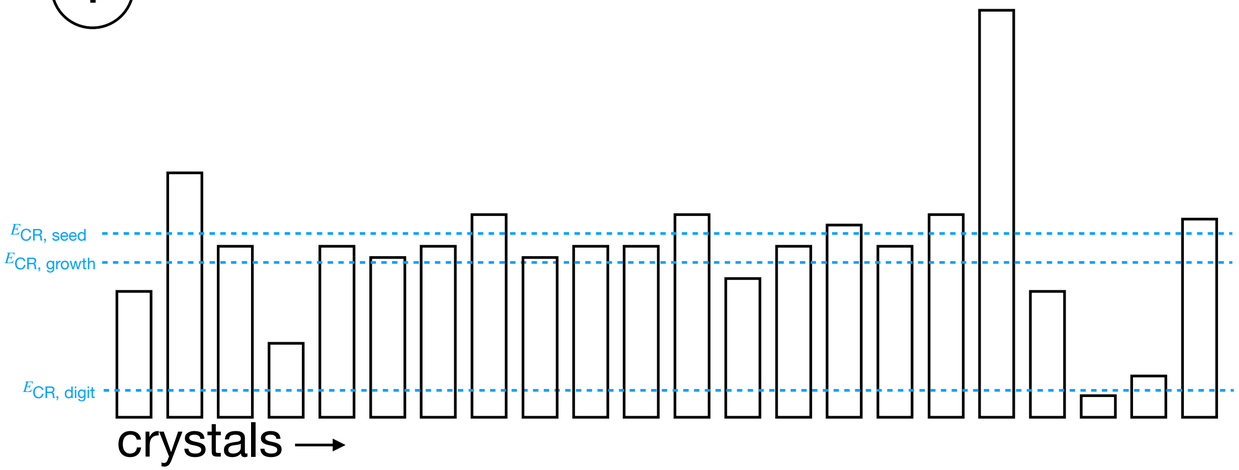
Introduction

- ECL reconstruction is based on BaBar calorimeter reconstruction with some adjustments for better robustness against beam backgrounds
  - Reduce combinatorics by creating topologically connected regions (CR) first
  - Energy depositions from one particle should be contained in the same CR
  - Ideally only truly overlapping particles end up in the same CR
- Actual and recent simulated beam backgrounds are much (!) higher than anticipated when designing the software (see BELLE2-NOTE-TE-2016-001)
  - Most CRs contained only one true particle and 30-50 hit crystals. They almost always had one or two local maximums as shower seeds
  - This is not true anymore, and various protection mechanisms in our algorithms work but result in deterministic, but hard to calibrate photon efficiency losses
- Challenges: Low energy photons, overlapping energy depositions, low p_t track-cluster matching, photon identification close to charged tracks, …
- **Changes to HLT and offline reconstruction are mandatory to recover efficiency losses and reduce code execution time**

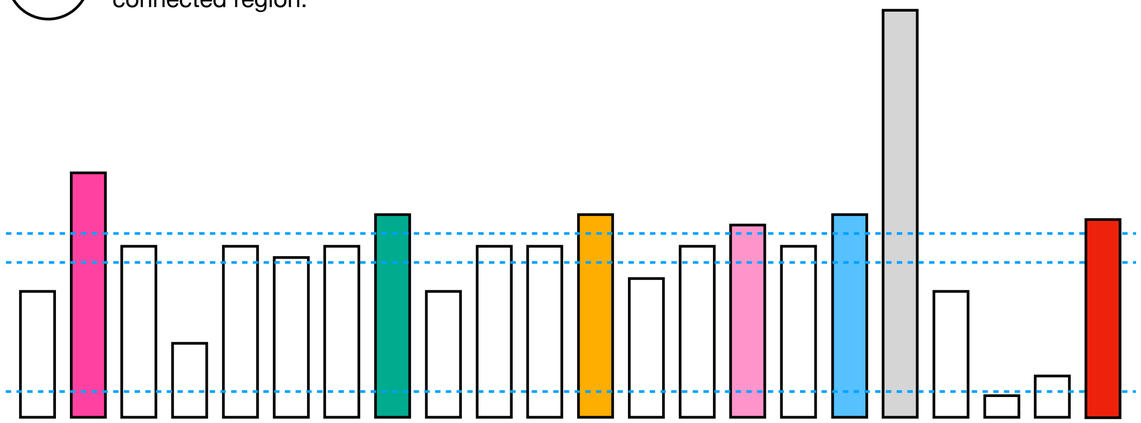
ECL clustering: Connected regions

$E_{\text{CR, seed}} \geq E_{\text{CR,growth}} \geq E_{\text{CR,digit}}$

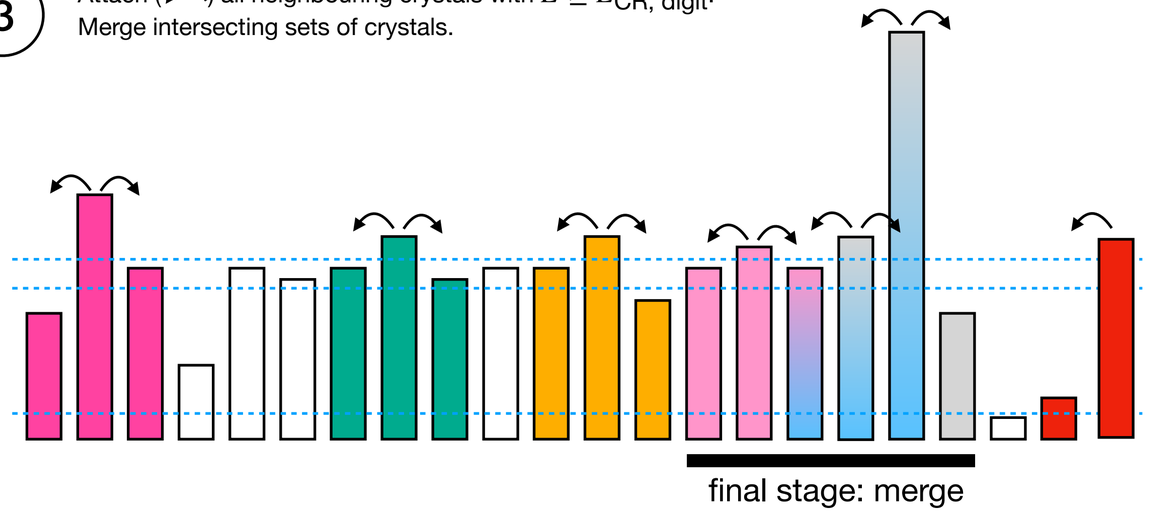
1 All crystals in the event.



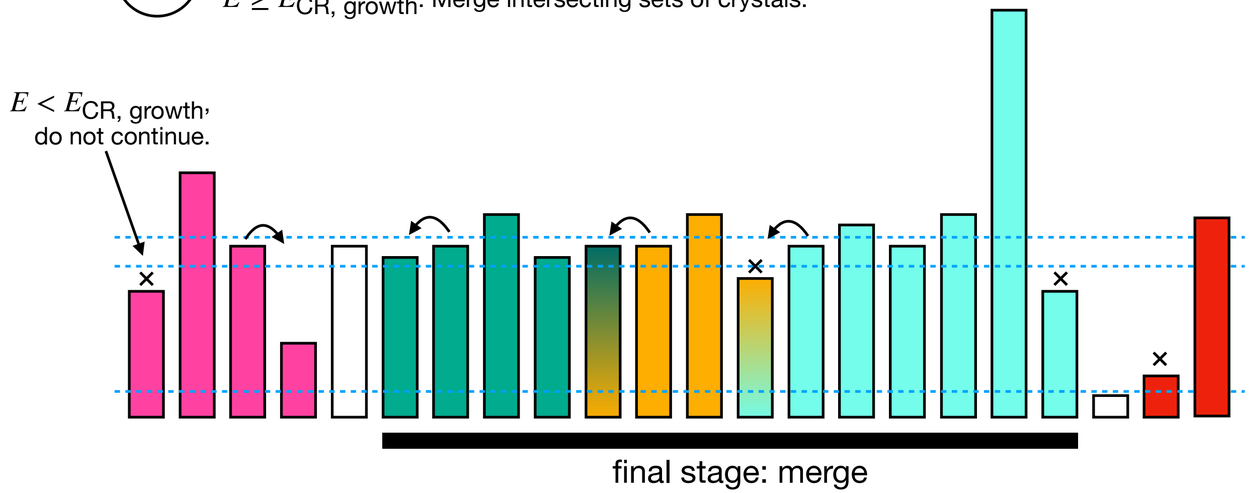
2 Every crystal with $E \geq E_{\text{CR, seed}}$ starts a connected region.



3 Attach (↶) all neighbouring crystals with $E \geq E_{\text{CR, digit}}$. Merge intersecting sets of crystals.



4 Continue attaching all neighbouring crystals with $E \geq E_{\text{CR, digit}}$ of all crystals with $E \geq E_{\text{CR, growth}}$. Merge intersecting sets of crystals.

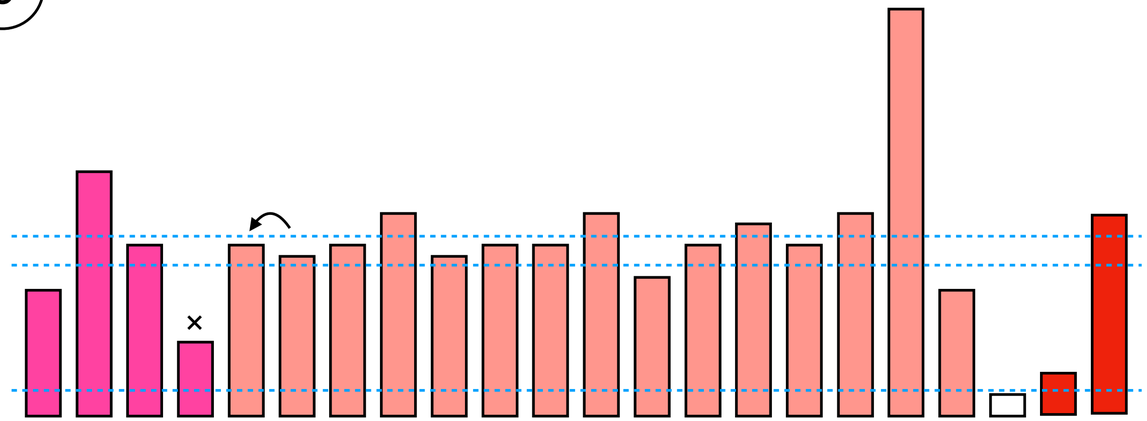


https://gitlab.desy.de/belle2/software/basf2/-/tree/main/ecl/modules/eclCRFinder/src?ref_type=heads

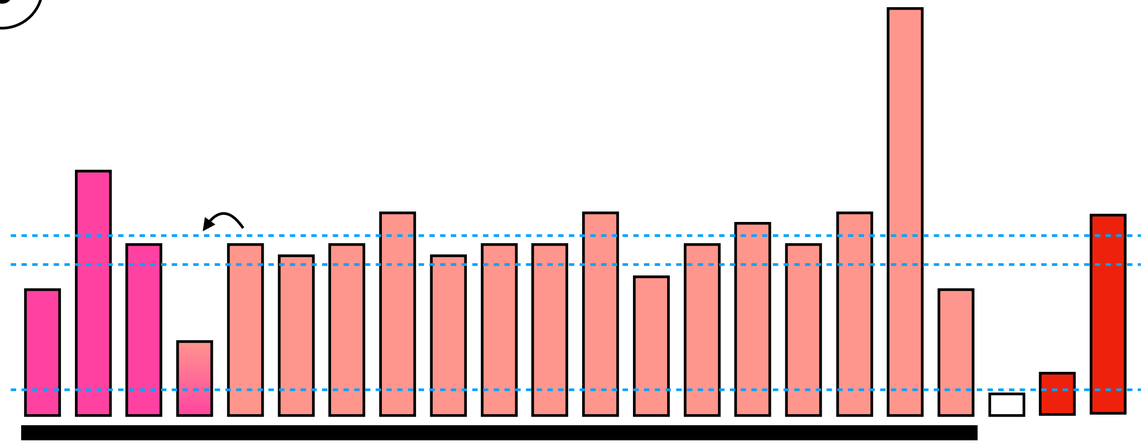
ECL clustering: Connected regions

$E_{\text{CR, seed}} \geq E_{\text{CR,growth}} \geq E_{\text{CR,digit}}$

5 Continue until there are no growth crystals left.

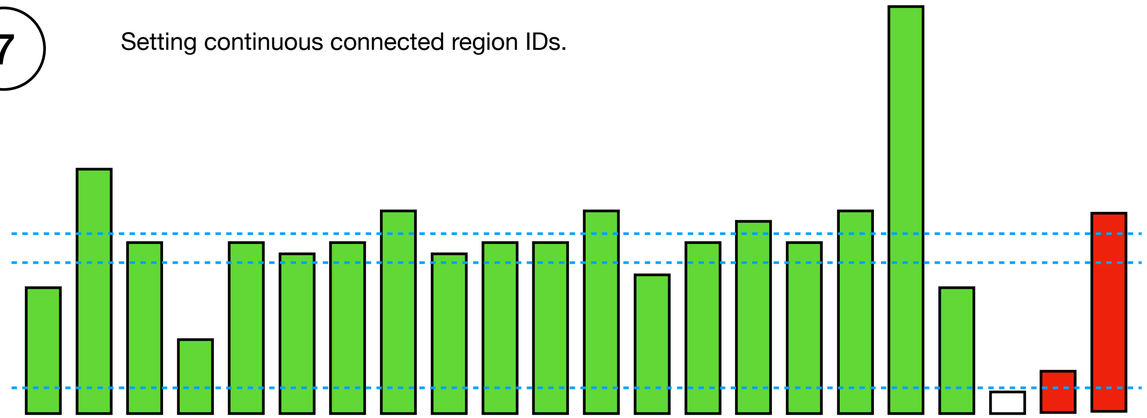


6 Continue until there are no growth crystals left.



final stage: merge

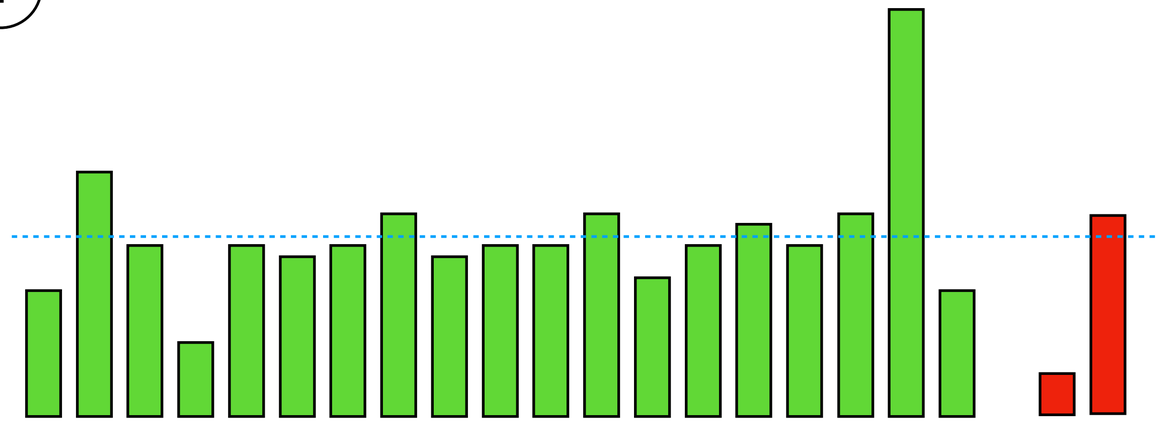
7 Setting continuous connected region IDs.



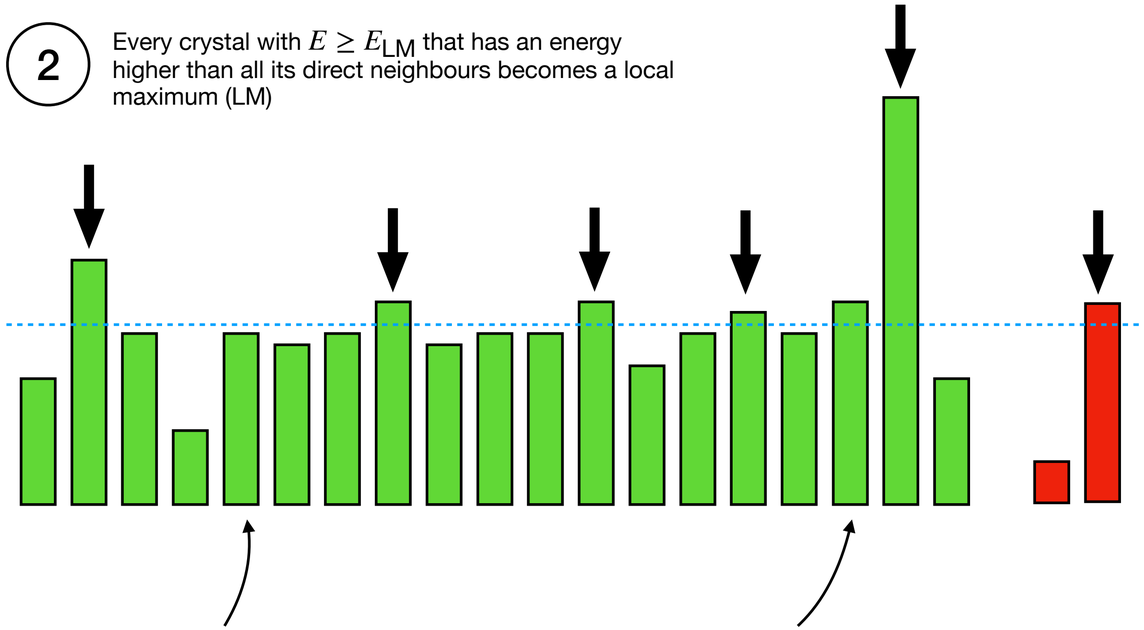
CRId = 1 CRId = 2

ECL clustering: Local Maximum Finder

1 All connected regions in the event.



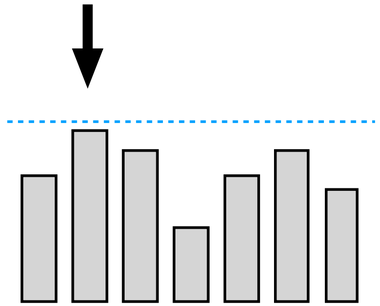
2 Every crystal with $E \geq E_{\text{LM}}$ that has an energy higher than all its direct neighbours becomes a local maximum (LM)



Has no neighbours with higher energy, but $E < E_{\text{LM}}$.

Above $E \geq E_{\text{LM}}$, but has neighbours with higher energy

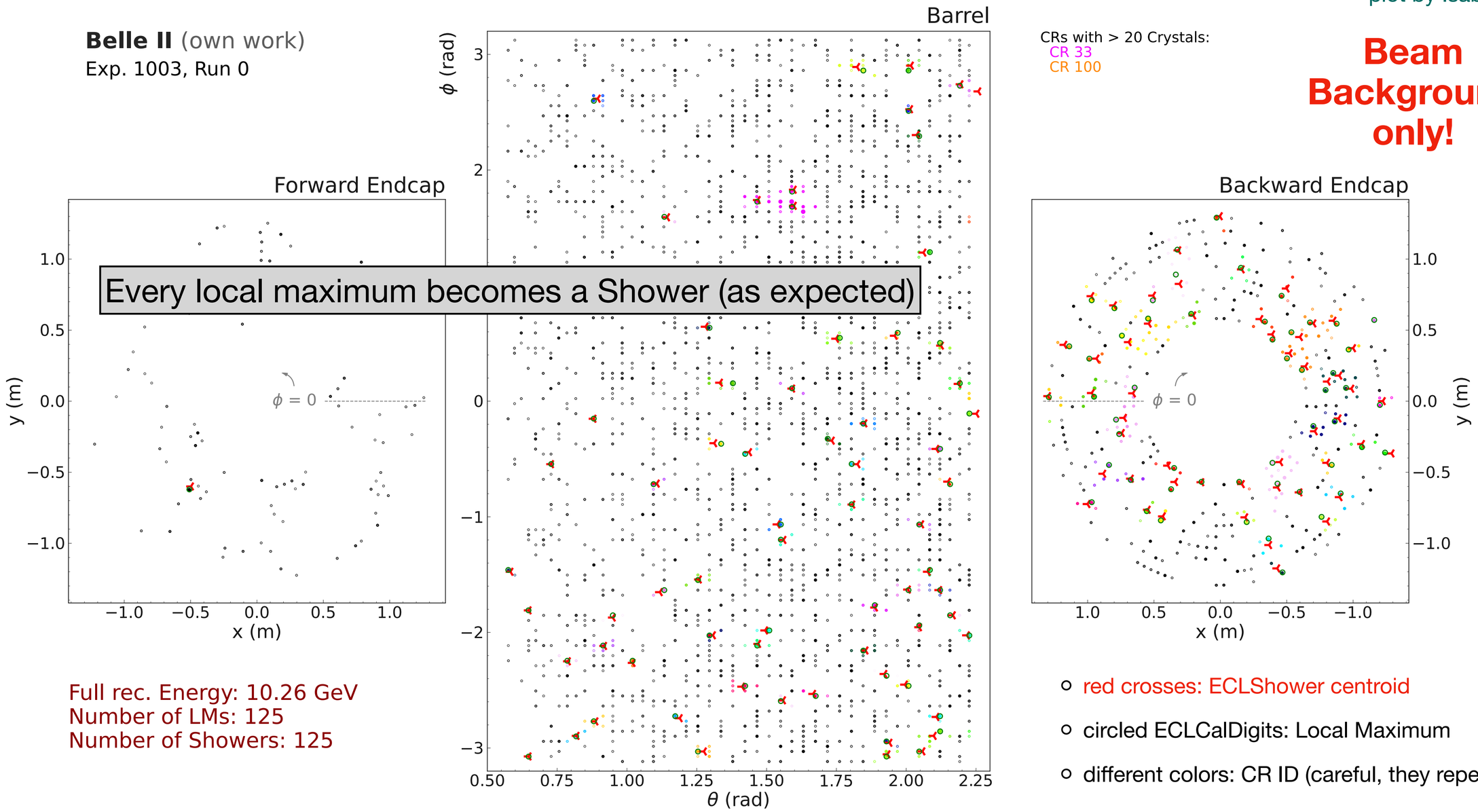
3 If a connected region has no LM (e.g. if $E_{\text{LM}} > E_{\text{CR, seed}}$), use the highest energy crystal in the CR as local maximum (the first from a loop over all ECLCalDigits if there are multiple) - **why we implemented this? I have no idea.** Current default values prevent this.



https://gitlab.desy.de/belle2/software/basf2/-/tree/main/ecl/modules/eclLocalMaximumFinder/src?ref_type=heads

ECL clustering event display exp1003 backgrounds

plot by Isabel Haide

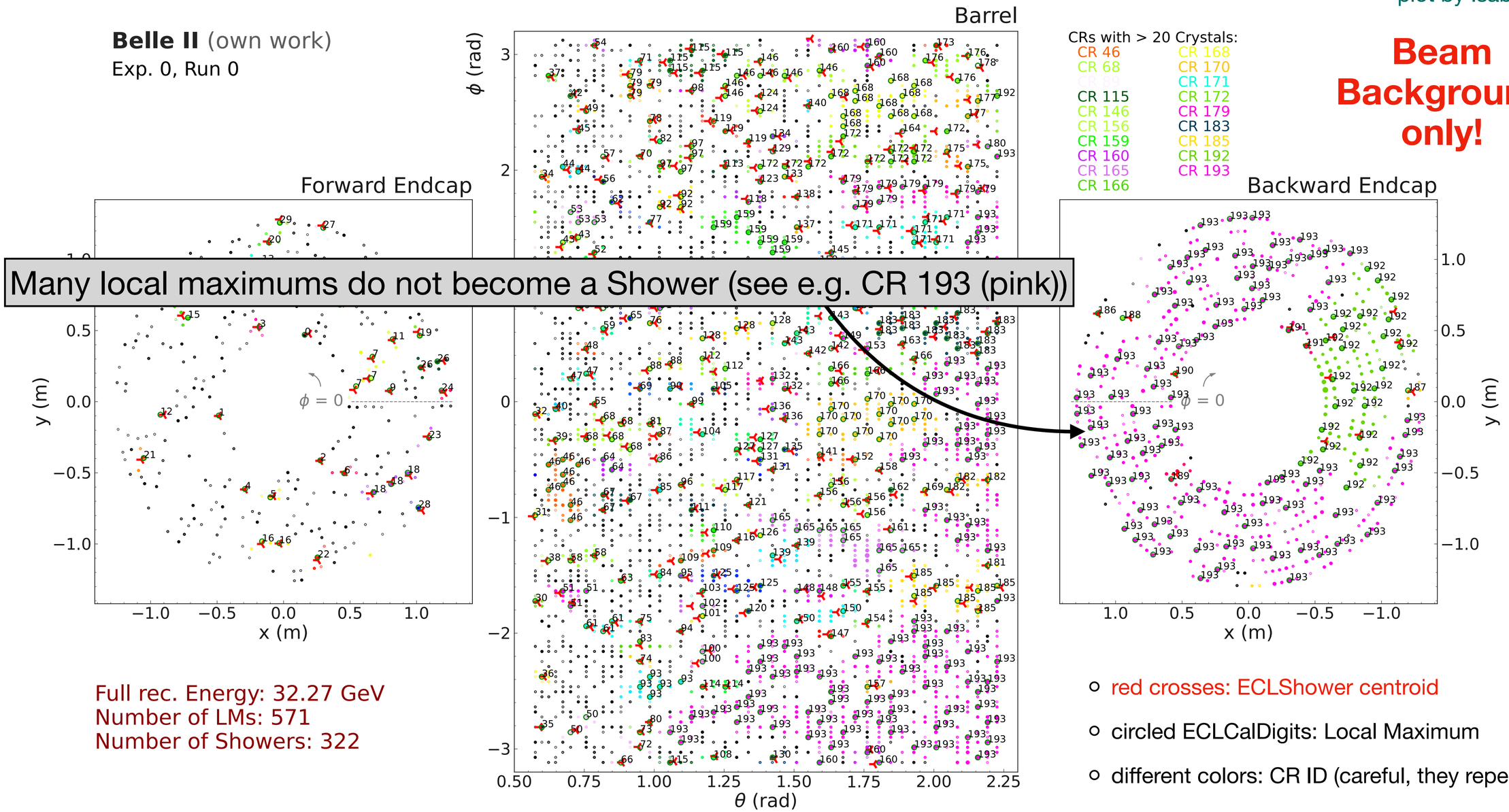


Beam Background only!

- red crosses: ECLShower centroid
- circled ECLCalDigits: Local Maximum
- different colors: CR ID (careful, they repeat)

ECL clustering event display exp0 backgrounds

plot by Isabel Haide

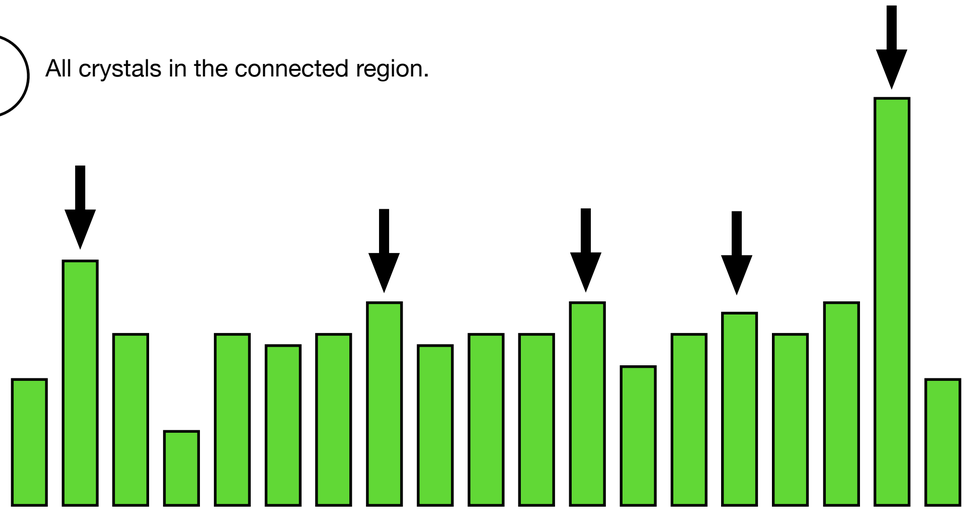


- red crosses: ECLShower centroid
- circled ECLCalDigits: Local Maximum
- different colors: CR ID (careful, they repeat)

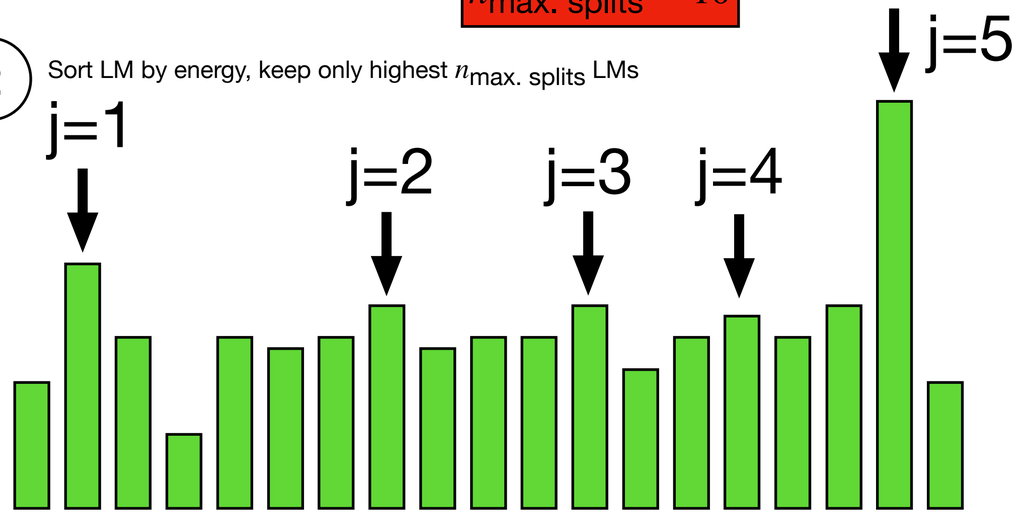
ECL clustering: Splitter N1 for $n_{\mathbf{LM}} > 1$

Current defaults:
$n_{\text{max. splits}} = 10$

1 All crystals in the connected region.

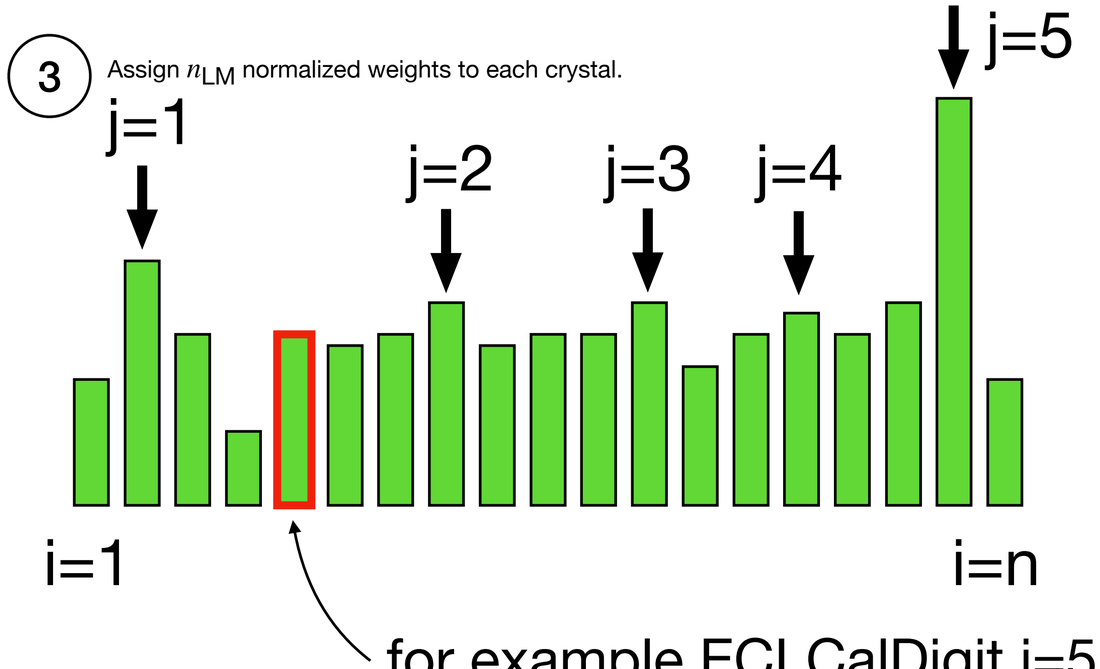


2 Sort LM by energy, keep only highest $n_{\text{max. splits}}$ LMs



ECL clustering: Splitter N1 for $n_{\mathbf{LM}} > 1$

Current defaults:
$a = 2.5/3.581$
threshold $= 2.5$ keV



for example ECLCalDigit i=5:

$w_{51} = 0.5$
$w_{52} = 0.45$
$w_{53} = 0.04$
$w_{54} = 0.006$
$w_{55} = 0.004$

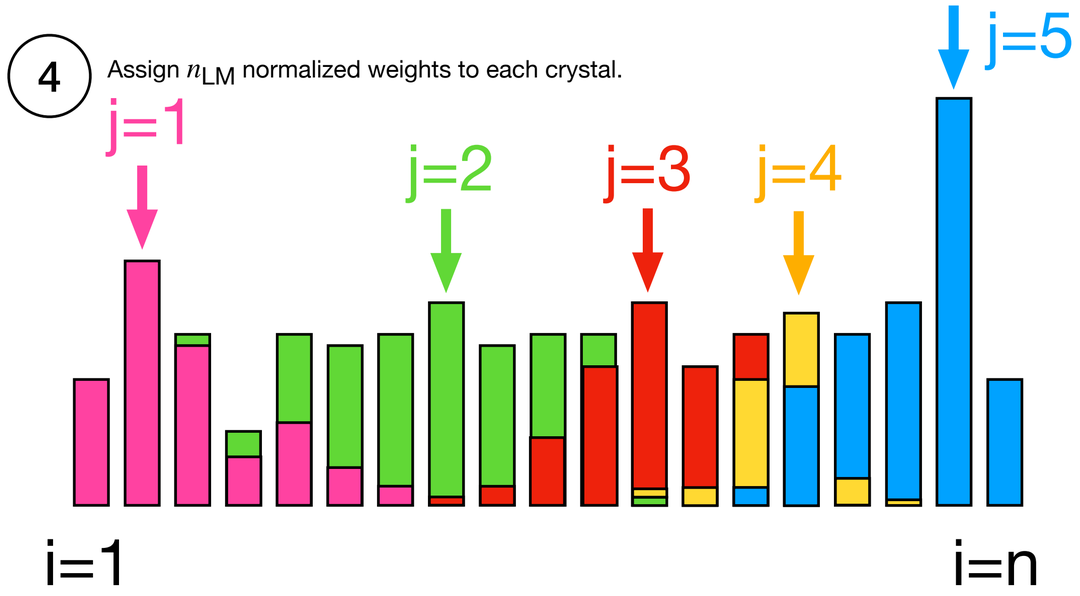
Weight **i**-th ECLCalDigit in the CR to the **j**-th LM with weight

$$w_{ij} = E_j \exp(a d_{ij}) / \Sigma_k E_k \exp(a d_{kj})$$

- a is a constant
- E_j is the LM energy
- E_k is the k-th ECLCalDigit energy
- d_{ij} is the distance between the **j**-th shower candidate centroid and the **i**-th ECLCalDigit

If w_{ij} is smaller than a threshold: $w_{ij} = 0$

ECL clustering: Splitter N1 for $n_{\mathbf{LM}} > 1$



These conditions remove a lot of potential shower candidates if there are too many energy depositions of similar magnitude because the reweighting becomes unstable.

The solution for this to limit the size of the CR (see Jonas talk for state-of-the art LM classification).

Repeat the weighting procedure from step 3, but use the updated centroid positions and ECLCalDigit weights.

1) Calculate the mean centroid shift of all LMs. If it is below 1mm, exit the procedure.

2) Calculate the weighted energy per crystal in the shower candidates. If this value is maximal in an **ECLCalDigit that is not identical to the LM, mark this shower candidate for deletion**.

3) If the highest weighted energy is **below a threshold $E_{\mathbf{min}}$, mark this shower candidate for deletion**.

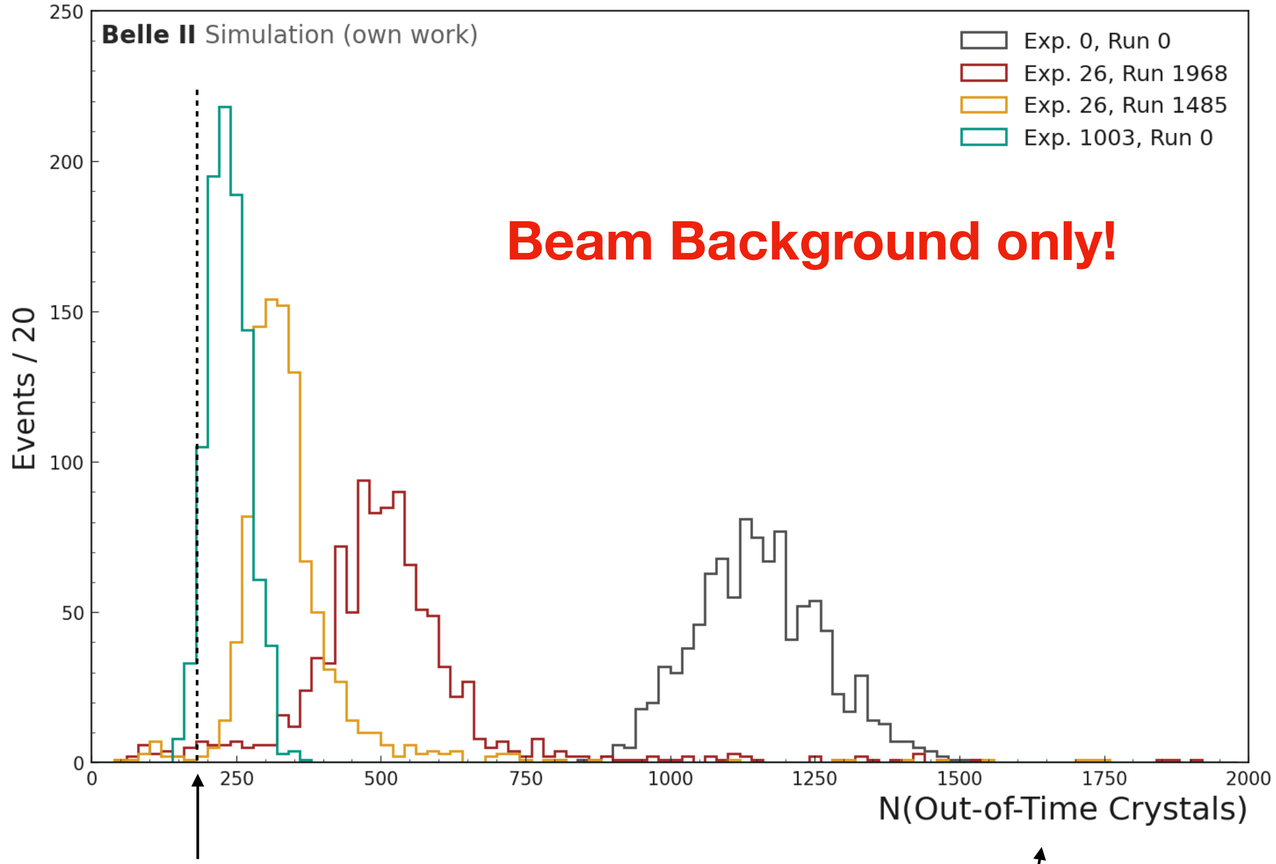
4) If any condition 2) or 3) is fulfilled, remove those shower candidates*.

*This can mean that we remove all shower candidates and leave an empty CR, which can be seen in the event display on slide 7 in CR193.

Technical (for experts)

plot by Isabel Haide

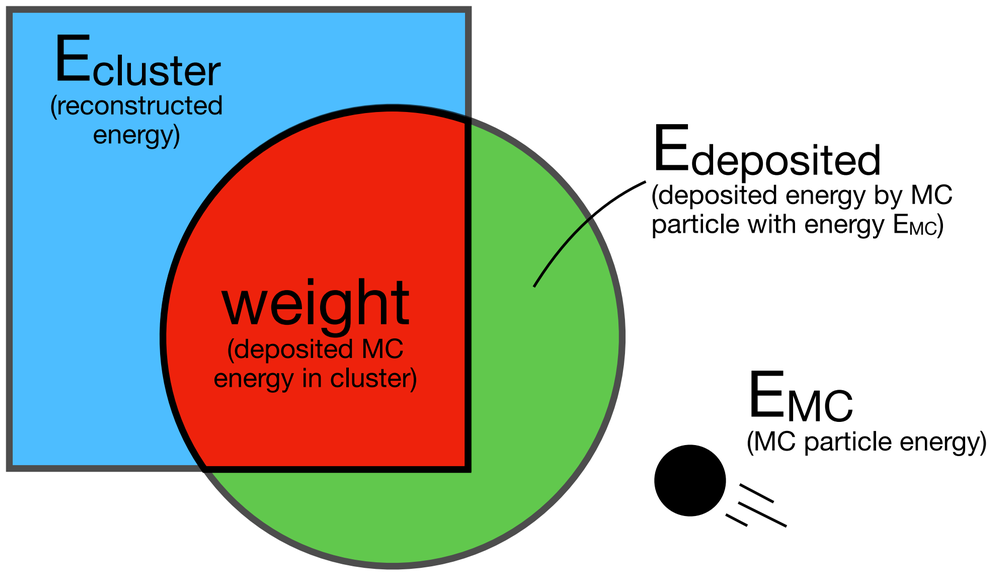
- Simulation and reconstruction (by Isabel), and analysis (by me) with release-08-01-00 with cherry-pick of commit 149f54b23b3c649a200dff87e1711d7a1e297dbe (https://gitlab.desy.de/belle2/software/basf2/-/merge_requests/3425/diffs?commit_id=149f54b23b3c649a200dff87e1711d7a1e297dbe)
- We will show studies for three different beam background conditions:
  - run independent exp0 (expected nominal high backgrounds)
  - run independent exp1003 (expected “early phase 3” backgrounds, rather low for ECL compared to actual 2023 conditions)
  - run-dependent exp26r1968 (one of the highest background runs in exp26)
- We will show studies for four different setting in the two modules ECLCRFinder and ECLLocalMaximumFinder:
  - A. $E_{seed} = 10\,\text{MeV}, t_{seed} = \infty, E_{growth} = 10\,\text{MeV}, E_{LM} = 10\,\text{MeV}$ (current baseline)
  - B. $E_{seed} = 20\,\text{MeV}, t_{seed} = 200\,\text{ns}, E_{growth} = 20\,\text{MeV}, E_{LM} = 20\,\text{MeV}$ (planned HLT and offline setting)
  - C. $E_{seed} = 20\,\text{MeV}, t_{seed} = 200\,\text{ns}, E_{growth} = 10\,\text{MeV}, E_{LM} = 20\,\text{MeV}$ (as B, but larger connected regions)
  - D. $E_{seed} = 20\,\text{MeV}, t_{seed} = \infty, E_{growth} = 20\,\text{MeV}, E_{LM} = 20\,\text{MeV}$ (as B, but not timing selections)

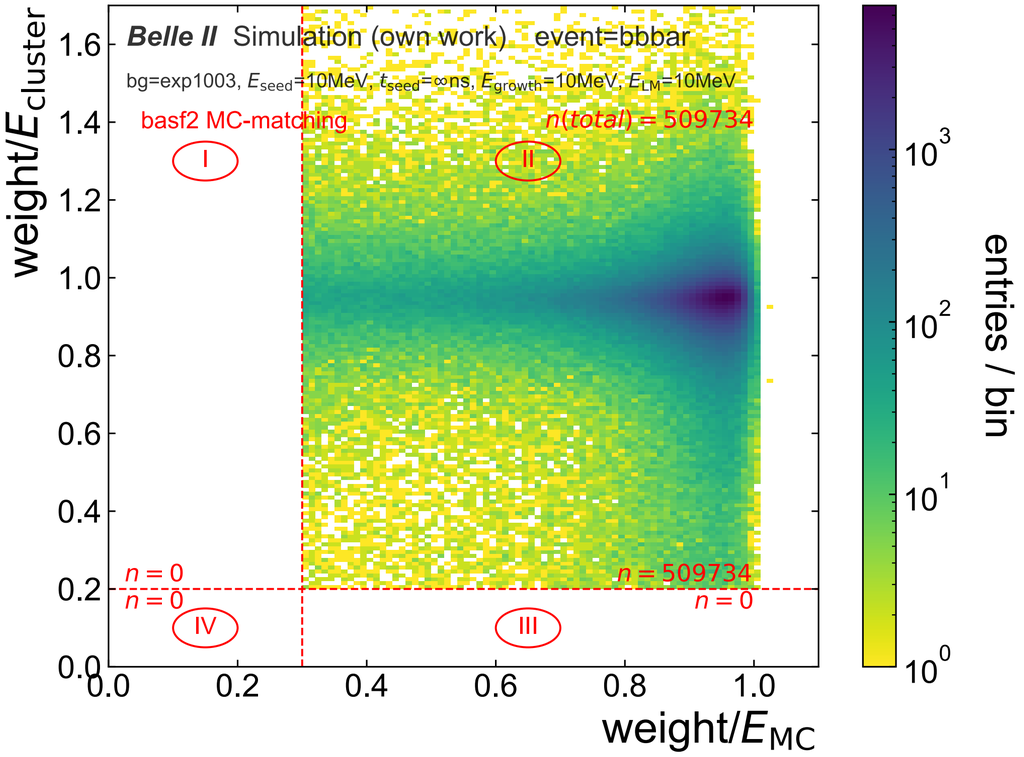


basf2 ECL reconstruction uses <N>=183 as proxy for “very high beam backgrounds” (based on release-05 beam backgrounds!)

basf2 variable **nECLOutOfTimeCrystals**

MC-Matching of photons, $E_{cluster}$ > 25 MeV, 50k B$\bar{B}$ events



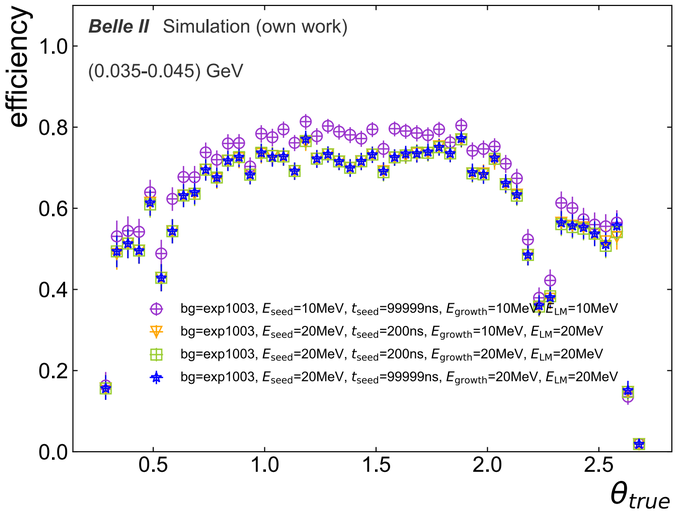


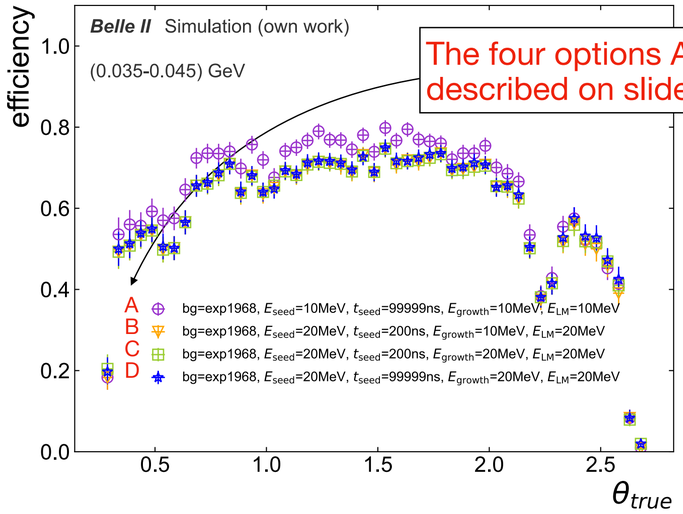
$$\text{efficiency} = \frac{\text{ECL cluster with } w/E_{\text{MC}} > 0.3 \text{ and } w/E_{\text{cluster}} > 0.2 \text{ and } mcPDG = 22}{\text{all generated photons}}$$

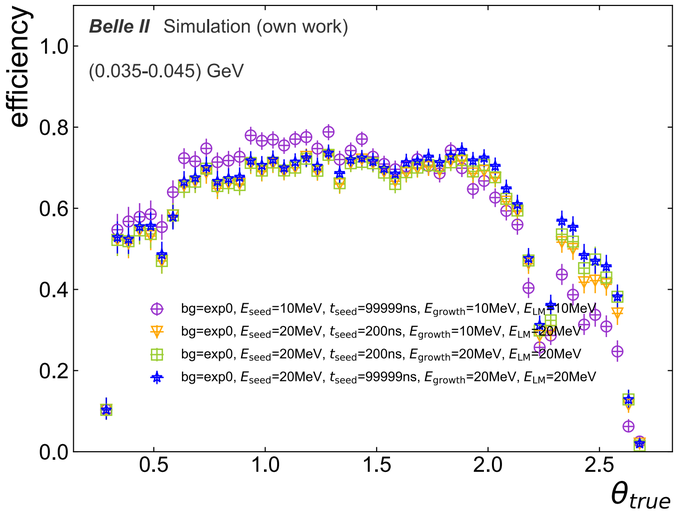
Photon efficiency for different beam backgrounds

exp1003 | r26exp1968 | exp0

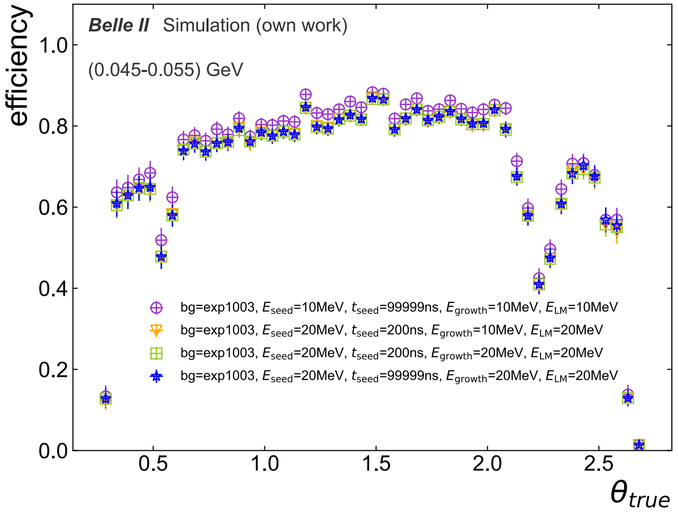
0.035-0.045 GeV

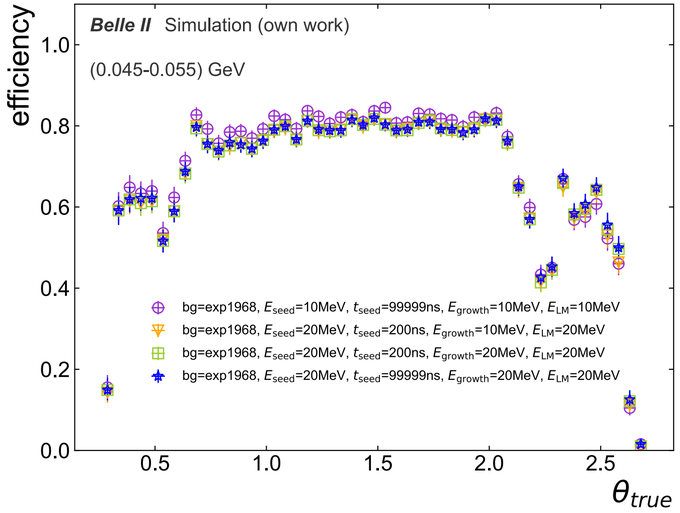


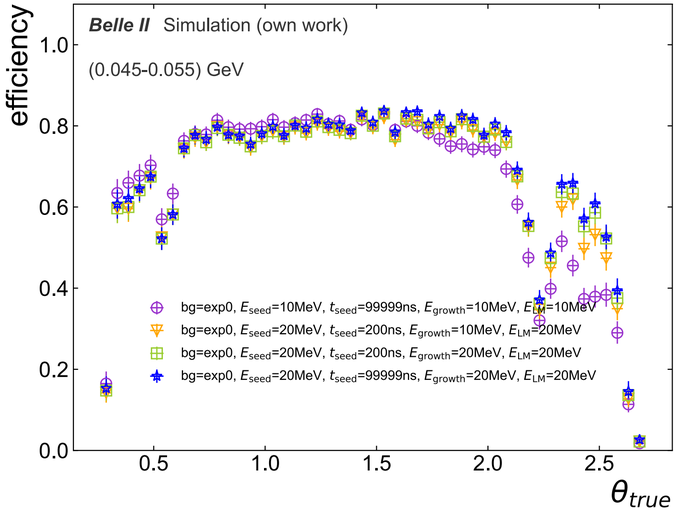




0.045-0.055 GeV



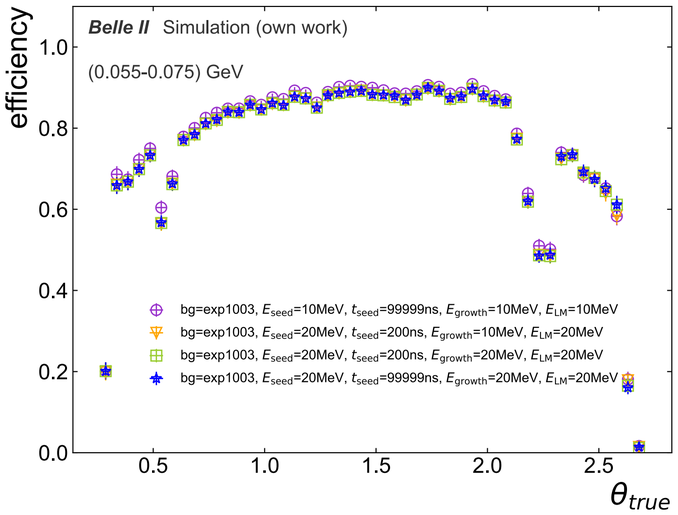


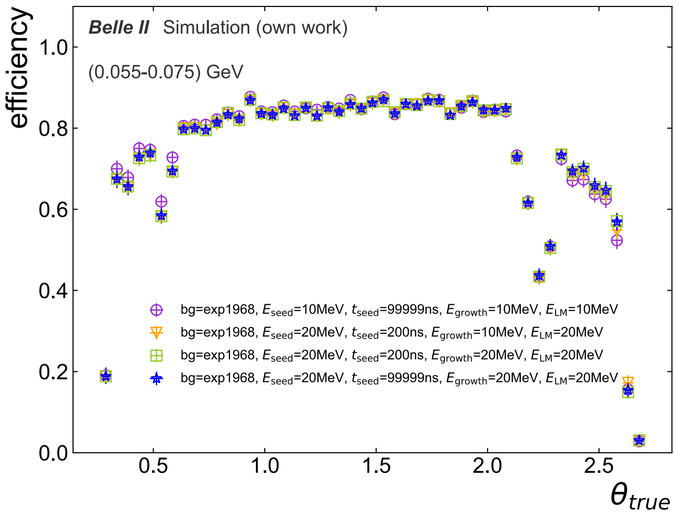


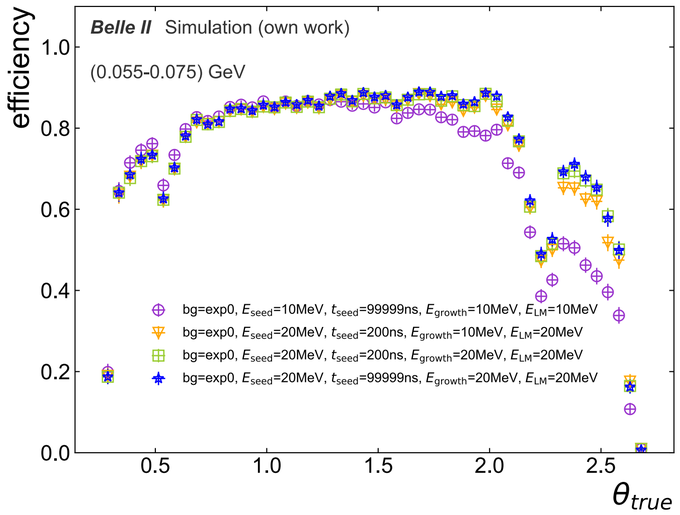
Photon efficiency for different beam backgrounds

exp1003 | r26exp1968 | exp0

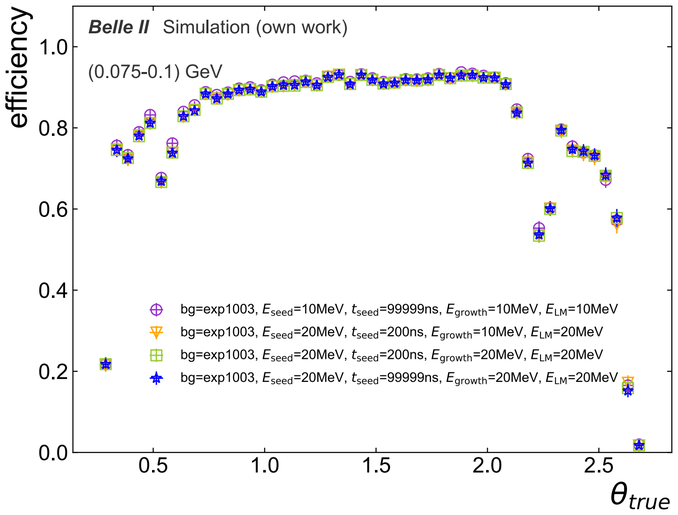
0.055-0.075 GeV

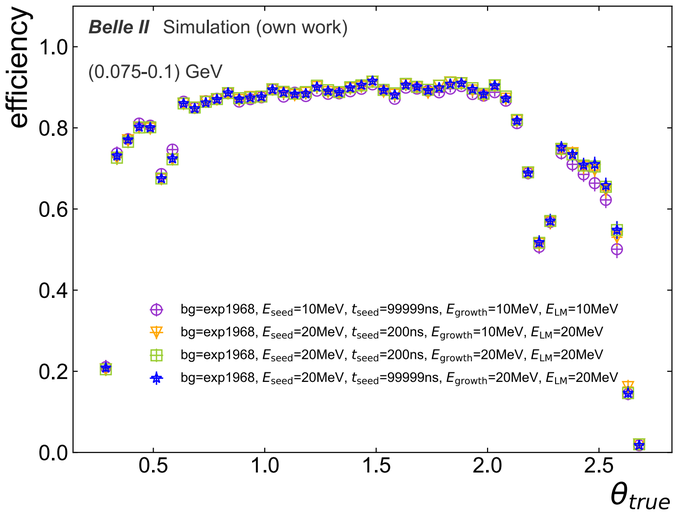


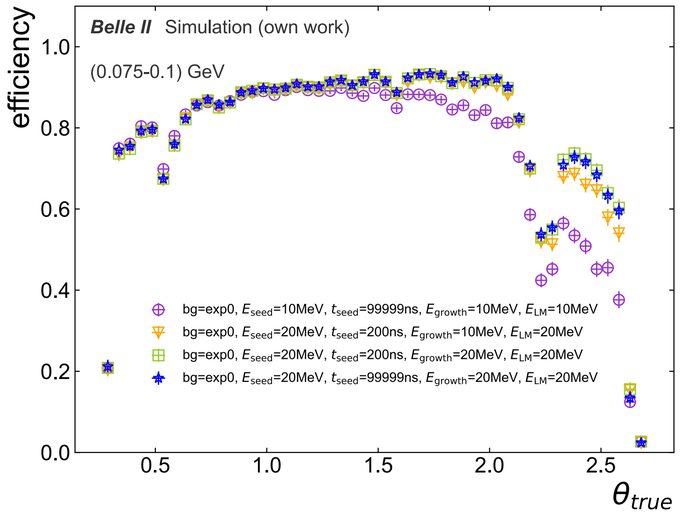




0.075-0.1 GeV



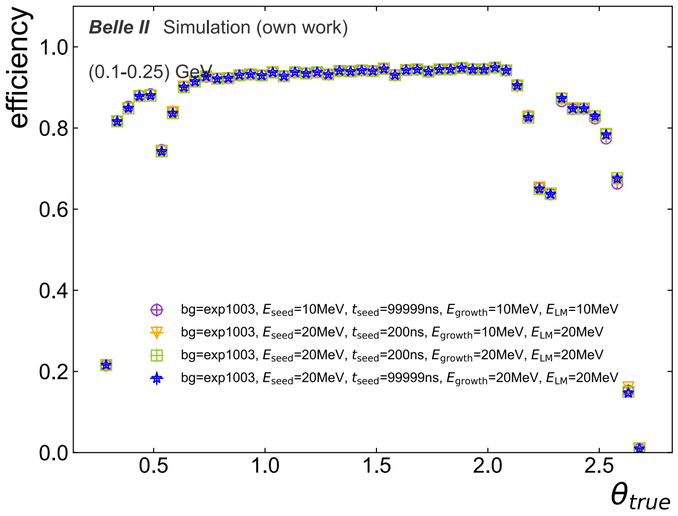


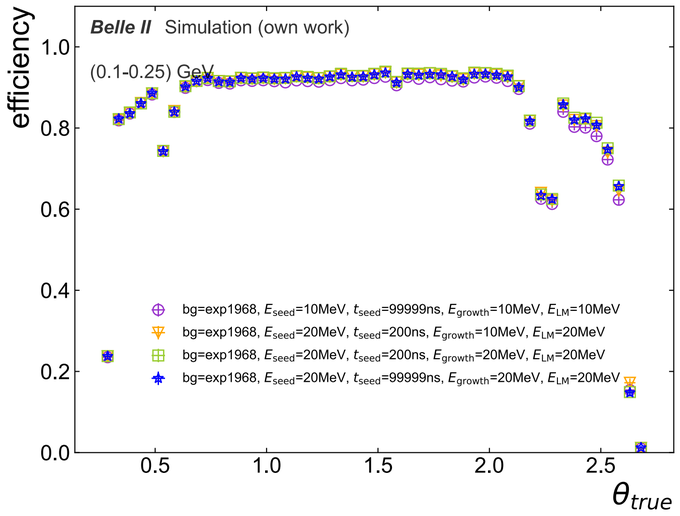


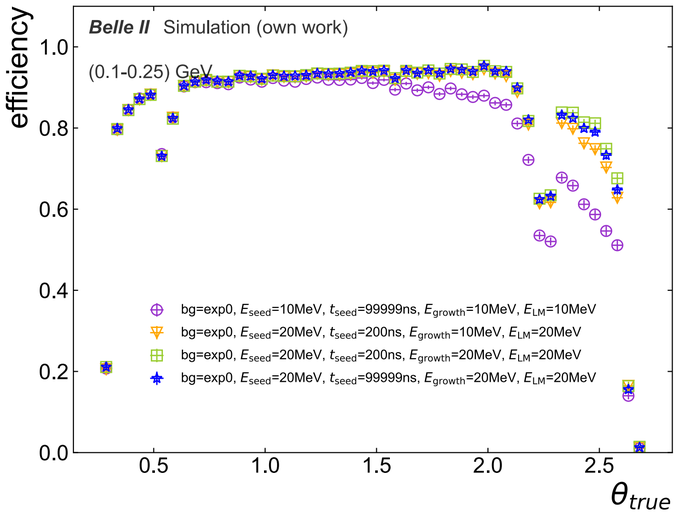
Photon efficiency for different beam backgrounds

exp1003 | r26exp1968 | exp0

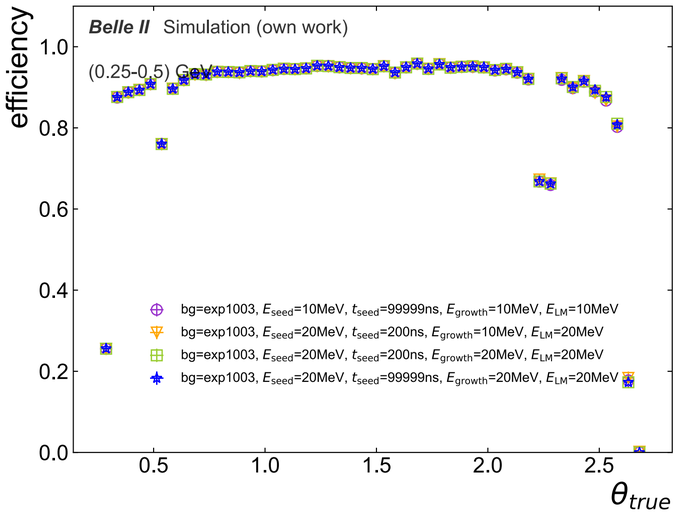
0.1-0.25 GeV

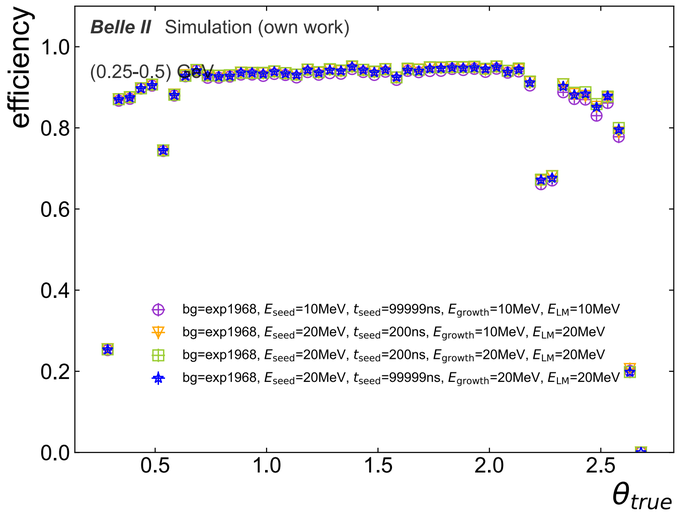


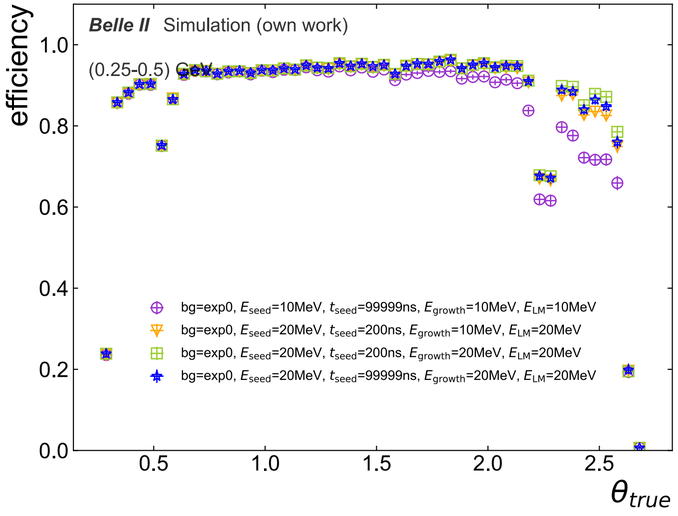




0.25-0.5 GeV



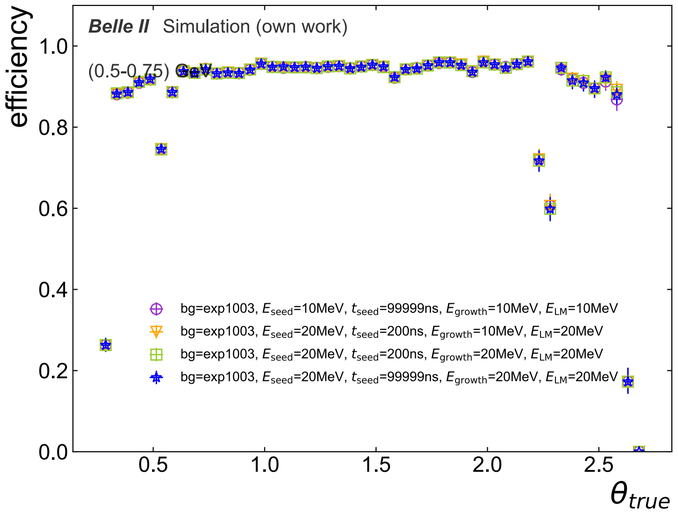


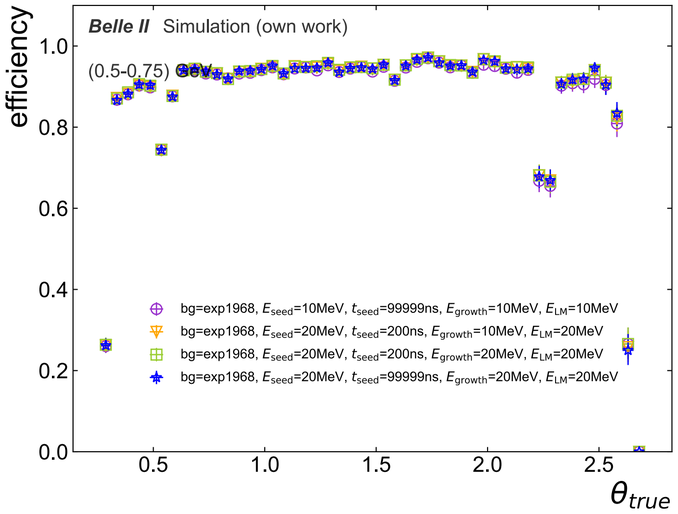


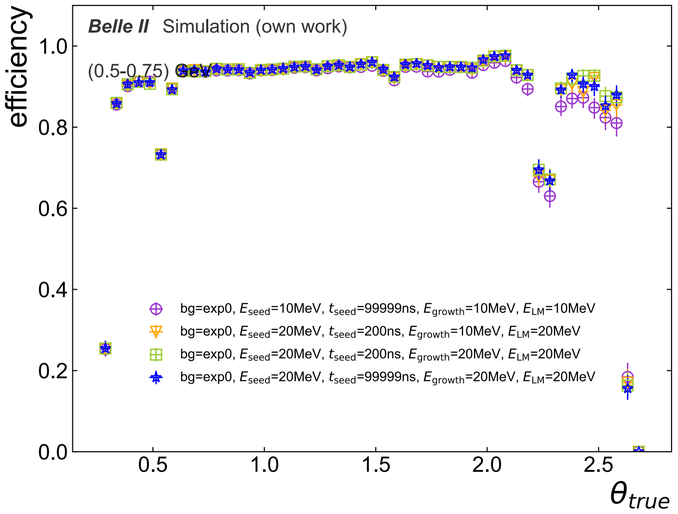
Photon efficiency for different beam backgrounds

exp1003 | r26exp1968 | exp0

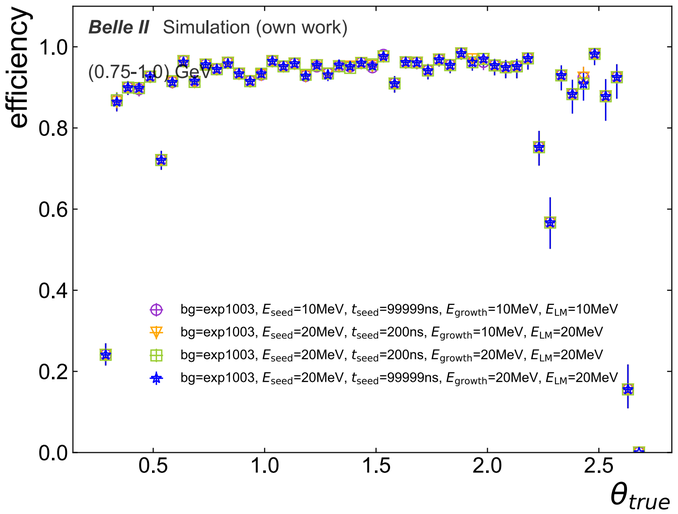
0.5-0.75 GeV

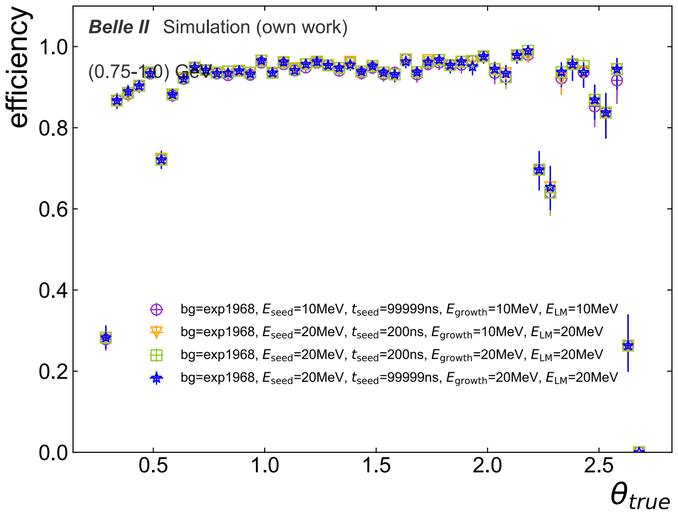


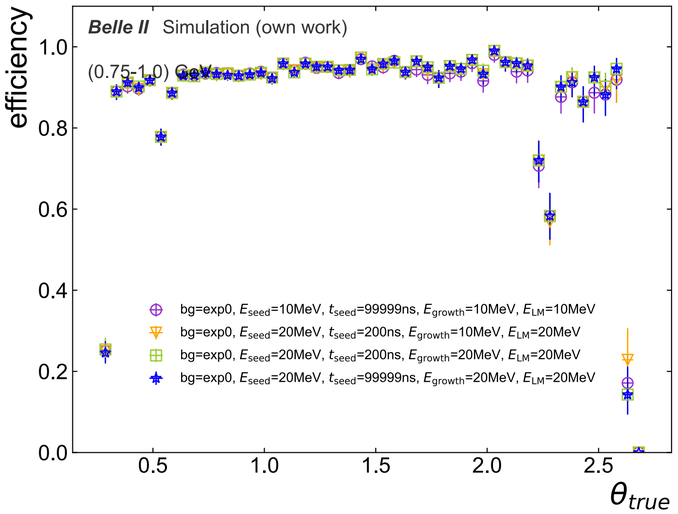




0.75-1.0 GeV







Conclusion

- The background conditions, especially close to injection, are extreme, **analysts need to filter events** that have a large number of out-of-time crystals **if they are sensitive to inefficiencies** (we do this in the single photon analysis for example)
- Post-reconstruction MVAs do not help for this challenge
- ECL will implement the following changes to reconstruction:
  - **Increase seed and growth thresholds to 20 MeV and apply a timing cut of 200 ns to seed crystals (via module parameters, no code changes needed)**
  - future: limit the weight distance of crystals possibly belonging to a LM to 5 CsI Moliere radii (~20cm)
  - future: allow LM shifts by one crystal (relax condition 3 on page 11)
  - medium future: LM classifier (Jonas), GNN reconstruction (F. Wemmer et al., ”Photon Reconstruction in the Belle II Calorimeter Using Graph Neural Networks”, https://link.springer.com/article/10.1007/s41781-023-00105-w)

Backup

Baseline clustering: CRFinder

Current defaults:
$E_{\text{CR, seed}} = 10\,\text{MeV}$
$E_{\text{CR, growth}} = 10\,\text{MeV}$
$E_{\text{CR, crystal}} = 0.5\,\text{MeV}$

- Three energy thresholds with enforced hierarchy $E_{\text{CR, seed}} \geq E_{\text{CR,growth}} \geq E_{\text{CR,digit}}$
  - Timing cuts are possible (need release-09 bug fix to use absolute timing cuts)
- Every ECLCalDigit with $E \geq E_{\text{CR, seed}}$ starts a connected region (CR) candidate
- Every direct neighbour (including corners) with $E \geq E_{\text{CR,digit}}$ is added to the CR candidate
- If a ECLCalDigit is shared by more than one CR candidate, the CR candidates are merged into a new CR candidate
- If any ECLCalDigit in a CR candidate has $E \geq E_{\text{CR,growth}}$, every direct neighbour (including corners) with $E \geq E_{\text{CR,digit}}$ is added to the CR candidate. If a ECLCalDigit is shared by more than one CR candidate, the CR candidates are merged into a new CR candidate.
- This process continues until no more crystals with $E \geq E_{\text{CR,growth}}$ exist.
- In a final step, continuous CR IDs from 1.. n are assigned to all n CRs.
- At the end, every ECLCalDigit belongs to no or exactly one CR (the design goal is to remove as many crystals in this step as possible to speed up the ECLN1Splitter).

Baseline clustering: Local Maximum Finder

Current default:
$E_{\text{LM}} = 10\,\text{MeV}$

- The Local Maximum Finder loops over all connected regions, ECLCalDigit that are in no CR are not considered.
- In each CR, every ECLCalDigit with energy $E > E_{\text{LM}}$ that has also a higher ECLCalDigit energy than all its direct neighbours becomes a local maximum.
- There is no relation between the $E_{\text{CR, seed}}$ or $E_{\text{CR, growth}}$, and E_{LM}, and especially $E_{\text{LM}} > E_{\text{CR, seed}}$ is technically possible.
- If a connected region has no LM (e.g. if $E_{\text{LM}} > E_{\text{CR, seed}}$, use the highest energy crystal in the CR as local maximum (the first from a loop over all ECLCalDigits if there are multiple)
- *Note: The number of LMs in the FWD, barrel and BWD are stored in the mdst dataobject EventLevelClusteringInfo.*

Baseline clustering: Splitter N1

- Splitter N1 is meant to provide optimal response for electromagnetic clusters and is optimized for photons between 30 MeV and 7 GeV.
- The main (and algorithmically only) challenge is to split energies if two particles deposit energy close by so that ECLCalDigit must be shared between two (or more) ECLShowers.
- The design idea is inspired by BaBar and based on an iterative weighting and calculation of shower positions until all shower positions are stable (usually in 1-2 iterations).
- The code was developed and tested on 2016 “high beam backgrounds”: Most CRs contained only one true particle and contained 30-50 ECLCalDigits. They almost always had 1 or 2 LMs, rarely more.
- The case $n_{\mathrm{LM}} = 1$ is so common, that this case that is always checked first.

https://gitlab.desy.de/belle2/software/basf2/-/blob/main/ecl/modules/eclSplitterN1/src/ECLSplitterN1Module.cc?ref_type=heads

Baseline clustering: Splitter N1 for $n_{\mathbf{LM}} = 1$

- The N1Splitter loops over all connected regions, ECLCalDigit that are in no CR are not considered.
- To get a measure of beam background levels, we divide, event-by-event, the number of out-of-time ECLCalDigit (see slide 3) by 183.
  - For all recent runs, this ratio is always ≈1.
- To get a first energy estimate, we get the energy sum in the 3×3 crystals centred at the LM.
- Using the background fraction and the first energy estimate [1], we sum the highest n crystals within up to 5×5 (excluding corners) ECLCalDigits for the raw energy estimate.
- We use all crystals within up to 5×5 (excluding corners) ECLCalDigits for all shower shape calculations and the position calculations [2].

https://gitlab.desy.de/belle2/software/basf2/-/blob/main/ecl/modules/eclSplitterN1/src/ECLSplitterN1Module.cc?ref_type=heads
[1] Number of optimal crystals: https://gitlab.desy.de/belle2/software/basf2/-/blob/main/ecl/dbobjects/include/ECLnOptimal.h?ref_type=heads
[2] Position calculation: https://gitlab.desy.de/belle2/software/basf2/-/blob/main/ecl/utility/src/Position.cc