Analysis of samples for IDAO

Giulia D’Imperio

Reconstruction and analysis meeting
21/01/21

Samples for International Data Analysis Olympiad (IDAO)

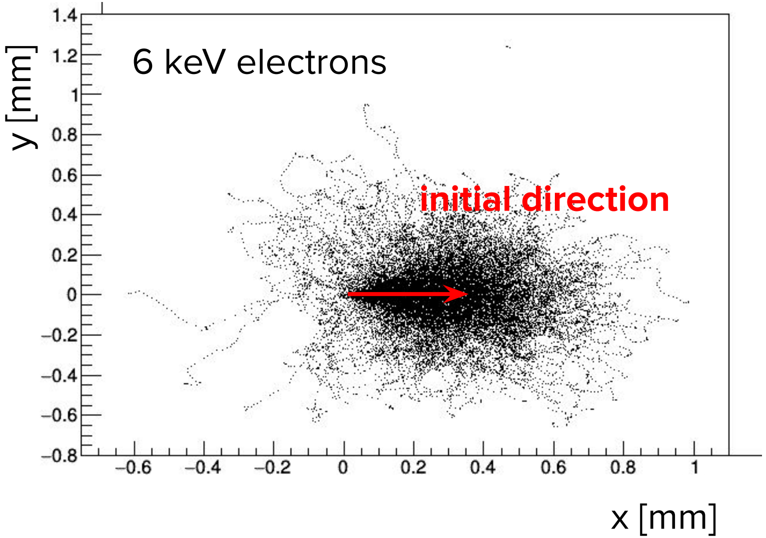
- First step simulation (not for IDAO)
  - MC-truth simulations with GEANT/SRIM (for electron/nuclear recoils) of the particles interactions in gas
    ➡ contain info of position and energy released for each interaction (x,y,z,dE)

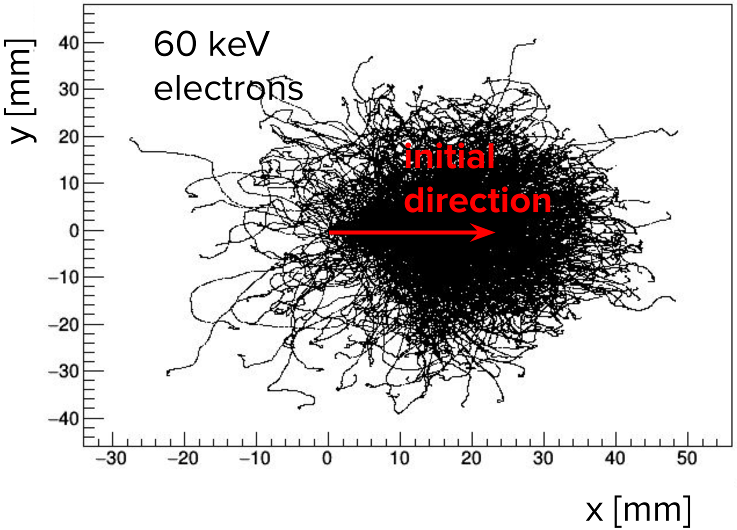
- For IDAO:
  - "Digitized" simulations of **signal only,** for electron recoils (ER) and nuclear recoils (NR): include detector effects like diffusion and GEM gain fluctuations
    ➡ chosen values: drift field E_drift = 930 V, distance from GEM d_GEM = 30 cm
  - Final images: **Signal + camera noise**

- Format of images: ROOT TH2 histograms

Geant4/SRIM simulation

- ER simulated with Geant4
- He NR simulated with SRIM
- 1000 events starting from the center
- Energies 1, 3, 6, 10, 30 keV
- Initial direction (1,0,0)

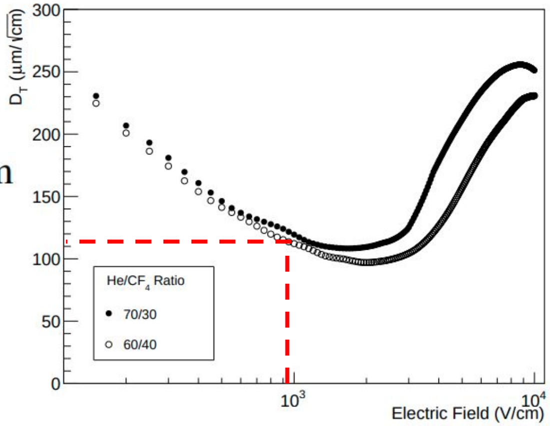




Digitization parameters

- Transverse diffusion from https://arxiv.org/abs/2007.00608 for an electric field of 0.93 kV/cm

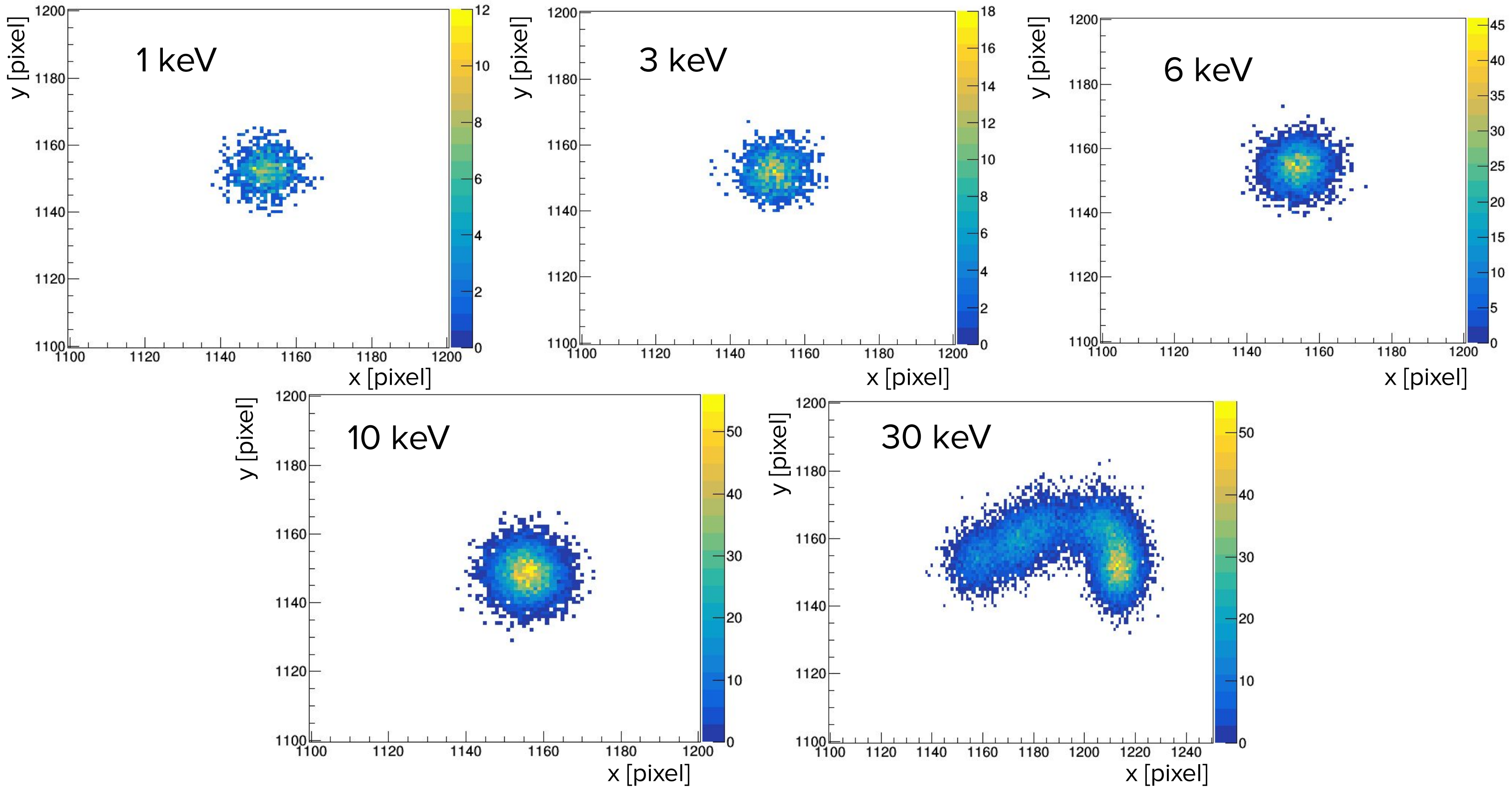
$$\sigma_T = \sqrt{\sigma_{T0}^2 \oplus D_T^2 \cdot z} \qquad D_T^{60/40} = 115 \ \frac{\mu m}{\sqrt{cm}} \qquad \sigma_{T0}^{60/40} = (280 \pm 60)\ \mu m$$



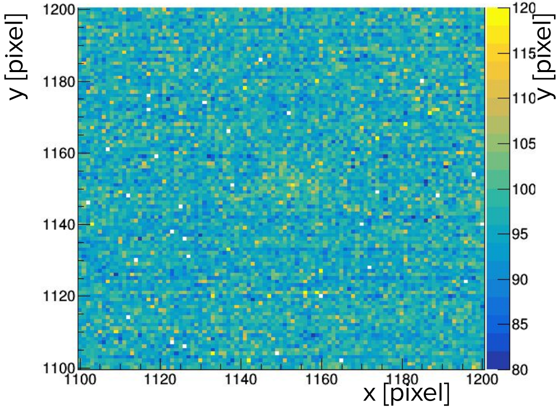
- Active area: 35 cm x 35 cm
- ORCA Fusion:
  - 2304 x 2304 pixels (1 pixel 6.5 um x 6.5 um)
  - Camera aperture 0.95
  - Sensor size 14.976 mm Orca Fusion
- Ionization potential: 46.2 eV (Garfield simulations 42-49 eV)
- Single GEM gain: 123 (see IEEE TRANSACTIONS ON NUCLEAR SCIENCE, VOL. 65, NO. 1, JANUARY 2018)
- light yield: 0.07 photons/electrons
- Sensor calibration ➜ 1 photon = 0.5 sensor counts
- Distance from the GEM: 30 cm

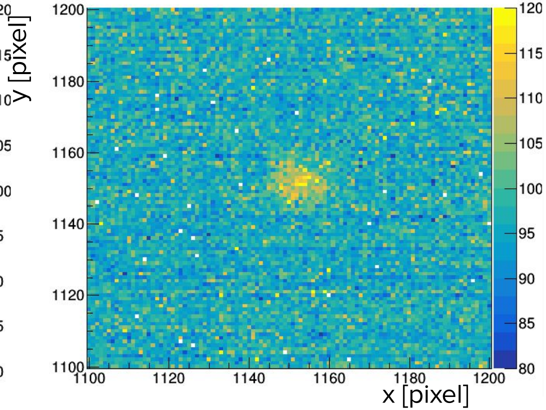
Cross checks of digitization

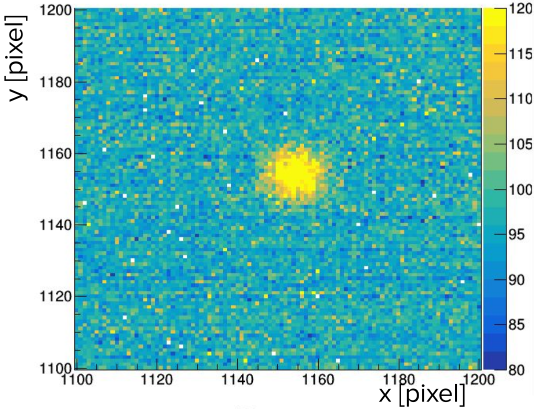
Digitized ER images (no camera noise)

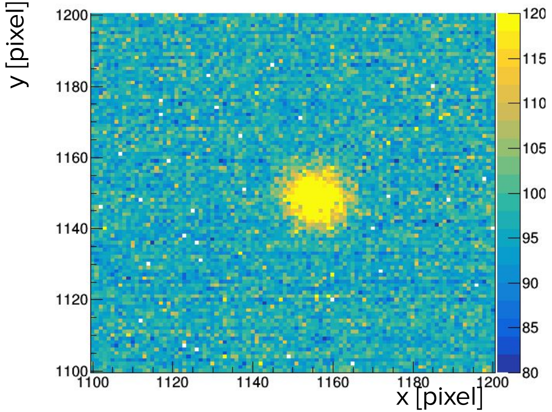


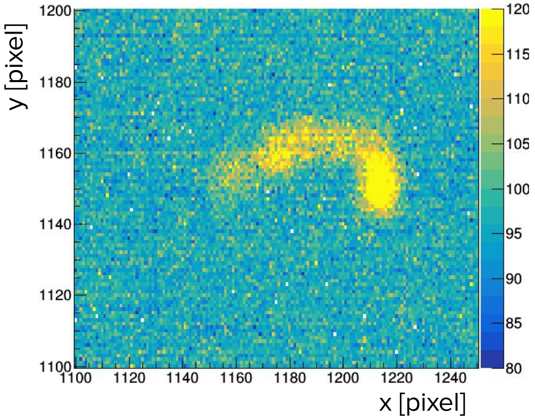
Digitized ER images (with camera noise)





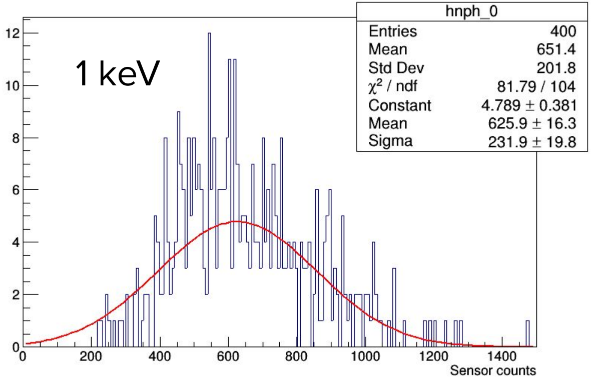


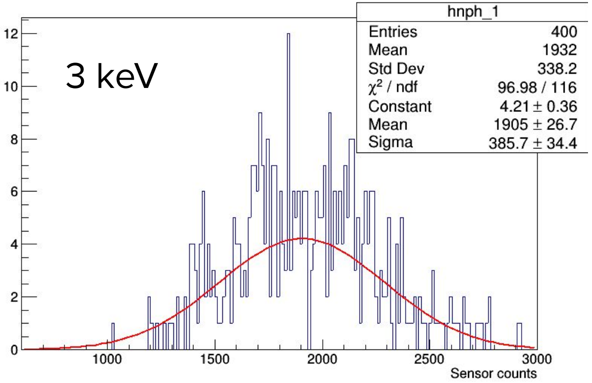


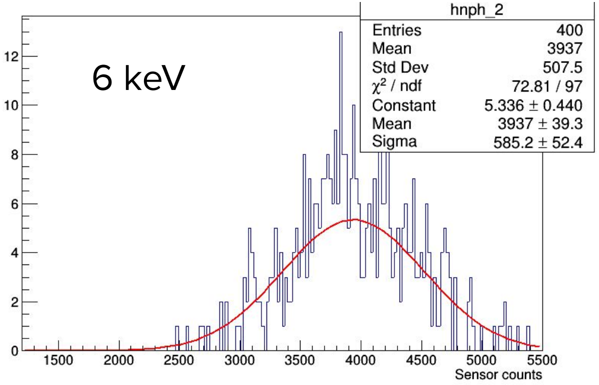


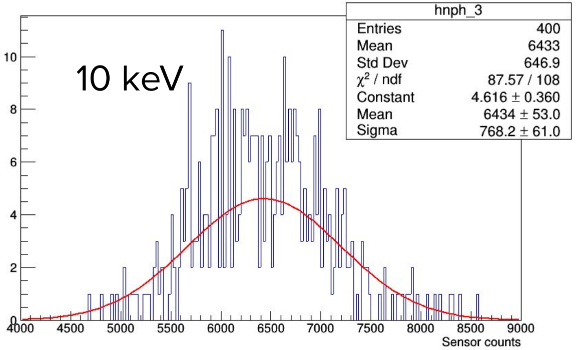
Sensor counts distributions

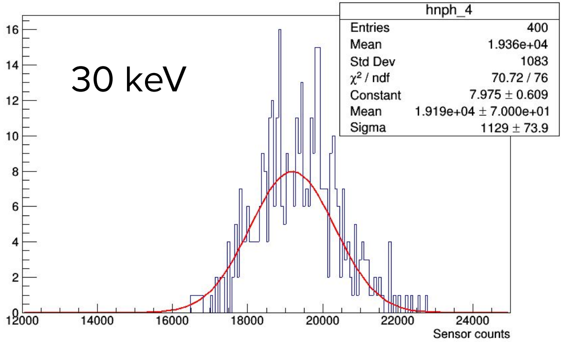
Energy resolution is dominated by gain fluctuations





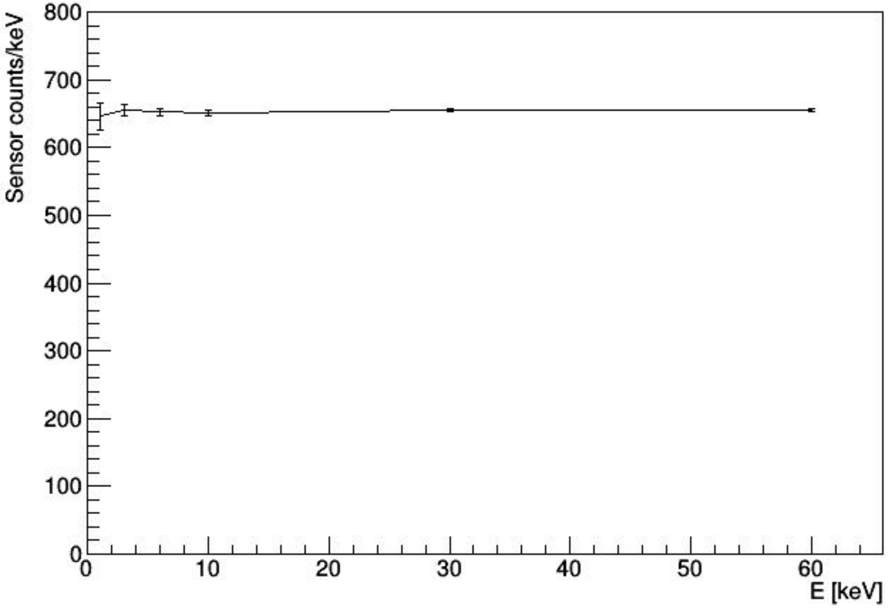


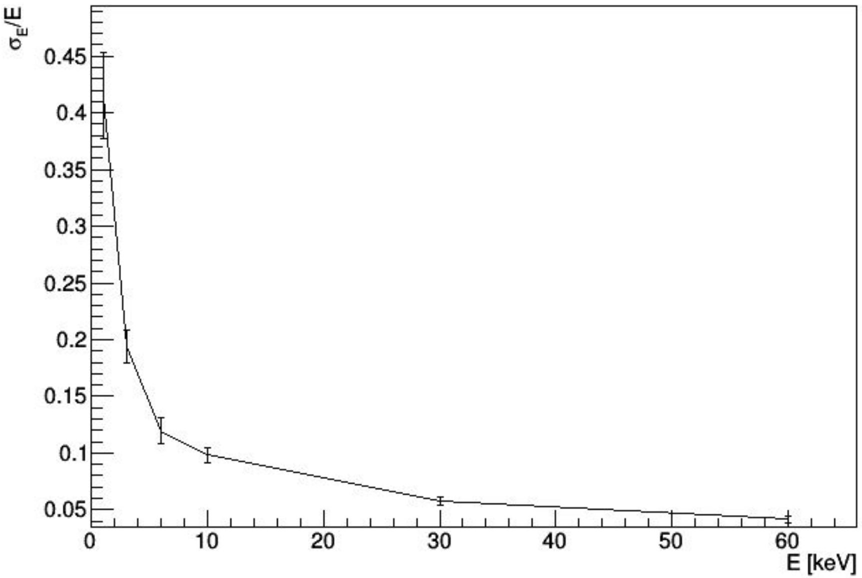




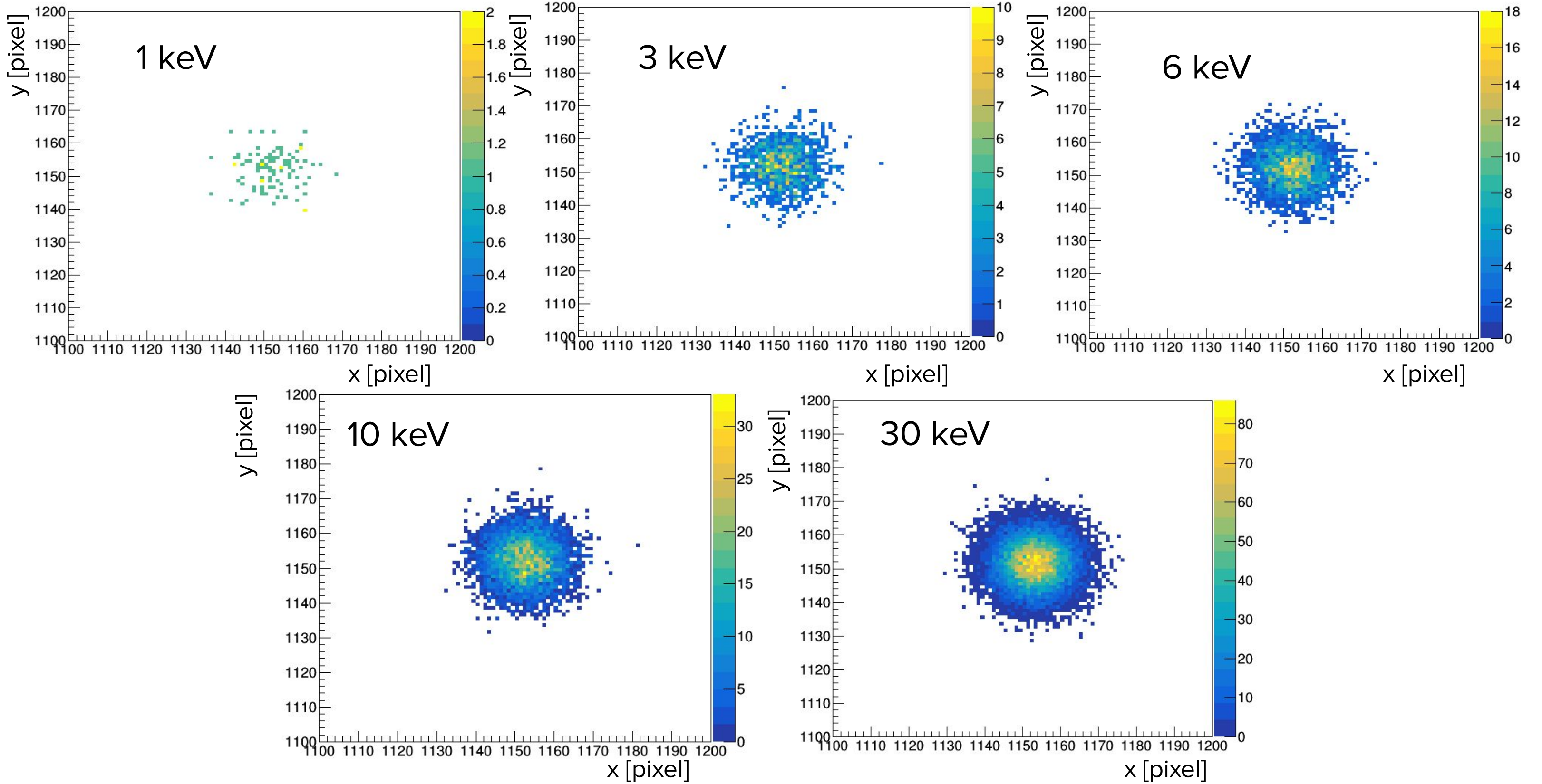
Light yield and energy resolution vs energy

- "Light yield" : average ~650 counts/keV
- Energy resolution ~40% at 1 keV, decreasing to 5% at 30 keV

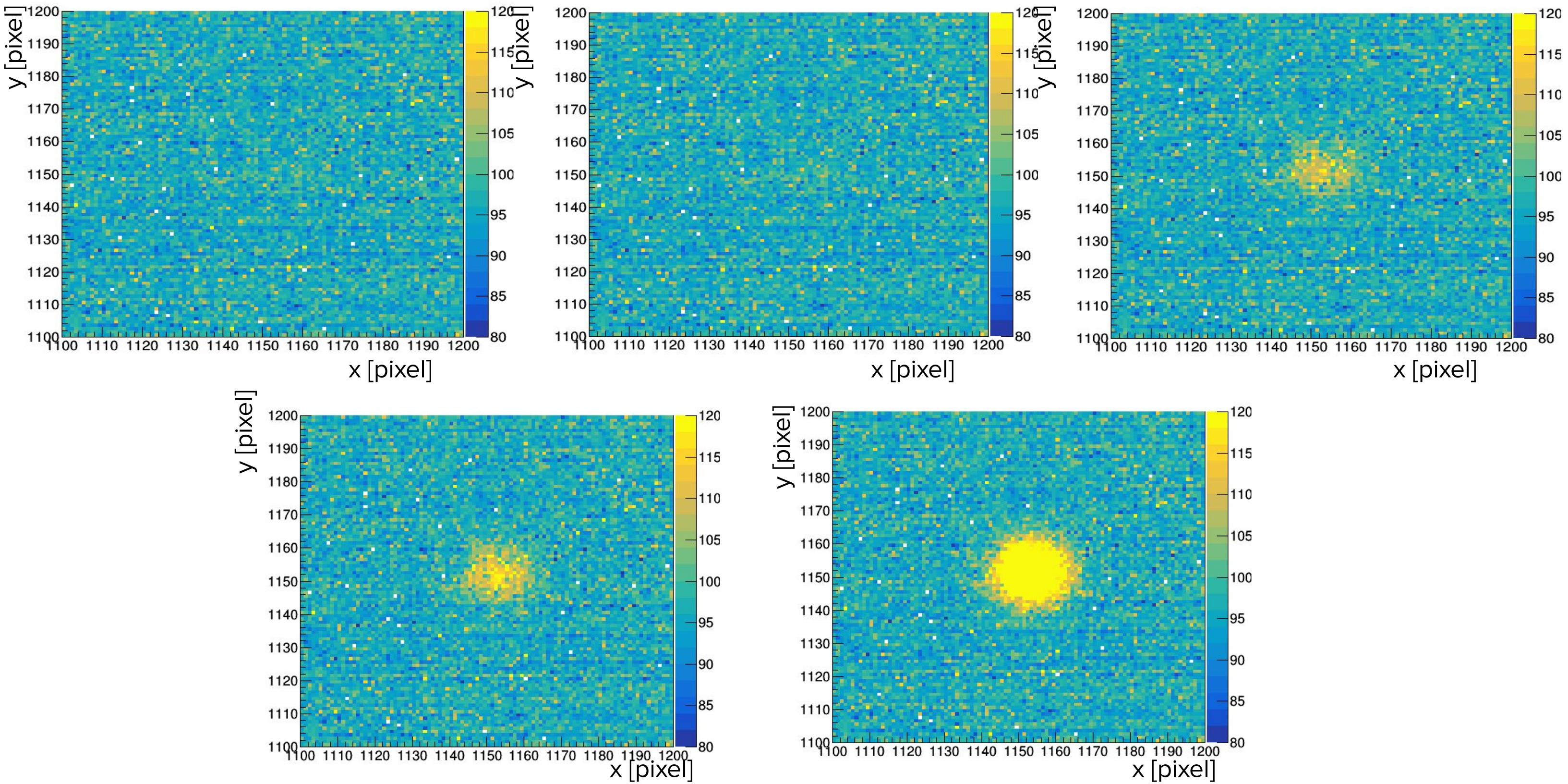




Digitized He NR images (no camera noise)

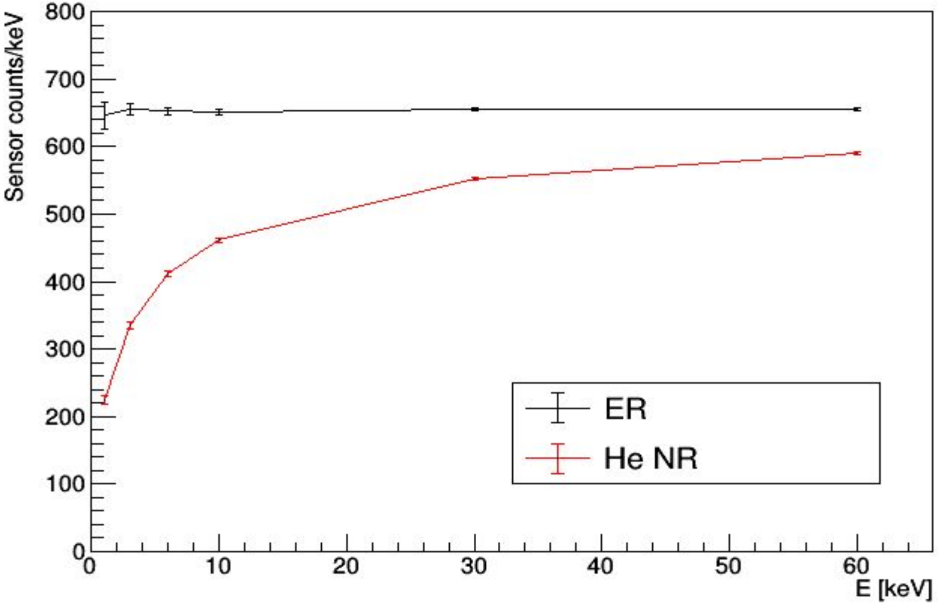


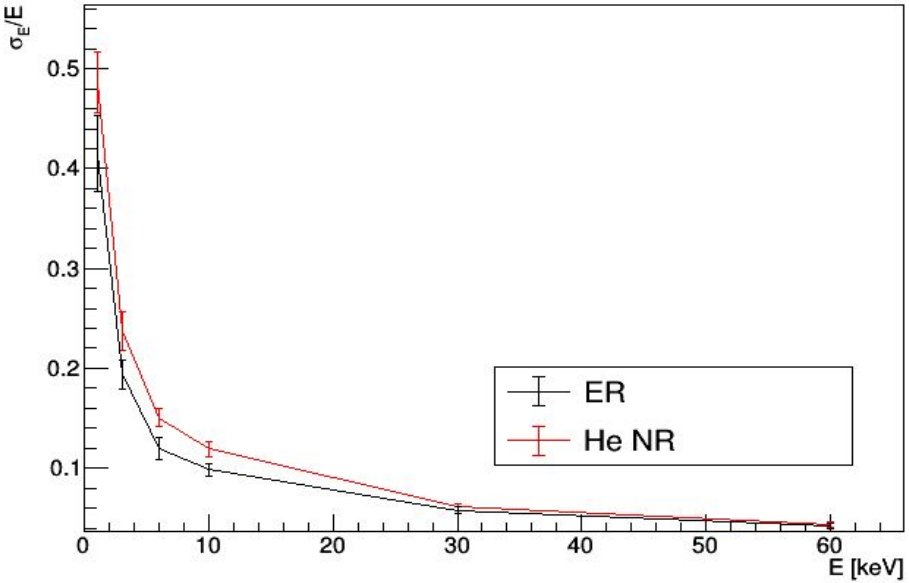
Digitized He NR images (with camera noise)



Light yield and energy resolution for He NR

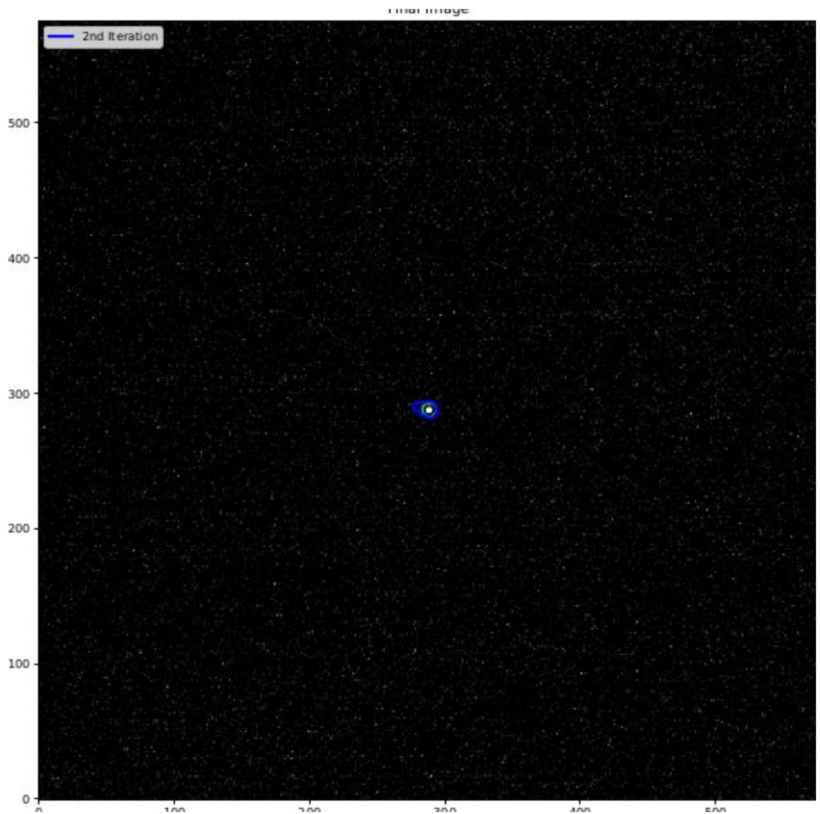
- “Light yield” is lower for NR because of the “quenching factor” (QF): not all energy is converted in light. Note that QF is a function of ion energy ➡ light yield is function of energy
- Energy resolution is higher for the NR due to lower light yield





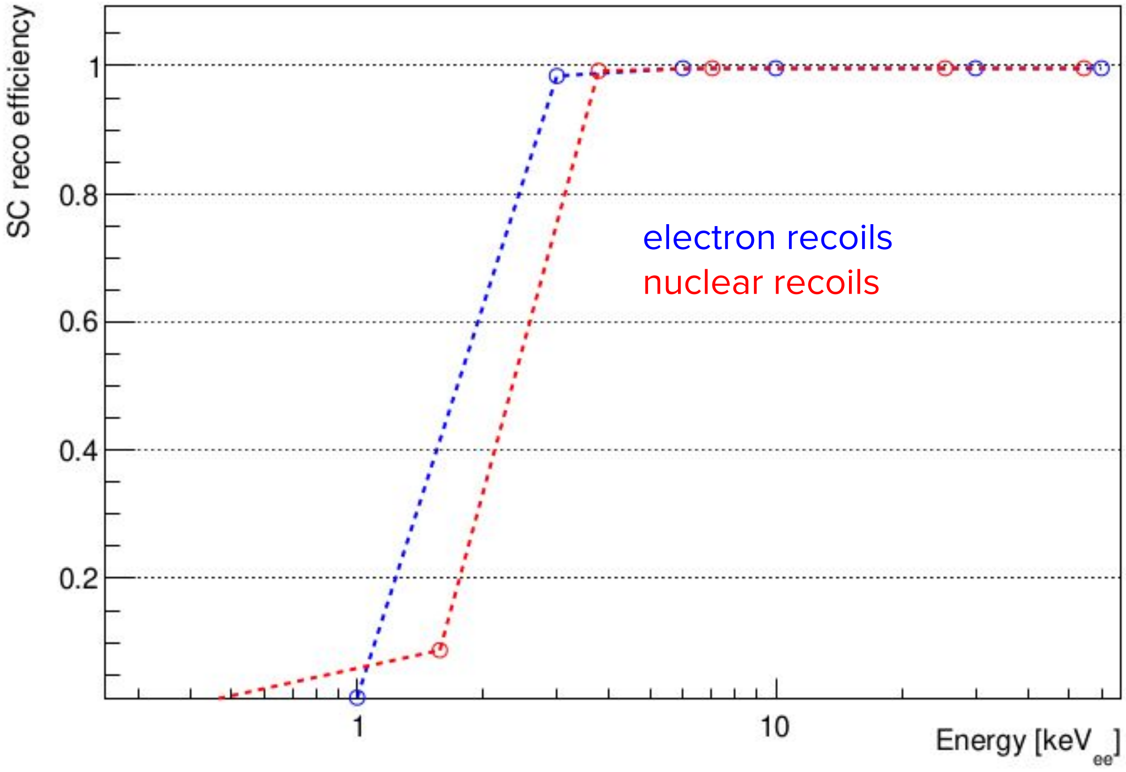
Reconstruction of MC samples

Reco example



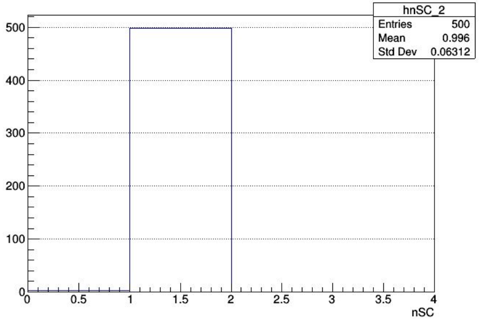
Nuclear recoil of 3 keV

Reconstruction efficiency of SC



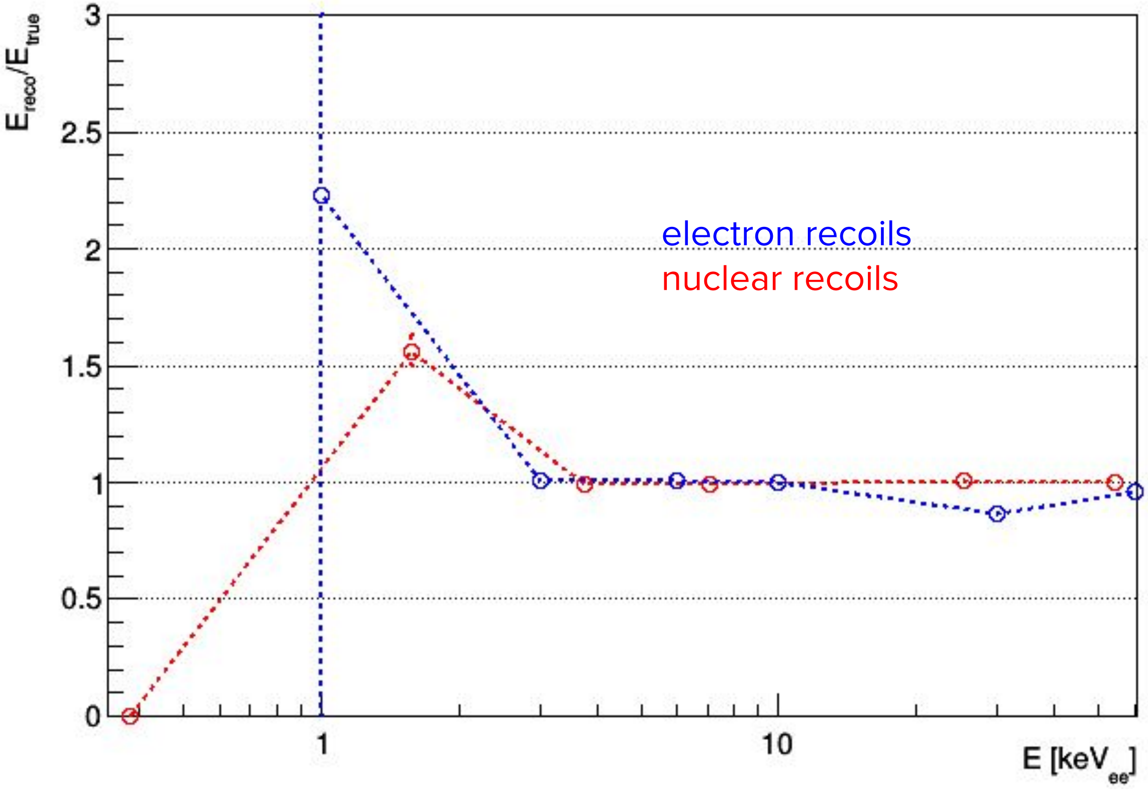
efficiency is defined as
eff = N_SC/N_tot

N_SC are the events with at least 1 reconstructed SC



Note that the comparison is in energy electron-equivalent (keVee)
➜ E [keVee] = E [keV] * QF

Reco sc_energy bias



Note that the comparison is in energy electron-equivalent (keVee)
➔ E [keVee] = E [keV] * QF

CYGNO reconstruction: summary of results

- Test of CYGNO reconstruction performance in the 1-60 keV range, using the samples produced for IDAO
- CYGNO reconstruction has very high efficiency for E> 3 keVee for both ER and NR
- CYGNO reconstruction starts to lose parts of the track for ER of E> 30 keV
- Energy bias in reconstruction is negligible for NR of E > 6 keV (~4 keVee)
- Probably need more samples between 1 and 3 keV(ee) to study better the threshold
- Start to look at variables for ER/NR classification

IDAO exercise

IDAO exercise

- Goal: Classify events in 10 categories: 5 energies x 2 type of particles
- Figure of merit: area of the ROC curve (1 is the “perfect” classification)

  ➜ need to reduce to a two-class problem:

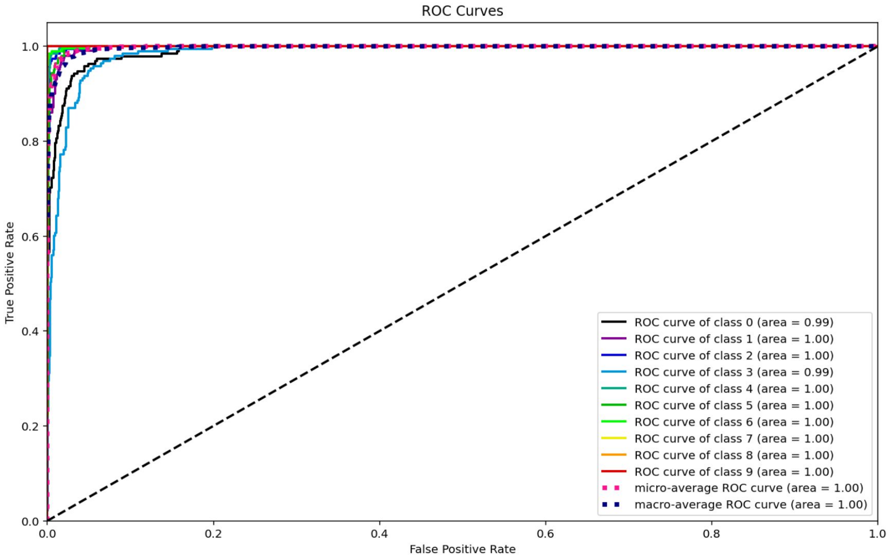
  “true positive rate” vs “false positive rate”

  - True positive rate : rate of events correctly assigned to the class
  - False positive rate : rate of events correctly rejected

IDAO preliminary results

The classes are:

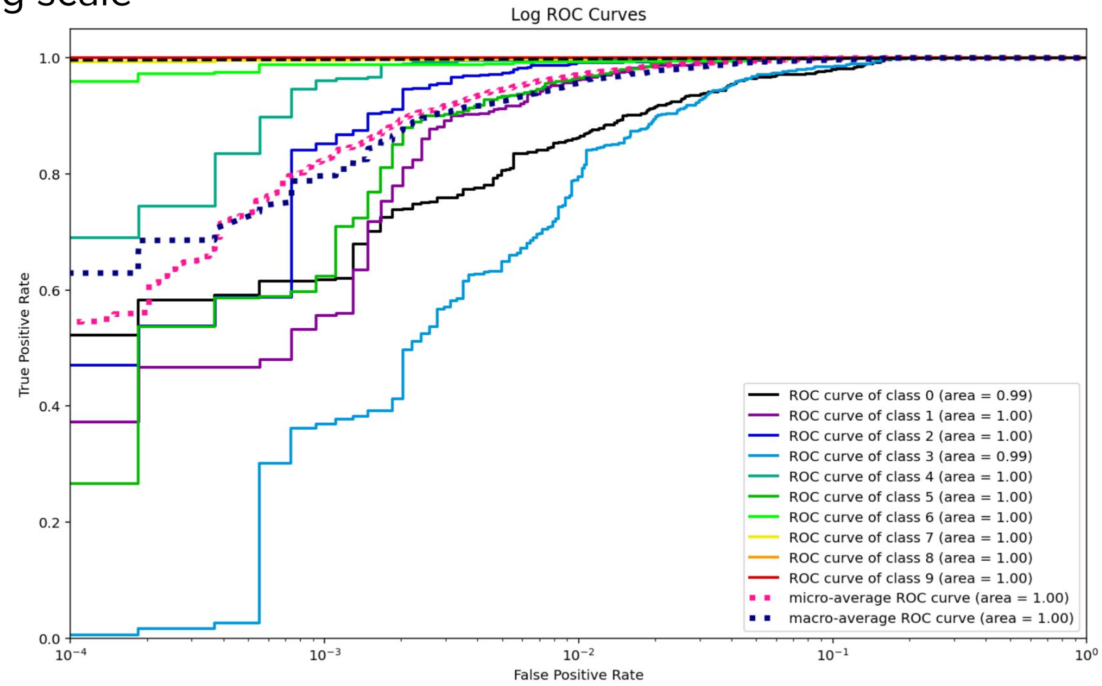
```
'''
CYGNO_60_40_ER_10_keV_930V_30cm_IDAO_crop:                    6,
CYGNO_60_40_ER_10_keV_930V_30cm_IDAO_crop_noBKG:              6,
CYGNO_60_40_ER_1_keV_930V_30cm_IDAO_crop:                     0,
CYGNO_60_40_ER_1_keV_930V_30cm_IDAO_crop_noBKG:               0,
CYGNO_60_40_ER_30_keV_930V_30cm_IDAO_crop:                    8,
CYGNO_60_40_ER_30_keV_930V_30cm_IDAO_crop_noBKG:              8,
CYGNO_60_40_ER_3_keV_930V_30cm_IDAO_crop:                     2,
CYGNO_60_40_ER_3_keV_930V_30cm_IDAO_crop_noBKG:               2,
CYGNO_60_40_ER_6_keV_930V_30cm_IDAO_crop:                     4,
CYGNO_60_40_ER_6_keV_930V_30cm_IDAO_crop_noBKG:               4,
CYGNO_60_40_He_NR_10_keV_930V_30cm_SRIM_IDAO_crop:            7,
CYGNO_60_40_He_NR_10_keV_930V_30cm_SRIM_IDAO_crop_noBKG:  7,
CYGNO_60_40_He_NR_1_keV_930V_30cm_SRIM_IDAO_crop:             1,
CYGNO_60_40_He_NR_1_keV_930V_30cm_SRIM_IDAO_crop_noBKG:   1,
CYGNO_60_40_He_NR_30_keV_930V_30cm_SRIM_IDAO_crop:            9,
CYGNO_60_40_He_NR_30_keV_930V_30cm_SRIM_IDAO_crop_noBKG:  9,
CYGNO_60_40_He_NR_3_keV_930V_30cm_SRIM_IDAO_crop:             3,
CYGNO_60_40_He_NR_3_keV_930V_30cm_SRIM_IDAO_crop_noBKG:   3,
CYGNO_60_40_He_NR_6_keV_930V_30cm_SRIM_IDAO_crop:             5,
CYGNO_60_40_He_NR_6_keV_930V_30cm_SRIM_IDAO_crop_noBKG:       5,
```



All ROC curve areas are very close to 1....

IDAO preliminary results

ROC curve in log scale



All ROC curve areas are very close to 1....

IDAO exercise: summary and discussion

- Algorithm used by IDAO organizers, based on machine learning, looks very powerful for this classification task
- The task could be easy because there is only electronic noise (not crowded images as in data), and just few energy points
- Maybe need to make the game more challenging.. Some possibilities
  - add noise?
  - create new MC samples from a continuum spectrum and add the energy reconstruction to the goals of the exercise
- We will have a discussion at the beginning of next week with Nikita et al. to decide how to proceed
- Can we compare the performance of our reconstruction to their algorithms?