THINK-II Workshop
Uncertainty-Aware Neural Network prediction and its use

Raphaël Bertrand (CPPM, Univ. AMU, CNRS/IN2P3)

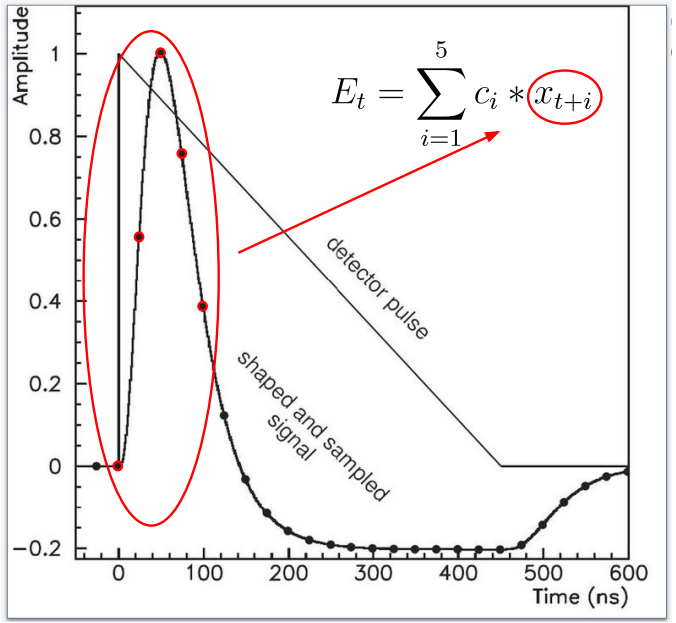






Energy reconstruction with Optimal Filtering algorithm

- **Electronic signal** produced
  - **Amplitude ∝ true deposited energy (E^{true})**
  - **Spans ~625 ns** (25 **B**unch **C**rossings)
  - **Shaped, sampled** and **digitised at 40 MHz**
    - ➢ **Bipolar ⇒ in-time pileup mitigation with out-of-time pileup**
    - ➢ Sampling designed to reach **max at 3rd sample** ideally

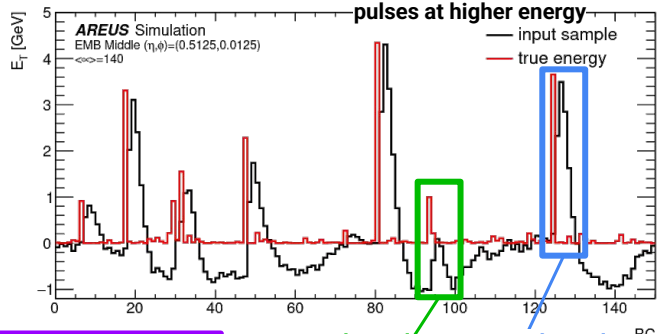


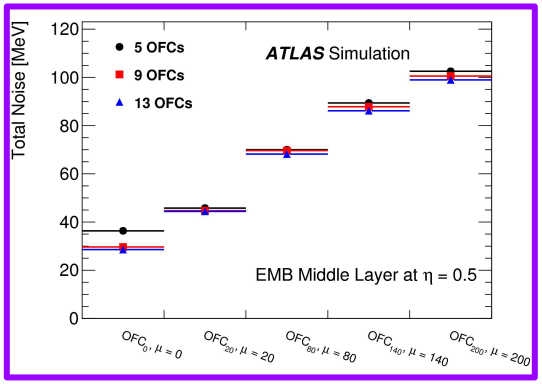
- **OF** algorithm → Energy reconstruction
  - **Unbiased** energy prediction with respect to the HS energy
  - **Weighted sum** of samples around the pulse peak
    - ➢ 4 to 5 samples used
  - **Max finder/Timing cut** to select the correct BC
    - ➢ OF also computes the timing shift

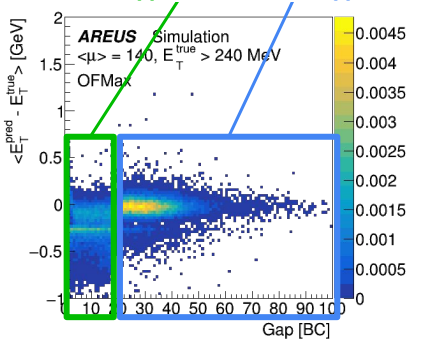
Limitations of the OF algorithm for HL-LHC

- HL-LHC ⇒ **Increased luminosity and pileup**
  - **Higher rate of overlapping pulses**
  - Increased pileup noise
- Limitations of Optimal Filtering
  - **Degraded energy resolution with overlapping pulses**
  - Increasing the number of samples (OFCs) gives limited improvement

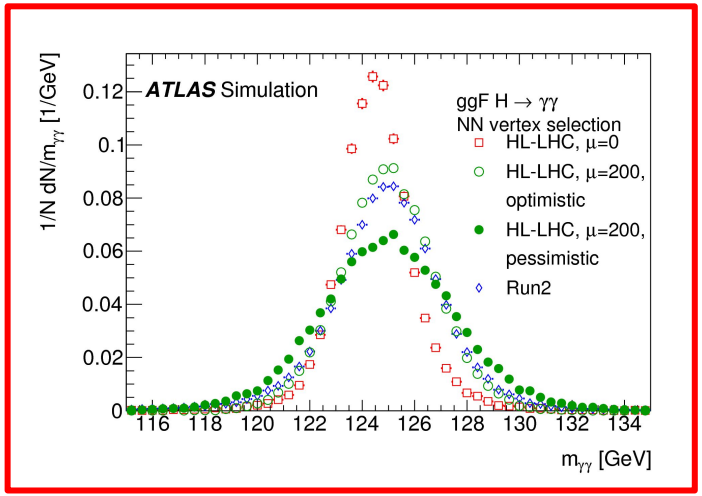
Simulated pileup with additional injected pulses at higher energy







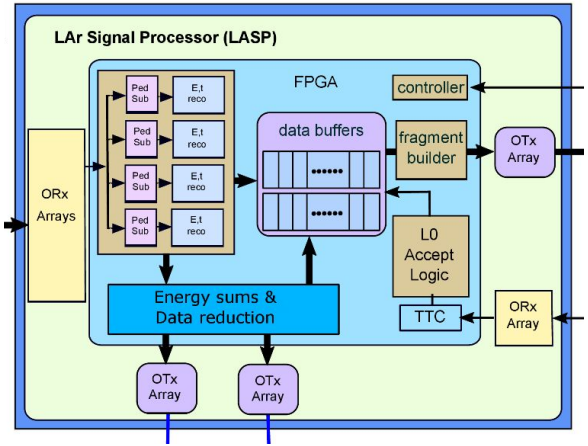
LAr Phase-II TDR



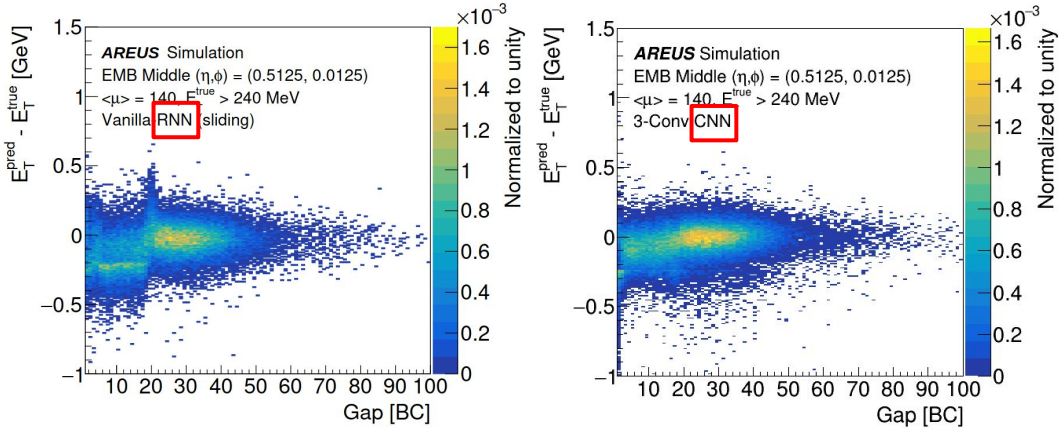
- Impact on **physics performance**
  - **Degraded energy resolution**
    - ➢ **Possible improvement** to obtain performance comparable to Run 2 (Optimistic)
    - ➢ No change (Pessimistic)

Hardware requirements for new reconstruction methods

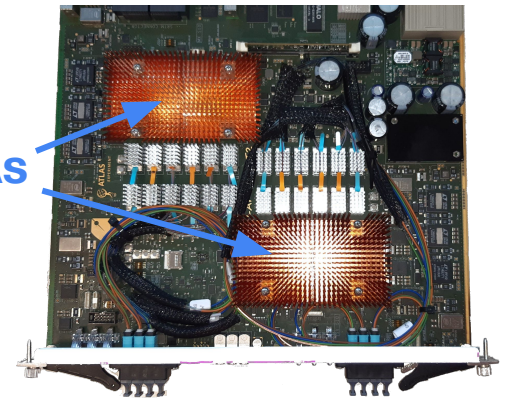
- Off-detector readout board (LASP) will carry out two **FPGAs for energy computation**
  - Opportunity to **embark more complex reconstruction algorithms**
- Reconstruction algorithm requirements :
  - **Online** computation (per BC)
  - **Max latency : ~125 ns** (used in trigger system)
  - **Fit in FPGAs** : **O(500)** Multiply-Accumulate operations (**MAC units**)
    - ➢ 4 to 5 MAC units required to implement OF
  - **384 channels per FPGA** (many algorithm instances needed)
- Neural networks preliminary studies
  - CNN and small RNN tested and implemented



[Comput.Softw.Big Sci. 5 (2021), s41781-021-00066-y](#)



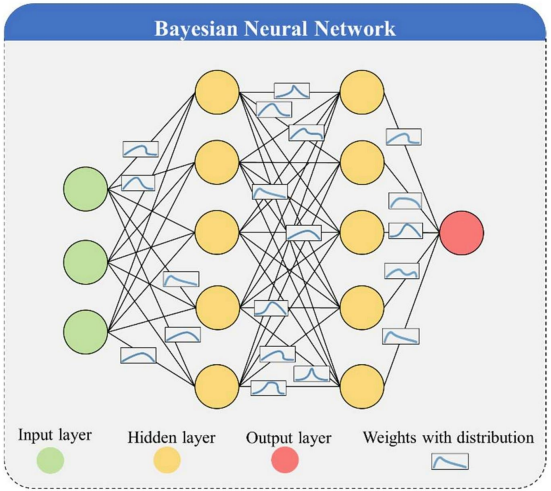
LASP board demonstrator



The Need for Uncertainty Prediction

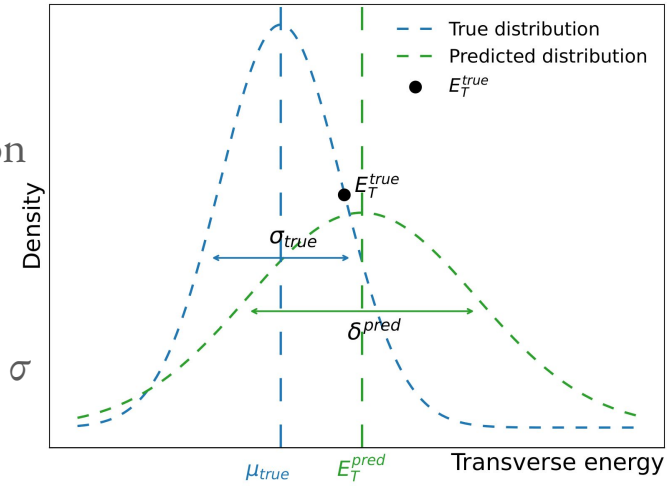
- **From average to event-by-event uncertainty** : current LAr calorimeter uncertainty estimates derived from large samples of events
  - ➢ **Goal** : estimate the uncertainty of each individual event
- **Adapt to changing LHC conditions** : accounts for variation in instantaneous luminosity and bunch-train structure
  - ➢ Potential better signal-to-noise estimation and improvement of data quality

- **Common approach – Bayesian Neural Networks (BNNs)** : **assign prior distributions to network weights**, which induce a predictive distribution over the output
  - ➢ Predictive distribution generally intractable ⟹ Require expensive sampling methods
  - ➢ **FPGA deployment incompatible** with the latency and resource constraints



Alternative Methods

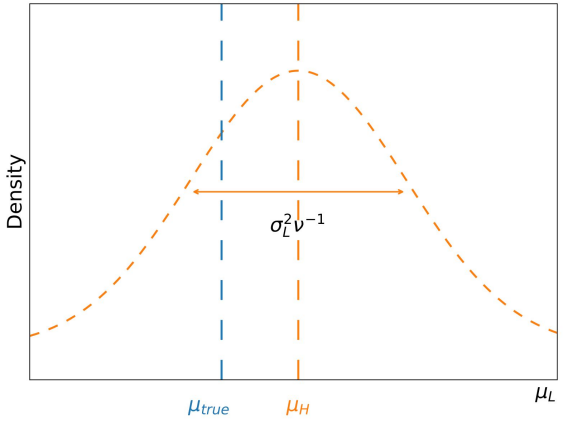
- **Direct uncertainty prediction** : train the neural network to predict both the energy and its associated uncertainty
  - ➢ **Gaussian parametrisation** : model the predictive distribution as a Normal distribution : $\mathcal{N}(\mu, \sigma^2)$
    - μ : predicted mean energy
    - σ : predicted uncertainty
    - For a calibrated Gaussian, ~68% of true values lie within $\mu \pm \sigma$
- **Two approaches** :
  - ➢ **Deep Evidential Regression (DER)** : predict a higher-order distribution over μ and σ
    - Assume the true energy is drawn from a Normal distribution
    - Separation between statistic and random uncertainties
  - ➢ **Direct Probabilistic Regression (DPR)** : predict μ and σ directly

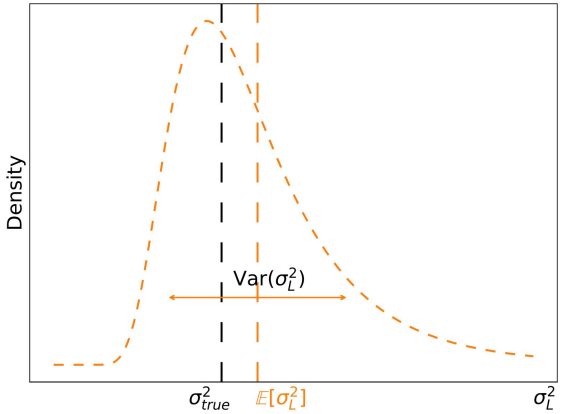


Deep Evidential Regression

- **Underlying energy distribution** : the true energy is modelled as $\mathcal{N}(\mu_{true}, \sigma^2_{true})$
  - ➢ $\mu_{true}, \sigma^2_{true}$ are unknown, network aim to predict estimators on them : μ_L, σ^2_L
- Estimators are considered as random variables :
  - ➢ $\mu_L \mid \sigma^2_L \sim \mathcal{N}(\mu_H, \sigma^2_L \nu^{-1})$
  - ➢ $\sigma^2_L \sim \Gamma^{-1}(\alpha, \beta)$
- **Four predicted parameters (μ_H, ν, α, β)** to describe predicted mean, error on mean, and uncertainty
  - ➢ Possible to implement in FPGA
- **Two sources of uncertainty**, added in quadrature :
  - ➢ **Aleatoric** (random) : intrinsic variability of the energy measurement, $\mathbb{E}[\sigma^2_L]$
  - ➢ **Epistemic** (systematic) : uncertainty in the estimated mean due to limited evidence, $Var(\mu_L)$
- $E_{pred} = \mu_H,$

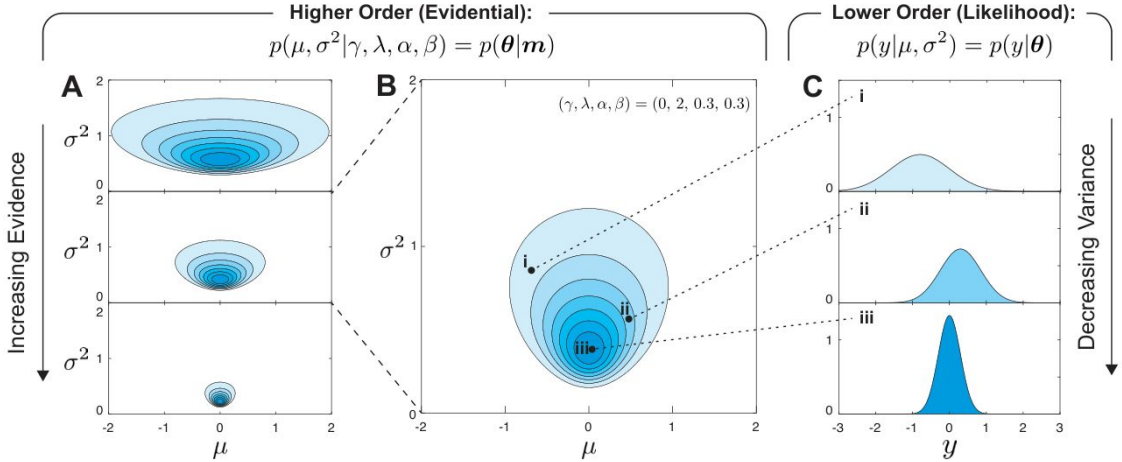
$$\delta^2_{pred} = \frac{\beta(1+\nu^{-1})}{\alpha - 1}$$





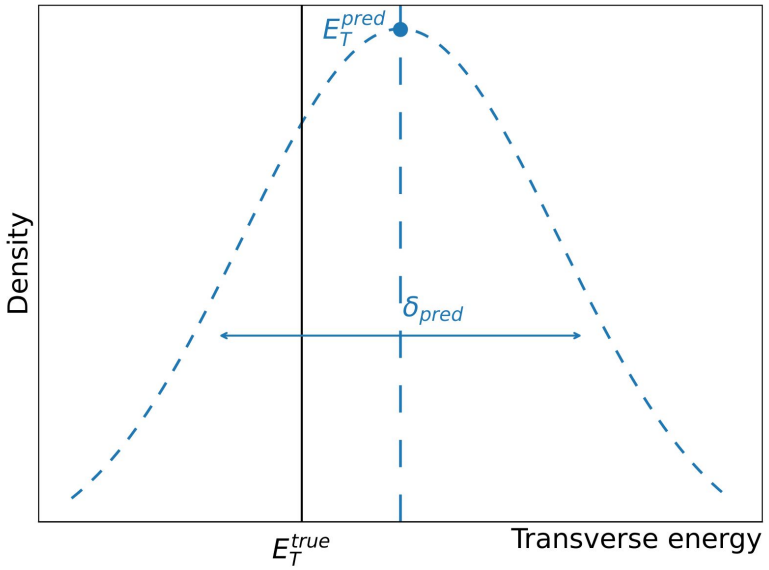
Training the DER model

- **Training objective** : learn the parameters of the high-level distribution so that the corresponding low-level Gaussian distributions assign high likelihood to the true energy
- **Marginal likelihood** : Account for all possible values of μ_L and $\sigma^2_L(1+\nu^{-1})$ by integrating them out
  - ➢ Account for probability for each parameters configuration
- **Analytic solution** of the marginal likelihood : a Student-t distribution
  - ➢ No sampling techniques required



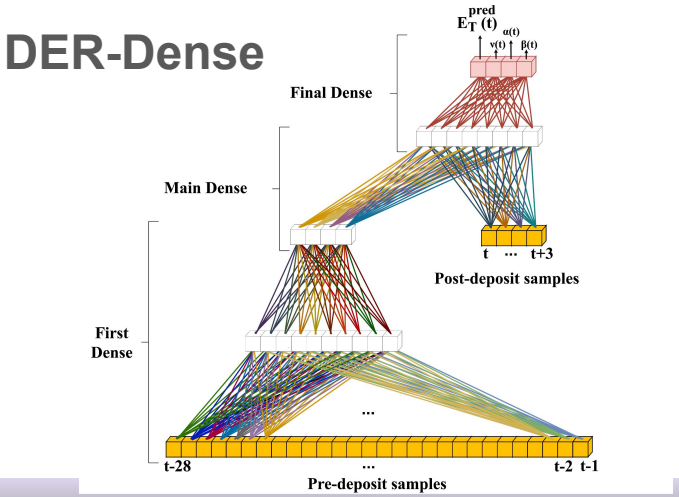
Direct Probabilistic Regression

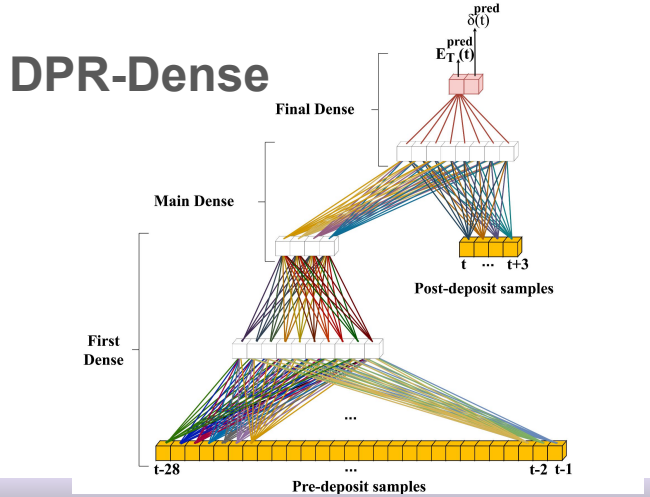
- **Probabilistic energy prediction** : the true energy is single value, while the model predicts a **Gaussian distribution centered on the predicted energy**, with the **uncertainty quantifying the expected deviation** between the predicted and true energies
- **Two deterministic outputs** : μ_{pred} and δ_{pred}
  - Possible to implement in FPGA
- **Uncertainty estimation** : δ_{pred} captures the total predictive uncertainty, without explicitly separating aleatoric and epistemic components
- **Training objective** : maximise likelihood of the Gaussian



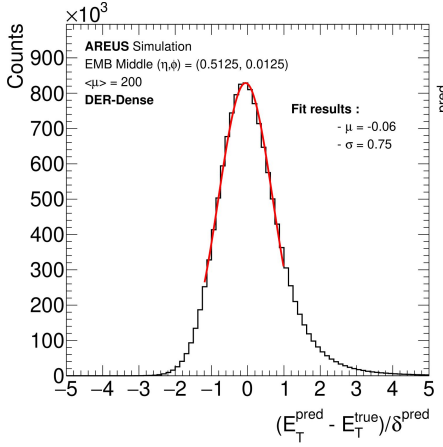
DER and DPR architectures

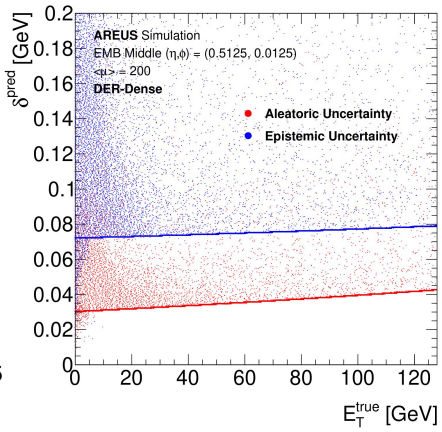
- **Simple network modification** : add additional output neurons to the final dense layer to predict the required parameters
- **DER-Dense output layer** : four outputs (μ_H,ν,α,β)
  - ➢ Constraints : $\nu > 1$, $\alpha > 0$, $\beta > 0 \Longrightarrow$ use of softplus as activation function
- **DPR-Dense output layer** : two outputs (μ_H, δ_{pred})
  - ➢ Constraint : $\delta_{pred} > 0 \Longrightarrow$ use of ReLU as activation function

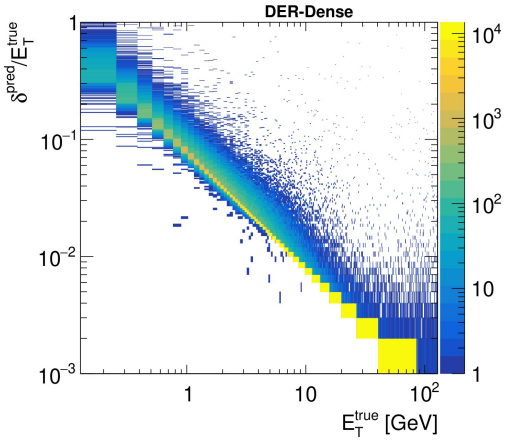


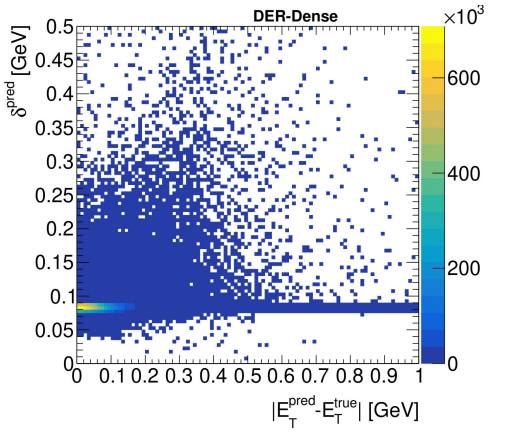


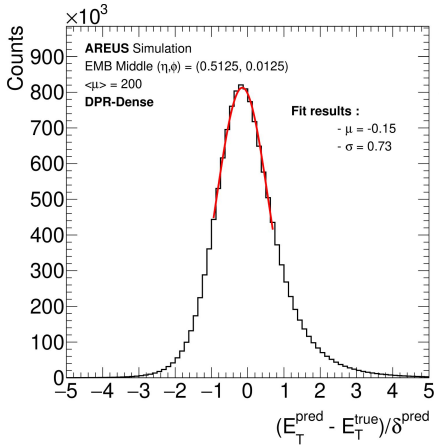
Uncertainty-Aware Models for Energy reconstruction

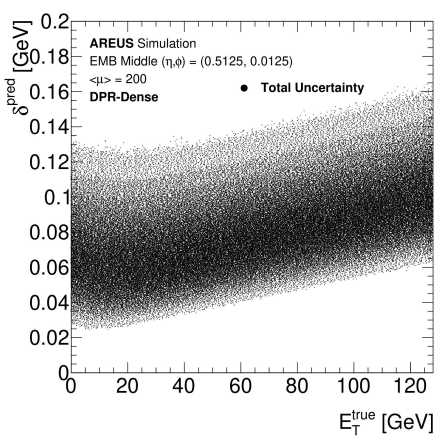


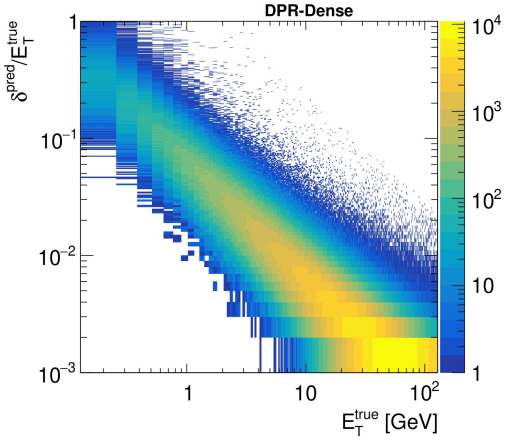


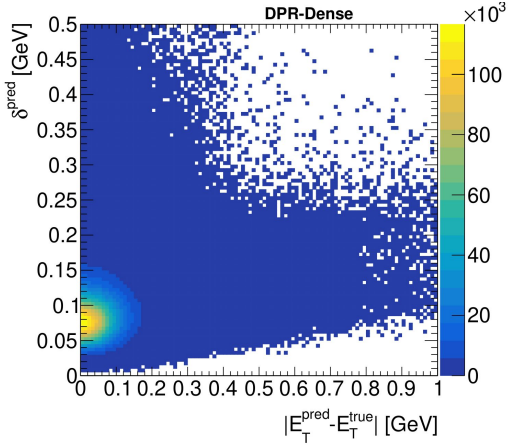






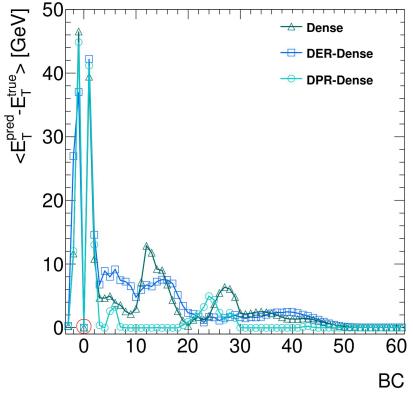


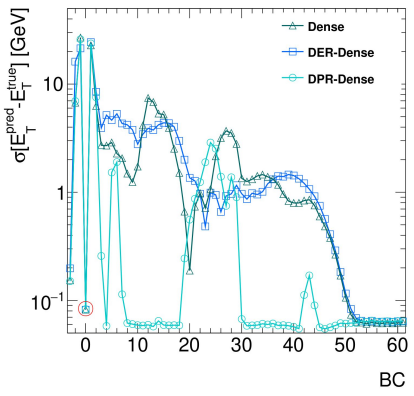


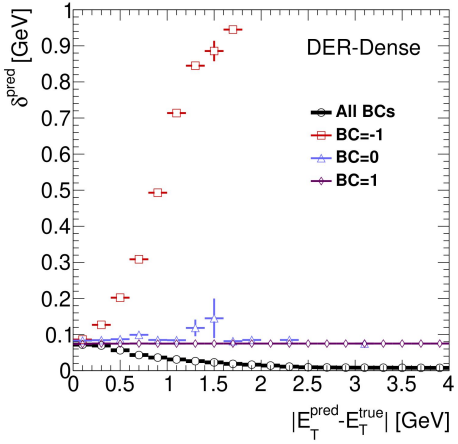


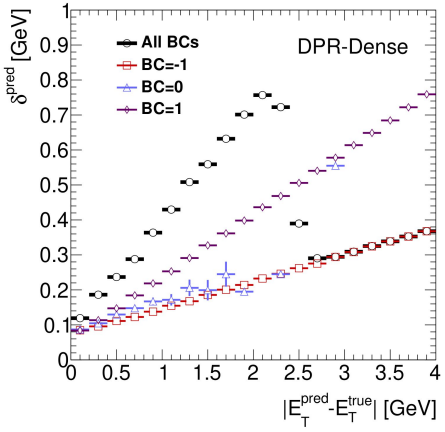
Exploiting the Uncertainty for Triggering

- **Current performance evaluation** : use of **hard-scattering (HS)** events
  - ➢ Perfect triggering assumption
- **Online operation** : energy reconstruction must be performed for every bunch crossings (BC)
- **Fake energy prediction** (using OF or networks) :
  - ➢ Large energy deposits can affect input samples across multiple BCs
  - ➢ How to associate an energy deposit with the correct BC ?

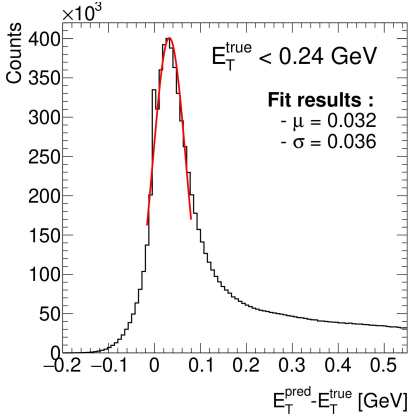


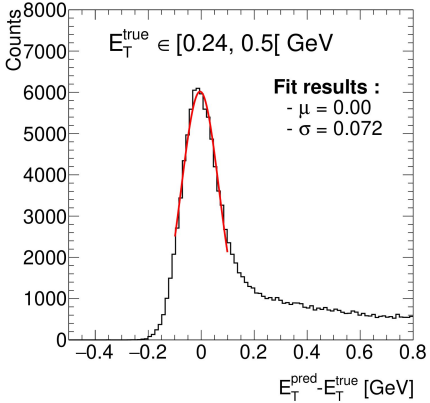


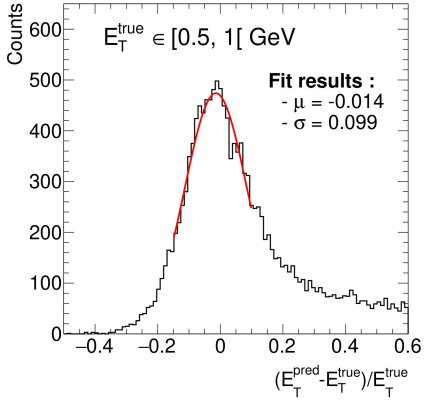


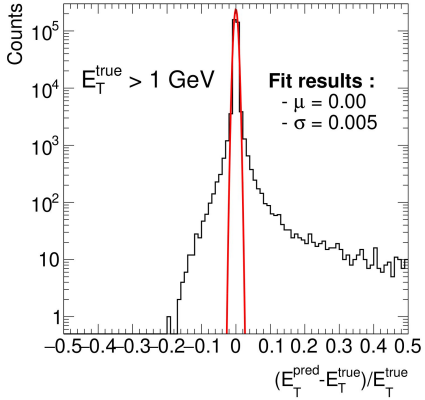


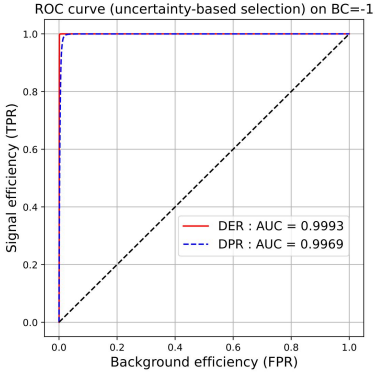
Triggering Performance

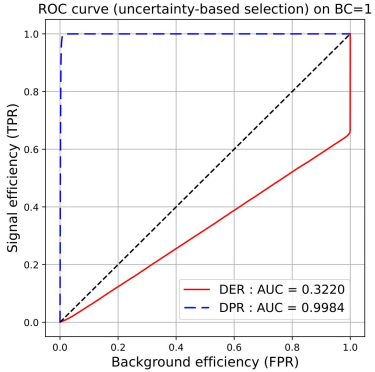




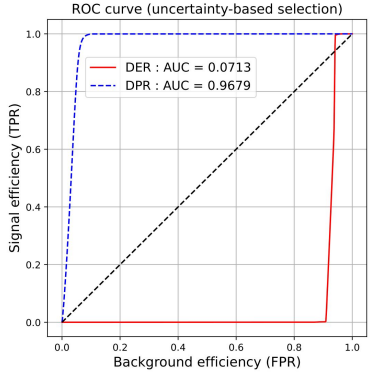












BC before

BC after

All BCs