Towards Realistic ePIC Simulations: Background Mixing, Digitization and Timeframe building

Sakib Rahman (BNL)
ePIC Workshop on Streaming Reconstruction
19 August 2026

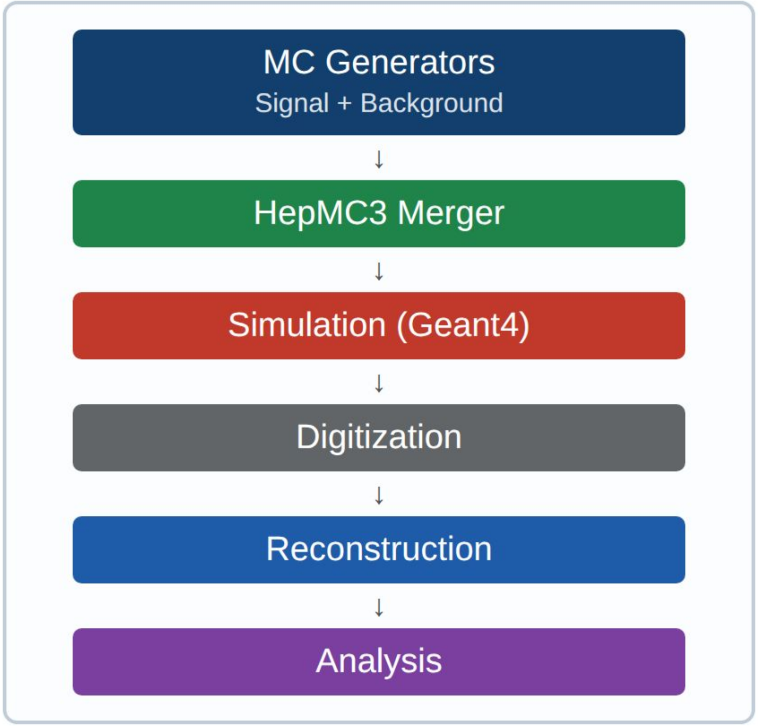
Introduction

For streaming reconstruction development, ePIC simulations must deliver outputs that look like realistic streamed data: timeframes mixing physics and backgrounds + 10Gbit/s detector noise

Bkg-Mixed Simulations
Sources & rates, HepMC3 merger, how to find campaign outputs and understand them

Digitization Including Noise Injection
Simulated data => Realistic detector response

Timeframe builder
Add timing structure to background mixed simulation outputs. Move “merging” downstream of “simulation”.



Current Production Workflow

Background Mixed Simulations: HepMC3 Merger

`1 Reco Event = 2`μ`s integration frame`
Exactly 1 Physics Event + Background Events
Sampled according to Respective Frequencies

Integration frame length defined by the readout time window of the MAPS

Event	signal	synrad	ebrems	etouschek	ecoloumb	p.b.gas
Event 1	●	●●●●	●			
Event 2	●	●●●●			●	●
Event 3	●	●●●		●		
Event 4	●	●●●	●			●
Event 5	●	●●●●	●			

Exactly 1 Signal Per Integration Frame Config

Background Mixed Simulations: Assumptions and Source Data

Assumptions
10000 Ahr vacuum and 10μm gold coating on central beam pipe

Frequencies and status codes tracked in github repo for 10x100 and 10x275 beam energy configurations. Being migrated to epic-devcloud.org.

Sampled background event rates & HepMC status-code offsets — 10×275 GeV

Process	Rate	Offset
Synchrotron radiation	**13.28 GHz**	2000
e− beam gas — bremsstrahlung	3.18 MHz	3000
e− beam gas — Coulomb	29.6 kHz	4000
e− beam gas — Touschek	233.5 kHz	5000
Proton beam gas	32.6 kHz	6000
Signal (e.g. DIS)	**~500 kHz**	**0**

Will update all background source files once new information is available from project

Background Mixed Simulations: Latest Campaign Outputs

Full and reco files for individual background sources

`rucio did list --short epic:*26.07.1* | grep BACK | sort`

```
epic:/FULL/26.07.1/epic_craterlake/BACKGROUNDS/BEAMGAS/electron/coulomb/EIC_ESR_Xsuite/dataprod_rel_1.0.1/10x275/10000Ahr/MachineRuntime50s
epic:/FULL/26.07.1/epic_craterlake/BACKGROUNDS/BEAMGAS/electron/coulomb/EIC_ESR_Xsuite/dataprod_rel_1.0.1/18x275/10000Ahr/MachineRuntime50s
epic:/FULL/26.07.1/epic_craterlake/BACKGROUNDS/BEAMGAS/electron/coulomb/EIC_ESR_Xsuite/dataprod_rel_1.0.1/5x100/10000Ahr/MachineRuntime50s
epic:/RECO/26.07.1/epic_craterlake/BACKGROUNDS/BEAMGAS/electron/coulomb/EIC_ESR_Xsuite/dataprod_rel_1.0.1/10x275/10000Ahr/MachineRuntime50s
epic:/RECO/26.07.1/epic_craterlake/BACKGROUNDS/BEAMGAS/electron/coulomb/EIC_ESR_Xsuite/dataprod_rel_1.0.1/18x275/10000Ahr/MachineRuntime50s
epic:/RECO/26.07.1/epic_craterlake/BACKGROUNDS/BEAMGAS/electron/coulomb/EIC_ESR_Xsuite/dataprod_rel_1.0.1/5x100/10000Ahr/MachineRuntime50s
```

Full and reco files for ~50k events for each background mixed configuration

`rucio did list --short epic:*26.07.1* | grep Bkg | sort`

```
epic:/FULL/26.07.1/epic_craterlake/Bkg_Exact1S_2us/GoldCt/10um/DIS/NC/10x100/minQ2=1
epic:/FULL/26.07.1/epic_craterlake/Bkg_Exact1S_2us/GoldCt/10um/DIS/NC/10x100/minQ2=10
epic:/FULL/26.07.1/epic_craterlake/Bkg_Exact1S_2us/GoldCt/10um/DIS/NC/10x100/minQ2=100
epic:/FULL/26.07.1/epic_craterlake/Bkg_Exact1S_2us/GoldCt/10um/DIS/NC/10x100/minQ2=1000
epic:/FULL/26.07.1/epic_craterlake/Bkg_Exact1S_2us/GoldCt/10um/DIS/NC/10x275/minQ2=1
epic:/FULL/26.07.1/epic_craterlake/Bkg_Exact1S_2us/GoldCt/10um/DIS/NC/10x275/minQ2=10
epic:/FULL/26.07.1/epic_craterlake/Bkg_Exact1S_2us/GoldCt/10um/DIS/NC/10x275/minQ2=100
epic:/FULL/26.07.1/epic_craterlake/Bkg_Exact1S_2us/GoldCt/10um/DIS/NC/10x275/minQ2=1000
epic:/FULL/26.07.1/epic_craterlake/Bkg_Exact1S_2us/GoldCt/10um/SIDIS/pythia6-eic/1.0.0/10x100/q2_0to1
epic:/FULL/26.07.1/epic_craterlake/Bkg_Exact1S_2us/GoldCt/10um/SIDIS/pythia6-eic/1.0.0/10x275/q2_0to1
epic:/RECO/26.07.1/epic_craterlake/Bkg_Exact1S_2us/GoldCt/10um/DIS/NC/10x100/minQ2=1
epic:/RECO/26.07.1/epic_craterlake/Bkg_Exact1S_2us/GoldCt/10um/DIS/NC/10x100/minQ2=10
epic:/RECO/26.07.1/epic_craterlake/Bkg_Exact1S_2us/GoldCt/10um/DIS/NC/10x100/minQ2=100
epic:/RECO/26.07.1/epic_craterlake/Bkg_Exact1S_2us/GoldCt/10um/DIS/NC/10x100/minQ2=1000
epic:/RECO/26.07.1/epic_craterlake/Bkg_Exact1S_2us/GoldCt/10um/DIS/NC/10x275/minQ2=1
epic:/RECO/26.07.1/epic_craterlake/Bkg_Exact1S_2us/GoldCt/10um/DIS/NC/10x275/minQ2=10
epic:/RECO/26.07.1/epic_craterlake/Bkg_Exact1S_2us/GoldCt/10um/DIS/NC/10x275/minQ2=100
epic:/RECO/26.07.1/epic_craterlake/Bkg_Exact1S_2us/GoldCt/10um/DIS/NC/10x275/minQ2=1000
epic:/RECO/26.07.1/epic_craterlake/Bkg_Exact1S_2us/GoldCt/10um/SIDIS/pythia6-eic/1.0.0/10x100/q2_0to1
epic:/RECO/26.07.1/epic_craterlake/Bkg_Exact1S_2us/GoldCt/10um/SIDIS/pythia6-eic/1.0.0/10x275/q2_0to1
```

Background Mixed Simulations: Limitation Addressed in 26.07.1

Issue

Any background source (e.g. synchrotron radiation) provides N simulated events on disk, far fewer than what a run consisting of many 2 μs integration frames consumes. Made worse by the fact that we were resetting the starting point for sampling to event 0 in each parallel job under a task.

Each job computes a large random skip

`skip = rate term + random term`

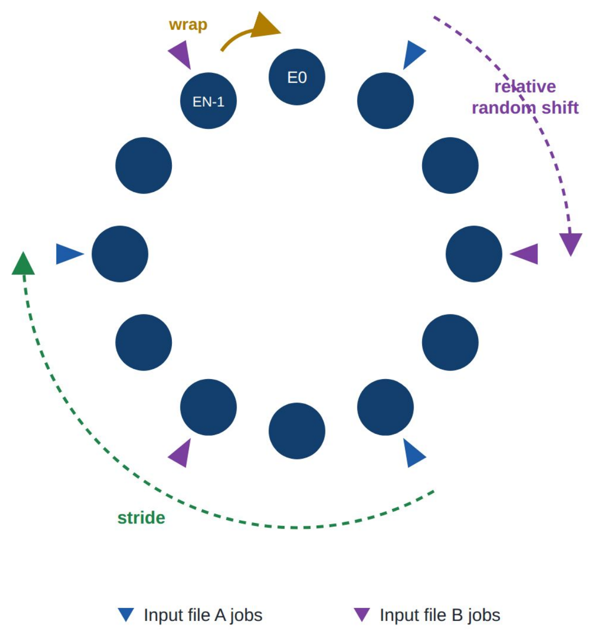
`rate term = int(SKIP_N_EVENTS×job id)+1`

keeps bkg consumption ∝ signal events processed

`random term =  int(rand()×2147483647)`

seeded per task and per signal-file hash to decorrelates parallel jobs

Circular buffer solution: parallel jobs enter at decorrelated random offsets and cycle indefinitely



Background Mixed Simulations: Outstanding Issues

Event cycling only resolves part of the issue since the background source pools are still finite and generation of certain backgrounds is very resource intensive

Pre-produced proton beamgas samples had reused seeds, resulting in repetition within source files

Need to centralize background events production within ePIC Production System

Gabor's talk demonstrating the feasibility of reproducing Andrii's synchrotron background generation framework results on GPU using Simphony, a fork of NVIDIA OptiX-based Opticks framework developed by Simon Blyth

Torre's talk demonstrating possible PanDA integration of Simphony workflows

GPU Offload Path

?

Generative AI Path

Background Mixed Simulations: Understanding Outputs

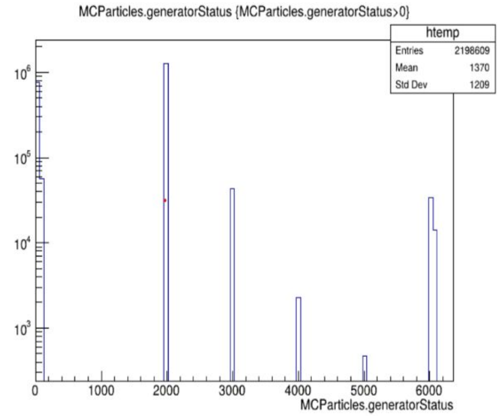
HepMC status codes of background particles are offset by background type

For example: A stable synchrotron photon will have status code

1 (stable) + 2000 (synchrotron) = 2001

Example: Minimal ROOT script to view counts of processes by status codes

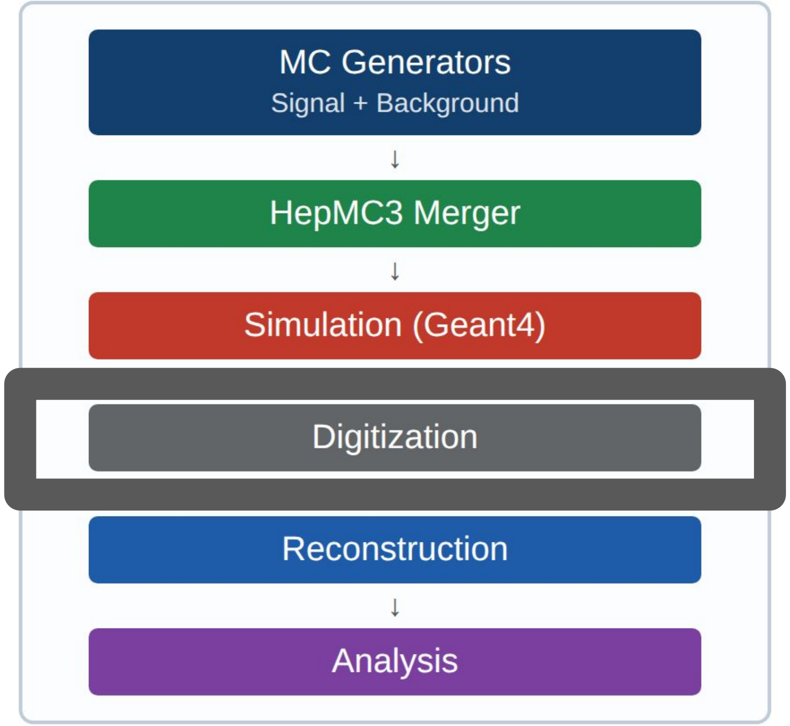
```
TChain *T=new TChain("events");
T->Add("root://dtn-eic.jlab.org//volatile/eic/EPIC/RECO/25.06.1/epic_craterlake/Bkg_1SignalPer2usFrame/Sy
T->Draw("MCParticles.generatorStatus", "MCParticles.generatorStatus>0");
gPad->SetLogy();
```



DD4hep / HepMC generatorStatus codes

Code	Meaning
0	Not from the generator — e.g. a **Geant4 secondary** created during simulation (provenance in *simulatorStatus*)
1	**Stable**, undecayed final-state particle — passed to Geant4 for tracking
2	**Decayed** by the generator — kept in the record; its decay products carry the physics
3	Documentation entry — internal generator history (e.g. hard-process partons)
4	Incoming beam particle
5–10	Reserved by the HepMC standard
11–200	Generator-dependent internal codes

Digitization: Generic Algorithms



Tracker hits

Energy deposits below a configured threshold discarded

Hit times smeared according to a time resolution

Calorimeter hits (e.g. fEMCal with simplified geometry)

Total energy smeared with a resolution function comprising stochastic, constant, and noise terms

All hits converted to an ADC value within a configurable dynamic range (including a fluctuating pedestal) and a TDC value derived as a smeared hit time

Digitization: Generic Algorithms

Algorithm	Models	Subsystem (s)	Status
CalorimeterHitDigi simple	Resolution-function smearing (stochastic, constant, noise terms), ADC with fluctuating pedestal, TDC from smeared time	BHCAL, FHCAL, EHCAL, …	Default reco: Placeholder smearing
CalorimeterHitDigi sipm	Light yield → detected photons (photon detection efficiency) → fired pixels with saturation → ADC incl. pedestal	EEEMCal, fEMCal	Default reco
SiliconTrackerDigi + RandomNoisePixel	Energy threshold, hit-time smearing; random noise-pixel injection into empty channels	MAPS Si vertex, barrel & endcap trackers	Default reco
MPGDTrackerDigi	Dedicated gas-detector response for micro-pattern gaseous detectors	MPGD barrel & endcap trackers	Default reco
PhotoMultiplierHitDigi	Photosensor response for single-photon detection (SiPM / HRPPD / MCP-PMT)	dRICH, pfRICH, DIRC	Default reco

Digitization: Advanced Algorithms

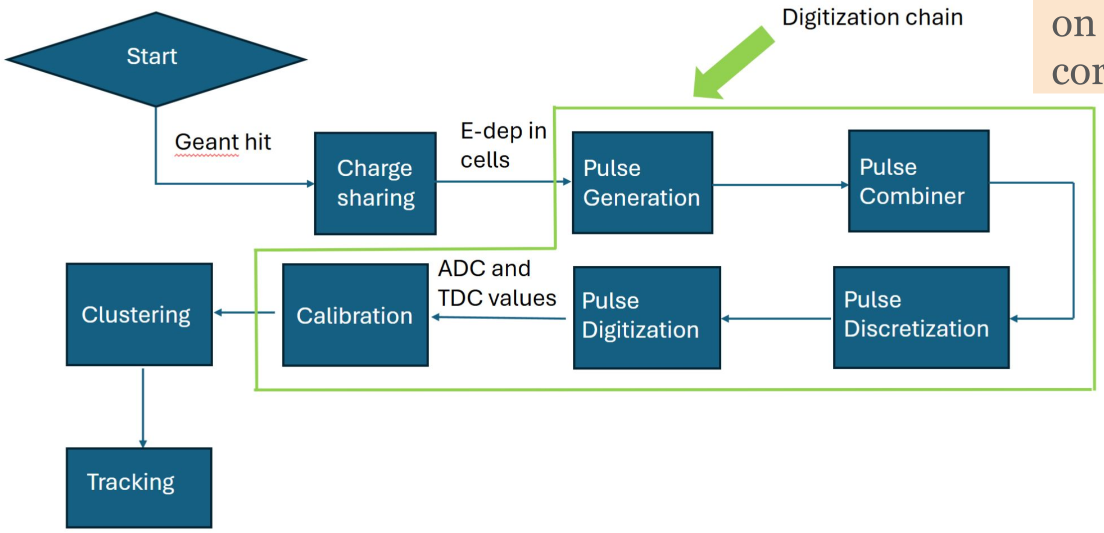
Charge sharing: deposited charge distributed over neighboring readout cells with a Gaussian profile, reproducing sub-cell position resolution in sensors

Pulse generation: per-cell deposits converted to analog waveforms with configurable shapes (Landau or user-defined)

Pulse combination: overlapping pulses on the same channel combined; optional correlated electronic noise injection

Pulse discretization: waveform resampled into fixed sampling windows matching the readout clock

Pulse digitization: final conversion to ADC and TDC values via emulations of the specific front-end ASICs



Modular waveform-based digitization chain in eicrecon

Digitization: Advanced Algorithms

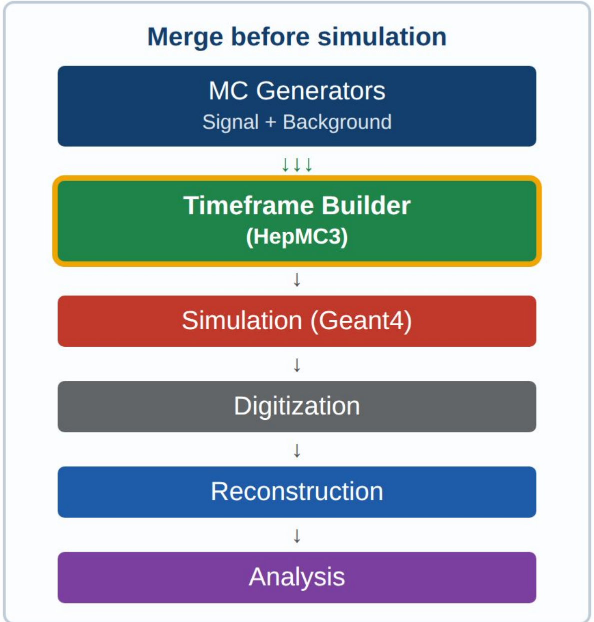
Stage	Barrel TOF	Endcap TOF	BIC	Low Q2 Tagger
Charge Sharing	x	x		x
Pulse Generation	x	x	x	x
Pulse Combination	x	x	x	x
Pulse Noise Injection			x	x
Pulse discretization	x	x		
ASIC digitizer	CFDROC	EICROC	CALROC	

Waveform-chain outputs are produced alongside the default collections but not yet consumed by default reconstruction

Background Mixed Simulations: Timeframe Builder

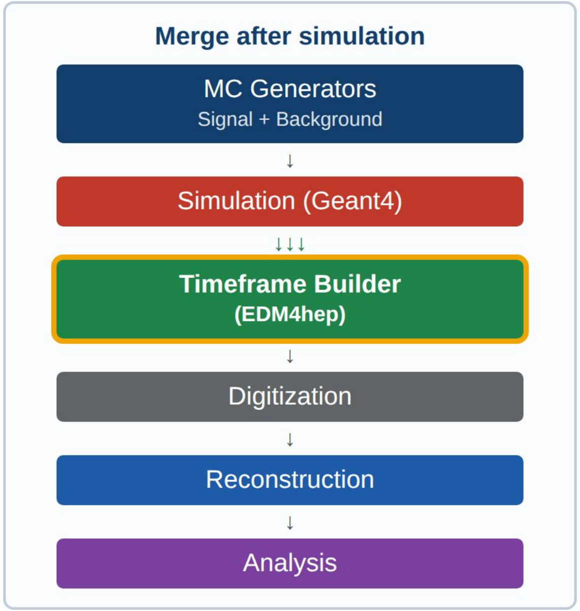
Legacy Mode

Replicate previous hepmc merger but with added timing structure (merging still happens pre-simulation)

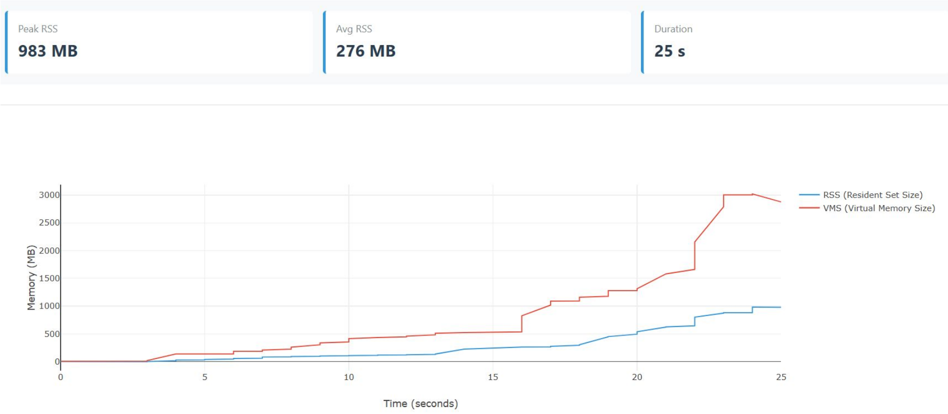


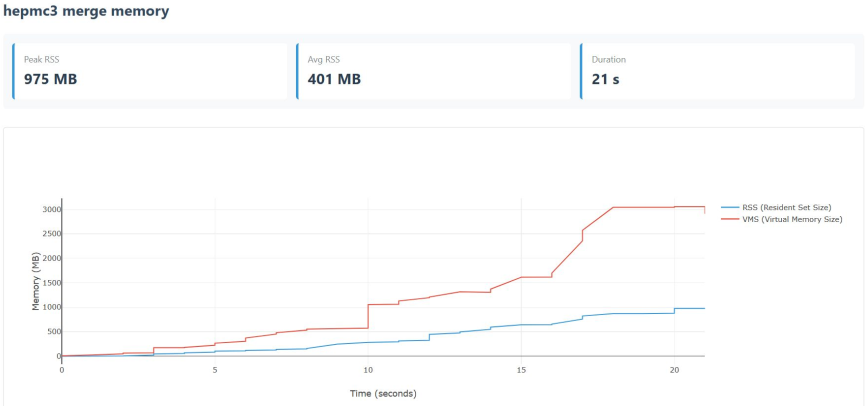
Post-simulation Merger Mode

Simulate background pools once and merge statistically at the 2 μs frame level with available edm4hep output root files (merging happens between simulation and digitization step)



Background Mixed Simulations: Timeframe Builder





Legacy Mode

Post-simulation Merger Mode

Currently, evaluating feasibility in terms of physics reproducibility and production resource limitations for both modes

Summary and Outlook

Background mixing: backgrounds sampled by rate into 2 μs timeslices; provenance via status-code offsets; FULL + RECO samples in rucio (26.07.1)

Digitization: Generic algorithms in default reco path + waveform chain with ASIC emulations (CFDROC, EICROC, CALOROC) in parallel path

Timeframe Builder merging moved after Geant4: simulate background pools once, reuse everywhere. Evaluating arrow IPC output backends to avoid storing intermediate full edm4hep files.

A = noise + background only · **B** = physics + noise + background

Realistic time sequences: alternate signal-free and signal-bearing periods with physics occurrence sampled from the collision rate, replacing today's exactly-1-signal-per-frame configuration.

Detector Noise: now per channel at digitization; signal-free periods call for standalone noise generation in empty frames

Advanced Algorithms: Implement full reconstruction path