

State-Space Model as Triggering and Reconstruction Algorithm in the **Project 8** Neutrino Mass Experiment

Hannah Binney, Luis Felipe Koehler Domingues,
Christina Reissel, Kieran Wall, **Yao Yin**,
Joseph Formaggio, Phil Harris, Elise Novitski

APS DPF CPAD Conference — RDC 5 TDAQ — Oct 21, 2026

PROJECT 8

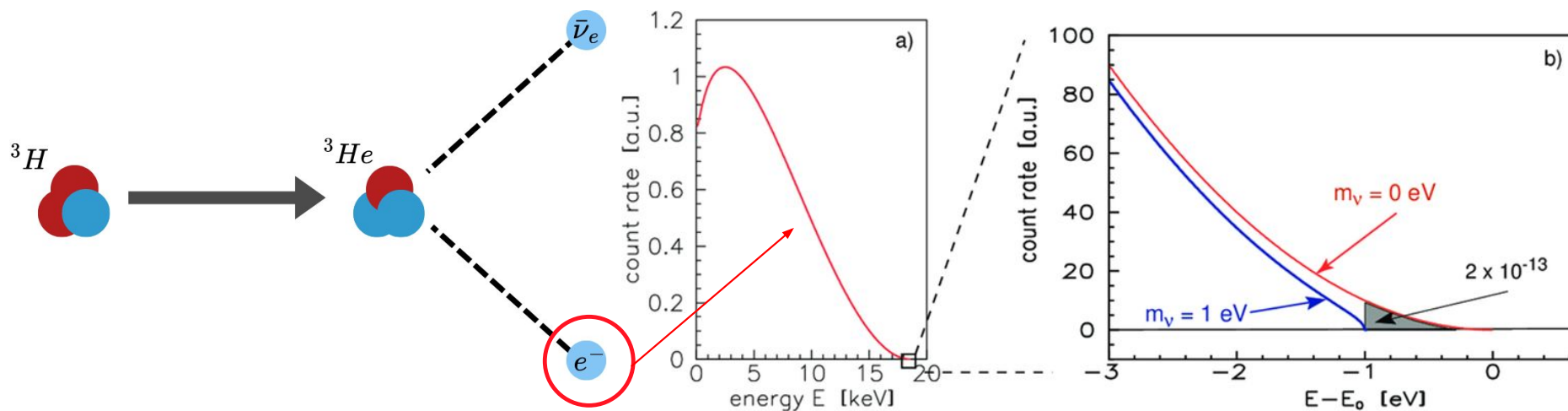


Outline

Intro to Project 8	Why SSM?	Architecture & Training	Performance & Application in DAQ Chain	Conclusion
--------------------------	----------	----------------------------	--	------------

Neutrino Mass with Precision Beta Spectroscopy

Neutrino mass can be measured directly through tritium beta decay: Energy carried away by the neutrino distorts the endpoint of the tritium beta decay energy spectrum.

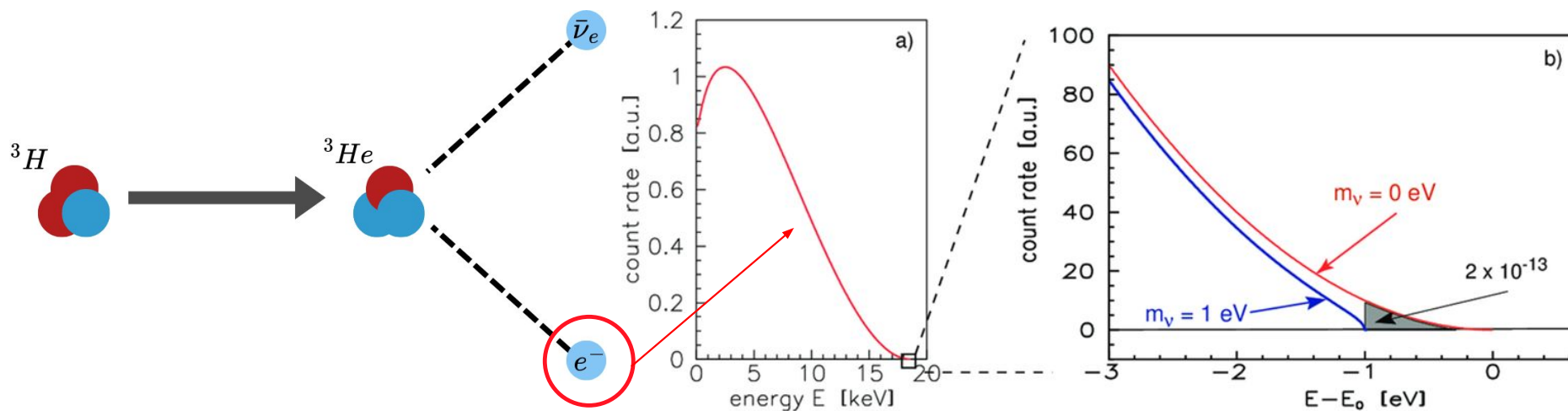


Plots: KATRIN Collaboration

Neutrino Mass with Precision Beta Spectroscopy

Neutrino mass can be measured directly through tritium beta decay: Energy carried away by the neutrino distorts the endpoint of the tritium beta decay energy spectrum.

keV-scale low-energy experiment, sub-ppm precision



Plots: KATRIN Collaboration

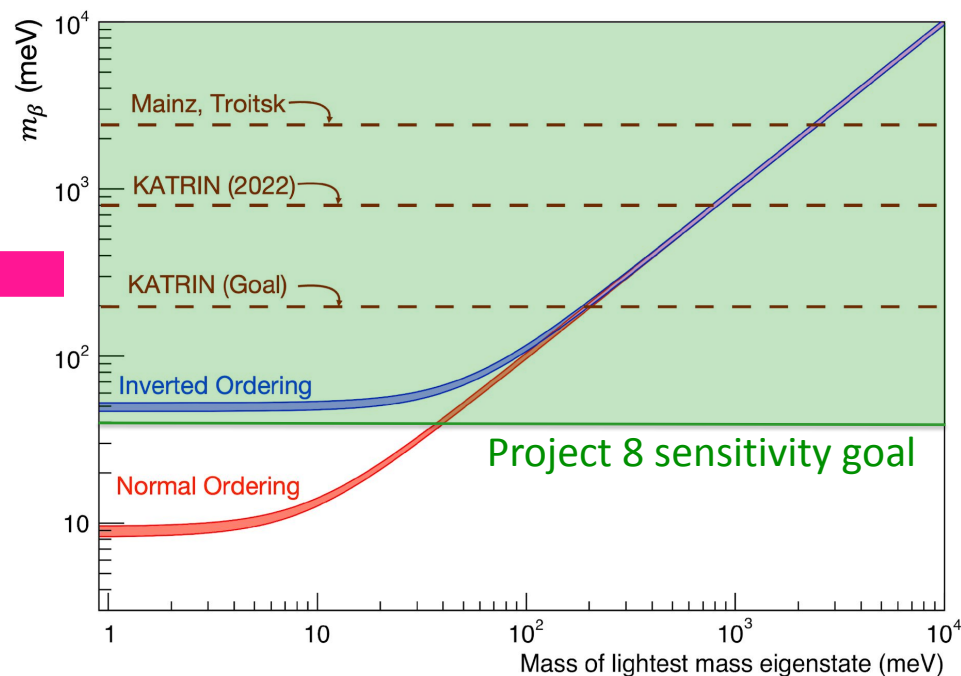
Neutrino Mass with Precision Beta Spectroscopy

Neutrino mass can be measured directly through tritium beta decay: Energy carried away by the neutrino distorts the endpoint of the tritium beta decay energy spectrum.

Project 8 Goal:

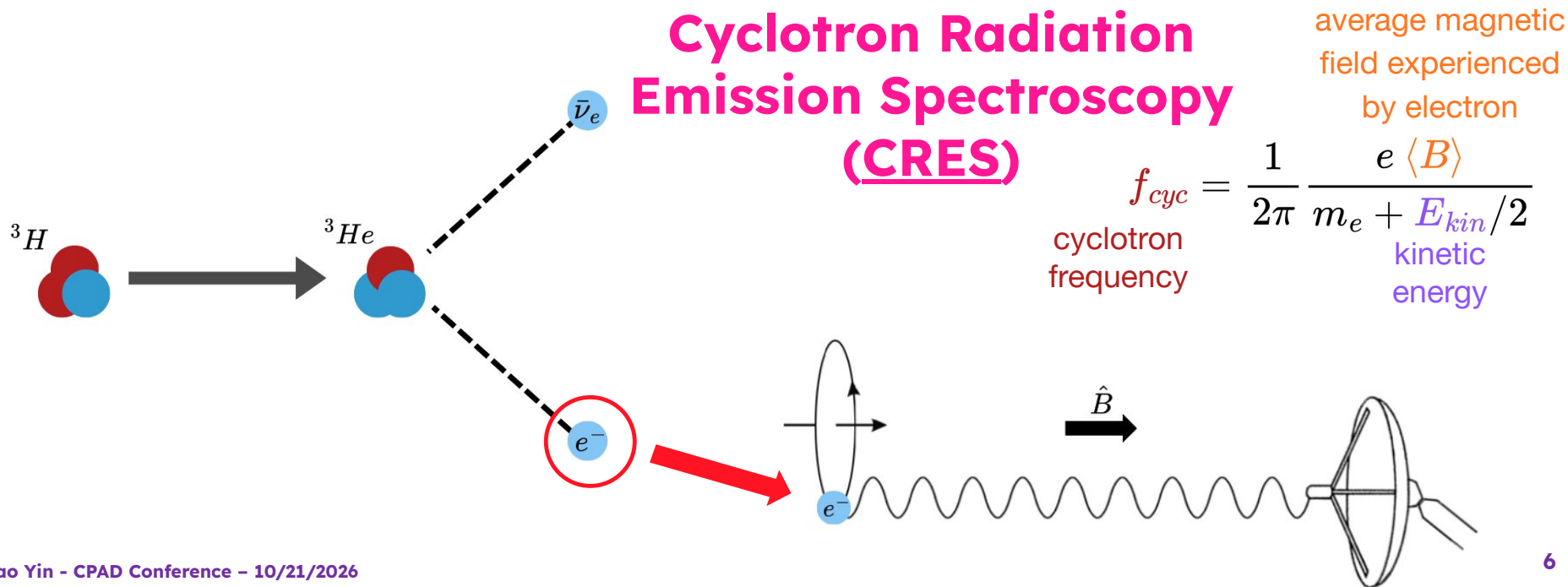
sensitivity to **40 meV**

measuring the mass or excluding
the inverted ordering.



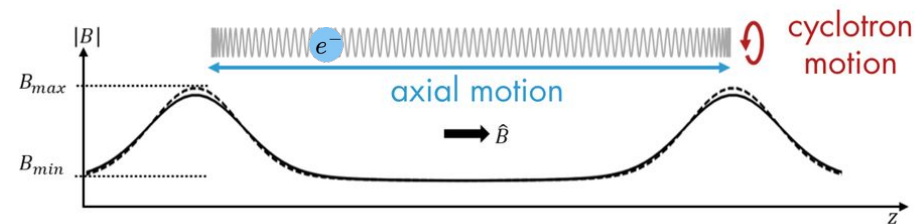
Neutrino Mass with Precision Beta Spectroscopy

Neutrino mass can be measured directly through tritium beta decay: Energy carried away by the neutrino distorts the endpoint of the tritium beta decay energy spectrum.

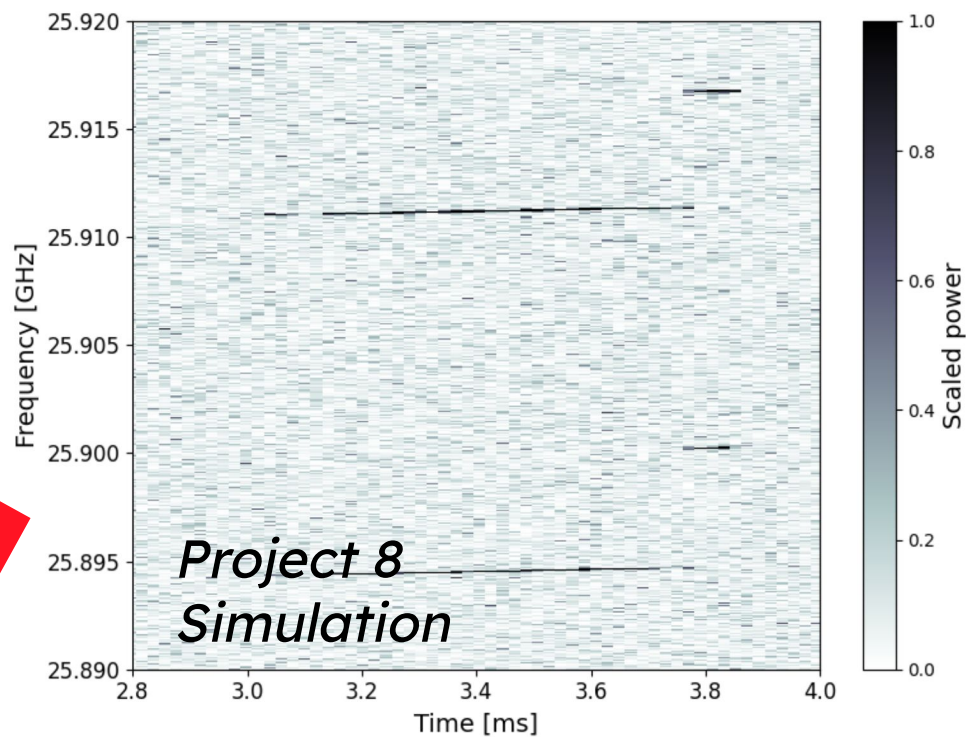
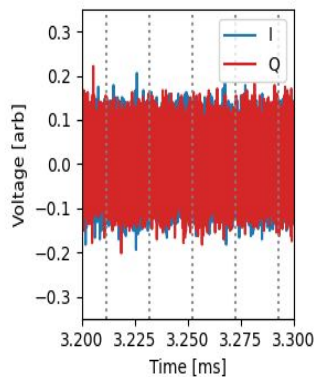
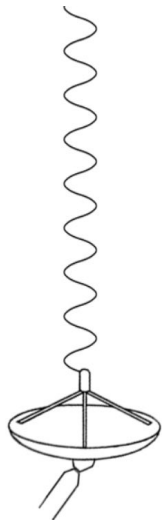


CRES Signals

Electron undergoes cyclotron and axial motion inside a magnetic trap.



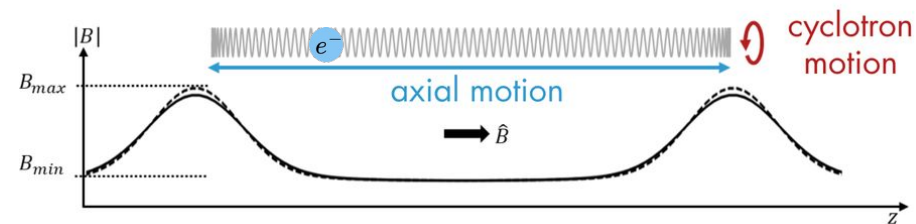
Voltage vs. time



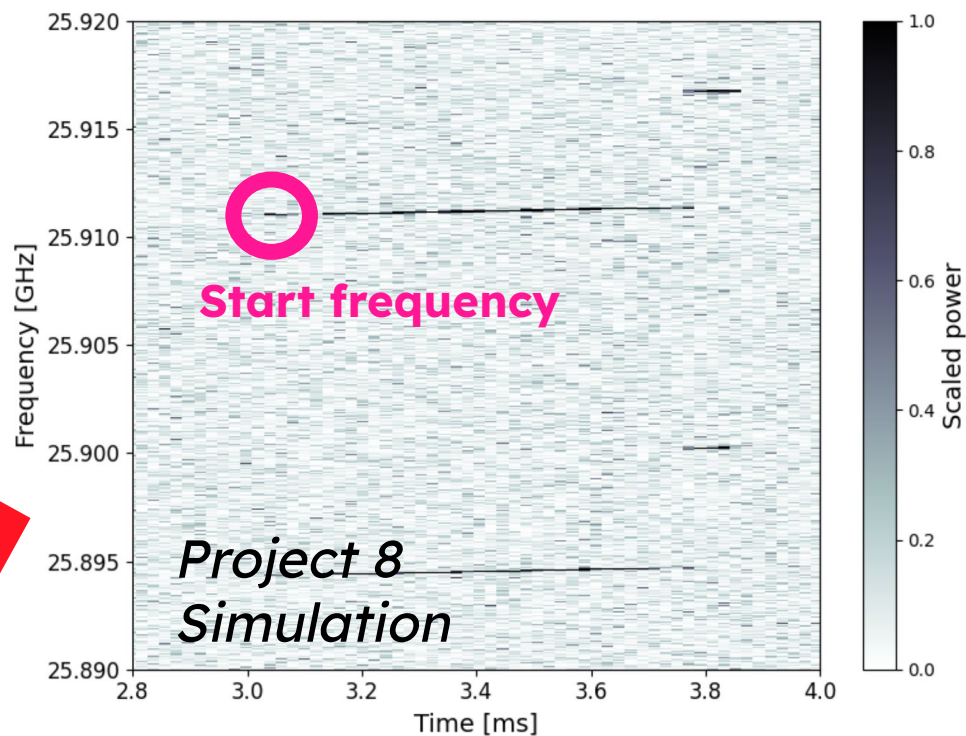
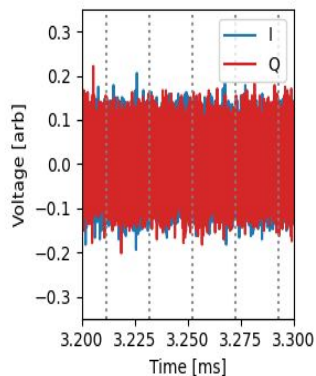
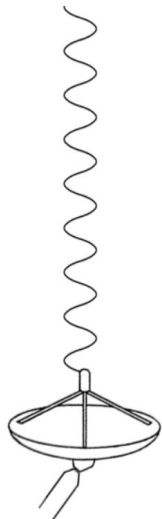
"Spectrogram"

CRES Signals

Electron undergoes cyclotron and axial motion inside a magnetic trap.



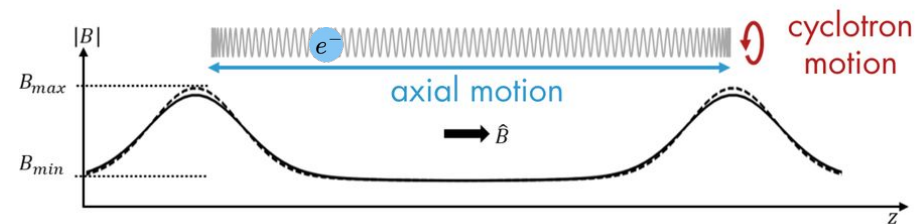
Voltage vs. time



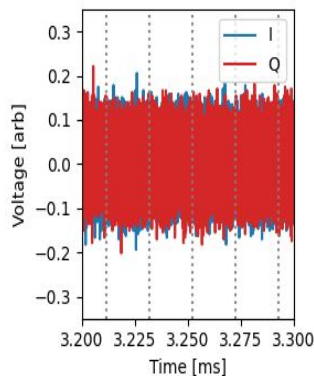
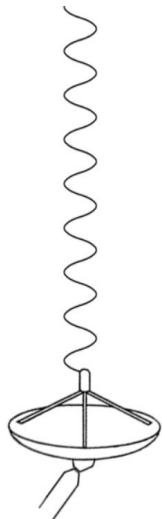
"Spectrogram"

CRES Signals

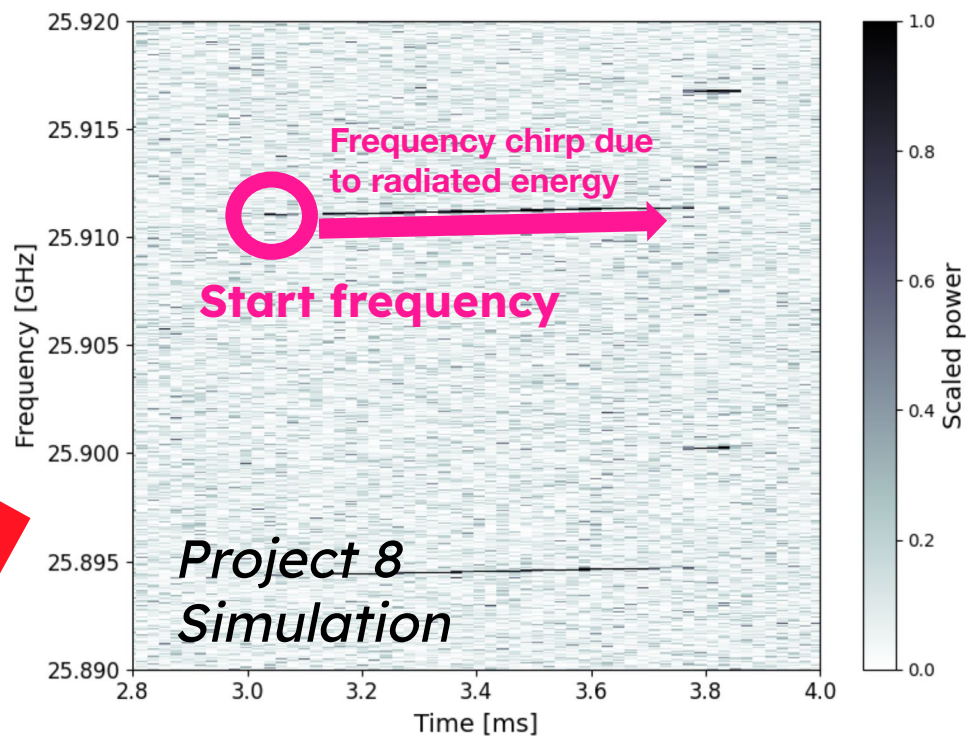
Electron undergoes cyclotron and axial motion inside a magnetic trap.



Voltage vs. time



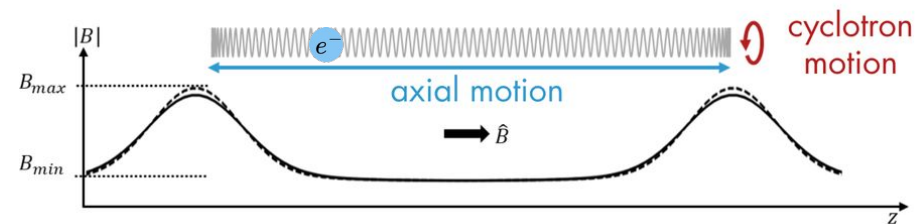
FT



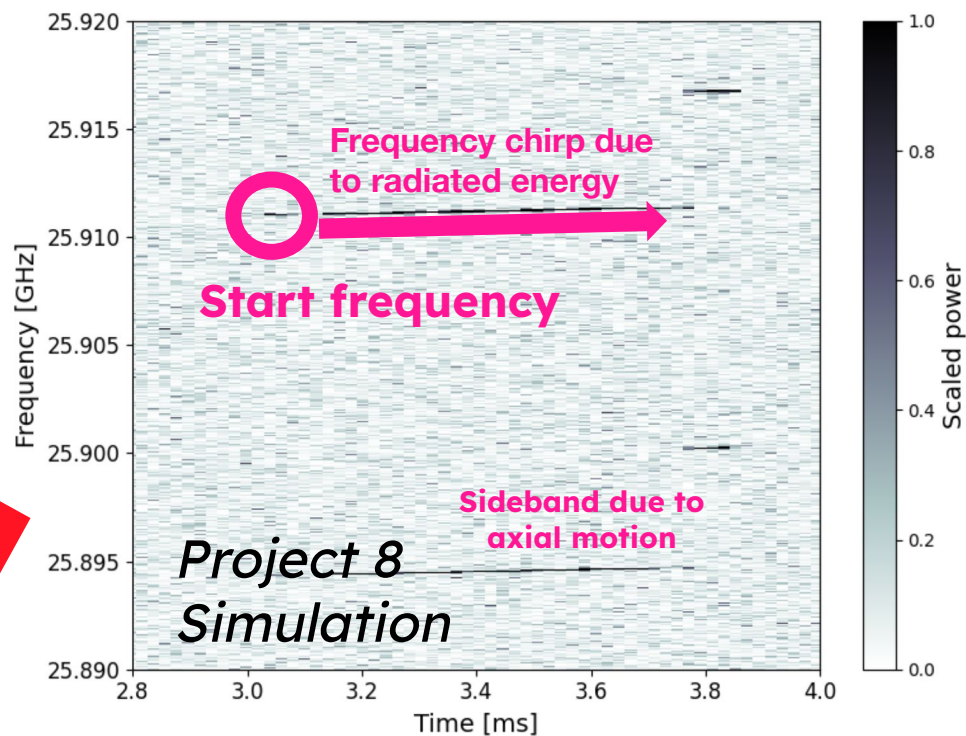
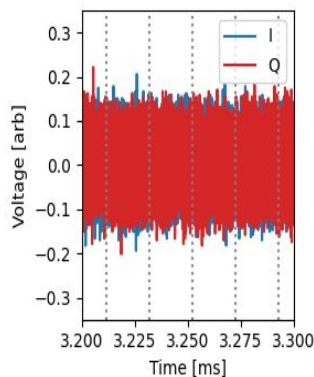
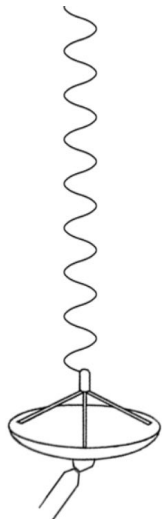
“Spectrogram”

CRES Signals

Electron undergoes cyclotron and axial motion inside a magnetic trap.



Voltage vs. time

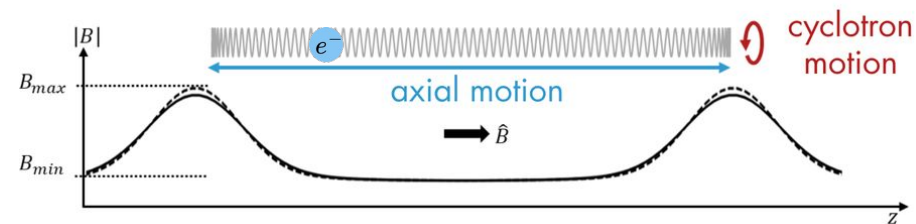


*Project 8
Simulation*

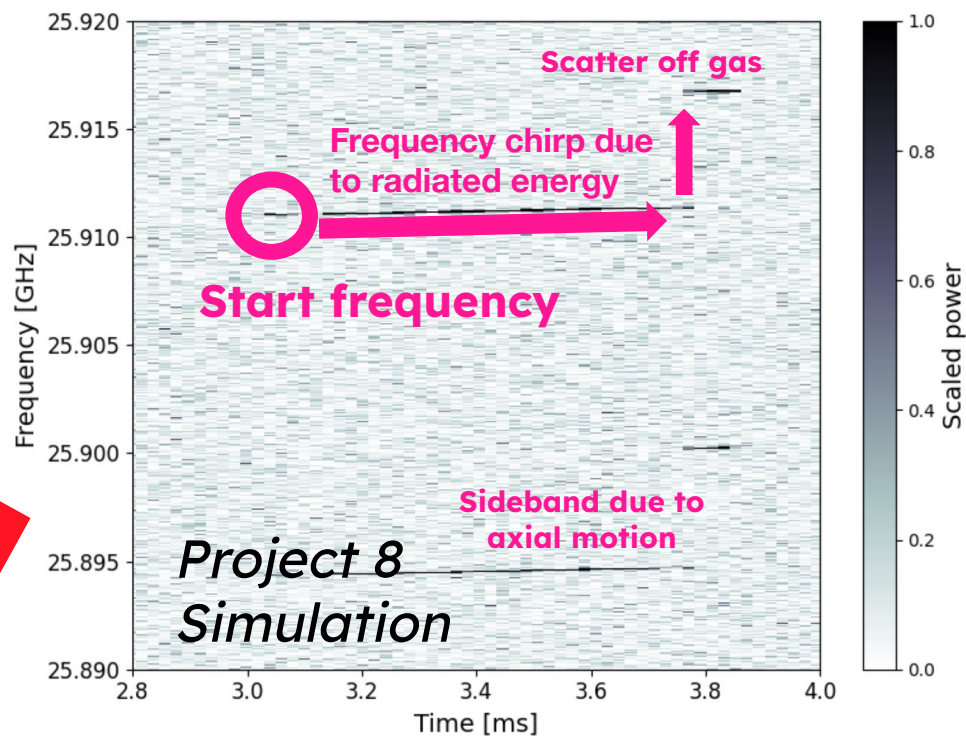
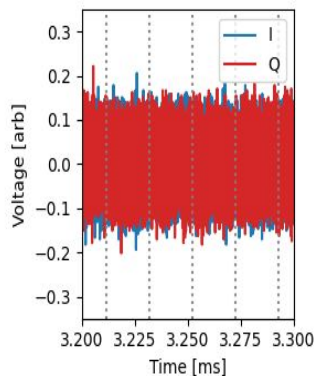
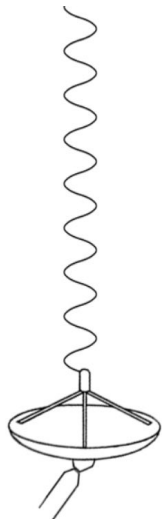
“Spectrogram”

CRES Signals

Electron undergoes cyclotron and axial motion inside a magnetic trap.



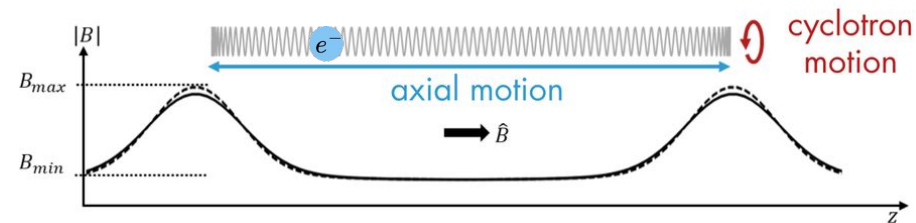
Voltage vs. time



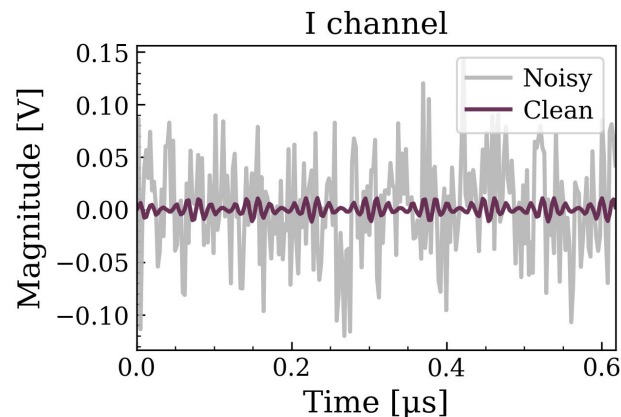
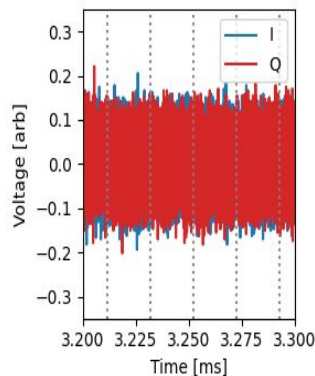
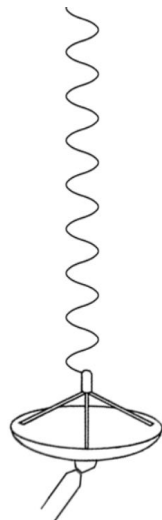
"Spectrogram"

CRES Signals

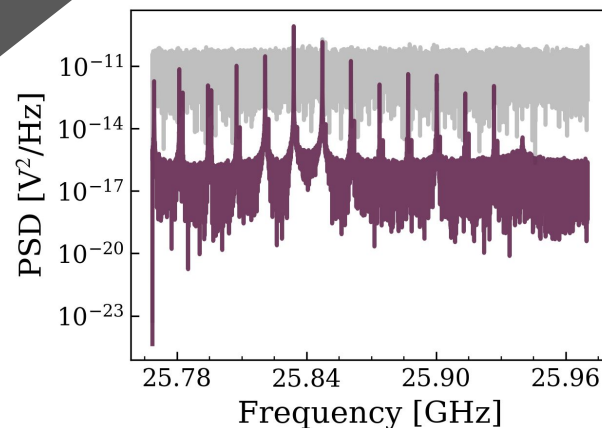
Electron undergoes cyclotron and axial motion inside a magnetic trap.



Voltage vs. time

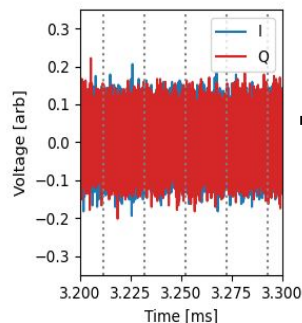


Low signal to noise ratio!



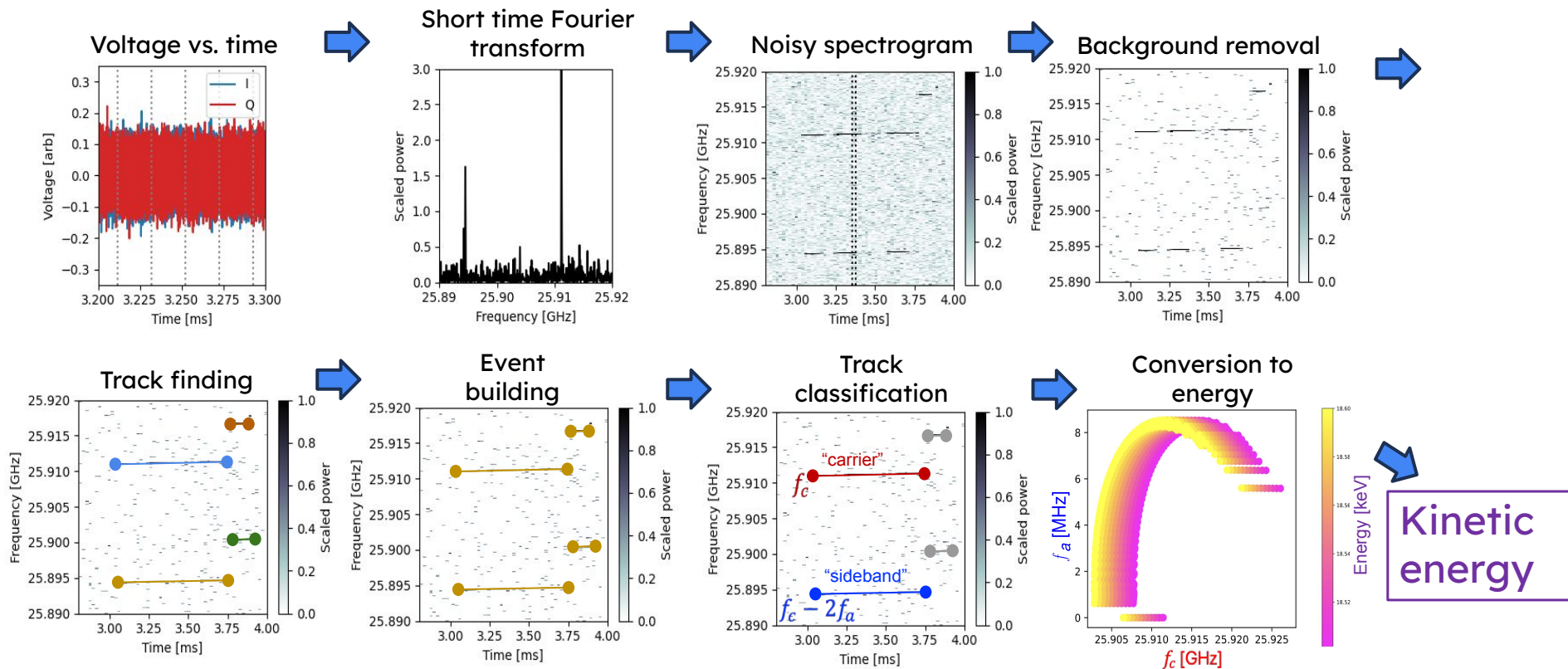
Project 8 event reconstruction

Voltage vs. time

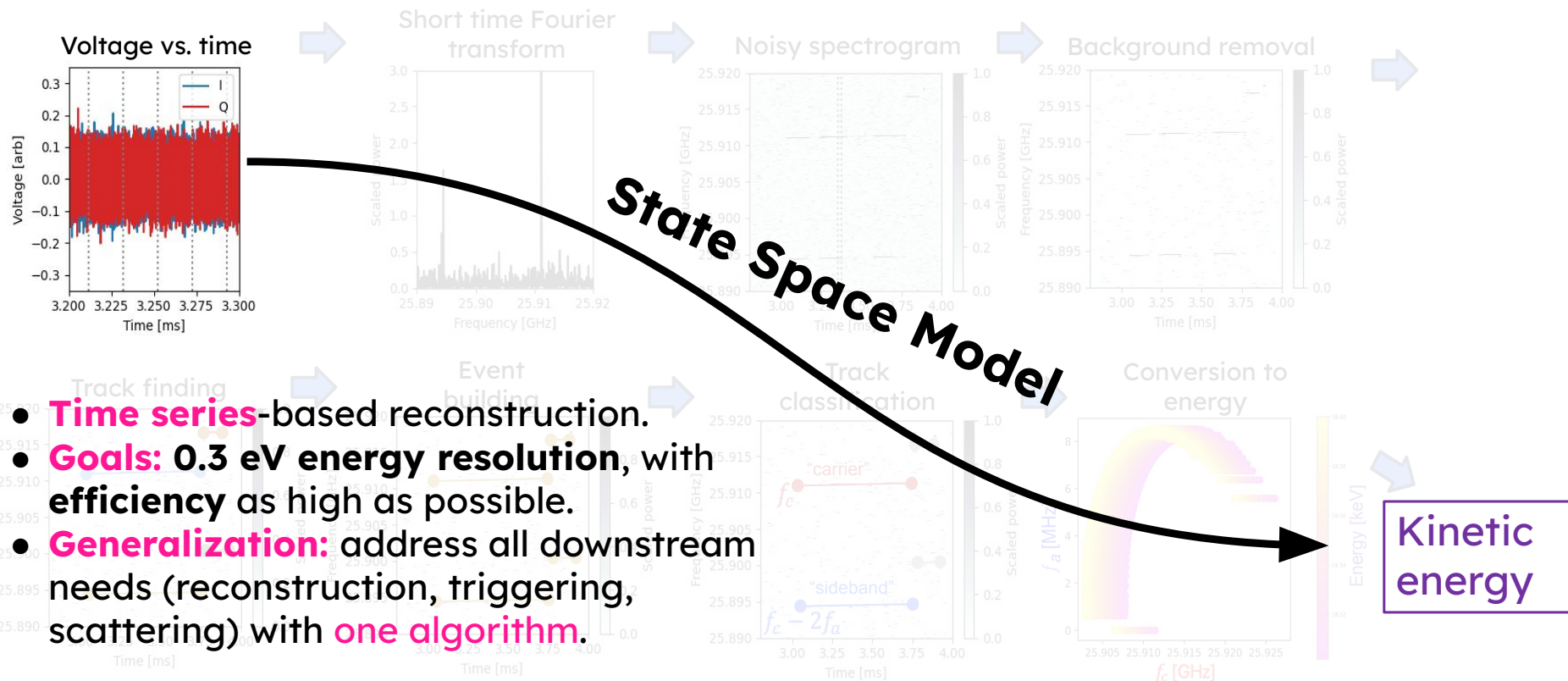


Kinetic
energy

Standard reconstruction algorithm



Reconstruction with State Space Model



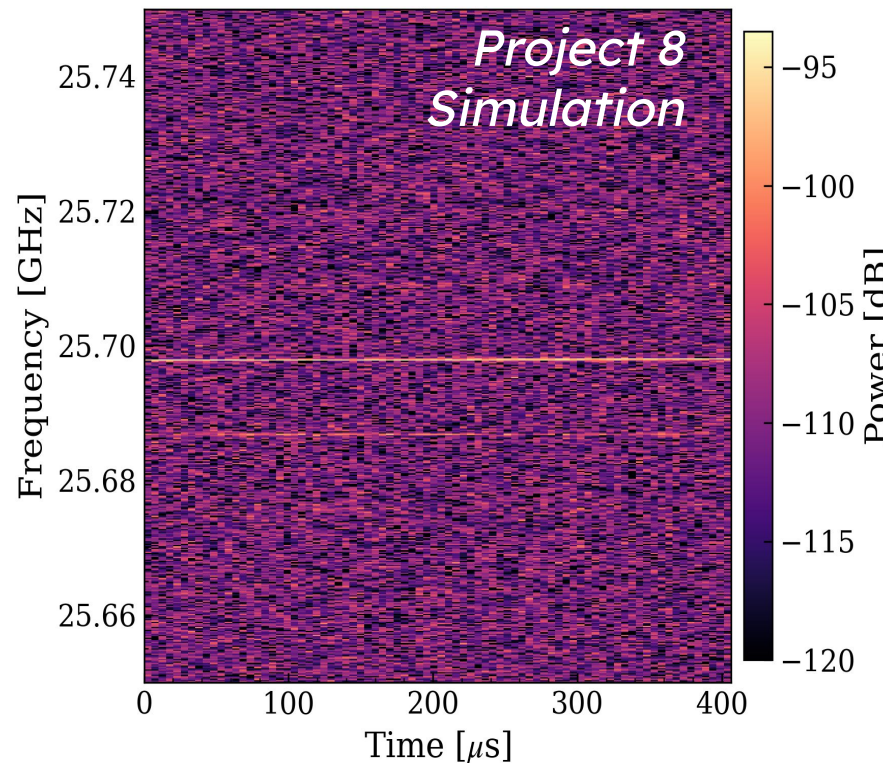
Why State-Space Models

- Our implementation has computational and memory complexity of $O(L)$. Linear scaling with long context!
- More see recently accepted NeurIPS paper: [PhyTS: A Benchmark for Scientific Time Series](#)
- Small model (~600k) for long sample size, FPGA hardware friendly

Window size	Samples in window	Transformer 1 attention head memory needed	Our SSM recurrent state memory needed
102 μ s	40 960	6.7 GB	0.4 MB

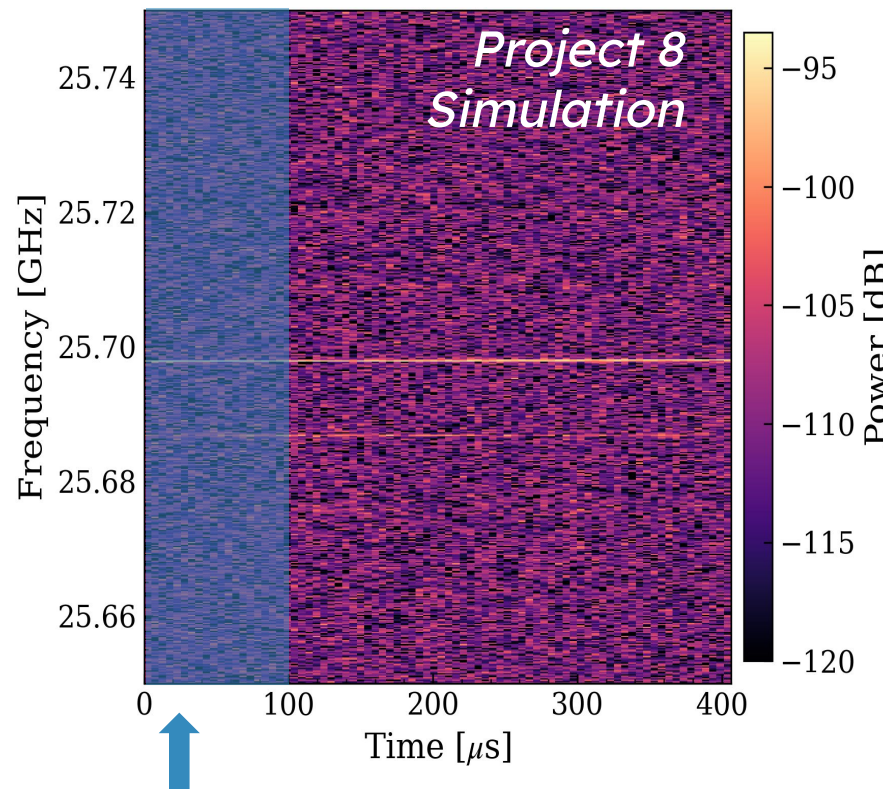
Training dataset

- **~100 k** high-fidelity **simulations** of realistic cavity CRES signals.
- Flat in energy around Tritium endpoint 18.5 – 18.7 keV. Uniform distributions in other observables
- Signal **~0.4 ms**, corresponding to **163 840 samples** → only **40 960** used for training and inference (102 μ s).
- **No pileup or scattering**, all signals start at $t = 0$ s.
- **No noise**; gaussian noise model added to I/Q during training.



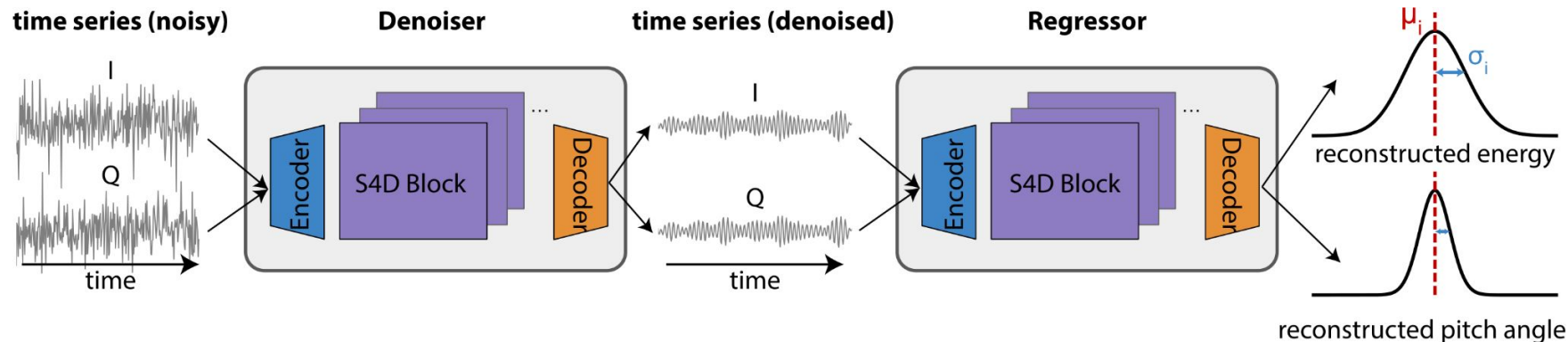
Training dataset

- **~100 k** high-fidelity **simulations** of realistic cavity CRES signals.
- Flat in energy around Tritium endpoint 18.5 – 18.7 keV. Uniform distributions in other observables
- Signal **~0.4 ms**, corresponding to **163 840 samples** → only **40 960** used for training and inference (102 μ s).
- **No pileup or scattering**, all signals start at $t = 0$ s.
- **No noise**; gaussian noise model added to I/Q during training.



track length used for
training and testing

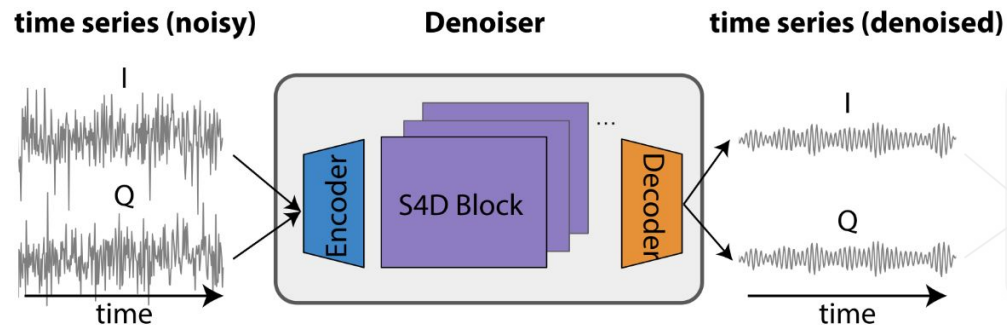
Model Architecture



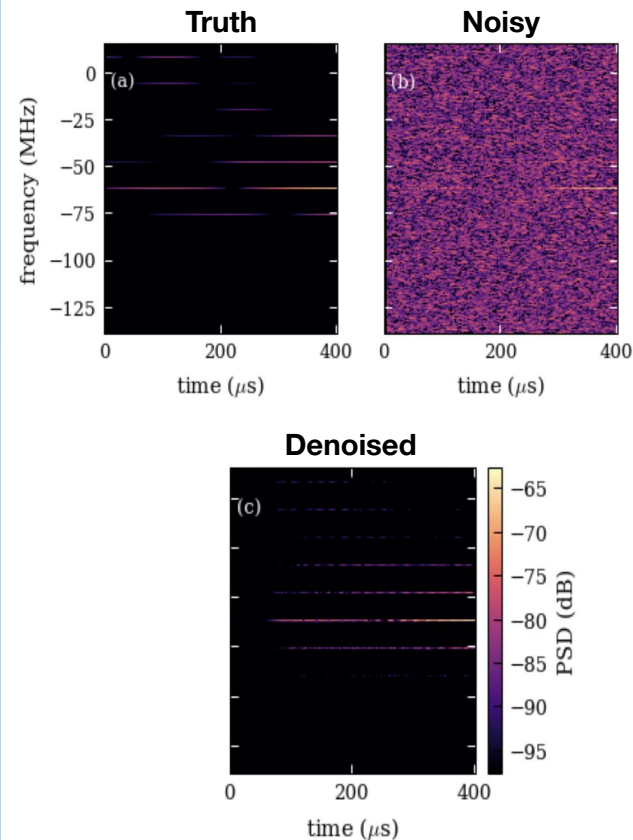
S4D, input dimension = 2, model dimension = 128, 6 layers, no dropout

~300k trainable parameters, mirrored between denoiser and regressor (total **~600k**)

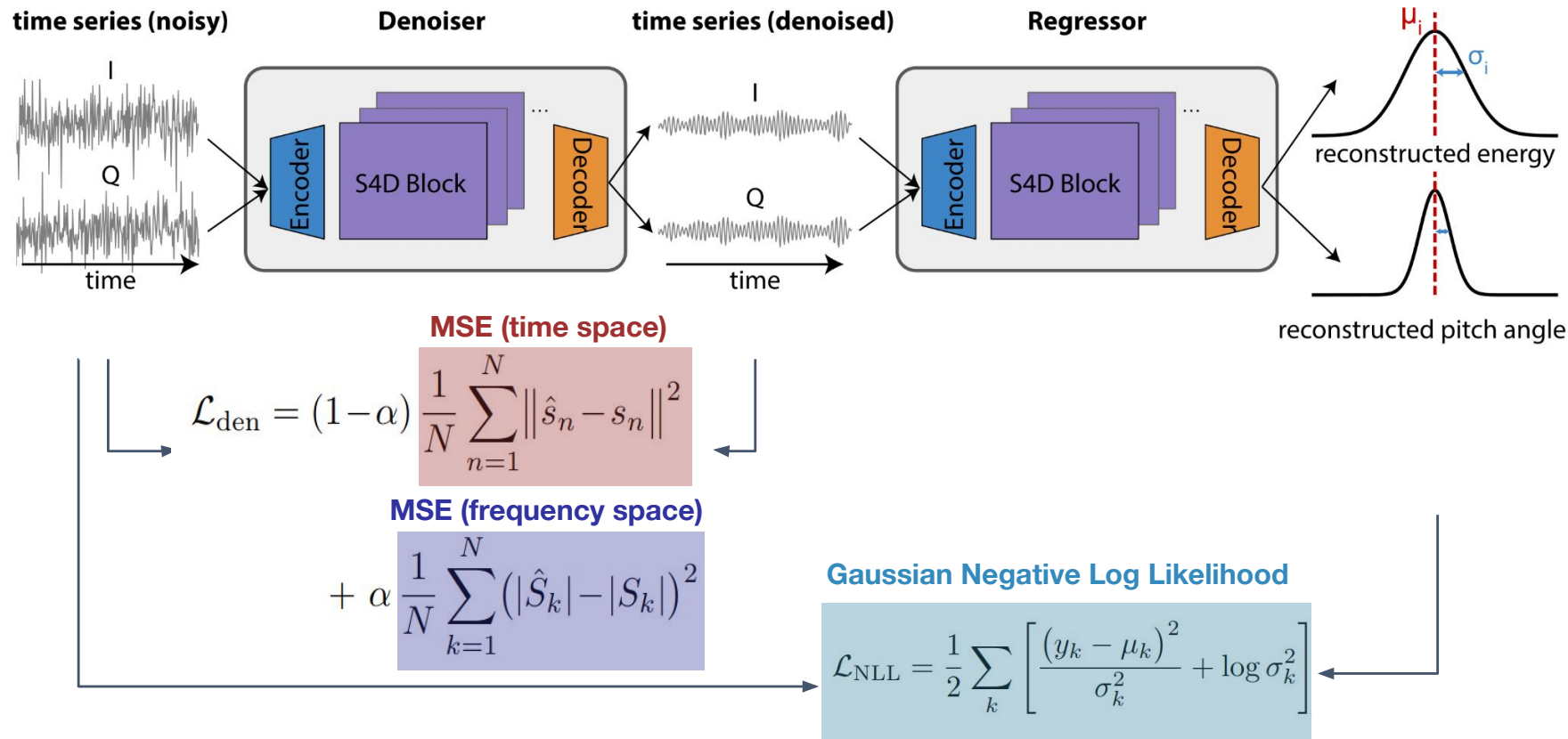
Model Architecture



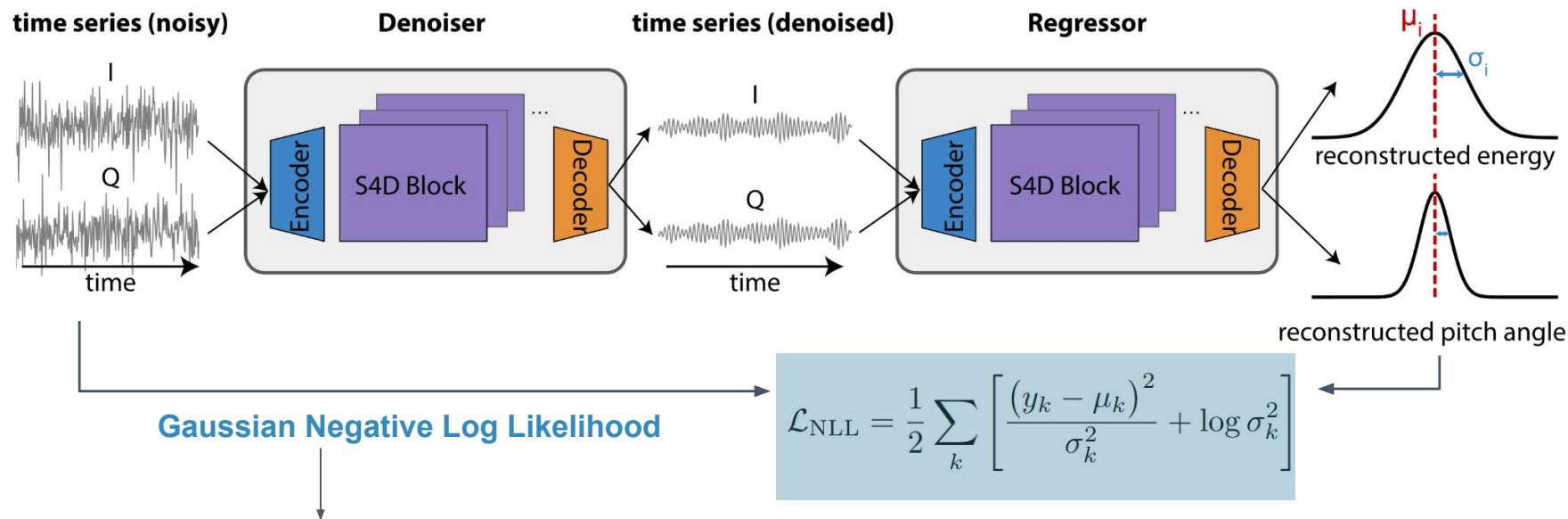
$$\mathcal{L}_{\text{den}} = (1 - \alpha) \underbrace{\frac{1}{N} \sum_{n=1}^N \|\hat{s}_n - s_n\|^2}_{\text{MSE (time space)}} + \alpha \underbrace{\frac{1}{N} \sum_{k=1}^N (|\hat{S}_k| - |S_k|)^2}_{\text{MSE (frequency space)}}$$



Model Architecture

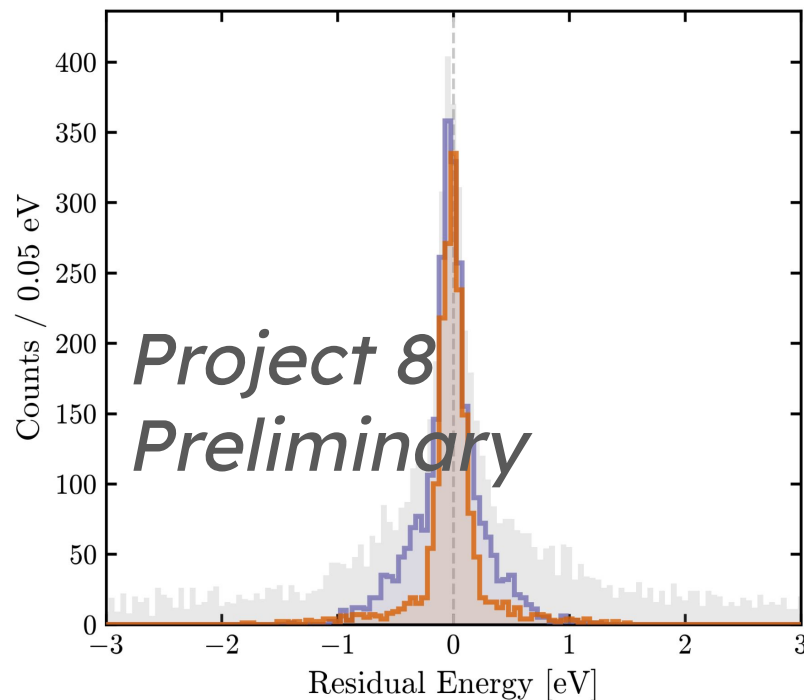


Model Architecture



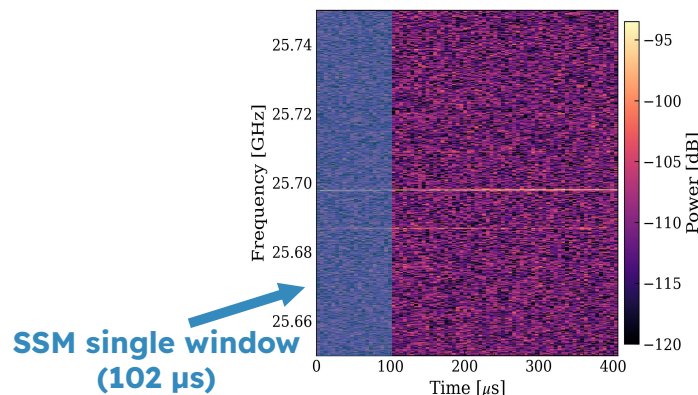
Allow us to reconstruct per-event mean + uncertainty for each observable (e.g. KE)

Electron Energy Reconstruction Performance

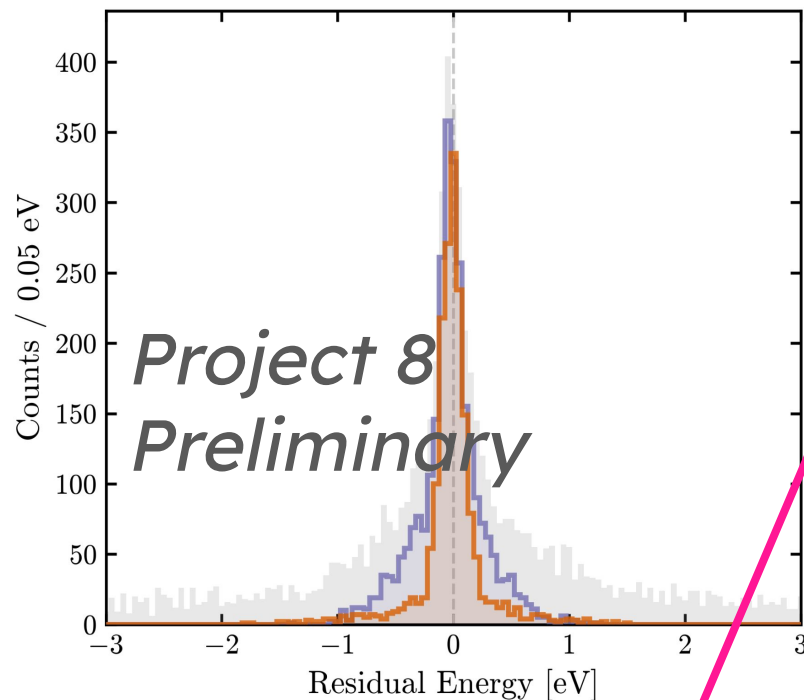


- SSM (all tracks)
std = 18.491 eV, eff = 100.0%
- SSM single window ($\sigma < 0.603$ eV)
std = 0.290 eV, eff = 25.6%
- Standard Algorithm
std = 0.296 eV, eff = 17.1%

- Compared with standard algorithm with identical test conditions.

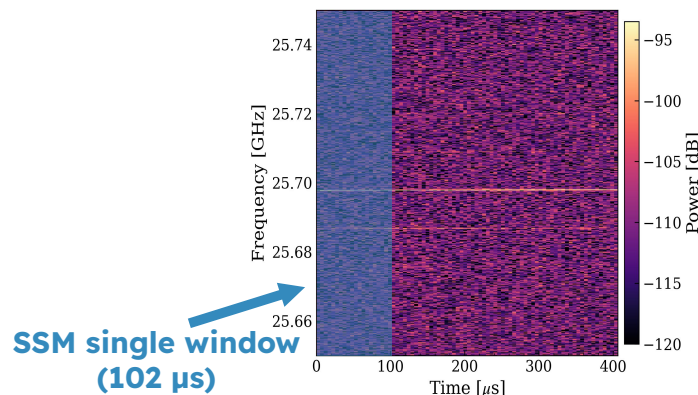


Electron Energy Reconstruction Performance

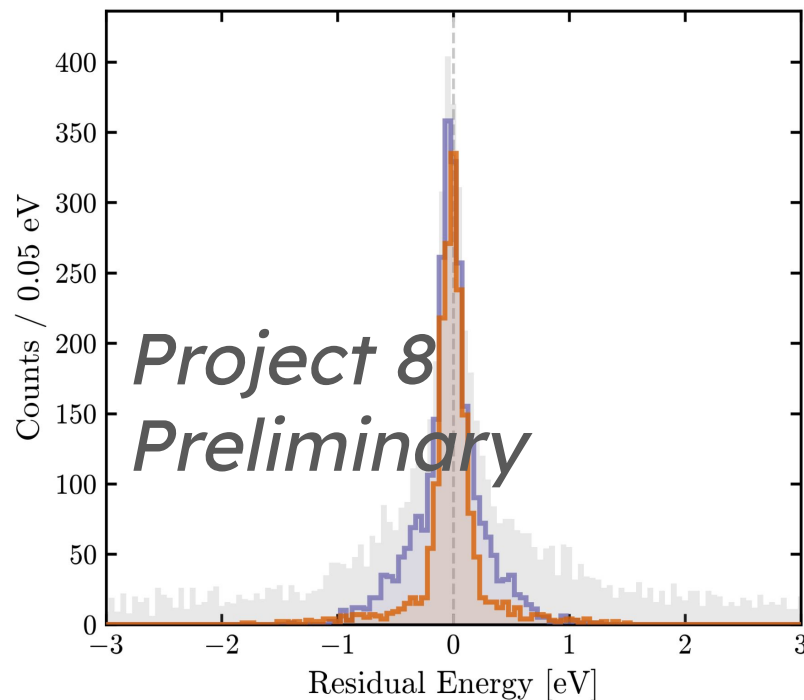


- SSM (all tracks)
std = 18.491 eV, eff = 100.0%
- SSM single window ($\sigma < 0.603$ eV)
std = 0.290 eV, eff = 25.6%
- Standard Algorithm
std = 0.296 eV, eff = 17.1%

- Compared with standard algorithm with identical test conditions.
- +50% recovered events at target resolution.**

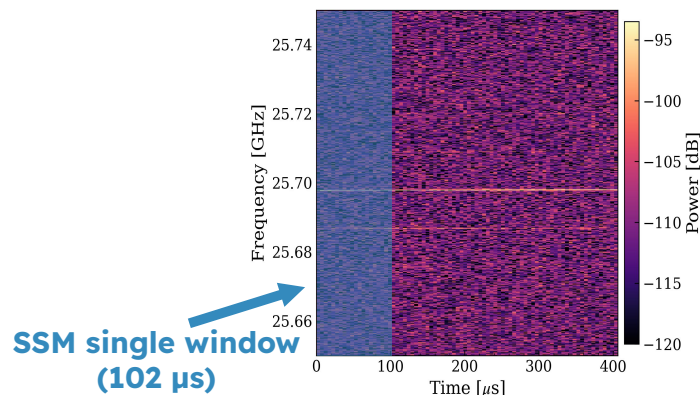


Electron Energy Reconstruction Performance

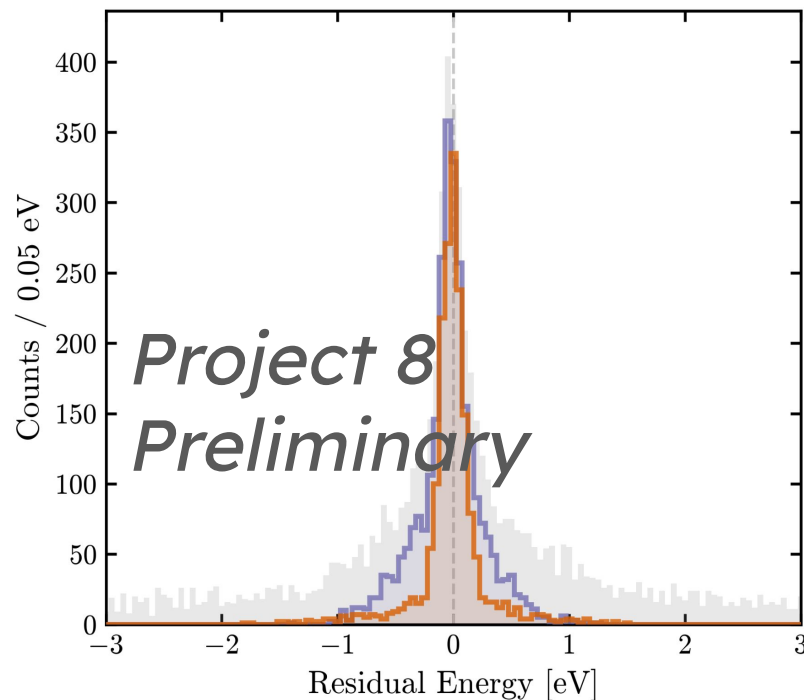


- SSM (all tracks)
std = 18.491 eV, eff = 100.0%
- SSM single window ($\sigma < 0.603$ eV)
std = 0.290 eV, eff = 25.6%
- Standard Algorithm
std = 0.296 eV, eff = 17.1%

- Compared with standard algorithm with identical test conditions.
- +50% recovered events at target resolution.**
- Same resolution achieved with only **1/4 of the data** available to the standard algorithm → **1.9x lower SNR.**

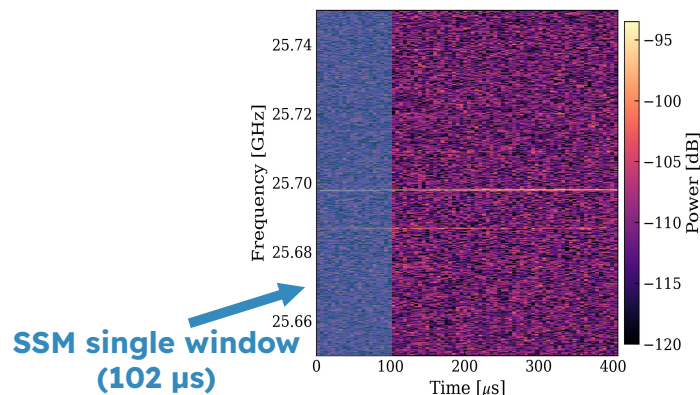


Electron Energy Reconstruction Performance



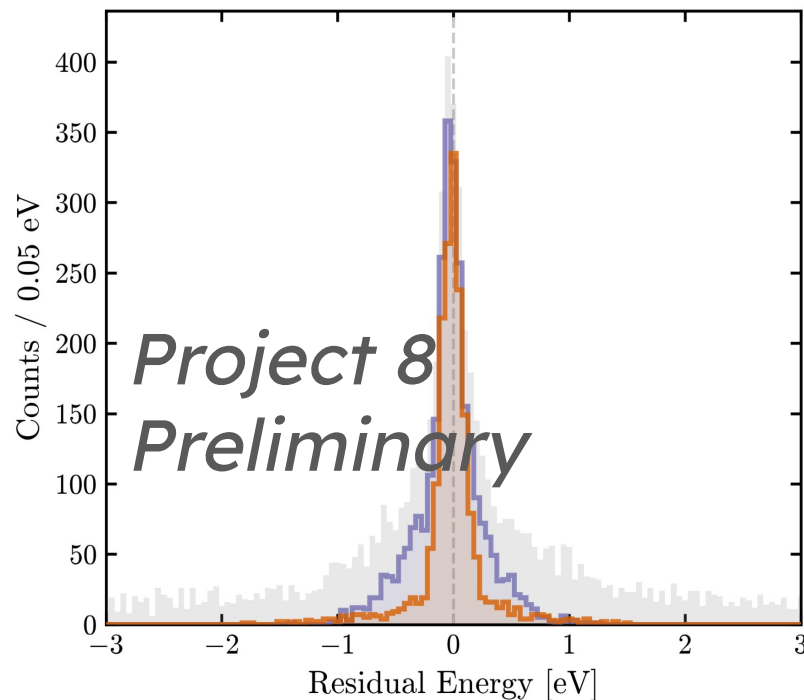
- SSM (all tracks)
std = 18.491 eV, eff = 100.0%
- SSM single window ($\sigma < 0.603$ eV)
std = 0.290 eV, eff = 25.6%
- Standard Algorithm
std = 0.296 eV, eff = 17.1%

- Compared with standard algorithm with identical test conditions.
- +50% recovered events at target resolution.**
- Same resolution achieved with only $\frac{1}{4}$ of the data available to the standard algorithm $\rightarrow$ **1.9x lower SNR.**
- Lower SNR Physics impact: possible to increase tritium gas density, thereby **increasing event rate & sensitivity**



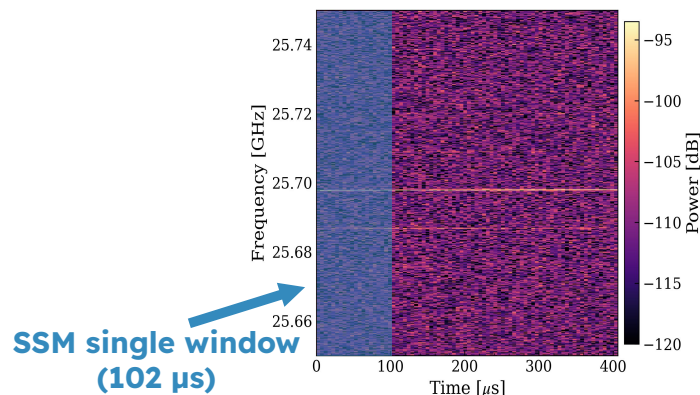
SSM single window
(102 μ s)

Electron Energy Reconstruction Performance



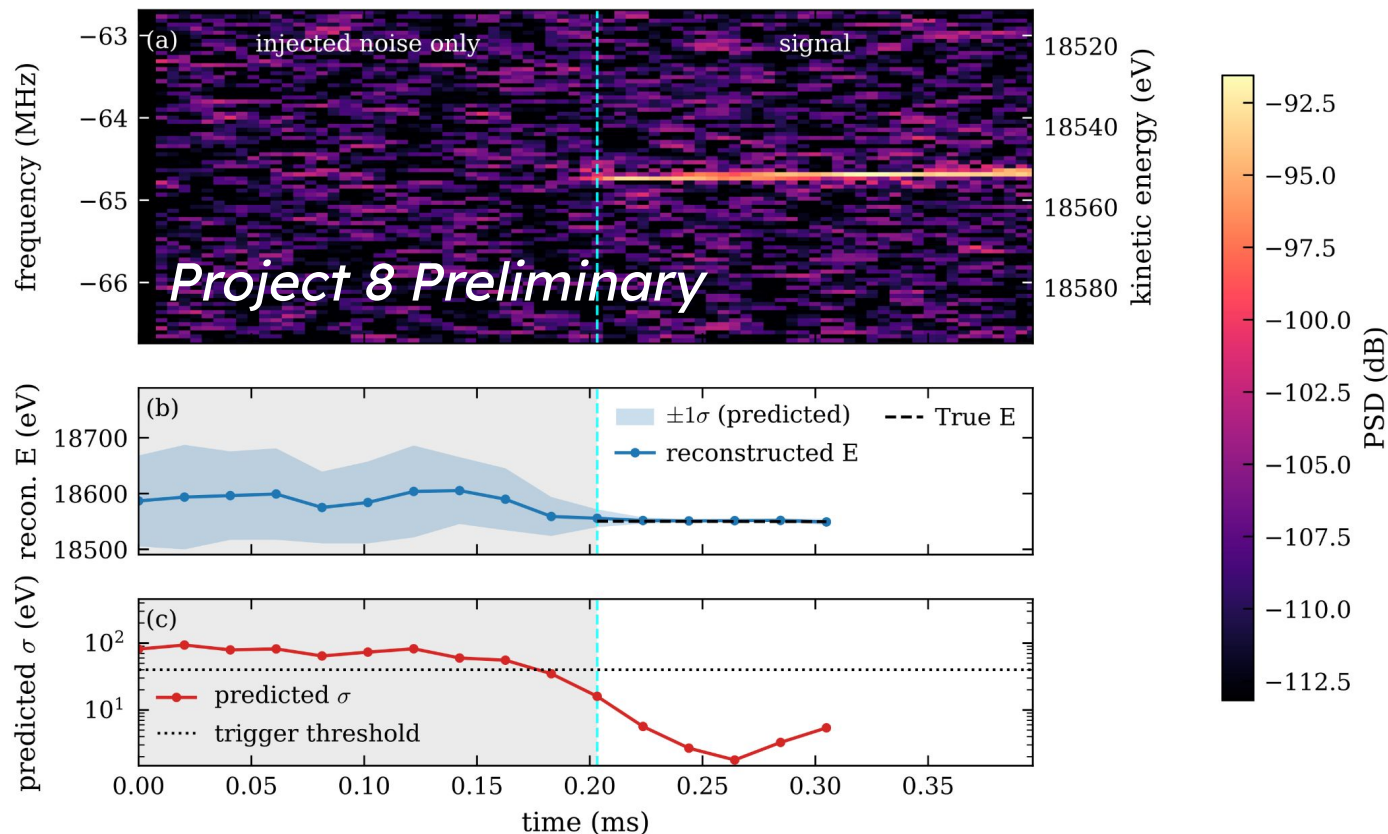
- SSM (all tracks)
std = 18.491 eV, eff = 100.0%
- SSM single window ($\sigma < 0.603$ eV)
std = 0.290 eV, eff = 25.6%
- Standard Algorithm
std = 0.296 eV, eff = 17.1%

- Compared with standard algorithm with identical test conditions.
- **+50% recovered events at target resolution.**
- Same resolution achieved with only $\frac{1}{4}$ of the data available to the standard algorithm $\rightarrow$ **1.9x lower SNR.**
- Lower SNR Physics impact: possible to increase tritium gas density, thereby **increasing event rate & sensitivity**
- Computational effort ($\sim 10k$ events):
 - Standard algorithm: **~ 1 hour (CPU)**
 - Our work: **~ 7.5 mins (RTX 5000 GPU)**



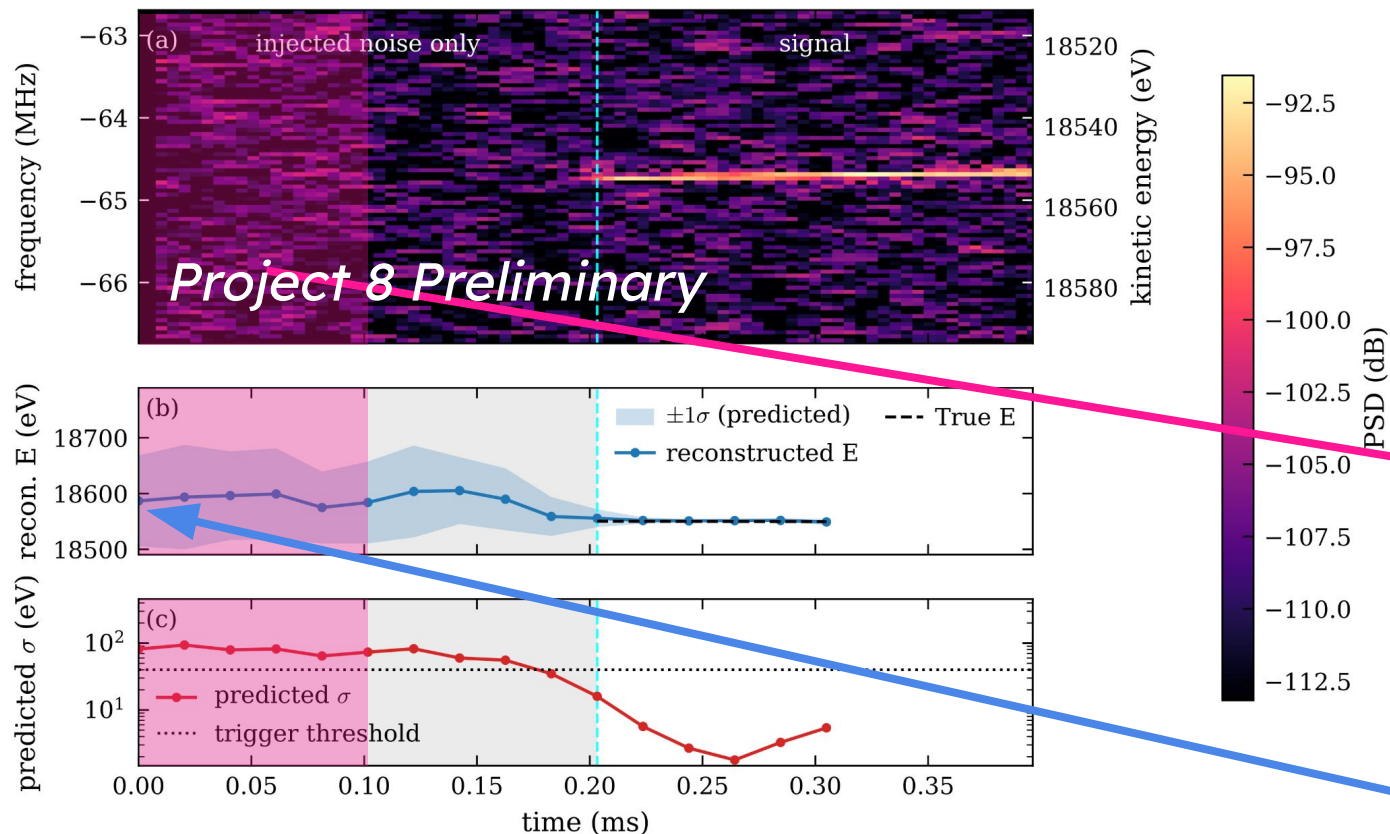
SSM as an event trigger

Trigger on a sudden decrease in reconstructed uncertainty.



SSM as an event trigger

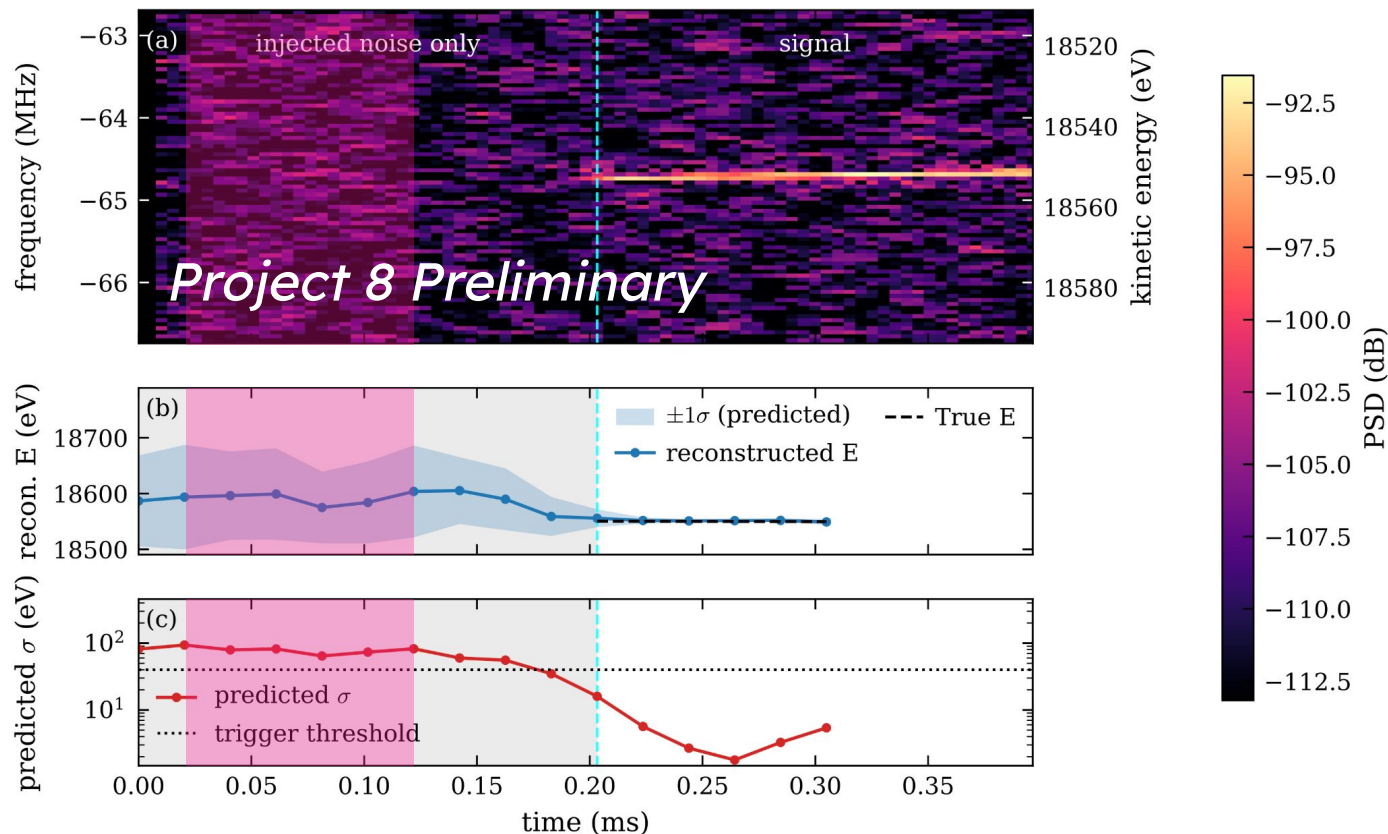
Trigger on a sudden decrease in reconstructed uncertainty.



Model takes
in 102 us of
data at a
time) $\rightarrow$
output E ,
 σ

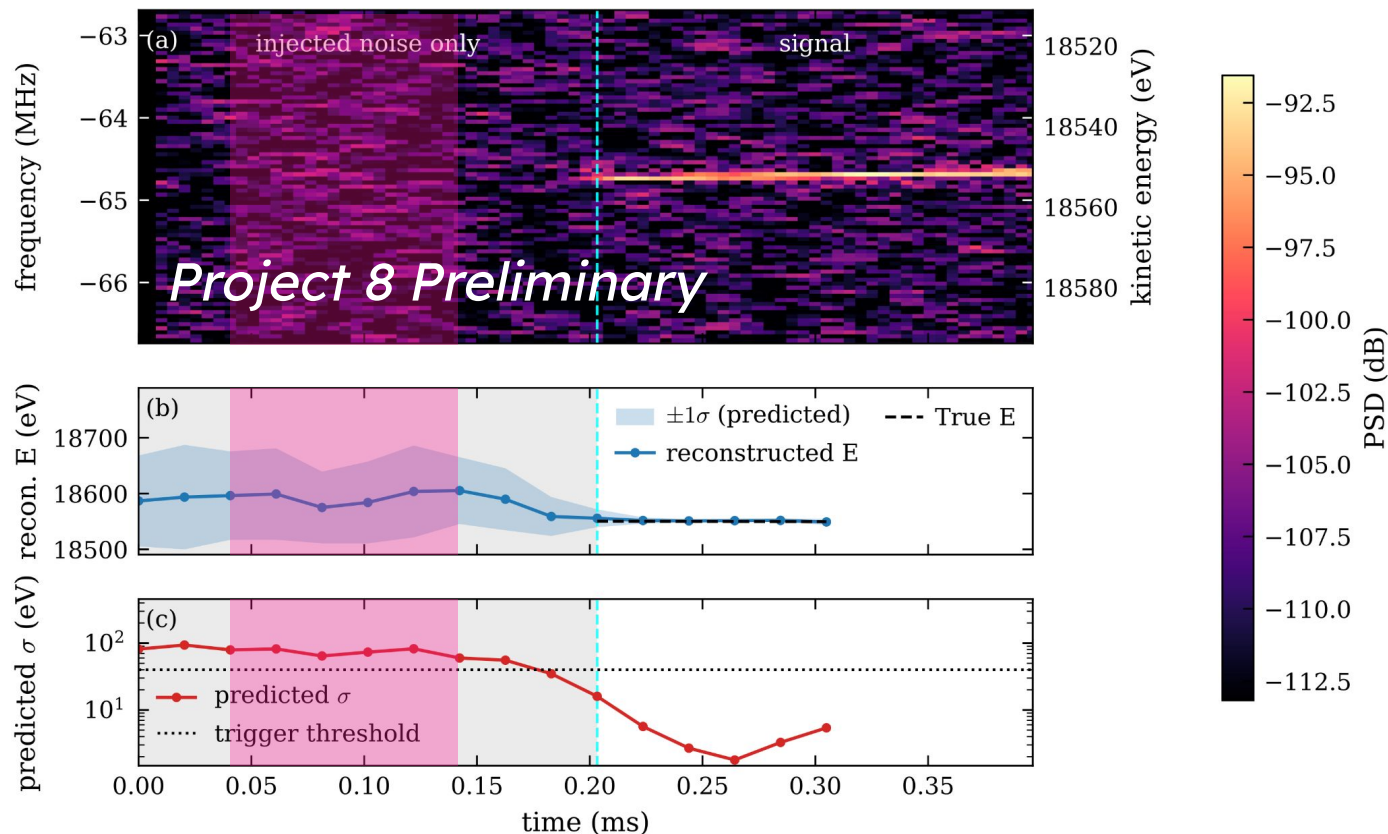
SSM as an event trigger

Trigger on a sudden decrease in reconstructed uncertainty.



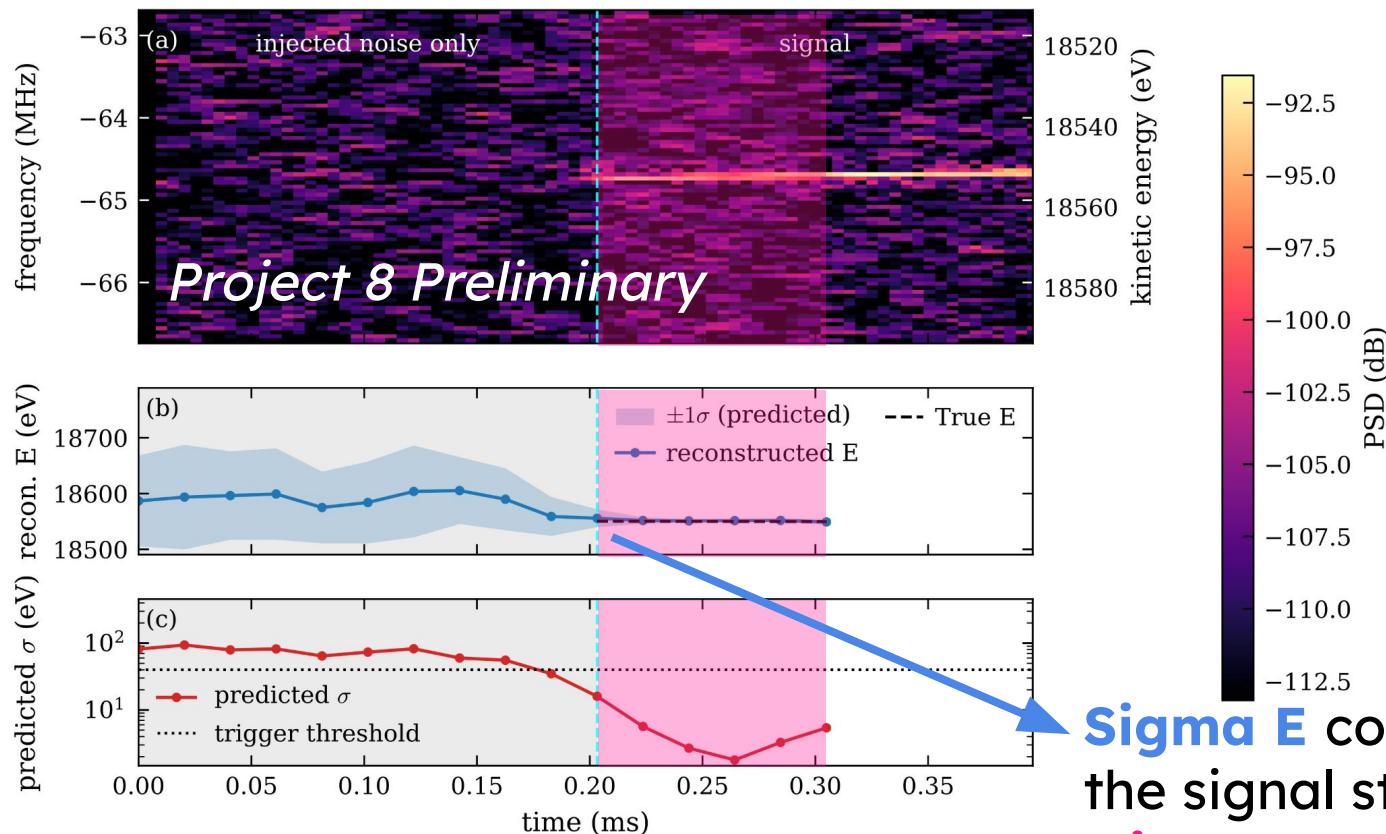
SSM as an event trigger

Trigger on a sudden decrease in reconstructed uncertainty.



SSM as an event trigger

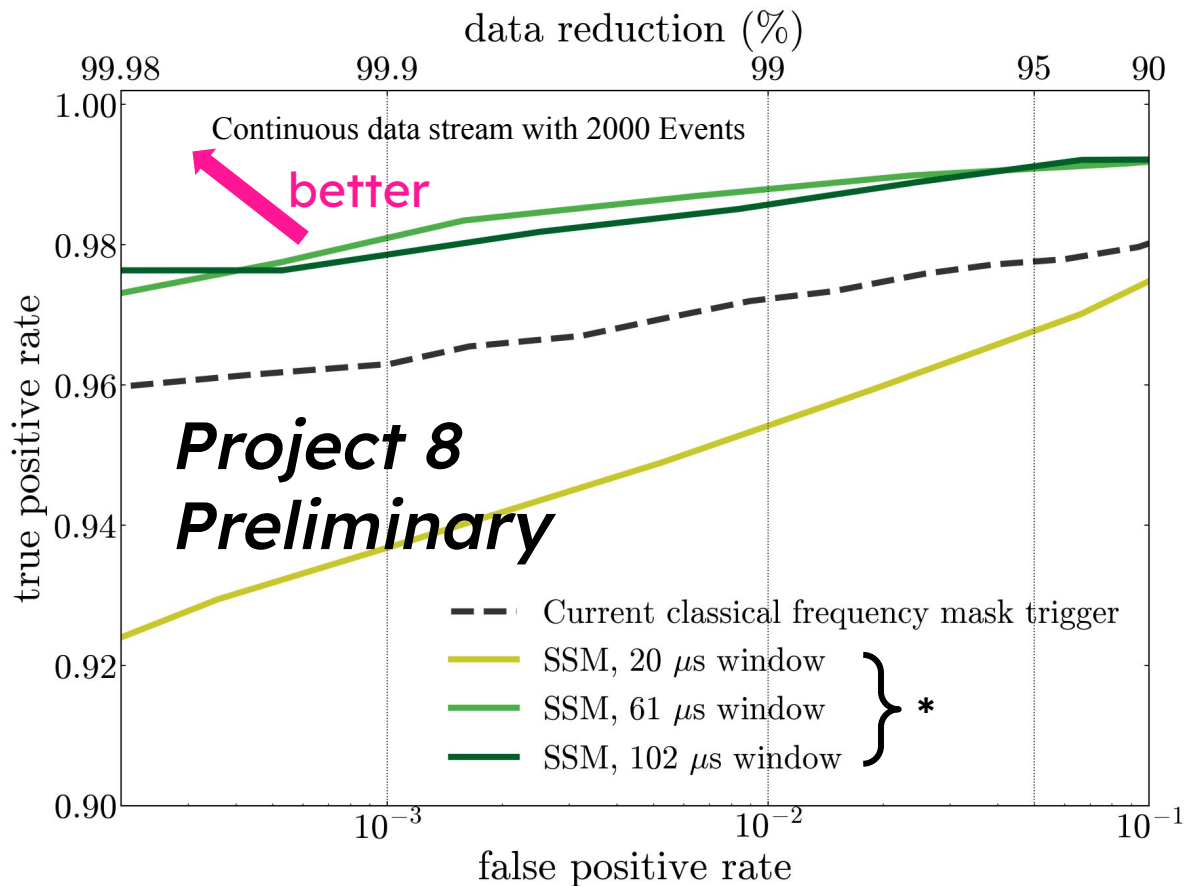
Trigger on a sudden decrease in reconstructed uncertainty.



Sigma E collapses when
the signal starts $\rightarrow$
trigger!

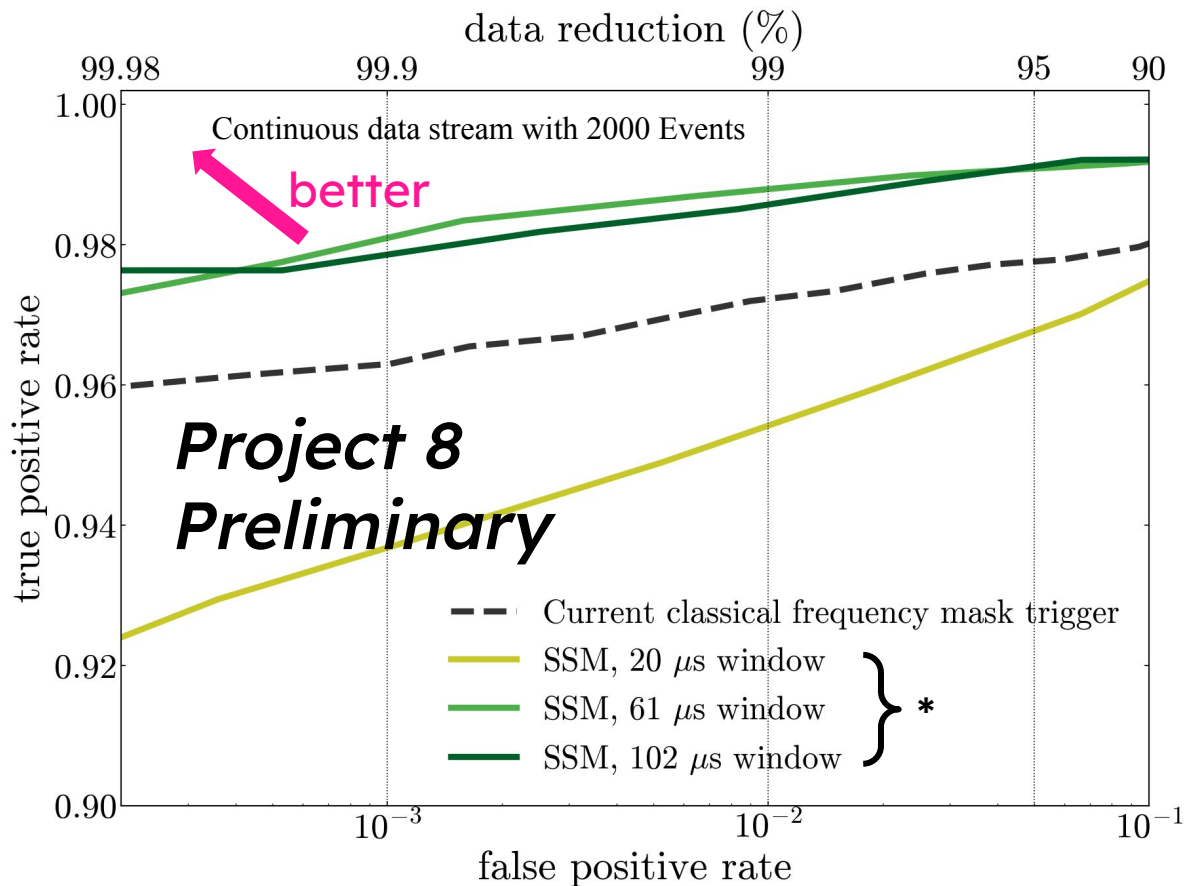
Comparison with classical trigger

- Classical Frequency Mask Trigger (FMT) triggers on frequency bin's power exceeding a threshold
- **Model never trained on noise-only records!**
→ out-of-distribution generalization!



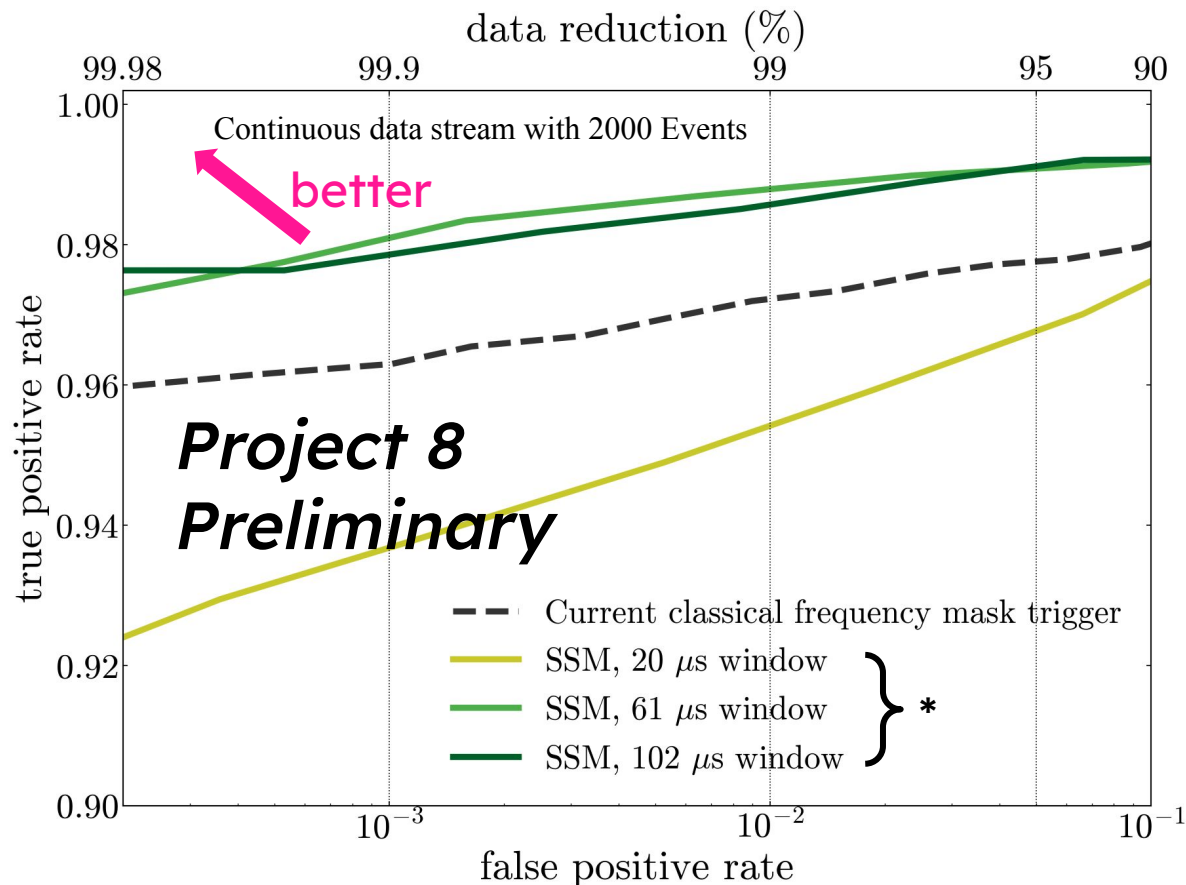
Comparison with classical trigger

- Classical Frequency Mask Trigger (FMT) triggers on frequency bin's power exceeding a threshold
- **Model never trained on noise-only records!**
→ out-of-distribution generalization!
- 61 μ s and 102 μ s SSMs' demands on GPU too high to be ran as a Level-1 trigger



Comparison with classical trigger

- Classical Frequency Mask Trigger (FMT) triggers on frequency bin's power exceeding a threshold
- **Model never trained on noise-only records!**
→ out-of-distribution generalization!
- 61 μ s and 102 μ s SSMs' demands on GPU too high to be ran as a Level-1 trigger
- Work in Progress: implement it as a high-level trigger after the FMT
- *SSM with classical event builder, not pure SSM!



Work In Progress: Path to SSM online triggering and reconstruction

SSM on 1 H200 GPU, batch 256, input data rate 403MHz:

	Throughput	Data reduction rate needed from Level-1 Trigger to be real time on 1 H200
Classical FMT (for comparison only, NOT on GPU)	Real time, 403 MHz	N/A (this is the L-1 Trigger)
SSM 20 μ s window	~7.9 MHz	~98%
SSM 61 μ s window	~2.3 MHz	~99.4%
SSM 102 μ s window	~1.2 MHz	~99.7%

Work In Progress: Path to SSM online triggering and reconstruction

SSM on 1 H200 GPU, batch 256, input data rate 403MHz:

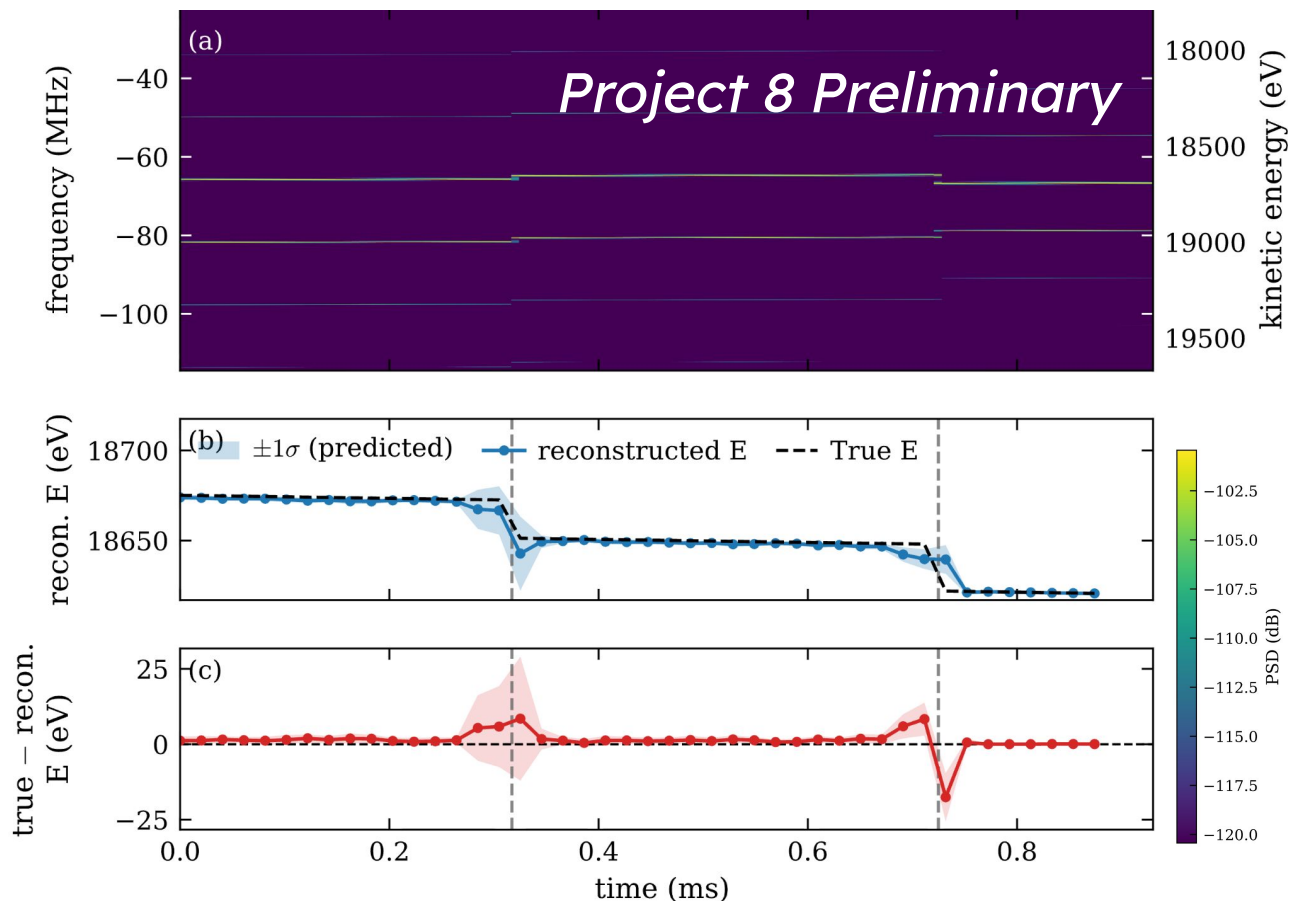
	Throughput	Data reduction rate needed from Level-1 Trigger to be real time on 1 H200
Classical FMT (for comparison only, NOT on GPU)	Real time, 403 MHz	N/A (this is the L-1 Trigger)
SSM 20 μ s window	~7.9 MHz	~98%
SSM 61 μ s window	~2.3 MHz	~99.4%
SSM 102 μ s window	~1.2 MHz	~99.7%

One model can double as High-level trigger + online energy reconstruction after the classical trigger!

Small model size suitable for dedicated FPGA: 2.4 MB in FP32, (not yet quantized) → Weights and state fits in on-chip memory. Work in Progress!

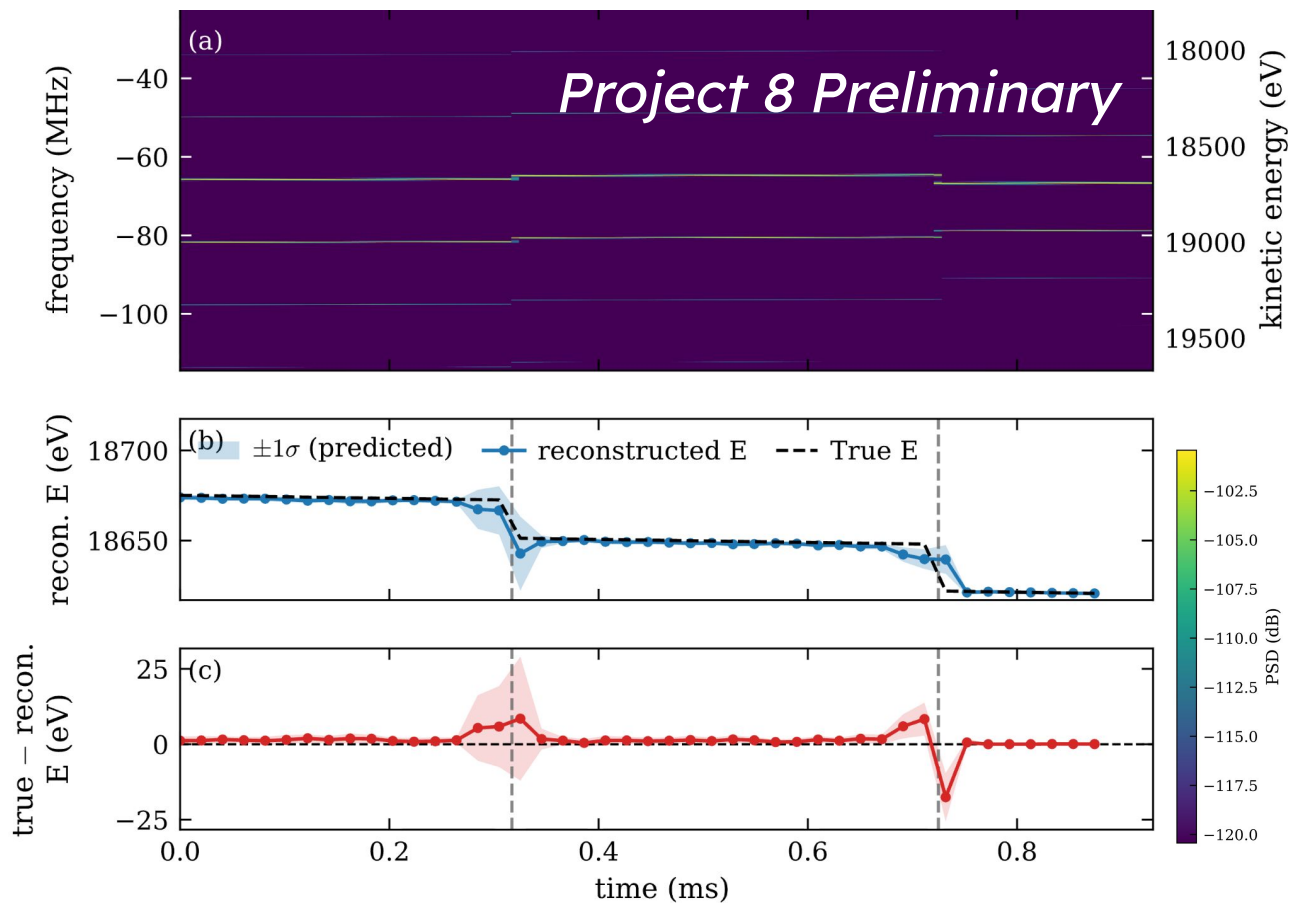
SSM task-specific heads

- Foundation model can be generalized with **specialized task heads** trained on top of it – e.g. detect when a scatter happens in its window



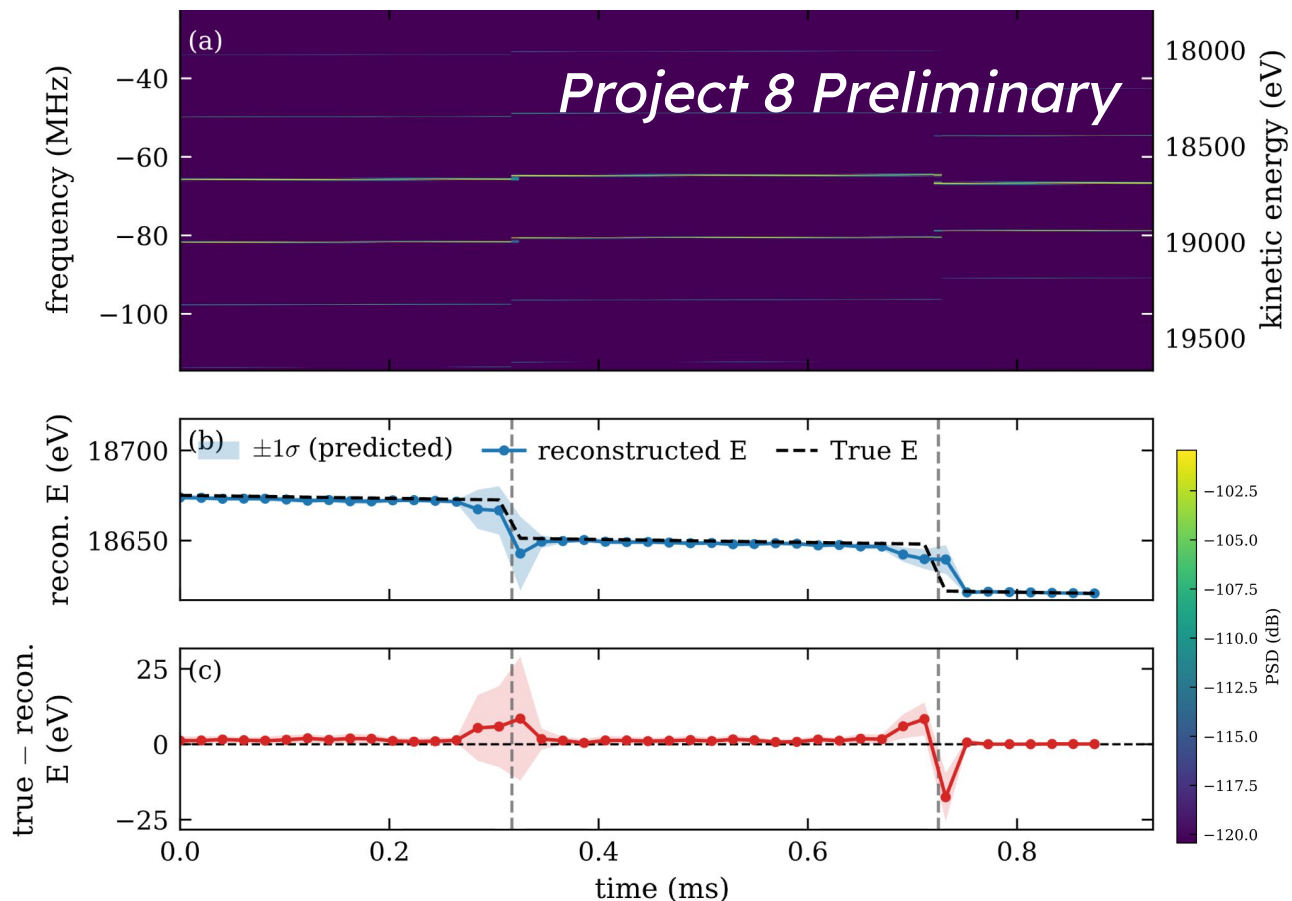
SSM task-specific heads

- Foundation model can be generalized with **specialized task heads** trained on top of it – e.g. detect when a scatter happens in its window
- Electrons scatter due to collisions with gas molecules in our apparatus

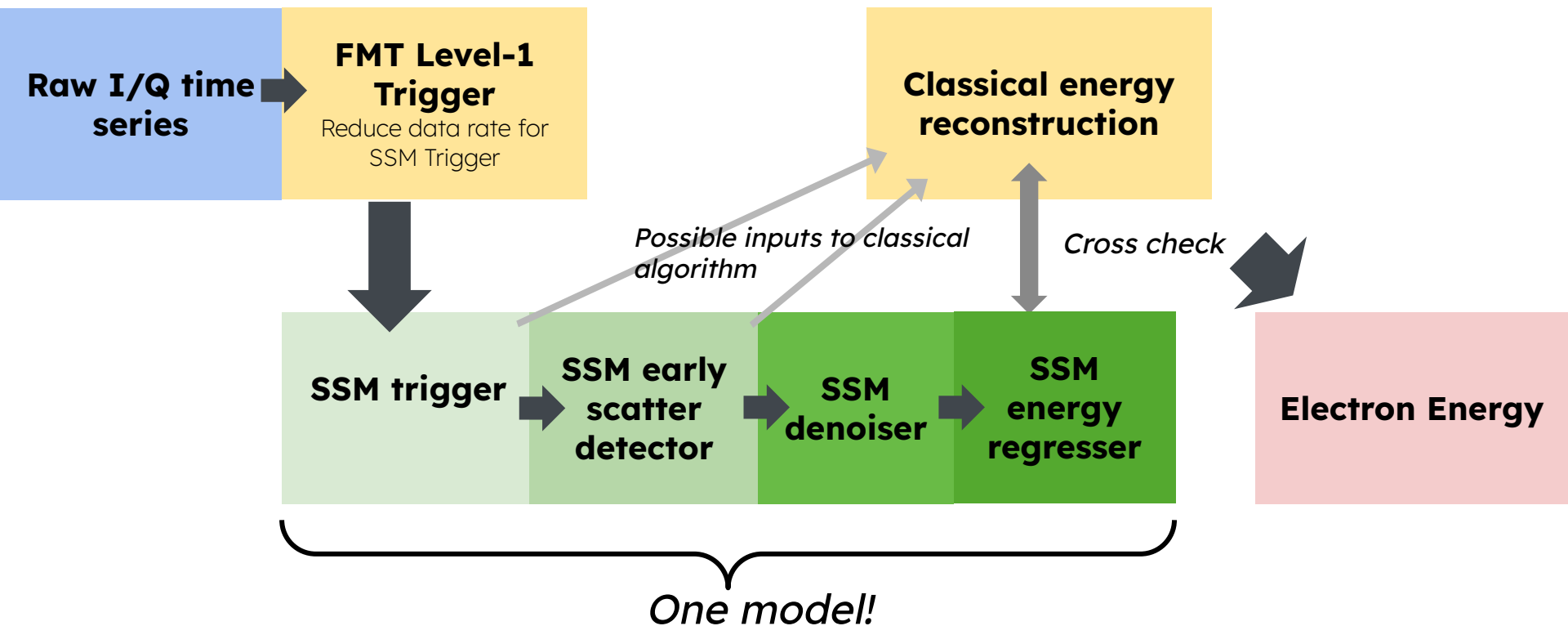


SSM task-specific heads

- Foundation model can be generalized with **specialized task heads** trained on top of it – e.g. detect when a scatter happens in its window
- Electrons scatter due to collisions with gas molecules in our apparatus
- Currently, classical algorithm cannot yet reconstruct events with scattering



Work In Progress: SSM in the DAQ Chain



Conclusion

- Demonstrated SSM architecture on Project 8 signal. Architecture **generalizable** to different sensors and physics experiments dealing with **time series** (See [NeurIPs paper](#))

Conclusion

- Demonstrated SSM architecture on Project 8 signal. Architecture generalizable to different sensors and physics problems dealing with **time series** (See [NeurIPs paper](#))
- Improvement in **both** reconstruction **efficiency & resolution** with a **small network**
 - Size suitable for FPGA implementations

Conclusion

- Demonstrated SSM architecture on Project 8 signal. Architecture generalizable to different sensors and physics problems dealing with **time series** (See [NeurIPs paper](#))
- Improvement in both reconstruction **efficiency** & **resolution** with a **small network**
 - Size suitable for FPGA implementations
- **Reconstructed uncertainty** as tuning: no hand-tuned power or frequency cuts during triggering or analysis steps

Conclusion

- Demonstrated SSM architecture on Project 8 signal. Architecture generalizable to different sensors and physics problems dealing with **time series** (See [NeurIPs paper](#))
- Improvement in **both** reconstruction **efficiency** & **resolution** with a **small network**
 - Size suitable for FPGA implementations
- **Reconstructed uncertainty** as tuning: no hand-tuned power or frequency cuts during triggering or analysis steps
- **Foundation** model **generalizable** to many downstream tasks: **trigger, scatter identification, and reconstruction**
 - Demonstrated **task-specific detector heads** for early-scatter detection
 - Potential to reuse parts of network to save memory and compute for variety of tasks

Conclusion

- Demonstrated SSM architecture on Project 8 signal. Architecture generalizable to different sensors and physics problems dealing with **time series** (See [NeurIPs paper](#))
- Improvement in **both** reconstruction **efficiency** & **resolution** with a **small network**
 - Size suitable for FPGA implementations
- **Reconstructed uncertainty** as tuning: no hand-tuned power or frequency cuts during triggering or analysis steps
- **Foundation** model **generalizable** to many downstream tasks: **trigger, scatter identification, and reconstruction**
 - Demonstrated **task-specific detector heads** for early-scatter detection
 - Potential to reuse parts of network to save memory and compute for variety of tasks
- Computational efficient: fast inference on a single GPU for offline reconstruction
 - WIP: Full online triggering paired with classical trigger

Conclusion



- Demonstrated SSM architecture on Project 8 signal. Architecture **generalizable** to different sensors and physics problems dealing with **time series** (See NeurIPs paper)
- Improvement in **both** reconstruction **efficiency** & **resolution** with a small network
 - Size suitable for FPGA implementations
- **Reconstructed uncertainty** as tuning: no hand-tuned power or frequency cuts during triggering or analysis steps
- **Foundation** model **generalizable** to many downstream tasks: **trigger, scatter identification, and reconstruction**
 - Demonstrated **task-specific detector heads** for early-scatter detection
 - Potential to reuse parts of network to save memory and compute for variety of tasks
- Computational efficient: fast inference on a single GPU for offline reconstruction
 - WIP: Full online triggering paired with classical trigger



Project 8 Collaboration, 2026

With thanks to our funding agencies:

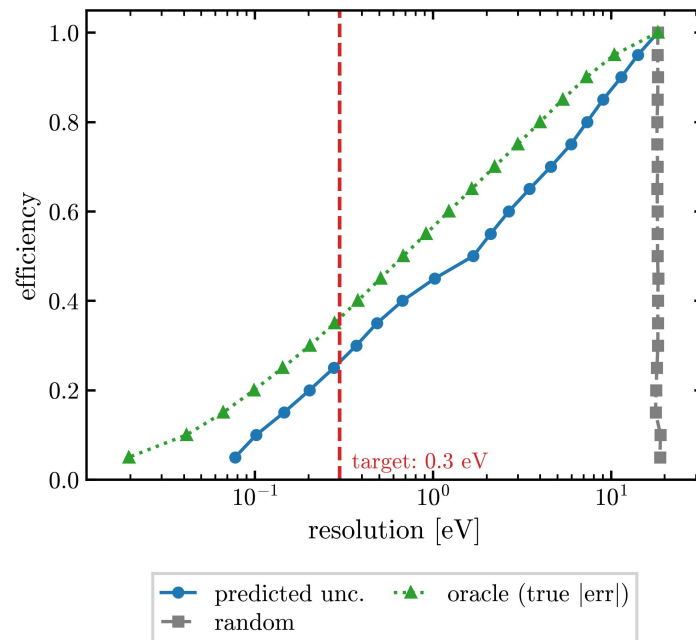
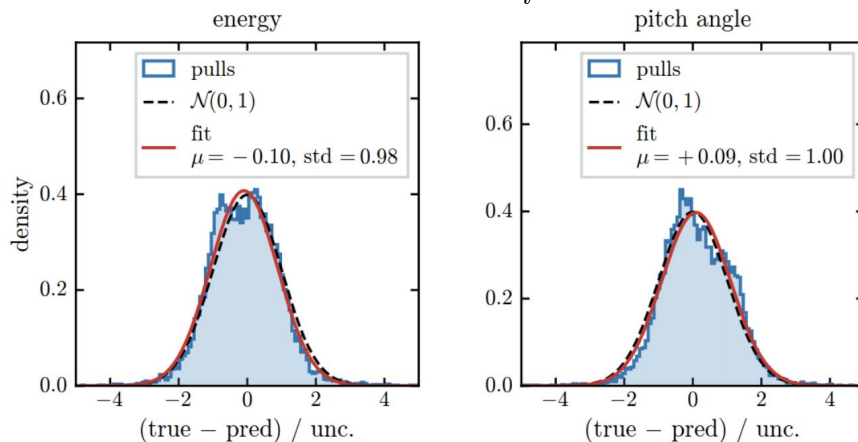


Backup

Uncertainty Calibration

- Pull distribution follow $N(0,1) \rightarrow$ can be used as tuning parameter and **replace manual efficiency cuts**.
- We can **choose a final resolution** and define an uncertainty threshold to **cut out events with high error**.

$$\text{pull}_i = \frac{\theta_{\text{true}} - \hat{\theta}_i}{\hat{\sigma}_i}$$



Signal-to-noise ratio comparison

- Signal-to-Noise ratio (SNR) defined per event as the ratio of summed signal power to noise power in the entire analysis window (5 or 20 bins)
 - longer analysis window has larger SNR
- Average input SNR:
 - Standard algorithm: 73
 - SSM: 38
- **Same resolution at factor ≈ 1.9 lower SNR**
- Physics impact: increase of gas density possible?!

