Dual-Decoder CNN Autoencoder for Unsupervised Anomaly Detection & Denoising in TbT-BPM Data

IN2P3/IRFU Machine Learning workshop, 26-28 Nov 2025

Charles Ndungú Ndegwa[1,2] Andres Gomez[1] Francesca Bugiotti[1] Barbara Dalena[3] Quentin Bruant[3] Valérie Gautard[3]

[1]CentraleSupélec / LISN / CNRS, Université Paris-Saclay
charles.ndungu-ndegwa@universite-paris-saclay.fr

[2]University of Nairobi

[3]DACM/LEDA, CEA

Outline

Roadmap

SuperKEKB context and problem

SOMA baseline

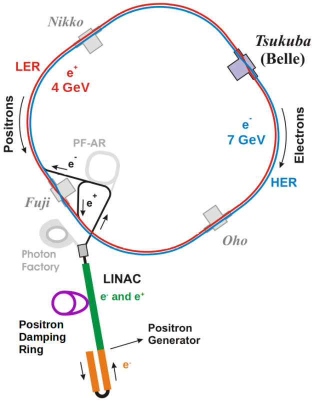
Dual-decoder autoencoder

Training, scoring, FFT alignment

Results
- Anomaly detection results
- Denoising results

Conclusion and Outlook

SuperKEKB as pathfinder for FCC-ee



SuperKEKB / Belle II

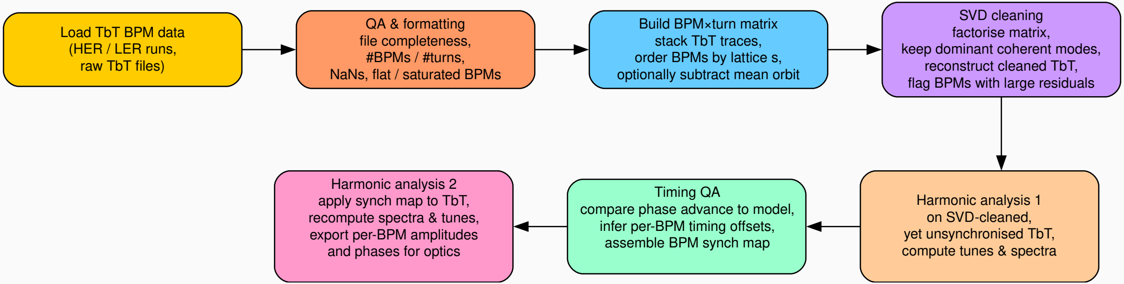
- $\sim$3 km e^+e^- collider at KEK Japan.
- Holds the luminosity record
- **Turn-by-Turn (TbT) BPMs** are key diagnostics:
  - **What:** Record transverse position (x, y) at every revolution period (T_{rev}).
  - **How:** Capture coherent betatron oscillations (excited by kickers/pingers).
  - **Why:** Enable harmonic analysis to extract optics (phase advance, β-functions, coupling).
  - **Challenge:** High noise levels require data cleaning for accurate online analysis.
- Optics and beam parameters close to FCC-ee regime (nano-beams, high current) $\Rightarrow$ test bench for FCC-style diagnostics.

Motivation and challenges

Data challenge	TbT BPM matrices show offsets, spikes, drifts and missing / duplicated turns; some BPMs are partially or fully faulty.
Physics impact	Distorted TbT signals bias tune spectra and phase advance, degrading optics reconstruction and masking true lattice errors.
Operational pain	Manual BPM vetting and ad-hoc cuts do not scale to repeated SuperKEKB optics scans and are unrealistic at FCC-scale BPM counts.
Objective	Build an unsupervised, reproducible module that scores BPM health and outputs denoised TbT data ready for harmonic analysis and optics tools.

Roadmap

SOMA data handling, QA, SVD and harmonic analysis



Keintzel, J. (2024). *SOMA: SuperKEKB Optics Measurement Analysis* [Computer software]. GitHub.
Malina, L., Dilly, J., Hofer, M., Soubelet, F., Wegscheider, A., Coello de Portugal, J., & Tomás, R. (2022). Harpy: A fast, simple and accurate harmonic analysis tool for TBT BPM data. *IPAC 2022 Proceedings*.
pylhc collaboration. (2025). *omc3 0.25.0 documentation*.

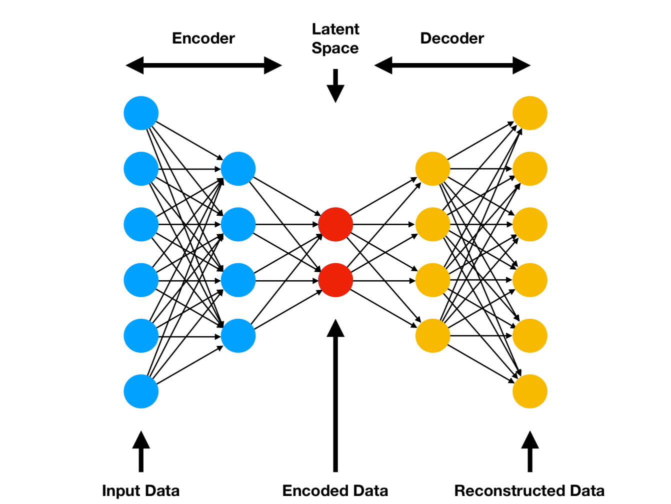
Roadmap

LISN LABORATOIRE INTERDISCIPLINAIRE DES SCIENCES DU NUMÉRIQUE

Autoencoder

Autoencoder principle

- Encoder compresses a TbT window $x(t)$ to a low-dimensional latent vector z.
- Decoder reconstructs $\hat{x}(t)$ from z and is trained to minimise $\|x - \hat{x}\|^2$ on (mostly) healthy data.
- Windows that do not match the learned normal pattern give larger reconstruction error: this becomes an unsupervised anomaly score.



Autoencoder Architecture Choice

Data Constraints

- Input: $N_{BPM} = 68$, $N_{turns} = 4000$.
- Total dimension per sample: $\approx 2.7 \times 10^5$ features.
- Dataset: Several hundred runs/day (multi-day).

Why 1D-CNN over MLP?

- **Curse of Dimensionality:** An MLP input layer would require flattening the matrix, resulting in > 17 million parameters. Computationally expensive, Prone to over-fitting.
- **Sensor Correlation:** We treat the 68 BPMs as **channels**. The 1D-CNN processes all sensors simultaneously, preserving spatial correlations.
- **Translation Invariance:** CNNs detect local temporal patterns (shapes) regardless of global drift across different days.

LISN LABORATOIRE INTERDISCIPLINAIRE DES SCIENCES DU NUMÉRIQUE

Data model and tune definition

TbT matrix

$$\mathbf{X} \in \mathbb{R}^{N \times T}, \quad \mathbf{X} = \big(x_i(t)\big), \; i = 1, \ldots, N, \; t = 1, \ldots, T.$$

Per-channel standardization (computed on a training set):

$$\mu_i = \frac{1}{T} \sum_{t=1}^{T} x_i(t), \quad \sigma_i = \sqrt{\frac{1}{T} \sum_{t=1}^{T} \big(x_i(t) - \mu_i\big)^2},$$

$$\tilde{x}_i(t) = \frac{x_i(t) - \mu_i}{\sigma_i + \varepsilon}, \qquad \tilde{\mathbf{X}} \in \mathbb{R}^{N \times T}.$$

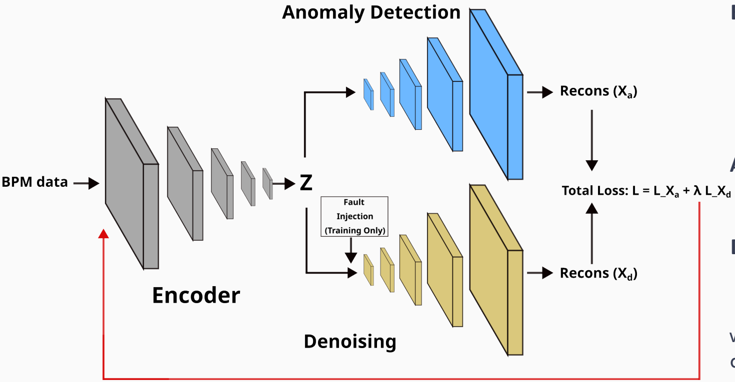
Tune extraction (post-analysis)

$$\mathcal{F}_i(f) = \sum_{t=0}^{T-1} w(t)\, \hat{x}_i^{\,d}(t)\, e^{-j2\pi f t/T}, \qquad Q_i = \arg\max_{f \in \mathcal{B}} |\mathcal{F}_i(f)|.$$

Symbols

- N: number of BPMs; T: number of turns.
- $x_i(t)$: TbT position of BPM i at turn t; μ_i, σ_i: mean and standard deviation for BPM i.
- $\tilde{x}_i(t)$: standardised TbT signal; $\hat{x}_i^{\,d}(t)$: denoised output.
- $w(t)$: Hann window; $\mathcal{B}$: tune search band; Q_i: fractional tune for BPM i.

Architecture: shared encoder and two decoders (DDAE)



Encoder E

- Four Conv1D blocks, kernel sizes $(7, 5, 5, 5)$, stride 2, channels $(16, 32, 64, 128)$.
- Maps $\tilde{\mathbf{X}}$ to a latent tensor Z.

Anomaly decoder D_a

$$\hat{\mathbf{X}}^{\,a} = D_a(E(\tilde{\mathbf{X}})).$$

Denoising decoder D_d

$$\hat{\mathbf{X}}^{\,d} = D_d\big(P(E(\tilde{\mathbf{X}}^{\mathrm{corr}}))\big),$$

where P is a 1×1 projection that limits denoiser capacity.

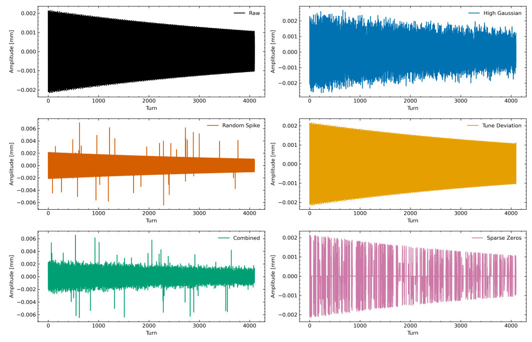
$$\mathcal{L}_a = \frac{1}{B}\sum_{b=1}^{B}\left\|D_a(E(\tilde{\mathbf{X}}_b)) - \tilde{\mathbf{X}}_b\right\|_F^2, \quad \mathcal{L}_d = \frac{1}{B}\sum_{b=1}^{B}\left\|D_d(P(E(\tilde{\mathbf{X}}_b^{\mathrm{corr}}))) - \tilde{\mathbf{X}}_b\right\|_F^2,$$

$$\mathcal{L} = \mathcal{L}_a + \lambda\,\mathcal{L}_d.$$

B: batch size; $\|\cdot\|_F$: Frobenius norm; λ: denoiser weight.

Roadmap

Training corruptions and objective



Examples of synthetic faults applied to healthy TbT traces.

Let $\{\tilde{X}^{(b)}\}_{b=1}^{M}$ be standardized training TbT matrices.

Corruption operator $\mathcal{C}_\omega$

- For each batch element and epoch, draw a configuration ω (fault type + parameters).
- Build corrupted input $\tilde{X}^{(b)}_{\text{corr}} = \mathcal{C}_\omega(\tilde{X}^{(b)})$.

Training objective

$$\mathcal{L} = \mathcal{L}_a + \lambda\,\mathcal{L}_d.$$

DDAE training loop (summary)

Inputs

- Training set $\{\tilde{X}^{(b)}\}$, batch size B.
- Hyperparameters: learning rate η, denoiser weight λ, corruption recipe $\mathcal{C}_\omega$.

Loop

1. Initialise parameters $(\theta_E, \theta_a, \theta_d)$.
2. For each epoch:
   2.1 Sample a mini-batch $\{\tilde{X}^{(b)}\}_{b=1}^{B}$.
   2.2 Build corrupted inputs $\tilde{X}^{(b)}_{\text{corr}} = \mathcal{C}_\omega(\tilde{X}^{(b)})$.
   2.3 Forward passes:
   $$\hat{X}_b^a = D_a(E_{\theta_E}(\tilde{X}^{(b)})), \quad \hat{X}_b^d = D_d\big(P(E_{\theta_E}(\tilde{X}^{(b)}_{\text{corr}}))\big).$$
   2.4 Compute $\mathcal{L} = \mathcal{L}_a + \lambda \mathcal{L}_d$.
   2.5 Adam update step on $\nabla \mathcal{L}$.
3. Early stopping on validation loss; Save the best checkpoint.

Optimisation, anomaly scores, and FFT alignment

Optimisation (training stage)

- Optimiser: Adam or SGD, learning rate η, weight decay, optional LR scheduler (`step`, `plateau`).
- Loss per batch: $\mathcal{L} = \mathcal{L}_a + \lambda\,\mathcal{L}_d$.
- Train up to E epochs; keep the checkpoint with minimum validation loss.

Anomaly scores at test time

- For each acquisition, load raw TbT matrix $X \in \mathbb{R}^{N\times C\times T}$, standardise to $\tilde{X}$.
- Forward pass through anomaly branch: $\hat{X}^{\,a} = D_a(E_\theta(\tilde{X}))$.
- Per-BPM reconstruction error

$$e_i = \frac{1}{CT}\sum_{c=1}^{C}\sum_{t=1}^{T}\left(\tilde{x}_{i,c}(t) - \hat{x}^{a}_{i,c}(t)\right)^2, \qquad i = 1,\dots,N.$$

- These e_i are the `Error` values written to a file and used to rank BPMs (top / bottom N).

FFT alignment (post-processing)

- For selected BPMs, export raw and denoised traces (inverse-scaled to physical units).
- Applies a Hann window, zero padding, and compute FFT-based metrics.

Inference workflow and relation to SOMA

Given a trained model and stored normalisation statistics (μ_i, σ_i):

1. For each test CSV file:
   1.1 Load TbT matrix $X \in \mathbb{R}^{N \times T}$ and standardise to $\tilde{X}$.
   1.2 Run forward pass to obtain $(\hat{X}^a, \hat{X}^d)$.
   1.3 Compute per-BPM errors e_i and save ranked anomaly tables and summary plots.
   1.4 Save raw and denoised traces (inverse-scaled) for selected "good" and "bad" BPMs for FFT analysis.
2. Run SOMA on the same acquisitions to obtain reference tunes and optics.
3. Use the exported traces and FFT metrics to:
   - check that denoising preserves the SOMA tune on clean BPMs (small $|\Delta Q|$),
   - quantify denoiser,
   - identify additional faulty BPMs beyond the provided audit list.

Roadmap

Anomaly detection results

SOMA QA outcome X plane (HER_2024_06_25_19_47_46)

BPM	Known bad	SOMA status
MQD3E18	✓	Removed
MQEAE20	✓	Removed
MQEAE25	–	Removed
MQEAE35	✓	Present
MQR2ORE	✓	Present
MQD3E23	✓	Present

✓= appears in audit list; "Removed" = rejected by SOMA QA.

Reading

- SOMA discards three BPMs here (MQD3E18, MQEAE20, MQEAE25).
- MQEAE35 and MQR2ORE remain in the optics fit.
- MQD3E8 is dropped earlier due to fewer turns.

Use for AE

- These labels provide a light-weight ground truth for anomaly scores.

LISN LABORATOIRE INTERDISCIPLINAIRE DES SCIENCES DU NUMÉRIQUE

Anomaly decision and metrics

Flagging rule

- Rank BPMs by score e_i; flag as anomalous if $e_i > \tau$ (chosen to keep top K BPMs).

Confusion counts

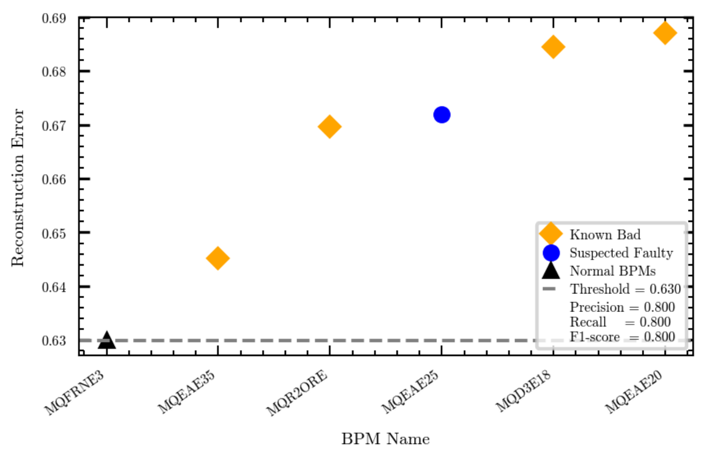
- TP: flagged and truly bad; FP: flagged and truly normal;
- FN: not flagged and truly bad; TN: not flagged and truly normal.

Metrics

$$\mathrm{Precision} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \qquad \mathrm{Recall} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}$$

$$\mathrm{F1} = \frac{2\,\mathrm{TP}}{2\,\mathrm{TP} + \mathrm{FP} + \mathrm{FN}}$$

DDAE vs SOMA (HER_2024_06_25_19_47_46)



Agreement (AE $\cap$ SOMA)

- MQEAE20, MQD3E18, MQEAE25.

AE-only (kept by SOMA)

- MQEAE35, MQR2ORE.

Key point

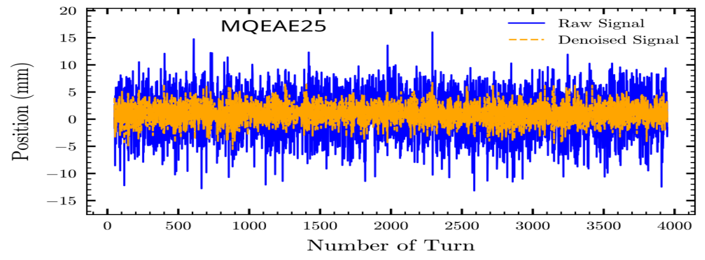
- MQEAE25: flagged by both DDAE and SOMA but not in the provided bad list $\Rightarrow$ candidate new fault.

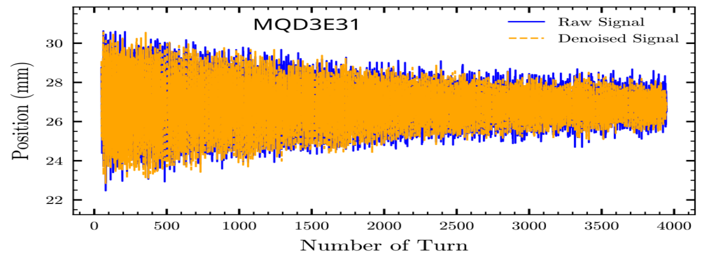
Use in operations

- Investigate BPMs with large DDAE error, even if SOMA accepts them.

LISN
LABORATOIRE INTERDISCIPLINAIRE DES SCIENCES DU NUMÉRIQUE

BPM waveforms: suspected fault vs good





Suspected faulty BPM: MQEAE25

- No clear exponential damping.
- Plateaus and spikes.
- Unstable phase; removed by SOMA.

Example good BPM: MQD3E31

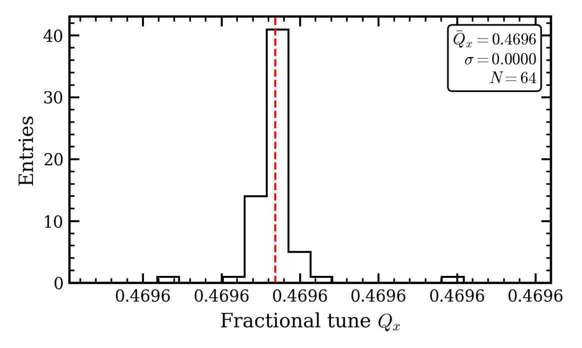
- Smooth decay and stable baseline.

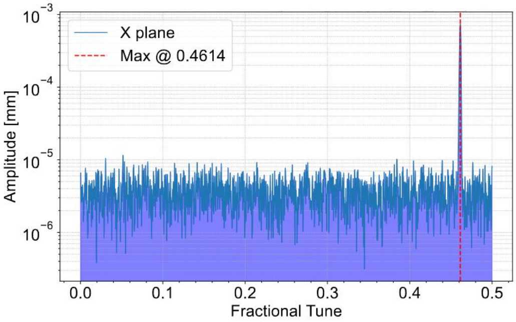
Takeaway

- Visual morphology matches high AE score and SOMA rejection.

Denoising results

SOMA baseline: tune definition and distribution





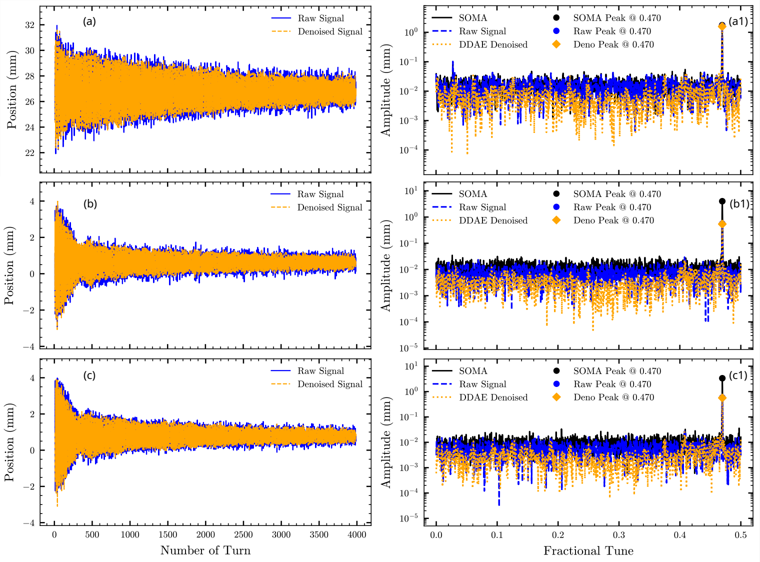
Across BPMs

- Fractional tunes Q_x from SOMA cluster tightly around $\bar{Q}_x \approx 0.4696$.
- This cluster acts as a reference for denoiser performance.

Per BPM

- Tune Q_i is taken as the peak of the HannFFT spectrum.
- Narrow, tall peak over a low floor $\Rightarrow$ reliable tune estimate.

Examples of good BPMs: peak preserved, floor reduced



Time-frequency exemplars (HER X-plane).

(a,a1) MQD3E31, (b,b1) MQI4E, (c,c1) MQI5E.

Left panels (ac) show turn-by-turn waveforms; right panels (a1c1) show FFT spectra.

FFT-based metrics for denoising

For each BPM, we compare the FFT of the raw TbT signal with the FFT of the DDAE-denoised signal.

Peak Retention Ratio

$$\mathrm{PRR} = \frac{\text{amplitude(max peak denoised)}}{\text{amplitude(max peak raw)}}$$

Uses the height of the main betatron peak before/after denoising.

SNR Improvement

$$\mathrm{SNR} = 10\log_{10}\left(\frac{\sum \text{signal power}}{\sum \text{noise power}}\right)$$

$$\Delta\mathrm{SNR} = \mathrm{SNR}_{\text{denoised}} - \mathrm{SNR}_{\text{raw}}.$$

Signal power from the main peak; noise power from the band outside the peak window.

Noise Reduction Ratio

$$\mathrm{NRR} = 1 - \frac{\sum \text{noise}_{\text{denoised}}}{\sum \text{noise}_{\text{raw}}}$$

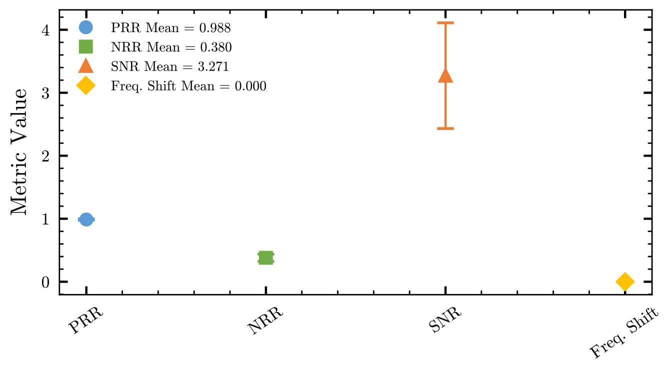
Sums of noise amplitudes outside a small window around the main peak.

Frequency Shift

$$|\Delta Q| = \left| f_{\text{max,raw}} - f_{\text{max,denoised}} \right|$$

$f_{\text{max,raw}}$ and $f_{\text{max,denoised}}$ are the tunes of the main peak before and after denoising.

Denoising metrics: low-error BPMs (healthy group)



Subset

- BPMs with the lowest DDAE reconstruction errors e_i.
- Metrics compare raw vs DDAE-denoised spectra.

Behaviour

- PRR ≈ 1: betatron peak height preserved.
- NRR > 0: broadband floor reduced.
- ΔSNR mildly positive for most channels.
- $|\Delta Q| < 10^{-3}$: tune alignment unaffected.

Conclusion

- The DDAE acts conservatively on healthy BPMs-noise floor is cleaned without altering tune location or amplitude.

LISN LABORATOIRE INTERDISCIPLINAIRE DES SCIENCES DU NUMÉRIQUE

Roadmap

Discussion and operational value

Anomaly detection

- Per-BPM reconstruction errors e_i from the anomaly branch give a ranked health indicator over all BPMs in an acquisition.
- High-error BPMs contain all entries from the provided "bad" list and additional candidates that were not previously flagged.

Denoising (FFTbased metrics)

- *Low-error BPMs (DDAE "good" subset):* PRR ≈ 1, NRR > 0, small positive $\Delta\mathrm{SNR}$, and $|\Delta Q| < 10^{-3}$, so the main betatron peak and tune are preserved while the broadband floor is reduced.

Operational impact

- Fewer manual cuts and more usable BPMs per optics scan after machine changes or upgrades.
- Stored statistics, corruption recipe, and model checkpoints provide an auditable, reproducible QA layer that can scale to FCC-ee BPM counts.

LISN LABORATOIRE INTERDISCIPLINAIRE DES SCIENCES DU NUMÉRIQUE

Next steps

- Run the dual-decoder module alongside SOMA on more SuperKEKB data (HER/LER, X/Y) to validate robustness.
- Model optimization.
- Add simple explainability tools (saliency, error heatmaps) to interpret high anomaly scores.
- Evaluate on simulated FCC-ee optics scans to stress-test scalability.
- Run the denoised data through Optics analysis.

Acknowledgments

Funding and collaborations

EAJADE (MSCA RISE, Grant 101086276)
SuperKEKB operations team (Y. Onishi)
AFD & Expertise France (UoN ESC mobility)

Contact

charles.ndungu-ndegwa@universite-paris-saclay.fr

Thanks

THANK YOU FOR YOUR ATTENTION!