The Application of Machine Learning in Schottky Spectra Analysis

Manuel Bradicic[1,2], *Kacper Lasocha*[1]
[1]CERN, Geneva, Switzerland,
[2]KTH, Royal Institue of Technology, Sweden

Abstract
Schottky Spectra, that is random fluctuations of the macroscopic beam properties, such as beam position or intensity, contain rich information on machine parameters crucial for preserving the beam stability. Historically, the Schottky noise signals have been a powerful diagnostics tool in many accelerators, but in the LHC the Schottky-based diagnostics is difficult due to the presence of strong coherent components in the measured spectra. The goal of this project was to analyse and interpret high-dimensional data of the LHC Schottky spectra using machine learning techniques.

Keywords
Schottky Spectra, Autoencoder, LHC, Beam Diagnostics, Machine Learning

1 Introduction

Schottky spectra span over a broad range of frequencies, however, for the purpose of this work, our focus is solely on a single pair of left and right transverse sidebands. A theoretical, noiseless transverse spectrum is presented in Figure 1. As described in the literature [1–4], the spectrum encodes the beam and machine parameters in the following manner:

1. The betatron tune (Q) determines the centre-of-mass frequencies of the two sidebands,
2. The bunch length ($4\sigma_{bl}$) is reflected in the summed widths of the sidebands,
3. The chromaticity (Q') is encoded in the relative difference of the sidebands' widths,
4. The synchrotron frequency (f_s) determines the distance between the sidebands' subpeaks,
5. The transverse emittance in the plane of measurement (denoted here as ϵ_x) is proportional to the cumulative power of the sidebands and has an impact on the shape and width of individual sidebands' subpeaks,
6. The transverse emittance in the plane other than the plane of measurement (denoted here as ϵ_y) has an impact on the shape and width of individual sidebands' subpeaks,
7. The bunch intensity (N) is proportional to the cumulative power of the sidebands,
8. The current in octupole magnets (I_{oct}) has an impact on the shape and width of individual subpeaks of the sidebands.

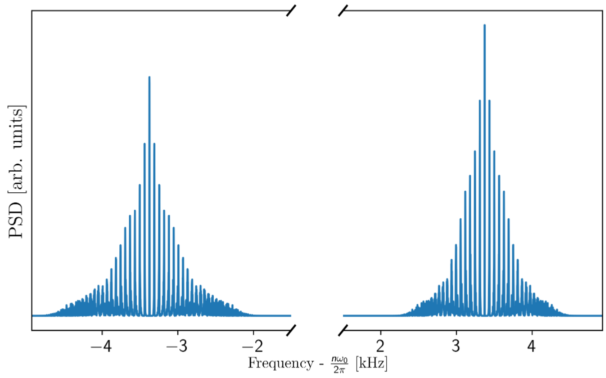


Figure 1: Transverse Schottky spectra.

We see that the multidimensional Schottky spectrum contains information on at least 8 distinct beam and machine parameters, each very important from the perspective of the accelerator operation. Their extraction is however not straight-forward, as the individual parameters interplay with each other. In addition, clean theoretical spectra are in reality distorted by the presence of unwanted spectral components. This study aims to address the challenges linked with the interpretation of the LHC Schottky signals by applying machine learning techniques, specifically autoencoders, to analyse high-dimensional Schottky spectra.

Autoencoders offer the potential to compress the data from a high-dimensional to a low-dimensional representation while preserving meaningful information. In a supervised setup, based on the simulated data, the network can be trained to reduce the dimensionality of the whole spectrum to the aforementioned beam parameters. Through this approach, this study investigates how machine learning can contribute to the advancement of the beam diagnostics in the LHC.

1.1 Outline

Section **2** discusses the methodology used, explaining the concept of an autoencoder, describing the datasets employed for training the models and giving the details on the training process. The following Section **3** provides a quantitative and qualitative assessment of the models' performance on both simulated and real LHC Schottky spectra. The obtained results are discussed in Section **4**, which concludes the study.

2 Methods

In this section, we outline the methodology employed to analyse and interpret high-dimensional data using machine learning techniques. A key focus is the use of autoencoders, which are designed to compress high-dimensional input data into a lower-dimensional latent space and then reconstruct the original data from this compressed representation.

2.1 Autoencoders

The autoencoder neural networks is a flexible and advantageous structure for exploiting low-dimensional features in high-dimensional data [5]. Specifically, it maps the original high-dimensional input vector $\mathbf{x} \in \mathbb{R}^j$ to a low-dimensional latent space $\mathbf{z} \in \mathbb{R}^r$ and then back to the high-dimensional space $\tilde{\mathbf{x}} \in \mathbb{R}^j$. The goal of the autoencoder is to map the output $\tilde{\mathbf{x}}$ back to the input $\mathbf{x}$ itself. It is achieved by minimizing the difference between input-output vectors, i.e. $||\tilde{\mathbf{x}} - \mathbf{x}||_2 \approx 0$ where $r \ll j$.

In our case, the input vector $\mathbf{x}$ of size j represents a combination of Schottky spectrum $\mathbf{x}_s$ and metadata $\mathbf{x}_m$. The metadata vector $\mathbf{x}_m$ contains the well-known parameters of the circular accelerator in the moment of acquisition of the Schottky spectrum, i.e. the bunch intensity and the current flowing in the octupole magnets. The encoding is denoted by:

$$\mathbf{z} = \phi(\mathbf{x}),$$

where $\mathbf{z}$ denotes the latent-space vector, representing the parameters 1-6 from the list in Section **1**. Decoding is denoted by:

$$\tilde{\mathbf{x}} = \psi(\mathbf{z}).$$

The goal of the model training is to find the autoencoder weights θ, that is the parameters of the whole model, such that the reconstructed spectrum $\tilde{\mathbf{x}}$ is as close as possible to the initial input vector $\mathbf{x}$:

$$\arg\min_{\theta} ||\mathbf{x} - \tilde{\mathbf{x}}||_2^2 = \arg\min_{\theta} ||\mathbf{x} - \psi(\phi(\tilde{\mathbf{x}}))||_2^2.$$

If parameters 1-8 from the list in Section **1** are known, as is the case when the studied spectra are synthetically created based on the Schottky theory, we can assess the partial performance of the model in solely encoding the Schottky spectrum into the latent space, without considering the final decoding back to the full-dimensional Schottky spectrum. Such an approach is particularly interesting from the perspective of beam diagnostics, as it reduces model to the fundamental role of deriving beam parameters from the Schottky spectra. In such a case, the model optimization will be based on a reduced loss function:

$$\arg\min_{\phi} ||\mathbf{y} - \mathbf{z}||_2^2,$$

where vector $\mathbf{y}$ represents the parameters describing the spectrum. The general structure of both implementations is outlined in the Figure 2.

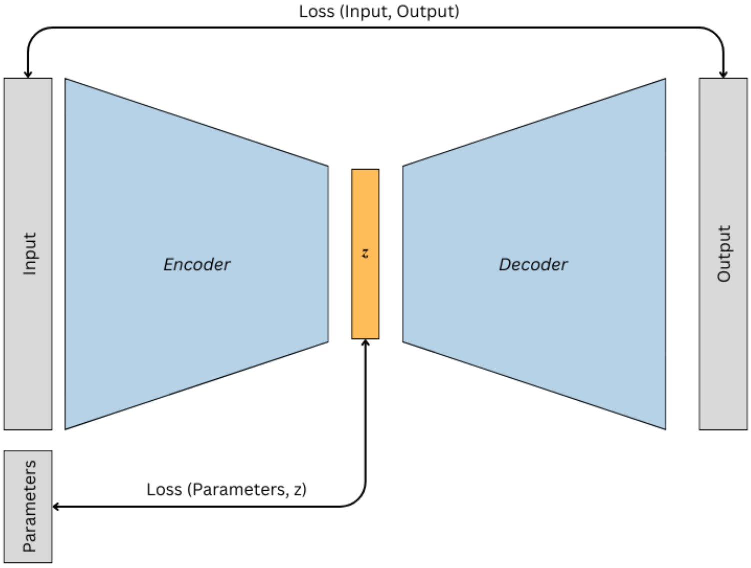


Figure 2: General structure of the implemented autoencoder. Encoders' performance is evaluated by calculating the error between parameters $\mathbf{y}$ describing the spectrum and latent space $\mathbf{z}$, whereas for the whole model, the error between input spectrum $\mathbf{x}_s$ and reconstructed spectrum $\tilde{\mathbf{x}}$.

2.2 Dataset

In this study, two datasets were used: the first, referred to as the LHC Dataset, consists of real Schottky Spectra measured at the Large Hadron Collider, while the second dataset contains artificially generated spectra based on Schottky theory, referred to as Synthetic Dataset. Both datasets contain a series of Schottky spectra, limited to 4094 points in the direct neighborhood of the transverse satellites.

The LHC Dataset consisted of spectra acquired in August 2024 during the STABLE phase of two LHC fills (Fill 9996 and Fill 10001). The dataset represents both measurement planes and two LHC beams, and consists of 3.5×10^5 instantaneous spectra, which after the required averaging reduces to 3.5×10^3 independent samples. An example spectrum from the LHC Dataset is presented in Figure 3.

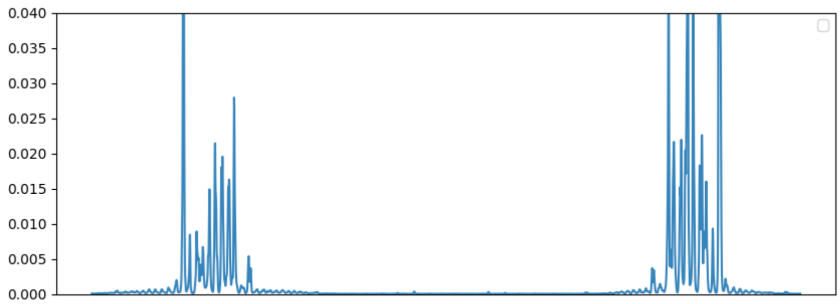


Figure 3: An example spectrum from the LHC dataset.

The Synthetic Schottky dataset was randomly generated based on the Schottky theory and consists of 5×10^6 independent spectra. For each spectrum, the values of beam and machine parameters listed in Section **1** are taken from uniform distributions covering the range of values that could be present at the LHC. To bring a synthetic spectrum closer to real observations, a nosing algorithm was utilized. The nosing function introduced a white noise of a random magnitude to the whole signal, and introduced centrally localized coherent components that imitated the spiking in satellites which could be observed in the real data. An example spectrum from the Synthetic Dataset is presented in Figure 4, while the step-by-step noise introduction procedure is illustrated in Appendix A.

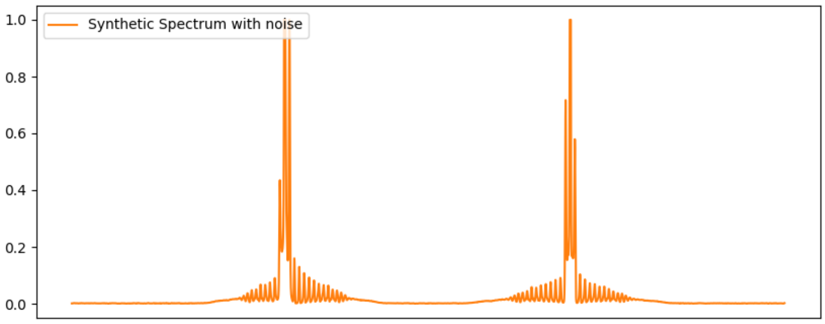


Figure 4: An example synthetic spectrum with noise added.

The primary distinction between the datasets is that the LHC Dataset is unsupervised, while the Synthetic Dataset is supervised. For the LHC Dataset, out of the parameters mentioned in Section **1**, only the octupole current and intensity can be assumed to be precisely known. Other parameters are known with a limited precision and accuracy. For the synthetic dataset, simulated based on known and well-determined parameters, the data is labeled, allowing for the training and evaluation of the spectra encoding into the latent space.

2.3 Experimental Setup

The architecture of the autoencoder proposed in this study comprises of 21 linear layers in total with the input-output dimensions $j = 4096$ as a combination of the input spectra $\mathbf{x}_s$ and the metadata vector $\mathbf{x}_m$, as presented in Figure 5. The metadata vector $\mathbf{x}_m$ represents the bunch intensity and octupole current of the corresponding spectrum, while the latent space z is limited to a total of 6 dimensions, corresponding to parameters 1-6 of the list in Section **1**. Each linear layer of the network employed a dropout rate of 20% and utilized a LeakyReLU activation function. The Adam optimizer was used with a learning rate set to 5×10^{-4}.

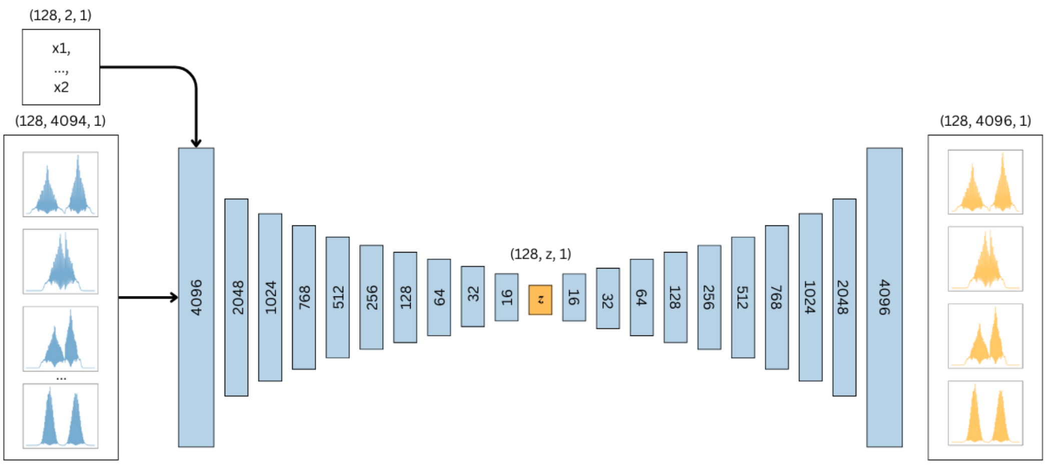


Figure 5: The proposed autoencoder comprises the encoder $\phi(\mathbf{x})$ and decoder $\psi(\mathbf{z})$ - left and right hand side of latent space $\mathbf{z}$, respectively.

The whole datasets were arranged into individual files with a batch size of 32, with the data being randomly shuffled. The training process was run in parallel using 4 concurrent processes (4 as the number of workers) on the Nvidia Tesla T4 GPU, resulting in 128 batches per iteration.

2.4 Training and evaluation

Eight models were trained in this study, differing with the training goals and the used cost function. **Model 1** was trained to simultaneously predict parameters 1-6 of the list given in Section **1**. Then, **Models 2-7** were trained to predict only one of these parameters. The only model that used the decoder part of the network was **Model 8**, which aim was to fully reconstruct the original Schottky spectrum.

All investigated models were firstly trained in a supervised manner using the synthetic dataset, and then qualitatively evaluated on the LHC dataset. The encoder's performance was analyzed based on the root mean square error (RMSE) between the true values and the predictions, with a confusion matrix for each parameter supplementary constructed. The performance of the **Model 8** was assessed based on mean square error (MSE) metric. The analysis of the encoder model's performance on the LHC Dataset was based on a comparison to the results from alternative diagnostic techniques used in the LHC.

3 Results

This section provides an analysis of the results achieved by the models on the Synthetic and LHC Datasets. For Synthetic Dataset, the RMSE metric for parameter prediction is provided. The parameter estimates for the LHC Dataset are compared to the results with other tools used for the diagnostics of the LHC beam parameters.

3.1 Synthetic Dataset

Table 1 displays the achieved RMSE score of an encoder, as a measure of the difference between the true parameters used for the creation of the Synthetic Dataset and the parameters infered from the Schottky spectra. As expected, the predictions of "dedicated" models outperform the general model. The obtained RMSE values are relatively low, with the exception of the transverse emittance in the plane other than the plane of the measurement.

Encoder Quantitative Analyisis						
Model	RMSE					
#	Q'	Q	f_s [Hz]	$4\sigma_{bl}$ [ps]	ε_x [µm]	ε_y [µm]
1	3.12	2.26×10^{-3}	0.761	70.2	0.296	1.41
2	2.01	-	-	-	-	-
3	-	4.31×10^{-4}	-	-	-	-
4	-	-	0.479	-	-	-
5	-	-	-	47.2	-	-
6	-	-	-	-	0.103	-
7	-	-	-	-	-	1.41

Table 1: RMSE per parameter for different models.

The performance of the models can be also monitored by investigation of the confusion matrices, that is 2-D histograms displaying the predicted value versus the ground truth, As seen in Figure 6, the predictions are laying close to the diagonal, showing that the obtained RMSEs are the result of the low precision of the measurement, rather than missing accuracy. Presented below 1-D histograms of the predicted and true values suggest however systematic imperfections of the trained models.

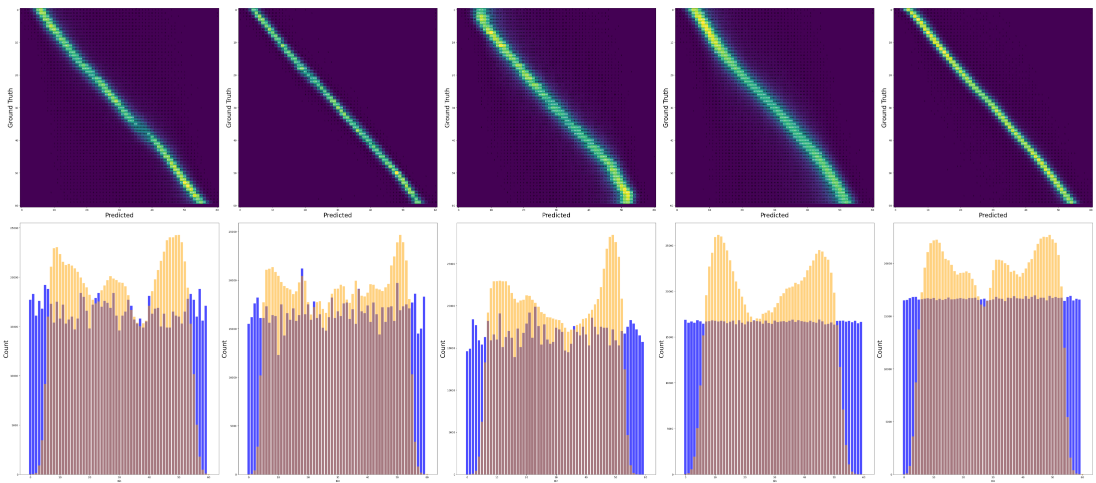


Figure 6: Upper: confusion matrices where, y-axis demonstrated ground truth values and x-axis prediction, in 60 bins. Lower: count per bin for the actual dataset (blue) and the predicted values (yellow). Both rows in the following order, left to right: chromaticity, betatron tune, synchrotron frequency, bunch length, transverse emittance in the plane of measurement.

The performance of a decoder trained on the Synthetic dataset was partially satisfactory. As can be seen in the examples given in Figure 7, the general shape of the spectrum was obtained, but the internal structure of the sidebands was not reconstructed.

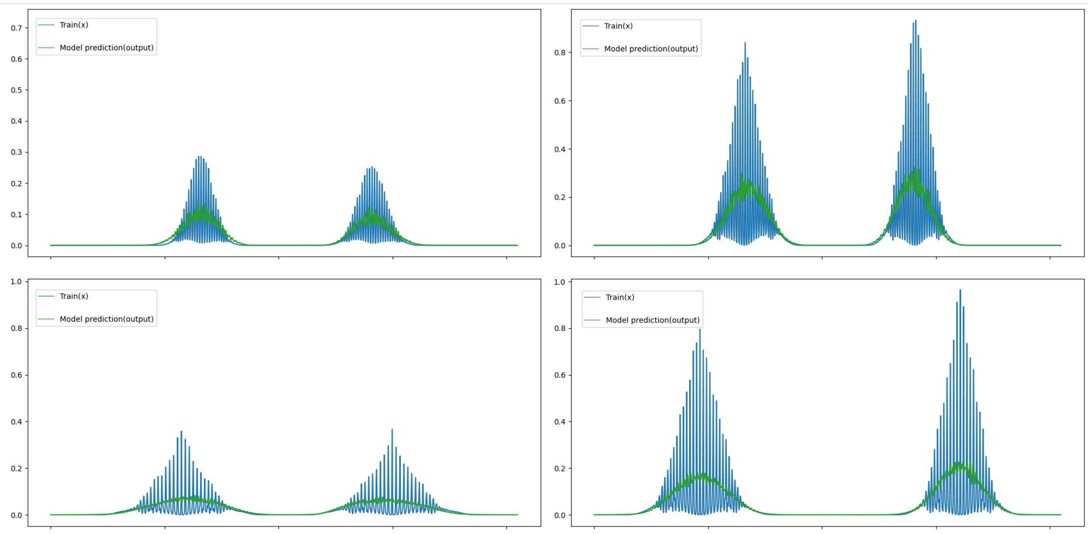


Figure 7: Reconstruction of a synthetic spectrum (in green) from the original spectrum (in blue) using the autoencoder.

3.2 LHC Dataset

Encoder models trained and evaluated on the Synthetic dataset were tested on the LHC dataset. The predictions of the **Models 2 - 6** were compared to the independent estimates of these parameters, resulting in the following values of RMSE:

- The betatron tune (Q): 2.12 $\times10^{-3}$ vs standard Schottky, 3.72 $\times10^{-3}$ vs BBQ,
- The bunch length ($4\sigma_{bl}$): 0.34 ns vs BQM,
- The chromaticity (Q'): 2.73 vs standard Schottky, 3.72 vs setpoint,
- The synchrotron frequency (f_s): 2.58 Hz vs prediction based on RF voltage,
- The transverse emittance (ϵ_x): 3.34 $\times10^{-6}$ vs BSRT

The detailed plots of the measurement in Fills 9996 and 10001 are presented in Appendix B.

4 Discussion and Outlook

The presented study investigated the application of machine learning techniques to analyze the Schottky signals acquired at the Large Hadron Collider and infer from them a set of beam and machine parameters, important from the perspective of beam operation. The findings demonstrated that autoencoders have a potential to be successful in obtaining this goal.

The models employed in this work were initially trained and evaluated using synthetic data, followed by testing on real data collected from the LHC. The results obtained with simulated Schottky spectra can serve as a reference for the ultimate diagnostic potential with the used model architecture. The predictions of chromaticity, betatron tune, synchrotron frequency and the transverse emittance in the plane of the measurement featured a good agreement with the true values. However, the same approach was unsuccessful in predicting the transverse emittance in the plane other than the plane of the measurement. This may suggest that the architecture used is not sufficiently complex and a richer architecture could achieve better results. The decoder, designed to reconstruct spectra from the latent space, also showed a limitation, as it generally only reconstructed only a general shape of the spectrum. This may suggest that the decoder's structure was not sufficiently complex or potentially the choice of loss function favors smoother reconstructions.

The models trained to achieve the lowest RMSE on the synthetic data were further evaluated on the real data obtained at the LHC. The results demonstrated varied performance across different parameters. For instance, the betatron tune yielded consistent and reliable results, though, predictions for beam one generally outperformed those for beam two. The chromaticity estimate was clearly correlated with the target value, but the absolute values deviated within a few units from the setpoint, as well as from the estimate provided by a standard Schottky spectra analysis. In contrast, predictions for bunch length were inaccurate in most cases, and estimates for both synchrotron frequency and transverse emittance failed to track their target values.

The partial success in predicting parameters that describe the real spectra can be explained by several factors. Firstly, the real data may differ significantly from the synthetic data, indicating that the artificial noise used in the synthetic data needs further refinement. This could be addressed by analyzing spectra where the model underperforms to better align the synthetic data with real conditions observed at LHC. Secondly, the model's performance may still be limited by its complexity when it comes to the real data, meaning that the current architecture may not sufficiently capture all the nuances of the real spectra. To address this, it may be necessary to investigate more advanced model architectures as well as systematic hyperparameter tunning. Finally, the theory used to produce the artificial spectra might not be complex enough to describe the conditions present in the LHC. The assumptions of the Gaussian longitudinal bunch profile, no impact of the beam-coupling impedance, and no higher-order chromaticity could limit the performance of the models and their applicability to the LHC Schottky spectra.

Bibliography

[1] Daniel Boussard. Schottky noise and beam transfer function diagnostics. Technical report, CERN, 1986.

[2] Swapan Chattopadhyay. Some fundamental aspects of fluctuations and coherence in charged-particle beams in storage rings. In *AIP Conference Proceedings*, volume 127, pages 467–623. American Institute of Physics, 1985.

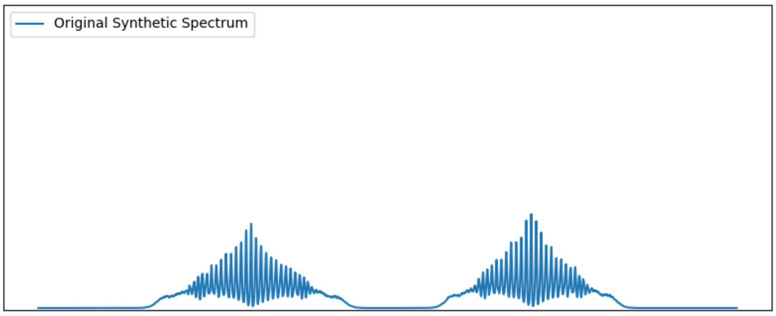
[3] Kacper Lasocha, Christophe Lannoy, Nicolas Mounet, and Diogo Alves. Theory of bunched beam transverse schottky signals with stationary rf system and high-order chromaticity. *Phys. Rev. Accel. Beams*, 27:112801, Nov 2024.

[4] Christophe Lannoy, Diogo Alves, Kacper Lasocha, Nicolas Mounet, and Tatiana Pieloni. Impact of octupoles on the Schottky spectra of bunched beams. *JACoW IPAC*, 2024:WEPG32, 2024.

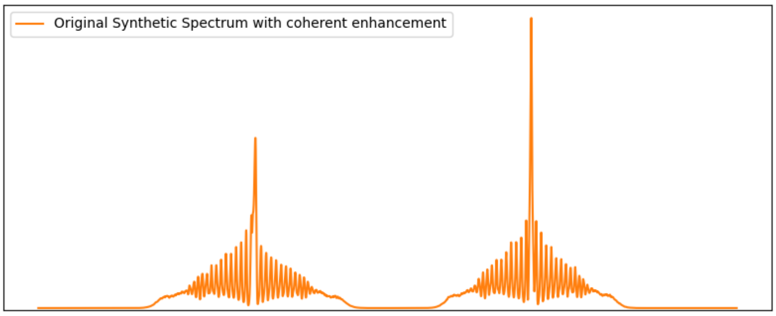
[5] Steven L Brunton and J Nathan Kutz. *Data-driven science and engineering: Machine learning, dynamical systems, and control*. Cambridge University Press, 2022.

Appendix

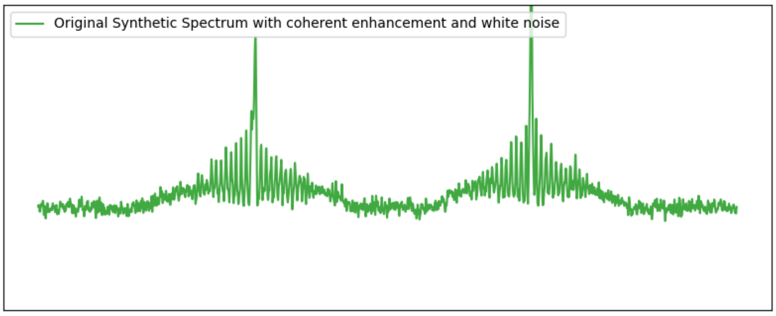
A Progression of Schottky Spectra with Coherent Enhancement and Noise Additions



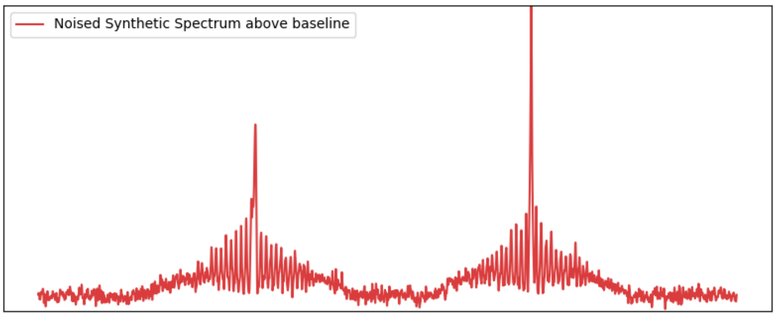
(a)



(b)



(c)



(d)

Figure A.1: Sequence of spectra from (a) the original Schottky spectra to (d) the original Schottky spectra with added noise; including steps of coherent enhancement (b) and white noise in (c).

B Encoder prediction on the LHC dataset

B.1 Fill 9996

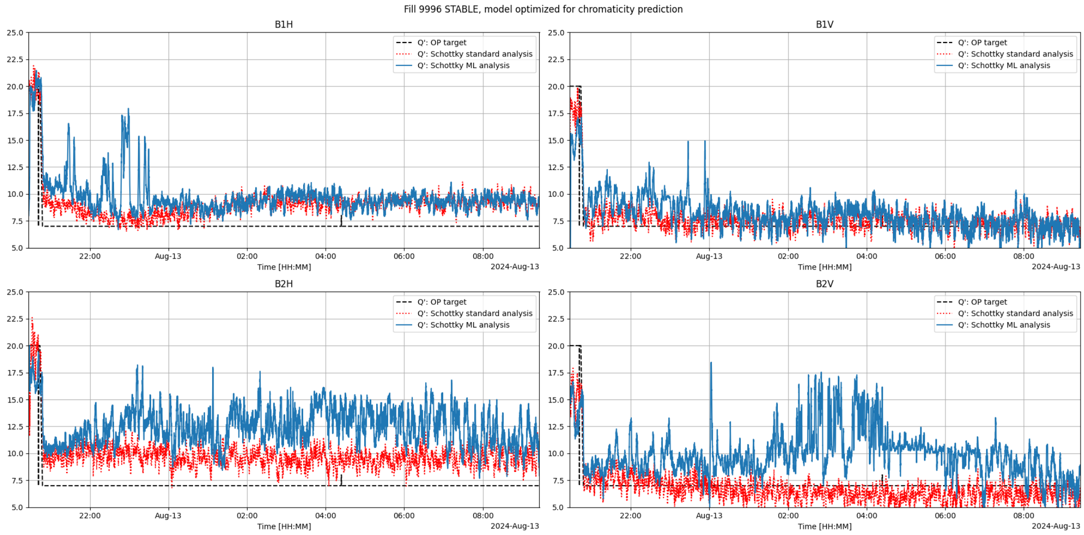


Figure B.1: Chromaticity

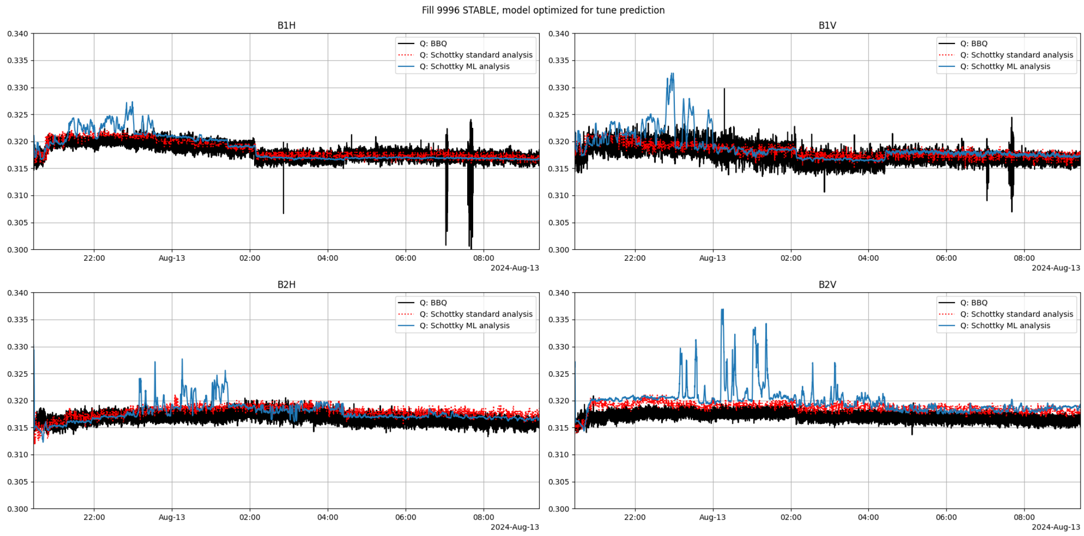


Figure B.2: Betatron tune

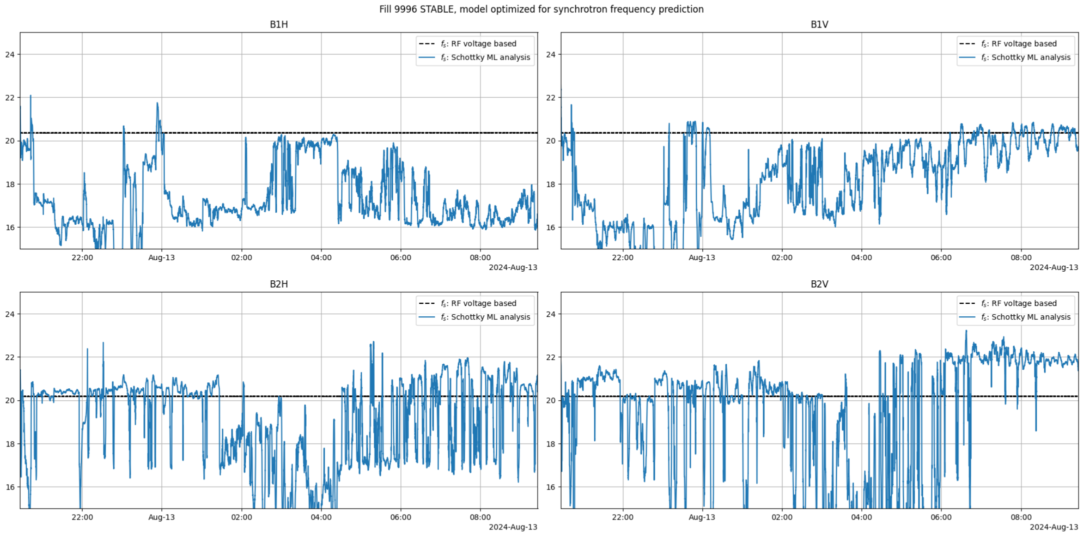


Figure B.3: Synchrotron frequency

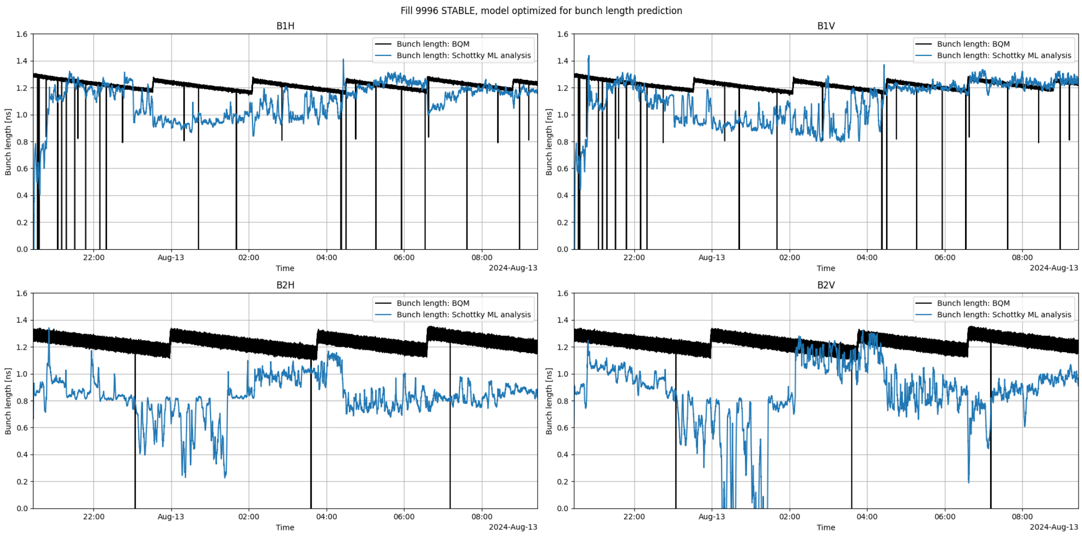


Figure B.4: Bunch length

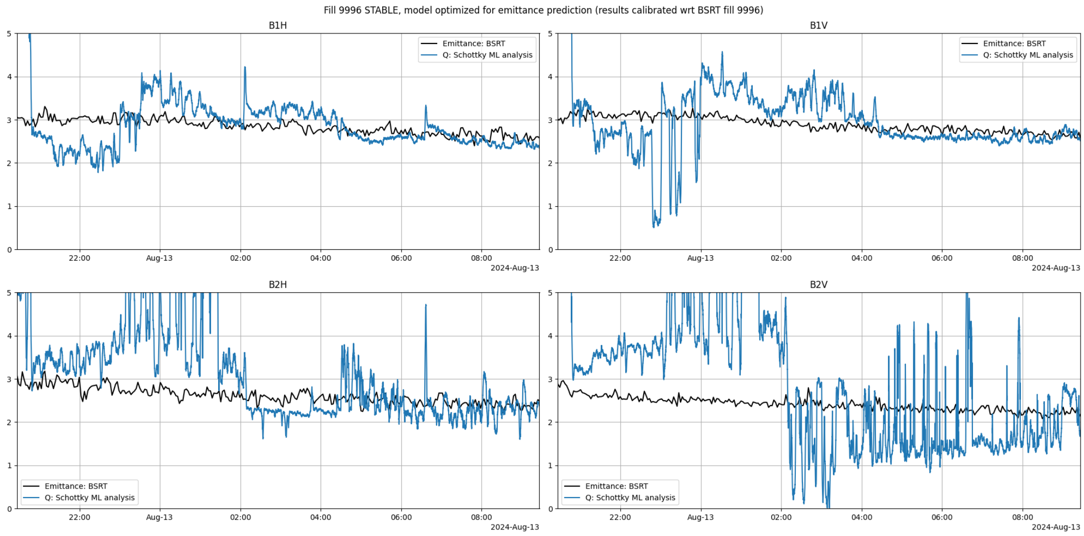


Figure B.5: Emittance

B.2 Fill 10001

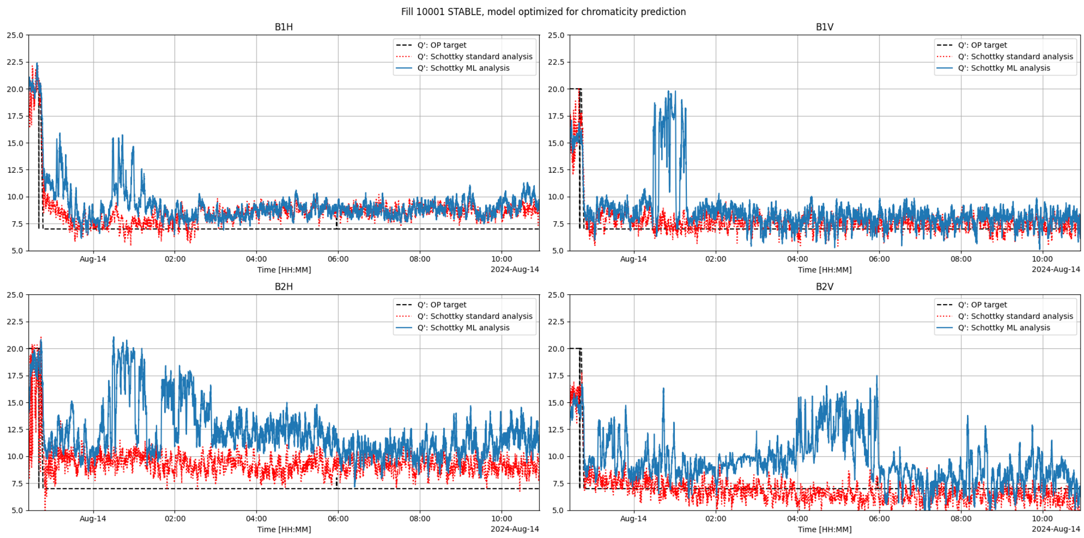


Figure B.6: Chromaticity

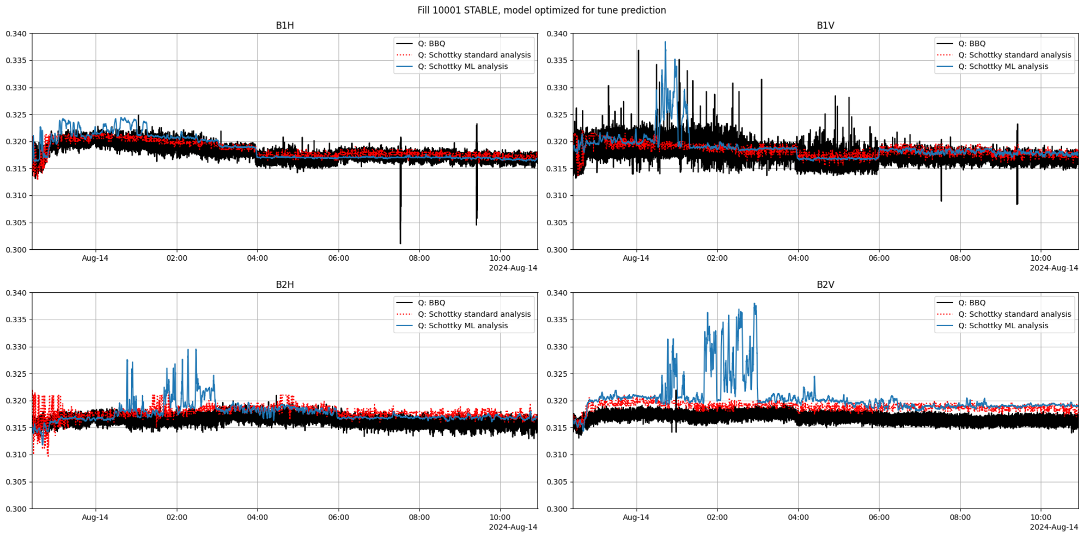


Figure B.7: Betatron tune

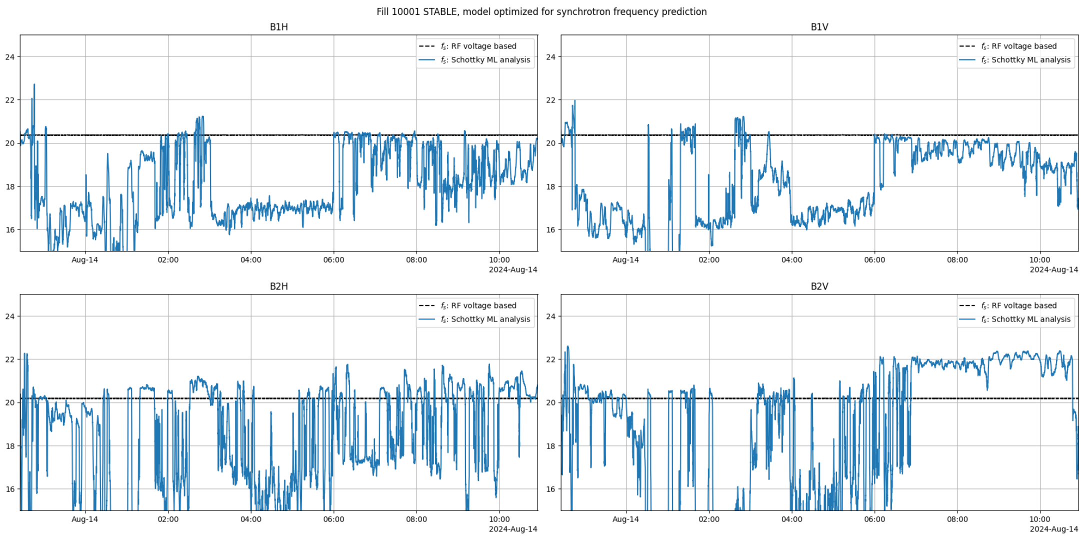


Figure B.8: Synchrotron frequency

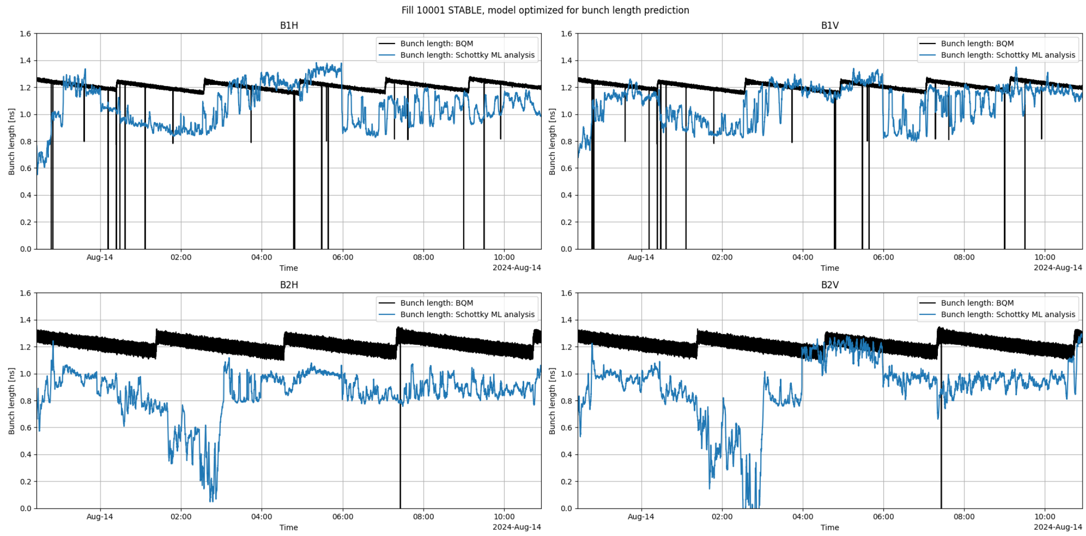


Figure B.9: Bunchlength

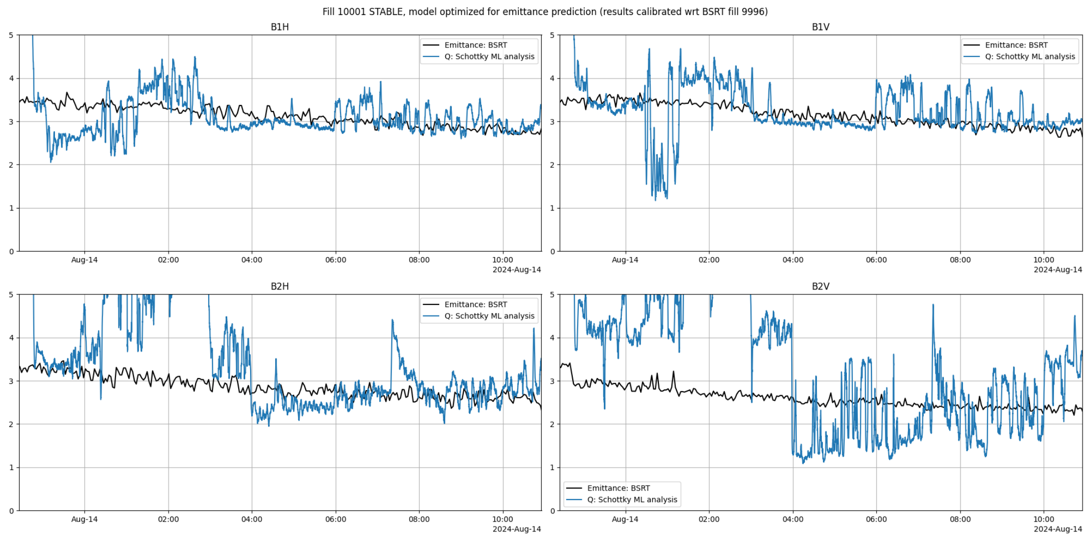


Figure B.10: Emitance