Università degli Studi di Padova

DIPARTIMENTO DI FISICA E ASTRONOMIA "GALILEO GALILEI"
Corso di Dottorato di Ricerca in Fisica

Tesi di Dottorato

Study and development of SPS slow extraction schemes and focusing of secondary particles for the ENUBET monitored neutrino beam

Candidato:
Michelangelo Pari

Relatori:
Prof. Andrea Longhin
Dott. Francesco Maria Velotti

Correlatore:
Prof. Gianmaria Collazuol

Anno Accademico 2019-2020

Abstract

The Enhanced NeUtrino BEams from Kaon Tagging (ENUBET) project aims at developing a first "monitored" neutrino beam, in which the neutrino flux could be measured with a $\mathcal{O}(1\%)$ precision. To do so, the secondary particles decay tunnel will be fully instrumented with compact segmented calorimeters, with the goal of tagging each $K^+ \longrightarrow \pi^0\, e^+\, \nu_e$ (K_{e3}) decay. As a consequence of this, a high precision measurement of the ν_e cross-section could be performed, since the uncertainty on the neutrino flux represents the current main limitation. A full instrumentation of the decay tunnel significantly influences the requirements on the extraction of the primary protons. The pile-up in the detectors rules out the fast-extraction scheme, and calls for the use of the slow resonant extraction, in which continuous spills of the length of several seconds are extracted from the accelerator. Such a long spill would require the use of static focusing devices on the beamline. However, given the low number of produced ν_e, speeding up the cross-section measurement by resorting to a magnetic horn for improved focusing represents an appealing idea. This would require to modify the slow extraction scheme in order to produce a new "pulsed" version of it, with pulse lengths of the order of some millisecond.
In the present work, such a pulsed slow extraction scheme is designed, implemented, and tested at CERN-SPS, with the goal of proving its feasibility according to the requirements of ENUBET. The obtained experimental results are validated with simulations and future possible improvements are investigated. In connection to this, a dedicated study on the frequency response to magnet ripples of the slow extraction process is undertaken: this problem is strictly connected with the performance of the pulsed slow extraction, and can also significantly contribute to improve the standard continuous-spill operation of the experiment (and any other fixed target facility). Both measurements and simulations are used to characterize the process and propose meaningful improvements. Finally, a framework for the simulation and optimization of the ENUBET magnetic horn is developed. This is used to start the investigation of the potential flux gain which could come from the use of the magnetic horn in the ENUBET beamline, coupled with the pulsed slow extraction.

Sommario

Il progetto Enhanced NeUtrino BEams from Kaon Tagging (ENUBET) ha come obbiettivo lo sviluppo di un fascio di neutrini "monitorato", in cui il flusso di neutrini in uscita possa essere misurato con una precisione dell'ordine del 1%. Per fare ciò, il tunnel di decadimento per le particelle secondarie sarà completamente instrumentato, con l'intento di misurare ogni decadimento $K^+ \longrightarrow \pi^0\, e^+\, \nu_e$ (K_{e3}). Questo implicherebbe la possibilità di portare a termine una misura ad alta precisione della sezione d'urto dei neutrini elettronici; infatti, ad oggi, la principale limitazione su tale misura è l'incertezza sul flusso di neutrini prodotto. Equipaggiare l'intero tunnel di decadimento con rivelatori di particelle pone dei vincoli significativi su come debbano venire estratti i protoni primari. Infatti, il tipico caso della estrazione veloce comporterebbe la saturazione dei rivelatori. Questo fa sì che sia necessario utilizzare la estrazione lenta (slow resonant extraction), in cui un flusso di protoni continuo viene estratto in un intervallo di tempo di alcuni secondi. Con tempi di estrazione così lunghi, utilizzare elementi di focalizzazione statici (cioè non impulsati) sembra essere l'unica opzione. Tuttavia, dato il basso numero di neutrini elettronici prodotti, velocizzare la misura della sezione d'urto utilizzando un cosiddetto "magnetic horn" per aumentare il numero di particelle focalizzate è una idea allettante. Fare questo significherebbe modificare lo schema di estrazione dei protoni in modo da creare una versione "impulsata" della estrazione lenta, con impulsi della lunghezza temporale di pochi millisecondi.
In questo lavoro, tale schema di estrazione lenta impulsata sarà progettato, sviluppato e testato al CERN-SPS, con l'obbiettivo di dimostrarne la fattibilità, compatibilmente con i vincoli posti da ENUBET. I risultati sperimentali ottenuti saranno validati con simulazioni, e verranno considerate e studiate nuove possibilità di miglioramento non testate all'acceleratore. Relativamente a questo, sarà effettuato uno studio della risposta in frequenza del processo di estrazione lenta. Questo problema, oltre ad essere connesso con i risultati sul lavoro dell'estrazione impulsata, è anche molto utile per migliorare la qualità dell'estrazione lenta nominale per ENUBET (e per qualsiasi altro esperimento che la utilizzi). Sia misure che simulazioni saranno usate per caratterizzare al meglio questo processo e proporre dei possibili miglioramenti. Infine, verrà sviluppato un insieme di simulazioni volte allo studio e all'ottimizzazione del "magnetic horn" per ENUBET. Questo permetterà di iniziare a valutare il possibile aumento di flusso derivato dall'utilizzo di tale strumento nella linea di fascio di ENUBET, accoppiato all'estrazione lenta impulsata.

ἐπάμεροι· τί δέ τις; τί δ΄ οὔ τις;
σκιᾶς ὄναρ ἄνθρωπος. ἀλλ΄ ὅταν
αἴγλα διόσδοτος ἔλθῃ, λαμπρὸν
φέγγος ἔπεστιν ἀνδρῶν καὶ
μείλιχος αἰών.

Πίνδαρος, Πύθια VIII

Contents

List of Figures

List of Tables

Chapter 1

Motivation and outline

The estimation of the produced neutrino flux from accelerator neutrino experiments has always been a challenging process, which makes up for a significant part of the systematic budget. While for observations of neutrino oscillations in disappearance mode (e.g.. $\nu_\mu \rightarrow \nu_\mu$) this error component can be reduced by the employment of a near-far detector system, for the new era of appearance mode oscillation experiments (e.g. $\nu_\mu \rightarrow \nu_e$) aimed at the detection of leptonic CP-violation this becomes more complicated, making this error an obstacle for the discovery.

A possible solution to the problem is proposed by the Enhanced NeUtrino BEams from Kaon Tagging (ENUBET) project. The goal of ENUBET is to develop a so called "monitored" neutrino beam, in which the decay tunnel is instrumented with fine-grained ($\mathcal{O}(10\ \mathrm{cm}^2)$) detectors in order to estimate the neutrino flux with a direct measurement of the neutrino production vertices decay products. This would allow to decouple the neutrino flux estimation from everything which is upstream of the decay tunnel (e.g. hadron production simulations, hadron production data and measurements etc.), and which usually significantly contributes to its uncertainty. It has been estimated that in such a way the measurement precision could reach the 1% level, which would allow for a direct short baseline cross-section measurement at the same order of uncertainty. This result would be of particular importance for the electron neutrino, for which precise cross-section measurements at the energy range typical of oscillation experiments (a few GeV) are missing.

From initial simulations ENUBET has shown that, among all the components of the secondary beam traveling along the decay tunnel, not only charged kaons are the main source of positrons through the K_{e3} decay channel (e.g. $K^+ \longrightarrow \pi^0\, e^+\, \nu_e$), but the produced positrons are also emitted at an angle significantly higher than all the other decay products of all the particles in the beam. This makes the positron from K_{e3} a significant fraction of all the particles hitting the decay tunnel walls. The practical goal of ENUBET

is then to discriminate each positron hitting the tunnel walls against any other event: this would allow for a direct estimation of the electron neutrino flux, potentially reaching the 1% precision.

The main problem related with the full decay tunnel instrumentation is pile-up: in order to maintain an efficient tagging of the positrons, the rate of particles hitting the target needs to be low enough ($\lesssim 500$ kHz/ cm^2). As a consequence, the fast-extraction of the primary protons typically employed in neutrino beams cannot be operated, since it would completely saturate the detectors. The slow resonant extraction becomes then the primary proton extraction method for ENUBET; in this case, intensities up to 10^{13} protons could be continuously extracted in a few seconds, as opposed to tens of microseconds for the fast-extraction case. One of the significant advantages of the fast extraction would be lost: the possibility of focusing the beam using a magnetic horn. The latter device is a pulsed focusing element (which cannot usually sustain pulses longer than a few ms) with a superior angular and momentum acceptance with respect to common static focusing elements as quadrupoles. With a nominal slow extraction, only static elements could be used on the beamline, and taking an electron neutrino cross-section measurement at the 1% of statistical error could take a significant amount of time, especially taking into account that charged kaons make up for only a small fraction of secondary flux, and the K_{e3} branching ratio is about 5%. For this reason, while keeping the development of a static beamline as the baseline option, ENUBET has decided also to investigate the possibility of using a magnetic horn, so to potentially speed up the measurement. The basic idea for a horn operation of ENUBET would be to modify the slow extraction scheme to a new pulsed version of it. By slow extracting a few millisecond-long proton pulses repeated during the $\mathcal{O}$(1s) nominal extraction time, a magnetic horn could potentially be used without exceeding the pile-up threshold, but still increasing the integrated flux due to its superior acceptance.

The theoretical design and operational implementation of this new type of extraction, together with the investigation of the possible horn-based operation of ENUBET, are the core topics of this Ph.D. work. In particular, the CERN-SPS is the accelerator used to develop and test the new extraction method. The SPS provides its main physics users (the experiments at the North Area) with a slow-extracted spill of protons at 400 GeV. This is ideal for the production of a significant amount of kaons at the design momentum of 8.5 GeV requested by ENUBET. According to the pile-up threshold estimations, the requested extracted pulse lengths would need to be between 2 and 10 ms, with a repetition rate of 10 Hz. These requirements are addressed with dedicated beam dynamics simulations and beam measurements, targeting the optimization of the temporal spill structure. The outcomes from these studies are then exploited to characterize and improve the spill quality of the standard slow extraction. Finally, a systematic investigation of the possible

kaon-flux gains which can be achieved with the horn, with respect to the static focusing case, is carried out, exploring potential horn geometries for ENUBET.

The structure of the work is organized as follows. In Chapter 2, the basic physics concepts are introduced, together with an overview of the ENUBET project and the SPS with its slow extraction. The implementation, experimental results, and simulation study of the new pulsed slow extraction method are detailed in Chapter 3. Chapter 4 is dedicated to the study of the frequency transfer process of the slow extraction, in terms of frequency ripples on the magnets. In Chapter 5, a full simulation and optimization framework for the magnetic horn is presented. Finally, conclusions and possible outlooks are discussed in Chapter 6.

Chapter 2

Introduction

2.1 Neutrino physics

The importance and relevance of neutrino physics as a probe for the Standard Model of particle physics (and beyond) is growing steadily over the last decades. An overview of this field, focusing in particular on neutrino oscillations, will be outlined in this section; for more details, several in-depth resources are available [1–4].

Only about two decades ago, the solution of the Solar Standard Model puzzle by the SNO experiment [5, 6] led to the decisive evidence of the neutrino oscillation phenomenon, with the important implication that neutrinos are massive particles.

According to the Standard Model, neutrinos are neutral fermions associated to each lepton as a part of the same SU(2)-doublet of the weak force. In particular, in the same way as for leptons, mass terms for the neutrinos can be generated from the Yukawa interaction with the Higgs boson. The mass generation implies that the lagrangian weak-interaction term $\mathscr{L}_{I,L}^{CC}$ for the leptonic charged current would take the form of:

$$\mathscr{L}_{I,L}^{CC} = \frac{g}{2\sqrt{2}}\,\bar{\nu}_i\,\gamma^{\mu}(1-\gamma^5)W_{\mu}\,\mathrm{U_{ij}}\,l_j + \mathrm{h.c.} \tag{2.1}$$

where ν and l are respectively the neutrino and lepton operators, (i, j) are mass operator indexes, and U is the Pontecorvo-Maki-Nakagawa-Sakata (PMNS) mixing matrix. Equation 2.1 could seem to lead to analogous consequences to the quark mixing case, determined by the Cabibbo-Kobayashi-Maskawa (CKM) mixing matrix. There is, however, a substantial difference: since the neutrinos can only interact via the weak force, only the flavor states can be experimentally observed, which are a linear combination of the mass eigenstates. This leads directly to the neutrino oscillation phenomenon, i.e. neutrinos can spontaneously oscillate to different flavors. On top of this, the neutrinos, being neutral massive fermions, are potentially allowed to be

Majorana particles [7]. It can be proven that this is a more fundamental representation than the Dirac one, implying that the most general expression of the PMNS mixing matrix is:

$$U = U^D D^M \tag{2.2}$$

with U^D being the unitary Dirac mixing matrix, and D^M a diagonal matrix containing the Majorana phases. In particular:

$$\begin{aligned}\mathrm{U}^D &= \begin{pmatrix}1 & 0 & 0\\ 0 & c_{23} & s_{23}\\ 0 & -s_{23} & c_{23}\end{pmatrix}\begin{pmatrix}c_{13} & 0 & s_{13}e^{-i\delta}\\ 0 & 1 & 0\\ -s_{13}e^{i\delta} & 0 & c_{13}\end{pmatrix}\begin{pmatrix}c_{12} & s_{12} & 0\\ -s_{12} & c_{12} & 0\\ 0 & 0 & 1\end{pmatrix}\\ &= \begin{pmatrix}c_{12}c_{13} & s_{12}c_{13} & s_{13}e^{-i\delta}\\ -s_{12}c_{23} - c_{12}s_{23}s_{13}e^{i\delta} & c_{12}c_{23} - s_{12}s_{23}s_{13}e^{i\delta} & s_{23}c_{13}\\ s_{12}s_{23} - c_{12}c_{23}s_{13}e^{i\delta} & -s_{23}c_{12} - s_{12}c_{23}s_{13}e^{i\delta} & c_{23}c_{13}\end{pmatrix},\end{aligned} \tag{2.3}$$

and

$$\mathrm{D}^M = \mathrm{diag}(1, e^{i\lambda_1}, e^{i\lambda_2}) \tag{2.4}$$

where $(s/c)_{ij} = (\sin/\cos)(\theta_{ij})$ and θ_{ij}, δ, and λ_i are real phases. One important consequence of the δ phase is to violate the leptonic CP symmetry, as it makes U^D a complex matrix. The probability of a flavor oscillation in vacuum, given a certain neutrino energy E and distance L from the neutrino source, can be expressed as:

$$\begin{aligned}\mathrm{P}_{\nu_\alpha\to\nu_\beta}(L,E) = \delta_{\alpha\beta} - 4\sum_{i>j}\Re[\mathrm{U}_{\alpha i}\mathrm{U}^*_{\beta i}\mathrm{U}^*_{\alpha j}\mathrm{U}_{\beta j}]\sin^2\frac{\Delta m^2_{ij}L}{4E} +\\ + 2\sum_{i>j}\Im[\mathrm{U}_{\alpha i}\mathrm{U}^*_{\beta i}\mathrm{U}^*_{\alpha j}\mathrm{U}_{\beta j}]\sin\frac{\Delta m^2_{ij}L}{4E}\end{aligned} \tag{2.5}$$

where $\Delta m^2_{ij} := m^2_i - m^2_j$, and (i,j) are the neutrino mass indexes. Several important implications can be proven from Eq. 2.5 and the expression of the mixing matrix (Eqs. 2.3 and 2.4):

- The Majorana phases get always canceled in the matrix product. The neutrino oscillation process is not sensible to the Majarona phases, and cannot be used to prove whether neutrinos are Majorana particles.
- The neutrino oscillation process can be used to measure all the parameters of the Dirac mixing matrix, and also the squared mass differences of the neutrinos.
- The CP-violating phase δ can only appear inside the imaginary matrix product term (i.e. third addendum of Eq. 2.5). This implies that only experiments in "appearance" mode (i.e. $\alpha \neq \beta$) are sensible to CP

violation, as opposed to experiments in "disappearance" mode (i.e. $\alpha = \beta$). In particular, it can be shown that for the case of anti-neutrinos the only difference with Eq. 2.5 is the switch of the sign of the third addendum, leading to a CP asymmetry of:

$$\mathrm{P}_{\nu_\alpha \to \nu_\beta} - \mathrm{P}_{\bar{\nu}_\alpha \to \bar{\nu}_\beta} = 4 \sum_{i>j} \Im[\mathrm{U}_{\alpha i} \mathrm{U}^*_{\beta i} \mathrm{U}^*_{\alpha j} \mathrm{U}_{\beta j}] \sin \frac{\Delta m^2_{ij} L}{4E} \tag{2.6}$$

- The oscillation probability fluctuates as a function of the distance L from the source. In particular, at the position of the source $L = 0$, no oscillation takes place.

While neutrino oscillation experiments are sensible to the squared mass differences, nothing can be inferred on the absolute values of the neutrino masses, which are still unknown and bound to very small values ($<$ eV) from cosmological constraints [3].

A neutrino oscillation experiment has the goal to measure the oscillation probabilities $\mathrm{P}_{\nu_\alpha \to \nu_\beta}(L, E)$: the source-detector distance L (and, when possible, the neutrino energy E) is chosen in order to maximize the oscillation probabilities for a defined pair of mass neutrinos. Typically, the situation is more complicated than the case of Eq. 2.5, as effects of neutrino-matter interaction and a spread of neutrino energies need to be taken into account. Neutrino oscillation experiments are divided according to the employed source of neutrinos, as solar, atmospheric, reactor, and accelerator experiments. Each one is mostly sensible only to a subset of the mixing matrix parameters and mass differences, because of the different neutrino flavors, energies, and source-detector distances. The combination of the results of a vast amount of experiments from the past decades allowed to uniquely determine Δm^2_{21} and s_{12}, while all the remaining parameters are affected by the unknown sign of Δm^2_{31}, which for now is also indistinguishable from Δm^2_{32}. This implies that two orderings (i.e. hierarchies) of the neutrino masses are possible, referred to as "normal hierarchy" ($m_1 < m_2 \ll m_3$) or "inverted hierarchy" ($m_3 \ll m_1 < m_2$). The most important open problems in neutrino physics which can be addressed by oscillation experiments are:

- determining the neutrino mass hierarchy;
- a high precision measurement of θ_{23}, so to determine its octant;
- observing the CP violation effect, and provide a precise measurement of the CP violating phase δ.

There is a significant correlation between the problems listed above, as the lack of knowledge on each one of them contributes to increased uncertainties on the others. To this end, an overall increase of precision on all

of the mixing angles and squared mass differences is also an important goal for future experiments. Concerning the CP violating phase, a very recent result from T2K has constrained the possible values of δ to the smallest 3σ intervals up to now, and ruled out the CP-conserving values at 95% confidence level [8]. However, the problem is still strongly dependent on the mass hierarchy, and the uncertainties are still large.

Addressing these issues represents a considerable challenge for the future (and present) neutrino oscillation experiments, as a significant increase in the measurement precision is required. In fact, the effects that need to be observed are subleading with respect to what has been measured in the past. To make an example, a measurement of the CP violation will require experiments to be run in neutrino appearance mode, as opposed to the disappearance mode mostly used in the last decades. The oscillation probability for $\nu_\mu \rightarrow \nu_e$ is only 5%, which is significantly smaller with respect to other disappearance probabilities that led to successful results in previous experiments (e.g. $\nu_e \rightarrow \nu_e$ for solar experiments is about 30%). On top of this, an even lower statistics will be available for the corresponding anti-neutrino process $\bar{\nu}_\mu \rightarrow \bar{\nu}_e$ (which has to be subtracted to $\nu_\mu \rightarrow \nu_e$), as the $\bar{\nu}_e$ cross section is significantly smaller than the ν_e one.

2.2 Accelerator neutrino beams

Accelerator neutrino experiments could provide a solution to the open problems in neutrino physics discussed in the previous section. This type of experiments consists into extracting protons from an accelerator onto a target, so to produce a secondary particle beam. This beam is then charge-selected and focused into a so called decay tunnel, in which the secondary particles are free to decay. A hadron dump at the end of the decay tunnel suppresses the propagation of hadrons and charged leptons, in such a way that only neutrinos are able to escape, producing a neutrino beam. One or more neutrino detectors are placed downstream the beamline, along the neutrino trajectory. A general schematic of a neutrino beam experiment is shown in Fig. 2.1.

Typically, most of the produced neutrino beam is composed by ν_μ, mainly generated by the reaction

$$\begin{aligned} p + A \rightarrow & \pi^\pm + X \\ & \hookrightarrow \pi^\pm \rightarrow \mu^\pm + \nu_\mu/\bar{\nu}_\mu \end{aligned} \tag{2.7}$$

where A is a target nucleus and X any reaction sub-product. Pions make up for the majority of secondary particles produced by the interactions of the protons with the nuclear target. The fraction of electron neutrinos produced in this type of beams is usually in the order of $\mathcal{O}(1\%)$ with respect to the muon ones. For this reason, accelerator neutrino experiments are mainly

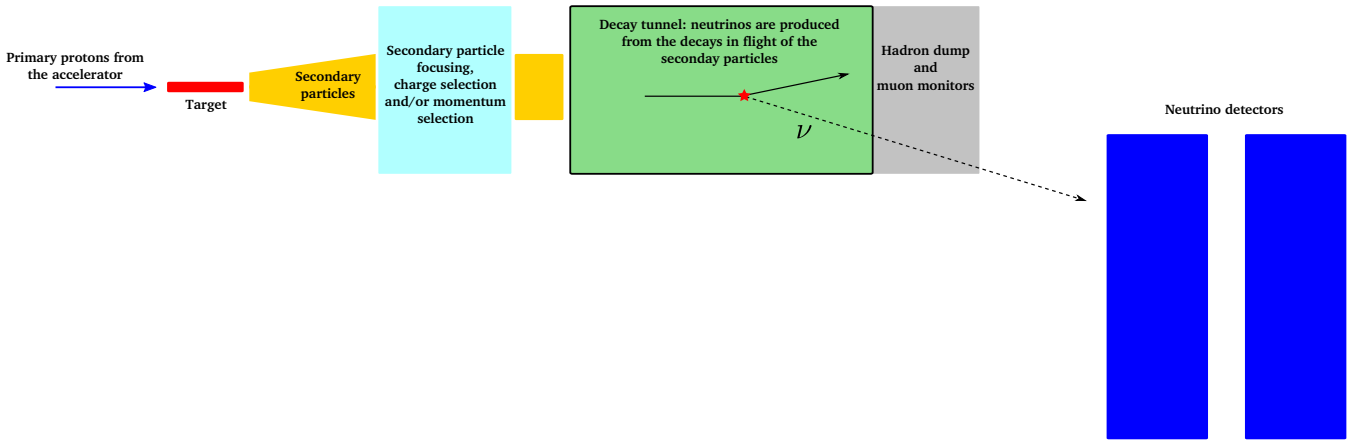

Figure 2.1: Schematic example of an accelerator neutrino experiment (not to scale).

used to detect oscillations of muon neutrinos, with typical neutrino energies ranging from 1 to 20 GeV and source-detector distances of hundreds of km. The final values of neutrino energy and distance are usually tuned so to reach the Δm^2_{31} oscillation maximum. It can be proven that the combination of the oscillation results from different accelerator neutrino experiments, working at different source-detector distances and in $\nu_\mu \rightarrow \nu_e$ appearance mode, could be used to tackle at the same time the mass-hierarchy determination and the measurement of the CP-violation phase [2], exploiting the matter-effect with the earth.

One of the crucial aspects for a successful operation of accelerator neutrino experiments is the particular implementation of the secondary beamline, which ultimately defines the characteristics of the produced neutrino flux [9]. One common approach is to implement a so called wide band beam (WBB). In this case, the charge-selection and focusing system for the secondary particles consists in one or more magnetic horns (i.e. pulsed focusing devices with high angular and momentum acceptance), collinear with the target and the decay tunnel. With this configuration, secondary particles in a large momentum and angular range are all focused into the decay tunnel, producing high neutrino fluxes with a broad energy spectrum. Wide band neutrino beamlines allow to reach very high neutrino fluxes, which are required for a precise measurement of the oscillation phenomenon related quantities. Examples of recent successful WBB are the ones of T2K [10], NuMI [11], and CNGS [12]. One possible way to restrict the neutrinos energy spectrum without losing the advantages of a WBB is to use the so called off-axis technique. In fact, it can be shown that the emission angle of a muon neutrino from a 2-body pion (or kaon) decay is strongly correlated to the

neutrino energy, while only loosely correlated with the parent energy [9]. In such a way, placing the neutrino detector at a certain angle with respect to the decay tunnel allows to select a more narrow neutrino energy spectrum.

A significantly different approach is to develop a narrow band beam (NBB) based on a careful focusing and momentum selection of the secondary particles. In this case, the secondary beamline becomes more sophisticated: an appropriate combination of quadrupoles, dipoles, and collimators is used to transport only a narrow momentum range of the secondaries down to the decay tunnel, so to produce neutrinos in a small energy range. Important advantages of this technique with respect to a WBB are the superior suppression of wrong charge particles and of the background coming from the collinearity of the target, beamline, and decay tunnel (called WBB background). NBB's allow for a very clean selection of the neutrino flux initial conditions, at the expenses of a substantial loss in statistics. This is the reason why most of these experiments have been used for precise measurements of the neutrino cross section and electroweak parameters, instead of the oscillation probabilities. For instance, the narrow band neutrino beams at the West Experimental Area of CERN-SPS make for a successful example, since they allowed to precisely measure neutrino cross-sections at different energies and other electroweak constants as the Weinberg angle θ_W [13–15]. In these types of experiments the neutrino detector is typically placed a very short distance (also referred to as short baseline) from the decay tunnel ($\mathcal{O}(100$ m$)$), as opposed to the typical long baseline oscillation experiments.

An important obstacle for present and future oscillation experiments based on accelerator neutrinos is to obtain a precise estimation of the produced neutrino flux. For this task, detailed simulations of the full beamline (from target interactions to secondary decays), constrained by experimental hadro-production data, short baseline flux measurements and/or muon monitoring at the hadron dump, are usually employed, leading to uncertainties $\gtrsim$ 5-10% of the total neutrino flux [16, 17]. These values are still too high to allow for a successful solution of all the oscillation-related open problems of neutrino physics described in Section 2.1.

One possible solution for long baseline neutrino experiments is to employ a combination of two neutrino detectors, one placed in the very vicinity of the decay tunnel (near detector) and the other at the desired oscillation distance (far detector). This technique allows to significantly decouple the measurement of the oscillation probabilities from the beamline-based estimation of the neutrino flux: the direct flux measurement at the near detector (at which very few neutrinos will oscillate) will be directly compared with the one at the far detector. This method works particularly well in disappearance mode, as the same particles are observed at the far and near detector. For instance, in the long baseline MINOS experiment, the neutrino flux systematics have been reduced to about 3% thanks to this technique [18]. The application

of this method for experiments in appearance mode (e.g. $\nu_\mu \rightarrow \nu_e$), is instead more problematic, as additional errors arise due to the different cross sections at the near and far detectors [19].

Conversely, the poor knowledge on the produced neutrino flux also directly affects the precision of cross section measurement for short baseline experiments. As the theoretical models for neutrino interactions with nuclear matter present significant problematics [20, 21], a solid experimental knowledge of the neutrino cross section would both benefit the theoretical understanding of the weak nuclear structure and the precision of neutrino oscillation experiments. This is especially important for the case of the electron neutrino. In fact, given the significantly low statistics of electron neutrinos inside the typical accelerator neutrino fluxes, direct ν_e cross-section measurements at the energies of interest for the future $\nu_\mu \rightarrow \nu_e$ experiments are still characterized by dominant uncertainties, which can reach values as high as $\sim 20\%$ [22].

2.3 The ENUBET project

The Enhanced NeUtrino BEams from kaon Tagging (ENUBET) project proposes to address the problem of a precise neutrino flux determination with the development of a "monitored" neutrino beamline, potentially allowing to reach precisions of $\sim 1\%$ [23–27].

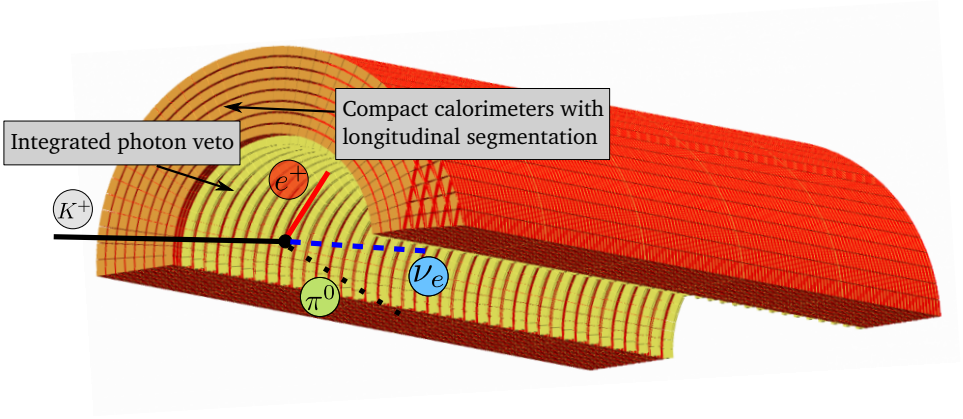

Figure 2.2: Sketch of a sector of the ENUBET instrumented decay tunnel.

In a monitored neutrino beam, the decay tunnel for neutrino production is instrumented with detector material. In the case of ENUBET, the full decay tunnel will be instrumented as an hollow cylindrical calorimeter, built up by several radial layers of compact calorimeters and a layer of photon

veto, as shown in Fig. 2.2.

The basic idea of ENUBET is that, with a proper focusing of the secondary particles so that the beam envelope is fully contained inside the decay tunnel, the main source of particles to hit the decay tunnel walls is given by the products of large angle kaon decays. In particular, the three body decay "K_{e3}" of charged kaons, defined as

$$K^+ \to \pi^0 \, e^+ \, \nu_e \qquad\qquad \mathrm{BR}(K_{e3}) = (5.07 \pm 0.04)\% \text{ [3]} \tag{2.8}$$

is shown to emit the positron at average angles which are about one order of magnitude larger with respect to the ones of most of the other dominant 2-body decays of the secondary particles (e.g. $\pi^+ \to \mu^+ \, \nu_\mu$). Moreover, the beamline parameters can be tuned so that the electron neutrinos produced from K_{e3} decays make up for the totality of the produced electron neutrino flux. Under these conditions, and given that K_{e3} is also the only source of positrons hitting the tunnel walls, a very precise estimation of the produced electron neutrino flux could be performed by monitoring the positrons at the decay tunnel. This technique would allow to decouple the problem of neutrino flux estimation from all of its most error prone components, as hadroproduction data and models, full tracking and interaction simulations, horn current and primary proton flux monitoring, and so on. At leading order, the only required information would be the direct measurement of positrons and the geometrical efficiency of positron detection: this would allow to significantly reduce uncertainties and predict the neutrino flux at the 1% level. The first goal of ENUBET would be to exploit the improved knowledge on the flux to carry out a high precision electron neutrino cross section measurement, which, as discussed in Section 2.2, covers a strategic role for the future neutrino oscillation experiments.

The main sources of background particles at the instrumented decay tunnel (also referred to as "tagger") are other large angle products from other kaon decays, while a possible contamination of the electron neutrino flux could be given by muon decays in flight.

The overall design of the ENUBET beamline and decay tunnel technology is driven by the requirement of reducing the systematics coming for the processes of Table. 2.1 to a level which will not impact the 1% precision on the electron neutrino flux. Some first important design parameters are the momentum of the secondary particles and the length of the full beamline. As kaons have a shorter lifetime than pions (and muons), a combination of high secondary momentum and a short beamline would reduce the contamination of electron neutrinos from muons to negligible levels. For instance, a beamline with a $\mathcal{O}(10$ m) focusing section followed by a decay tunnel of 40 m would stop on the hadron dump a significant fraction of undecayed pions and muons, while allowing the decay of enough kaons. It can be shown that

Decay mode	Branching Ratio	Short notation
$\pi^+ \to \mu^+ \nu_\mu$	$\simeq 100\%$	
$\hookrightarrow \mu^+ \to e^+ \nu_e \bar{\nu}_\mu$	$\simeq 100\%$	
$K^+ \to \mu^+ \nu_\mu$	63%	$K_{\mu 2}$
$K^+ \to \pi^+ \pi^0$	20.7%	$K_{\pi 2}$
$K^+ \to \pi^+ \pi^+ \pi^-$	5.6%	$K_{\pi 3}$
$K^+ \to \pi^0 \mu^+ \nu_\mu$	3.3%	$K_{\mu 3}$
$K^+ \to \pi^+ \pi^0 \pi^0$	1.7%	

Table 2.1: Other decay channels which could play a role for a successful operation of ENUBET.

while pion decays contribute to a negligible level to the number of particles hitting the tagger, both 2 and 3-body decay channels of kaons are instead generating a non negligible background of pions, muons, and photons. To face this issue, the detector technology at the decay tunnel is chosen as a radial and longitudinally segmented structure of compact calorimeters, which can be used to discriminate between a positron, a charged pion, and a muon, according to the particular development of an electromagnetic shower (e^+, γ) with respect to an hadronic one (π^+), or a track of a minimum ionizing particle (μ^+). In order to discriminate between a positron and a photon, a first radial layer of photon veto is employed. The momentum of the secondary particles can also be tuned to improve the e^+/π^+ discrimination. To this end, the final choice of the design momentum of the ENUBET beamline has been driven by the combined requirements of $\nu_{e,K}/\nu_{e,\mu}$ contamination, e^+/π^+ discrimination, total number of produced kaons and electron neutrino flux, and relevance for future oscillation experiments. This resulted in a design momentum of 8.5 GeV, with a 5-10% momentum bite (i.e. momentum spread, typically refers to half of the full momentum range), which leads to a flux of electron neutrinos peaked at an energy of 4 GeV, with a 3 GeV full-width half-maximum. The latter range covers the energy range of interest for DUNE[1]. The requirements of a narrow energy range and a low background at neutrino detector for a high precision cross section measurement imply that ENUBET should be implemented as a NBB, as discussed in Section 2.2.

The large scale of the ENUBET decay tunnel (40 m length for a $\lesssim 1$ m radius) requires the detector technology employed to be cost effective, while the working environment also demands a certain radiation hardness and reliability. For these reasons, the calorimeter modules have been chosen to be of the type of sampling scintillator calorimeters. The particular design of the calorimetric modules and photon veto detectors is the result of a

[1]Other designs with secondary energies centered on the Hyper-Kamiokande range are also being considered.

full R&D program. The main goal of good particle discrimination properties has been paired with many more aspects, which have been investigated trough measurements and simulations [28–31]. Among these, the particular scintillator to be employed, the resistance to radiation and neutron fluxes, the best coupling solution with the detector readout electronics, and so on. The final design for a calorimetric unit has been chosen to be a 11 cm-long module, composed of five 1.5 cm-long iron absorbers and five 0.7 cm-long EJ-204 plastic scintillator tiles, with a transverse area of 3×3 cm^2. A single module consists in about 4.3 radiation lengths (for electromagnetic-shower development) and 0.45 nuclear interaction lengths. The photon veto has been chosen to be a doublet of scintillator tiles. Each scintillator tile is read independently with optical fibers, which are then bundled for each module and coupled to a Silicon Photomultiplier (SiPM). An overview is shown in Fig. 2.3.

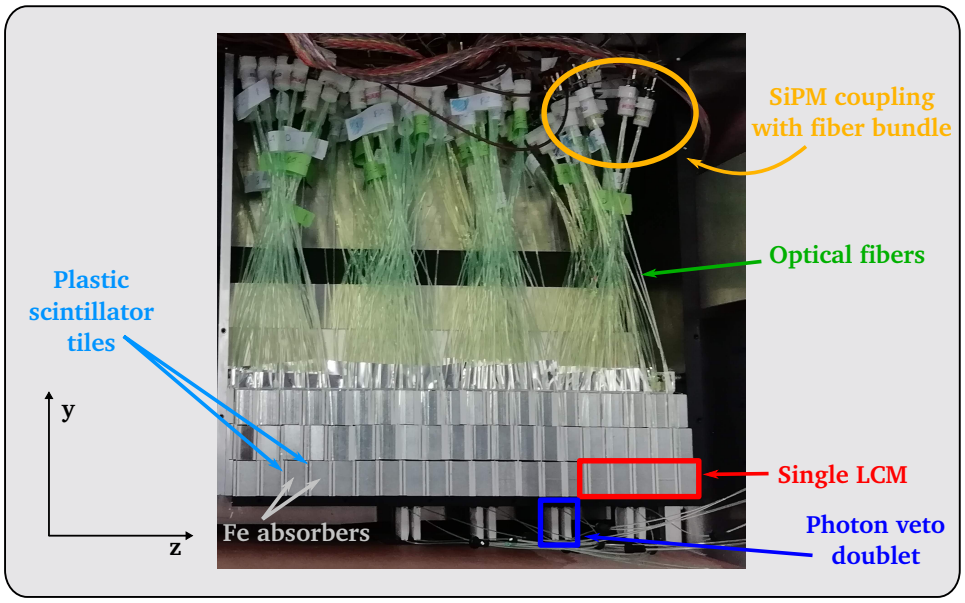

Figure 2.3: A small prototype of the ENUBET tagger tested with a secondary particle beam at CERN-PS (beam direction along the z axis). The term LCM stands for Lateral readout Compact Module, which is how each calorimetric unit is referred to.

While the description of ENUBET outlined up to now has focused on the production of an electron neutrino flux, everything still holds for the case of an electron anti-neutrino flux, which can be produced just by inverting the polarity of the beamline.
Moreover, the operation of ENUBET is not only limited to electron neutrinos, but could be employed for the production of a muon neutrino flux, with

an equally high precision. This can be achieved thanks to the advantages of a narrow band beam. In fact, since the secondary particle momentum is contained into a small range, the energy spectrum of the muon neutrinos produced from pions ($\nu_{\mu,\pi}$) will be well separated from the one of muon neutrinos from kaons ($\nu_{\mu,K}$, mostly coming from $K_{\mu 2}$). The number of produced $\nu_{\mu,K}$ could be already estimated using the measured positrons and the K_{e3} branching ratio. Moreover, since the precision on the K_{e3} branching ratio is limited to $\sim 0.8\%$ (Eq. 2.8), the additional information from muons and pions at the tagger (mainly coming from kaon decays, see Table 2.1) could be used to further reduce the precision on the estimation of the $\nu_{\mu,K}$ flux.

Concerning muon neutrinos, ENUBET could also mitigate another important problem typical of neutrino experiments, which is the energy reconstruction of an event at the neutrino detector [32]. In fact, the combination of a NBB with the off-axis method (seen in Section 2.2), would allow to constrain the muon neutrino energy according to the radial distance from the beam axis, exploiting the strong kinematic correlation between neutrino energy and emission angle in the 2-body decay. This would allow to reduce the coupling of the neutrino energy estimation to methods characterized by significant uncertainties. An example of the application of this method can

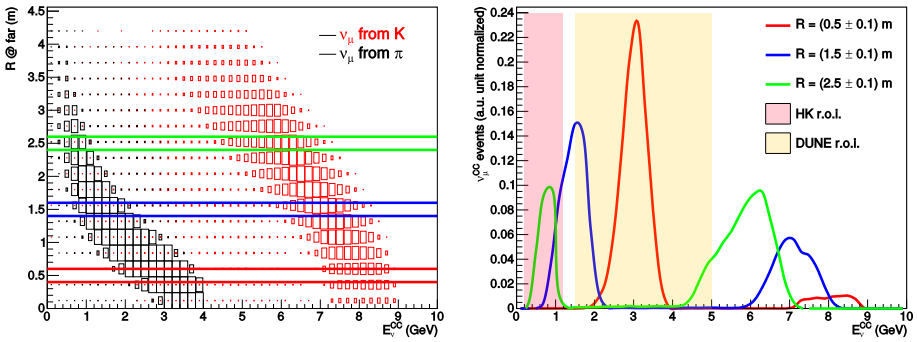

Figure 2.4: Left: correlation of ν_μ^{CC} energy with the radial distance from the beam axis at the neutrino detector. Right: energy distributions for the corresponding radial ranges selected from the left plot. From [25].

be seen in Fig. 2.4, which also shows how the muon and kaon neutrino spectra can be efficiently separated, due to the narrow momentum bite of the secondaries. The technique of exploiting the advantages of a narrow band beam together with off-axis method (more commonly used for WBB) has been referred to as Narrow Beam Off-Axis (NBOA) technique, and more details can be found in [24, 25, 33].

Instrumenting the decay tunnel comes at the price of strict constraints on the particle rates, so to avoid detector saturation. In fact, pile-up and

saturation can severely influence the performance of ENUBET, and have to be kept at negligible levels. Dedicated studies of waveform processing and signal reconstruction are currently under development so to better characterize the performance limitations related to pile-up. For the time being, it can be assumed that the rate of particles hitting the decay tunnel walls needs to remain at least below 500 kHz/cm^2.

This requirement makes the use of fast-extraction of the primary protons, which is the extraction scheme most commonly used in neutrino physics, not compatible with a successful ENUBET operation. The fast-extraction method is used to extract the full circulating intensity (e.g. up to $\sim 4.5 \times 10^{13}$ protons for CERN-SPS) from a synchrotron in a time typically of tens of microseconds: this would completely saturate the instrumented decay tunnel. For this reason, ENUBET will resort to a slow resonant extraction scheme (see Section 2.4), in which the same proton intensity is continuously extracted on a time span of several seconds, bringing the particle rate at the tagger down by several order of magnitudes and fulfilling the pile-up constraints. The only potential drawback of this particular extraction scheme, would be to make impossible the use of magnetic horns for the focusing of the secondary particles (see Chapter 5). As mentioned in Section 2.2, a magnetic horn is a pulsed focusing device with a high angular and momentum acceptance. The current pulses typically reach values of several 100 kA, and can easily compromise the device if prolonged for times longer than a few milliseconds. Without a magnetic horn, the ENUBET beamline would be composed by standard normal-conducting accelerator magnets (i.e. dipoles and quadrupoles), with the advantage of a significant simplification of its implementation and operation, and a particle rate which could reach very low values (according to the selected extraction duration). However, the low statistics due to the small number of kaons and even smaller number of K_{e3} events could lead to very long times for performing an electron (anti)neutrino cross-section measurement at 1% of statistical precision (e.g. up to a few years). For this reason, exploiting the superior acceptance of a magnetic horn represents a possibility which should not be ignored. One way to use such a device while still fulfilling the requirements on the maximum particle rate would consist in the development of a novel slow extraction scheme where the proton intensity is extracted in millisecond-long pulses, repeated regularly over several seconds.

An example of the two possible proton extraction schemes envisaged by ENUBET is shown in Fig. 2.5, where the used proton intensities refer to the CERN-SPS. In particular, the CERN-SPS (see Section 2.5) has been chosen as a potential proton driver for ENUBET: the high proton momentum of 400 GeV, combined with a maximum intensity of 4.5×10^{13} particles, allows to maximize the number of produced kaons per proton on target (POT), while the activities of the CERN Neutrino Platform project (as the presence of the protoDUNE prototypes at the North Experimental Area) represent

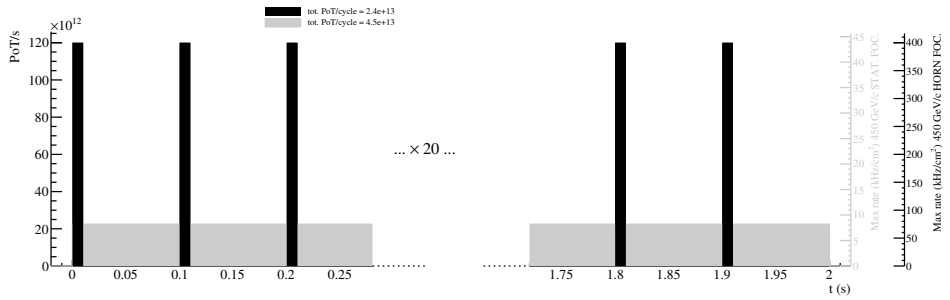

Figure 2.5: Sketch of the two possible extraction schemes compatible with the ENUBET project. In light gray, a standard continuous slow extraction. In black, the newly proposed pulsed slow extraction scheme, here with 10 ms pulse-length and 10 Hz of repetition period. As first presented in [26].

important synergy opportunities. For this reason, ENUBET has planned to undertake a dedicated study on the SPS slow extraction, so to characterize its properties and develop the new pulsed scheme, with a first goal of slow extracting 2-to-10 ms-long pulses, repeated at 10 Hz. These dedicated slow extraction studies will be covered in Chapters 3 and 4.

Thanks to a flux of secondary particles higher than early expectations, and because of the significant complications related with the optimization and operation of a magnetic horn, ENUBET has mainly focused its efforts (up to now) on the design and optimization of the static version of the beamline. The current version of the beamline is the result of an in-depth simulation study, with the ultimate goal of finding the best configuration in terms of signal to noise ratio at the instrumented decay tunnel.

A variety of different aspects have been taken into account, as the total length of the beamline, the magnet apertures and collimation approaches, the level of tertiary background transported at the decay tunnel, and so on, leading to the result shown in Fig. 2.6. In terms of optics, a quadrupole triplet is used as first focusing element, followed by a dipole-quadrupole-diple section aimed at the narrow momentum and charge selection (i.e. 8.5 GeV$\pm$(5-10)%), and a final quadrupole just before the decay tunnel in order to match the beam envelop inside the full tunnel length. Collimation covers a very important role, as a large amount of tertiary particles are produced along the beamline and transported to the tagger. In particular, a 5-cm thick Tungsten foil after the target and a 2.65 m long Inermet180 block before the tunnel have been proven necessary to suppress the number of background positrons arriving at the detector to a sustainable level.

For the case of the magnetic horn option of ENUBET, a new dedicated study anologous to the one which brought to the beamline of Fig. 2.6 should be performed. In fact, given the particular properties of such a device, it

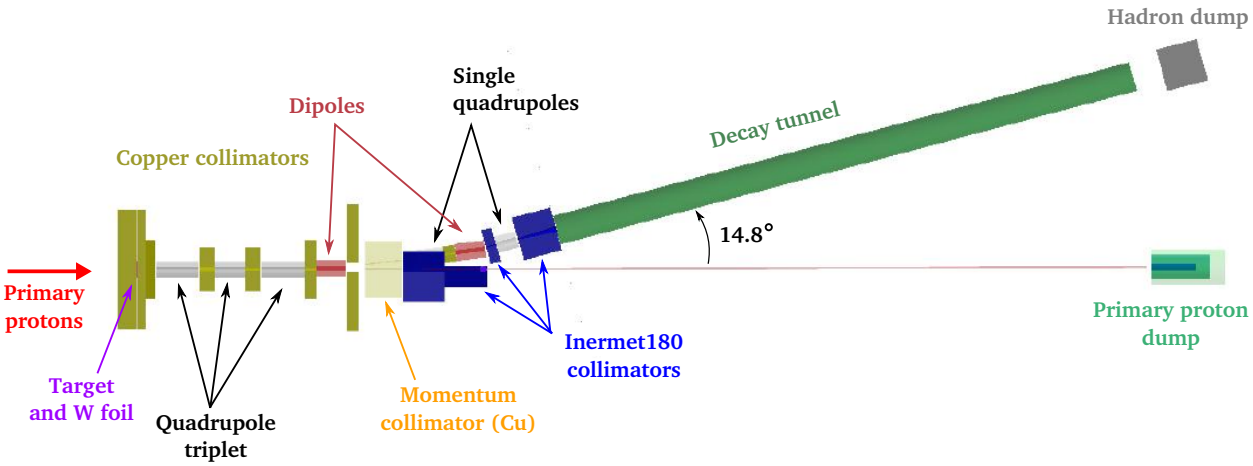

Figure 2.6: Current version of the static ENUBET beamline. Some elements are partially hidden by collimators (as the target and W folis, and the 4$^{\text{th}}$ quadrupole), but are signaled by the arrows on the figure.

is not given that the same beamline configuration found for the static case would also be optimal for the horn option. This will be further discussed in Chapter 5, in the context of the horn optimization study of ENUBET.

2.4 Slow extraction from synchrotrons

Some basic concepts of accelerator physics are necessary to contextualize the framework of slow extraction, and will be introduced in this section. For more details, a vast amount of in-depth resources is available [34–39].

In a synchrotron, the ideal static magnetic fields of all of its elements form the so called reference (or design) orbit, defined as the orbit which exactly repeats itself after every turn. The coordinate system used in accelerator physics is defined with respect to this orbit, as a local right-hand coordinate system of the Frenet-Serret type. In particular, the space can be divided into longitudinal and transverse dimensions, where the longitudinal dimension is defined by the tangent versor $\hat{s}$ to the reference orbit at the particle location on the orbit, while its orthogonal space forms the x and y coordinates. A schematic view of the coordinate system is shown in Fig. 2.7.

Because of imperfections and magnet errors the orbit repeating itself does not coincide exactly with the ideal case of the design orbit: in this case the more general term “closed orbit” is used instead.

The s coordinate along the reference orbit is used as the independent variable for the equations of motion, which are usually obtained under the realistic assumption in which the transverse velocities of the particles are

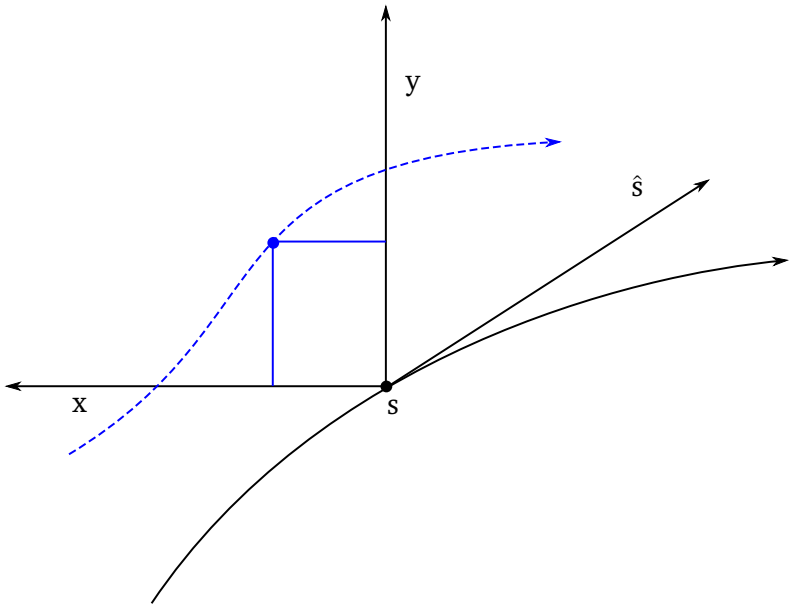

Figure 2.7: Local curvilinear reference system employed in accelerator physics. The curved black line is the reference orbit of the particle accelerator, while the curved dashed blue line is an example of the trajectory of a particle.

negligible with respect to the longitudinal ones. This implies:

$$\frac{d}{dt} = v\frac{d}{ds} \tag{2.9}$$

where v is the total velocity of the particle. The relation of Eq. 2.9 greatly simplifies the equations of motion. In the absence of accelerating cavities in the machine, the transverse motion of a particle can be decoupled from the longitudinal one: since this is the case for the present work, only the transverse motion will be considered, defined by the coordinates:

$$\begin{pmatrix} x \\ x' \\ y \\ y' \end{pmatrix} \tag{2.10}$$

where the primed quantities refer to the derivative with respect to the longitudinal coordinate s. In the presence of normal linear magnets, it can be shown that the two transverse planes (x and y) can be decoupled, and the action of every magnet can be expressed as a 2×2 transfer matrix for each plane. The combined action of all magnets (and drift spaces) for a full turn in the accelerator can then be expressed by the so called one turn matrix

$M(s)$. It can be shown that for a stable motion in any of the transverse planes the one turn matrix takes the form of:

$$M(s) = \begin{pmatrix} \cos 2\pi\, Q + \alpha(s) \sin 2\pi Q & \beta(s) \sin 2\pi\, Q \\ -\gamma(s) \sin 2\pi\, Q & \cos 2\pi\, Q - \alpha(s) \sin 2\pi\, Q \end{pmatrix} \tag{2.11}$$

where α, β, and γ are uniquely defined functions given the accelerator optics and elements, and are usually referred to as Twiss or Courant-Snyder functions. The quantity Q is the (betatron) tune of the machine, and does not depend on s. In particular, it is possible to rewrite the one turn matrix of Eq. 2.11 as:

$$\begin{pmatrix} \beta^{1/2} & 0 \\ -\alpha\, \beta^{-1/2} & \beta^{-1/2} \end{pmatrix} \begin{pmatrix} \cos 2\pi\, Q & \sin 2\pi\, Q \\ -\sin 2\pi\, Q & \cos 2\pi\, Q \end{pmatrix} \begin{pmatrix} \beta^{-1/2} & 0 \\ \alpha\, \beta^{-1/2} & \beta^{1/2} \end{pmatrix} \tag{2.12}$$

which shows how the turn by turn phase space motion could be mapped to a simple rotation with a change of basis. From Eq. 2.12 it follows that the tune represents the numeric frequency of the turn by turn motion of a particle, or, conversely, $2\pi\, Q$ represents the phase advance per turn. In particular, Eq. 2.11 leads to phase space portraits of the shape of ellipses, while the change of basis of Eq. 2.12 turns the ellipses into circles. In the latter basis, the new coordinates are typically referred to as normalized coordinates, $(\tilde{x}, \tilde{x}')$ and $(\tilde{y}, \tilde{y}')$. The difference between the phase space and normalized phase space is shown in Fig. 2.8.

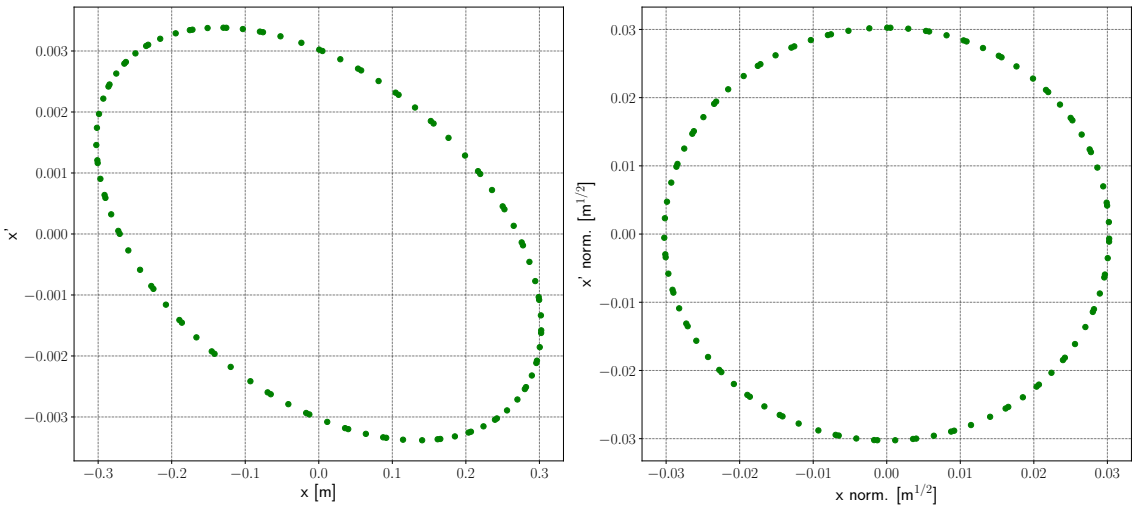

Figure 2.8: Turn by turn phase space portraits for the x plane. Left: standard coordinates. Right: normalized coordinates.

The particular values of the Twiss functions at the considered longitudinal location determine the parameters of the ellipses, i.e. its shape and

orientation. The turn by turn motion along the x (or y) coordinate can be expressed as:

$$x_n \propto \sqrt{\beta}\sin(2\pi Q_x\, n + \phi_0) \tag{2.13}$$

where n is the turn index and ϕ_0 the initial phase. From Eq. 2.13 it follows that the beam envelope along the accelerator is proportional to $\sqrt{\beta(s)}$, where it can be proven that the proportionality constant is an invariant of the motion, used to define the emittance ε_x.

The formalism constructed up to now can be expanded to particles with different momentum from the design one. The quantity typically used to express the momentum deviations of the particles is:

$$\delta_p = \frac{p - p_0}{p_0} \tag{2.14}$$

where p_0 is the design momentum of the accelerator and p the momentum of a generic particle. The focusing and bending effects of the magnets depend on δ_p: the closed orbit of an off-momentum particle will be shifted with respect to the on-momentum one according to the so called dispersion function, and the Twiss-functions and tune will variate too. In particular, the change of tune as a function of the momentum is called chromaticity and defined as:

$$\xi = \frac{dQ}{d\delta_p} \tag{2.15}$$

which in a linear approximation (typically assumed) implies that the tune variation with momentum is $\xi\,\delta_p$. In order to arbitrarily control the value of the chromaticity of an accelerator, non-linear magnets are usually introduced into the lattice, as sextupoles (for the first order chromaticity) and octupoles (for the second order chromaticity).

The introduction of non-linear elements complicates the picture shown above, and can eventually introduce unstable motion under particular conditions. Moreover, not only non-linear elements, but also linear and non-linear errors present on the linear magnets of the accelerator can drive instabilities.

It can be shown that every multipole field component can drive a particular resonance in the accelerator, leading to increasingly divergent oscillations and the consequent loss of the beam. The resonance condition for the transverse plane can be defined as:

$$p = qQ_x + rQ_y \tag{2.16}$$

where p, q, and r are integers. In particular, by only considering the x-dimension (i.e. $r = 0$), it follows that any integer or fractional value of the tune can potentially lead to a resonance, where each irreducible m/n resonance is driven by a $2n$ multipole component (e.g. $\propto 1$ resonances are driven by dipoles, $\propto 1/2$ by quadrupoles and so on). It can be also observed

that only the fractional part of the tune is what defines the problem of Eq. 2.16, since any integer part is re-absorbed into the left member of the equation. This generates a series of so called resonance lines in the transverse tune space, as shown in Fig. 2.9, most of which are carefully avoided in operation.

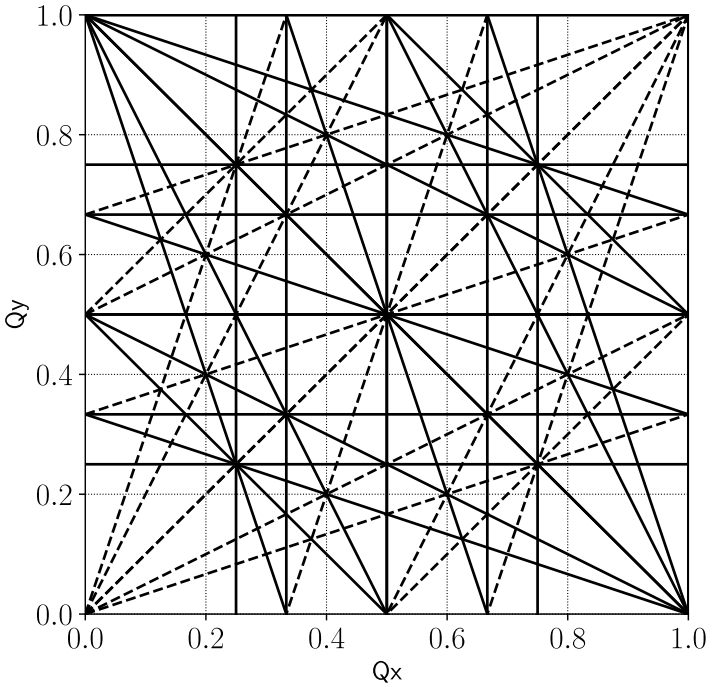

Figure 2.9: All the resonance lines in the fractional tune space up to the 4^{th} order.

To see how the phase space portraits can significantly change under resonant conditions, with respect to the stable ellipses of Fig. 2.8 and Eq. 2.13, the simple case of a linear accelerator with a single thin non-linear magnet can be considered, as shown in Fig. 2.10.

In this case, the one turn evolution consists in the application of the linear one turn matrix (or a simple $2\pi Q$ rotation in normalized phase space) followed by the angular kick $\Delta x'$ of the additional magnet. This motion is described by a discrete difference equation, but under a $1/n$ resonant condition, an approximate n-turns Hamiltonian can be obtained [40–43], which allows for a powerful analytical investigation of the problem.

The Hamiltonians reported in Fig. 2.11 show how no stable area is present for the case of a linear half-integer resonance, while in the presence of non-linear elements the phase space gets divided into a stable central region and unstable external regions.

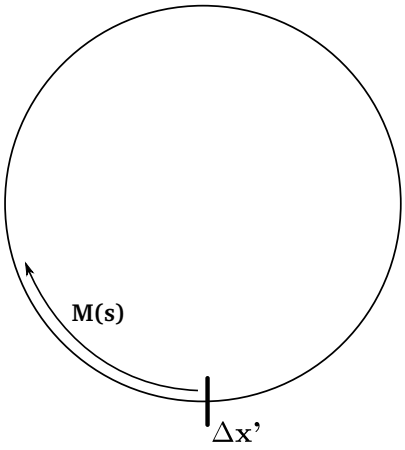

Figure 2.10: A linear lattice characterized by the one turn matrix $M(s)$ containing a single additional thin element.

The latter feature can be exploited in order to produce a resonance in a controlled way, with the goal of continuously extracting all the particles from the accelerator by gradually shrinking the stable phase space area. This is the basic concept behind the slow resonant extraction from synchrotrons.

This work will focus in particular on the case of a third-integer resonance driven by sextupole magnets (Fig. 2.11, top plot), as it represents the slow extraction method currently employed at CERN-SPS, and one of the most widespread implementations of the slow extraction. Only the fundamental concepts of the third-integer resonant extraction will be described here, while a fully detailed treatment can be found in [44, 45].

The main working scheme of this type of extraction can be observed in Fig. 2.12. It can be shown from the 3-turns Hamiltonian that the phase space separatrix takes the shape of three straight lines joining into a triangle. The beam is gradually extracted by shrinking the stable area, making the particles drift along the outgoing separatrix arms. An electrostatic septum is used to deviate the extracted particles into the extraction channel. The sextupoles used to drive the resonance are dedicated extraction sextupoles, as opposed to the chromatic ones: they can reach significantly stronger strengths than the latter ones, and are typically placed in small dispersion regions to minimize their chromatic effect.

Several important quantities which characterize this type of extraction are defined in the following:

- When more than one extraction sextupole are present along the ring, a so called virtual sextupole can be used to simplify the formalism. The virtual sextupole embodies the effect of all the extraction sextupoles

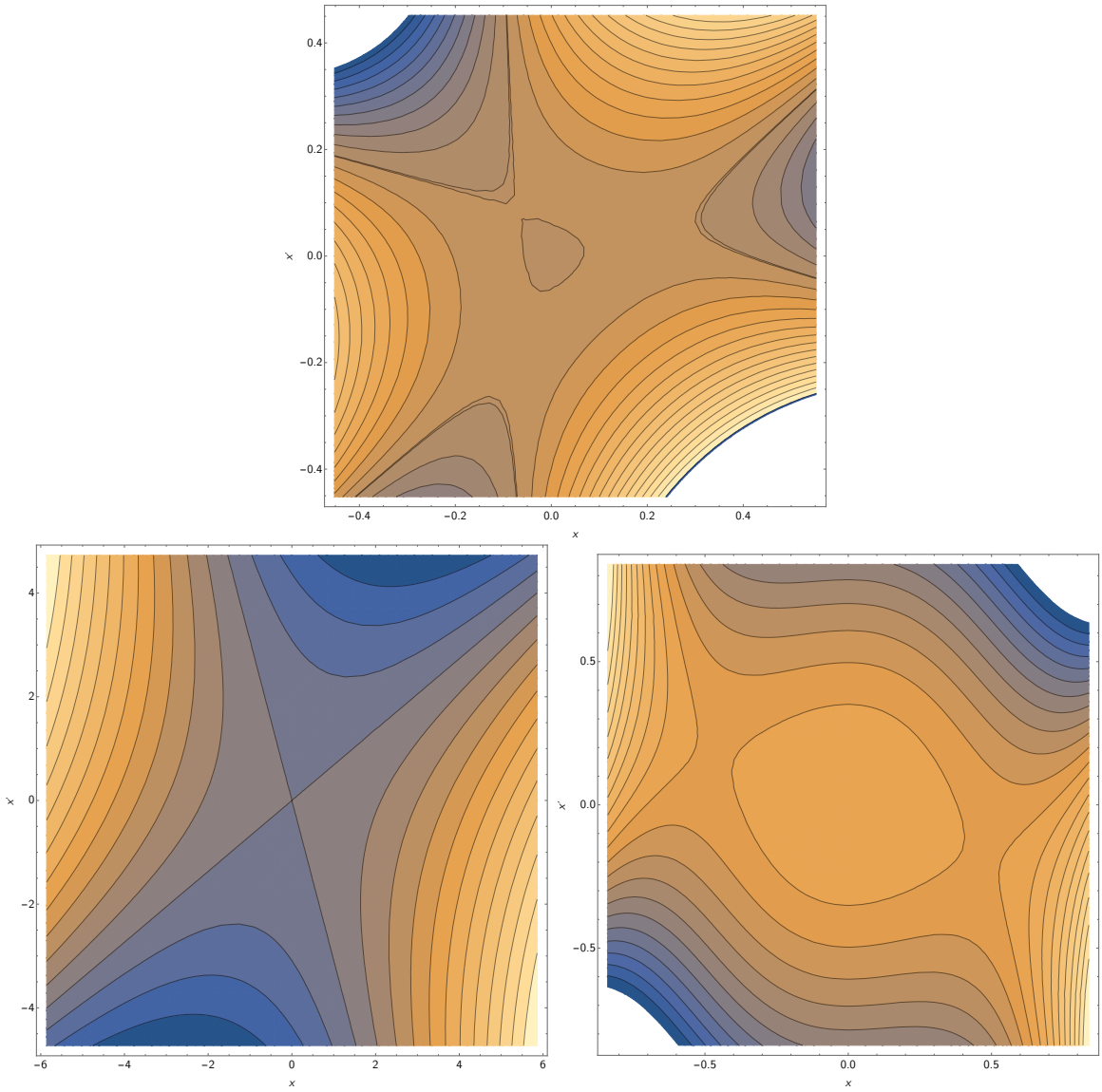

Figure 2.11: Phase space portraits under different resonant conditions, and with different magnet kicks. Top: third-integer resonance driven by a sextupole magnet. Bottom-left: linear half-integer resonance. Bottom-right: half-integer resonance in the presence of an octupolar component. The color scale indicates the Hamiltonian value (arbitrary units).

into a single one, which is defined by an equivalent strength V_{ss} and phase location V_{ϕ}.

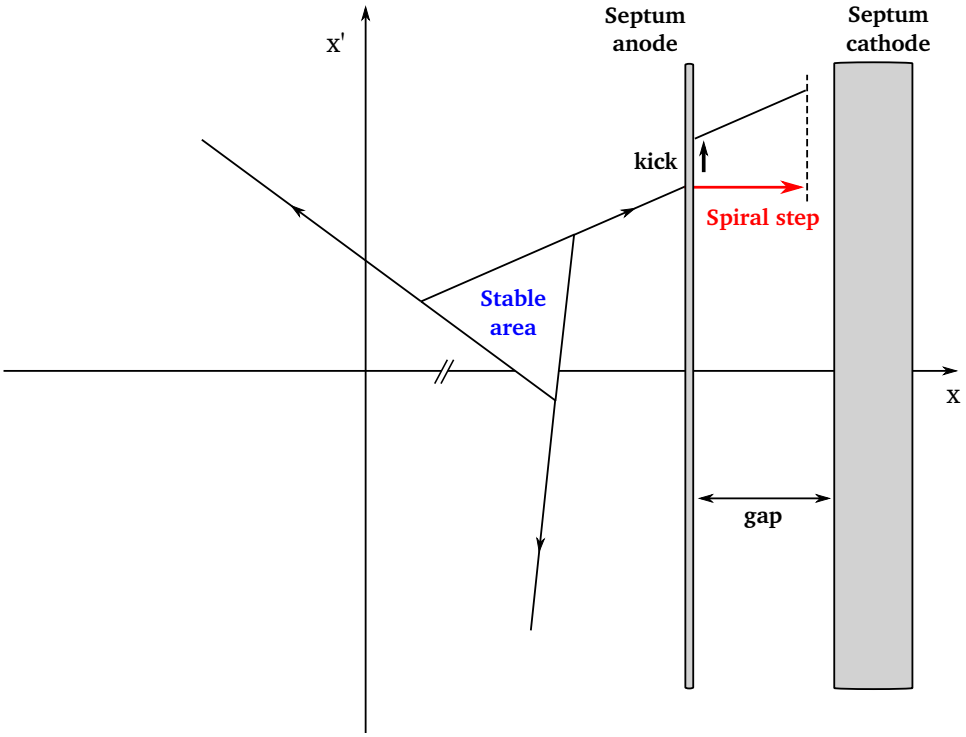

Figure 2.12: Schematic view of the slow extraction phase space (at the electrostatic septum position) for a third-integer resonance.

- The stable area in phase space can be expressed as:

$$A_s^2 = \frac{48\sqrt{3}\pi}{V_{\mathrm{ss}}^2}(\delta Q)^2 \tag{2.17}$$

where δQ is the distance of the particle from the resonant tune. Equation 2.17 can be used to find a useful instability criterion for each particle:

$$|\delta Q| < A\sqrt{\frac{V_{\mathrm{ss}^2}}{48\sqrt{3}\pi}} \tag{2.18}$$

where A is the phase space amplitude of the particle. Equation 2.18 allows to map the extraction problem to a 2-dimensional tune-amplitude space (or analogously δ_p-amplitude), which can be represented in what is called a Steinbach diagram, and shown in Fig. 2.13.

- The spiral step (referred to Fig 2.12) is defined as the maximum horizontal leap that a particle takes into the electrostatic septum gap, and it can be defined as:

$$\Delta X_{\mathrm{ss}} = \frac{3}{4}V_{\mathrm{ss}}\frac{X_{\mathrm{ES}}^2}{\cos\theta} \tag{2.19}$$

where X_{ES} is the electrostatic septum position and θ is the extraction separatrix angle. This quantity needs to remain smaller than the septum gap.

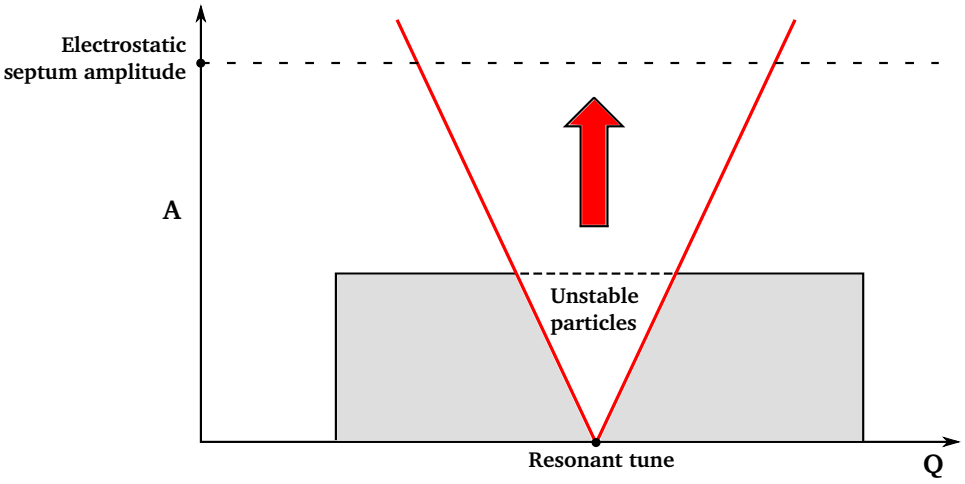

Figure 2.13: Schematic view of the Steinbach diagram: the rectangle area represents the particles in the tune-amplitude space.

In order to extract all the beam, the Steinbach triangle of Fig. 2.13 has to be swept across the entire particle distribution. When the tune distribution of the particles is significantly larger than the tune length of the unstable region (referred to as stopband) the extraction is called "momentum extraction", as particles are extracted along their momentum distribution. In the opposite case, particles would be extracted along their amplitude distribution, and the extraction would be referred to as "amplitude extraction". Many different techniques for driving the particles into resonance are available, and some of them will be discussed in Section 2.5.

Finally, it follows from Fig. 2.12 that a number of particles will be lost on the finite width of the septum blade. This effect is unavoidable, but needs to be kept under control during operation, as it represents one of the most critical issues of the slow extraction. One additional problem comes from the fact that unstable particles with different momenta have different separatrices (as implied from the different amplitudes of Fig. 2.13), and so the effective separatrix which will intersect the septum blade will have a finite width (e.g. assuming zero dispersion at the septum). One possible way to minimize the losses would be to match the machine optics so to superimpose all the extraction separatrices for different momenta, and make them intersect the septum blade at the angle of minimum losses. This matching technique is referred to as Hardt condition [46]. On top of this, many alternative loss-reduction techniques are available, of which some will be discussed

in the following section.

2.5 The slow resonant extraction at CERN-SPS

The Super Proton Synchrotron (SPS) is the second largest accelerator at CERN, with a circumference of about 6.9 km. The SPS lattice is composed by normal conducting magnets, whose structure is organized in 108 FODO periods, and 6 super-periods. Each super-period contains a Long Straight Section (LSS) for special insertions and dedicated operations. A 25 GeV proton beam is injected from the CERN Proton Synchrotron (PS) in LSS1, and accelerated up to a maximum of 450 GeV. Since its construction (terminated in 1976), the SPS has mostly been used for providing beam for fixed target experiments. Currently, up to 4.5×10^{13} protons of 400 GeV can be extracted in 4.8 s from LSS2 to the North experimental Area (NA) fixed target experiments and user facilities. In particular, the NA contains the CERN Neutrino platForm (CENF) extension of the experimental hall EHN1. This facility is fed by two dedicated tertiary beamlines [47, 48], and it currently hosts the protoDUNE detectors [49] and some of the T2K ND280-Upgrade prototypes [50]. An overview of the accelerator and some of its related experiments is shown in Fig. 2.14.

The slow extraction currently implemented at the SPS can be defined as a chromatic-based third integer resonant extraction, where the beam is extracted along its momentum distribution (i.e. momentum extraction, as defined in Section 2.4).

Parameter	**Value**
Momentum	400 GeV/c
Emittance	1.9×10^{-8} m
One turn time	23 µs
Chromaticity	-26.67
Resonant tune	$26.6\bar{6}$
Start of tune ramp	26.62
End of tune ramp	26.72
Flat top duration	4.8 s
Momentum range (δ_p)	3×10^{-3}
Virtual sextupole strength	169.3 $\mathrm{m}^{-1/2}$
Closed orbit bump (max)	48 mm
Electrostatic septum pos.	68 mm

Table 2.2: Main slow extraction parameters of the SPS accelerator. When applicable, each parameter refers to the horizontal plane.

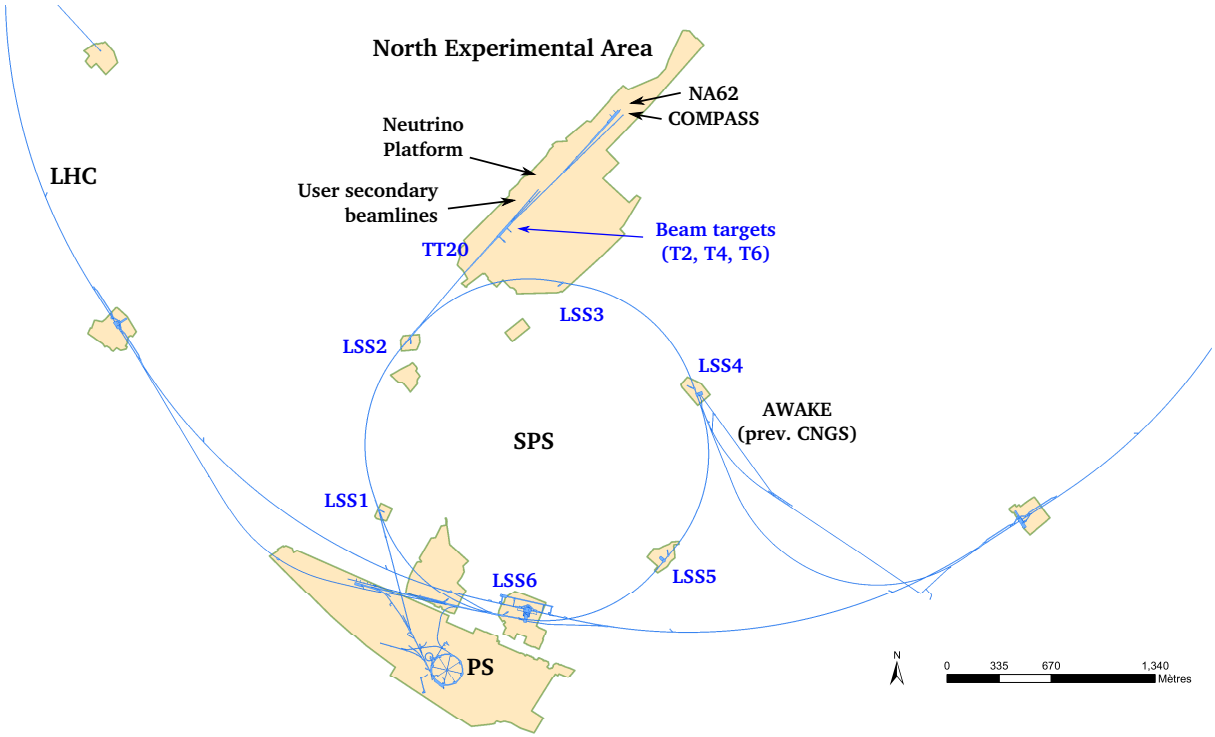

Figure 2.14: Aerial view of the SPS, with its connected experimental facilities and accelerators. The beam is injected from the PS in LSS1, while it can be extracted from LSS2, 4, and 6. The slow extraction only takes place from LSS2 into the TT20 transfer line, which brings the beam to the NA. LSS6 and 4 are used for LHC-filling or for dedicated experimental facilities as AWAKE (in LSS4, previously occupied by CNGS) or HiRadMaT (LSS6).

Table 2.2 reports the main extraction parameters of the SPS. The extraction takes place on the horizontal plane, hence, the emittance, chromaticity, tune, close orbit maximum, and septum position are referred to the horizontal plane. During the extraction, an extraction bump brings the closed orbit up to a maximum of 48 mm with respect to the reference orbit, in correspondence of the electrostatic septum. This brings the beam at about 20 mm from the wires of the electrostatic septum, making it the principal aperture limitation of the machine. Given the rather high (negative) value of chromaticity, this quantity will often be reported as normalized chromaticity, i.e. normalized to the machine tune value: the normalized chromaticity of the SPS (χ) is then -1.

In order to remove any unwanted high frequency component from the beam, the RF is switched off before extraction, and the beam is debunched. However, before being debunched, the beam undergoes a so called "RF gym-

nastic" procedure [51, 52], with the purpose of increasing the spread of the momentum distribution, and shaping it as close as possible to a uniform one.

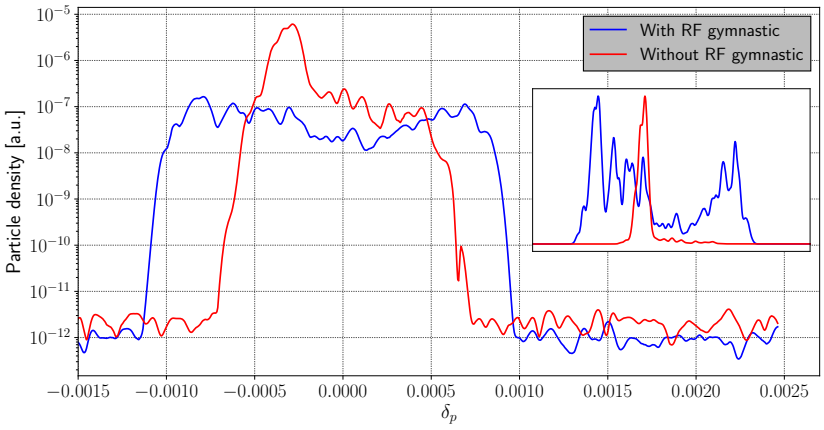

Figure 2.15: Momentum distribution (at extraction) with and without RF gymnastic measured with a Schottky pick-up. Main figure: two distributions in a logarithmic scale, small figure: linear scale, with a renormalization to the maxima for visualization purposes.

Figure 2.15 shows a recent measurement of the momentum distribution of the SPS with (blue) and without (red) the RF gymnastic. The measurement has been taken in 2018 using the technique of the Schottky pick-up [53], for which the power spectrum of a coasting beam signal measured by a pick-up can be mapped to the momentum distribution of the beam through the relation:

$$\frac{\Delta f}{h f_0} = -\eta_s \, \delta_p \tag{2.20}$$

where f_0 is the revolution frequency, h is the observed harmonic number of f_0, and η_s is the current slip factor of the machine. It is possible to observe from Fig. 2.15 that, without the RF gymnastic, almost all the particles would be contained in a momentum range less than 1/4 of range obtained with the RF gymnastic. At the SPS particles are extracted along their momentum distribution (i.e. momentum extraction): a large and uniform momentum spread allows to have an extraction ramp as linear as possible, and does not require to set very high values of chromaticity in order to reach the desired tune span (or conversely, it allows to reach higher values of tune span for the same chromaticity.

Given the momentum of Fig. 2.15 and the fact that the extraction tune ramp is adjusted in order to obtain a flat spill, a uniform momentum distribution with a linear tune ramp is assumed for the simulation models of the

extraction. Using the latter models, it is possible to visualize the Steinbach diagram of the SPS, as reported in Fig. 2.16.

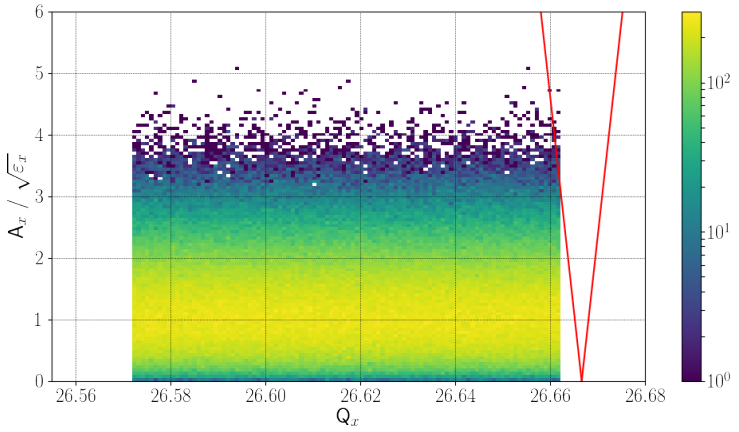

Figure 2.16: Steinbach horizontal tune diagram for the SPS slow extraction, at the start of flat top. The color scale indicates the number of particles in the tune-amplitude space.

Figure 2.16 shows the particle density in the tune-amplitude space and the Steinbach triangle at the start of flat top. Using the parameters of Tab. 2.2 the resonance stopband can be computed using:

$$|\delta Q| = A \sqrt{\frac{V_{\mathrm{ss}}^2}{48\sqrt{3}\pi}} \tag{2.21}$$

where V_{ss} is the virtual normalized sextpole strength, and for the amplitude A the representative value of $2\sqrt{\varepsilon_x}$ is used, yielding:

$$\begin{aligned} S_{\mathrm{B}}^{Q} &\simeq 0.003 \\ S_{\mathrm{B}}^{\delta_p} &= \frac{S_{\mathrm{B}}^{Q}}{\xi} \simeq 10^{-4} \end{aligned} \tag{2.22}$$

where S_{B}^{Q} is the stopband in the tune space, and $S_{B}^{\delta_p}$ is the stopband in the momentum (δ_p) space. In particular, the momentum stopband determines the irreducible size of the extraction separatrix at the electrostatic septum, given that the dispersion at that location is small but not zero (~ -0.2 m). It can be also noticed that the SPS slow extraction is a momentum extraction in a good approximation, given the fact that the stopband-width is several tens of times smaller than the full distribution, as evident from Fig. 2.16 and Eq. 2.22. The extraction at SPS is implemented in two possible ways, which are described in the following.

The first extraction implementation is a standard tune sweep extraction (referred as QSWEEP), where the tune of the machine is swept across the resonance by ramping all the main quadrupoles, while the other machine elements are kept constant. By doing so the main optical functions in the ring are changing along the flat top. In particular, a variation in time of the phase advance between virtual sextupole and electrostatic septum makes the extraction separatrix rotate, implying that the Hardt condition cannot be matched for the full extraction. One possible way to address this problem would be to use a dynamic bump (i.e. an extraction bump that scales along the flat top) as successfully employed for the operation of the JPARC slow extraction [54]. Such a solution has previously been attempted also at the SPS [55], for then being replaced by the second extraction method which will be described later in this section.

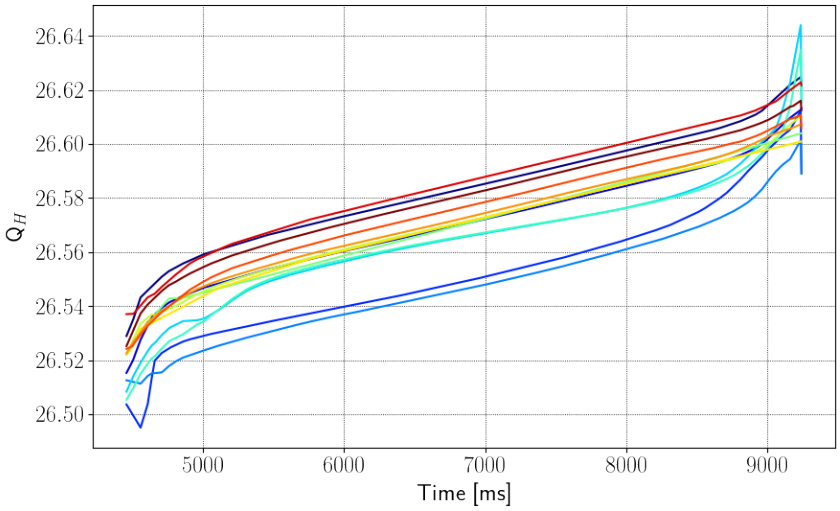

Figure 2.17: Example of trimmed flat top tune ramps for the QSWEEP operation of the SPS.

Typical flat-top tune ramps used for the QSWEEP extraction are shown in Fig. 2.17. Thanks to the high level interface of the LHC Software Architecture (LSA) [56], tune ramps like the ones reported in Fig. 2.17 can be directly trimmed into the machine, and are automatically converted into the appropriate combination of magnet currents using a set of calibration tables.

The flat-top tune ramp is chosen in order to produce an extracted spill as flat as possible. It can be observed from Fig. 2.17 that significant variations are present between tune ramps of different days (the reported ramps refer each one to a different day, chosen over a period of two months): this is because the machine conditions can be subjected to drifts, driven by several reasons, as super-cycle changes, different magnetic hysteresis cycles, and so

on [57–59].

In order to repeatedly adjust the tune ramp for producing a flat extracted spill, a control-room application, named Autospill, is employed [60].

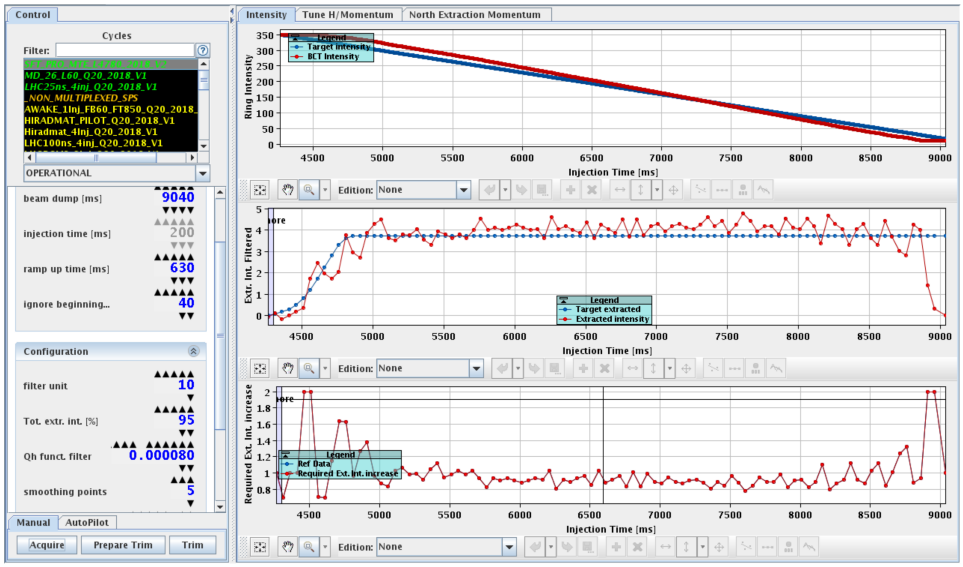

Figure 2.18: Example of the interface of the Autospill application for spill smoothing, after a successful iteration.

Figure 2.18 shows the interface of the Autospill application for the spill shape optimization. The application is based on an iterative feed-forward approach, meaning that changes in the flat-top ramp are trimmed on a machine cycle based on data collected at the previous cycle (as opposed to a feedback system, in which changes are applied instantly). The algorithm works in the following way:

1. The demanded spill profile is set by the user via the configuration parameters (blue curve in the middle canvas of Fig. 2.18).

2. The last extracted spill signal is computed using the circulating intensity measured with a DC Beam Current Transformer (BCT) (the top canvas of Fig. 2.18 shows the measured BCT signal in red, while the reconstructed spill is shown in red in the middle canvas).

3. The last tune ramp trimmed in the machine is read, and the point-to-point ratios between the demanded spill signal and the measured one (bottom canvas of Fig. 2.18) are used to locally scale the slope of the tune signal as:

$$\frac{dQ_{\text{new}}}{dt}(t) = \frac{dQ_{\text{last}}}{dt}(t) \times \frac{S_{\text{ref}}(t)}{S_{\text{meas}}(t)} \tag{2.23}$$

where Q is the horizontal tune, t the considered time instant, and S_{meas} and S_{ref} respectively the measured and demanded spill signals. As for the case of the algorithm the problem is discrete, t can be replaced with an array index `i`. Additionally, the slope correction factor can be multiplied by a scaling factor, which can be used as a hyper-parameter for better convergence. The new tune ramp computed according to Eq. 2.23 is then trimmed into the machine through the LSA interface, and the algorithm repeats from point 2 until convergence is reached.

The algorithm of Eq. 2.23 derives from a simple expression for the extracted spill signal. By observing Fig. 2.16 it is clear that, during the extraction, the tune-amplitude distribution of the particles will move into the resonance (i.e. the Steinbach triangle is fixed) with a speed equal to the tune ramp speed. In an equivalent way, for the QSWEEP extraction, the Steinbach triangle will be swept across a fixed momentum distribution, as shown in Fig. 2.19.

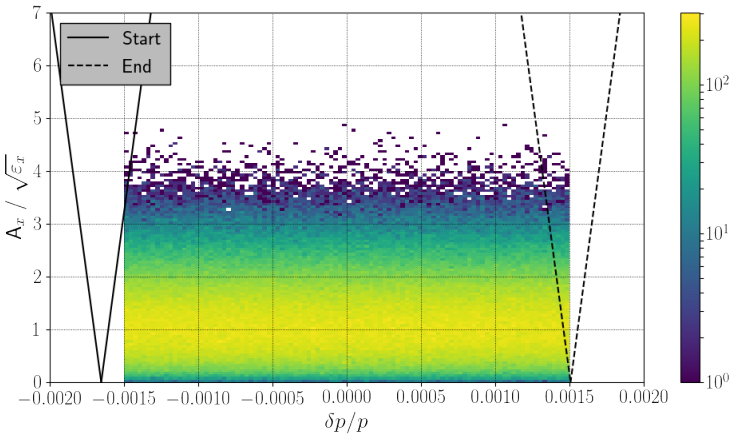

Figure 2.19: Steinbach diagram in the momentum space δ_p for the SPS slow extraction. The initial (continuous black line) and final (dashed black line) positions of the Steinbach triangle according to the SPS tune ramp are shown.

This means that, for each time instant t along the ramp, there will be one resonant momentum δ_p^R on the momentum distribution defined as:

$$\delta_p^R(t) = \frac{Q_R - Q(t)}{\xi} \tag{2.24}$$

where $Q(t)$ is the machine tune, Q_R the SPS resonant tune of $26.6\bar{6}$, and ξ the machine chromaticity (considered as a constant in good approximation). By assuming an instantaneous particle extraction and an ideal momentum

extraction, the number of extracted particles in a time instant t along the ramp can be expressed as:

$$\begin{aligned} dN &= P\big(\delta_p^R\big)\, d\delta_p^R \\ &= P\big(\delta_p^R(t)\big)\, \frac{\partial \delta_p^R(t)}{\partial t}\, dt \\ &= -\frac{P\Big(\frac{Q_R - Q(t)}{\xi}\Big)}{\xi}\, \frac{\partial Q(t)}{\partial t}\, dt \end{aligned} \tag{2.25}$$

where P is the momentum density distribution, and Eq. 2.24 has been used to obtain the final result. Equation 2.25 can be used to define a particle density as a function of the machine tune as:

$$K(Q) := -\frac{P\Big(\frac{Q_R - Q}{\xi}\Big)}{\xi} \tag{2.26}$$

Using the last step of Eq. 2.25 and applying the definition of Eq. 2.26, the number of extracted particles per seconds is found to be:

$$\frac{dN}{dt} = K\big(Q\big)\, \frac{dQ}{dt} \tag{2.27}$$

This expression represents a good approximation for the macro-structure of the extracted spill shape (i.e. low frequency components of the spill signal). Equation 2.27 is at the basis of the iterative algorithm of Eq. 2.23, which uses the derivative of the tune as the optimization variable for reaching the optimum spill. Typically, convergence is reached within ~ 10 iterations, and the optimization is re-issued when the spill quality degrades significantly (on a days to week basis), as mentioned above. A particular case of the result of Eq. 2.27, is that for a uniform momentum distribution, a linear ramp will lead to a flat spill (as it is the case in simulations). It is possible to see from Fig. 2.17 that the central part of the ramps is not far from a linear ramp. An example of an obtained spill can be observed in the middle canvas of Fig. 2.18 (red curve).

The second, and very recent, slow extraction implementation at the SPS is the so called Constant Optics Slow Extraction (COSE) [61]. This type of extraction eliminates the problem of the variation of the machine optics during the flat top. As mentioned before, ramping the main quadrupoles and keeping the remaining elements fixed changes the optics of the machine along the flat top, resulting in a variable separatrix presentation at the electrostatic septum. One way in which this problem is often addressed is by exploiting a so called betatron core slow extraction. This type of extraction employs a betatron core magnet (i.e. a toroidal ferromagnetic material wrapped with coils) placed around the beam pipe, in order to generate a DC electric field

in the direction of the beam [62, 63]. In such a way, the typical debunched beam used in slow extraction can be gradually accelerated into resonance, without the need of ramping any magnet.

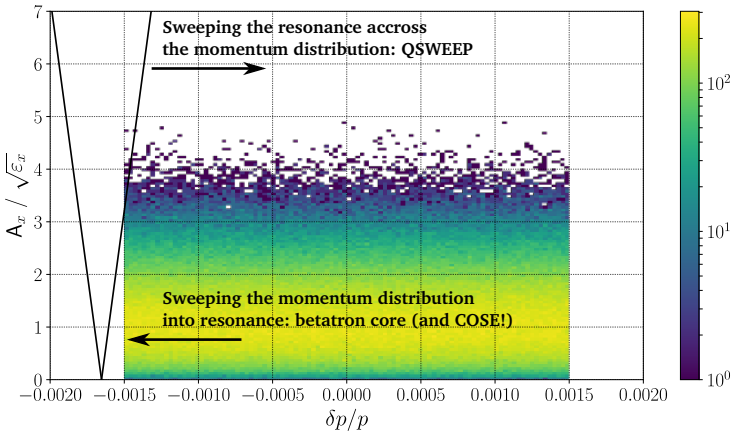

Figure 2.20: Differences of the Steinbach diagram evolution along the flat top for a QSWEEP-like extraction and COSE/betraton core-like one.

Figure 2.20 summarizes the difference between a betatron core-like extraction and a QSWEEP-like one, according to the evolution of the Steinbach diagram during flat top. In a quadrupole sweep extraction, the resonance (i.e. the Steinbach triangle) is swept across a fixed momentum distribution. In a betatron core one, instead, the resonance is fixed, and the full momentum stack is accelerated (or decelerated) into it. In formula:

$$\delta_p(t) = \frac{p(t) - p_0}{p_0} \tag{2.28}$$

where $p(t)$ is the momentum of the particle, and p_0 is the fixed design momentum.

Typically, the betatron core extraction is used for small medical machines, where the momenta are of hundreds of MeV/c. Accelerating the full momentum stack of Fig. 2.20 into resonance at the SPS would imply to accelerate the particles over $\sim$ 1 GeV/c of momentum range, making the development of a dedicated betatron core magnet prohibitive.

A clever way to obtain an equivalent result is to move the momentum distribution into resonance by sweeping the design momentum of the machine p_0, while keeping constant the momentum p of each particle:

$$\delta_p(t) = \frac{p - p_0(t)}{p_0(t)} \tag{2.29}$$

The design momentum is in fact a parameter defined by the machine, in relation to the gradient field of every magnet. In particular:

$$p_0 \propto B\rho \tag{2.30}$$

where $B\rho$ is the magnetic rigidity. Since the geometric strength of each magnet type is normalized by $B\rho$, sweeping each gradient field proportionally to $B\rho$ will keep the machine optics fixed, while shifting the momentum distribution. For the COSE implementation at the SPS, all the magnets of the machine are simultaneously swept thanks to the LSA framework, which allows to control the high level parameter "momentum" (i.e. $B\rho$), in an analogous way to what is done for the horizontal tune in the QSWEEP extraction.

It can be proven that the overall spill shape can be defined by an expression analogous to Eq. 2.27, where, in this case, the tune is replaced by the design momentum (more details are presented in Appendix A). This implies that COSE can be optimized in the same way in which QSWEEP is (i.e. using the Autospill iterative approach based on Eq. 2.23), just by acting on a different machine parameter, as it has been proven with dedicated tests and operation.

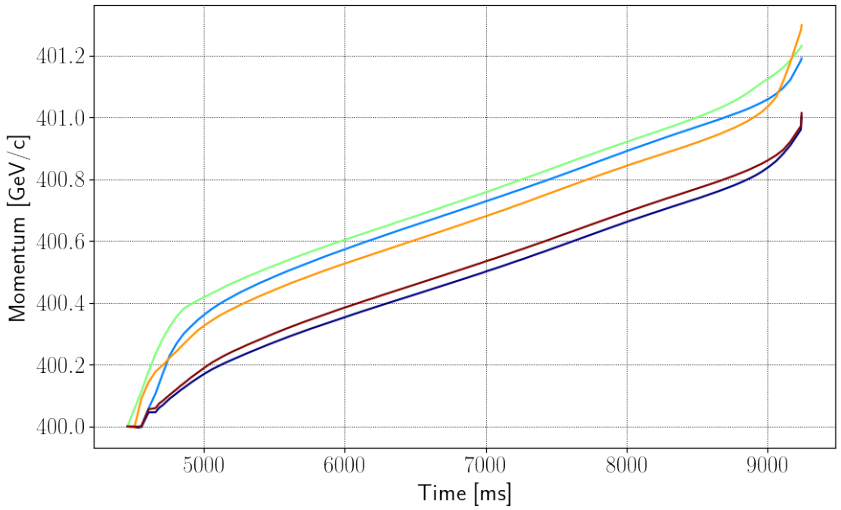

Figure 2.21: Example of trimmed flat top momentum ramps for the COSE operation of the SPS (selected over a period of one month).

Figure 2.21 shows typical flat-top momentum ramps trimmed in the machine for COSE operation. It can be observed that the momentum ramp is positive, i.e. the design momentum is increased. This is in agreement with the fact that the momentum distribution needs to be "decelerated" into resonance (Steinbach diagram of Fig. 2.20). Despite a quantitative comparison

with the tune ramps of Fig. 2.17 would not be reliable due to the machine drifts, the qualitative agreement between the shapes of the momentum and tune ramps can be observed in Fig. 2.22.

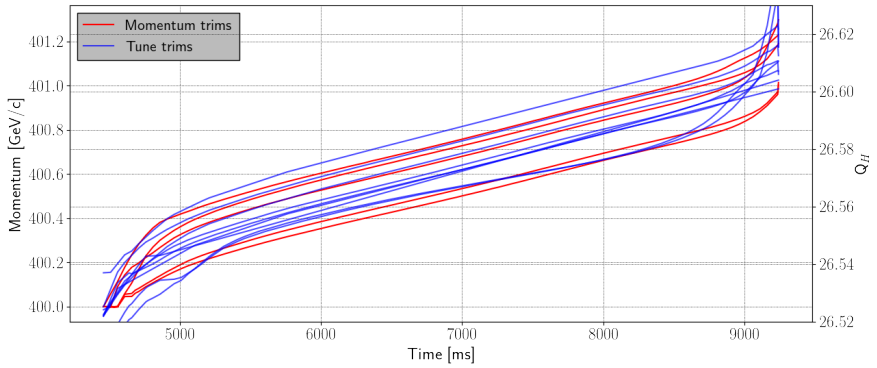

Figure 2.22: Comparison between trimmed tunes for QSWEEP extraction (blue) and trimmed momenta for COSE extraction (red).

The good qualitative agreement confirms that the overall shape of the trimmed ramps (for extracting a flat spill) is ultimately given by the shape of the momentum distribution $P(\delta_p)$, as expected from Eq. 2.27 and its COSE counterpart.

The correct holding of a constant optics during the flat top (and in particular, of a fixed separatrix presentation at the electrostatic septum) has been confirmed both by simulations and measurements, as for instance observing the time-profile of the losses during the operation of slow extraction with a silicon bent crystal for septum shadowing [64]. Ultimately, the COSE extraction allowed for a successful test of the technique of separatrix folding for losses reduction at the electrostatic septum [65]. This technique exploits the detuning effects of strong octupole fields in order to bend back the extracted separatrix [66–68], so to increase the spiral step as much as possible. Without a precise and fixed presentation of the separatrix at the electrostatic septum the beam could be easily lost with harmful consequences, due to the effects of the strong octupole fields.

Given the many validations, COSE is being used as default operational slow extraction method for the SPS since September 2018.

After having described in detail the current extraction implementations at the SPS, it is also useful for the present work to be aware of what has been done in the past at the SPS concerning the slow extraction.

Back in the seventies, well before the LHC era, the SPS was slow-extracting the beam towards two experimental areas: the North Area (NA) and the West Area (WA, now dismantled). Given the two fixed-target based

facilities awaiting the beam, the SPS extraction cycle was more complicated than today: typically, two flat-tops were present, at two different energies of $\sim$ 250 and $\sim$ 450 GeV. During the first flat-top, part of the beam was slow extracted at the WA, since it could not sustain more than 250 GeV particles. The remaining beam was then accelerated to 450 GeV, and extracted to the NA. During the first flat-top, because of the successive re-acceleration of the beam, the RF was left ON during slow extraction. Moreover, the RF was used in the real time spill feedback system, for optimizing the spill shape by acting on the beam momentum. Different types of resonant extraction were used for both the flat tops: third-integer, half-integer coherent and non coherent, and, occasionally, integer [69]. The third integer extraction was implemented in a similar way as the QSWEEP of today: 4 extraction sextupoles were used and the beam was brought towards resonance by sweeping the current in the main quadrupoles, extracting spill lengths no longer than 3 s. The half-integer extraction was instead used for the so called "fast resonant extraction" [70]. The name comes from the fact that the spill length was always shorter than 3 ms, reaching minima of 70 µs for the coherent case. In this extraction type the beam was first brought near a half-integer tune by the main quadrupoles of the machine, and 4 extraction octupoles were switched on in order to define a stable phase space area. The tune was then swept across resonance by one dedicated extraction quadrupole, connected to a capacitor bank which was rapidly discharged, later in time replaced with a thyristor generator able to clip the discharge ramp [71]. The spill length was typically regulated by changing the chromaticity, or, if necessary, acting on the discharge ramp. The same half-integer based fast-resonant extraction could be also operated in a coherent mode, in which the full beam was kicked out of the stable region onto the separatrix by a fast kicker, and then shaved out by the electrostatic septum in a small number of turns (with a minimum of 4 turns). This type of extraction was generating shorts bursts 23 µs-long, separated from one another by 23 µs, i.e. the SPS revolution time. This coherent resonant extraction was sometimes also implemented by using an integer resonance, with the advantage of removing the 23 µs spacing between the extracted bursts. In the latter case, the stable phase space was typically generated using quadrupoles.

After the WA was upgraded for 450 GeV operation (1983), a new SPS slow extraction cycle with a single flat top has been implemented. In this cycle, a third-integer extraction was performed to slow-extract the beam simultaneously to the WA and NA [72]. To achieve this, the six-sextant symmetry of the machine had been broken, creating two halves with slightly different focusing quadrupoles strengths. In such a way, the two halves of the machine could have two different horizontal tunes, allowing to achieve the same phase space presentation at the two electrostatic septa. A part from the "tune split", the slow extraction implementation was otherwise roughly the same as the QSWEEP of today, but with the spill shape optimized by a

feedback system connected to a servo-quadrupole [73], instead of the present Autospill.

Once the WA has been dismantled the tune-split has been removed, and the same third-integer resonant extraction has continued to be performed solely for the NA, bringing us full circle to the beginning of this section.

Chapter 3

A pulsed slow extraction from CERN-SPS

The present chapter will focus on the main practical goal of this work: developing the pulsed slow extraction scheme envisaged by the ENUBET project and described in the previous chapter. It is important to outline how this work has been structured in the three years available. In particular, the Long Shutdown 2 (LS2) of CERN dictated a complete shutdown of all the CERN complex from the end of 2018 to 2021. This implied that the only available time for developing and testing the new extraction scheme at the SPS would have covered only part of the first year of the present study. For this reason, the first year of this project has been entirely dedicated to machine operations and measurements, with the ultimate goal of proving the feasibility of the pulsed extracted scheme, according to the ENUBET requirements. To this, it has followed a phase of simulation, validation and investigations of possible improvements of the achieved experimental results.

The described structure of the present work is reflected in the organization of the sections of this chapter. Section 3.1 describes how the new extraction scheme has been designed (Subsection 3.1.1), implemented (Subsection 3.1.2), and tested in the machine (Subsections 3.1.3 and 3.1.4). In particular, Subsection 3.1.4 will describe how the first goal of the ENUBET project is successfully reached through the use of an iterative optimization of the new extraction scheme. Following the full experimental characterization of Section 3.1, Section 3.2 will describe the developed simulation model of the pulsed slow extraction, compare its results with the experimental data, and investigate in simulation possible improvement methods. Promising results are found which will be interesting to be tested when the SPS will be restarted. Finally, Section 3.3 offers some reflections on the achieved results which are described up to this point. In particular, the limitations and operability of the possible improvements found from simulation

are discussed, with a focus on their inter-operation with other advantageous extraction methods for loss reduction which have been successfully tested and validated at the SPS concurrently to this work.

3.1 The burst mode slow extraction

As described in Section 2.3, the type of extraction needed for the ENUBET experiment would be a pulsed version of the slow extraction. The full intensity circulating in the machine should be extracted in several 2 to 10 ms-long bursts, possibly repeated at a repetition rate of 10 Hz. Looking back at the slow-extraction configurations implemented in the SPS in the past and described Section 2.5, the most similar one could be identified as the fast resonant extraction: in this case the intensity was extracted in a single pulse of less than 3 ms, sometimes composed of sub-pulses, 23 µs apart, if the extraction was coherent. However, a continuous repetition of ms-scale slow extracted bursts was never performed before: this new type of extraction will be referred to as "burst mode slow extraction".

The operational implementation of such an extraction should be simple and as close as possible to the current slow extraction scheme of the SPS, in order to exploit the available tools, and the reliability gained from years of operational experiences.

3.1.1 A deterministic design of the extraction scheme

A good starting point is to think about a fast resonant extraction implemented for the current third-integer scheme, for example, the QSWEEP extraction.

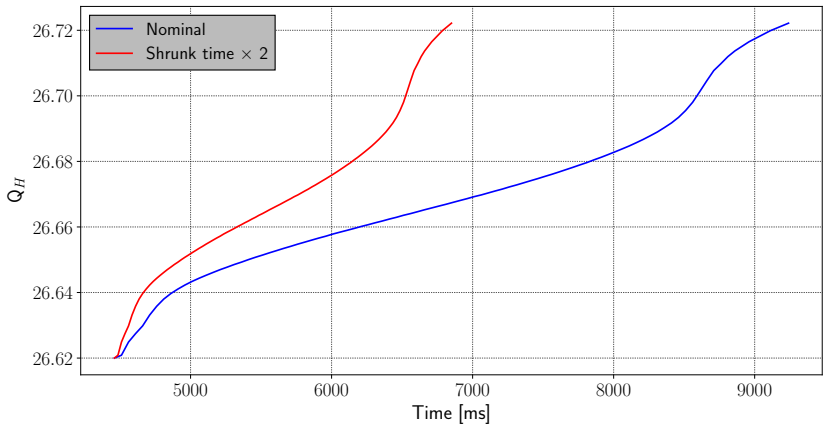

Figure 3.1: Nominal extraction tune ramp (blue) compared to the same tune ramp after applying a time-shrinking of a factor 2 (red).

Figure 3.1 shows how a faster extraction (red curve) can be implemented from the current one (blue curve). If $Q(t)$ is the nominal flat-top machine tune at the time instant t (blue curve), the faster tune signal of the red curve has been obtained using the time-shrinking transformation:

$$t \longmapsto \alpha t \tag{3.1}$$

where α is a real parameter > 1.

By applying the transformation of Eq. 3.1 to Eq. 2.25 or Eq. 2.27, the obtained spill expression is:

$$\frac{dN}{dt} = \alpha \times \left(K\big(Q\big) \frac{dQ}{dt} \right) \tag{3.2}$$

which corresponds to α times the extracted spill signal without the transformation of Eq. 3.1 (i.e. Eq. 2.27). If the starting time t_s of the flat-top is taken as the time origin, so that $t_s = 0$, then the new time interval will be:

$$t \in \left[0, \frac{\Delta T}{\alpha} \right] \tag{3.3}$$

where ΔT is the nominal flat-top length. By putting together Eq. 3.2 and 3.3 it is clear that the spill corresponding to the new shrunk tune signal will maintain the same shape as with the nominal tune, but it will be α times more intense and α times shorter. This can also be justified by a more intuitive reasonment: generating a new tune ramp using the transformation of Eq. 3.1 corresponds to speed up the ramp by a factor α: the same tune sweep over the momentum distribution will take place at α times the speed. This implies that the same spill signal will be extracted in $1/\alpha$ of the time, and for the conservation of the number of particles, the spill will have to be more intense by a factor α. In fact, since the total tune range remains the same, the same number of particles is extracted.

This transformation gives the significant advantage of preserving the spill shape optimization from the Autospill algorithm: if the optimized spill is flat with the nominal tune, it will also be flat with a shrunk version of it. This result is used in order to build the burst mode slow extraction: instead of applying a global time-shrinking of the tune, the tune is divided in N intervals of length Δt (i.e. the burst repetition period, or simply burst period), and each interval is time-shrunk according to Eq. 3.1. Putting it in formula:

$$Q(t + n\,\Delta t) \longmapsto \begin{cases} Q\Big(\frac{\Delta t}{\lambda}\, t + n\Delta t\Big) & t \in [0, \lambda] \\ f(t + n\Delta t) & t \in [\lambda, \Delta t] \end{cases} \tag{3.4}$$

where Δt is the repetition period of the bursts, λ is the length of a single burst, n is the index of the burst period ($n \in \mathbb{N},\ n \in [0, N]$), and f a function satisfying the following constraint for each burst period n:

$$f(t + n\Delta t) \leqslant Q\Big((n+1)\,\Delta t\Big) \quad \forall\, t \in [\lambda, \Delta t] \tag{3.5}$$

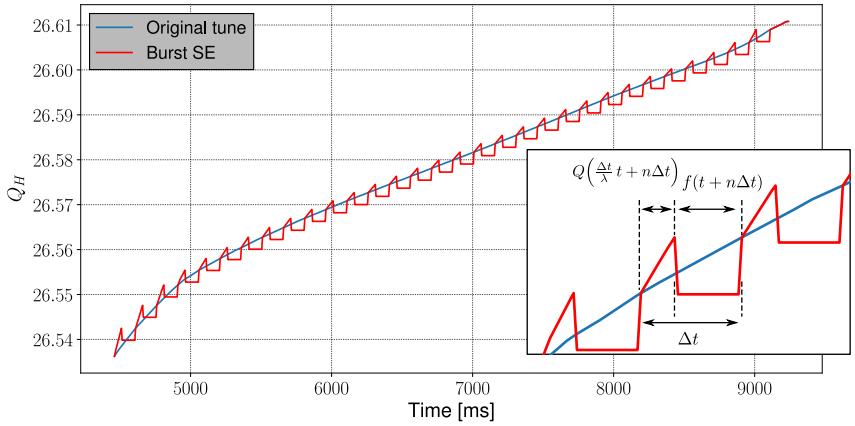

Figure 3.2: Nominal flat-top tune ramp (blue) and corresponding burst mode slow extraction tune (red) generated using Eq. 3.4.

Figure 3.2 shows an example of burst extraction tune ramp (red curve), constructed according to Eq. 3.4, and the corresponding nominal extraction tune (blue curve). The zoomed plot in Fig. 3.2 can be used to visualize Eq. 3.2: on each burst period Δt the nominal tune is time-shrunk in order to extract a single burst of length λ, hence the time-shrinking constant α (referred to Eqs. 3.1, 3.2, and 3.3) becomes $\Delta t/\lambda$. Once the λ seconds long burst has been extracted, no other particle is extracted until the next burst period: in this region the new tune function could take any value fulfilling Eq. 3.5.

The reason of Eq. 3.5 is that, under the assumptions in which Eq. 2.27 has been derived, once the tune sweeps over a certain momentum region, it depletes it of all the particles. In such a way, if the tune ramp comes back to a region which has been already crossed, nothing will be extracted anymore. This implies that in order not to extract any particle the tune value needs to be kept lower than the last extraction tune. This notion was not included in the derivation of Eq. 2.27, which can be extended to:

$$\frac{dN}{dt} = \begin{cases} K(Q)\,\frac{dQ}{dt} & \text{if } \frac{dQ}{dt} > 0 \\ 0 & \text{otherwise} \end{cases} \tag{3.6}$$

implying that the spill satisfies the previous expression of Eq. 2.27 only if the tune ramp is a monotonic function of time, which is usually the case during standard operation, as can be observed from Fig. 2.17.

All of what has been carried out up to now using the flat-top tune ramp $Q(t)$ holds in a good approximation also by using the momentum extraction ramp $p_0(t)$ (i.e. COSE scheme), given the results of Section 2.5 and Appendix A. Nevertheless, despite the significant advantages of the COSE

scheme, the burst mode slow extraction will be implemented using QSWEEP as the underlying extraction scheme, and hence using the technique outlined in Eq. 3.4 and shown in Fig. 3.2. The reason for this is driven by practicality: the burst extraction ramp (Fig. 3.2) could be a strain for the magnet current generators to follow, and so, minimizing the number of magnets to ramp will maximize the probability of success and save debugging time. Nevertheless, a successful operation of the burst extraction in the QSWEEP scheme is a necessary condition for it to work in the COSE scheme. The extrapolation to COSE is only a matter of implementation in the SPS control system and it is seen as future step, which will follow the success of this initial implementation.

3.1.2 Implementation

The implementation of the new extraction scheme into the SPS follows the design described above. In particular, a dedicated Graphical User Interface (GUI)-based control room application has been developed for the task. The application has been written in the Python programming language, exploiting a Python wrapper (PyJAPC) of the Java API for Parameter Control (JAPC). This software framework is a higher level client interface to the Control Middleware (CMW), which allows for a unified control of the accelerator hardware [74–76].

The algorithm implemented in the developed application works as follows:

1. Insert the input parameters for the demanded spill characteristics and tune change (as the burst length λ, burst period Δt, the shape of f-function, and more).

2. The current flat-top tune is read and processed (as per Eq. 3.4) according to the input parameters in order to generate the new burst extraction tune.

3. The proposed tune is shown and compared to the nominal one (as in Fig. 3.2). If confirmed, the new tune is trimmed into the machine, for the selected extraction cycle.

An example of the developed application (referred to as `burstControl`) and the possible tune changes is shown in Fig. 3.3. It can be observed that four alternative burst extraction tunes are available to be chosen as new machine tune (they change in the shape of the f function, referred to Eq. 3.4), and compared to the nominal flat-top tune. In particular, a total of 14 tune configuration parameters are available from the settings menu.

Figure 3.4 shows the comparison between the nominal slow extracted continuous spill and a burst-extracted spill obtained by using the developed control application of Fig. 3.3 during dedicated beam tests in the SPS. The

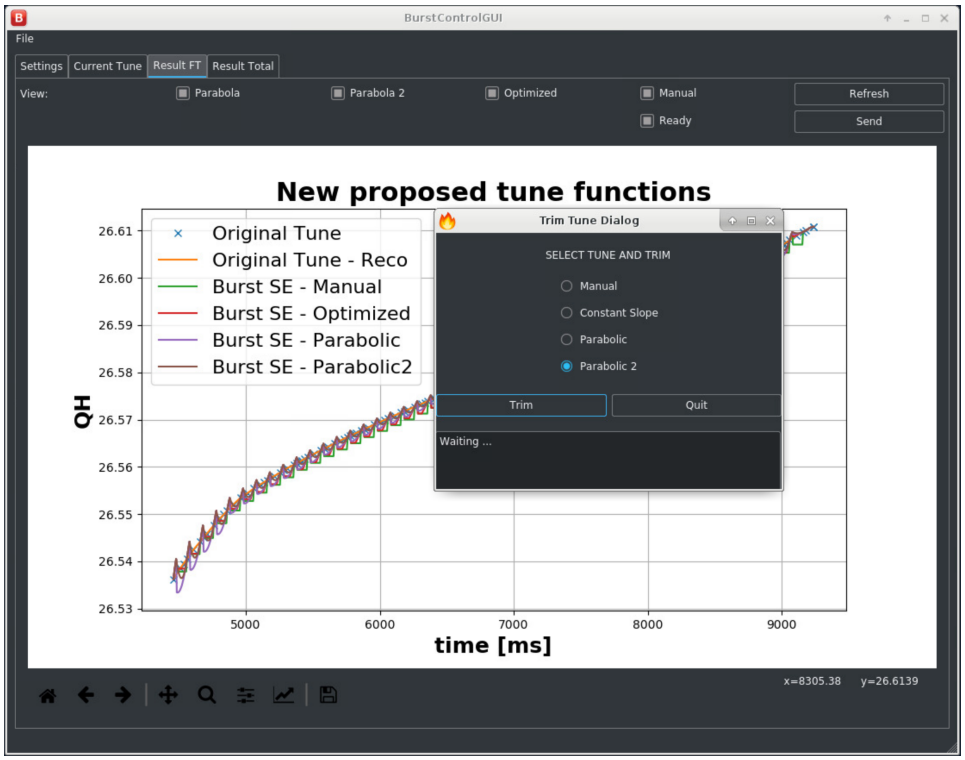

Figure 3.3: Screenshot of the developed control room application for burst extraction operation during the phase of tune trimming.

spill profiles have been measured using a secondary emission intensity monitor placed on the extraction line and sampled at 2 kHz; the calibration of the digitized signal to units of p^+/s is done by using the integrated extracted intensity measured with the BCT along the ring. It can be noticed that there is about a factor 5 between the average peak intensity of the bursted spill and the continuous one. This is well expected from the results of Subsection 3.1.1, given that the bursted spill was selected to have 50 ms pulses over a 250 ms period, yielding a time-shrinking factor of $\Delta t/\lambda = 250/50 = 5$.

Before starting to characterize the performance of the burst mode slow extraction, it is important to first discuss the reason why so many different f functions (referred to Eq. 3.4) are available in the developed control-room application, as shown in Fig. 3.3. Moreover, for each different function shape, numerous configuration parameters are available. From the spill expression of Eq. 3.6, there should not be any difference in the result (i.e. no particles are extracted) for any different shape of f, provided that it fulfills Eq. 3.5. However, in the real case, it has been observed that f plays an important role in the hardware response of the magnet circuit and current generators to the demanded input signal. This hardware response can be so bad to

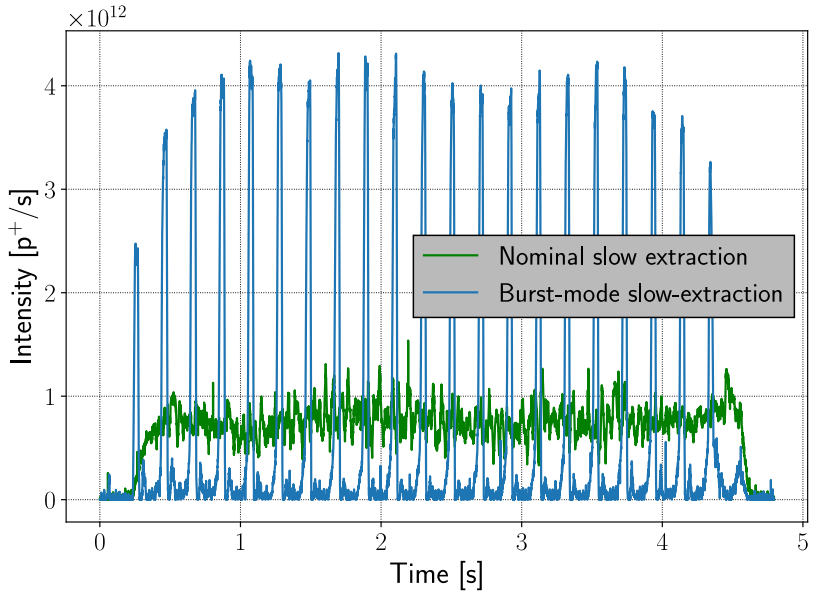

Figure 3.4: Example of a nominal extracted spill (green curve) compared with a burst-extracted spill (blue curve) of 50 ms pulses over 250 ms periods.

significantly alter the shape of the extracted spill.

An example of this effect can be observed in Fig. 3.5, which shows the presence of a significant overshoot and ringing on the measured main focusing quadrupoles current, repeated twice for each burst period. In particular, the overshoots at the start of each increasing tune ramp have detrimental effects for the spill: the overshoot pulse is extracting almost half of the burst intensity in the fist small fraction of the expected burst length, making the shape of the burst irregular. This phenomenon is due to the power converters regulator loop not being able to properly follow the unusual reference current shape. Attenuating this effect does not seem to be a straightforward task, and different methods have been tried during the measurements.

Figure 3.6 shows an example of different tune functions which have been used in attempt to attenuate the wrong power converters response. As it can be observed from Fig. 3.5, the change of slope right before the start of the burst seems to trigger a strong overshoot. This has been handled by imposing

$$\frac{df}{dt}\Big((n+1)\,\Delta t\Big) = \frac{dQ}{dt}\Big((n+1)\,\Delta t\Big) \tag{3.7}$$

and adding other possible attenuation factors as:

- making the same slope ramp start early along the f function (e.g. red curve in Fig. 3.6);

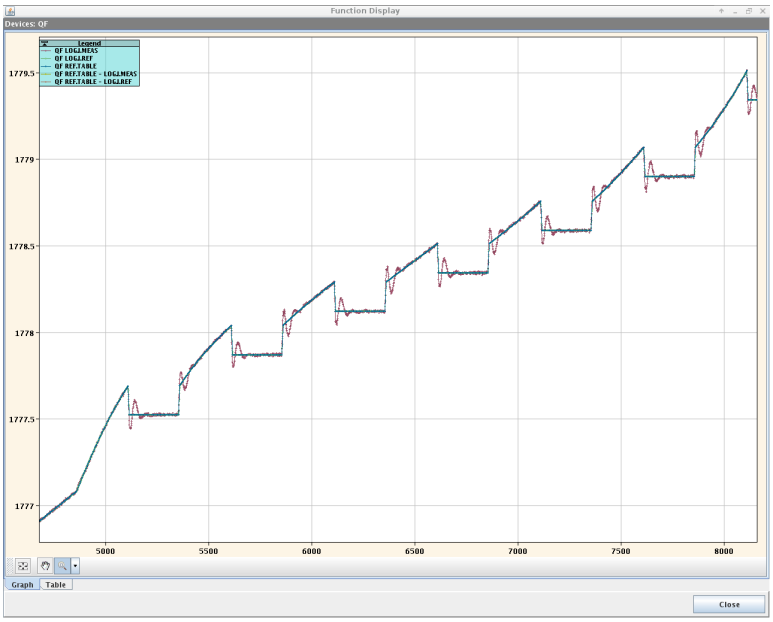

Figure 3.5: An example of a bad hardware response to the generated tune function during burst mode slow extraction tests at the SPS. In green: demanded current for the main focusing quadrupoles. In red: measured current of the main focusing quadrupoles.

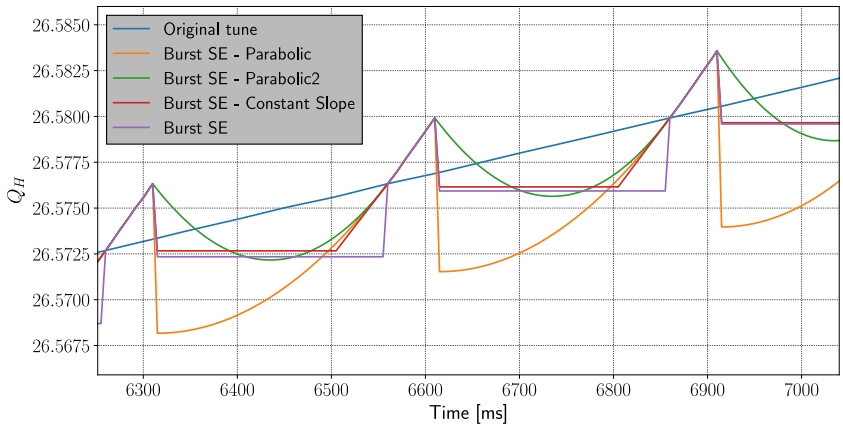

Figure 3.6: Different burst extraction tune functions used to tackle the problem of wrong power converters response.

- replacing the linear ramp with a smooth function, as a parabola (e.g. orange curve in Fig. 3.6);
- imposing the full f function to be a smooth parabolic function and not piecewise anymore (e.g. green curve in Fig. 3.6).

In particular, the reason why a parabolic shape has been chosen is because the underlying Digital to Analog Converter (DAC) for the reference current generator performs a second order polynomial reconstruction of the input signal; higher order polynomials are not guaranteed to be correctly followed [1].

From machine tests it has been observed that there is not a single solution that minimizes the problems all the times, but the different solutions work best for different configurations of burst mode slow extraction. For this reason, care has been used to select the best f function before each burst mode slow extraction measurement (compatibly with how many shapes were currently implemented at the time, and with the available time for measurements).

3.1.3 Machine tests and experimental characterization

The new slow extraction scheme has been tested in the SPS in a dedicated Machine Development (MD). The extracted intensity has been set to $\sim 4 \times 10^{12}$ protons per spill, which is about an order of magnitude less than the nominal operational intensity. The time profile of each extracted spill signal has been measured using the secondary emission monitor previously described.

The first parameter which has been monitored in order to characterize the performance of the new extraction scheme has been the dumped intensity. This parameter is defined as the remaining intensity in the machine after the slow extraction. These remaining particles are then cleanly disposed on the SPS internal dump.

Contrary to the simple model used in Section 2.5 and based on the assumption of an instantaneous momentum extraction (which, nevertheless, is a good approximation for the overall shape of the continuous spill, since it excludes high frequency components), particles in phase space take a finite amount of time for being extracted, and so present a certain inertia to respond to rapid tune changes. For this reason, at each extraction cycle, a fraction of particles will remain inside the ring. As described and tested in [77], this fraction can be expressed as:

$$\frac{I_{\text{rem}}}{I_{\text{tot}}} = \int_0^{A_0} a\, e^{\frac{-a^2}{2\sigma^2}}\, da = \sigma^2 \left(1 - e^{\frac{-A_0^2}{2\sigma^2}}\right) \tag{3.8}$$

where I_{rem} and I_{tot} are respectively the remaining and total ring intensity, a is the phase space amplitude, σ is the beam size, and A_0 is the minimum particle amplitude that gets extracted. A_0 depends on the extraction

[1] Private communication with A. Nicoletti.

parameters, and can be written as:

$$A_0 = \sqrt{3 - 2\sqrt{2}}\,\frac{2}{k_s \beta_s}\sqrt{2\pi \frac{dQ}{dt}} \tag{3.9}$$

where k_s is the extraction sextupole strength, β_s is the betatron function at the sextupole, and Q the machine tune, assumed to be a linear ramp. By putting together Eqs. 3.8 and 3.9, it follows that the faster the tune ramp, the more particles will remain into the machine. This is a relevant issue for the burst mode slow extraction, given that it is based on a time-shrinking transformation of the tune function, which has the effect of locally multiplying the derivative of the tune by the time-shrinking factor α, as shown in Eq. 3.2 and Fig. 3.1. In particular, by neglecting the f function pieces of the burst tune, and approximating the nominal slow extraction tune ramp $Q_0(t)$ to a linear one, the dumped intensity should approximately follow:

$$\frac{I_{\text{rem}}}{I_{\text{tot}}}(\alpha) \propto (1 - e^{-b\,\alpha}) \tag{3.10}$$

where α is the time shrinking factor, and b is a constant defined as:

$$b = (3 - 2\sqrt{2})\,\frac{64\pi}{(k_s \beta_s)^2 2\sigma^2}\frac{dQ_0}{dt} \tag{3.11}$$

These results imply that the burst mode slow extraction has the potential to significantly increase the dumped intensity (factors α of 10 or more are to be expected): this is an aspect to carefully optimize to make the burst-mode slow extraction operational. In order to investigate the behavior of the dumped intensity, a dedicated burst extraction run has been performed by maintaining a fixed burst length of 50 ms, and gradually increasing the repetition period, so to increase the time-shrinking factor α. The f function component has been selected to be as steep as possible (compatibly with the hardware response), in order to minimize its influence on the behavior of Eqs. 3.10 and 3.11.

The dumped intensity has been measured using the ring BCT, and its evolution as a function of the shrinking factor α (i.e. burst period over burst length) has been fitted with a function of the type of Eq. 3.10, but including an offset to take into account other external effects. The result can be observed in Fig. 3.7. In particular, the fit parameter b (referred to Eq. 3.10) has been used to extrapolate the average equivalent tune ramp speed dQ_0/dt by inverting Eq. 3.11. The obtained result is compared with the average speed of tune variation in the SPS, as a validation test. The value obtained from the fit turns out to be the same order of magnitude of the expected ramp slope, proving that such a simple model can also help to describe the behavior of the non-extracted intensity for the burst mode extraction. The results of Fig. 3.7 are not encouraging, given that a 10% of dumped intensity

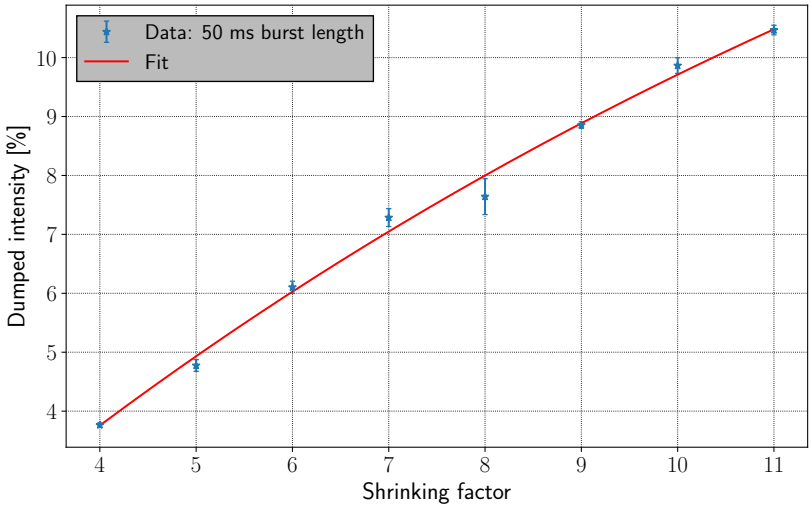

Figure 3.7: Dumped intensity (as a fraction of the total circulating intensity) as function of the time-shrinking factor α for a burst extraction run with fixed burst length of 50 ms.

is reached at a shrinking factor of 10 (which is the baseline for the ENUBET requirements). This could be too much for the SPS in regular operation and possible mitigations shall be evaluated.

In order to reduce as much as possible the dumped intensity, it has been attempted to act on the particular shape of the f function, so to induce a so-called "momentum cleaning": by slowing down the slopes of the non-extracting sections of the tune ramp, the particles not extracted during the previous pulse would then be extracted.

An example of a tune function featuring "momentum cleaning" is shown in Fig. 3.8: by comparing it with the tune function of Fig. 3.2 it is evident how more of the non-extracted particles will be cleaned out from the machine due to the slower tune speed between bursts.

The result of this test can be observed in Fig. 3.9, compared with the previous data and expected behavior. In particular, the dumped intensity can be lowered significantly, especially at high shrinking factors, and, more importantly, the behavior foreseen by Eq. 3.10 does not seem to hold anymore. This follows directly from the fact that components of the tune functions other than the burst ramps are now significantly contributing to the extraction of particles.

The same technique has been used for the successive burst extraction tests, where different values of demanded burst length and period have been explored. Figure 3.10 shows an overview of the results.

It can be observed from Fig. 3.10 that no other burst extraction run

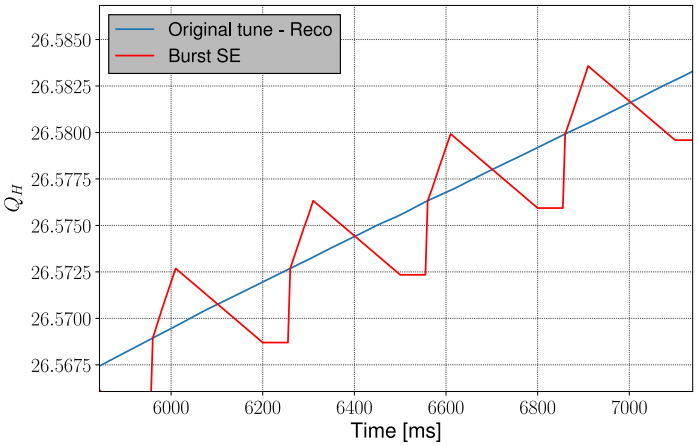

Figure 3.8: Example of a burst extraction tune ramp used for the "momentum cleaning" operation.

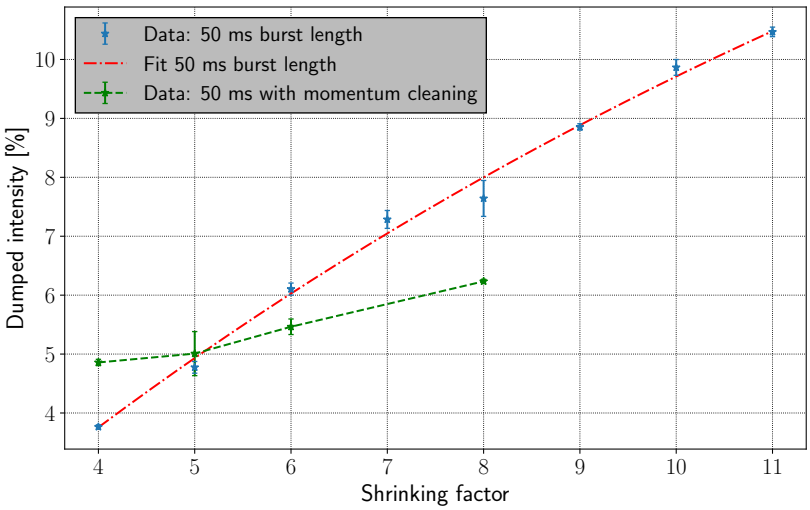

Figure 3.9: Comparison between the dumped intensity behavior as a function of the shrinking factor with (green) and without (blue) the momentum cleaning approach.

has reached dumped intensity values as high as the initial one (blue curve). More importantly, a trend of dumped intensity reduction as a function of the selected burst length can be observed, i.e. smaller demanded burst lengths seem to have a lower dumped intensity, until the reach of the floor of $\sim 3\%$ of the continuous extraction. This observed behavior could be explained by the fact that for smaller burst lengths λ, the burst period value Δt for

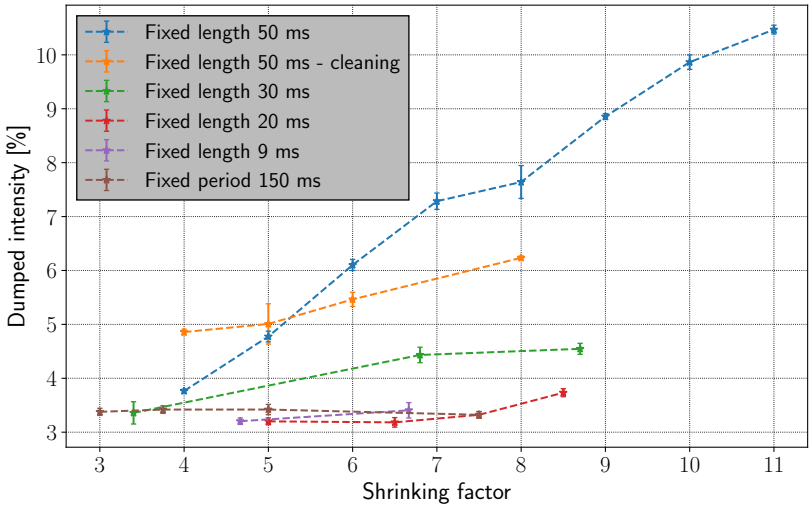

Figure 3.10: Measured dumped intensity behavior for different burst extraction runs performed during the MD's at the SPS.

reaching the same shrinking factor is also smaller, implying a higher number of extracted bursts N_b:

$$N_b = \frac{\Delta T}{\Delta t} \tag{3.12}$$

where ΔT is the flat top length. With a higher number of extracted bursts there is both less extracted intensity per burst and a higher number of non-extracting tune regions in which to apply the momentum cleaning technique. These two facts combined significantly contribute to a more effective removal of the remaining particles from the ring.

This is an important achievement for the burst mode slow extraction requirements provided by the ENUBET experiment, where the repetition period should be of 100 ms, and the burst length in the range of 10 to 2 ms. Finally, the particles extracted between two bursts represent a small fraction of the bursts intensities, and should not influence the correct operation of the experiment. This is both because the strong focusing system is not pulsed in such moments, and also because a gated veto can be placed, if necessary.

The second parameter which has been used to characterize the operational performance of the burst extraction is a measure of the temporal length of each extracted burst. This quantity will be referred to as "effective

burst length", and defined as:

$$\lambda_{\text{eff}}^{n} = \frac{\left(\int_{-\Delta t/2}^{\Delta t/2} s(t + n\Delta t + t_0)\, dt\right)^2}{\int_{-\Delta t/2}^{\Delta t/2} s^2(t + n\Delta t + t_0)\, dt} \tag{3.13}$$

where $s(t)$ is the extracted spill, n is the burst index inside the spill, and t_0 is the center of the first extracted burst. Equation 3.13 is based on the common "effective spill length" and "duty factor" parameters typically used in the framework of slow extraction, and defined in [78]. The latter parameters are often used to measure the quality of the full extracted spill: the effective spill length of a square pulse corresponds to its exact length (i.e. 100% usable length), while the value decreases for pulses with irregular amplitudes. The same holds for Eq. 3.13, which can also be used as a measure of the length of an extracted pulse. In particular, it is useful to consider the example of a Gaussian-like pulse fully contained in the integration interval; its effective length will approximately be:

$$\frac{\left(\int_{-\Delta t/2}^{\Delta t/2} e^{-t^2/2\sigma^2}, dt\right)^2}{\int_{-\Delta t/2}^{\Delta t/2} e^{-t^2/\sigma^2}\, dt} \simeq 2\sqrt{\pi}\sigma \simeq 3.5\,\sigma \tag{3.14}$$

which corresponds to about the 92% of its area, and proves to be a realistic estimation of the "usable" length of an extracted pulse.

For all the burst extraction configurations tested at the SPS, Eq. 3.13 is applied to each burst of every extracted spill; the averaged results are reported in Fig. 3.11. This shows the burst extraction performance for different demanded burst duration and as a function of the shrinking factor.

Two important trends can be observed from Fig. 3.11:

- A slight increase of the effective burst length seems to take place for increasing shrinking factors.
- The smaller the demanded burst length, the more distant from it is the obtained effective burst length.

The first point can be explained by the previous observations on the dumped intensity. In fact, the higher is the shrinking factor, the faster will be the burst tune ramp: this implies that for higher shrinking factors a bigger fraction of particles will be extracted between the pulses. The effective burst length will then result to be slightly larger.

The second point is instead a direct consequence of the finite time-spread of a group of extracted particles. Even if an instantaneous tune change

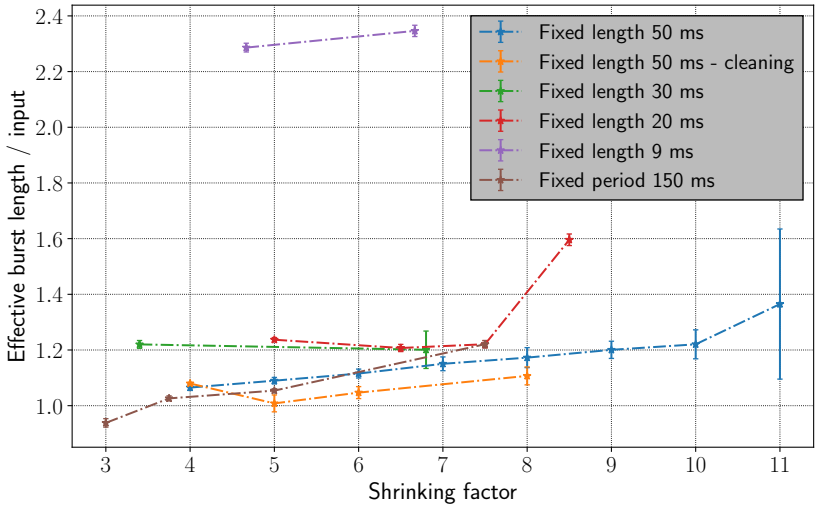

Figure 3.11: Measured effective burst length normalized to the demanded one for different burst extraction runs performed during the MD at the SPS.

is applied in order to have a δ-like pulse, the extracted particles will be distributed on a finite length interval: this is at the basis of the frequency transfer process of slow extraction, which will be the main topic of Chapter 4. To this effect, also the power converter response will contribute, since the regulator loop will have difficulties to follow the demanded reference current for smaller pulses (i.e. higher frequencies).

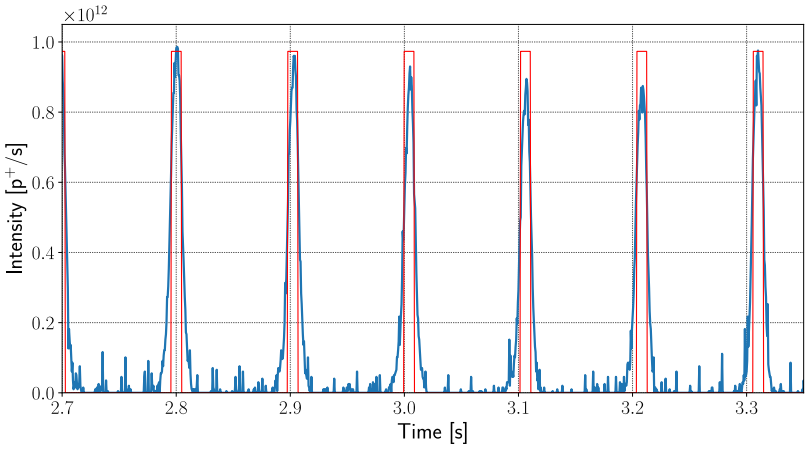

Figure 3.12: Measured spill (blue) obtained by demanding 9 ms as burst length, superimposed with a 9 ms square wave (red) for reference.

Figure 3.12 shows a clear example of the latter effect, where the obtained pulses by imposing a 9 ms length are about two times wider than the expected value (a square wave of 9 ms pulses would be the ideal expected output). This is a measured spill from the "fixed length 9 ms" data points reported in Fig. 3.11, and which in fact present a ratio of about 2.3 between the measured effective burst length and the demanded one.

Given that the main dependence on the measured effective burst length parameter is on the demanded burst length, it is reasonable to average the results belonging to data with different shrinking factors and the same demanded burst length. This allows to express the effective burst length as a function of the demanded burst length, as shown in Fig. 3.13.

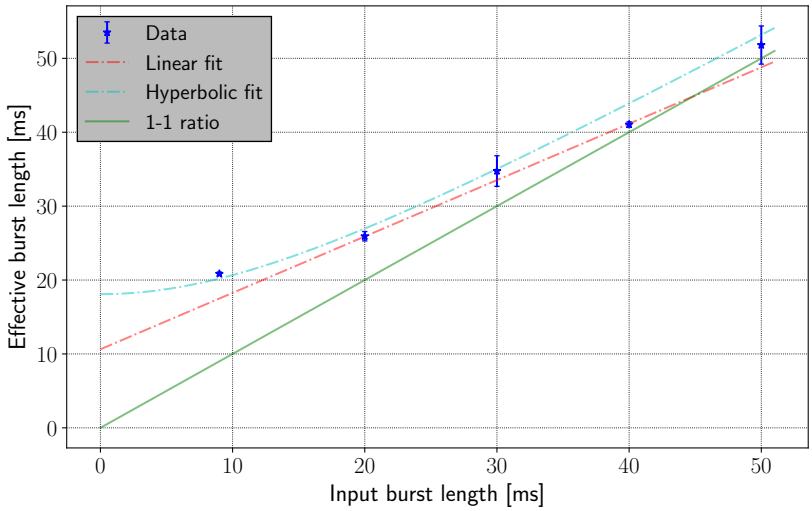

Figure 3.13: Measured effective burst length as a function of the demanded one (referred as input), fitted with a linear and an hyperbolic function, and compared to the 1-1 law.

It would be natural to think the 1-to-1 ratio law to be an asymptote for the behavior of the effective burst length data at large demanded burst lengths. It is in fact already evident from Fig. 3.13 that the burst lengths points larger than 40 ms are very close to the ideal behavior. The points with a smaller demanded burst length deviate significantly from the ideal situation instead, as expected from the considerations made above. Understanding how these points deviates from the ideal behavior is important to define the performance limits of the burst extraction: for instance, in Fig. 3.13, the data has been fitted with a linear and a hyperbolic function. The choice of the latter has been driven by the consideration of the 1-to-1 line to be an asymptote. The two laws predict two very different scenarios. According to the hyperbolic model, a burst length saturation at about ~ 18 ms would

make this value the minimum achievable burst length. The linear model, instead, would allow to reach progressively smaller lengths by lowering the demanded one.

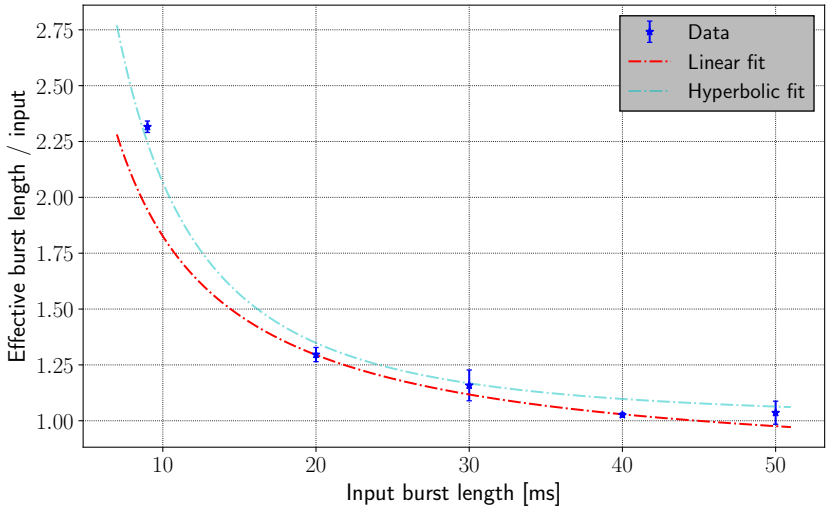

Figure 3.14: Ratio of the effective and demanded burst length as a function of the demanded burst length (referred as input), fitted with the same linear and hyperbolic functions of Fig. 3.13.

The same behaviour can be visualized taking the ratio of the measured and ideal effective burst lengths as a function of the ideal one (Fig. 3.14). The same linear and hyperbolic fits of Fig. 3.13 are reported, and both predict a divergence at zero input lengths.

3.1.4 An iterative approach to the burst extraction

The results from Figs. 3.13 and 3.14 have shown that a simple deterministic approach to the burst extraction (i.e. setting a demanded length and shaping the tune accordingly) is not enough to reach the ENUBET goal of 10-to-2 ms-long pulses. Moreover, the observed behavior of the effective burst length could imply an impossibility of reaching such goal, given the current machine configuration. It is then crucial to investigate what happens by demanding shorter and shorter burst lengths, with the aim of reaching the first experiment goal of 10 ms-long pulses.

This task can be performed in two ways: one is to manually scan the space for lower demanded burst lengths by using the developed application, while the second way would be to use an iterative approach based on an automatic spill optimization, until convergence is reached. This last approach is exactly what the Autospill control-room application (described in Sec-

tion 2.5) is performing for the operational continuous spill on a daily basis. Exploiting a well established operational tool to optimize and control the burst extraction scheme would be a great advantage in terms of stability and future operability of the new extraction scheme; hence, this has been the chosen approach.

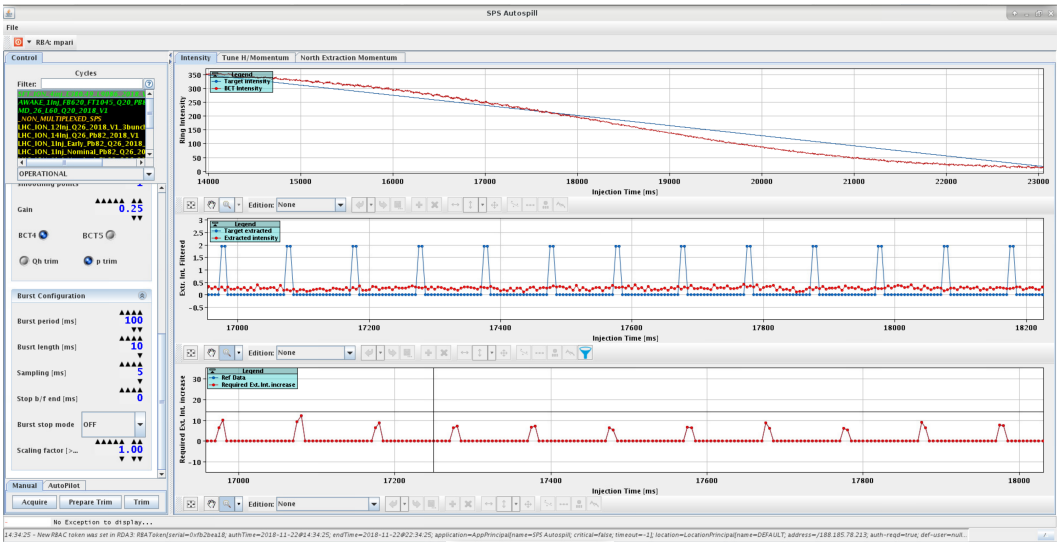

Figure 3.15: Screenshot of the upgraded Austospill application used during the last burst extraction tests at the SPS.

The Autospill application has then been upgraded in order to fully support the optimization of the burst mode slow extraction[2]. An example of some of the new developed features is visible in Fig. 3.15, which shows the demanded burst extraction spill (blue, middle plot) and the current flat spill (red). On the left side of the plots, a few of the burst extraction configuration parameters added in the upgrade can be observed.

The procedure outlined for the last burst extraction MD has been the following:

1. optimize the continuous extracted spill using the standard Austospill mode;

2. use the `burstControl` Python application set to a burst length and burst period of respectively 10 and 100 ms (i.e. the first goal of the ENUBET experiment).

3. enable the burst mode of the Autospill application (set on the same 10 over 100 ms burst operation) in order to optimize the tune set at point 2 to the demanded values.

[2]The application is written in Java and directly uses the native Java libraries for accelerator control of CERN, contrary to the developed Python application.

The reason why the `burstControl` application is used to set the initial condition for the optimization process is to simplify the task, minimizing the time needed and the possible complications from a bad power converter response. The latter issue, in fact, is not taken into account by the Autospill algorithm, while it has been proved to be well handled by the Python application.

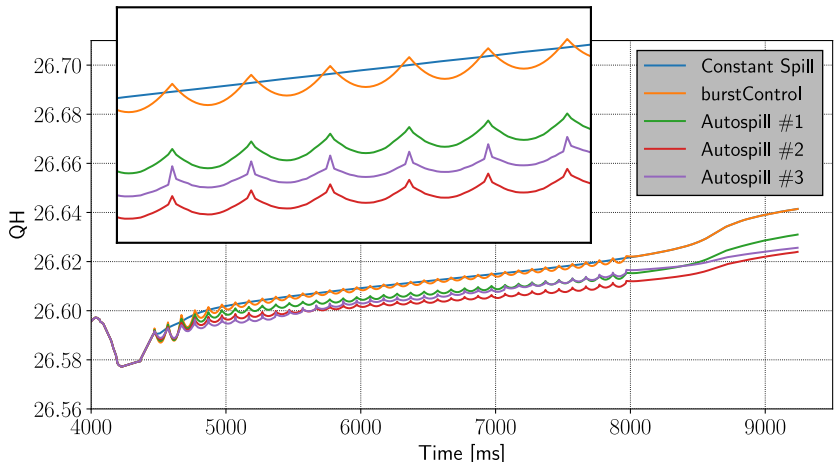

Figure 3.16: Flat-top tunes of a successful burst extraction spill optimization. The small plot shows a zoom of the tune functions of each optimization step.

Convergence to the demanded burst length value has been achieved during the tests, and Fig. 3.16 shows the evolution of the flat-top tune for every step of the optimization. In particular, the numerical optimization successfully shrinks the extraction parts of the tune in order to reduce the effective burst length, while keeping unchanged the parabolic shapes of the non-extracting tune regions initially generated with `burstControl`.

As Fig. 3.17 shows, a satisfactory convergence is reached in only 3 iterations: the average effective burst length is brought to about 10 ms from an initial value of 19 ms. This is an important result, and several conclusions can be drawn from it:

- Since the burst extraction parameters used in this optimization were 10 ms-long bursts repeated over 100 ms periods, this means that the feasibility of the first goal of the ENUBET experiment has been proven.
- The successful result has been reached only by acting on the machine horizontal tune through one operational control-room application (Autospill), and another one which has proven to be stable and reliable in the previous tests (`burstControl`). This implies that eventually using the burst mode slow extraction as an operational scheme would be rather straightforward.

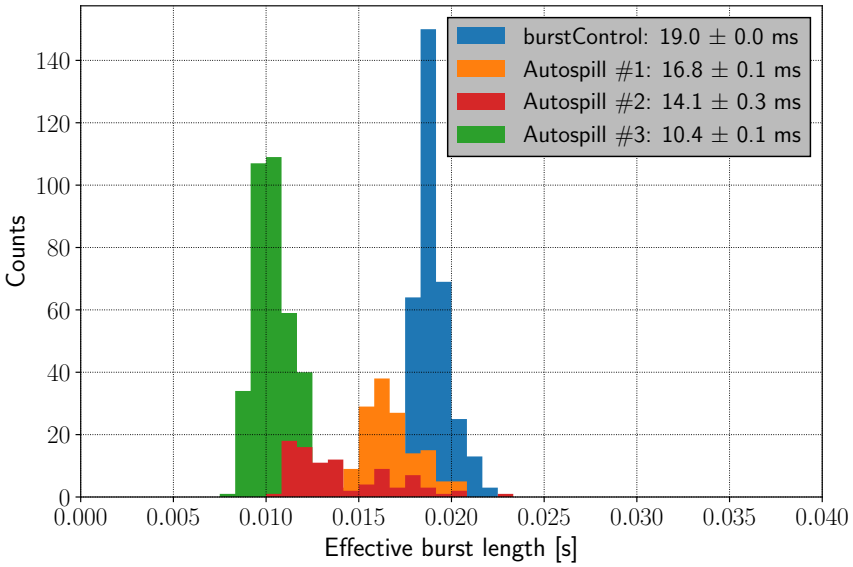

Figure 3.17: Distribution of the effective burst length for each extracted burst during the successful spill optimization with Austopill. The average of each histogram is reported in the legend.

- A convergence in only 3 iterations shows that there is margin for improvement: more iterations could not have been taken just for a matter of available time, but in principle the result would continue to improve.

- The reaching of the 10 ms goal, other than being a success for the experiment, proves that the behavior of the effective burst length parameter as a function of the demanded one is not the hyperbolic one from the fitted function of Figs. 3.13 and 3.14. Otherwise, such a results could not have been reached, given the foreseen early saturation of the effective burst length.

The success of the burst extraction optimization can be also be observed from the measured spill profile, reported in Fig. 3.18.

During the burst extraction machine developments the losses at the electrostatic septum have also been monitored. This is because particle losses at the electrostatic septum blade are one of the most important parameters for the operation of the slow extraction: any increase of losses with respect to the nominal case could represent a showstopper for any new extraction scheme.

Concerning the burst mode slow extraction, being it built on top of the existing QSWEEP extraction of the SPS, and being it based on a horizontal flat-top tune change, no increase of integrated losses is expected in principle.

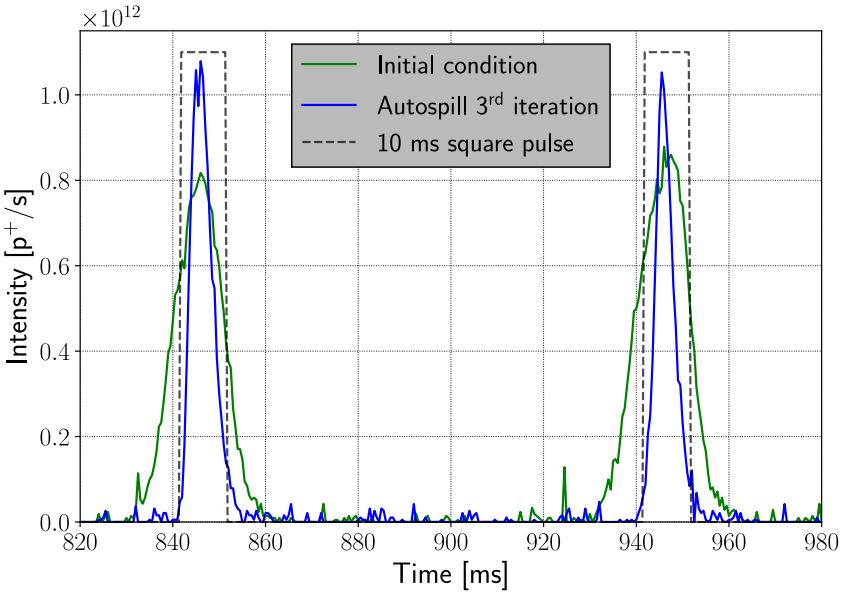

Figure 3.18: Comparison of the measured extracted bursts before and after the Autospill optimization of the burst extraction spill. The bursts are superimposed to a 10 ms square pulses for reference.

The experimental results, reported in Fig. 3.19, confirm this expectations. The reported data consist in the integrated losses at the 5 electrostatic septum modules, plus the first collimator downstream the electrostatic septum.

While the iterative approach to the burst mode slow extraction has been proven to be very promising, and allowed to reach the first goal of ENUBET, some limitations and issues have also been observed, and will be discussed in the following.

The main limitation is related to the evolution of the intensity of each burst during the Autospill optimization: the height of the bursts seem to be significantly randomized by the process.

In Fig. 3.20 a visual example of this randomization is evident from the measured spill intensity profiles: the burst heights at the end of the optimization reach amplitudes as big as three times the initial ones. In particular, observing the distribution of the heights of all the extracted bursts for each optimization step, it can be seen how the spread is increased up to about a factor 5 at the end of the optimization, as shown in Fig. 3.21.

Since the deterministic implementation of the burst extraction is shown to keep an acceptable spread of the burst heights (the absolute value of the σ's from Fig. 3.21 contains also the spread from beam intensity fluctuations between different extracted spills, which anyway is similar for every iteration

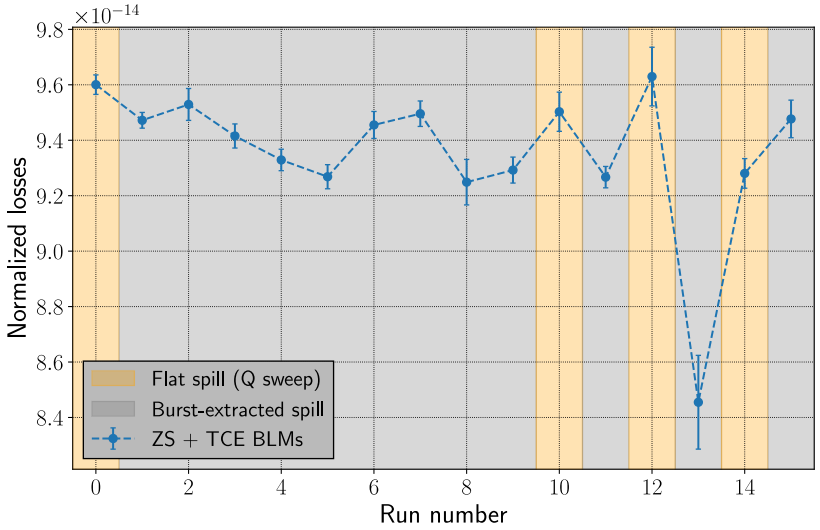

Figure 3.19: Integrated extraction losses (normalized to the extracted intensity) during the burst extraction tests. In the yellow areas the nominal continuous extraction was being operated, while in the gray areas the burst extraction was active. The dip corresponding to the run number 13 was due to a manual reduction of the extracted intensity during the tests.

step), the problem is intrinsically related with the Autospill application.

In order to fully understand and correct this issue it will be needed to further debug the application under real conditions, and this will only be possible after the end of the shutdown. However, another limitation of the Autospill for the burst extraction operation is likely to be related with this problem.

As described in Section. 2.5, the measured spill signals used for the feed-forward algorithm are not coming for the secondary emission monitor sampled at 2 kHz (e.g. spills of Fig. 3.20), but from DC BCT data, sampled at 5 ms. This is particularly problematic for the burst mode slow extraction, as to achieve effective bursts lengths of 10 ms (or less) the needed burst tune ramps have to be smaller than 5 ms. Despite seeming counter intuitive at first, it is still possible to reach tune ramps significantly smaller than the minimum sampling time, as shown in Fig. 3.22 and proven by the success of the burst length optimization. However, this offers less degrees of freedom for the optimization, and makes the application blind to burst intensity variations at scales shorter than 5 ms, making not feasible to reach shorter lengths.

On top of this, the spill signal obtained from BCT data is also noisier than the one measured by the secondary emission monitor. This does

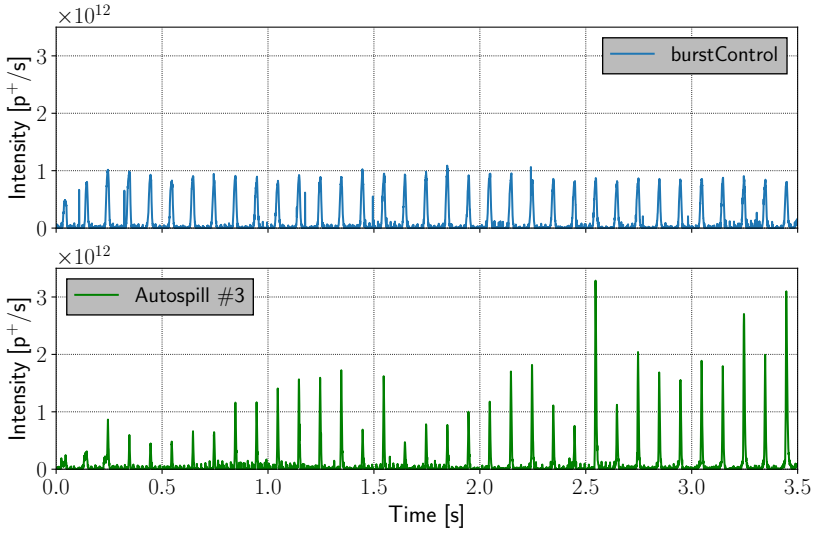

Figure 3.20: Top: spill created with the `burstControl` application as initial condition for the iterative optimization. Bottom: spill obtained at the third iteration of the Autospill, with the successfully optimized burst length.

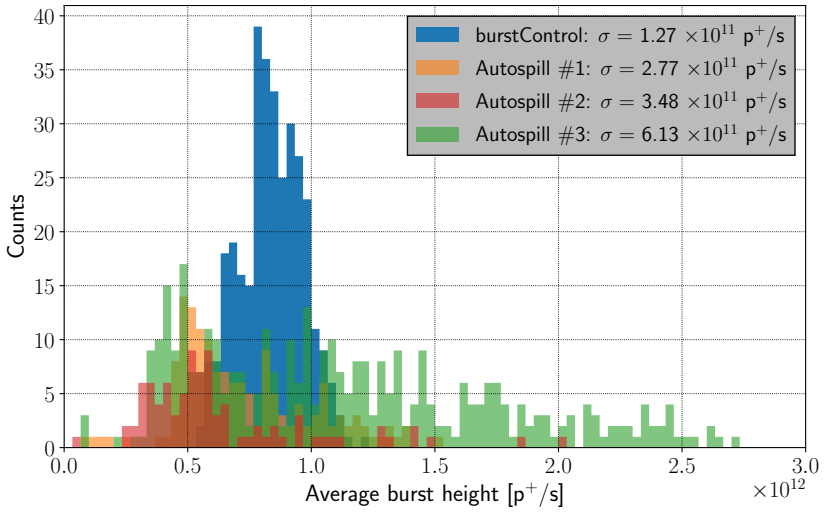

Figure 3.21: Evolution of the burst height distribution during the burst extraction spill optimization process. The standard deviation (σ) of each distribution is reported in the legend.

not represent a problem for the standard continuous spill extraction: the measured BCT signal can be under-sampled and filtered with smoothing al-

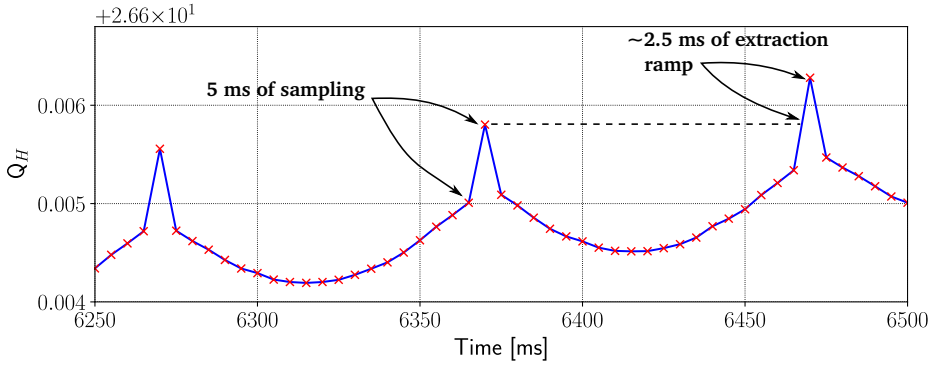

Figure 3.22: Detail of the tune ramp of the final Autospill iteration: the sampling points used to generate the tune signal are marked in red.

gorithms without losing useful information on the spill structure. In the case of the burst mode slow extraction both filtering and under-sampling are not viable options, and the noise can contribute to fluctuations in the height of the bursts, as evident from Eq. 2.23.

Figure 3.23: Example of the high level of noise on the measured spill (red curve, middle plot) during the burst mode slow extraction Autospill test.

An example of such effect can be observed in Fig. 3.23: the strong noise on the measured spill (red curve, middle plot) affects significantly the heights of the burst proposed for the tune trim (red curve, bottom plot). This phenomenon gets worse for smaller extracted intensities, where the noise on the BCT increases: this was actually the case during the Autospill tests for burst extraction, where the intensity was almost 2 orders of magnitude lower than the nominal one, due to requirements from other parallel MD's.

One possible solution which will have to be investigated for the future

is to further upgrade the Autospill control-room application to use the secondary emission monitor spill signal for the spill optimization. In this case, both the problems described above could be potentially solved, given that the secondary emission monitor has a sampling time of 0.5 ms, and a lower level of noise. This solution will certainly need to be investigated in order to solve the issue with the bursts intensities observed in Figs. 3.20 and 3.21, and to allow to reach burst lengths shorter than 5 ms.

As a final experimental limitation, which influences both the iterative and the deterministic burst extraction implementations, there is the fact that the regulator loop of the quadrupole power converters cannot work with sampling times shorter than 3 ms (already, power converters response problems have been observed with such sampling too, as reported in Fig. 3.5), for a matter of convergence. Nevertheless, as shown in Fig. 3.22, this is not preventing to reach extracting ramps shorter than 3 ms, but it will require a more sophisticated handling of the burst extraction implementation, in order to reach an optimum result.

3.2 Simulations and possible improvements for the burst mode extraction

After the efforts dedicated to the implementation, proof of principle, and experimental characterization of the burst mode slow extraction at the SPS described in Section 3.1, a simulation model of the new extraction scheme has been implemented. This model is based on a MADX [79] simulation of the SPS, which has been successfully validated with data on several different occasions since its original development [52, 61, 64, 65]. The main goal of the simulation work on the burst extraction is to investigate different configurations of the extraction scheme which would allow to reach the full 2-to-10 ms burst length range specified by the ENUBET project. In order to do this, the obtained experimental results will be first compared against the simulation to verify the agreement, for then investigating and simulating new possible improvements which could not have been tested before the CERN accelerator complex shutdown.

The simulation model of the QSWEEP extraction for the SPS uses a uniform momentum distribution for the particles and consequently a linear ramp for the main quadrupoles: this implies that the extracted spill will have ideally the shape of a square pulse, as per Eq. 2.27. This simplifies significantly the code, and automatically makes the simulated extracted spill correspond in shape to the result of the Austospill iterative optimization process. The burst extraction ramps are then built on top of the linear quadrupole ramps, using the concept developed in Subsection 3.1.1. Following the same idea of using a simplified approach that embodies the main characteristics

of the process, the sophisticated and frequently modified burst extraction ramp shapes of the type of Fig. 3.6 are not reproduced in simulation one to one, as neither are all the different burst period/length combinations tested in the first measurements and summarized in Figs. 3.10 and 3.11. Taking into account all of this would add too many parameters to the problem, and increase significantly the simulation times without adding much to the conclusions already drawn from measurements. The simulations will instead be focused on a burst extraction with a fixed repetition period of 100 ms, while trying to reduce the achievable effective burst length as much as possible, aiming for a 10 to 2 ms burst length: these are the extraction characteristics of interest for the ENUBET project.

The first thing that needs to be investigated is the dependence of the effective burst length on the part of the tune ramp between bursts, referred as f in Fig. 3.2 and Eq. 3.4. As it has been observed from the experimental results on the dumped intensity with and without momentum cleaning (Fig. 3.9), the shape of f influences the number of extracted particles between two ideal pulses: this is consequently inducing a variation of the effective burst length. This particular dependence could not have been scanned during the measurements due to the limited amount of available time and the main focus on minimizing the power converter overshoots (for which the f function plays an important role).

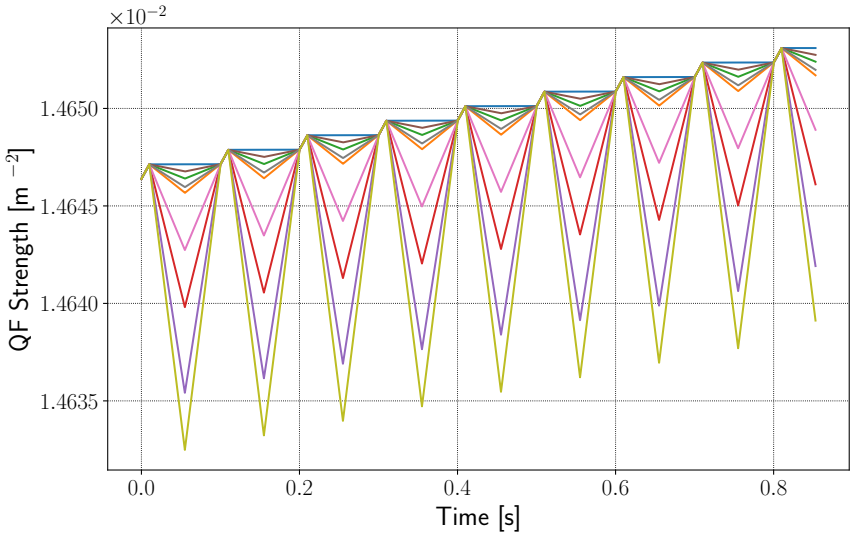

Figure 3.24: Simulated burst extraction quadrupole ramps for the same burst length and burst period parameters of 10 and 100 ms, but with different f functions.

The f function of the simulated burst extraction ramp is chosen to be a symmetric piece-wise linear "V"-shaped function, characterized by a single

parameter $\varkappa$ defined by the relation:

$$\varkappa \cdot Q\Big((n+1)\,\Delta t\Big) = \min_{t\in[\lambda,\Delta t]} f(t+n\Delta t) \;\Big|\; \varkappa \leqslant 1 \tag{3.15}$$

where the formalism is referred to Eq. 3.4. Equation 3.15 simply means that $\varkappa$ equals to the ratio between the minimum of the f function, and the final value of the last extracting ramp. For this reason $\varkappa$ will be referred to as fractional tune comeback. Figure 3.15 shows an example of different simulated focusing quadrupole ramps for different values of $\varkappa$: the smaller the value of $\varkappa$, the faster the f function, while a value of $\varkappa = 1$ implies f to be an horizontal line.

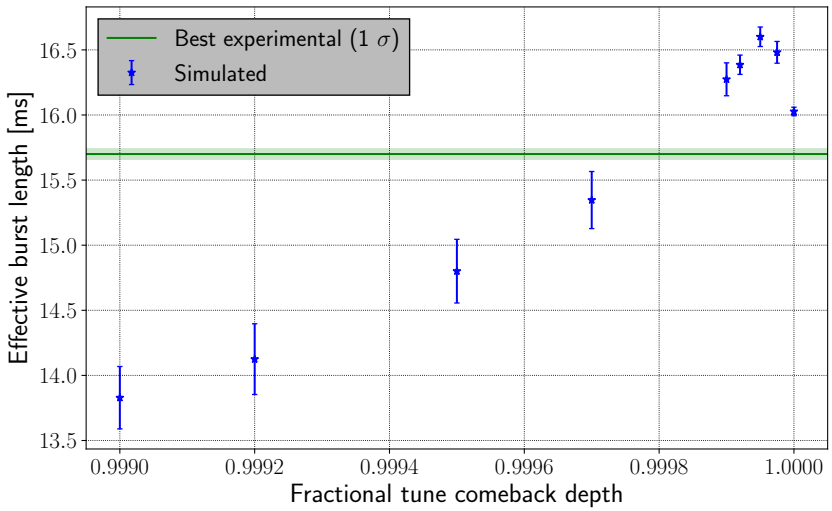

Figure 3.25: Simulated average effective burst length as a function of $\varkappa$ (fractional tune comeback depth), with fixed burst length and period of 10 and 100 ms, respectively.

A scan for different values of $\varkappa \in [0.999, 1]$ is simulated with the fixed burst extraction parameters of 10 ms burst length and 100 ms burst period. The average effective burst length is computed from the simulated spills for each value of $\varkappa$. The minimum value of $\varkappa = 0.999$ represents a realistic lower limit case according to what has been observed during the machine tests. The result of this scan is reported in Fig. 3.25 and compared with the best experimental value of effective burst length obtained for burst period and length respectively of 100 and 10 ms. It can be observed that the maximum effective spill length is not obtained for $\varkappa = 1$ (i.e. horizontal line), but for a slow double ramp: this is because a larger particle momentum space is crossed with a tune slow enough to extract most of the particles still in the

machine. The experimental case was often presenting a $\varkappa$ around the maximum of Fig. 3.25, given that the dumped intensity was minimized. Despite the significant added layer of complications present in the experimental case (e.g. bad power converter responses, more complex tune shapes, etc.), the best experimental point falls inside the range of minimum and maximum effective burst lengths foreseen by the simulated scan. However, this is not valid in the general case, which is usually characterized by a larger effective spill length, as it will be further discussed in the following.

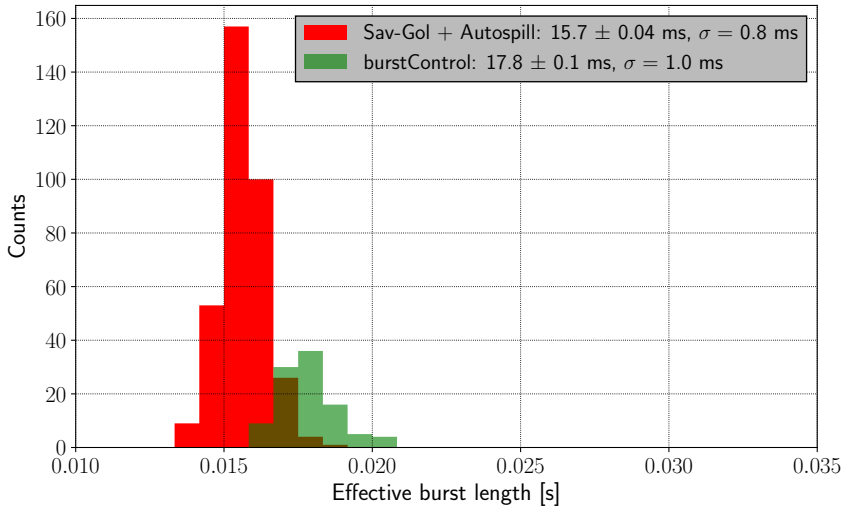

Figure 3.26: Best experimental configuration for effective burst length (red histogram) compared with the average case (green histogram). The histograms averages and standard deviations are shown in the legend.

The difference between the typical case and the best experimental point is shown in Fig. 3.26: while the typical cases are obtained used the deterministic algorithm previously discussed, the best experimental point has been obtained via a "one-shot" burst extraction setting with Autospill, where the measured continuous spill had been smoothed using a Savitzky-Golay filter [80]. The latter result has the drawback of significantly randomizing the intensities of the extracted bursts (at least by a factor 2 with respect to the `burstControl` case), in line with what observed in the previous section and shown in Figs. 3.21 and 3.20. On the other hand, it is evident from the mean value of the histograms of Fig. 3.26 that the typical case of the experimental effective burst length lies above the maximum value foreseen by the simulated scan of Fig. 3.25. This result is not surprising. As mentioned above, the typical $\varkappa$ value of the experimental case is near the maximum of Fig. 3.25. Adding to this a more complex tune shape with in average a lower tune speed (as for example a parabolic shape) and some overshoots

and ringing from the power converters, it follows that more particles should be extracted between two bursts in the experimental case with respect to the modeled one.

The results of the scan reported in Fig. 3.25 can be used to define a range in $\varkappa$ delimited by the two values which correspond to the minimum and maximum effective burst length, referred to as $\varkappa_{\text{min}}$ and $\varkappa_{\text{max}}$. For each of these values, a scan of the burst length parameter (keeping the burst period fixed at 100 ms) can be simulated: the result is shown in Fig. 3.27.

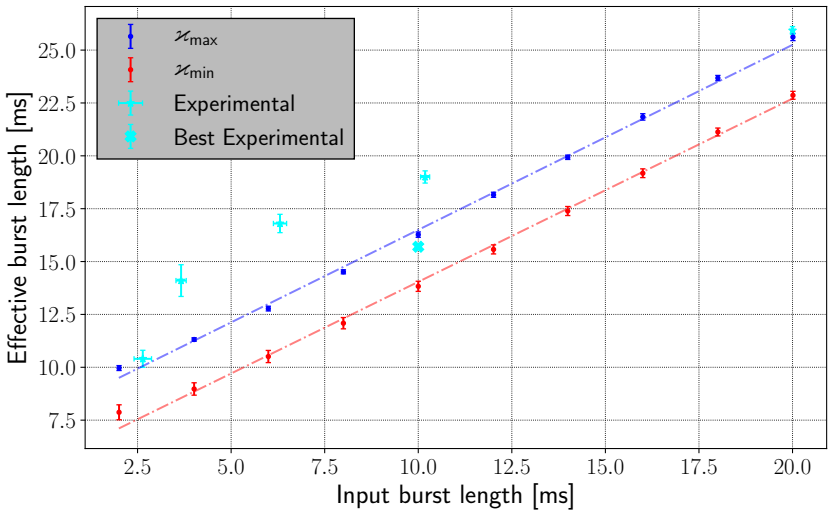

Figure 3.27: Simulated burst length scan for 100 ms burst period (red and blue points) compared to the experimental data (cyan points).

The two set of points obtained from the scan (red and blue in Fig. 3.27) define a region in which experimental data should approximately lie in if neglecting hardware effects and more complex tune shapes. The experimental data in the range of interest is also reported on the plot: for the Autospill optimization data (where there is no such a thing as demanded burst length) an effective demanded burst length is computed using a technique based on Fig. 3.22.

Two important remarks can be drawn from the result of the scan and its comparison with data reported in Fig. 3.27:

- The two set of points obtained from the simulated burst length scan (red and blue) show that the behavior of the effective burst length as a function of the input one is linear in a good approximation: slight saturation effects appear to be visible below $\sim$ 2.5 ms, but without significant deviations from a linear law in the range of interest.

- The data points in good agreement with the simulation results show

that it is possible to minimize the hardware effects and tune shape influence, and follow a simple model. When not compatible with the simulation, the experimental points lie above the simulated range, as expected from the effects of a bad hardware response and complex tune shapes.

The simulation results of Fig. 3.27 have been obtained using the nominal slow extraction parameters of the SPS, i.e. the same used for the burst mode slow extraction tests. It follows from the simulation results that under these conditions the minimum effective burst length that could realistically be reached in operation is between 10 and 7 ms.

One possible way to reduce the minimum achievable burst length would be to act on the extraction sextupole strength of the machine. The 3-turns amplitude growth of an unstable particle which moves along one of the phase space separatrix arms is:

$$\Delta A = \frac{3}{4} V_{\mathrm{ss}} A^2 \tag{3.16}$$

where A is the normalized phase space amplitude, and V_{ss} the virtual extraction sextupole strength. It is clear from Eq. 3.16 that both the speed and the acceleration of the motion along the separatrix are locally proportional to the virtual sextupole strength. Using the typical SPS amplitude distribution, the time-to-extraction of every particle can be computed using Eq. 3.16. A visualization of the computation is shown in Fig. 3.28: a particle is considered extracted when its trajectory crosses the extraction amplitude of the machine (corresponding to the electrostatic septum).

By repeating this process for different virtual sextupole strengths, the advantage gained from increasing this parameter can be understood.

The result is reported in Fig. 3.29, which shows the time-to-extraction intervals distribution for the value of the nominal SPS sextupole strength scaled by a factor 1, 2, and 4. The important thing to observe is that increasing the sextupole strength not only shortens the average time that particles take to be extracted, but shrinks also the distribution, as it can be seen in more detail from the small upper plot and legend of Fig. 3.29. In particular, by using Eq. 3.16, the difference in time-to-extraction $\Delta\tau_{ij}$ between two different amplitudes A_i and A_j from Fig. 3.28 can be approximated as:

$$\Delta\tau_{ij} \simeq \frac{4}{3}\frac{1}{V_{\mathrm{ss}}}\left(\frac{\Delta A_{ji}}{A_i\, A_j}\right) \tag{3.17}$$

where it can be noticed the inverse proportionality to the virtual sextupole strength. The process used here is a simple approximation of a momentum based slow extraction (with zero dispersion at the septum), as it is the case for the SPS, where Figs. 3.28 and 3.29 represent the extraction of a thin momentum slice (i.e. containing the full amplitude distribution). In the

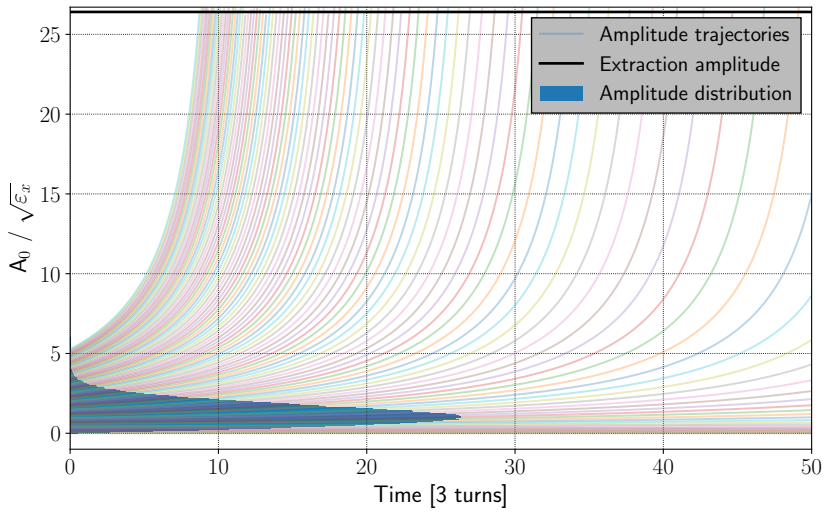

Figure 3.28: Amplitude trajectories of unstable particles along the separatrix during extraction (colored lines). The typical SPS amplitude distribution corresponding to a Gaussian particle distribution is also shown.

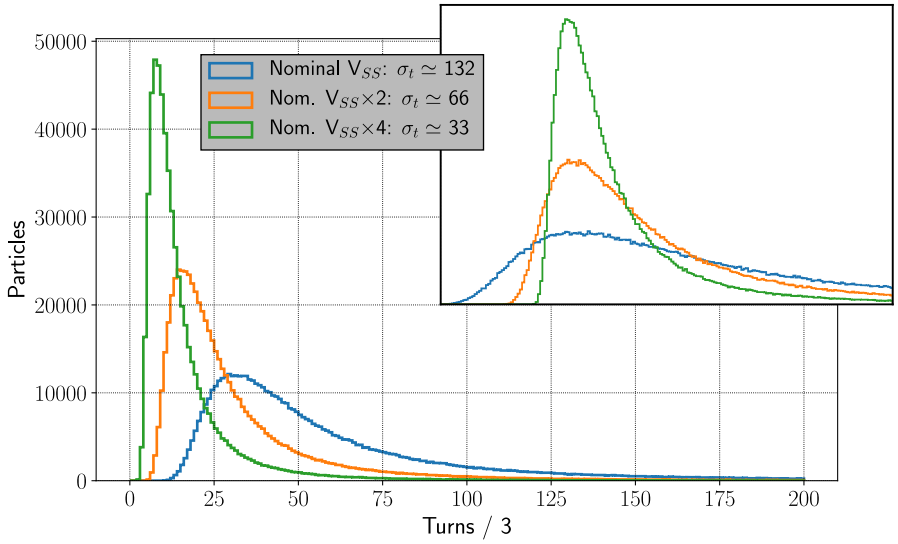

Figure 3.29: Time-to-extraction distributions corresponding to the situation of Fig. 3.28, obtained for different values of virtual sextupole strength. The small plot shows the distributions aligned to their center, while the standard deviations are reported in the legend.

real case, particles are not only moving along the separatrix arms, but also on the sides of the phase space triangle and on more external phase space

trajectories. A full analytical computation can be found in [44], but does not change the trend observed in Fig. 3.29.

The shrinking of the time distances is what is needed to achieve shorter burst lengths in the burst mode slow extraction. For this reason, a simulated scan of the demanded burst length parameter analogous to the one reported in Fig. 3.27 is performed with the nominal extraction sextupole strength of the SPS scaled by a factor 2 and 4.

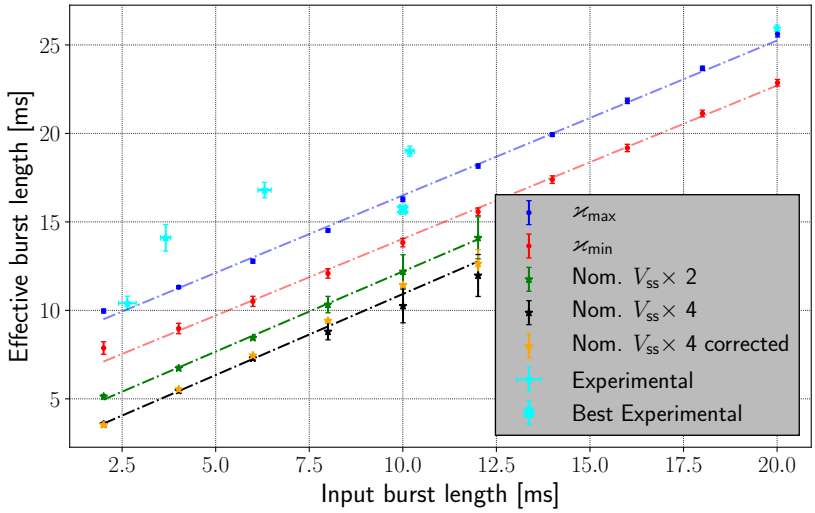

Figure 3.30: Burst length scan: the results relative to the V_{ss} increase are also reported. The "corrected" points refer to a spiral step correction.

The results of the scan, reported in Fig. 3.30, confirm that increasing the sextupole strength does indeed help in reaching smaller burst lengths. In particular, the minimum burst length for a 2 ms input gets reduced of about a factor 1.5 for each two-fold increase of the sextupole strength, hence jumping from a minimum of $\sim$ 8 ms for the nominal case, to $\sim$ 5 ms with the two-fold increase, and $\sim$ 3.5 ms for the four-fold increase. The dependence of the effective burst length to the sextupole strength can be observed in Fig. 3.31.

In particular, the shorter the demanded burst length the bigger is the fractional reduction of the effective burst length for higher sextupole strengths. This is well in line with the simple example which led to the results of Fig.3.29 and Eq. 3.17, which are assuming the extraction of a thin slice in the momentum distribution: the shorter the demanded burst length, the better this simple approximation will hold. This implies that a reduction of a factor V_{ss} would be the asymptotic limit for infinitely short demanded lengths. The result of Fig. 3.31 explains why an attempt to increase the sextupole strength did not bring any observable result during the machine tests of the burst ex-

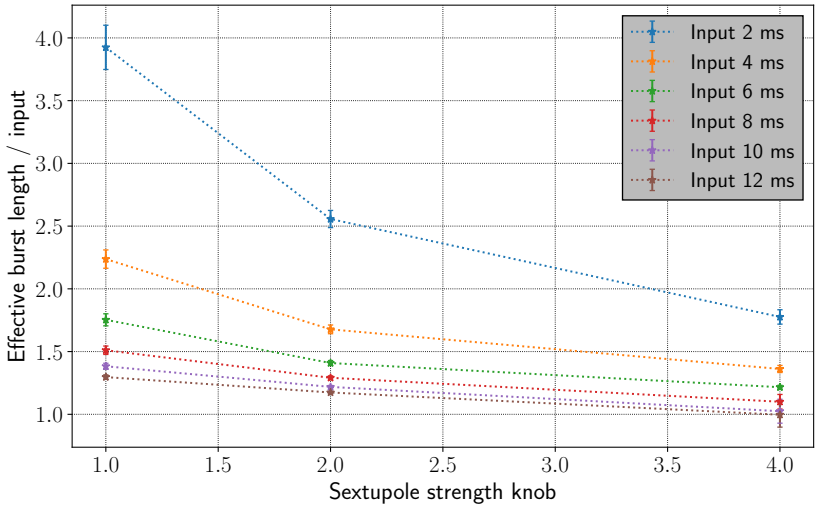

Figure 3.31: Effective burst length to demanded burst length ratio as a function of the sextupole strength scaling factor (where 1 is the nominal strength). The reported set of points are grouped according to the same demanded burst length, as indicated in the legend.

traction. The demanded burst length was 10 ms, and the sextupole strength was increased only of a factor $\lesssim 1.5$. Adding the small improvement foreseen by Fig. 3.31 to the hardware effects present in the experimental case explains why no systematic reduction was observed.

One important issue related with increasing the sextupole strength is the consequent increase in the spiral step, which is defined as the maximum horizontal jump of a particle into the electrostatic septum (Eq. 2.19, reported here for brevity):

$$\Delta X_{\mathrm{ss}} = \frac{3}{4} V_{\mathrm{ss}} \frac{X_{\mathrm{ES}}^2}{\cos\theta} \tag{3.18}$$

where X_{ES} is the electrostatic septum position and θ is the extraction separatrix angle. In the SPS nominal extraction, the extraction bump is adjusted in order for the spiral step to be justified inside the electrostatic septum cathode-anode gap (Fig. 2.12). When scaling the sextupole strength, the spiral step will scale of the same factor, as per Eq. 3.18, and the beam can eventually hit the cathode. This happens for both the simulated cases: the extracted separatix is shaved off by the septum cathode, as it is possible to notice from the left plot of Fig. 3.32.

The extraction bump is then re-adjusted in order to reproduce the same nominal spiral step for the two cases of sextupole strength increase. The result for the four-fold sextupole strength increase is reported in the left

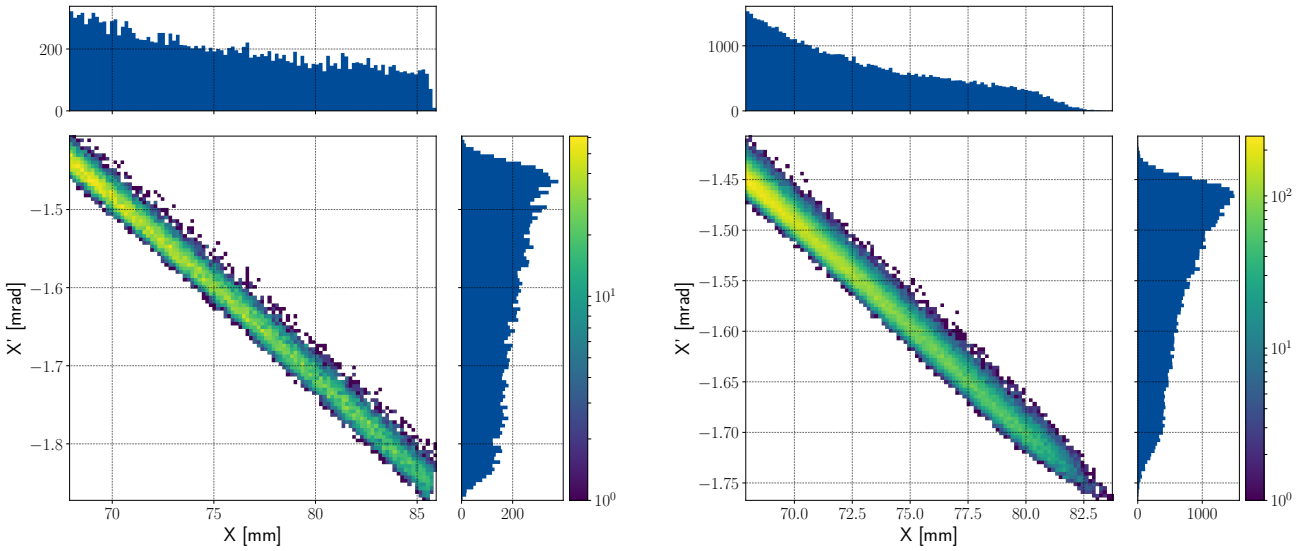

Figure 3.32: Extracted beam at the electrostatic septum for a sextupole strength four times the nominal one. Left: keeping the extraction bump in the nominal configuration. Right: correcting the extraction to fit the extracted beam into the septum aperture.

plot of Fig. 3.32. In particular, the extraction bump needs to bring the closed orbit closer to the septum, and so reducing the value of the extraction amplitude of Fig. 3.28. The orange points of Fig. 3.30 represent the scan results for the $\times 4$ sextupole strength increase with a corrected extraction bump: they show that the burst length behavior is not affected by this correction, being the two simulated data sets (with and without the bump adjustment) compatible within the error. The reason of this can be routed to Eq. 3.17, which shows that the extraction time difference between two particles does not depend on the extraction amplitude value.

One important drawback of this procedure is the increase of extraction losses at the electrostatic septum blade. By reducing the extraction amplitude, the density of particles that hit the septum blade increases, as it can be observed by comparing the X-projections of the two plots of Fig. 3.32. In particular, the losses on the septum blade for the two-fold and four-fold extraction sextupole strength scaling increase respectively by about a 20 and 50% with respect to the nominal case. On top of this, another problem consist in the fact that the closed orbit can get too close to the septum blade.

As shown Fig. 3.33, the acceptance (defined as the distance of the closed

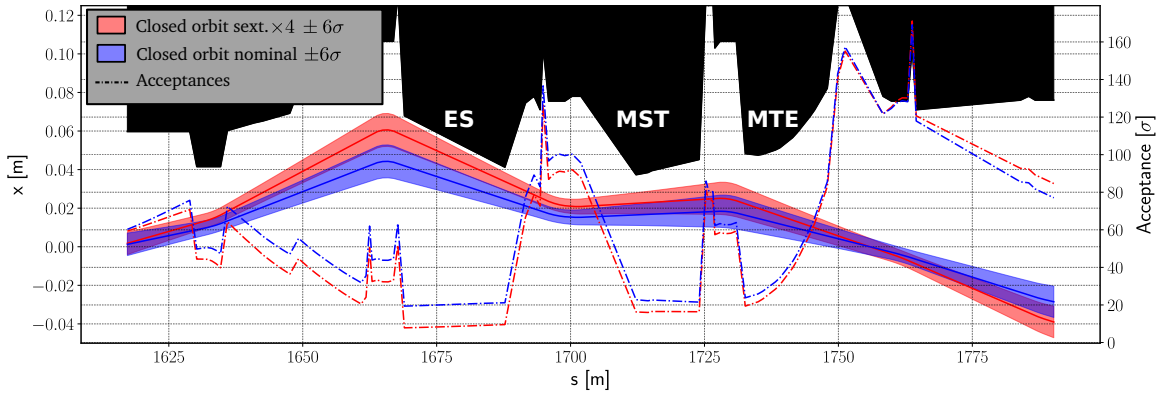

Figure 3.33: Effect on the closed orbit and acceptance of the bump scaling for the four-fold sextupole strength increase. The elements names ES, MST, MSE refer respectively to the electrostatic septum, and to the first and second magnetic septa.

orbit from the machine apertures in units of beam sigmas) at the beginning of the electrostatic septum jumps from about 19.4 to 7.8 sigmas when increasing the sextupole strength by a factor 4. Both the significant acceptance reduction and losses increase induced by the factor 4 larger sextupole strength make this particular configuration not a real possibility for operation.

An alternative technique to further reduce the minimum effective burst length achievable in operation would be to use amplitude extraction. As described in Section 2.5 the SPS nominal slow extraction is a momentum extraction (e.g. Fig. 2.16). In an amplitude extraction the beam would be extracted instead along its amplitude distribution, meaning that the chromaticity should be reduced nearly to zero. Figure 3.34 shows a typical example of a Steinbach diagram in the case of amplitude extraction (where the chromaticity has not been set exactly to zero in order to better visualize the particle density).

The significant advantage in terms of reduced effective burst length can be understood by considering the simple example of Figs. 3.28 and 3.29: only a slice of the amplitude distribution is extracted, which implies that the extraction time distributions of Fig. 3.29 will be extracted a slice at a time, and so with a minimal time spread. With the same premises as before, the actual situation is more complicated [44], but does not change the overall trend of significant time spread reduction.

The amplitude extraction is simulated at the SPS by matching the chromaticity to zero, and leaving the remaining extraction optics the same. One consequence of the amplitude extraction is that the spill will take the shape of the beam amplitude distribution, following an analogous expression as

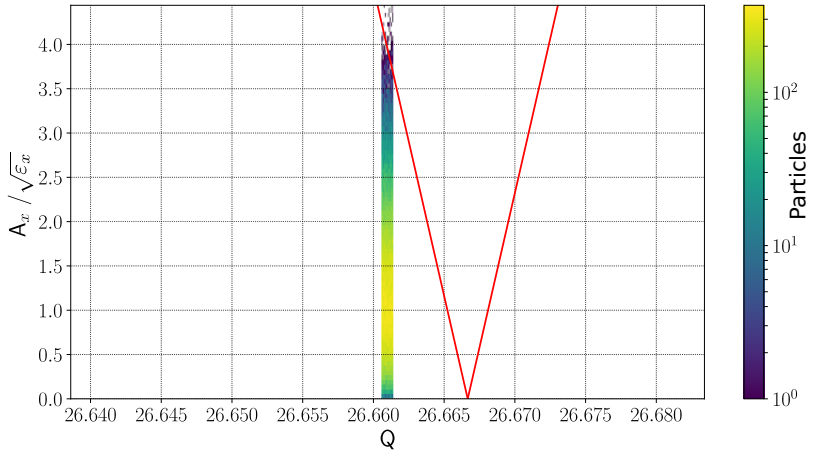

Figure 3.34: Steinbach diagram for a small chromaticity (i.e. amplitude extraction): the beam will be extracted along its amplitude distribution.

Eq. 2.27 for the momentum extraction. Since the phase space distribution is a 2-dimensional Gaussian, this means that the tune ramp needs to be optimized in order to extract a flat spill, as done for the operational extraction in the experimental case. The results of this process are reported in Fig. 3.35, where both the simulated amplitude extraction spill and the corresponding extraction ramps are shown.

Simulating an effective burst length scan using the amplitude extraction leads to the very promising results of Fig. 3.36.

The remarkable result is that the achieved effective burst length for a demanded 2 ms is even lower than what has been obtained for the case of the four-fold increase in sextupole strength: about 3.1 ms against 3.5 ms. Considering that the rest of the machine parameters have been left untouched (extraction sextupole strength included), this result seems very promising for reaching the full range requested by the ENUBET experiment without any increase of extraction losses. One thing that has to be mentioned, anyway, is how the slope of the fitted linear law on the amplitude extraction points of Fig. 3.36 is not aligned with the other results, and, moreover, the higher demanded length points seem to saturate and stop following the linear law at all. In order to better understand what this implies, it is useful to look at the values of the effective burst length normalized by the demanded one, as shown in Fig. 3.37.

In this case, it is evident that, while for all the other points the asymptotic ratio for large demanded lengths is 1 (as expected), for the case of the amplitude extraction the ratio gets below 1 just after the small value of 5 ms of demanded burst length, for then decreasing even further. This does not mean that the full intensity of a burst is extracted in a time smaller than

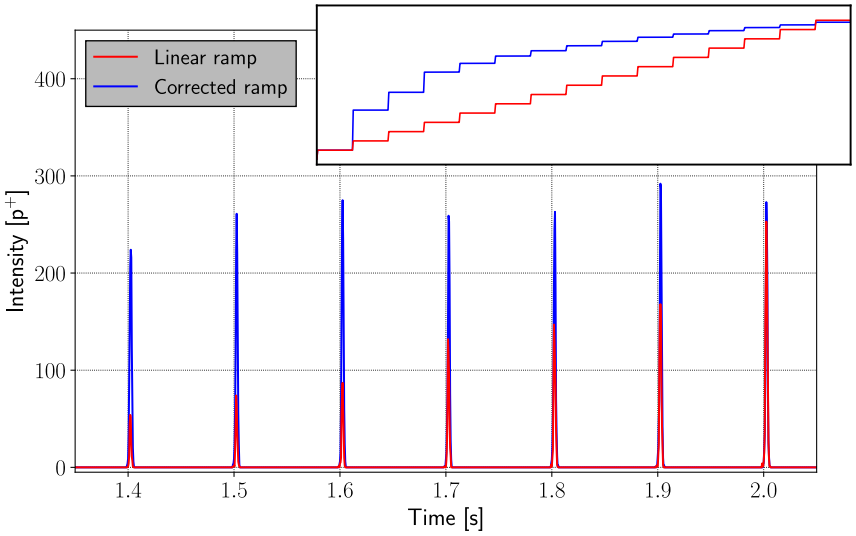

Figure 3.35: Simulated burst mode spill with amplitude extraction with a linear extraction ramp (red) and an optimized one (blue). The upper plot shows how the focusing quadrupole ramp has been reshaped to achieve a constant intensity of the bursts.

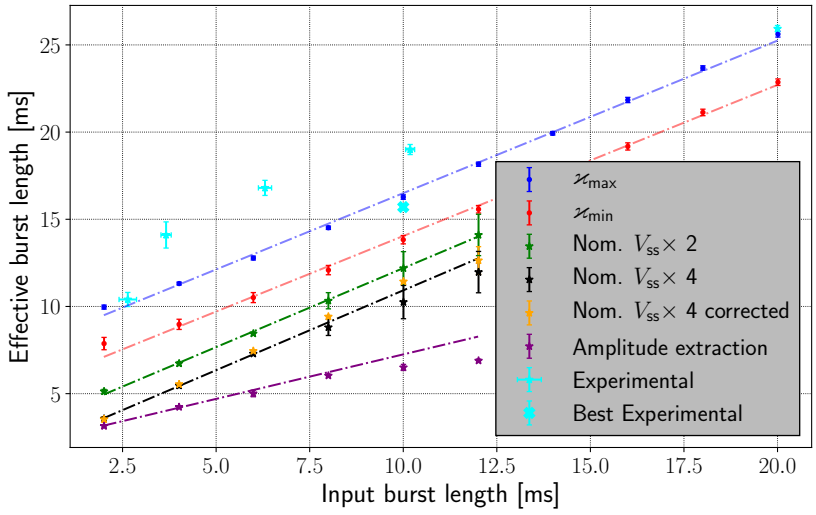

Figure 3.36: Results of the simulated burst length scan for amplitude extraction (purple) compared to the previous data.

the demanded one (which would violate causality), but that the quality of the extracted burst degrades significantly. In fact, as discussed earlier, the employed effective burst length parameter is also a measure of the quality of

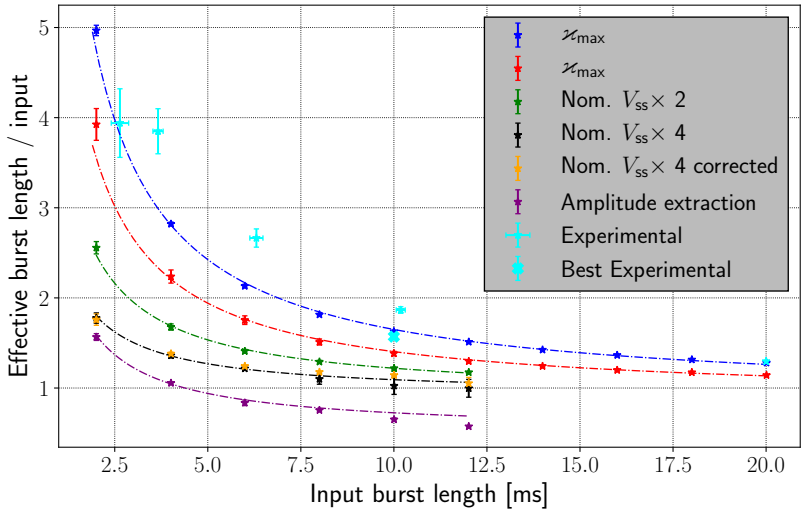

Figure 3.37: All the results from the burst length scans with the effective burst length normalized by the corresponding value of demanded burst length.

a pulse: up the point the extracted bursts are regular (Gaussian to square like), the parameter estimates their actual width in a good approximation, but when strong asymmetries or noise are present, the spill quality sensing of the parameter will intervene by decreasing the value of the result. This is a powerful way to quickly detect when a reasonable burst quality is lost, as it can be easily noticed from Fig. 3.37, provided that the actual shapes of the bursts are monitored in order to avoid false negatives (i.e. bad quality bursts with low effective spill length which are taken as good).

Increasing the demanded burst length in the amplitude extraction results in the degradation of the burst pulses, as shown in Fig. 3.38. For demanded burst lengths larger than 3 ms, most of the intensity gets extracted in a first peak, for then being followed by a lower intensity trail of pulses. By increasing the burst length, a wider amplitude range will be crossed in a single shot. This means that the difference in extraction time of the highest amplitude particles and the ones located on the fixed points of the resonance with respect to the others will increase.

The result of Fig. 3.38 and 3.37 suggests that the amplitude extraction is a suitable and convenient solution for very small demanded burst length, which is where the chromatic extraction fails to provide adequate bursts.

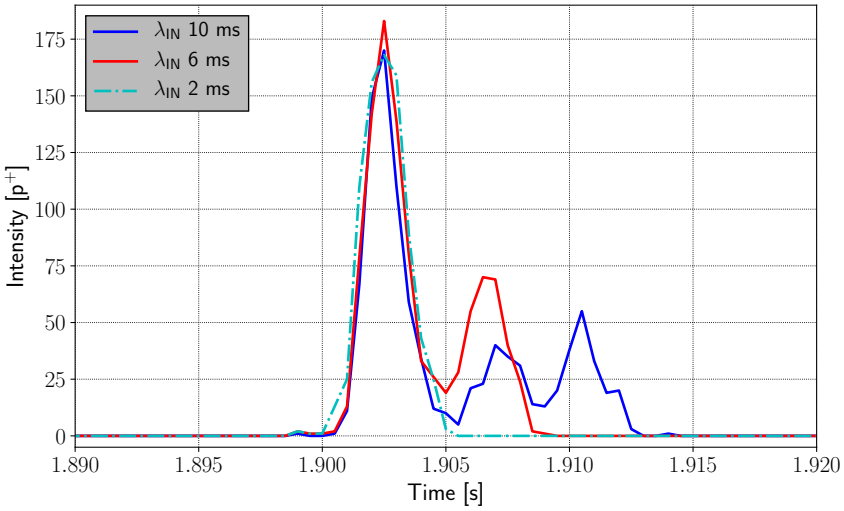

Figure 3.38: Example of the burst quality degradation for the extracted pulses when using amplitude extraction. The legend specifies the different demanded burst length to which each pulse correspond..

3.3 Conclusions

In this Chapter the new extraction scheme envisaged by the ENUBET project has been successfully developed, tested, and simulated. In particular, the machine tests have characterized the scheme in terms of its limitations and found out that its implementation can be performed in a straightforward way, based on a horizontal tune change. The residual intensity in the machine after extraction does not seem to represent an issue for the range of interest of ENUBET, while the length of the extracted bursts is the most sensible parameter. The problem of convergence to the demanded value of burst length has been addressed by employing an iterative approach based on a fully operational control-room application; this allowed to reach the first experimental goal of 10 ms-long extracted pulses, repeated at a frequency of 10 Hz. The issues related with the iterative process have been identified (as the limited sampling available and randomization of the burst intensities), and ideas on how to address them have been proposed and will be tested in the future. Given the experimental observations on the dependence of the extracted pulses on the particular tune shape used for the extraction, the possibility to use the full tune shape parameters for an optimization algorithm is also being considered as a future improvement.

A simulation model of the burst extraction has been developed, finding a satisfactory agreement with experimental data despite the numerous com-

plications present in it. The simulation model has been used to investigate possible methods for further reducing the extracted burst length.

The first method consists in increasing the extraction sextupole strength, which proved to reduce the minimum achievable burst length from 8 to 5 ms for a factor 2 increase of the extraction sextupole strength. A drawback of this technique is an increase in the extraction losses due to a reduction of the distance of the closed orbit to the electrostatic septum needed to keep the spiral step constant. This increase reaches quickly prohibitive values (e.g. 20% for a factor 2 and 50% for a factor 4), showing that increases significantly bigger than a factor 2 have to be excluded. However, it is important to notice that the burst mode slow extraction could be used in combination with the very promising losses reduction methods which have been successfully tested at the SPS during 2018. The first one is the electrostatic septum local shadowing with a bent silicon crystal, which has been proven to achieve losses reductions between 20 and 40% [64, 81]. In an analogous way, the use of a passive scatterer for electrostatic septum losses reduction has been proved to reduce losses by a 15% [82]. These two schemes only consist in the insertion of a mechanical device in the machine, placed upstream the electrostatic septum: they do not depend on the used extraction scheme. This would allow to reach higher values of sextupole strengths, while keeping the losses close to nominal. Another technique which could be even more advantageous is the one of separatrix folding through octupoles [66–68]. This clever method exploits octupole magnets in order to avoid that the extraction separatrix is shaved off by the septum cathode when high values of sextupole strength are used; hence with no need of correcting the spiral step. It has been proven at the SPS that this technique can lead to loss reductions as high as 40% [65]. Since the reason for the loss reduction of this technique is an increase of the sextupole strength, it would be natural to couple it to a burst mode slow extraction scheme with increased sextupole strength. In fact, differently from the case of crystal shadowing (and diffuser), in which a losses reduction method is operated to cancel the losses increase from the burst extraction, the separatrix folding technique would allow to operate the burst mode slow extraction with increased sextupole strength with the same losses reduction of 40% reached for the nominal extraction case. The one possible complication is that only the COSE scheme is suitable for the separatrix folding technique. As mentioned in Section 3.1, a COSE implementation of the burst mode slow extraction presents the hypothetical hardware-side showstopper of some of the magnets not being able to follow the burst extraction ramp. However, its design and software-side implementation would be as straightforward as the tested QSWEEP version. Given the significant advantages which would follow from this scheme, it could be worth it for the future to test whether the SPS magnets can handle a COSE implementation of the burst mode slow extraction.

The second method for effective burst length reduction investigated in simulation has been the amplitude extraction. The results proved the method to be very promising for short demanded burst lengths, where results even better than the case of a four-fold sextupole strength increase can be achieved, leading to minimal effective burst lengths around 3 ms. One positive aspect of this method is that it does not interfere with losses at the electrostatic septum, being the chromaticity the only extraction parameter which gets modified. In case the COSE scheme would turn out to be not compatible with the burst extraction, the amplitude extraction method could be paired with a dynamic bump in order to fulfill the Hardt condition and minimize losses [54, 55]. The magnets will have to be able to follow the burst extraction ramp, but in this case only a significantly smaller number of magnets (the extraction bump ones) will have to comply, as opposed to the case of the COSE scheme.

Finally, most of the range required by the ENUBET experiment of 2 to 10 ms-long bursts seems to be achievable exploiting both the techniques explored in simulations: amplitude extraction for the lower end, and momentum extraction (with the possible aid of increased sextupole strength) for the upper end. The latter case of momentum extraction has also been experimentally proven to be able to reach the upper end of the ENUBET burst length range for the nominal SPS extraction parameters.

Chapter 4

Study of the slow extraction frequency response of CERN-SPS

The quality of the extracted spill in terms of frequency noise will be the main subject of this chapter. The presence of unwanted frequency components on the extracted spill (typically coming from the main power supplies) is a very old and important issue for slow extraction, which can lead to a significant deterioration of the quality of the beam used by the experiments. In the following, the process of slow extraction will be studied using a transfer function formalism, in order to understand and characterize the frequency response of the SPS slow extraction. This subject is deeply connected to the previous study on the burst extraction, as the very same burst extraction process could be seen as an ideal input signal (the demanded tune signal) passing through a transfer function which attenuates its high frequency components, and hence generating a burst extracted spill with wider pulses than expected. A good understanding of the frequency response properties for standard extraction is also required to understand special cases as the burst extraction. While the efforts of this chapter will be focused on suppressing unwanted frequency components, an understanding on how to amplify them can then be easily extrapolated.

The present chapter is structured in six main sections. Section 4.1 will introduce the problem of frequency transfer in slow extraction. In Section 4.2 different simulation models of the process will be developed, and important aspects of it will be investigated exploiting the developed models. In Section 4.3, the data from dedicated measurements at the SPS will be used to validate the simulation models, while in Section 4.4, another validation will be attempted by using operational data from physics runs. Section 4.5 will show how the developed simulation models could be used to look for

machine optima in terms of ripple suppression, just by scanning a few key slow extraction parameters. Moreover, it will be shown how one of the developed models can be easily applied to different accelerators: the MedAustron accelerator [83] is used as an example, proving how the frequency transfer properties change from a machine to another. Finally, Section 4.6 will summarize the obtained results, and point out some important differences between the developed simulation models and the real case.

4.1 Introduction

The experiments using the slowly extracted particles from synchrotrons are very sensitive to any flux variation of the outgoing particles: any perturbation in the temporal structure of the spill could possibly lead to problems in the particle detection process, as increased counting uncertainties or detector saturation. This is also true for the presence of any frequency component on the spill, from the power supply 50 Hz harmonics to the 200 MHz radio frequency (RF) bunched structure. At the SPS, to minimize the effects of the latter frequency, the beam is debunched during extraction and carefulness is applied in the quadrupole currents setting, in order to avoid spikes at the beginning of the flat top which have been shown to expose the remaining bunched structure on the extracted spill. Anyway, the main power supply currents are the principal responsible for a large part of the harmonics seen on the extracted beam structure. Unfortunately there is no easy solution to minimize their effect in the outgoing particle flux, in fact, this is a topic of decades long studies in many of the slow extraction based accelerators around the world [84–90]. In order to optimize the SPS spill quality, both the macro and micro structure are targeted separately [59]. Currently, an iterative feed-forward algorithm acting on all the main quadrupoles [60] is employed to optimize the macro structure of the extracted spill to that of a square pulse (i.e. Autospill, described in Section 2.5). This algorithm does not have a significant effect on the power supply ripples because it is mostly sensible to large scale variations of the spill signal. In the past, a real-time feedback system acting on a set of four correcting quadrupoles was being used for the task [73]; this system had to be decommissioned since the variation introduced on the extracted beam trajectory was leading to mis-steering on the targets [60] and consequent loss of useful intensity for the experiments. The ripple content of the spill is targeted by another iterative feed-forward system which uses a set of two servo-quadrupoles to minimize the main four 50 Hz harmonics measured on the spill by a secondary emission monitor sampled at 2 kHz. Both the iterative feed-forward methods described here apply the corrections estimated from the measured spill signal at the subsequent slow extraction machine cycle until the optimum solution is reached. This process can take several machine cycles to converge (usually about 10)

and the stability of the reached solution relies upon the reproducibility of the magnets response and power supply ripples in the SPS. It has been observed that a change in machine super-cycle modifies the magnets response to current [59], and also that the phase and amplitude stability of the 50 Hz harmonics can be subjected to significant changes in a days to weeks basis [91]: both these facts trigger sudden changes in the spill quality, and call for periodical re-iterations of the feed-forward correction methods. The regulator feedback loop of the main power converters also attempts to suppress the harmonics components by injecting opposite phase harmonics, but it was observed experimentally that this introduces low frequency noise, which was finally worse for the experiments (the loop gain is high enough only at low frequencies, e.g. $\lesssim 20$ Hz). The beam dynamics of the slow extraction process itself acts as a low pass filter to current ripples. This remarkable fact comes from the time-to-extraction distribution of unstable particles in phase space, of which a full theoretical derivation is carried out in [44]. Earlier studies at the SPS, dating back to 1983, already identified such an effect to be best represented by a second order low pass filter with a time delay [92, 93]. These results have been used for the improvement of the previously used spill feedback system. More recent studies at the SPS [58, 94] seemed to confirm the past observed low pass filter behavior, and have shown that no significant increase in extraction losses should be observed for realistic amplitudes of focusing quadrupole ripples. An analogous campaign at the J-PARC Main Ring used the current-to-spill transfer function in order to improve the ripple-correcting feedback system there employed [95]. It has also been proven that the feedback system could be replaced by a promising new approach, in which a real-time estimation of the tune variation made from the measured quadrupoles and dipoles currents is used to inject a counteracting current signal on some dedicated quadrupoles [96]. This approach could be considered as a real-time feed-forward system, particularly advantageous when the ripples are not stable from one machine cycle to another. Another possible way to improve the spill quality is to increase the smoothing effect of the time intervals distribution of the extracted particles, by acting on the machine parameters (from a frequency point of view, it corresponds to increasing the attenuation of the low pass filter of the extraction process). Recent studies at GSI SIS-18 synchrotron have shown how a reduced extraction sextupole strength and beam emittance can help suppressing the ripples on the extracted spill, due to the increased smearing of the extraction time intervals distribution [97, 98]. Intuitively, the importance of the extraction sextupole strength on the response to ripples is related to the fact that it defines the acceleration of unstable particles in phase space, and with it the spread of the extraction intervals distribution. This key fact has also been observed at the SPS when attempting to maximize the response to quadrupolar pulses for the burst-mode slow extraction [99], as described in detail in Section 3.2 (e.g. Fig. 3.29). Modulation techniques based on

the RF, as transverse RF noise injection, RF phase modulation, or simply bunched extraction, are also well established ways to attenuate the power supply ripples, but they will not be covered here. Inspired by the results of the latter techniques [100], a method of injecting a dominant quadrupolar ripple at a frequency higher than the main harmonics and above the cut-off frequency of the low pass filter effect by a definite factor, has been shown to improve the overall spill quality at GSI [101].
In summary, the knowledge of the slow extraction low pass filter effect covers an important role in order to exploit at best the methods of spill quality improvement.

4.2 Modeling and simulation of slow extraction frequency response

A ripple on the input current of the main magnets will be transferred to the extracted spill according to a certain transfer function, which includes both purely hardware and beam dynamics effects. Overall, the problem will be treated as follows:

$$i(t) = i_0(t) + \Delta i(t) \longrightarrow s(t) = s_0(t) + \Delta s(t) \tag{4.1}$$

Where $i(t)$ indicates the input signal, composed by an ideal component i_0 and a perturbation Δi; the same holds for the spill, $s(t)$. In the typical case i_0 is a nearly linear ramp and s_0 a constant, while Δi and Δs are a combination of discrete harmonics components and continuous noise. The present study will focus on the amplitude transfer function T_a of the noise signals, which is defined as:

$$T_a(\omega) = \left|\left| \frac{\widetilde{\Delta s}(\omega)}{\widetilde{\Delta i}(\omega)} \right|\right| \tag{4.2}$$

Where $\widetilde{\Delta s}$ and $\widetilde{\Delta i}$ are the Fourier transforms of Δs and Δi, respectively. Eq. 4.2 represents the full transfer function of the slow extraction process, which is exactly what the model will try to represent.

4.2.1 Transfer function block model of the SPS

The slow extraction process comprises a series of different transformations that can be organized in the fashion of a cascaded circuit model, as shown in Fig. 4.1.

As the SPS slow extraction is a quadrupole driven resonant extraction, and the resonance is on the horizontal plane, the main contributors to the spill quality are the focusing quadrupoles. In particular, the contributions from other active elements, as extraction sextupoles and main dipoles, have

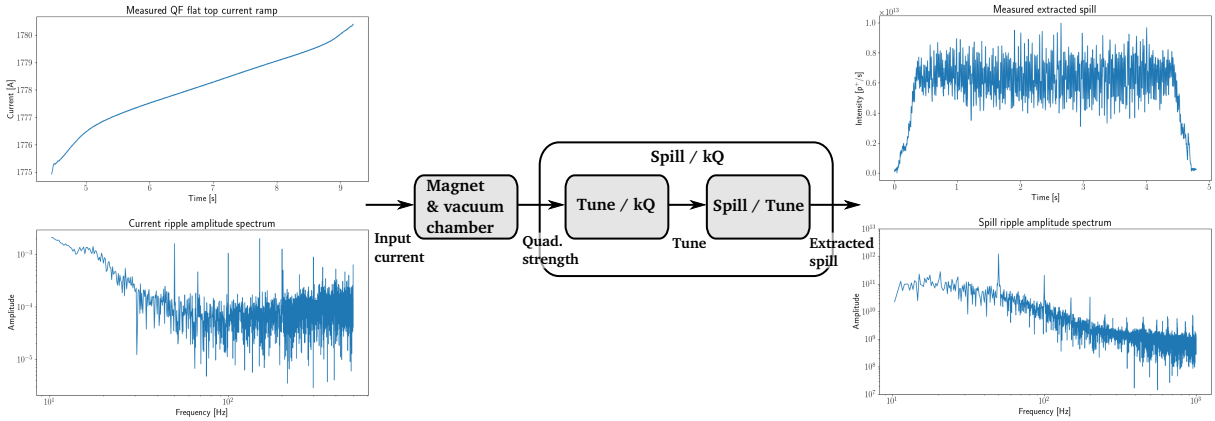

Figure 4.1: Schematic model of the slow extraction process in terms of its transfer function blocks. The plots on the left are a typical example of the focusing quadrupole current ramp (up) and the spectrum of its ripple component (down). The plots on the right are a typical operational extracted spill (up) and the spectrum of its ripple component (down), computed in the fashion of Eq. 4.1.

been shown in literature to be significantly lower [58, 94, 97]; the same has been verified also for the defocusing quadrupoles by using MADX. For these reasons the input signal for the present work will be assumed to be the current of the main focusing quadrupoles. An example of a measured current signal of the main quadrupoles can be observed in the top left plot of Fig. 4.1 (corresponding to $i(t)$, in the formalism of Eq. 4.1), while the spectrum of its ripple component (i.e. $||\widetilde{\Delta i}||$) is reported in the bottom left. In Fig. 4.1-right, a full extracted spill and its ripple spectrum (i.e. respectively $s(t)$ and $||\widetilde{\Delta s}||$) are shown.

The first gray block of Fig. 4.1 represents the effects of the magnet losses and shielding from the vacuum chamber: the input current is converted in the effective magnetic field which is seen by the beam. This particular block embodies all the hardware effects which cannot be simulated using a tracking code. The contribution from these effects are discussed in the final part of this chapter.

The second and third blocks represent the slow extraction process. They can either be seen as one transformation, or decomposed in two. In the following subsections, details on these two blocks are given in order to characterize the slow extraction frequency response.

It is important to point out that despite the process is being represented in the fashion of a cascaded circuit model, no initial assumption of linearity and cascability of each block has been made.

Magnet strength to spill simulation results

The second and third gray blocks of Fig. 4.1 can be simulated together (i.e. block "Spill/kQ") using a MADX implementation of the SPS slow extraction [52]. Many studies based on this model have shown impressive agreement with data [61, 64, 65]. In order to build a transfer function of the process, single frequency sinusoidal ripples are injected one at a time in addition to the nominal quadrupole function used to drive the resonance. The amplitude of the corresponding frequency component is then measured on the spill signal, and normalized by the input amplitude. Each ripple has the following expression:

$$r(t) = A_K \cdot \sin(2\pi f t + \phi) \tag{4.3}$$

where A_K is the amplitude (in units of the quadrupole strength), f is the frequency (in Hertz) and ϕ is a phase shift. Since the focus of the present study is on the amplitude of the transfer function, ϕ is assumed to be constant and equal to 0 without loss of generality. The harmonic analysis of the ripples on the spill are performed subtracting the DC component. In order not to introduce any bias in doing so, a fiducial time interval is selected around the center of the spill to exclude the two transients at the beginning and end of the extraction. The spectral components of the perturbation are then obtained either by using a Fast Fourier Transform (FFT) or the Numerical Analysis of Fundamental Frequencies (NAFF) algorithm [102, 103]. The NAFF algorithm has shown the most accurate results, hence it was chosen as the default method for the following analysis.

Figure 4.2 shows the transfer function obtained for different values of the input sinusoidal amplitude A_K. Several remarks can be drawn from the result:

1. Below a certain input amplitude ($\simeq 2.5 \cdot 10^{-8}$ m^{-2}), all the obtained transfer functions are comparable to each other, and no amplitude dependence is observed. Above such amplitude, a non-linear behavior emerges, making the transfer function dependent on the input amplitude.

2. A zero is present only for the small input amplitudes.

3. All the obtained transfer functions behave as a low pass filter, with a cut-off frequency around ~ 100 Hz.

The first two points can be readily derived by considering a uniform momentum distribution of the protons, as simulated in the MADX model. In this case, assuming an instantaneous extraction of particles, the spill can be approximated by the expression:

$$\frac{dN(t)}{dt} = \begin{cases} C\,\frac{dQ(t)}{dt} & \text{if } \frac{dQ}{dt} > 0 \\ 0 & \text{otherwise} \end{cases} \tag{4.4}$$

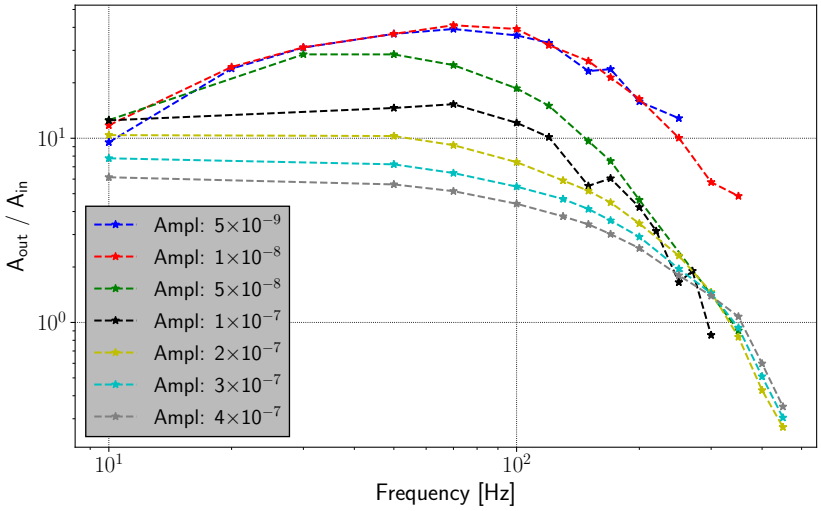

Figure 4.2: Simulated fixed amplitude transfer functions for the full extraction process. The injected amplitudes are reported in the legend and are expressed in units of m^{-2}, i.e. normalized quadrupole strength.

where dN/dt represents the number of extracted particles per second, Q is the horizontal tune of the machine, and C a constant. It is possible to explain observation 1 by noticing that Eq. 4.4 is a linear differential equation only if the tune is a monotonic function. In this case, the equation is that of a derivator circuit, which explains the presence of a zero for small amplitudes. The last observation, which is not foreseen by the ideal model of Eq. 4.4, is related to the interval distribution of the extracted particles. This behavior is in line with literature, previous studies, and machine observations, as previously detailed in Section 4.1.

Even with the simple model of Eq. 4.4, useful insights can be obtained by considering the case of a linear tune sweep superimposed to a sinusoidal ripple:

$$Q_{\text{eff}}(t) = \frac{\Delta Q}{\Delta t}\, t + Q_i + A_Q \cdot \sin(2\pi f t) \tag{4.5}$$

where Q_i is the initial point of the ideal tune ramp, ΔQ the ideal tune variation, Δt is the flat top duration and A_Q is the amplitude of the ripple in tune. Imposing monotonicity on $Q_{\text{eff}}(t)$ yields:

$$A_Q\, 2\pi f \leqslant \left|\frac{\Delta Q}{\Delta t}\right|. \tag{4.6}$$

Using the value of $\Delta Q/\Delta t$ for the SPS in Eq. 4.6 and converting from tune to quadrupole strength, it is found that $A_K \simeq 2.4 \cdot 10^{-8}$ m^{-2} delimits the

linear from non-linear behavior of Eq. 4.4 for $f \geqslant 50$ Hz: this is in good agreement with the results shown in Fig. 4.2.

The available slow extraction MADX model can be exploited to gain a better insight on the problem of linear and non-linear transfer function. Injecting two ripples at 50 and 70 Hz with equal amplitude, first below and then above the amplitude threshold of Eq. 4.6, the different behavior can be observed: the results are summarized in Fig. 4.3.

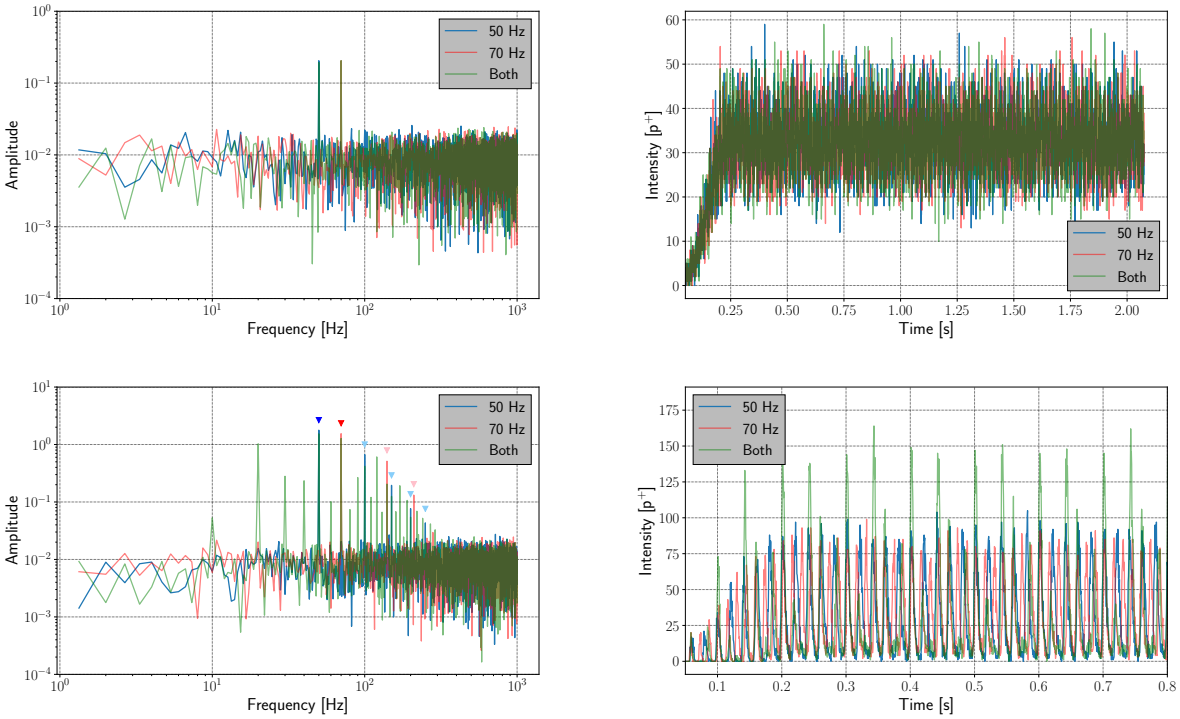

Figure 4.3: Ripple signal spectra (left) and corresponding extracted spills (right) obtained by injecting a 50 and 70 Hz ripple signals of equal amplitude, both separately and together. Top plots: ripple amplitudes below linearity threshold. Bottom plots: above linearity threshold.

The two top plots of Fig. 4.3 show the results for the case of ripple amplitudes below threshold. It can be noticed that the extracted spill is in good approximation a square pulse (plus a noise component), and that the superposition principle holds, being the spectral component the same both when injecting the ripples separately and together. The two bottom plots show instead the case of input amplitudes above threshold. It is evident that the spill is no more a square-pulse-like signal, but is split in several bursts. This is a sufficient condition for harmonics of the main injected frequency to appear in the spectrum of the spill. In particular, in the cases of single-frequency ripples (blue and red curves) the main spectral components and

their harmonics are highlighted by markers (full color for main frequencies and light colors for harmonics). Observing the spectrum of the case with the combined ripples (green line), it can be noticed that the superposition principle does not hold anymore.

By referring to the transfer functions reported in Fig. 4.2 it is possible to notice that both 50 and 70 Hz are located before the corner frequency of the low pass filter, meaning that it is not acting on these frequencies. In order to investigate whether the low pass filter effect modifies the behavior observed in Fig. 4.3, the same superposition exercise is repeated for the frequencies of 180 and 200 Hz, which are locate well above the cut-off frequency. Figure 4.4 summarizes the results.

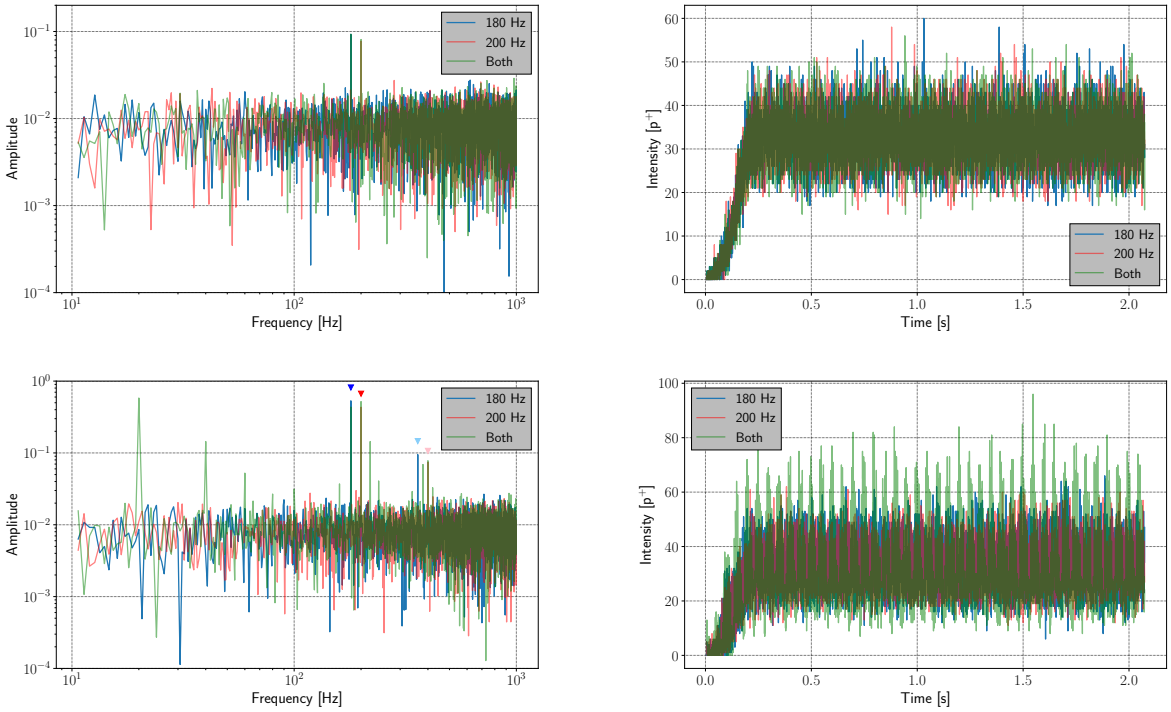

Figure 4.4: Ripple signal spectra (left side) and corresponding extracted spills (right side) obtained by injecting a 180 and 200 Hz ripple signals with equal amplitude, both separately and together. Top plots: ripple amplitudes below the linearity threshold. Bottom plots: above the linearity threshold.

It is possible to notice from Fig. 4.4 that the trend is similar to the previous case of injection of 50 and 70 Hz: the superposition principle holds below the amplitude linearity threshold, while it is broken otherwise. Anyway, a very important difference can be observed for the non-linear input amplitudes: the effect of the low pass filter prevents the spill (bottom-right

plot) from splitting into bursts, despite the injected input amplitude being the same of the previous case. Moreover, the bottom-left plot shows that only the first harmonic of the main frequencies is generated, as opposed to the case of Fig. 4.3. This comparison clearly shows how the low pass filter effect of the slow extraction process can be beneficial for the quality of the extracted spill: if properly exploited, it can be a very powerful tool to suppress the power supply ripples.

These results show that the transfer function is amplitude-independent (and so implying a linear system, with all the benefits coming from it) only for small enough ripples satisfying the monotonicity of the tune ramp. In the opposite case, the system becomes non-linear, and a unique transfer function for the process cannot exist anymore (in the proper sense of the term). For the non-linear case, an operative procedure in the style of Fig. 4.2 can be developed to construct an empirical amplitude-dependent transfer function, for example, by using only the attenuation of the main frequency component of single-injected ripples. The procedure could be expanded to the harmonics of the main frequency, for an increased precision, but keeping in mind the violation of the superposition principle observed in Fig. 4.3 and 4.4.

As a natural continuation of the previous results, it is interesting to observe, from a transfer function formalism point of view, the spill quality improving technique implemented at GSI SIS-18 [101]. As already described in Section 4.1, by injecting a dominant frequency ripple above the low pass filter cut-off frequency, it is possible to reduce the amplitudes of the lower frequency components (below the cut-off frequency, and hence more disruptive). The violation of the superposition principle, which has been shown in Fig. 4.3 and 4.4, is a necessary condition for this effect to take place. A simple example can be built for the case of the SPS by injecting a sinusoidal ripple of 350 Hz with amplitude well above the non-linearity threshold, together with other smaller frequency ripples in the linear regime. Intuitively, the high amplitude 350 Hz ripple will dominate the tune modulation, cutting away part of the smaller ones; since for the case of the SPS 350 Hz is above the low-pass filter cut-off frequency, the effect on the extracted spill of such ripple will be strongly attenuated.

Figure 4.5 shows the result of the test. All the amplitudes of the linear ripples are significantly reduced when they are injected together with the non-linear one. On the opposite side, the amplitude of the dominant 350 Hz signal does not change significantly when superimposed to all the other ripples. It can already be observed from this simple example that how much a ripple is reduced depends on its frequency. A transfer function for the attenuated linear ripples can be computed, and it is shown in Figure 4.6 (green curve).

The new transfer function of the superimposed ripples falls in between the two of the correspondent single injected signals (blue and red, respectively for linear amplitudes and non-linear 350 Hz wave). The strongest attenuation

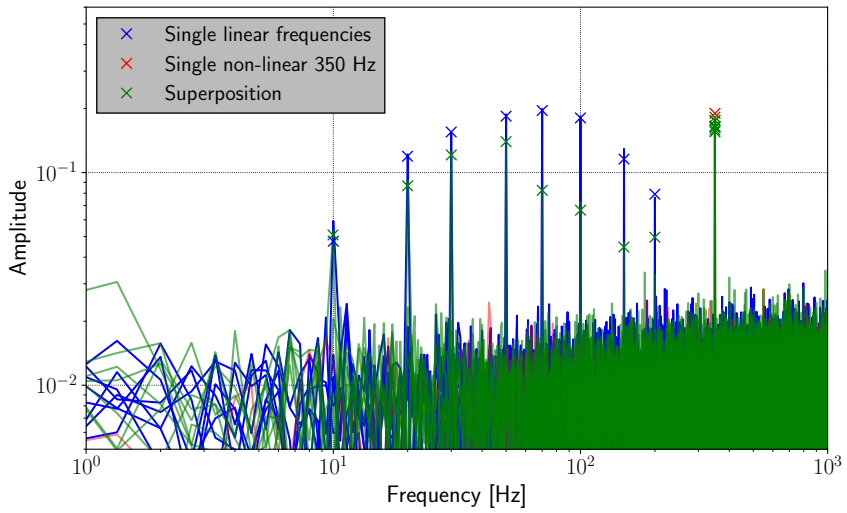

Figure 4.5: In blue: spectra of single injected ripples in the linear regime (constant amplitude of 0.005×10^{-6} m^{-2}). In red: spectrum of single injected 350 Hz ripple in the non-linear regime (amplitude of 0.2×10^{-6} m^{-2}). In green: superposition of the 350 Hz ripple with the linear ones, injected one at a time. The markers show the amplitude of each spectral component.

with respect to the linear transfer function (blue) seems to be reached between 70 and 200 Hz, where the ratio of the two reaches values of about 0.4. Going beyond this simple example, the picture gets more complicated and interesting, being the attenuation dependent on multiple parameters, as the position and amplitude of the dominant non-linear ripple, and the frequency distance of the attenuated modulations from the dominant ripple. On top of this, the amplitudes of the attenuated ripples can become a parameter too, especially if they start to cross the non-linearity threshold. This could be treated as an optimization problem in order to find the best combination of parameters for the maximum attenuation of the desired frequencies.

The advantage of using such a technique to improve the spill quality comes especially when the experiments using the extracted beam are not sensible to frequency components above a certain threshold, and so a new dominant frequency can be introduced without consequences (synchrotron modulation and bunched extraction can attenuate the ripples in the same way too). Unfortunately, this is not the case for the SPS, in which every strong frequency component disturbs the experiments detectors. For this reason, further studies are needed to evaluate the actual benefit of this technique for the SPS case, and it will not be covered in this thesis.

Before continuing with the modeling of the SPS frequency response, it is

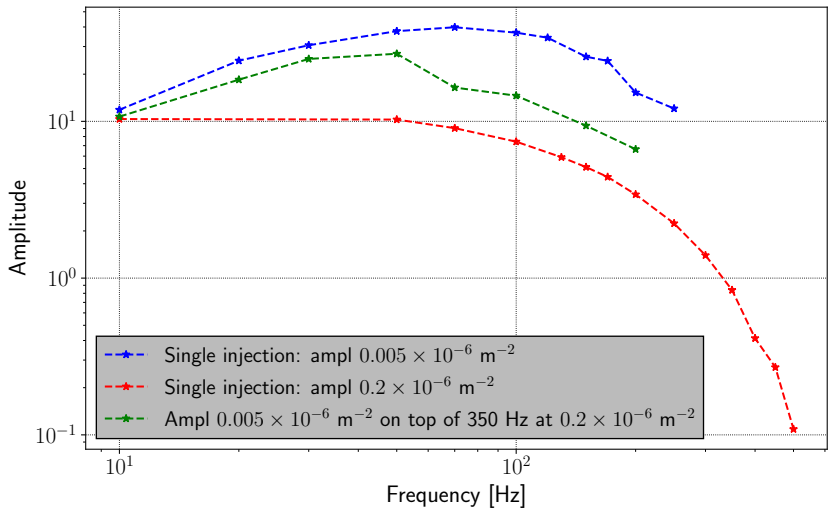

Figure 4.6: The blue and red curves show respectively the linear transfer function and the 0.2×10^{-6} m^{-2} fixed amplitude one, referring to Fig. 4.2. The green curve shows the transfer function for linear ripples injected together with a 350 Hz modulation at an amplitude of 0.2×10^{-6} m^{-2}.

important to make some considerations about the linearity violation condition (i.e. the opposite of Eq. 4.6), which can be derived from Eq. 4.4. It is understood that Eq. 4.4 does not contain the low-pass filter effect, since it has been derived assuming an instantaneous particle extraction (the low-pass filtering comes from the finite time-to-extraction of every particle in phase space). The question of whether only one of these effect (i.e. either Eq. 4.4 or the low-pass filter) would be the source of the non-linearity arises naturally. If for example the non-linear behavior would only come from the "rectifier" behavior of Eq. 4.4, it could be possible to restore linearity by changing the input signal from $Q_{\text{eff}}(t)$ to:

$$Q'_{\text{spill}} = \begin{cases} C \frac{dQ_{\text{eff}}(t)}{dt} & \text{if } \frac{dQ_{\text{eff}}}{dt} > 0 \\ 0 & \text{otherwise} \end{cases} \tag{4.7}$$

where Q'_{spill} corresponds to the ideal extracted spill generated by the input tune signal Q_{eff}, in the assumption of instantaneous extraction (i.e. Eq. 4.4). By reformulating the problem in such a way, it could be possible to bypass the "derivator-rectifier" component of the slow extraction and only consider the low pass filter part.

Figure 4.7 shows an example of the new input signal (top plot, blue line) and its Fourier spectrum (bottom plot), for a non-linear sinusoidal ripple of 50 Hz. The spectrum does not consist anymore in a single frequency component, but counts a number of harmonics of the main frequency as foreseen for

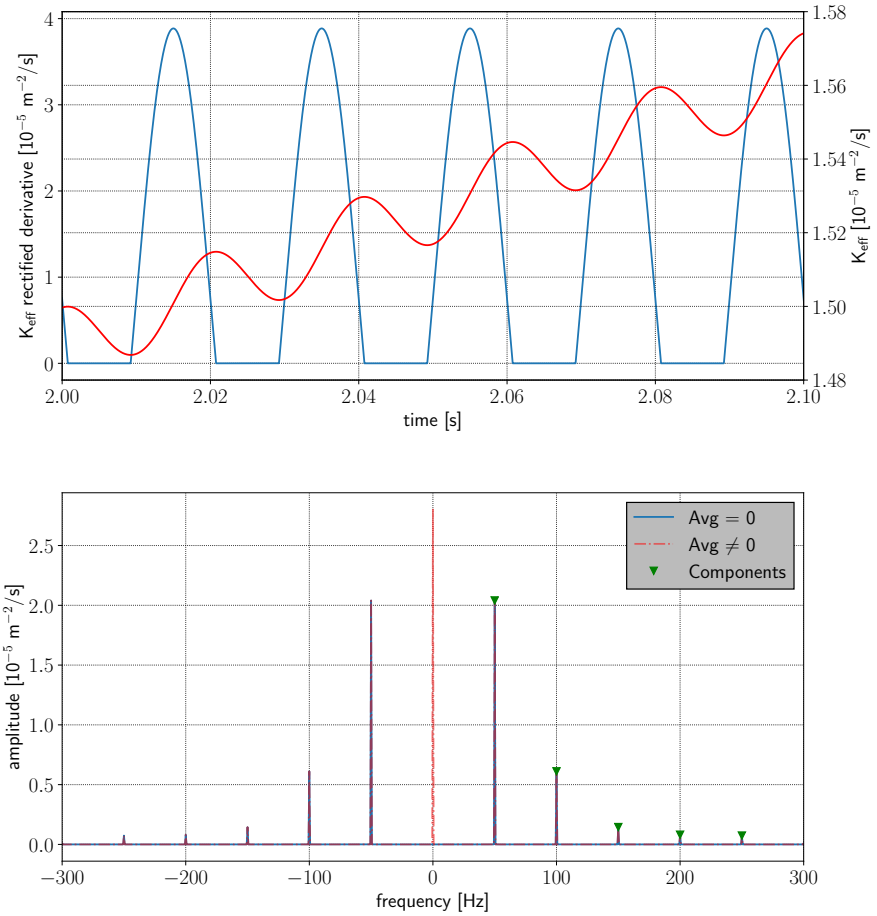

Figure 4.7: The top plot shows the input quadrupole strength ramp (red) and its rectified derivative (blue), as per Eq. 4.7. The bottom plot shows the spectrum of the recitified derivative: subtracting the average (blue) and not (dotted, red).

a periodic signal passing through a non-linear system. Regarding the computation of the transfer function using the rectified derivative input signals, only the first Fourier component will be used (i.e. the fundamental frequency, and the dominant one) for the sake of simplicity. In case of restored linearity, all the other transfer functions computed using the harmonics should give the same result.

Figure 4.8 shows the results of computing the slow extraction transfer function by using the rectified derivative of the quadrupole strength as input signal. Once again it is evident a non-linearity for the case of ripple amplitudes above the non-monotonicity threshold. In particular, the two smallest simulated amplitudes (blue and orange curves) fulfill the linearity condition,

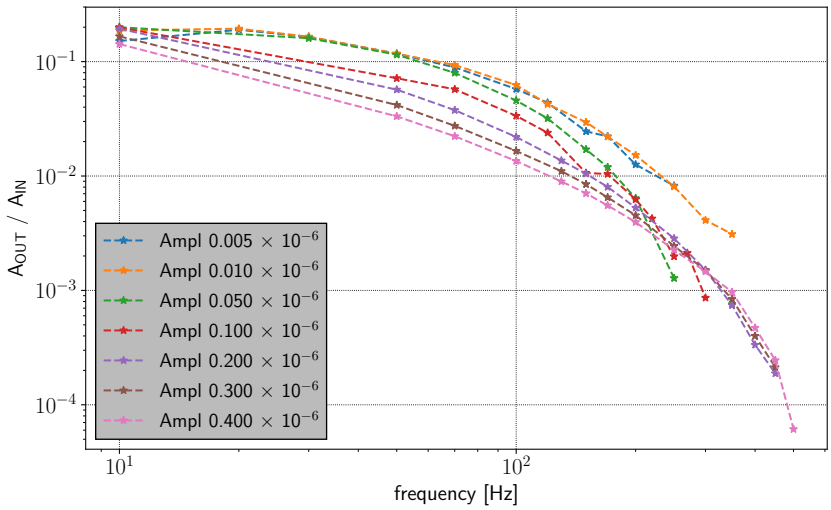

Figure 4.8: Fixed amplitude transfer functions from the rectified derivative of the focusing quadrupole strength to the extracted spill. The injected sinusoidal amplitudes are reported in the legend in units of m^{-2}.

and their transfer function are fully compatible in the exact same fashion of Fig. 4.2. Above the linearity threshold, the transfer function is amplitude dependent, as in Fig. 4.2. This results shows that the non-linearity is not only coming from Eq. 4.4, but it is also present in the low-pass filter effect of the slow extraction. For this reason, the attempted change of input signal does not seem enough to restore linearity, and since it carries with it some complications with respect to the differential approach stated in Eqs. 4.1 and 4.2 (as having to process the input ripples together with the tune ramp), it will not be used here.

One final question which the latter technique might be able to answer, and whose answer is not implied by the findings of Fig. 4.8, is whether the change of input of Eq. 4.7 could explain the attenuation effect observed in Fig. 4.6. The hypothesis is that computing the corresponding input frequencies by using Eq. 4.7 would make the green and blue transfer function of Fig 4.6 collapse to the same transfer function; which would lead to an easy way to correctly predict the attenuation effects.

Figure 4.9 shows that the difference between the amplitudes of the single injected and the attenuated (using 350 Hz) ripples is only a constant scaling factor of ~ 2 (top plot). Computing the transfer function using the latter amplitudes yields the result shown in the bottom plot. It is possible to notice that this technique does not correct for the different shapes of the two transfer functions: the difference of about a factor two is now moved to the lower frequency region with respect to Fig. 4.6. This result shows that the

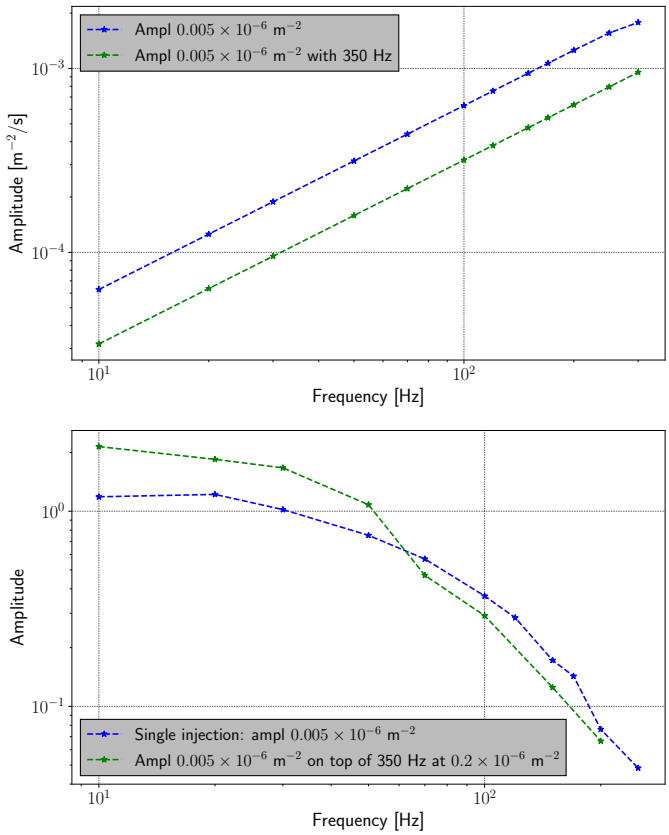

Figure 4.9: Top: input amplitudes as per Eq. 4.7 of fixed amplitude scan in the linear regime, with and without the superposition of the dominant 350 Hz signal. Bottom: transfer functions corresponding to Fig. 4.6 computed using the input amplitudes on the left.

change of input of Eq. 4.7 is not enough even for explaining the behavior of the ripple attenuation by injection of dominant high frequency components.

Magnet strength to tune simulation results

The second gray block of Fig. 4.1 represents the transfer function from focusing quadrupole to horizontal tune. It is possible to verify from simulation that a sinusoidal ripple on the main focusing quadrupoles leads to a same-frequency sinusoidal ripple on the tune ramp with good approximation. The only deviation coming from a residual small linear dependence of the tune on the focusing strength itself. Given this, a proportionality law between the focusing quadrupole ripple and the correspondent tune one will be assumed. The proportionality constant has been computed using the MADX model of the SPS, and also cross-checked from the calibration tables used for the

machine operation.

It is anyway useful to evaluate the tune spectrum distortion coming from the injected ripples. To do this, the available MADX model of the SPS is used in order to obtain the tune spectrum from the betatron oscillations of particles at the electrostatic septum. The tune spectra for different injected sinusoidal ripples of the type of Eq. 4.3 are then analyzed, with the goal of characterizing the behavior of the ripple-generated tune components.

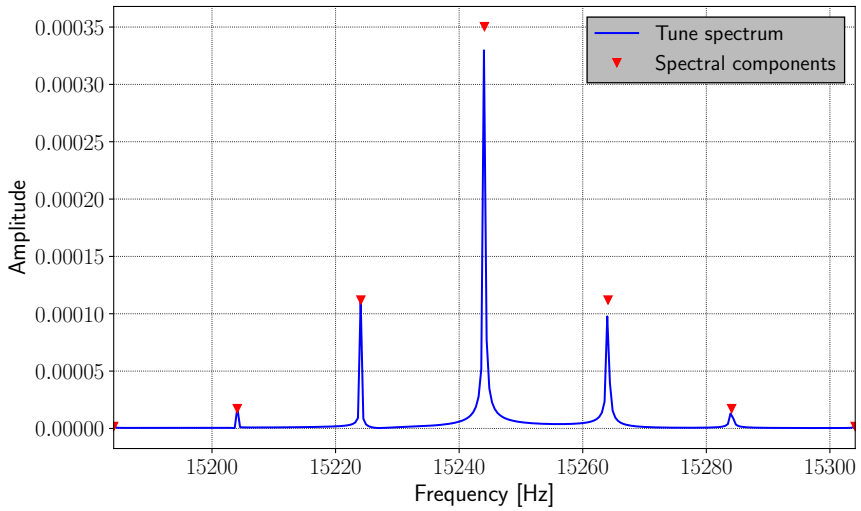

Figure 4.10: Simulated tune spectrum with the presence of an injected sinusoidal ripple of amplitude 10^{-7} m^{-2} and frequency of 20 Hz.

Figure 4.10 shows a simulated horizontal tune spectrum in the presence of a sinusoidal ripple on the focusing quadrupole strength. The main central peak (baseband) corresponds to the static tune value of the machine, i.e. the baseband tune, while the smaller peaks around it (sidebands) are a consequence of the injected ripple. An ideal sinusoidal ripple superimposed to the machine tune can be written in the form of:

$$Q(n) = Q_0 + A_Q \cos(2\pi f \Delta T \cdot n) \tag{4.8}$$

where $Q(n)$ is the tune as a function of the turn number n, Q_0 is static tune, A_Q the tune ripple amplitude, ΔT the one turn time and f the injected ripple frequency in Hertz. Being the horizontal tune the turn-by-turn oscillation frequency of the transverse horizontal coordinate of a particle with non-zero closed orbit, Eq. 4.8 means that the current ripple can be seen as a frequency modulation, as defined in [104]. The normalized horizontal

coordinate $\hat{x}$ will then behave as:

$$\begin{aligned}\hat{x}(n) &= \sin\Big(2\pi Q_0 n + \frac{A_Q}{Q_f}\sin(2\pi Q_f n)\Big) \\ &= J_0\left(\frac{A_Q}{Q_f}\right)\sin(2\pi Q_0 n) + \sum_{m=1}^{\infty} J_m\left(\frac{A_Q}{Q_f}\right)\times \\ &\times\Bigg[\ (-1)^{m+1}\sin\Big(2\pi n\left(Q_0 + mQ_f\right)\Big) + \\ &\quad + (-1)^m \sin\Big(2\pi n\left(Q_0 - mQ_f\right)\Big)\ \Bigg]\end{aligned} \tag{4.9}$$

where $Q_f = f\Delta T$ and J_m is the m-th order Bessel function of the first kind. Equation 4.9 shows the presence of the base-band and side-bands. In particular, the amplitude of the m-th pair of side-bands can be expressed as:

$$J_m\left(\frac{A_Q}{Q_f}\right) = \frac{1}{2^m\, m!}\left(\frac{A_Q}{Q_f}\right)^m + \mathcal{O}\left(\left(\frac{A_Q}{Q_f}\right)^{m+2}\right) \tag{4.10}$$

For small values of A_Q/Q_f the first pair of sidebands will be the dominant one, and its amplitude will be inversely proportional to the ripple frequency f. The same ripples used for the total strength-to-spill model have been injected in the present simulation, and the amplitude of the first sidebands pair has been used as the output signal.

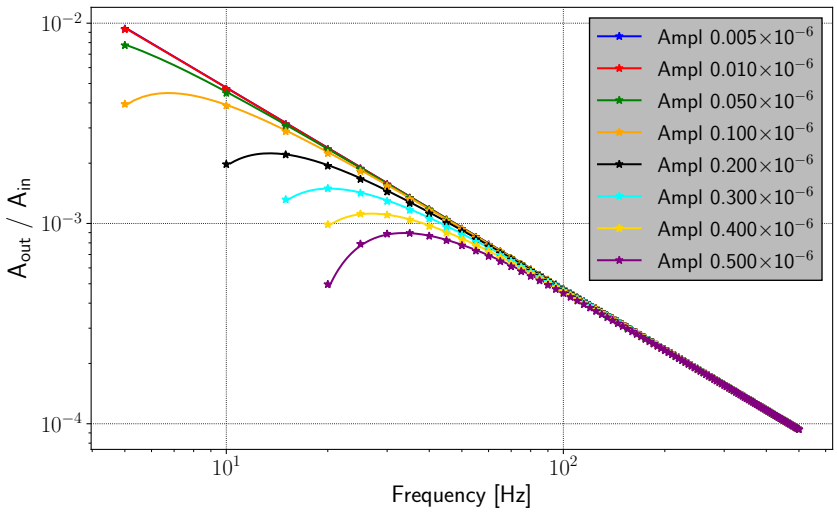

Figure 4.11: Amplitude of the first sideband pairs normalized by the input sinusoidal ripple amplitude. The continuous lines are not a simple interpolation, but the correspondent Bessel functions $J_1(A_Q/Q_f)$.

As foreseen, the points of Fig. 4.11 shows an asymptotic behavior $\propto 1/f$. Moreover, the continuous lines are the numerically computed Bessel functions $J_1(A_Q/Q_f)$ where A_Q is computed assuming $A_Q \propto A_K$: the agreement seems to confirm the initial assumption.

Tune to spill simulation results

In order to isolate the results of the tune-to-spill transfer function block (referred to Fig. 4.1) a 2D Henon map [105] model has been developed. The model can be expressed as follows:

$$\begin{pmatrix} x \\ x' \end{pmatrix}_{n+1} = \mathbf{R}\Big(2\pi Q_n\Big) \begin{pmatrix} x \\ x' + V_{\mathrm{ss}}\, x^2 \end{pmatrix}_n \tag{4.11}$$

where $(x, x')_n^T$ are the normalized horizontal phase space coordinates at the n-th turn, V_{ss} is the normalized virtual sextupole strength for the SPS, $\mathbf{R}(\theta)$ is a rotation matrix of angle θ, which in Eq. 4.11 coincides with the linear one-turn map, and finally Q_n represents the tune value at the n-th turn, which will have an expression similar to the one defined in Eq. 4.5. In order to inject amplitudes of tune ripples corresponding to the focusing quadrupole strength amplitudes injected in the previous MADX models, the conversion constant discussed above is used.

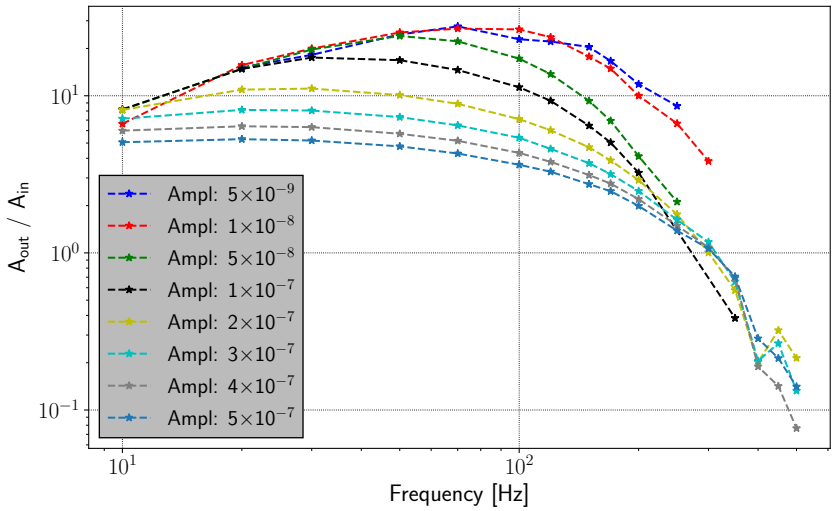

Figure 4.12: Transfer functions for different sinusoidal ripple amplitudes obtained from the tune-to-spill Henon map-based simulation.

Figure 4.12 shows the simulation results. These results are very similar to what has been obtained using the full SPS MADX model described in sub-

section 4.2.1 and reported in Fig. 4.2. In particular, the same considerations apply also for this case.

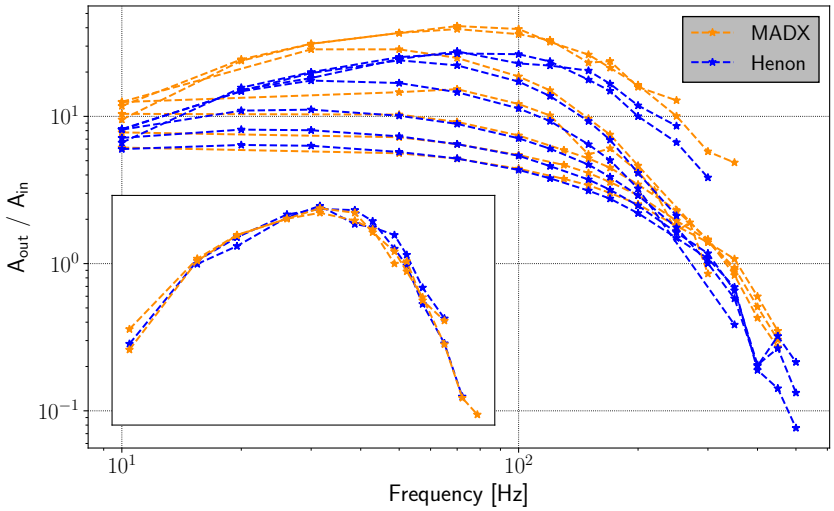

Figure 4.13: Fixed ripple amplitude transfer functions obtained for the Henon map simulation and for the full strength-to-spill MADX model of Fig. 4.2. The small plot shows the full compatibility between the Henon and MADX small-amplitudes transfer function after re-scaling the Henon result by a constant.

Figure 4.13 shows a direct comparison of the fixed amplitude transfer functions obtained from the MADX extraction model and the Henon-based model (respectively, the results of Fig. 4.12 and 4.2). It can be noticed that the main difference comes from a not perfect conversion of the ripple amplitude between the two models: the conversion constant appears to be amplitude dependent. The small-amplitudes linear transfer functions of the two models can be re-scaled to be almost equal, as shown in the small plot inside Fig. 4.13, only adjusting the conversion factor prioritizing the reconstruction of the small-amplitude part.

4.2.2 Effective transfer map

Using the results obtained up to now, an empirical semi-analytical model can be built in order to predict the frequency response of the slow extraction from a given current signal. The transfer functions extracted from the MADX

model have been fitted using different low pass filter functions of the type:

$$F(s) = \frac{K \cdot f(s)}{\Pi_i(1 + a_i s + b_i s^2)}$$
$$f(s) = \begin{cases} s & \text{linear ripples} \\ 1 & \text{non-linear ripples} \end{cases} \tag{4.12}$$

where $s \in \mathbb{C}$, a_i, b_i and K are real coefficients, and $f(s)$ introduces a zero for the linear case. Every obtained transfer function can be fitted with a function of the type of Eq. 4.12, as shown in Fig. 4.14.

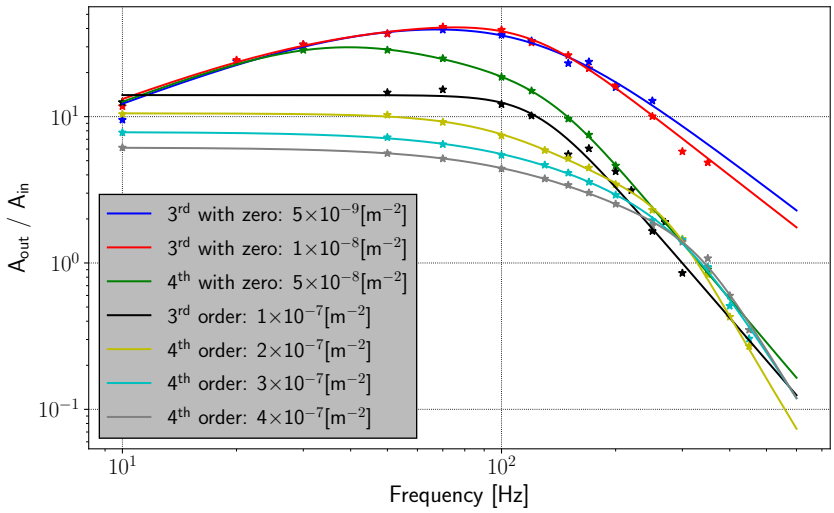

Figure 4.14: Fixed amplitude transfer functions fitted with curves of the type of Eq. 4.12. The legend refers to the order of the low-pass filter which gives the best fit.

The order of the low pass filter grows with the input amplitude, starting from a 2-nd order asymptotic behavior for the linear case, to a 4-th order of the maximum simulated amplitude. For the non-linear case, the transfer functions are only relative to the main frequency (i.e. there are also its harmonics, as shown in Fig. 4.3 and 4.4) and the spill can potentially be split in several pulses: the linear and non-linear models shall not be compared directly since they refer to two different conditions.

For the case of the linear small-amplitude transfer function, a good fit is given by the function:

$$F_{3z}(\omega) = \frac{K \cdot \omega}{\sqrt{\left(1 + \left(\frac{\omega}{\omega_c}\right)^2\right)\left(\left(1 - \left(\frac{\omega}{\omega_c}\right)^2\right)^2 + 4\zeta^2\left(\frac{\omega}{\omega_c}\right)^2\right)}} \tag{4.13}$$

with the effective pole at $\omega_c/2\pi \simeq 90$ Hz for the case of the SPS (eventually, two poles can be defined for more flexibility).
In the non-linear regime, the transfer function will depend also on the input amplitude, hence a 2-dimensional map is needed to represent a complete transformation from current to spill. The variation with amplitude for fixed-frequency transfer function points $\mathrm{tf}(a|f)$ approximates a power law function:

$$\mathrm{tf}(a|f) = c \cdot a^{\kappa} \tag{4.14}$$

where c and κ are real constants different for every f, and a is the the injected ripple amplitude.

Figure 4.15 shows the simulated fixed-frequency transfer functions, fitted with a power law of the type of Eq. 4.14. In particular, the behavior does not follow the fitted function as well as for the case of the fixed-amplitude transfer functions of Fig. 4.14: this has to be taken into account for the construction of an effective analytical 2-D map. Hence, the full-non-linear transfer function is obtained using a double logarithmic linear interpolation of the fitted low pass filter functions of Fig. 4.14 along the amplitude dimension. This corresponds to a power law interpolation of the type of Eq. 4.14:

$$\begin{aligned} \mathrm{tf}(f|a_i) &= c_{ij}\, a_i^{k_{ij}} \\ \mathrm{tf}(f|a_j) &= c_{ij}\, a_j^{k_{ij}} \end{aligned} \tag{4.15}$$

where $\mathrm{tf}(f|a)$ if the fixed-amplitude transfer function of input amplitude a as a function of the frequency f, a_i and a_j are a pair of adjacent input ripple amplitudes, and c_{ij} and k_{ij} are real valued functions of f (i.e. $c_{ij}(f), k_{ij}(f) : \mathbb{R} \to \mathbb{R}$) which are obtained by solving the system of Eq. 4.15.

Figure 4.16 shows the obtained normalized quadrupole strength to spill non-linear transfer map, where the abscissas and ordinates are respectively the injected ripple frequency and amplitude (in units of normalized quadrupole strength), while the color code shows the transfer function value. The same procedure could be repeated for the other harmonics of the main injected frequency on the extracted spill, obtaining a 2-D transfer map for every harmonic.

The non-linear transfer map model holds for single-injected non-linear ripples, while it should be revised in case of superposition of ripples, given the violation of the superposition principle shown in Figs. 4.3 and 4.4. In particular, while from the latter plots the main injected amplitudes seem to remain the same, this is not guaranteed in general (as shown by the superposition of linear and non-linear ripples in Fig. 4.5). On top of this, also more frequency components, as the difference of the two injected frequencies with its harmonics, will appear.

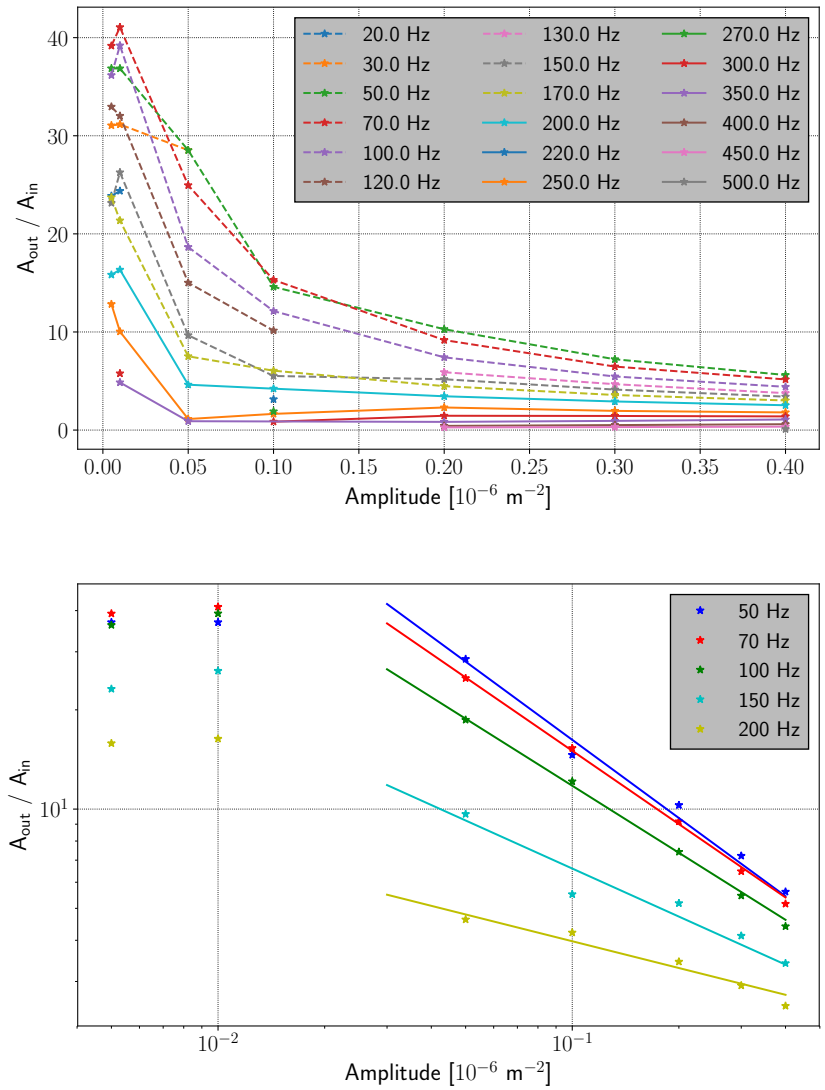

Figure 4.15: Fixed-frequency transfer functions as a function of the input ripple amplitude in linear scale (top), and double logarithmic scale fitted with a power law on the final points (bottom).

4.3 Dedicated ripple injection measurements

A dedicated data taking campaign was carried out at the SPS in order to validate the above-developed model. The measurements were taken by connecting a voltage waveform generator to the power converters of the main focusing quadrupoles, so to inject a sinusoidal perturbation at arbitrary frequency and amplitude. The circulating intensity in the machine was set

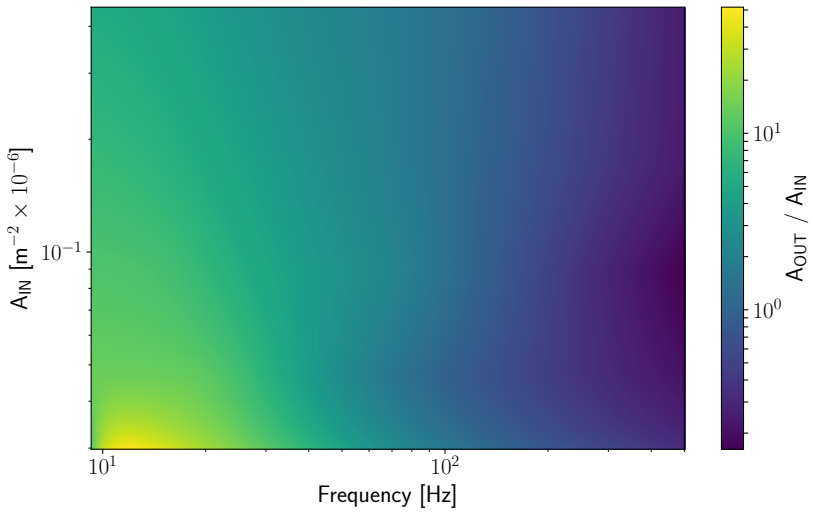

Figure 4.16: 2-dimensional transfer map obtained from the fitted transfer functions of Fig. 4.14, and a bilogarithmic linear interpolation.

to the value of $\sim 4 \times 10^{11}$ protons, that is about two orders of magnitude smaller than nominal, for machine protection reasons related to other parallel measurements. The recorded quantities for each extraction cycle are:

- amplitude and frequency of the injected voltage ripples from the waveform generator.
- measured input current of the focusing quadrupoles and its correspondent reference current, both sampled at 1 kHz.
- time profile of the extracted proton spill, measured with a secondary emission monitor placed just after the extraction septa. The beam intensity signal is sampled at 2 kHz. A typical measured spill profile is reported in Fig. 4.17.

From Fig. 4.17 it is evident that the spill signal is rather noisy and partially saturated at the lower level. This is due to the very low circulating intensity, which unfortunately could not be increased.

The noise on the data, Δi (Eq. 4.1), is calculated by subtracting the reference current from the measured one, and performing a spectral analysis. In order to compare the experimental data with simulations, it is necessary to convert the experimentally injected ripple amplitudes from units of current to quadrupole strength: this has been done using available calibration tables (see for example [56]).

Figure 4.18 shows an overview of the measurements. In particular, all the injected ripples fall into the non-linear regime, and therefore this data can

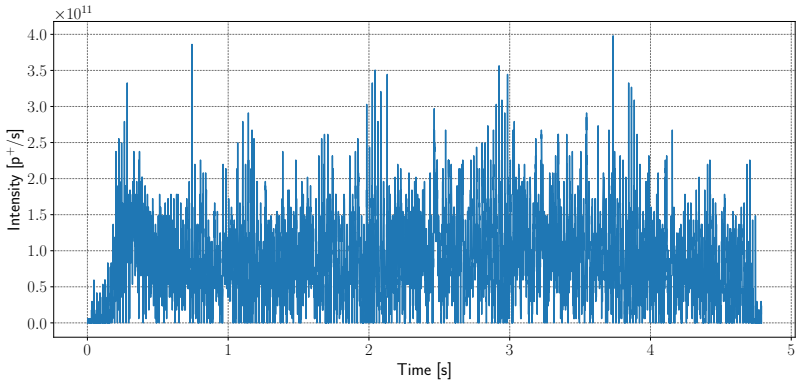

Figure 4.17: Typical spill (without injected ripples) measured by the secondary emission intensity monitor during the ripple injection MD.

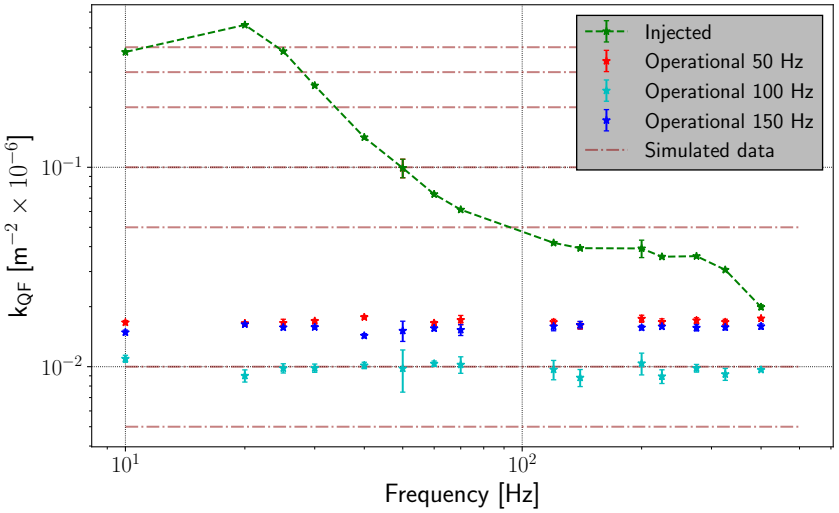

Figure 4.18: The green points (connected by dashed lines) show the strength-equivalent ripple amplitude and frequency of the sinusoidal ripple injected using the waveform generator. The red, light blue and blue points show the 50, 100 and 150 Hz harmonics components on the measured current coming from the power supply ripples. The horizontal lines are the ripple amplitudes simulated in Subsection 4.2.1. All the points that generated the green curve have been injected as voltage ripples of fixed 240 mV amplitude.

be used to validate both the full MADX simulation model and the empirical 2D transfer map developed above. All the injected data shown in Fig. 4.18 has been used as input for the simulations, both using the MADX model and the empirical 2D transfer function of Fig. 4.16.

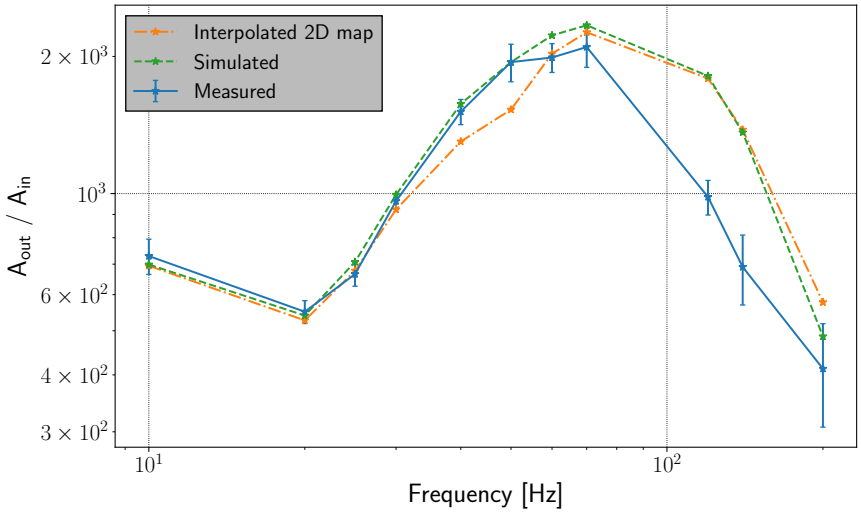

Figure 4.19: Comparison between the transfer function points from the experimental data (blue continuous line), the full MADX simulation model (dashed green line), and the 2D empirical transfer map (dot-dashed orange line).

Figure 4.19 shows the comparison between the transfer function points obtained from experimental data, MADX simulations, and 2D empirical transfer map (all from single frequency analysis). It is possible to notice that for all but two frequency values (120 and 140 Hz) the experimental data points are in a very good agreement with both the models. Moreover, the MADX simulation results are almost perfectly agreeing with the empirical map in all but two points, where anyway the fractional difference is smaller than 10%. These results show the robustness of the developed empirical map, especially considering that the map has been built using only 5 fixed-amplitude transfer functions (the ones of Fig. 4.2). By simulating more fixed-amplitudes transfer functions, the accuracy of the empirical transfer map can be improved up to a desired level. Moreover, the results of Fig. 4.19 show that no hardware effects (i.e. vacuum chamber shielding and magnet losses, referred to Fig. 4.1) are evident in the frequency region of interest, as expected from hardware estimations (more details in Appendix B).

The reason why less injected frequency points are shown in Fig. 4.19 with respect to Fig. 4.18 is that for injected frequencies above 200 Hz different methods used to process the spill spectra lead to systematically different results; this fact is most probably due to the reaching of the noise floor, which is particularly high because of the poor quality of the spill signals. Finally, the reason of the low compatibility between data and model for the

points at 120 and 140 Hz is not clear: given the good agreement for the rest of the points, it could be related to some unnoticed problem during the data taking.

4.4 Operational data

As for the dedicated measurement campaign described in Section 4.3, also for operational data the measured current, reference current and spill intensity profile are recorded for every extraction cycle. The extracted intensity is about 3×10^{13} protons per spill. For this reasons, operational data makes for another useful opportunity to validate the developed simulation models.

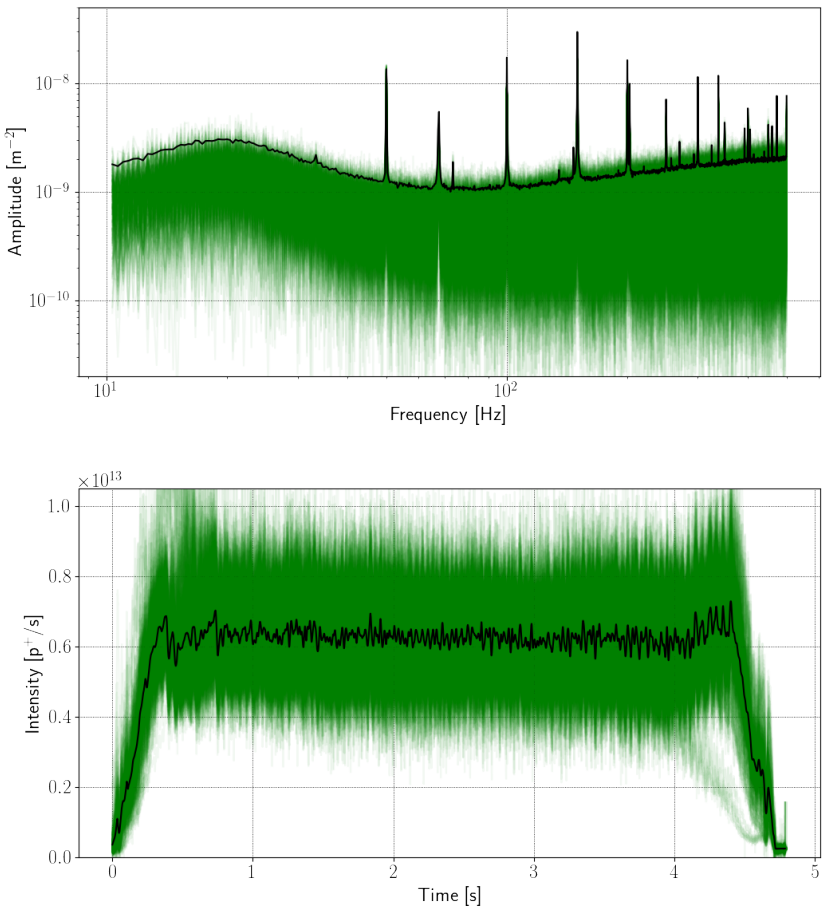

Figure 4.20: Average (black lines) and single measurements (green lines) for the ripple quadrupole current (top plot) and extracted spill (bottom plot) during a physics run.

Figure 4.20 shows the average spectrum of the focusing quadrupole ripple current (converted in units of quadrupole strength) and the corresponding measured spills during a stable physics run of two days. As introduced in Section 4.1, the 50 Hz and its harmonics are the frequency components impacting the spill the most during operation. Their typical amplitudes are below 2×10^{-8} m^{-2} (the averaged values of the 50, 100 and 150 Hz peaks are also shown in Fig. 4.18). These amplitudes are significantly smaller than those injected during the dedicated measurements, and, according to the simple case of Eq. 4.6, they should be below the linearity threshold of the transfer function.

Unfortunately, in this regime, the measured current reported in Fig. 4.20 cannot be directly used to compute the transfer function. To explain why, it is first useful to describe the current-logging system. Two Direct-Current Current Transformers (DCCT) are measuring the focusing quadrupole currents at every cycle. Two Analog to Digital Converters (ADC) are used to sample the DCCT's signals, and the final logged current signal is the average of the two, converted into Ampere. For a successful transfer function computation, it is crucial to know whether the measured current signal corresponds to the actual current fed to the magnets. To this purpose, several measurements of the DCCT sampled signals have been taken with no current on the magnets, in order to estimate the measurement noise of the DCCT-ADC chain.

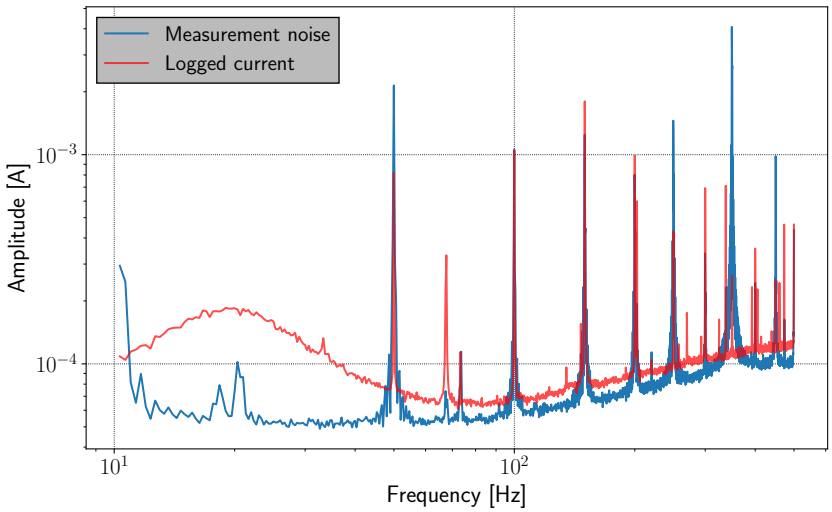

Figure 4.21: Average spectrum of the DCCT measurement noise (blue) compared with the average spectrum of the logged ripple currents during operation (red).

Figure 4.21 shows the average measurement noise spectrum from the

DCCT's compared with the average spectrum of the logged current of operational data. Above 50 Hz, the continuous parts of the two spectra (as opposed to the harmonics peaks) take exactly the same shape, with a small scaling factor probably due to different measurement conditions with the magnets ON and OFF. The amplitudes of the discrete spectral components (50 Hz harmonics and other peaks) measured with and without the main current are not directly related by a fixed scaling factor, but their values are mostly within the same order of magnitude. The lack of information on the relative phase does not allow a precise estimation of the real amplitudes present on the magnets, using the available data. For this reason, only the continuous transfer function will be computed from the operational data, as opposed to the single frequencies discrete transfer function computed up to now. Under linearity conditions, the two methodologies should lead to comparable results.

To correctly compute the continuous transfer function, the measurement noise needs to be removed from the logged current, in order to have an input spectrum as realistic as possible.

One possible way is to process directly the two amplitude spectra of Fig. 4.21, and apply a subtraction of squares on the two. In particular, assuming valid the superposition principle for the measurement noise and magnet current signals, the measured current amplitude spectra $|M(\omega)|$ will satisfy the equation:

$$|M|^2 = |S|^2 + |N|^2 + 2\cos(\phi_N - \phi_S)|S||N| \tag{4.16}$$

where $S(\omega)$ and $N(\omega)$ are the Fourier transforms of the magnet current and measurement noise signals, while $\phi_S(\omega)$ and $\phi_N(\omega)$ their respective phases. For an average of several measurements the mixed term could be neglected, by assuming its average to be small due to random phase differences (this approximation does not hold for the harmonics, where the phase difference is most probably fixed). The current signal amplitude spectrum could then be approximated as:

$$|S| \simeq \sqrt{|M|^2 - |N|^2} \tag{4.17}$$

paying attention to correctly handle negative values, which will especially appear in correspondence of the harmonic peaks. The solution of Eq. 4.17 contains the two approximations of valid superposition principle between measurement noise and current signal, and vanishing mixed term in Eq. 4.16, which may not hold in the real case.

In order to cross check the result from Eq. 4.17, another method could be employed to compute the actual current amplitude spectrum fed to the magnets.

Figure 4.22 shows the theoretical frequency response between measurement noise and actual magnet current; it is based on a model of the magnet

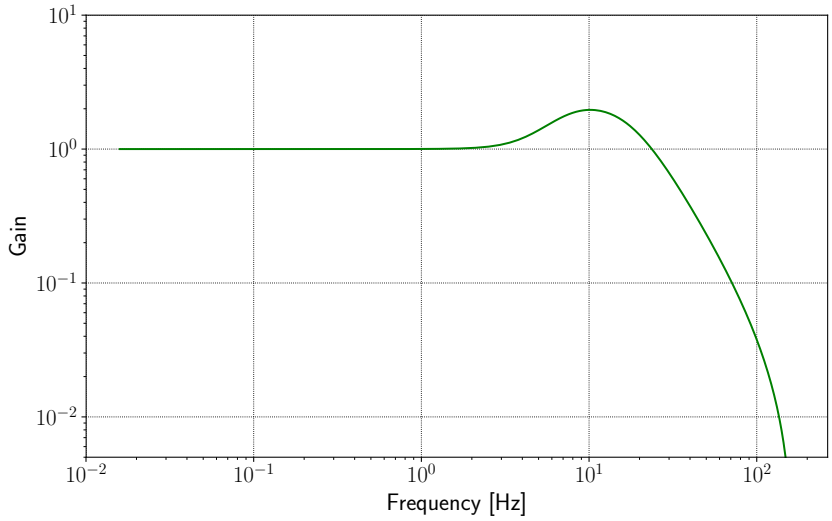

Figure 4.22: Theoretical frequency response function between measurement noise and magnet current, courtesy of Michele Martino.

at small-to-medium size voltages, which was reproduced quite closely by measurements at the end of 2017 [1]. The frequency range of the model is limited to about 167 Hz, i.e. the Nyquist frequency of the control system. By applying this frequency response function to the measurement noise spectrum of Fig 4.21, and keeping into account the presence of a white noise floor for large enough frequencies, the expected current in the magnets could be obtained.

Finally, in order to compare the operational data with the model, a simulated version of the operational continuous transfer function is obtained by using the measured current as an input for the MADX simulation. It has been previously observed that in case of ripples with amplitudes above the linearity threshold, the obtained transfer function is strongly dependent on the input signal. For the operational input current (e.g. Fig. 4.21) the only components of the spectrum that could potentially break the linearity are the main harmonics. As observed from Fig. 4.21, the harmonics present on the magnets should have the same order of magnitude of the ones seen in the measured current: this implies that the linearity condition should hold. Under these conditions, the simulated transfer function should not be strongly dependent on the input, meaning that, even if the simulated current is dominated by measurement noise, the correct result should be obtained.

The latter hypothesis is tested by simulating the measured current with

[1] The model and measurements have been carried out by Miguel Cerqueira Bastos, Michele Martino, *et al.* at CERN.

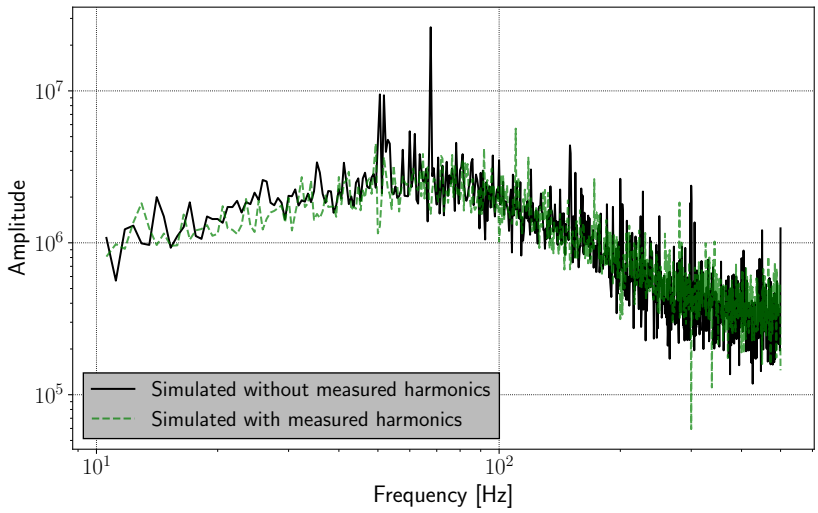

Figure 4.23: Comparison between simulated operational transfer function obtained by including the harmonic peaks in the simulation (dashed green line) or not (black line).

and without the main harmonics, and comparing the two resulting transfer functions. Figure 4.23 shows that no significant difference can be observed, confirming the hypothesis of linearity.

Figure 4.24 shows the comparison between experimental and simulated continuous transfer function without any correction applied to the input measured current (i.e. using the spectra of Fig. 4.20). The two results are not in good agreement: the experimental transfer function lacks the initial zero, its pole is at a lower frequency, and its low pass filter order is higher than the one of the simulated case.

Figure 4.25 shows the comparison between experimental and simulated continuous transfer function when the correction from Eq. 4.17 is applied. In this case, the poles of the two transfer functions are very close, and also the difference in the low pass filter slope is reduced with respect to Fig. 4.24. The initial zero, visible in the simulated data, is still not visible in the experimental transfer function.

Finally, Fig. 4.26 shows the comparison between experimental and simulated continuous transfer functions when the input magnet current is computed from the measured noise spectrum of Fig. 4.21 using the frequency response function of Fig. 4.22. As previously mentioned, a white noise floor has been used as an additional parameter to obtain this result. It can be observed that, in this case, the simulated and measured transfer functions are in good agreement: also the initial zero seems to be reproduced. The harmonic peaks visible in the measured transfer function are an artifact of

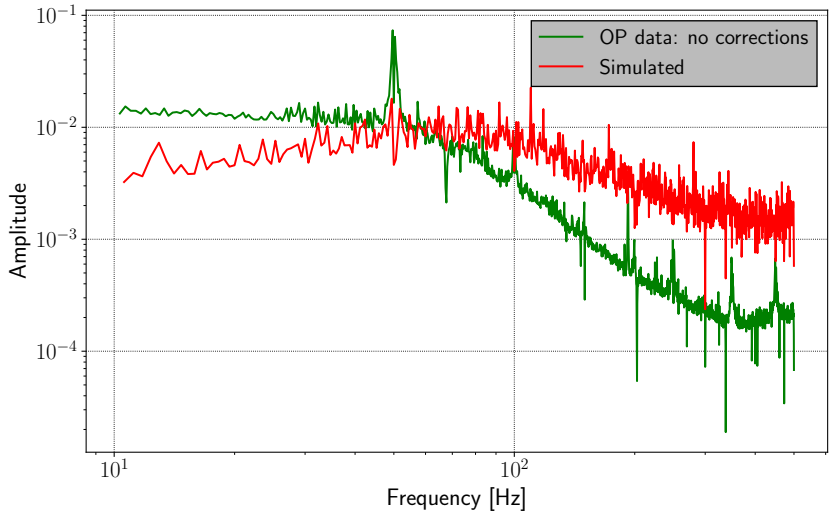

Figure 4.24: Measured operational spill transfer function (green) compared with the simulated one (red). No correction has been applied to the measured input current in this case.

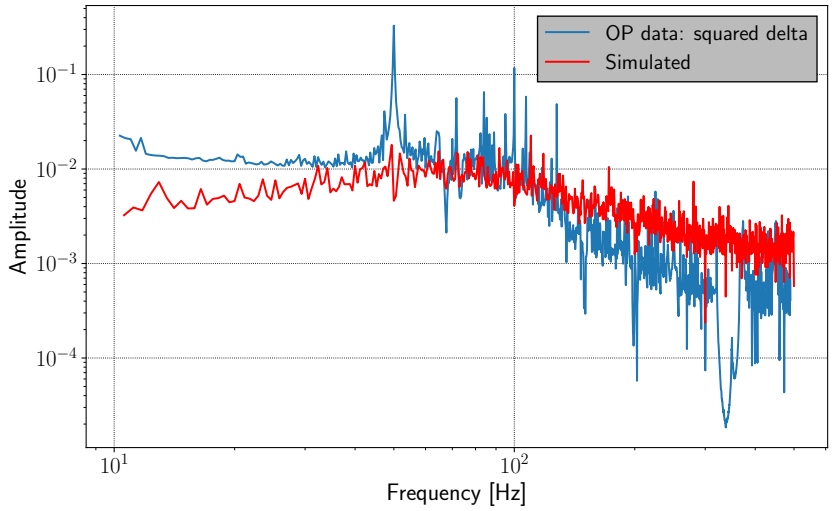

Figure 4.25: Measured operational spill transfer function (blue) compared with the simulated one (red). The difference of squares correction of Eq. 4.17 has been applied to the measured input current.

the analysis process: only the continuous spectrum has been used to obtain the result.

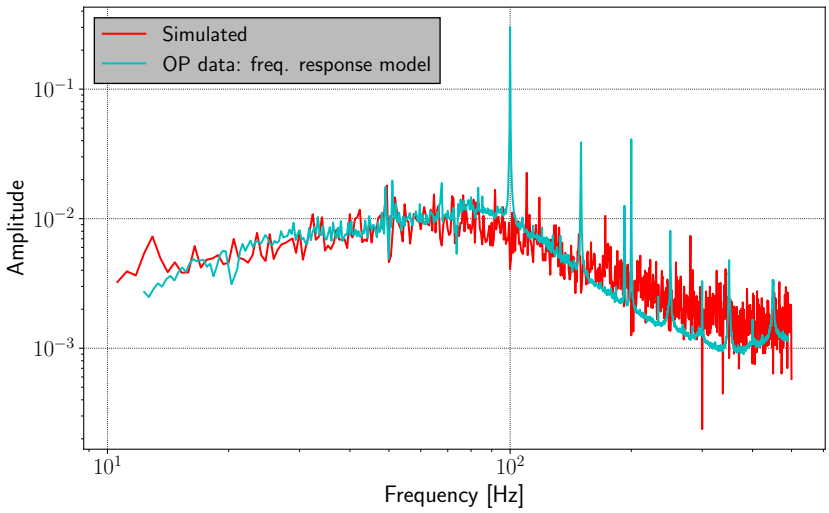

Figure 4.26: Measured operational spill transfer function (light blue) compared with the simulated one (red). The expected frequency response model from measurement noise to magnet current has been applied to obtain the measured input current.

Other systematic effects may be responsible for the non-perfect agreement between measurements and simulations, such as the servo-spill system used to suppress the main 50 Hz components (this system is always active for operational data, while it has been turned off for the dedicated measurements of Section 4.3). Also, it cannot be excluded the presence of strong ripples with frequencies higher than 500 Hz, being 500 Hz the Niquist frequency of the current measurement system. These open points will be investigated after the Long Shutdown 2, in order to provide a final validation for the results reached in this chapter.

As a final remark, one could wonder why all the measured-current problems listed and faced for the operational data have not been addressed also for the dedicated measurement campaign described in Section 4.3. The main reason is beacause, in the latter case, most of the injected ripples are significantly higher (Fig. 4.18) than the ones already present on the measured current (e.g. Fig. 4.20 up), and they start getting closer only from 200 Hz onward.

This is highlighted by looking at the histogram of the measured current for each injected ripple, as shown in Fig. 4.27. From 70 Hz onward the distribution starts to morph into a Gaussian, but below that a dominant sinusoidal behavior is clearly distinguishable. Nevertheless, most of the ripples injected during the MD are not multiples of 50 Hz, hence still allowing for a clean

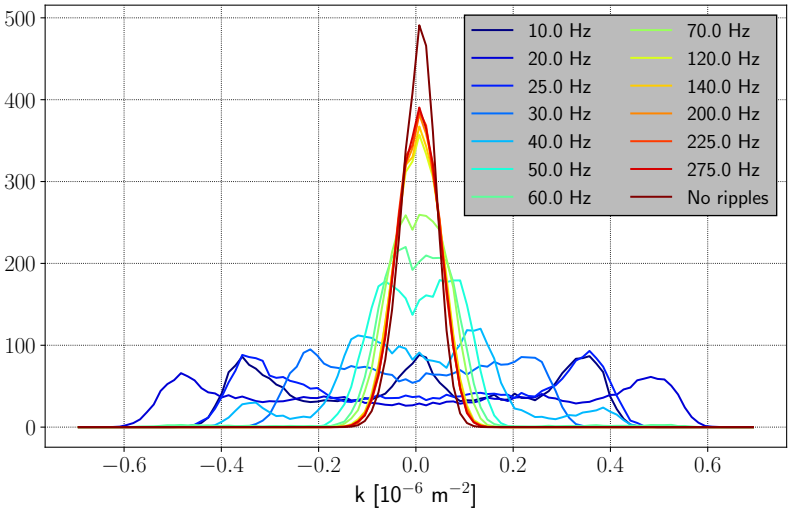

Figure 4.27: Histogram of the measured ripple current waveforms during the injection MD, and comparison with a measured current with no injected ripple.

identification of their amplitude. For the higher frequencies, the problems related to the very low spill signals start to dominate the transfer function computation, and the input current noise becomes a subleading effect.

4.5 Model predictions and different scenarios

The simulation model developed and validated in the previous sections has been built using the nominal optics and extraction parameters of the SPS. In this section the developed simulation model will be used to explore different machine parameters, looking for the possibility to improve the spill micro structure of the slow extraction.

The parameters that can be varied and affect the spill quality are: the machine chromaticity ξ and the extraction sextupole strength (using the virtual normalized sextupole strength formalism, V_{ss}). As shown for the burst mode slow extraction, increasing the virtual sextupole strength increases the particles acceleration in phase space and reduces the spread of the interval distribution. Decreasing it will have the opposite effect, hence increasing the attenuation of high frequencies, as also shown in [97, 98]. Regarding the chromaticity, its effect is not unique. Let first consider a sinusoidal ripple $r(t)$ superimposed to an ideal tune ramp, with a uniform momentum distribution of the particles. Under the assumptions of a monotonic tune and

instantaneous extraction, the spill can be expressed as:

$$\frac{dN}{dt} = K\,\frac{dQ}{dt} = K\,\frac{\Delta Q}{\Delta T} + K\,\frac{dr}{dt} \tag{4.18}$$

where K is the particle density in the tune space, which in case of a uniform particle distribution becomes:

$$K(q) = \begin{cases} \frac{N_{\text{tot}}}{|\xi|\Delta\delta_p} & \text{if } q \in [Q_{\min}, Q_{\max}] \\ 0 & \text{otherwise} \end{cases} \tag{4.19}$$

with N_{tot} being the total number of particles, $\Delta\delta_p$ the length of the uniform momentum distribution, and $Q_{\text{max/min}} = Q_0 + |\xi|\delta_{p,\text{max/min}}$. Using Eq. 4.19 it is then clear that the last addendum at the last member of Eq. 4.18 is inversely proportional to the chromaticity. In a typical situation in which the slow extraction has a fixed length and a fixed number of particles, the total tune variation ΔQ will be a function of the chromaticity as:

$$\Delta Q = |\xi|\Delta\delta_p \tag{4.20}$$

hence, if A_r is the sinusoidal ripple amplitude, the ratio between the rippled and flat components (from Eq. 4.18) of the spill becomes:

$$\frac{K\,2\pi f\,A_r}{K\Delta Q/\Delta T} = \frac{2\pi f\,A_r\Delta T}{|\xi|\Delta\delta_p} \propto \frac{1}{|\xi|} \tag{4.21}$$

meaning that the higher the chromaticity the more attenuated the ripple on the spill gets.

On top of this, by rescaling the tune variation according to Eq. 4.20 (i.e. in order to maintain a fixed extraction duration and number of extracted particles), the slope of the tune ramp $\Delta Q/\Delta T$ increases proportionally to the chromaticity. This implies that increasing the chromaticity the extraction process becomes intrinsically more resilient to ripples, given that the faster is the tune ramp the larger the ripple amplitude needed to significantly affect the spill. In particular, the ripple non-linearity (i.e. full modulation) threshold increases. To summarize, an increase in chromaticity would both reduce the ripples on the extracted spill and increase the amplitude threshold for a ripple to break the linearity condition. These results have been obtained by only using the simple derivator model of the extracted spill (Eq. 4.4 and 4.18), which does not include the low-pass filter effect. Conversely, nothing can be inferred from the latter model regarding the effect of the sextupole strength, which is based on the variation of the transit times (Fig. 3.29 and Eq. 3.17). By dividing the problem into a derivator-rectifier (Eq. 4.7) and a low pass filter, as attempted in Section 4.2.1, the variation in chromaticity would mainly affect the former, while the variation in sextupole strength the latter.

The Henon map model developed in Subsection 4.2.1 is used for performing the parametric scan. This simulation model is significantly faster than MADX: for a single frequency point one goes from about 10 hours to 5 min of simulation time[2]. The chosen range of the parameters for the scan is:

$$\begin{aligned} \xi &\in [2 \times \xi_0,\ \xi_0/2] \\ V_{\mathrm{ss}} &\in [V^0_{\mathrm{ss}}/2,\ 2 \times V^0_{\mathrm{ss}}] \end{aligned} \tag{4.22}$$

where ξ_0 is the nominal machine chromaticity, of value $\simeq -1 \times 26.67$ and V^0_{ss} is the nominal virtual extraction sextupole strength, of value $\simeq 169.3\ \mathrm{m}^{-1/2}$.

The scan points are taken over a 10×10 uniform grid over the parameter space defined by Eq. 4.22. For every (V_{ss}, ξ) pair, a full transfer function of 13 frequency points (from 10 to 500 Hz) is simulated using the Henon map model. The simulated ripple amplitude has been chosen small enough to fulfill the linearity condition, so to allow to neglect the dependence on amplitude of the transfer function.

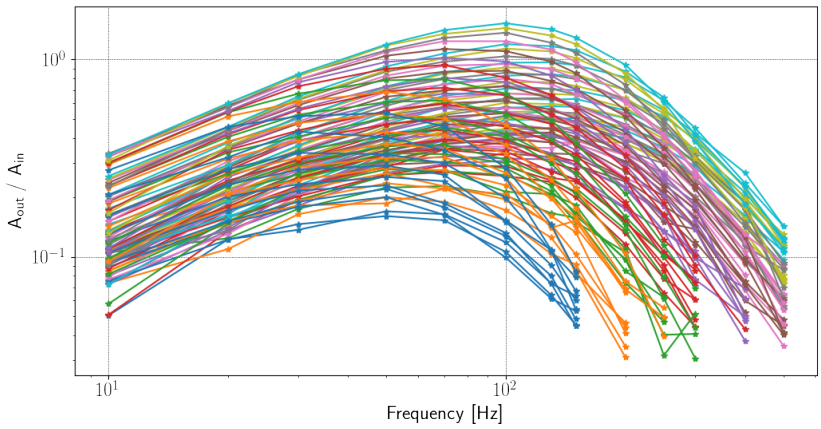

Figure 4.28: Overview of all the obtained transfer functions from the $V_{\mathrm{ss}} \times \xi$ parameter scan.

Looking at all the transfer functions obtained from this grid scan (Fig. 4.28), one can see that about one order of magnitude is spanned along the Y-axis, and the position of the pole spans a few tens of Hertz. This shows that significant improvements in ripple suppression can potentially be achieved.

For a better characterization of best sextupole strength-chromaticity combinations, 2D maps of the suppression of each of the main harmonics are reported in Fig. 4.29. The color code indicates the transfer function value normalized to the nominal one. Attenuations of the 50 Hz component of a

[2]depending on how the simulations are run and how many points are simulated in parallel the advantage can grow even further; on average, there's more than a factor 100 between the two.

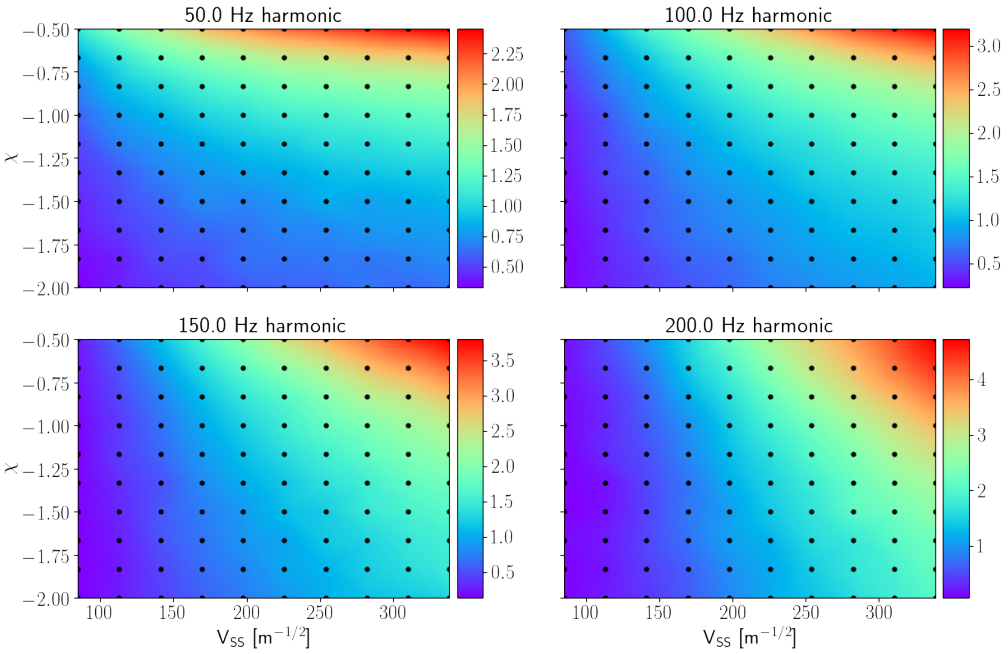

Figure 4.29: Interpolated $2D$ maps of the main harmonics suppression factors in the $V_{\text{ss}} \times \xi$ scan. The variable χ is defined as ξ/Q_0. The color scale shows the ratio between the obtained transfer function value and the nominal one for the considered harmonic component. The black points shows the scan grid.

few folds seem achievable, up to a maximum suppression of about a factor 3. The attenuations increase for higher harmonics. The results of Fig. 4.29 confirm what expected: higher chromaticity (in absolute value) and weaker sextupole strength increase the attenuation of the slow extraction transfer function.

A direct parametrization of the transfer function can be performed by observing the position of the pole and the value of the maximum as a function of the chromaticity and sextupole strength. Both these parameters strongly determine how good is the obtained transfer function in terms of ripple suppression.

Concerning the transfer function maximum, Fig. 4.30 shows that its value depends both on the chromaticity and on the sextupole strength, and that a suppression higher than a factor 2 is within reach.

Figure 4.31 shows the one dimensional projections of Fig. 4.30 for the sextupole strength and chromaticity. The plot on top shows the fixed V_{ss}

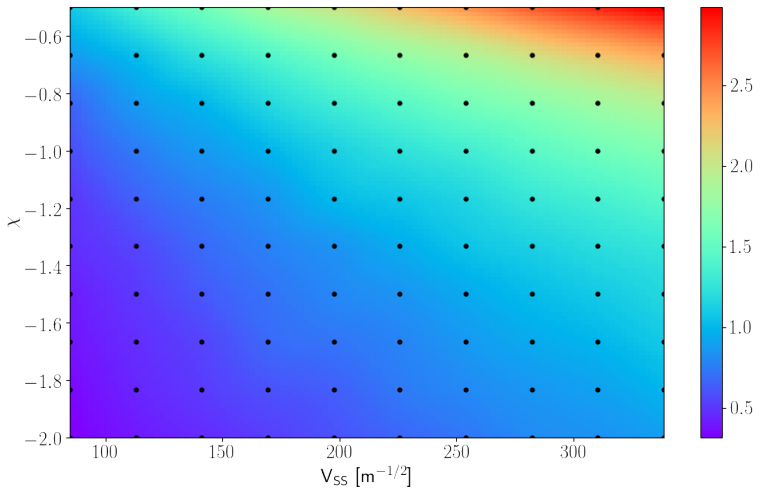

Figure 4.30: Interpolated 2D map of the maximum value of the transfer function for the $V_{ss} \times \xi$ parameter scan. The color scale indicates the maximum normalized by the maximum of the current SPS working point.

curves as a function of the normalized chromaticity χ, while the plot on the bottom shows the fixed χ curves as a function of V_{ss}. The continuous lines are the results of a fit of each curve. In particular, for the fixed V_{ss} curves the following exponential law has been used:

$$m(\chi)\Big|_{V_{ss}} = e^{a\cdot(\chi-b)} + c \tag{4.23}$$

where a, b and c are real parameters which depend on V_{ss}. For the fixed chromaticity curves it has been used a linear function:

$$m(V_{ss})\Big|_{\chi} = k \cdot V_{ss} + q \tag{4.24}$$

where k and q are real parameters depending on χ.

From all the fit results shown in Fig. 4.31 it would be possible to model the dependence of the fit parameters a, b, c, k, and q from Eq. 4.23 and 4.24 and extrapolate an effective analytic model which could be used for fast estimations.

Figure 4.32 shows the evolution of the a, b and c parameters from Eq. 4.23 as a function of V_{ss}. The continuous lines have been fitted over the points and it has been found that in first approximation a can be considered a constant, while b and c follow a linear law.

Figure 4.33 shows the evolution of the k and q parameters from Eq. 4.24 as a function of χ. The continuous lines have been fitted over the points and it has been found that both k and q follow an exponential law of the

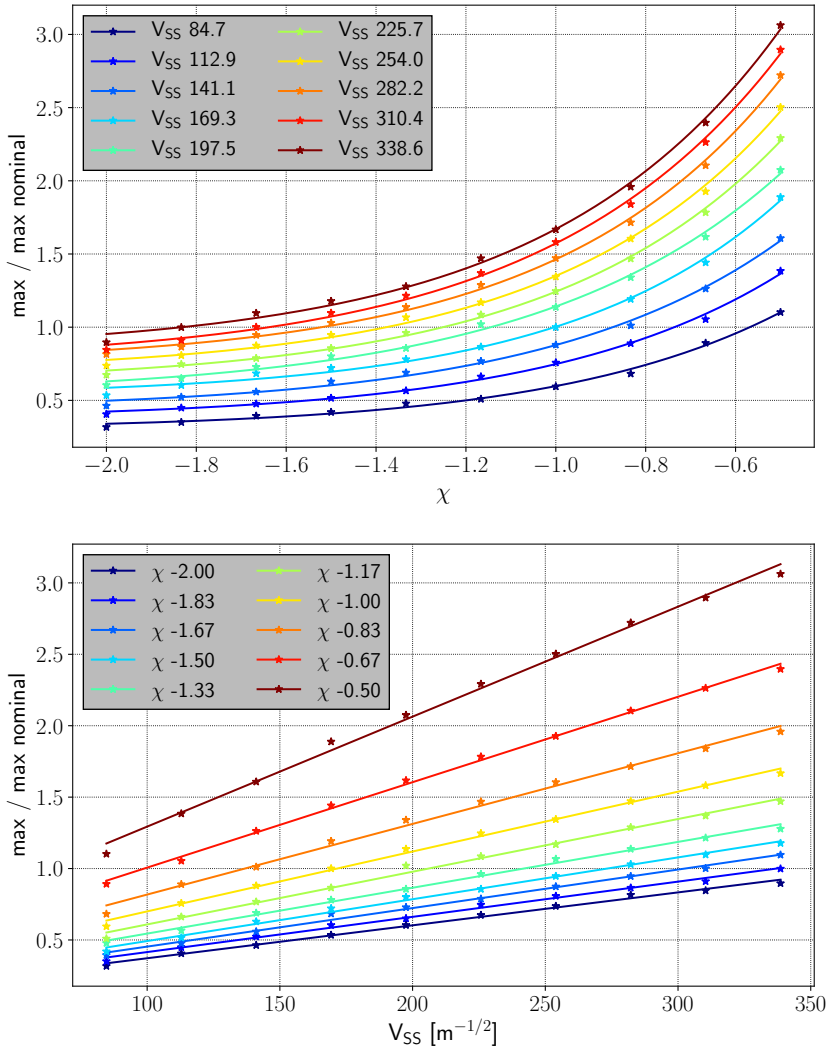

Figure 4.31: 1D projections of the transfer function maxima. Top: fixed V_{ss} curves as a function of χ. Bottom: the opposite.

type of Eq. 4.23. Putting together the results shown in Fig. 4.32 and 4.33, both Eq. 4.23 and Eq. 4.24 can be expanded into analytical two dimensional functions $m(\chi, V_{\mathrm{ss}})$.

The same procedure applied for the transfer function maxima can be also applied for the poles. The position of the pole as a function of the chromaticity and sextupole strength is shown in Fig. 4.34.

The result is not as smooth as the one of Fig. 4.30; this is because estimating the correct position of the pole is a more error prone process, especially when automatized to many transfer functions which can significantly vari-

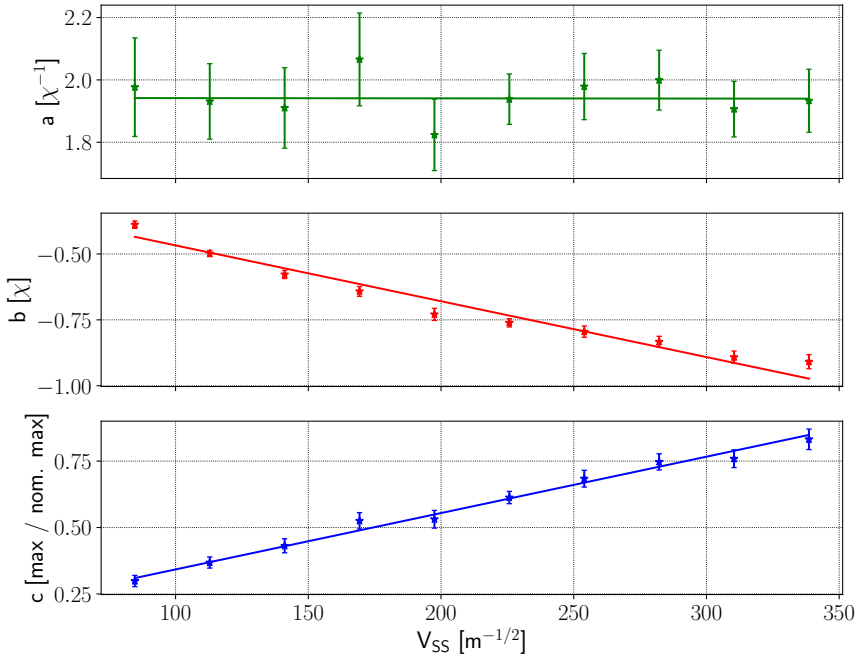

Figure 4.32: Evolution of the parameters a, b and c from Eq. 4.23 as a function of V_{ss}.

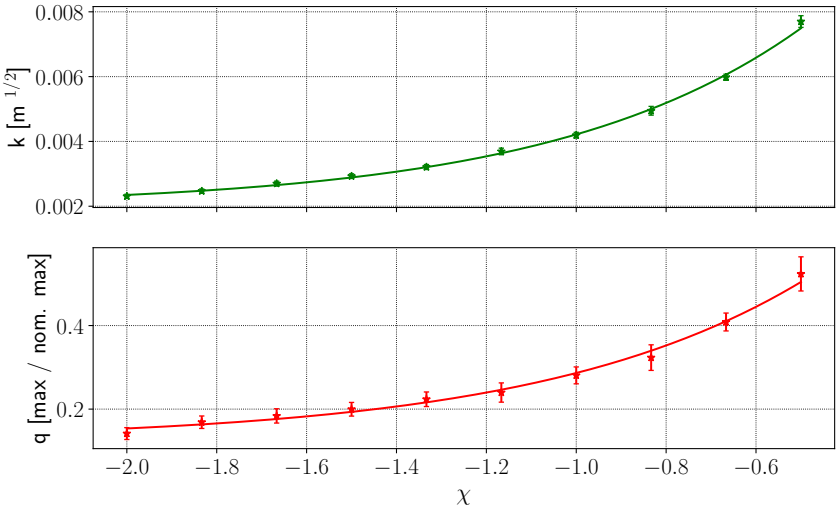

Figure 4.33: Evolution of the parameters k and q from Eq. 4.24 as a function of χ.

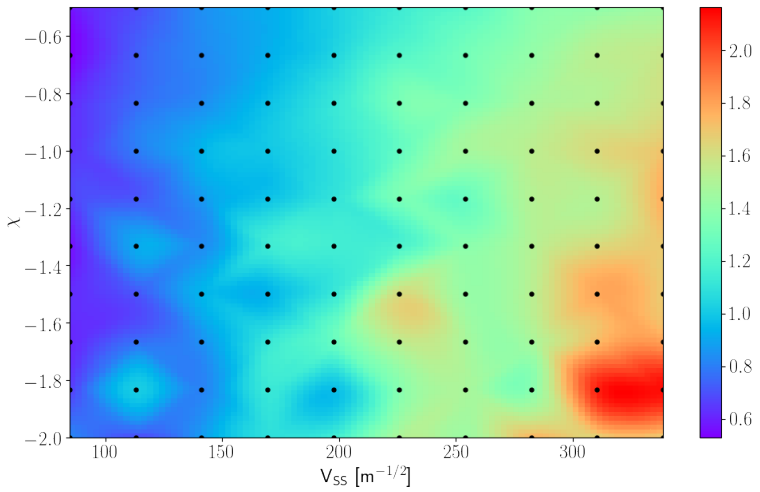

Figure 4.34: Interpolated 2D map of the pole of the transfer function for the $V_{\mathrm{ss}} \times \xi$ parameter scan. The color code indicates the position of the pole normalized by the one of the current SPS working point.

ate in shape. Anyway, the main dependence is on the V_{ss} parameter, being the gradient mostly horizontal. A reduction of the position of the pole of about a factor 0.7 seems within reach. For reference, the nominal pole is located at about 90 Hz. As for the case of the maximum, in order to characterize the dependence of the transfer function pole on both chromaticity and sextupole strength, it is useful to project the simulated points to one dimensional curves.

Figures 4.35 and 4.36 show the one dimensional projections of the transfer function pole map respectively for the sextupole strength and chromaticity. The curves are rather noisy in both figures: this makes it more difficult to find a good fit for the data in order to extrapolate a model, as done in Fig. 4.31. In this case, both the 1D projections have been fitted using a linear model of the type of Eq. 4.24.

Figures 4.37 and 4.38 show the evolution of the linear fit parameters (where k is the slope and q the intercept) respectively for the case of fixed V_{ss} curves and fixed chromaticity curves. Both the results point at the fact that the main dependence of the transfer function pole is a linear one on the virtual sextupole strength, with a small linear dependence on the chromaticity which could be neglected in first approximation, leading to:

$$p(V_{ss}, \chi) \simeq k \cdot V_{ss} + q \tag{4.25}$$

where the parameters k and q could be estimated using both the results from Fig. 4.38 and 4.37, neglecting the weak dependence on χ (which any-

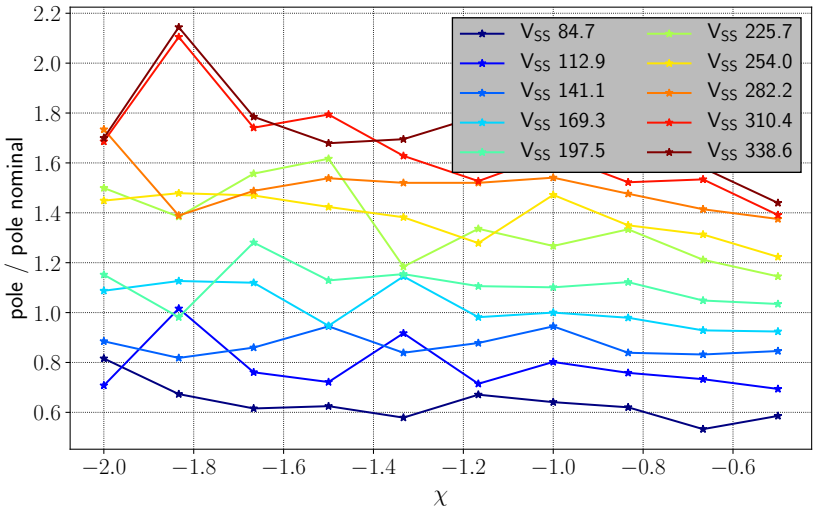

Figure 4.35: Fixed V_{ss} curves of the transfer function pole (normalized to the nominal one) as a function of the chromaticity.

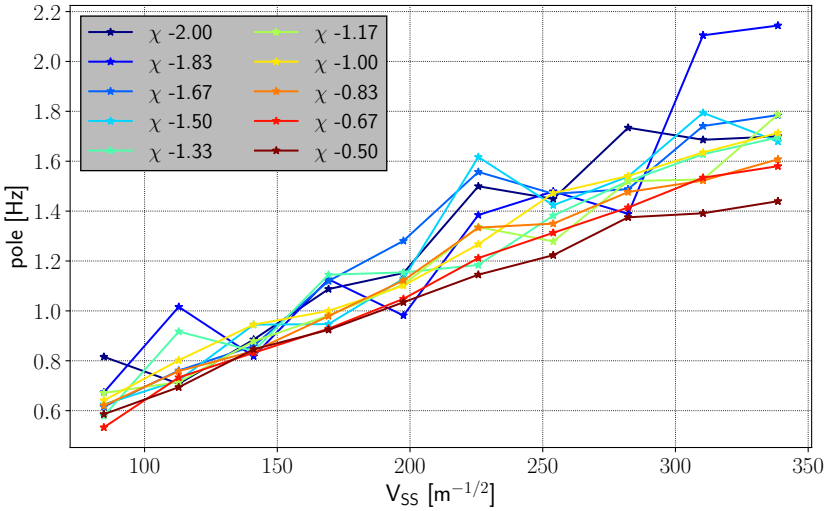

Figure 4.36: Fixed-chromaticity curves of the transfer function pole (normalized to the nominal one) as a function of the virtual sextupole strength.

way could very easily be added for increased accuracy). Equation 4.25 is also in line with the results of the simple model which led to Eq. 3.17, where an inverse proportionality of the time-to-extraction differences to the virtual sextupole strength is found.

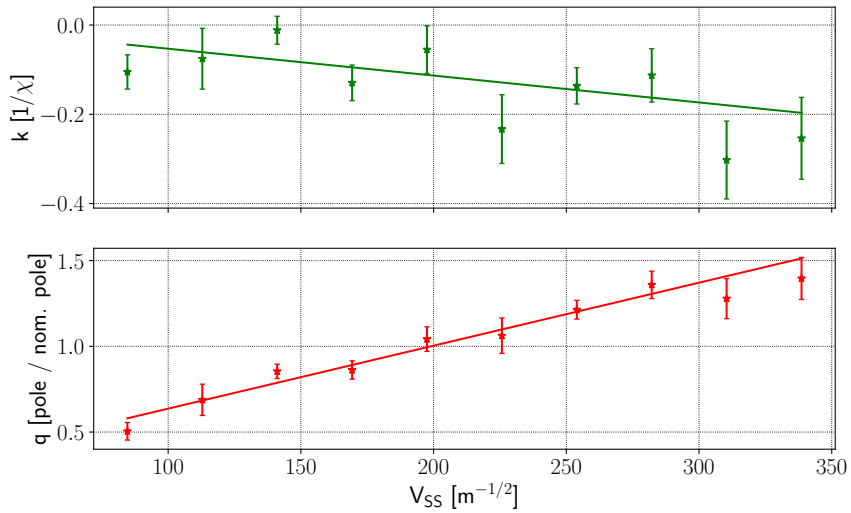

Figure 4.37: Dependence on V_{ss} of the linear fit parameters k and q for the linear fits of the curves of Fig. 4.35.

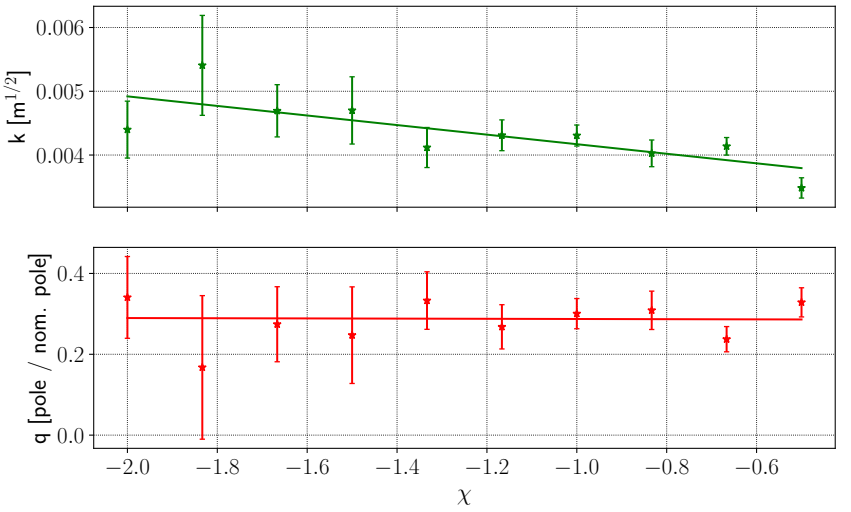

Figure 4.38: Dependence on χ of the linear fit parameters k and q for the linear fits of the curves of Fig. 4.36.

All the results presented in this section have shown that the frequency transfer properties of the slow extraction process could be significantly optimized varying the main accelerator parameters.

The Henon map-based model used for the extraction parameter scan can also be used for different accelerators, without the need of using full lattice files or writing new MADX code: it just needs the global and slow extraction specific machine parameters. In fact, this model has been used to investigate the main characteristics in frequency domain of the MedAustron slow extraction [83], in the context of a recent CERN-MedAustron collaboration[3]. This can also show the dependence of the slow extraction frequency response to the machine design parameters, e.g. tune, circumference, etc. The comparison between the SPS and MedAustron parameters is summarized in Table 4.1. Many of these parameters are significantly different between the two machines. In particular, the one turn time, which differs by a factor ~ 50, the tune ramp speed, with an order of magnitude difference, and the emittance, with almost two order of magnitudes. Both the accelerators are based on a third-integer resonant chromatic extraction.

Parameter	**CERN-SPS**	**MedAustron**
Momentum	400 GeV/c	$\leqslant$ 250 MeV/c
One turn time	23 µs	420 ns
Chromaticity	26.67	4
Total tune sweep	0.1	0.02
Flat top duration	4.8 s	9 s
Momentum range (δ_p)	3×10^{-3}	5×10^{-3}
Virtual sextupole strength	169.3 $m^{-1/2}$	29.8 $m^{-1/2}$
Emittance	1.9×10^{-8} m	6.6×10^{-7} m

Table 4.1: Main slow extraction parameters of the CERN-SPS and MedAustron accelerators, for proton extraction only.

The parameters of Table 4.1, together with the optics at the electrostatic septum and its transverse position, are enough to simulate the MedAustron slow extraction transfer function with the Henon map model.

The fixed-amplitudes transfer functions are obtained for the MedAustron slow extraction, as shown in Fig. 4.39. The low frequency ripples corresponding to 3.5×10^{-5}, 1.4×10^{-4} and 3.5×10^{-4} have been injected during dedicated measurements on ripple injection carried out at MedAustron[4]: the ripple amplitude of 3.5×10^{-5} is in the same order of magnitude of the power supply harmonics already present on the magnets. It can be observed that the transfer function pole is located at about 5 kHz, with respect to the $\sim$ 100 Hz pole observed in the SPS. The main reason for this difference can be traced back to the revolution period: the factor between the poles

[3]P. A. Arrutia Sota, A. De Franco, M. A. Fraser, F. Kuhteubl, C. Kurfuerst, M. Pivi
[4]by P. A. Arrutia Sota, A. De Franco, M. A. Fraser, F. Kuhteubl, M. Pivi

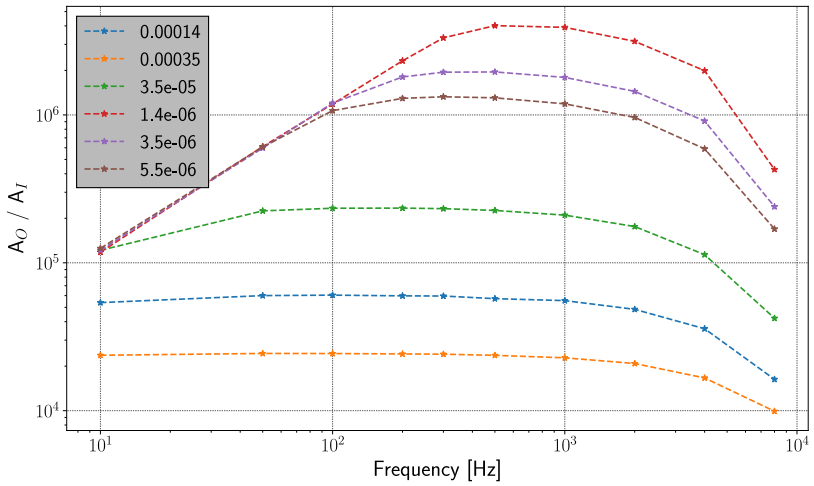

Figure 4.39: Obtained fixed-amplitude transfer functions from the MedAustron Henon map model. The injected sinusoidal ripple amplitudes are in units of tune and indicated in the legend.

is roughly the same factor between the revolution periods, i.e. about 50. This result is well expected from theory [44]. In fact, the threshold value for the ripple attenuation is directly related to the time a particle spends on resonance before being extracted, the so called dead-time. This quantity is expressed in units of turns (e.g. Fig. 3.29) and does not change significantly from one machine to another, with typical values of hundred of turns. The revolution period is what gives a dimension in units of time to the problem, consequently defining the cut-off frequency of the transfer function. Apart from the significant variation on the scale, the behavior of the transfer functions of Fig 4.39 is very similar to what observed for the SPS: a zero is present for small ripple amplitudes (and frequencies), according to the linearity condition of Eq. 4.6. After a certain amplitude threshold, the zero disappears to leave space for the full non-linear behavior, which can be observed in the three transfer functions with the highest amplitudes.

Another important observation from Fig. 4.39 is that the linearity is broken even for small amplitude transfer functions. In fact, the three smallest amplitude transfer functions (1.4, 3.5 and 5.5×10^{-6} in tune amplitude) are compatible with each other only up to less than 100 Hz, while the same ripples injected in the SPS would still be in the linear regime, and hence leading to equivalent curves. The reason for this is the one order of magnitude difference in the speed of the tune ramp that can be observed from Tab. 4.1. The slower tune ramp of the MedAustron accelerator makes it more sensitive to ripples, and reduces significantly the amplitude range for the validity of

the linear frequency transfer, as can be easily estimated using Eq. 4.6. On top of this, the significantly higher frequency of the pole (i.e. when the low-pass filter effect starts to act) makes it easier for a ripple to fully modulate a spill. In fact, as it has been shown for the SPS, only when the low-pass filter effect start to act can a linearity violating ripple be prevented from fully modulating the spill (e.g. see the difference between Fig. 4.3-bottom and 4.4-bottom); otherwise, Eq. 4.4 becomes a good approximation for the spill, which will be split into bursts. By adding this fact to the linear dependence on the ripple frequency for the condition of Eq. 4.6, it is clear that a wider range of ripples has the possibility of fully modulating the extraction. As discussed previously for the chormaticity-sextupole strength scan, increasing the chromaticity in order to be able to speed up the tune ramp could be highly beneficial in this case. These latter considerations imply that the spills in the flat part of the 3 amplitudes injected in the dedicated measurements (lower three curves of Fig. 4.39) are all fully modulating the spills. This is in fact the case for the simulated spills, and was also observed on the data. As a final remark, it has to be said that also a bigger emittance plays a role in increasing the sensitivity to ripple, given the fact that high amplitude particles are extracted significantly faster. From Tab. 4.1 it can be seen that the difference in emittance is almost of two orders of magnitudes between the two machines.

In the future, these results will be fully validated with the aid of a dedicated MADX model and machine measurements by the CERN-MedAustron collaboration. For now they are an interesting example on how the frequency transfer properties can scale from one accelerator to another, but preserving the same fundamental behavior. Moreover, this example shows also how well the SPS does in terms of ripple suppression: the pole is located at about 100 Hz and the horizontal emittance is very small, as a consequence of the $\simeq 6.9$ Km ring circumference and 400 GeV flat-top particle momentum.

4.6 Conclusions

In this chapter the frequency transfer properties of the slow extraction for the SPS have been investigated. A SPS slow extraction MADX model has been used to simulate the process, showing that for small ripples the problem is linear, and a transfer function can be defined. For amplitudes and frequencies combinations above the linear threshold it is still possible to build an effective model. By only simulating a limited number of fixed amplitude transfer functions in the non-linear regime, a semi-analytical map can be constructed in order to instantly predict the spill frequency response to a given input current modulation. Both the full MADX model and the semi-analytical one have been validated using data from dedicated measurements.

A satisfying agreement has been achieved for almost all the experimentally injected ripples up to 200 Hz, the only problems being related to a very low intensity of the spill signal. Operational data was also used to validate the developed model, showing difficulties to obtain a precise estimation of the 50 Hz harmonics on the input current. To overcome this, the model is validated using a continuous transfer function approach. Different methods for removing the measurement noise from the logged currents have been tested, obtaining results which are not far from the simulated ones. Anyway, a level of uncertainty on the operational data still remains, due to the presence of other possible tune modulating effects which are not taken into account in the model. All of these uncertainties will be further investigated in future measurement campaigns, while some further details and possible limitations of the developed model are discussed in Appendix B.

A Henon map-based model of the frequency transfer is also developed, showing a very good agreement with the MADX one. The model presents great advantages for the simulation times and it is used for a scan of the virtual sextupole strength and chromaticity. The results of the scan show that further attenuation of the current ripples would be possible for smaller sextupole strength and higher chromaticity.

Concerning the injection of a dominant high frequency ripple as an alternative spill-quality improvement technique, the MADX simulation has confirmed its potential, but further studies will be needed to understand its compatibility with the experiments requirements.

Finally, this work opens the possibility to have an online prediction of the effect of a current variation on the spill, and hence help the design and operation of slow extraction systems. For instance, this will be fundamental for the SPS to guarantee the high-quality spill demands from the North Area experiments, and preserve the SPS world leading quality of its fixed target facilities.

Chapter 5

The magnetic horn for the ENUBET project

The possibility of using a magnetic horn to increase the secondary particle flux of the ENUBET experiment is strictly connected with the employed proton extraction scheme. As the burst mode slow extraction has been successfully modeled and tested, studying the potential flux gains achievable with a horn becomes the next step of the project. The result will be significantly dependent on the particular design of the employed magnetic horn. Reaching a satisfactory design is a complex task, characterized by a high number of parameters and constraints. In this chapter, a dedicated framework for a full automated optimization of the ENUBET magnetic horn will be developed, leading to first possible horn candidates and estimates of kaon-flux gains.

The structure of the chapter is divided in five main sections. Section 5.1 introduces the magnetic horns and their use in neutrino beamlines (Subsection 5.1.1), for then focusing on the particular case of ENUBET and its requirements, specifying the scope of the present work (Subsection 5.1.2). Section 5.2 describes the modeling and simulation of the ENUBET magnetic horn, also showing the geometric parametrizations chosen for the study. The developed optimization framework is detailed in Section 5.3, where a discussion on the possible figures of merit is reported (Subsection 5.3.1), followed by a description of the employed optimization algorithm (i.e. the genetic algorithm, Subsection 5.3.2), framework implementation details (Subsection 5.3.3), and selected constraints and speed-up approximations (Subsection 5.3.4). The results of the optimization runs are presented in Section 5.4, where the beamline-independent results are first analyzed for the two selected horn geometries (Subsections 5.4.1 and 5.4.2), followed by the outcomes of a full-beamline tracking of the horn-focused particles (Subsection 5.4.3). Section 5.5 summarizes the findings of the chapter, discussing the possible next

steps of the magnetic horn design and optimization for ENUBET.

5.1 Introuction

5.1.1 Magnetic horns

Since its invention by Simon Van Der Meer in 1961 [106] and its first development for a neutrino beam [107], the magnetic horn has always played a key role in maximizing the neutrino flux for accelerator neutrino experiments. A magnetic horn consists of a "folded" conducting sheet on which a current flows in the longitudinal direction, e.g. entering from the inner side of the conductor and exiting from the outer side. This implies that a toroidal magnetic field is generated in the region enclosed between the inner and outer conductors.

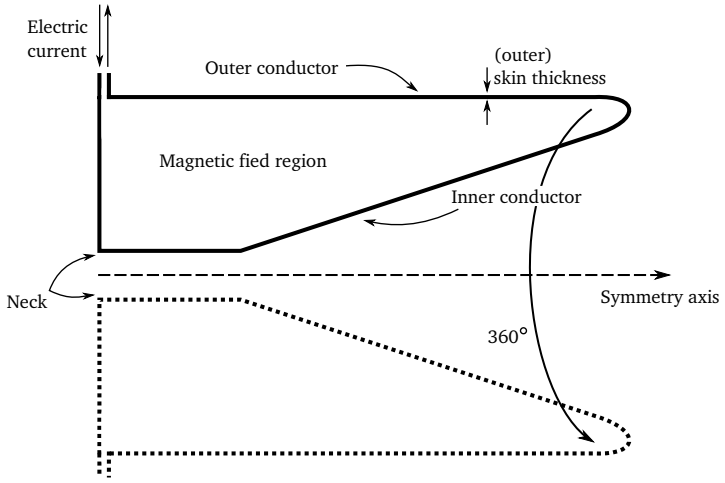

Figure 5.1: Cross-sectional schematic view of a magnetic horn.

As shown in Fig. 5.1, a magnetic horn can be thought as the 2π rotation of a closed circuit around a symmetry axis. The symmetry axis corresponds with the longitudinal beam direction, and so with the axis of the proton target, which is usually positioned upstream (and/or inside) the horn. The cylindrical symmetry of the problem defines an azimuthal magnetic field B_ϕ inside the region enclosed by the conductors[1], which can be readily computed using the Ampére's law:

$$B_\phi = \frac{\mu_0 I}{2\pi r} \tag{5.1}$$

[1] the magnetic field inside the conductor material is usually neglected in a good approximation

where I is the total current flowing in the horn and r is the radial distance from the symmetry axis. The magnetic field of Eq. 5.1 represents overall a good approximation for the experimental case, as shown, for instance, in [108] (where the skin effect is also reported to extend the field range into part of the conductor).

From the schematic drawing of Fig. 5.1 and Eq. 5.1 it follows that the magnetic horn is not the standard accelerator magnet: in order to be bent by the magnetic field, particles have to cross the device itself, making it intrinsically lossy. Moreover, the magnetic field is strongly non linear, and it also couples the X and Y transverse phase space dimensions. Any analytical treatment holds only under approximations [9, 109], not only for the non-linear field, but also because particles can enter and exit the magnetic field region in many different ways. For this reason, the best way of quantitatively predicting the effect of a magnetic horn on the beam is through numerical simulations.

The geometric shape of the horn conductor can significantly influence the horn overall focusing, and it can be shown that different families of horn geometries have different focusing effects [9]. However, the overall focusing action of a magnetic horn is that of a "rough" focusing with high acceptance: it is not a device for a precise point-to-parallel (and vice versa) focusing, but it focuses a broad spectrum of particle momenta and transverse angles. These qualities make the magnetic horn very advantageous for wide band neutrino beams (WBB), in which it focalizes most of the secondaries emitted from the target, increasing the fluxes typically by one order of magnitude with respect to a bare-target beam. Another advantage of the horn is that while it focuses particles of one sign, it has a defocusing effect for the particles of the opposite sign, hence helping in the charge-selection of the beam. Both the focusing and charge-selection properties of the horn can be further increased by using horn doublets or triplets (e.g. as done for quadrupoles): every horn after the first will further focus, narrow the spread, and charge-select the beam. This technique is still being used in many recent wide band neutrino beamlines, e.g. T2K [10], NuMI [11], CNGS [12], in which two or three horns with increasing transverse dimensions are following the proton target, and focusing the beam directly into the decay tunnel, allowing for a significant neutrino flux increase.

The development and operation of a magnetic horn is not easy, and it entails several limitations and problems. The typical currents needed to achieve a satisfactory focusing of the secondaries are ranging between 100 and 400 kA, which implies that the device can only be used for short intervals of time due to an otherwise excessive heating which could irreparably damage the conductor. The heating comes both from the Joule dissipation (combined with the skin effect) of the injected current and from the interaction of the secondary particles in the conductor. The latter source of

heating is fought by choosing a skin thickness as small as possible (typical values in the millimeter range) and a low-Z material for the conductor, as for instance particularly optimized Aluminum alloys, which will also have to withstand very high levels of radioactivity. On top of this, each horn needs to be coupled with a rather cumbersome water cooling system, by which water is sprayed on the hottest parts of the horn via nozzles, and flows into pipes around the full device. Other potential issues are related with the mechanical forces exerted by the strong magnetic field, especially in the region of the horn neck, and the consequent vibrational modes due to the fact that the field is pulsed. This can cause significant stresses in the material, and limit the lifespan of the device. All of this issues are the reason why the typical current pulses of the horns are within $\sim$ 100 µs and $\sim$ 1 ms. The power supply is usually based on a capacitor bank discharge, which produces a unipolar half-sinusoidal current pulse. In order to keep the magnetic field as constant as possible, the beam is extracted in times which are orders of magnitude shorter than the power supply current pulse. For instance, for the case of CNGS [12], the 3-4 ms-long current pulse is used to focus a 10 µs fast-extracted proton spill.

Given the complexity of the problem, finding the best horn geometry for a particular experiment is not trivial, and requires dedicated numerical simulations exploiting particle tracking and interaction codes. The problem is usually treated as an optimization problem, initially taking into account the hardware limitations in the form of geometrical constraints, and maximizing for the particular requirements of each case, as done for example in [110–112].

5.1.2 The ENUBET requirements and the scope of this work

The main goal of the ENUBET project is to carry out a precise measurement of electron neutrino cross section. The low number of electron neutrinos produced in the energy range of interest makes this task a potentially long process (up to a few years, depending on the beam availability and statistical precision), and every possible way to speed it up should be carefully considered. Using a magnetic horn coupled with the burst mode slow extraction developed in Chapter 3 is the leading solution for such a problem.

For the case of ENUBET, however, the use of a magnetic horn would not be as standard as for most of the other neutrino beams, for two main reasons:

- ENUBET is a narrow band neutrino beam (NBB), with a reference momentum of 8.5 GeV and a 5-10% momentum bite. For a NBB the advantage of the broad momentum acceptance of the magnetic horn is lost due to the small selected momentum range, while only the advantage in the angular acceptance remains.

- Being a monitored neutrino beam, ENUBET requires the beam to be slow extracted. In particular, the candidate extraction for the horn would be the burst mode slow extraction, with spill lengths between 2 and 10 ms.

The latter point is the most critical one. Magnetic horns have always been used coupled to a fast-extraction of the primary protons ($\mathcal{O}$(10 µs) of spill lengths), so the situation is rather peculiar. The main issue with the millisecond-long spill would be the consequent length of the current pulse, which could lead to severe overheating. On the positive side, the intensity of each pulse would be significantly lower (at least one order of magnitude) than the typical fast-extracted ones, leading to a smaller contribution from particle interactions. Concerning the designed 10 Hz repetition rate, a magnetic horn pulsed at 12.5 Hz, and its corresponding power supply, had already previously been designed and carefully investigated in the context of the European Neutrino Super Beam (EUROnu Super Beam) [113, 114]. However, the focusing of tens of millisecond-long pulses repeated at 10 Hz goes probably beyond the state of the art of magnetic horns, and will require a dedicated in-depth study.

The goal of this work is to start an in-depth investigation of the potential flux increase coming from using a magnetic horn in the ENUBET beamline, and to develop a full framework for the horn geometry optimization given the experiment requirements. Only as a future next step, once an optimized geometry and a satisfactory flux increase will be reached, the issues discussed above will be fully addressed.

5.2 Horn modeling and simulation

A standalone simulation of the ENUBET magnetic horn has been implemented in GEANT4 [115–117], based on a previous model used at the beginning of the project. The simulation implements the horn-target system with a high number of parameters allowing to control the geometry, the materials, and other technical aspects of the model.

A view of the GEANT4 simulation can be seen in Fig. 5.2, where it can also be observed that positive particles are focused and negative particles defocused, as expected in the nominal operation of ENUBET (i.e. electron neutrino flux produced from positive kaons). This will be the configuration used in this study, but the polarity could very easily be inverted when needed, so to investigate the anti-neutrino operation.

The target employed for this work is a graphite rod with a 10 mm diameter and 140 cm length. This configuration is based on the results of a previous study, which used the number of secondary kaons as a figure of merit for the design of the target.

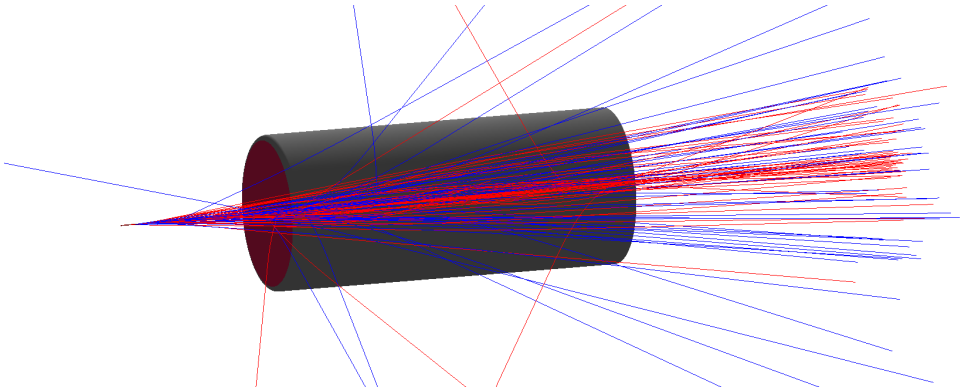

Figure 5.2: Three-dimensional view of the simulated target-horn complex. Pion tracks are shown for reference (blue π^- and red π^+).

In order to speed up the simulation, the primary protons-target interactions have been decoupled from the horn (and beamline) model, and simulated with FLUKA [118, 119]. The results have been stored into a database, and are loaded on demand into the horn simulation.

While it is possible to add a second horn in the simulation, in this work only a single magnetic horn will be considered. Adding a reflector would significantly increase the length of the beamline, which will still need dipoles and quadrupoles to precisely select the particles momenta and tilt the charged kaons with respect to the neutral ones. A second horn could then lead to a higher number of early kaon decays because of the increased length of the full beamline.

Two different main horn geometries have been parametrized and implemented in the simulation model.

The first one is a horn of the type first developed by Palmer [120], and similar to the one used by MiniBOONE [108]. This geometry has also been more recently further developed and optimized for the EUROnu Super Beam [110, 113].

The details of the parametrization are shown in Fig. 5.3. In particular, in addition to the reported parameters there are also the electric current I and the conductor thickness t. This particular parametrization has been selected because it has already been developed and built for other single-horn beamlines [108, 110, 113]. One practical concern of such geometry is that the target could eventually be located inside the horn neck (the narrowest part and most subjected to magnetic stresses) to maximize the flux gain. This could lead to several potential issues during operation, especially given the particular case of the ENUBET proton extraction.

Another well established horn parametrization which has been imple-

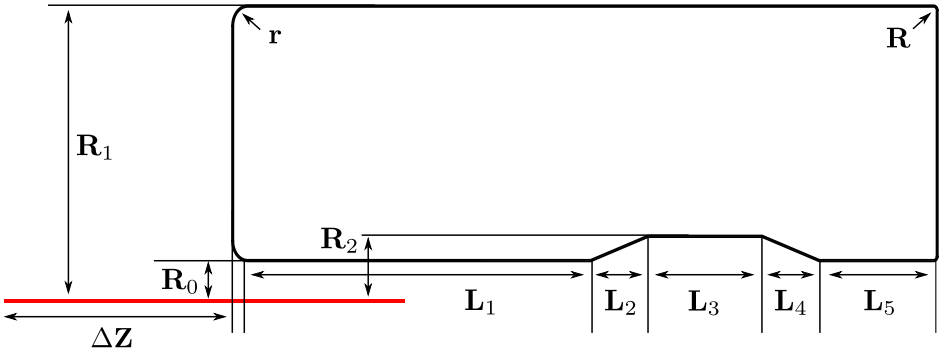

Figure 5.3: Schematic view of the MiniBOONE-style horn geometric parametrization.

mented is the double parabolic one. This particular geometry should have the characteristics of being able to focus a narrow region of particle momentum with high angular acceptance [9], and it is currently being used in the NuMI beamline [11] and considered for nuSTORM [112, 121].

The particular parametrization used for ENUBET has been slightly expanded with respect to the standard fully quadratic profile [9]. In particular, by defining r as the transverse radial coordinate and z the longitudinal one, the chosen profile follows the equation:

$$r = a\,z^2 + b\,z + c \tag{5.2}$$

where a, b and c are real coefficients. Given that the parabolic profile has to pass by two points only, one degree of freedom remains unconstrained, and it is used a new configuration parameter. In particular, this new parameter is chosen to be the radial coordinate of the parabolic minimum, i.e.:

$$r_{\min} = -\frac{b}{2\,a} \tag{5.3}$$

The implemented double parabolic geometry is reported in Fig. 5.4, in which also the two parabolic minimum parameters of Eq. 5.3 are shown, and referred to as m_1 and m_2. The net effect of a significant variation of $r_{\min}$ from the ideal parabolic case of $b = 0$ is to approximate a conical horn shape, as shown in Fig. 5.5.

Once the configuration parameters for the horn have been selected, the simulation performs the tracking of each secondary particle across the space and saves into a ROOT [122] file the coordinates, momenta, and type of every particle crossing the final longitudinal position of the horn.

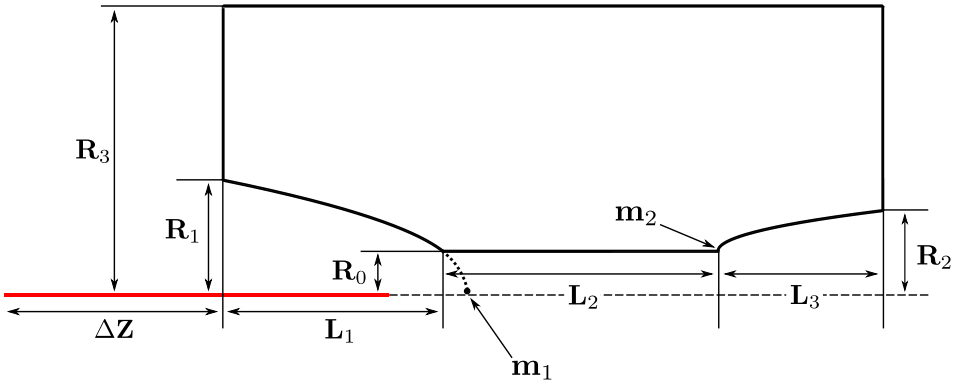

Figure 5.4: Schematic view of the double-parabolic horn geometric parametrization.

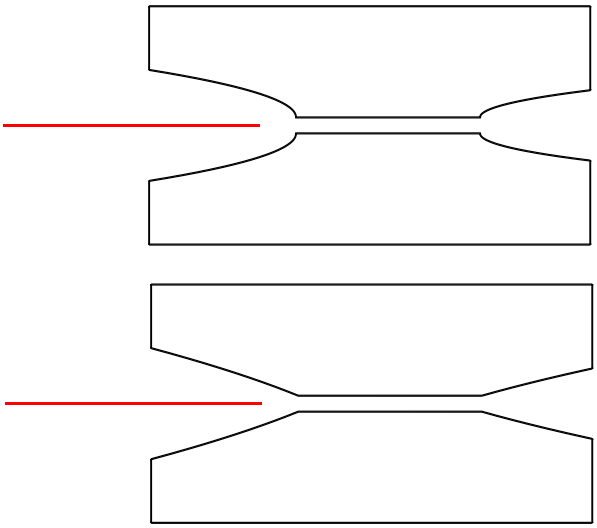

Figure 5.5: Top: standard double parabolic horn profile ($r_{\text{min}} = 0$). Bottom: "conic-like" profile obtained by decreasing r_{min} to negative values.

5.3 Optimization setup

5.3.1 Figure of merit and starting point

The geometry of the horn needs to be optimized so to maximize a certain figure of merit (FOM), or otherwise called objective function. For the ENUBET experiment, this quantity is related with the number of K^+ which are ultimately focused into the decay tunnel: this directly translates into the produced electron neutrino flux. So, with a straightforward definition, the

figure of merit for the horn could simply be defined as the number of K^+ entering the instrumented decay tunnel. However, the main problem with this definition is that it is strongly beamline-dependent.

Up to now, all the secondary beamlines studied and developed for ENUBET assume a nominal slow extraction of $\mathcal{O}(1\text{ s})$ duration, and employ only static focusing elements (namely, dipoles and quadrupoles). The simplest approach would be to insert the magnetic horn between the target and the start of the static beamline, but it is not given that this would be the best solution, and significant changes could be needed. The design of a dedicated beamline for the horn goes beyond the scope of this study: for this reason a beamline-independent figure of merit is chosen as the number of K^+ in the selected momentum bite at a fixed distance from the horn, and contained in the typical acceptance of a static element. From the studies on the static beamline, the latter acceptance corresponds to 15 cm of transverse radius (the typical aperture of existing quadrupoles) and a transverse angle of about 1.5 degrees (after which it has been found that particles are not transported anymore in the static beamline). To assess the performance of the horn with respect to the pure static case, this value can be normalized by the same figure of merit computed without the horn: a simple ray-tracing from the target to the first quadrupole. This latter normalized quantity will be used as a horn-specific figure of merit, and can be expressed as:

$$\Phi(z_Q, \theta_M, r_M) = \frac{N^h_{K^+}(z_Q, \theta_M, r_M)}{N^s_{K^+}(z_Q, \theta_M, r_M)} \tag{5.4}$$

where $N^{(h/s)}_{K^+}$ are the number of positive kaons (with and without the horn) at a distance z_Q from the target, and contained into a maximum radius and angle of respectively r_M and θ_M. The latter limits are fixed to the values detailed above, while z_Q to a realistic value of $\mathcal{O}(1\text{ m})$.

The starting point of this work are the results from a past manual scan of the horn parameters performed using the geometry of Fig. 5.3. The horn was directly assumed to be part of the static beamline, and the number of K^+ at the entrance of the decay tunnel was used as a figure of merit. The results of this scan showed a $\sim$ 5 times increase in the K^+ at the decay tunnel entrance (i.e. downstream end of the beamline).

The best configuration found during the scan (re-implemented in the new simulation) is shown in Fig. 5.6. If this horn configuration is used to compute the beamline-independent figure of merit defined above (Eq. 5.4), the result is about a value of 2. This value is significantly lower than the factor 5 found in the former study, and confirms how the full beamline affects the figure of merit. More details on this will be observed later in this chapter, but, for now, an optimization framework will be built for the horn alone.

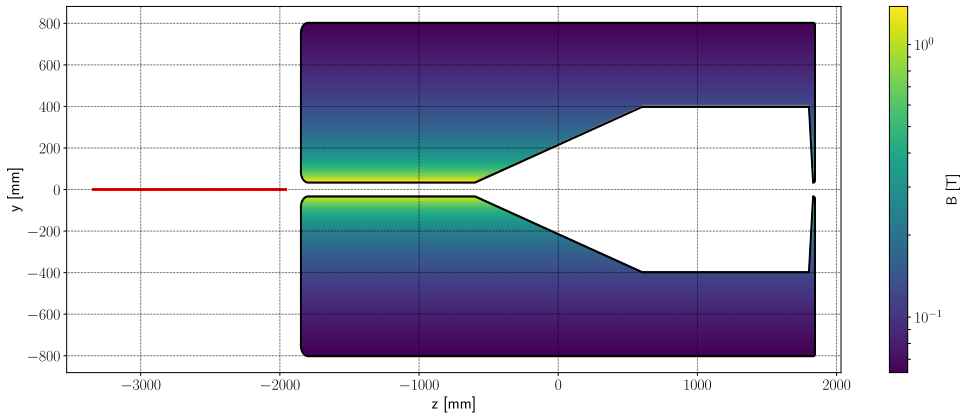

Figure 5.6: Geometry and magnetic field of the best horn found in the first manual scan performed in 2016.

5.3.2 Optimization algorithm

Instead of a manual scan, performing a large scale and fully automated optimization is the goal of this second phase of the dedicated study on the magnetic horn option for ENUBET.

The problem of horn optimization involves a large number of parameters. For the case of the MiniBooNE-like geometry reported in Fig. 5.3, the number of optimization parameters could be as high as 13, while for the double parabolic profile of Fig. 5.4 it would reach 12. On top of this, the non-linearity of the magnetic field and the complex shape of the conductors do not allow to predict much information on the proprieties of the objective function (intended as function from the parameters space to the figure of merit), in terms of expected number of local maxima and relative locations, gradient, and so on.

These particular characteristics make the class of population-based optimization methods very suitable for this problem [123]. The latter algorithms allow to scan vast parameter spaces in search for a global optimum without easily getting trapped in local minima. One of the most popular of such methods is the genetic algorithm, which has also recently been proven successful for the case of nuSTORM [112]. Given its versatility and straightforward implementation, it will be chosen for the first optimization study of the ENUBET magnetic horn.

The genetic algorithm falls into the category of heuristic methods[2], inspired by the process of natural section. The algorithm is based on a very

[2]the certain convergence to the global optimum in the parameter space cannot be mathematically proven, but it is often reached in the practical case

general set of rules, which could easily be applied to many different problems and highly customized. It is probably easier to understand its basic working principle with an example case, as shown in Fig. 5.7.

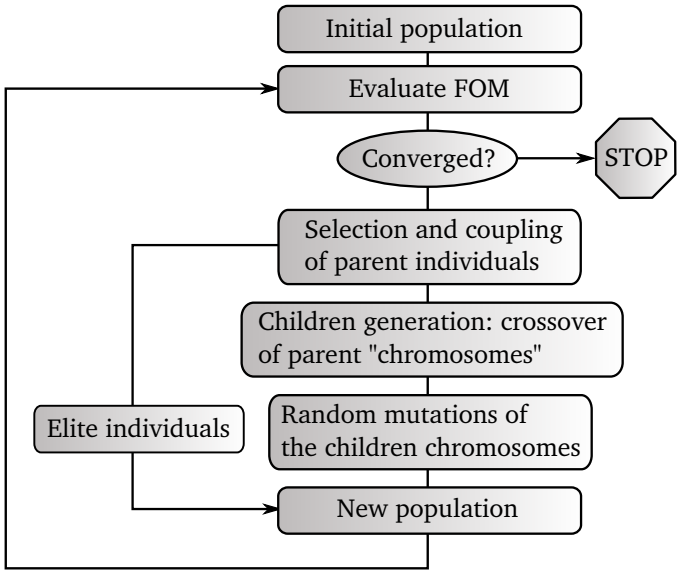

Figure 5.7: A schematic example of the flow of a genetic algorithm implementation.

In particular, every step of the algorithm reported in Fig. 5.7 can be implemented and customized in several ways, with the result depending on the particular problem, and with no strict guidelines. Using terms inspired from the natural selection process, one individual is defined as a point in the parameter space, while its chromosomes constitute a particular encoding of all the parameter values. The individuals of a generation are combined according to their figure of merit, so to form a new generation for the successive iteration of the algorithm. Of particular importance is what is referred as "elitism" [124], which consist on allowing the survival of the best performing individuals to the next generation. In fact, it can be proven that elitism is a necessary condition for the stochastic convergence of the algorithm [125]. The fulfillment of constraints (of key importance for the horn case) can also easily be implemented inside the scheme of Fig. 5.7: a simple penalty method [123] (i.e. associating a low figure of merit) or a logic check will exclude from the result forbidden regions of the parameter space.

5.3.3 Optimization framework

A dedicated optimization library has been developed for the task, containing the most popular implementations of the steps of Fig. 5.7 and data logging

routines. More in detail, for the particular problem of the horn, it has been chosen to use real chromosomes (i.e. parameter-space vectors), paired with uniform and interpolation crossover, and both gaussian and uniform offspring mutations. Of particular importance is the uniform mutation operation, which contributes significantly to avoid getting trapped in local optima. Concerning the constraints (in particular, global inequality constraints), an attempt to replace individuals not satisfying them is made before the evaluation of the objective function, so to save simulation time. If the latter approach fails, a penalty method is used. The application of the developed routines on a test function is reported in Fig. 5.8.

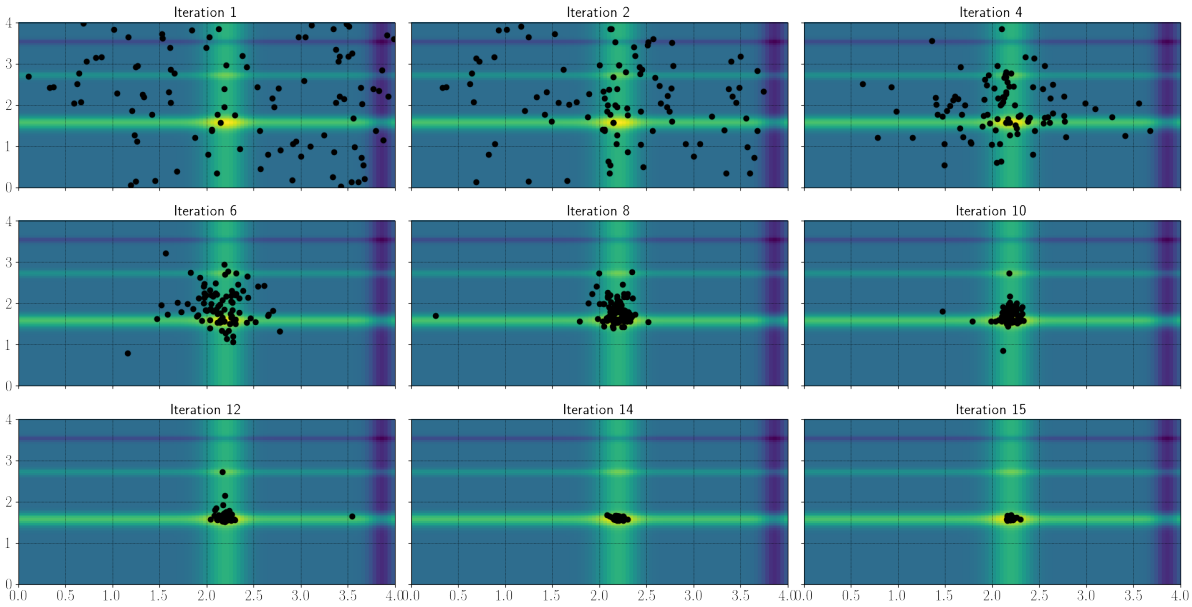

Figure 5.8: Example of the convergence of the genetic algorithm on a 2-dimensional test function with one local and one global optimum, and several lines of continuous local optima. The color scale is proportional to the function value.

Every black dot represents an individual in the parameter space, and each iteration follows the flowchart of Fig. 5.7. In particular, it can be observed how the local optimum of the problem is correctly identified and successively discarded for the global optimum. Concerning the convergence criteria, there are no strict guidelines for the case of the genetic algorithm, being a stochastic-based heuristic method. One possibility is, for instance, to monitor the differential evolution of the average objective function (eventually, for a selected number of individuals). It is important to keep in mind that the convergence of the algorithm does not imply convergence to the global optimum. The latter significantly depends on the values of the optimization "meta-parameters" (i.e. the configuration parameters of each optimization

step of Fig. 5.7), and it is problem dependent. The starting population of individuals is generated using a latin hypercube sampling method [126] in order to cover the parameter space as efficiently as possible.

A software framework has been developed to fully manage the optimization process. The main idea has been to keep a straightforward approach to the problem, and exploit as much as possible the available resources at the computing centre CC-IN2P3 [127], based on the Grid Engine batch-queuing system. A simplified schematic of the framework structure is shown in Fig. 5.9. In particular, many different specific libraries and scripts (as the Python genetic routines, `C++` data processing functions, and Bash book-keeping and job scheduling scripts) are all managed by what is called the optimization manager code, which, once launched, has been designed to perform a full optimization without the need of any supervision.

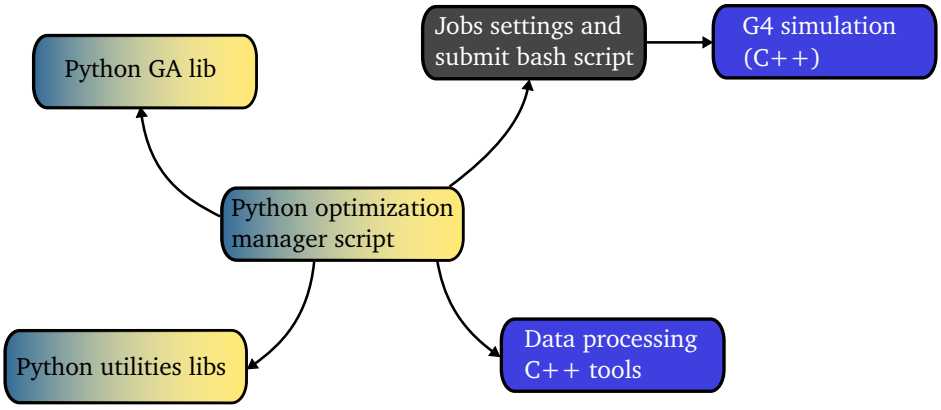

Figure 5.9: Schematic of the optimization framework software structure. Each block represent a dedicated library or script, while the arrows symbol a "uses" relation.

The bottleneck of the process in terms of timing is the GEANT4 magnetic horn simulation: for this reason, this part of the software is submitted on the computer cluster via Grid Engine. This allows to submit as parallel jobs all the different individuals of the current population (i.e. all the different horn candidates), hence making the elapsed time for a single iteration roughly equal to the time of a single horn GEANT4 simulation (without taking into account eventual queue delays and waiting times). The flow diagram of the optimization process is shown in Fig. 5.10. The optimization manager code defines this high level loop structure, while from Fig. 5.9 it follows that every part is performed by a different unit of the framework. By executing the main loop as a daemon process, the optimization has been proven to run

undisturbed for days.

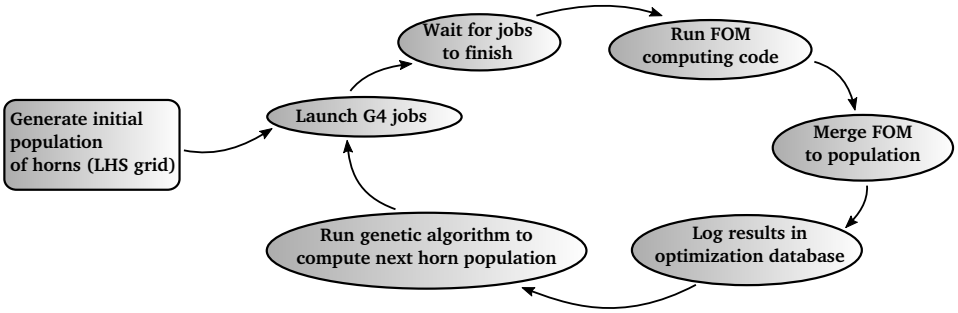

Figure 5.10: Schematic of the flow of the optimization framework: the process will repeat until the exit condition, which is usually the algorithm convergence. The handling of potential failure cases which would interrupt the loop have been implemented in the code after many tests.

While for the time being only the genetic algorithm has been implemented, the developed optimization framework of Figs. 5.9 and 5.10 is completely general, and could be run with any other optimization algorithm of choice. Expanding the optimization to different algorithms is in fact one of the objectives for the future. The same can be said for different figure of merits, or different GEANT4 simulations, which would only require to replace/modify a single unit of the framework.

5.3.4 Horn constraints and approximations

As previously mentioned, the parameter space of the horn configuration can have a number of dimensions as high as 13. In order to to speed-up the optimization and increase the chance of convergence to a global optimum, some less influential degree of freedom has been removed. For instance, the horn conductor thickness will be kept at a reasonable value of 4 mm, while the **R** and **r** parameters of Fig. 5.3 are fixed to a value of a few centimeters. These constraints make the number of dimensions for the MiniBooNE-like horn geometry equal to 10, while the one for the double-parabolic case becomes 11 (including also the parabolic minima $\mathbf{m}_1$ and $\mathbf{m}_2$, referred to Fig. 5.4).

Other inequality constraints are enforced on each remaining horn configuration parameter, based on realistic values or on safety limits to ease the horn construction. For instance, the horn current is limited to a maximum of 350 kA, while its total length to a maximum of 3.5 m. The length of the horn, other than for its construction, needs to be limited also to keep the ENUBET beamline as short as possible.

The enforcement of the previous constraints is not enough to significantly speed-up the optimization process, which would take a few hours to reach a statistical precision of the 2% on the output flux of selected K^+, which are used for the computation of the figure of merit. This could make an optimization loop to last from two weeks to a month, depending on the number of iterations and requested precision on the figure of merit. In order to avoid such prohibitive times, some more aggressive approximations are made in order to significantly speed-up the GEANT4 simulation of the horn.

The first approximation is a momentum cut below the lower end of the selected momentum bite (i.e. 8.5 GeV $\pm$10%), applied both at the level of the input target file and of the GEANT4 stepping action. An upper cut on the momentum is not nearly as effective, since more that 90% of the produced secondary particles populate the region below the cut, as it can be observed in Fig. 5.11.

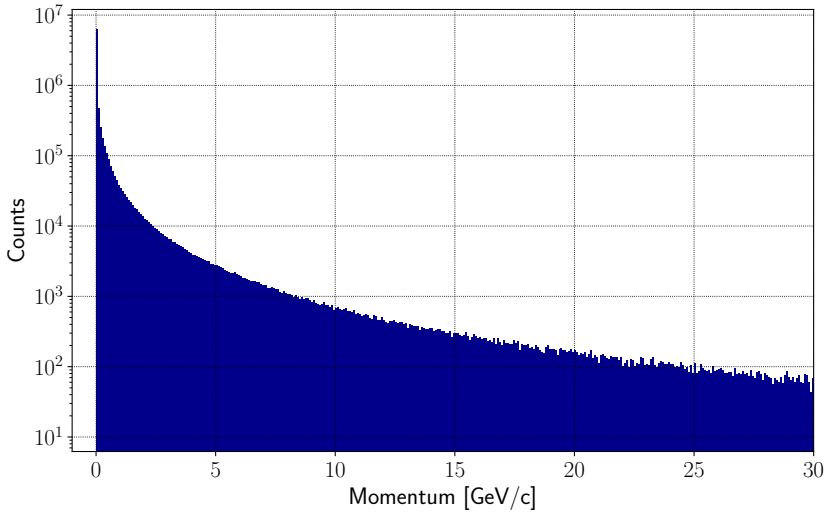

Figure 5.11: Momentum distribution of the all the secondary particles produced by the protons interaction on the target, for a number of protons of target of 10′000.

The second approximation is to remove all the particles which are not kaons. This approximation removes about the 90% of the particles remained from the previous cut, as it can be observed by looking at Fig. 5.12. In particular, the most abundant particle type are pions.

While the lower cut on momentum cannot produce any variation in the figure of merit (because of the momentum conservation), it is not given that removing all the non-kaon particles would not modify the figure of merit, because of reactions induced by the interactions with the horn conductor. In

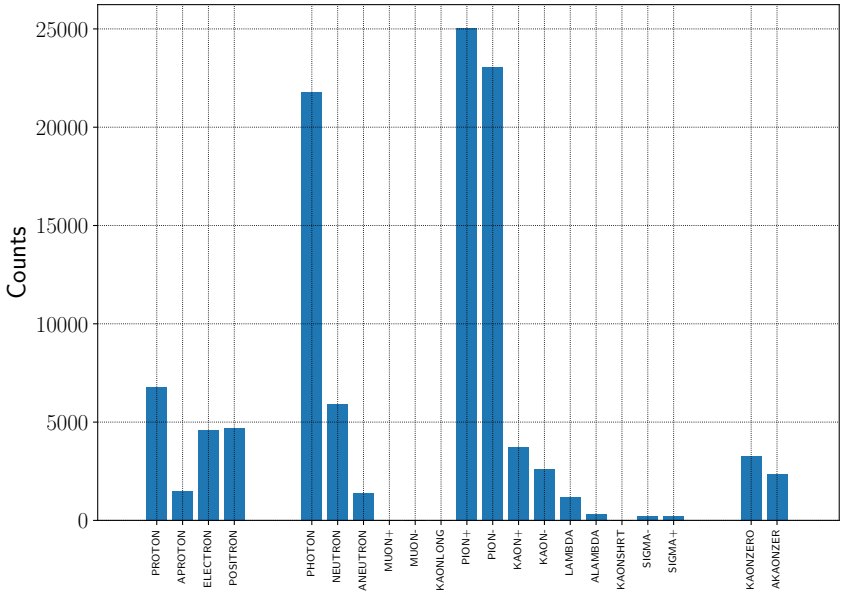

Figure 5.12: Secondary particle types produced for a number of protons of target of $10'000$, and with the previous momentum cut applied to the data (more exotic particle types in very low number are not shown).

fact, a slight decrease in the value of the figure of merit is expected. Comparing the selected K^+ distributions for the two cases shows that the difference is small, as it can be observed in Fig. 5.13. In particular, by computing the figures of merit for the two cases, the two values are off only by a 1.5% (about two times the statistical error, and so still compatible), which is acceptable given the gain in simulation times.

Given these results, each horn configuration in the optimization is run with both the previous cuts applied, and with a number of protons on target of $100'000$, which corresponds to an equivalent statistical precision of about 2% on the number of selected K^+. Under such conditions, the time taken by a full GEANT4 simulation is $\sim$ 5 minutes. The population size for the genetic algorithm is chosen to be of 500 individuals. Considering that convergence is reached in about $\mathcal{O}(100)$ iterations, and that the batch system can be subjected to queue waiting times, the elapsed time for a full optimization becomes 1-to-4 days.

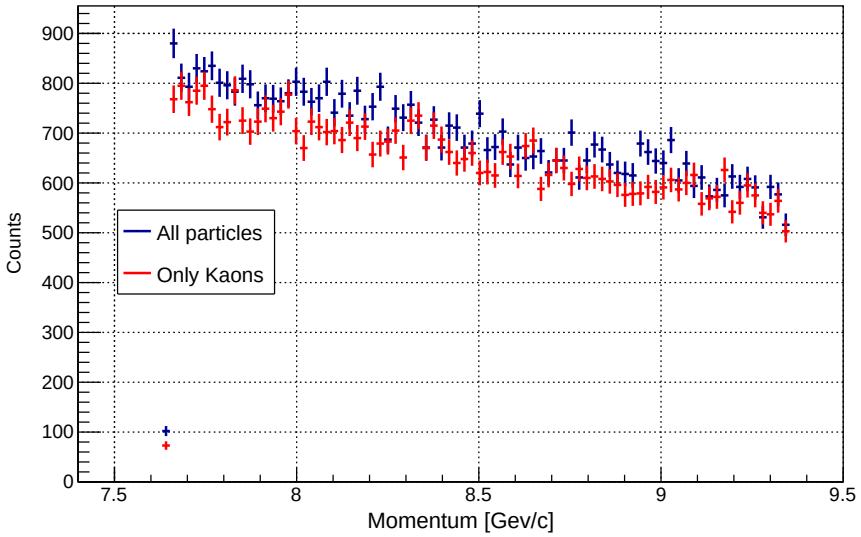

Figure 5.13: Comparison of the momentum distribution of the selected K^+ with the particle cut option (red) or without (blue).

5.4 Optimization results

5.4.1 Horn optimization for MiniBooNE-like geometry

The first test of horn geometry optimization which has been performed is for the MiniBooNE-like horn parametrization of Fig. 5.3. It has been observed that a convergence is reached after about 150 iterations, with a population of 500 horns. By looking at the evolution of the average figure of merit (used as a convergence indicator) reported in Fig. 5.14, it can be seen that its final value overtakes a factor 3. This already confirms that the optimization framework performs better than the manual scan shown before, which reached a figure of merit of a factor 2.

To better understand the results of the optimization, it is useful to observe the evolution of the best horn (in terms of figure of merit) during each iteration. Figure 5.15 shows that the maximum reached figure of merit is about a factor 3.2, with 310 kA of current and $\lesssim 3$ m of length.

It is good to observe that the current does not converge to the set maximum value of 350 kA, despite higher values than the final one have been explored by the algorithm. The same can be seen for the neck radius R_0 (see Fig. 5.3) which does not converge to the lower bound of 1.5 cm. In fact, as mentioned in Subsection 5.1.1, both the current and the neck radius significantly contribute to determine if a particular horn design could be made operational: extreme values should be avoided, especially for the case

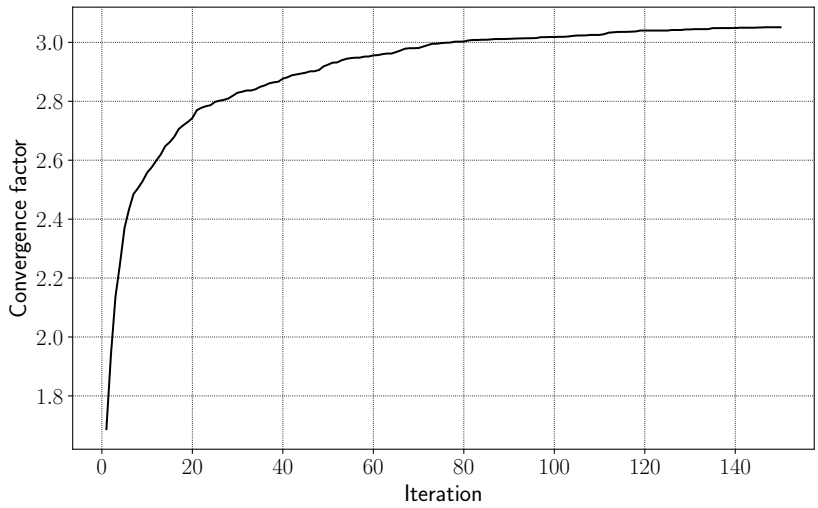

Figure 5.14: Evolution of the average of the figure of merits of the best 50 individuals.

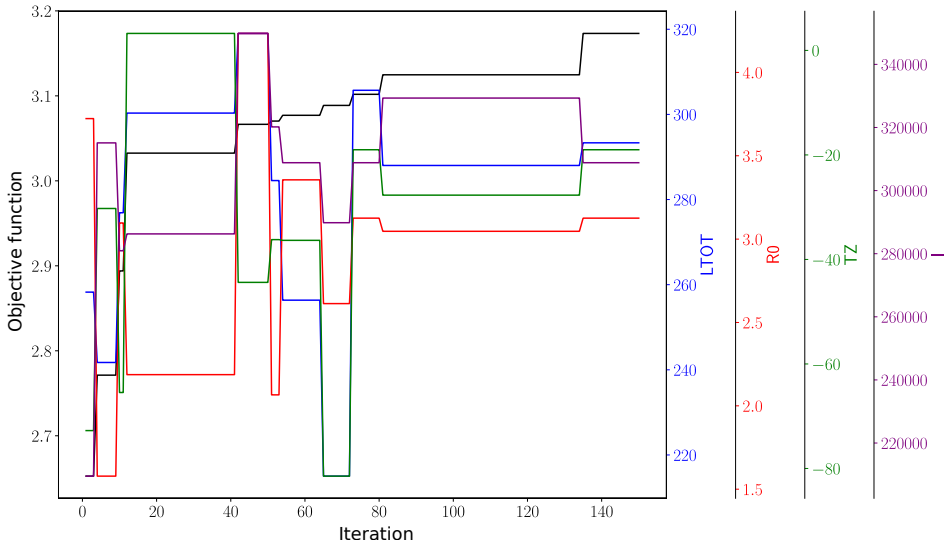

Figure 5.15: Evolution of the most significant parameters of the best horn for every iteration, where the distances are in centimeters and the current in Ampere. In particular, LTOT is the total longitudinal length, R0 the neck radius, TZ the target shift, and I the current.

of ENUBET. The only possible issue could lay in the final value of ΔZ (referred as TZ in Fig. 5.15), which implies that the target lies deep inside the horn. This can be better observed from Fig. 5.16. While a similar solution

has been found also in [110], and used in [11] and in several other instances in the past, it is worth it, in order to minimize operational failures, to try to keep the target as external as possible without disrupting too much the horn performance.

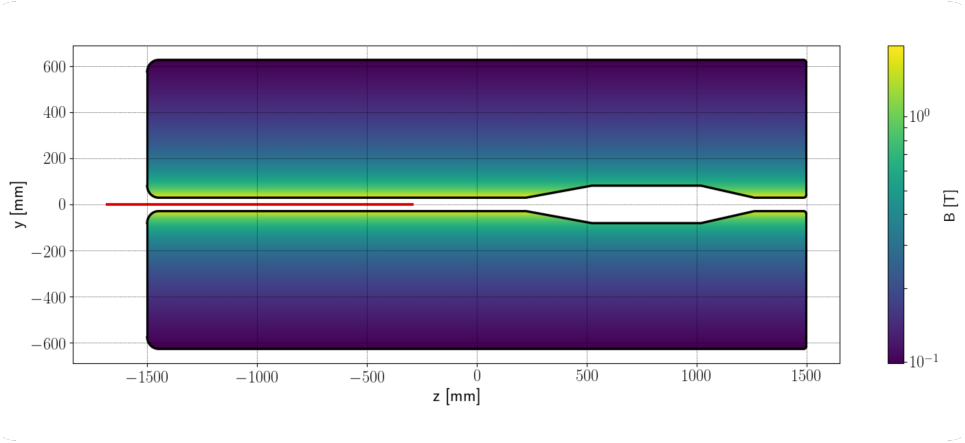

Figure 5.16: Cross-sectional view of the horn with the highest figure of merit found by the optimization algorithm. The color scale shows the magnetic field values (in Tesla).

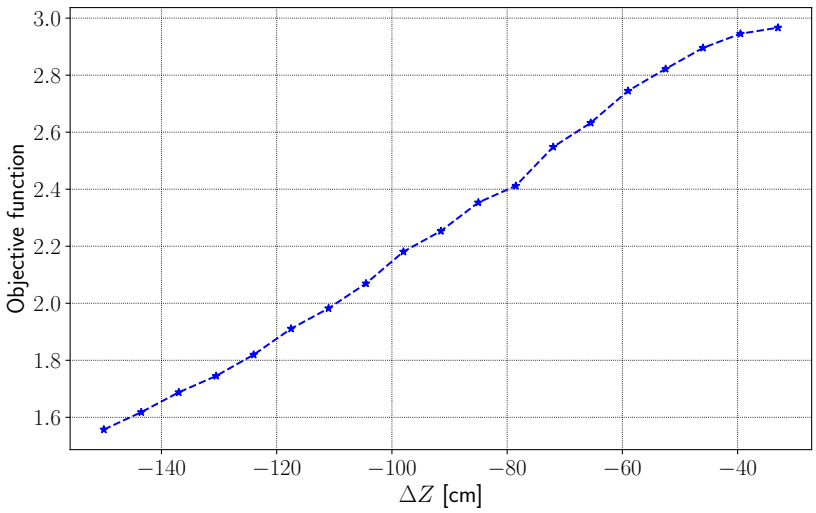

Figure 5.17: Value of the figure of merit as a function of the target shift parameter ΔZ, performed for the best horn geometry of Fig. 5.16.

Sliding the target out the horn while keeping the remaining parameters fixed has a strong effect on the figure of merit, which decreases from $\simeq 3.2$ to $\simeq 1.6$, as shown in Fig. 5.17. It can be observed that the behavior is fairly

linear up to reaching a saturation when the target is almost fully contained in the horn.

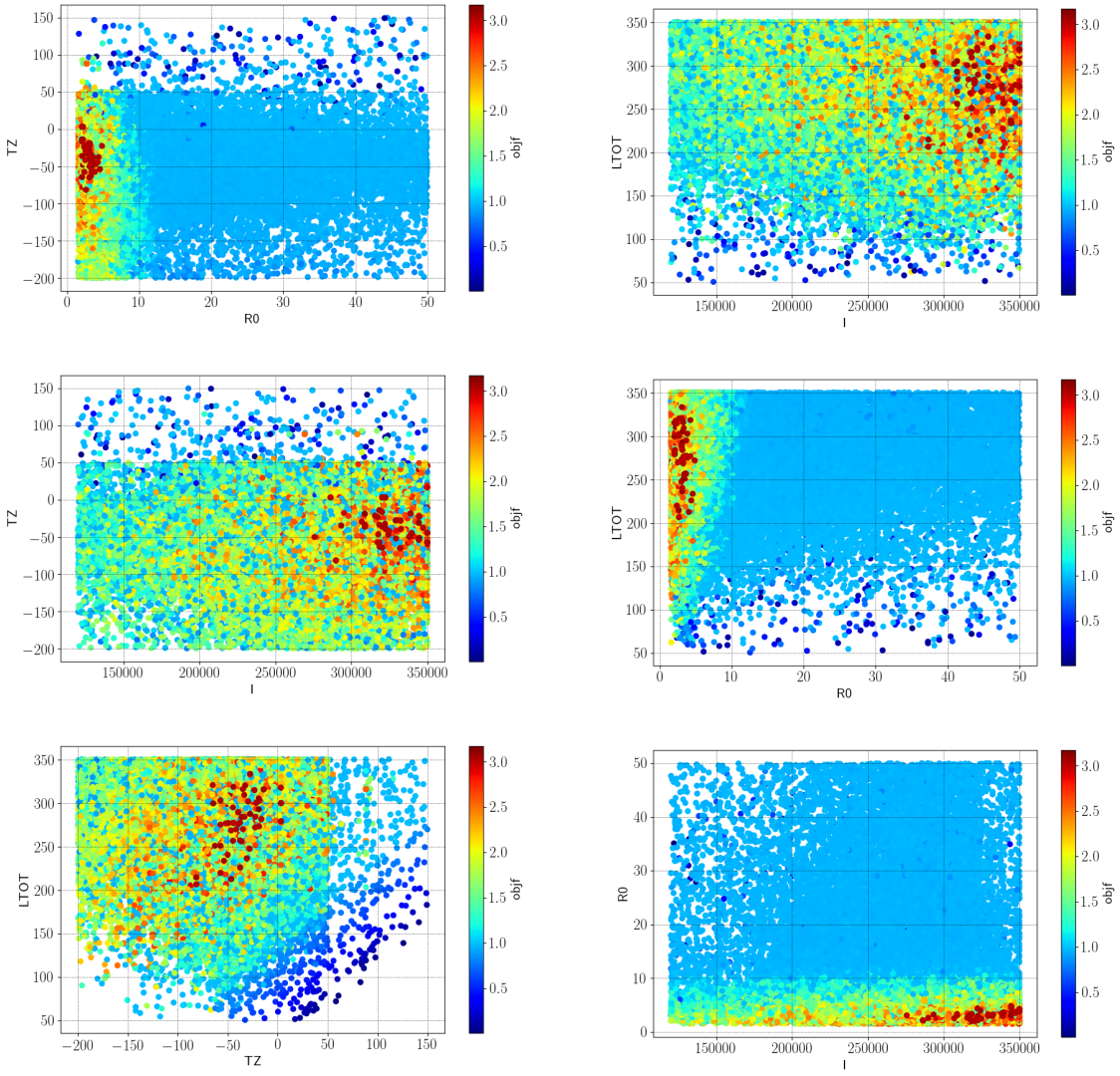

Figure 5.18: Two dimensional scatter plots of some of the optimization variables. The color scale corresponds to the objective function value.

In fact, the value of ΔZ and the other reported parameters are significantly influencing the figure of merit, as shown in the scatter plots of Fig. 5.18. The highest values of figure of merit concentrate only at specific values of these parameters, instead of being evenly spread in the space (as it is the case for most of the other not reported parameters).

The collected optimization data is very useful to investigate possible trade-off solutions which comply to different constraints. One example could be to find a horn with the target completely external. Instead of modifying the parameters of the best individual, as done in Fig. 5.17, it is more convenient to look into the parameter space for the best candidate satisfying the constraint. For instance, from Fig. 5.19, it follows that some configurations with a fully external target (N.B. the target length is 140 cm) and a figure of merit between 2.5 and 3 have been found.

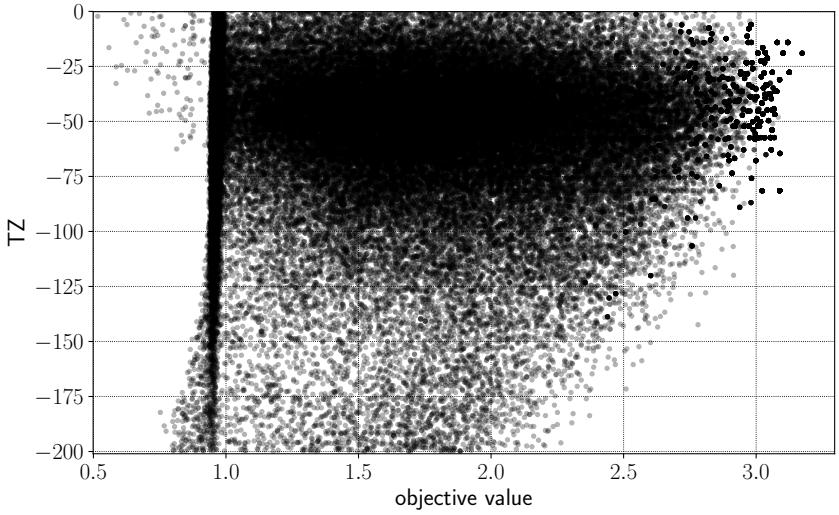

Figure 5.19: Target shift (in centimeters) plotted against the corresponding value of the figure of merit.

Given that the horn found from the initial ENUBET campaign, and shown in Fig. 5.6, also had an external target, but a figure of merit of 2, Fig. 5.19 confirms again that the automatic optimization framework yields better results than the previous manual scan. The best candidate found from Fig. 5.19 has a figure of merit of $\sim$ 2.8, and its cross-section can be observed in Fig. 5.20. With respect to the best horn reported in Fig. 5.16, it is significantly more compact (2 m of length and 30 cm of radius), but presents a higher value of the electric current.

If an external target would have to be chosen as an option, this constraint could be set directly into the genetic algorithm in order to to run a dedicated optimization. This would allow to improve the results even further. The same could be done for any other constraint, once the hardware requirements are more clear.

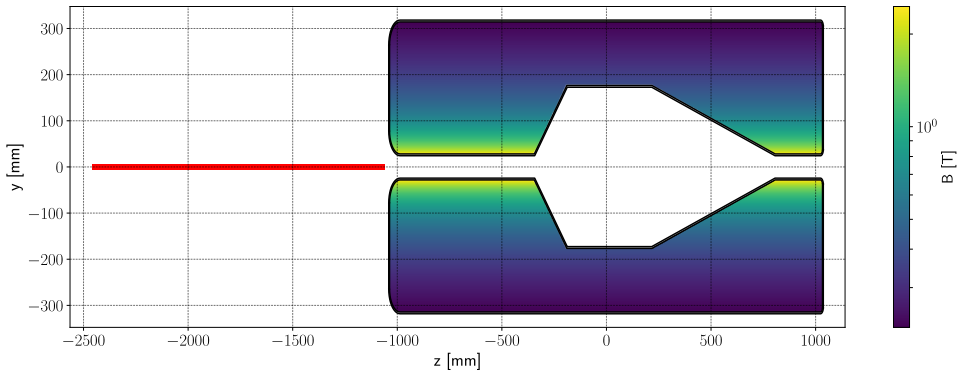

Figure 5.20: Best horn candidate found imposing a constraint of an external target on the logged optimization data.

5.4.2 Horn optimization for double-parabolic geometry

One advantage of the double-parabolic geometry with respect to the previous one, is that it offers an intermediate solution to the problem of the target inside the horn. In particular, it would be possible to make the target never enter the neck-section of the horn, while still being inside the parabolic side, as it can also be seen from the schematic of Fig. 5.4. This constraint is inserted into the double-parabolic optimization run, in addition to the previous ones.

The evolution of the best parameters reported in Fig. 5.21 shows that the final horn is 2.6 m long, i.e. about 40 cm shorter than the best case for the previous geometry. The current is slightly higher, but still it is good to notice that it did not converge to the absolute maximum of 350 kA. One issue with this result is that the final value of **R1** (see Fig. 5.3) is about equal to the one of the neck (**R0**), which implies that the optimization converged to a horn not really different from the one of Fig. 5.16, as it can be seen in Fig. 5.22. Even the figure of merit take approximately the same value of 3.2.

The result of Fig. 5.22 hints at the fact that a MiniBooNE-like geometry with the target free to enter the horn leads probably to the maximum figure of merit for the case of 8.5 GeV particles. This can be justified by the very high magnetic field values which are reached in the neck area. However, solutions of the type of Fig. 5.22 could easily be discarded by imposing a more strict lower limit for the values of **R1**. Figure 5.23 shows that it is possible to achieve high figures of merit also for big values of the opening radius of the double-parabolic horn.

By requesting an opening radius larger than 50 cm from Fig. 5.23, the horn candidate shown in Fig. 5.24 is found to be best performing. The figure

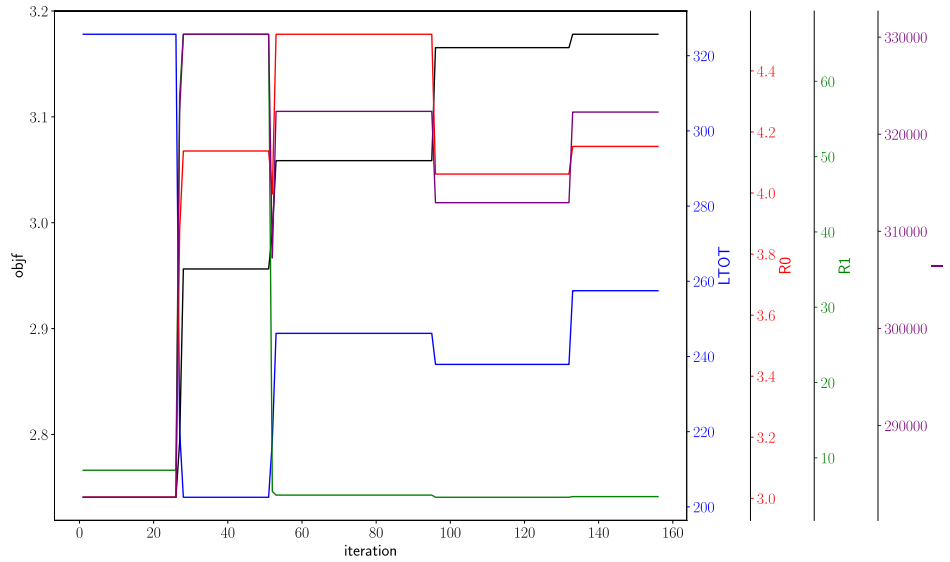

Figure 5.21: Evolution of the main parameters of the best horn for every iteration, where the lengths are in units of cm and the current in Ampere.

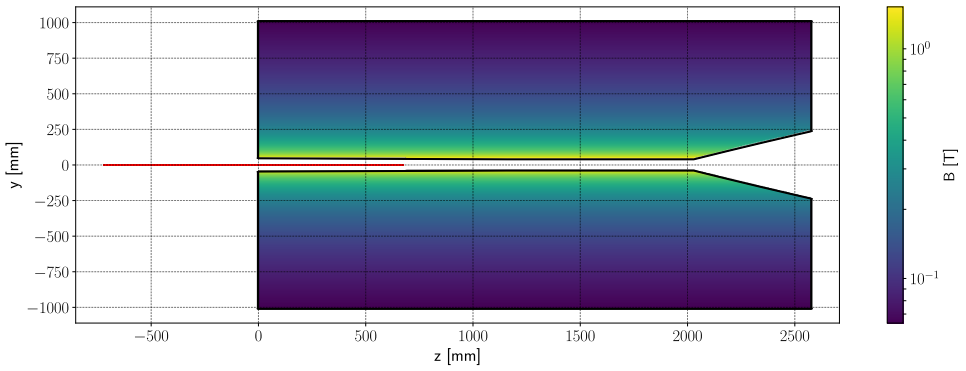

Figure 5.22: Cross-sectional view of the best horn obtained from the double-parabolic geometry optimization.

of merit of this horn is $\simeq 3$, which makes it a good trade-off solution. More in detail, it has a higher figure of merit than the other trade-off solution shown in Fig. 5.20, and also a smaller value of the current. It can also be observed that the horn has a conic-like aperture, given by the added **m1** parameter shown in Fig. 5.4. Concerning the large value of the transverse external radius, it can be proven that it is not a dominant parameter for the figure of merit, and it could eventually be reduced without significantly affecting the performance.

In order to attempt to improve this result even further, a dedicated op-

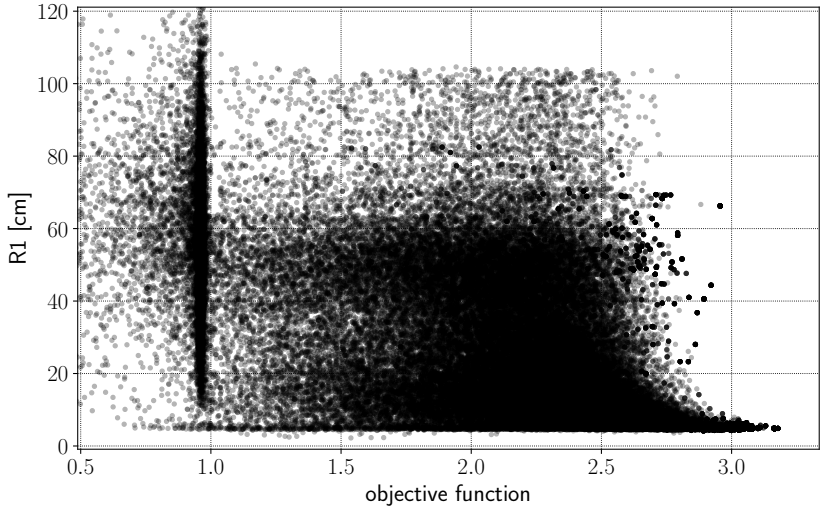

Figure 5.23: Opening radius of the double-parabolic horn against the corresponding figure of merit.

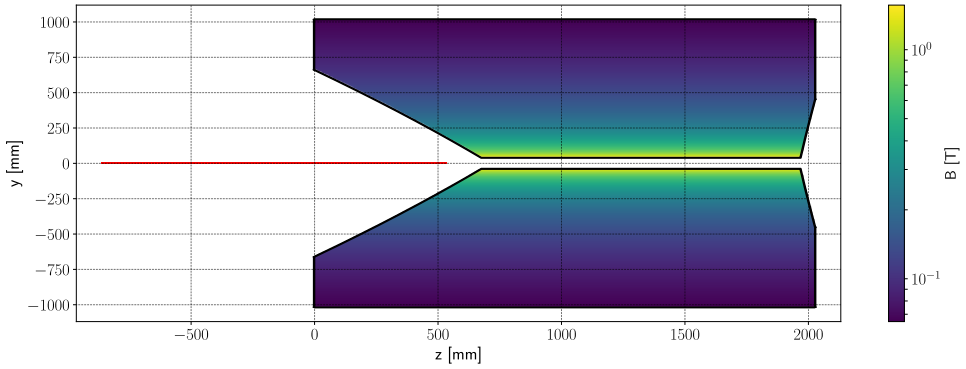

Figure 5.24: Cross-sectional view of the best double-parabolic horn obtained imposing an opening radius higher than 50 cm.

timization run could be launched by enforcing a realistic minimum value of the opening radius, once it would have been decided according to mechanical investigations.

5.4.3 Full beamline tracking

In order to have a complete picture of the achievable flux gain, different horn candidates have been used to generate the input secondary particle flux for the current version of the static beamline of ENUBET. The obtained results

in terms of kaon fluxes are shown in Table 5.1, while a comparison between the gain in K^+ at the decay tunnel and the horn figure of merit is reported in Fig. 5.1.

$K^+[10^{-3}/\text{POT}]$	**Total**	**Momentum bite**
Static beamline	0.21	0.19
Horn 0	0.11	0.09
Horn 1	0.35	0.3
Horn 2	0.16	0.14
Horn 3	0.32	0.3
Horn 4	0.27	0.23

Table 5.1: Numbers of K^+ at the entrance of the decay tunnel for the case of possible horn candidates inserted on the nominal static beamline. The horn candidate number 0 is the one from the past manual optimization. All the others are obtained from the genetic algorithm optimizations. Numbers 1 and 2 are respectively the best MiniBooNE-type horn and a trade-off with external target; similarly, 3 and 4 are the best parabolic horn and a trade-off.

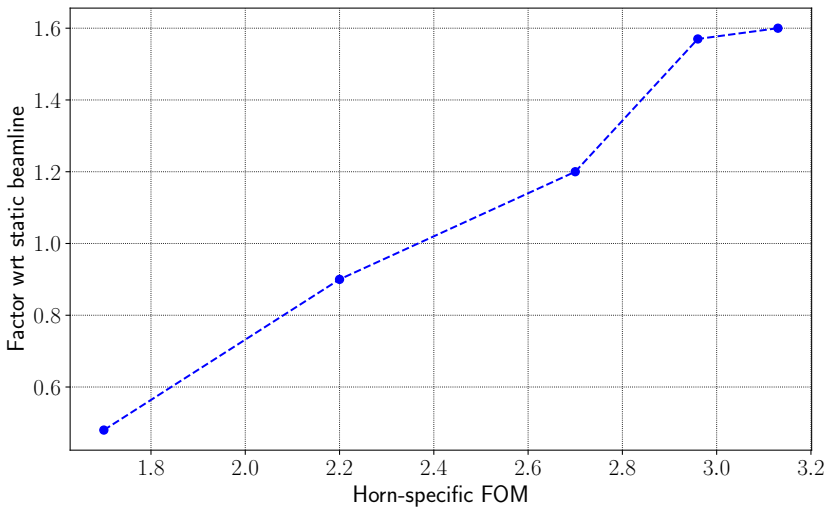

Figure 5.25: Fraction of momentum bite K^+ with respect to the static beamline as a function of the horn-specific figure of merit used in the optimization.

The comparison shows that the figure of merit obtained from the horn optimization is significantly reduced when tracking the focused secondaries along the full beamline. This does not come as a surprise: the current version of the static beamline of ENUBET is designed to have an aggressive

collimation, and elements semi-apertures never bigger than 15 cm. When looking at the results from the first study, discussed in Section 5.3.1, the opposite result was found: a value of 2 on the horn figure of merit was leading to ~ 5 on the full beamline. The reason of this is because the static beamline at the time was significantly different than the current one: only one dipole was present, the aperture of the elements was bigger, and less collimation was employed. These latter conditions allow to better exploit the potential of a horn, which would give its maximum gain for a wide band beam. In fact, the main effect of the horn is to reduce the average divergence of the beam, but at the cost of increasing is transverse dimension.

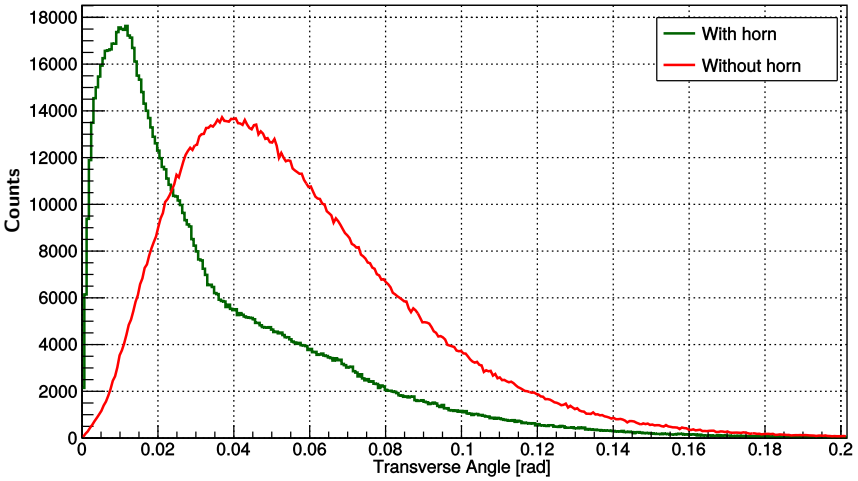

Figure 5.26: Distribution of the transverse angles of K^+ in the momentum bite at the first quadrupole, with and without magnetic horn.

The distributions of the transverse angle and transverse radius of the K^+ focused by a horn can be observed respectively in Figs. 5.26 and 5.27, compared with the bare-target case. In particular, the transverse angles are reduced on average by more than a factor two, while about the opposite happens to the average transverse radius. Given that the horn-specific figure of merit takes into account the aperture and angular acceptance of a quadrupole, a net gain of a factor $\gtrsim 3$ is still potentially achievable, even with the maximum aperture and angular acceptance of ENUBET. It is evident from Fig. 5.27 that the reason why the full tracking in the current static beamline produces smaller gains is related to the increased dimension of the beam, which is shaved off by the many collimators and elements. In order to prove the significant dependence of the flux gain on the beamline, some tests have been performed by increasing the aperture of some collimators to the one of the quadrupoles. The different tested configurations and relative results are listed in Table 5.2. The horn-specific figure of merit of the em-

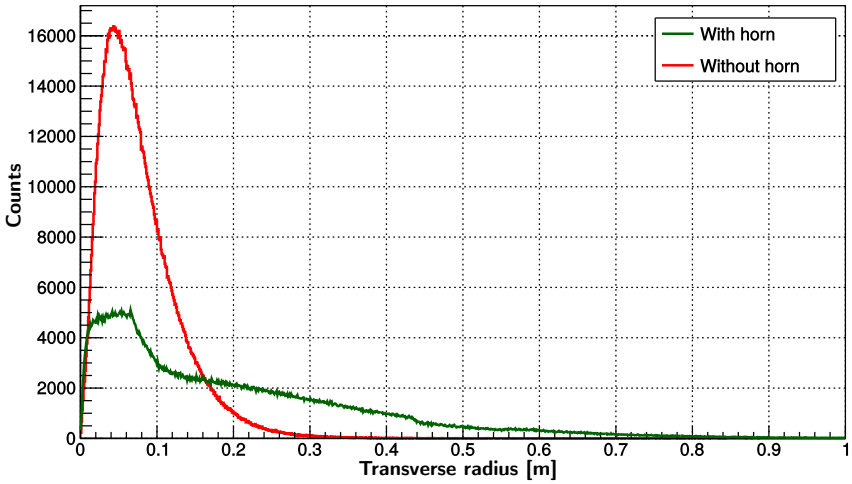

Figure 5.27: Distribution of the transverse radius of K^+ in the momentum bite at the first quadrupole, with and without magnetic horn.

ployed horn configuration was $\simeq 3.2$, which seems to be kept constant for the case in which a strict collimation is enforced only at the first quadrupole triplet. This is in fact in agreement with the definition of the horn-specific figure of merit. The maximum achievable gain, obtained by increasing all the collimator apertures, reaches about a factor 7, proving that the horn performance is severely influenced by the beamline design.

	K^+ gain at tunnel
Nominal beamline	1.6
Collimators opened to quad. apertures	6.8
All collimators opened but last one	4.7
All collimators opened but triplet ones	3.3

Table 5.2: Gain in the number of K^+ at the entrance of the decay tunnel for different collimation configurations.

The issue with all the modifications reported in Table 5.2 is that they substantially increase the secondary particles hitting the instrumented decay tunnel walls, hence increasing the background and worsening the discrimination of K_{e3} events. This confirms that a dedicated investigation on the best beamline configuration for the case of a horn should be undertaken.

5.5 Conclusions and next steps

The main goal of this work has been reached: a full optimization framework for the magnetic horn has been developed and successfully tested, leading to the first results on potential kaon flux gain. A figure of merit defined on the focusing acceptance of a nominal ENUBET quadrupole has been used in order to evaluate the performance of the horn without a full beamline tracking. The values of this figure of merit reached a maximum gain factor of $\gtrsim 3$, which hints at a possible satisfactory increase of flux. A strong dependence on the beamline has been noticed by comparing results from different beamline configurations. However, the effect of the beamline on the horn-specific figure of merit seems to be monotonic (e.g. Fig. 5.17), and so preserving the optima found by the optimization. The current version of the static beamline significantly reduces the horn gain found in the optimization to a maximum of a factor 1.6, given the aggressive collimation.

The results on the full beamline tracking have been obtained in the simplest approach, i.e. by keeping the beamline identical to the static one, but adding the horn at the beginning. This is not an ideal solution in order to fully exploit the potential of a magnetic horn, and specific solutions will be investigated in the future.

The first step is to match the beamline optics (and eventually the elements positions) to the horn. In fact, the horn significantly modifies the phase space presentation at the first beamline element (as shown in Fig. 5.28): this should be accompanied by a dedicated optics optimization phase.

Assuming that the effect of the beamline is order-preserving, the optics matching could be applied on the phase space generated by the best horn from the optimization framework, still guaranteeing that it is a maximum for the figure of merit. On top of this, the positions of the beamline elements could also be used as parameters to further optimize the beamline for the horn. To this end, additional information about the particle distribution in phase space at the downstream end of the horn could also be used to better steer the optimization. This could either be added as a set of constraints or as new figures of merit. In such a way it would be possible to have more control on the optimization in terms of the consequent optics matching of the beamline, and eventually it would change a single-objective optimization problem to a multi-objective one.

The problem could further be expanded by designing a completely different beamline dedicated to the horn-focusing option. While this task would take more time than just performing the steps described above, it could allow to better exploit the horn potential, so to reach the highest possible gain.

Once a satisfactory gain in flux would have been reached, it will be necessary to assess the systematics in the K_{e3} events identifications brought by the use of a magnetic horn.

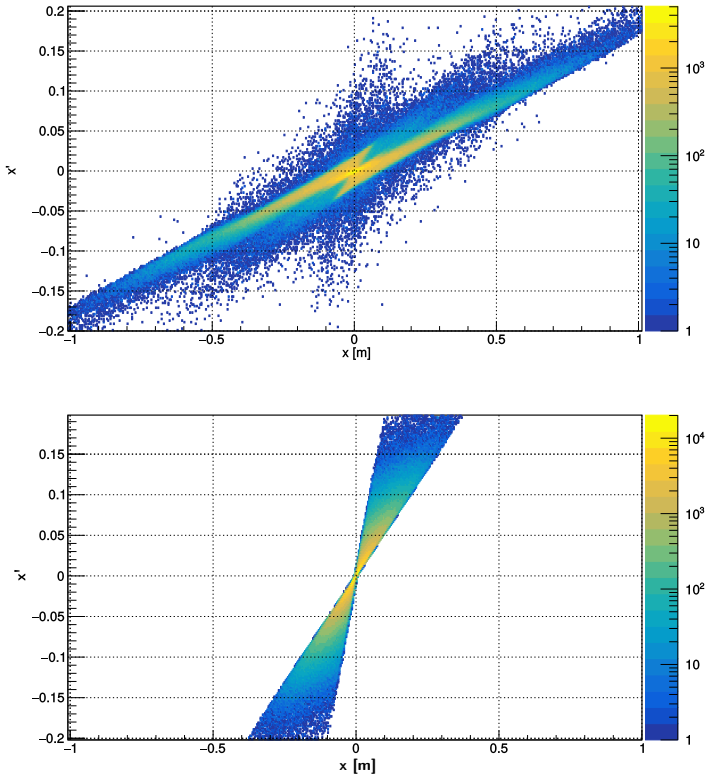

Figure 5.28: Horizontal phase space of the K^+ in the momentum bite at the first quadrupole. Top: with the horn. Bottom: without the horn.

Despite having already constrained the parameters to be compatible with engineering requirements, a more refined study on the hardware-side aspects of a horn for the ENUBET experiment will be undertaken.

Chapter 6

Conclusions and outlook

The ENUBET project plans to significantly improve the precision on the neutrino flux estimation for future accelerator neutrino beams. To do so, the typically passive decay tunnel will be fully instrumented, with the goal of tagging the K_{e3}, $K_{\mu 2}$, and $K_{\mu 3}$ decays of positive kaons.

Instrumenting the decay tunnel poses stringent constraints on the secondary particle flux, in order to avoid pile-up at the detectors. This makes it impractical to use the fast extraction scheme of the primary protons, typically employed in neutrino beams, and requires the use of the slow resonant extraction. A standard slow extraction with an intensity of $\mathcal{O}(10^{13})$ protons extracted over an interval of a few second would allow ENUBET to work under a critical pile-up threshold. Anyway, such a long extracted spill would force the use of static elements for the beamline. Typically, in neutrino beamlines, a magnetic horn is used to increase the secondary particle fluxes, thanks to its superior acceptance with respect to standard optical elements.

This work is based on the idea of the ENUBET project to modify the standard slow extraction scheme to a pulsed one, in order to still be able to use a magnetic horn, while maintaining a below-threshold operation of the experiment. This would have the significant advantage of reducing the time needed for ENUBET to perform a high precision neutrino cross-section measurement.

The pulsed version of the slow extraction, which has been called burst mode slow extraction, has been designed, simulated, and tested at CERN-SPS. This type of extraction has been implemented as extension of the third-integer resonant extraction of the SPS, acting on the horizontal tune function to control the temporal structure of the extracted spill. This led to the extraction of the full intensity in pulses of demanded temporal length (some tens of milliseconds) and repeated at a regular frequency, distributed along the 4.8 s 400 GeV flat-top. The main characteristics and performance of the burst extraction have been studied in simulation and tested in the

machine, allowing to identify its critial points. In particular, the finite time-to-extraction of particles in phase space and its distribution are making it difficult to control exactly the temporal length of each extracted burst. A solution, based on an iterative approach, has been proposed and proven successful by machine tests. This allowed to produce a burst length of 10 ms, reaching the first goal of the ENUBET project.

A dedicated simulation of the novel extraction method has been used to first validate the experimental results, and then to improve them toward the goal of covering the full range requested by ENUBET, i.e. 2 to 10 ms of burst length. In particular, two possible solutions have been found to further reduce the burst lengths down to 3-4 ms. The first is to increase the extraction sextupole strength, while the second is to use a so called amplitude extraction, by reducing the chromaticity to very small values. These two techniques should be selected depending on the target burst length. The simulated improvements will also be tested in the future at the SPS to fully validate the results. Other operational improvements, as upgrading the tools to control and optimize this type of extraction have been presented.

It was shown that the main timing-related problematics observed in the newly developed extraction method are directly related to a more general problem: the frequency response to magnet ripples of the slow extraction. That is, how any current signal on the accelerator magnets (here focusing on the quadrupoles) is transferred on the extracted spill, affecting its quality (e.g. strong presence of 50 Hz harmonics). Given the importance of this topic for both experiments which use the nominal slow extraction, and for those that will use the burst extraction, a dedicated study was carried out. The existing MADX simulation model of the nominal SPS slow extraction has been adapted to characterize the transfer function of the process. The problem has been found to be equivalent to a linear system for small enough ripple amplitudes (and frequencies), while becoming non linear above a certain threshold. Using the simulation, an effective semi-analytic transfer map has been constructed. Both the full MADX and the semi-analytic map have been successfully validated with dedicated measurements at the SPS. A simple model of the process based on a Henon-map has been developed, and its results proved to be in good agreement with the MADX model. Given the reduced computational cost of this model, it has been used to investigate possible alternative slow extraction configurations for ripple reduction. In particular, a significant dependence on both the sextupole strength and chromaticity could allow a factor two reduction of the transfer function height and pole. This dependence has been modeled, showing a connection with the results of the burst mode slow extraction. These potential improvements found in simulation will be deployed operationally on the SPS after the CERN Long Shutdown 2.

The last part of this work has been dedicated directly to the design of a magnetic horn for ENUBET, with the goal to fully profit of the burst slow extraction advances. A standalone GEANT4 model and an automatic optimization framework, based on a genetic algorithm, have been developed and tested. The goal of starting a systematic optimization study of the horn geometry, so to investigate the maximum reachable flux gains, has been achieved. First optimization runs have proven the flexibility and speed of the process, and shown that different trade-off solutions are possible. The genetic algorithm has proved capable of selecting improved horn configurations with respect to the previously considered ones. It has also been observed how the particular beamline used has the potential of significantly affecting the gain in kaon flux. This has made clear that the next steps on the horn optimization will be dedicated to the beamline. Both the optics and the arrangement of elements can be used to further improve the gain in kaons provided by the horn, so to develop a horn-based beamline option for ENUBET. The practical constraints related to the mechanical, thermal, and electrical aspects of the horn operated with the burst extraction will be also further studied in the future.

Appendix A

Low-frequency spill expressions for betatron core and COSE schemes.

In Section 2.5, an expression for the spill signal for the tune sweep extraction has been developed, under the assumptions of low-frequency and ideal momentum extraction. The result (Eq. 2.27) is reported here for brevity:

$$\frac{dN}{dt} = K\big(Q\big)\,\frac{dQ}{dt} \tag{A.1}$$

where $K(Q)$ is the particle tune distribution. It will be shown in the following how similar expressions can be obtained also for the case of the betatron extraction and COSE, highlighting the differences between the two.

As discussed in Section 2.5, in a betatron core extraction, the resonance is fixed, and the full momentum stack is accelerated (or decelerated) into it. In formula:

$$\delta_p(t) = \frac{p(t) - p_0}{p_0} \tag{A.2}$$

where $p(t)$ is the momentum of the particle, and p_0 is the fixed design momentum. In particular, $p(t)$ is the only time depend variable, which is changed by the DC voltage of the betatron core magnet as:

$$p(t) = p^* + \beta_p(t) \tag{A.3}$$

where p^* is the initial particle momentum, and β_p is the additional momentum increase from the betatron core electric force. With respect to the QSWEEP case, both the momentum distribution (Fig. 2.20) and the tune distribution (Fig. 2.16) of the particles are moving into the resonance, since both the particle tune and the particle momentum (Eq. A.2) are time dependent. Since the time dependence of the particle distribution is just a rigid

shift, and the distribution shape remains unchanged, the problem could be mapped to a static momentum distribution using the initial particle momentum δ_p^* as the new variable, defined as:

$$\delta_p(t) = \delta_p^* + \frac{\beta_p(t)}{p_0} \tag{A.4}$$

and so, following Eq. 2.25 for the QSWEEP case:

$$\begin{aligned} dN &= P\big(\delta_p^{*,R}\big)\, d\delta_p^{*,R} \\ &= P\big(\delta_p^{*,R}(t)\big)\, \frac{\partial \delta_p^{*,R}(t)}{\partial t}\, dt \\ &= -\frac{P\Big(\frac{Q_R - Q}{\xi} - \frac{\beta_p(t)}{p_0}\Big)}{p_0}\, \frac{\partial \beta_p(t)}{\partial t}\, dt \end{aligned} \tag{A.5}$$

where $\delta_p^{*,R}(t)$ is the value of the initial momentum of a particle currently on resonance, defined as:

$$\delta_p^{*,R}(t) = \frac{Q_R - Q}{\xi} - \frac{\beta_p(t)}{p_0} \tag{A.6}$$

and Q is the (now constant) machine tune. By defining a particle distribution as a function of the betatron momentum increase

$$D(\beta_p) := -\frac{P\Big(\frac{Q_R - Q}{\xi} - \frac{\beta_p}{p_0}\Big)}{p_0} \tag{A.7}$$

the spill signal can be expressed as

$$\frac{dN}{dt} = D\big(\beta_p\big)\, \frac{d\beta_p}{dt} \tag{A.8}$$

where this time the spill signal is determined by the sweep velocity of the betatron core momentum increase, in the same style ad Eq. A.1.

For the COSE method, it has been shown how the momentum distribution of the particles is shifted into resonance by acting on the design momentum of the machine as (Eq. 2.29):

$$\delta_p(t) = \frac{p - p_0(t)}{p_0(t)} \tag{A.9}$$

The sweeping of the design momentum can be defined as a function of the initial design momentum p_0^* and a fractional momentum increase ε:

$$p_0(t) = p_0^* \big(1 + \varepsilon(t)\big) \tag{A.10}$$

where ε is the only time dependent variable. Using Eq. A.10, the particle momentum (Eq. A.9) can be re-written in terms of the time independent variable δ_p^*, defined as the particle momentum relative to p_0^*:

$$\delta_p(t) = \frac{\delta_p^*}{1+\varepsilon(t)} - \frac{\varepsilon(t)}{1+\varepsilon(t)} \tag{A.11}$$

Differently from the case of Eq. A.4, Eq. A.11 shows that in this case the time dependence of the momentum distribution is not only a translation of the average value, but also a shrinking of the distribution. While it is possible to remove the average shift, it is not possible in general to remove the shrinking, which introduces a time dependence in the distribution P.

Following Eq. A.5:

$$\begin{aligned} dN &= P_\varepsilon\big(\delta_p^{*,R}\big)\, d\delta_p^{*,R} \\ &= P_{\varepsilon(t)}\big(\delta_p^{*,R}(t)\big)\, \frac{\partial \delta_p^{*,R}(t)}{\partial t}\, dt \\ &= \left(\frac{Q_R - Q}{\xi} + 1\right) P_{\varepsilon(t)}\left(\frac{(1+\varepsilon(t))(Q_R-Q)}{\xi} + \varepsilon(t)\right) \frac{\partial \varepsilon(t)}{\partial t}\, dt \end{aligned} \tag{A.12}$$

where $\delta_p^{*,R}(t)$ is

$$\delta_p^{*,R}(t) = \frac{(1+\varepsilon(t))(Q_R-Q)}{\xi} + \varepsilon(t) \tag{A.13}$$

and Q is the (now constant) machine tune. By defining a particle distribution as a function of the design momentum increase parameter ε:

$$E(\varepsilon) := P_\varepsilon\left(\frac{(1+\varepsilon)(Q_R-Q)}{\xi} + \varepsilon\right) \tag{A.14}$$

the spill signal can be expressed as:

$$\frac{dN}{dt} = E\big(\varepsilon\big)\, \frac{d\varepsilon}{dt} \tag{A.15}$$

which shows that the spill signal is determined by the sweep velocity of the design momentum parameter. In particular, in the usual case in which ε is small, Eq. A.11 could be expanded into:

$$\delta_p(t) \simeq \delta_p - \varepsilon(t) + \mathcal{O}(\delta_p^2) \tag{A.16}$$

and the time dependence on the momentum distribution could be dropped, leading to a solution analogous to the betatron core case.

Using the SPS parameters of Table 2.2 in Eq. A.12, it can be noticed that a positive increase of ε is required to drive the particles into resonance, which is actually the case in operation (Fig. 2.21).

Appendix B

Possible limitations of the frequency transfer model

It is important to be aware that in the real case there can be effects not included in the developed simulation model of Chapter 4. The vacuum chamber shielding is one of these (also depicted in Fig. 4.1). Once the ripples reach a high enough frequency, the eddy currents in the vacuum chamber inside the magnet are partially screening the field, acting as another low pass filter. This effect could be visible when comparing simulated and experimental data: for the present case, anyway, no evidence of such effect has been found (up to 200 Hz for the dedicated measurements case). This result is in agreement with the predictions for the SPS focusing quadrupoles vacuum chambers, for which the transfer function cut-off frequency of the shielding effect is foreseen to be at about 4 kHz. For different accelerators the vacuum chamber shielding could play an important role in ripple attenuation, as for example in the MedAustron case of Fig. 4.39, a shielding at the kHz level would help in the suppression of the otherwise undisturbed high frequency ripples. The same considerations apply for the magnet losses, which again are expected to be negligible at the frequencies of focus.

Another difference is that all the developed simulations have used the assumption of a linear tune ramp and a uniform momentum distribution. This is almost never the case in practice: in the SPS, the momentum distribution takes a shape comparable to two partially overlapped Gaussians, and as a consequence the tune ramp is adjusted in order to extract a flat spill.

Figure B.1 shows a typical extraction tune ramp at the SPS during the 4.8 s flat top. The only non-negligible slope variations are at the beginning and end of the spill (where the momentum distribution assumes low values): in the procedure of selecting a fiducial time range at the center of the spill in order to cut out the rising and falling edges these two parts get mostly discarded too. A possible complication is that the ramp profile can variate from a day to another, given the fact that the spill is frequently re-optimized

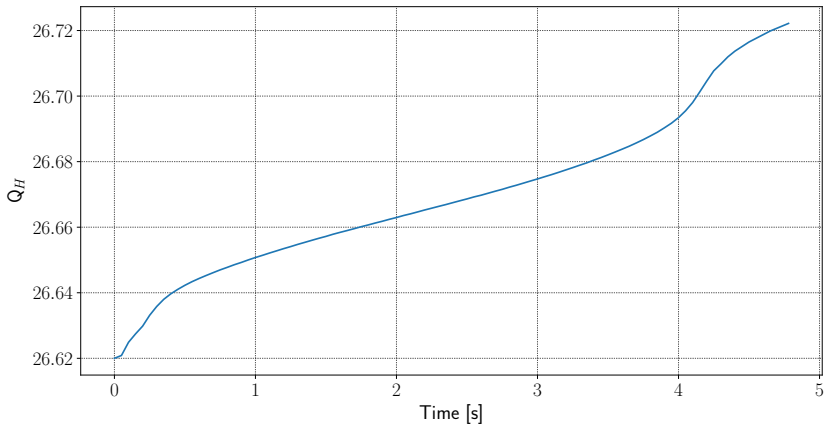

Figure B.1: Typical horizontal tune ramp at flat top at the SPS..

due to supercycle changes. Anyway, the overall good agreement of the simulation model with the data from the dedicated measurements showed no sign of a systematic difference given by such effect (while the uncertainties on the operational data still make it more complicated to confirm such a result). This represents a good news, since introducing the experimental momentum distribution into the simulation model and adjusting correctly the tune ramp would complicate its implementation. It is anyway interesting to quickly show how the problem would change by not neglecting these effects. As done in the previous section, the instantaneous extraction spill of Eq. 4.4 is taken as the starting point. The latter equation is also assuming a uniform momentum distribution and a linear tune ramp. If the momentum distribution is changed to a non-uniform one, the tune ramp $Q(t)$ leading to a flat spill will have to solve the following differential equation:

$$K(Q)\,\frac{dQ}{dt} = \frac{N_{\mathrm{tot}}}{\Delta T} \tag{B.1}$$

where K is the particle density distribution in the tune space, Q is the tune ramp, N_{tot} the total number of particles, and ΔT the flat top duration. For a non-uniform K, Eq. B.1 can almost never be solved analytically. It can be noticed that in the case in which dQ/dt is a constant (i.e. linear ramp), this model would predict the extracted spill to take the shape of the momentum distribution. This latter fact has been verified by measurements at the SPS to be true in a good approximation, hinting at the fact that this simple model is a good approximation for the overall spill shape prediction (i.e. low frequency region). In the operational case, the solution of Eq. B.1 is reached iteratively by the Autospill control-room application, as explained in Section 2.5. Assuming the solution of Eq. B.1 to be found, and calling it $Q_0(t)$, the foreseen spill signal in the presence of a sinusoidal ripple $r(t)$ will

be:

$$\begin{aligned} \text{Spill}(t) &= K(Q_0 + r) \times \frac{d(Q_0 + r)}{dt} \\ &\simeq K(Q_0) \frac{dQ_0}{dt} + K(Q_0 + r) \frac{dr}{dt} + \left.\frac{dK}{dQ}\right|_{Q_0} \frac{dQ_0}{dt} r + \mathcal{O}(r) \end{aligned} \tag{B.2}$$

where in the last passage K has been expanded at first order in r. The first term at the last passage of Eq. B.2 is by definition a flat spill, the second term is an expansion of what was previously the derivator behavior (i.e. zero of the transfer function), and the last term is proportional to the ripple itself. In particular, the spectrum of the second term is now the convolution of the ripple derivative spectrum and the tune density one. By performing a quick numeric computation, it is possible to verify that for the case of the SPS the dominant behavior is still that of a derivator, and any other complication added in Eq. B.2 is rather small. Anyway, Eq. B.2 is only valid for a ripple in the linear regime; on the opposite case the rectifier condition of Eq. 4.4 would need to be added. With a non-linear tune ramp, the ripple could become non-linear only in some parts of the ramp, complicating the situation. In an extreme cases as such, or also when the non-linear effects of Eq. B.2 would stop being negligible, switching formalism to the one Eq. 4.7 (where C is now $K(Q)$) could be a convenient choice.

Acknowledgements

This thesis is the product of three very busy years, which stress-tested my limits and pushed me out of my "comfort-zone" a significant amount of times (mostly, for the best). The final thesis document and the annexed research are just one of the products of all of this, along with many precious experiences and personal developments. As the latter would make for an acknowledgement section of their own, but seem to me too personal to be included here, the present acknowledgements will pertain mainly to the research-related aspects of this experience, which are already enough to fill this page (and part of the next).

There are several people who made this journey fruitful and precious, and I will try to thank them all. I will start from my supervisors, Andrea Longhin and Francesco Velotti, for both of whom I've been the first PhD student. Andrea always involved me in stimulating projects, giving me space to explore a variety of different topics, always guided by the will of understanding and improving everything that is possible. The guidance he gave me by exchanging questions, suggestions, and ideas has been fundamental for this work. I also had the pleasure to share with him many other experiences, such as conferences, data taking, shifts, dinners, drinks, etc. He has always been a great guide and company. Francesco allowed me to discover CERN and accelerator physics, and, in particular, the very interesting world of slow extraction. His guidance in this thesis has been crucial, not only inspiring me with his knowledge in the numerous discussions we had, but also managing the tricky task of keeping me focused and on-track on the actual goals of the work. He also gave me the opportunity to participate in many activities in one of the largest accelerators around the globe, for which I will always be thankful for. A special thanks goes to Gianmaria Collazuol, who is the one that first introduced me to the fascinating world of experimental physics (with the focus on *experimental*), and has been a mentor ever since. I am very grateful for all the opportunities he granted me and the time shared together: courses, laboratory experiences, discussions, projects, trips, elaborated plans, overly optimistic task-lists and time-schedules, shifts, beers, kebabs and many more. Continuing along my research path of these years, I have to thank many people I had the pleasure to work with both at CERN and in the ENUBET project. From CERN, many thanks go to B. Goddard,

V. Kain, M. Fraser, W. Bartmann, all the colleagues from BTP, the SPS operators, and all the other people I had the occasion to share some of my time with. From ENUBET, I thank the whole collaboration for the continuous inspiration I received from their work and dedication; in particular, I also thank the people I had the pleasure to work more closely with, for their support and help: F. Terranova, F. Pupilli, G. Brunetti, V. Mascagna and E. Lutsenko. Many thanks go also to S. Peggs and T. Satogata, for getting me up to speed on the field of accelerator physics with a beautiful course, and in the record time of two weeks. Stepping away from the mere topics of the thesis, I also have to thank the many people that helped me in other projects I have been involved in these years: M. Cicerchia, J. Walker, M. Lamoureux, L. Esposito and many more. Thank you also to all the people I had the pleasure to work or share some of my time with in the many experiences related to this PhD that I forgot to mention up to now, please feel included. Finally, any PhD is entangled in bureaucracy, and without some help sometimes it could be difficult to escape it alive: many thanks to everyone who helped with this spiny subject, both from Padova and CERN.

An invaluable support and warmth has come from a vast amount of personal experiences and the people in my life. Even if I have kept these acknowledgements focused on the work, one thing I have to say: thank you so much, you are the ones that kept me going and made everything worth it.

References

[1] C. Giunti and C. W. Kim. *Fundamentals of Neutrino Physics and Astrophysics*. Oxford University Press, Apr. 2007. DOI: 10.1093/acprof:oso/9780198508717.001.0001.

[2] F. Suekane. *Neutrino Oscillations. A Practical Guide to Basics and Applications*. 1st ed. Lecture Notes in Physics. Springer, 2014. DOI: 10.1007/978-4-431-55462-2.

[3] Tanabashi et al. "Review of Particle Physics". In: *Phys. Rev. D* 98 (3 2018), p. 030001. DOI: 10.1103/PhysRevD.98.030001.

[4] S. Bilenky. *Introduction to the Physics of Massive and Mixed Neutrinos*. 1st ed. Lecture Notes in Physics. Springer-Verlag, 2010. DOI: 10.1007/978-3-642-14043-3.

[5] Q.R. Ahmad et al. "Direct evidence for neutrino flavor transformation from neutral current interactions in the Sudbury Neutrino Observatory". In: *Phys. Rev. Lett.* 89 (2002), p. 011301. DOI: 10.1103/PhysRevLett.89.011301. arXiv: nucl-ex/0204008.

[6] SNO Collaboration. "The Sudbury Neutrino Observatory". In: *Nucl. Instrum. Methods A* 449 (2000), pp. 172–207.

[7] M. D. Schwartz. "Quantum Field Theory and the Standard Model". In: Cambridge University Press, 2013. Chap. 10. ISBN: 9781107034730.

[8] K. Abe et al. "Constraint on the matter-antimatter symmetry-violating phase in neutrino oscillations". In: *Nature* 580 (7803 Apr. 2020), pp. 339–344. DOI: 10.1038/s41586-020-2177-0.

[9] S. E. Kopp. "Accelerator neutrino beams". In: *Physics Reports* 439.3 (2007), pp. 101 –159. ISSN: 0370-1573. DOI: https://doi.org/10.1016/j.physrep.2006.11.004.

[10] K. Abe et al. "The T2K experiment". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 659.1 (2011), pp. 106 –135. ISSN: 0168-9002. DOI: https://doi.org/10.1016/j.nima.2011.06.067.

[11] P. Adamson et al. "The NuMI neutrino beam". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 806 (2016), pp. 279 –306. ISSN: 0168-9002. DOI: https://doi.org/10.1016/j.nima.2015.08.063.

[12] K. Elsener et al. *The CERN neutrino beam to Gran Sasso (NGS): conceptual technical design*. CERN Yellow Reports: Monographs. Geneva: CERN, 1998. DOI: 10.5170/CERN-1998-002. URL: http://cds.cern.ch/record/359028.

[13] P. Berge et al. "Total neutrino and antineutrino charged current cross section measurements in 100, 160, and 200 GeV narrow band beams". In: *Zeitschrift für Physik C Particles and Fields* 35 (4 Dec. 1987), pp. 443–452. DOI: 10.1007/BF01596895.

[14] A. Blondel et al. "Electroweak parameters from a high statistics neutrino nucleon scattering experiment". In: *Zeitschrift für Physik C Particles and Fields* 45 (3 Sept. 1990), pp. 361–379. DOI: 10.1007/BF01549665.

[15] A. Grant and J. M. Maugain. *High intensity version of the narrowband beam N3*. Tech. rep. CERN-EF-BEAM-83-2. Geneva: CERN, 1983. URL: https://cds.cern.ch/record/150346.

[16] L. Aliaga et al. "Neutrino flux predictions for the NuMI beam". In: *Phys. Rev. D* 94 (9 2016), p. 092005. DOI: 10.1103/PhysRevD.94.092005.

[17] MINERνA Collaboration. "Neutrino Flux Predictions for the NuMI Beam". In: *Phys. Rev. D* 94 (2016).

[18] M. Bishai. "Determining the Neutrino Flux from Accelerator Neutrino Beams". In: *Nucl. Phys. B* Proc. Suppl. (2012), pp. 210–214.

[19] M. Day and K. S. McFarland. "Differences in Quasielastic Cross Section of Muon and Electron Neutrinos". In: *Phys. Rev. D* 86 (2012).

[20] J. A. Formaggio and G. P. Zeller. "From eV to EeV: Neutrino cross sections across energy scales". In: *Rev. Mod. Phys.* 84.3 (2012).

[21] L. Alvarez-Ruso, Y. Hayato, and J. Nieves. "Progress and open questions in the physics of neutrino cross sections at intermediate energies". In: *New Journal of Physics* 16.7 (2014), p. 075015. DOI: 10.1088/1367-2630/16/7/075015.

[22] X. Bu. "Measurement of Electron Neutrino Charged-Current Inclusive Cross Section in $1-3$ GeV energy region with the NOνA Near Detector". In: (2015). eprint: arXiv:hep-ex/1601.01213.

[23] A. Longhin, F. Terranova, et al. *NP06/ENUBET annual report for the CERN-SPSC*. Tech. rep. CERN-SPSC-2020-009. SPSC-SR-268. Geneva: CERN, 2020. URL: https://cds.cern.ch/record/2714046.

[24] F. Acerbi et al. "A high precision neutrino beam for a new generation of short baseline experiments". In: (Jan. 2019). Input document for the European Particle Physics Strategy. Dec. 2018. arXiv: 1901.04768 [physics.ins-det].

[25] F. Acerbi et al. *The ENUBET project*. Tech. rep. CERN-SPSC-2018-034. SPSC-I-248. Geneva: CERN, 2018. URL: https://cds.cern.ch/record/2645532.

[26] A. Berra et al. *Enabling precise measurements of flux in accelerator neutrino beams: the ENUBET project*. Tech. rep. CERN-SPSC-2016-036. SPSC-EOI-014. Geneva: CERN, 2016. URL: https://cds.cern.ch/record/2221984.

[27] A. Longhin, L. Ludovici, and F. Terranova. "A Novel Technique for the Measurement of the Electron Neutrino Cross Section". In: *The European Physical Journal C* 75 (4 Apr. 2015). DOI: 10.1140/epjc/s10052-015-3378-9.

[28] G. Ballerini et al. "Testbeam performance of a shashlik calorimeter with fine-grained longitudinal segmentation". In: *Journal of Instrumentation* 13.01 (2018), P01028–P01028. DOI: 10.1088/1748-0221/13/01/p01028.

[29] F. Acerbi et al. "Irradiation and performance of RGB-HD Silicon Photomultipliers for calorimetric applications". In: *Journal of Instrumentation* 14.02 (2019), P02029–P02029. DOI: 10.1088/1748-0221/14/02/p02029.

[30] F. Acerbi et al. "Polysiloxane-based scintillators for shashlik calorimeters". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 956 (2020), p. 163379. ISSN: 0168-9002. DOI: https://doi.org/10.1016/j.nima.2019.163379.

[31] F. Acerbi et al. "The ENUBET positron tagger prototype: construction and testbeam performance". In: *Journal of Instrumentation* 15.08 (2020), P08001–P08001. DOI: 10.1088/1748-0221/15/08/p08001.

[32] A. M. Ankowski and C. Mariani. "Systematic uncertainties in long-baseline neutrino-oscillation experiments". In: *Journal of Physics G: Nuclear and Particle Physics* 44.5 (2017), p. 054001. DOI: 10.1088/1361-6471/aa61b2.

[33] F. Pupilli et al. "The ENUBET narrow band neutrino beam". In: *PoS* NOW2018 (2018). Ed. by Antonio Marrone, Alessandro Mirizzi, and Daniele Montanino, p. 030. DOI: 10.22323/1.337.0030.

[34] S. Peggs and T. Satogata. *Introduction to Accelerator Dynamics*. Cambridge: Cambridge University Press, 2017. DOI: 10.1017/9781316459300.

[35] H. Wiedemann. *Particle Accelerator Physics*. 4th ed. Springer International Publishing, 2015. DOI: 10.1007/978-3-319-18317-6.

[36] *CAS - CERN Accelerator School : 5th General Accelerator Physics Course: Jyväskylä, Finland 7 - 18 Sep 1992. CAS - CERN Accelerator School : 5th General Accelerator Physics Course*. 2 volumes, consecutive pagination. CERN. Geneva: CERN, 1994. DOI: 10.5170/CERN-1994-001. URL: https://cds.cern.ch/record/235242.

[37] D. A. Edwards and M. J. Syphers. *An Introduction to the Physics of High Energy Accelerators*. John Wiley & Sons, Ltd, 2008. ISBN: 9783527617272. DOI: 10.1002/9783527617272.

[38] A. W. Chao, K. H. Mess, M. Tigner, and F. Zimmermann. *Handbook of Accelerator Physics and Engineering*. 2nd ed. Singapore: World Scientific, 2013. DOI: 10.1142/8543.

[39] S. Y. Lee. *Accelerator Physics*. 3rd ed. Singapore: World Scientific, 2011. DOI: 10.1142/8335.

[40] Y. Kohayashi and H. Takahashi. "Improvement Of The Emittance In The Resonant Beam Ejection". In: *6th International Conference on High-Energy Accelerators*. 1967, pp. 347–351.

[41] Y. Kobayashi. "Theory of the resonant beam ejection from synchrotrons". In: *Nuclear Instruments and Methods* 83.1 (1970), pp. 77 –87. ISSN: 0029-554X. DOI: https://doi.org/10.1016/0029-554X(70)90537-9.

[42] S. G. Peggs. "Hamiltonian Theory Of The E778 Nonlinear Dynamics Experiment". In: *2nd ICFA Advanced Beam Dynamics Workshop*. Apr. 1988.

[43] S. G. Peggs and R. M. Talman. "Nonlinear Problems in Accelerator Physics". In: *Ann. Rev. Nucl. Part. Sci.* 36 (1986), pp. 287–325. DOI: 10.1146/annurev.ns.36.120186.001443.

[44] L. Badano et al. *Proton-Ion Medical Machine Study (PIMMS), 1*. 1999. URL: http://cds.cern.ch/record/385378.

[45] M. Benedikt. "Optical design of a synchrotron with optimisation of the slow extraction for hadron therapy". PhD thesis. Vienna, Tech. U., 1997.

[46] W. Hardt. *Ultraslow extraction out of LEAR*. Tech. rep. CERN-PS-DL-LEAR-Note-81-6. Geneva: CERN, 1981. URL: https://cds.cern.ch/record/1025914.

[47] N. Charitonidis and I. Efthymiopoulos. "Low energy tertiary beam line design for the CERN neutrino platform project". In: *Phys. Rev. Accel. Beams* 20 (11 2017), p. 111001. DOI: 10.1103/PhysRevAccelBeams.20.111001.

[48] M.S. Rosenthal et al. "Commissioning Results of the Tertiary Beam Lines for the CERN Neutrino Platform Project". In: *Proc. 10th International Particle Accelerator Conference (IPAC'19), Melbourne, Australia, 19-24 May 2019*. (Melbourne, Australia). International Particle Accelerator Conference 10. Geneva, Switzerland: JACoW Publishing, 2019, pp. 3738–3741. ISBN: 978-3-95450-208-0. DOI: doi:10.18429/JACoW-IPAC2019-THPGW064.

[49] B. Abi et al. "First results on ProtoDUNE-SP liquid argon time projection chamber performance from a beam test at the CERN Neutrino Platform". In: (July 2020). arXiv: 2007.06722 [physics.ins-det].

[50] C. Giganti, T. Lux, and M. Yokoyama. *NP07: ND280 Upgrade project*. Tech. rep. CERN-SPSC-2020-008. SPSC-SR-267. Geneva: CERN, 2020. URL: https://cds.cern.ch/record/2713578.

[51] F.M. Velotti et al. "Characterisation of the SPS Slow-extraction Parameters". In: *Proc. of International Particle Accelerator Conference (IPAC'16), Busan, Korea, May 8-13, 2016*. (Busan, Korea). International Particle Accelerator Conference 7. JACoW, 2016, pp. 3918–3921. DOI: doi:10.18429/JACoW-IPAC2016-THPOR055.

[52] F. M. Velotti. "Higher brightness beams from the SPS for the HL-LHC era". PhD thesis. Lausanne: Ecole Polytechnique Federale del Lausanne, 2017, p. 162. DOI: 10.5075/epfl-thesis-7498. URL: http://infoscience.epfl.ch/record/226174.

[53] P. Forck, P. Kowina, R. Singh, and M. Wendt. "Beam Measurements Using Schottky Signal Analysis". presented at IBIC2017 in Grand Rapids, MI, USA, unpublished. Geneva, Switzerland, 2018.

[54] M. Tomizawa et al. "Slow extraction from the J-PARC main ring using a dynamic bump". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 902 (2018), pp. 51 –61. ISSN: 0168-9002. DOI: https://doi.org/10.1016/j.nima.2018.06.004. URL: http://www.sciencedirect.com/science/article/pii/S0168900218307174.

[55] L.S. Stoel et al. "Progress Toward a Dynamic Extraction Bump for Slow Extraction in the CERN SPS". In: *Proc. 9th International Particle Accelerator Conference (IPAC'18), Vancouver, BC, Canada, April 29-May 4, 2018*. (Vancouver, BC, Canada). International Particle Accelerator Conference 9. JACoW Publishing, 2018, pp. 838–841. DOI: doi:10.18429/JACoW-IPAC2018-TUPAF055.

[56] D. Jacquet, R. Gorbonosov, G. Kruk, and P.P. Mira. "LSA - The high level applicaion software of the LHC and its performance during the first 3 years of operation". In: *Proc. 14th International Conference on Accelerator and Large Experimental Physics Control Systems, (ICALEPCS'14), San Francisco, CA, USA, 2013.* (San Francisco, California, USA). International Particle Accelerator Conference 10. JACoW Publishing, 2013, pp. 1201–1204. DOI: doi:10.18429/JACoW-IPAC2019-WEPMP035.

[57] V. Kain et al. "SPS Slow Extracted Spill Quality During the 2016 Run". In: *Proc. of International Particle Accelerator Conference (IPAC'17), Copenhagen, Denmark, 14-19 May, 2017.* (Copenhagen, Denmark). International Particle Accelerator Conference 8. JACoW, 2017, pp. 627–630. DOI: https://doi.org/10.18429/JACoW-IPAC2017-MOPIK049.

[58] F. M. Velotti et al. "Observations of SPS Slow-Extracted Spill Quality Degradation and Possible Improvements". In: *Proc. 9th International Particle Accelerator Conference (IPAC'18), Vancouver, BC, Canada, April 29-May 4, 2018.* (Vancouver, BC, Canada). International Particle Accelerator Conference 9. JACoW Publishing, 2018, pp. 761–764. DOI: doi:10.18429/JACoW-IPAC2018-TUPAF035.

[59] F.M. Velotti et al. "Characterisation of SPS Slow Extraction Spill Quality Degradation". In: *Proc. 10th International Particle Accelerator Conference (IPAC'19), Melbourne, Australia, 19-24 May 2019.* (Melbourne, Australia). International Particle Accelerator Conference 10. JACoW Publishing, 2019, pp. 2403–2405. DOI: doi:10.18429/JACoW-IPAC2019-WEPMP034.

[60] V. Kain, K. Cornelis, and E. Effinger. "New Spill Control for the Slow Extraction in the Multi-Cycling SPS". In: *Proc. of International Particle Accelerator Conference (IPAC'16), Busan, Korea, May 8-13, 2016.* (Busan, Korea). International Particle Accelerator Conference 7. JACoW, 2016, pp. 1371–1373. DOI: doi:10.18429/JACoW-IPAC2016-TUPMR051.

[61] V. Kain et al. "Resonant slow extraction with constant optics for improved separatrix control at the extraction septum". In: *Phys. Rev. Accel. Beams* 22 (10 2019), p. 101001. DOI: 10.1103/PhysRevAccelBeams.22.101001.

[62] L. Badano and S. Rossi. *Characteristics of a betatron core for extraction in a proton-ion medical synchrotron.* Tech. rep. CERN-PS-97-019-DI. Geneva: CERN, 1997. URL: http://cds.cern.ch/record/327305.

[63] L. Falbo, E. Bressi, S. Foglio, and C. Priano. "Betatron Core Slow Extraction at CNAO". In: *Proc. 9th International Particle Accelerator Conference (IPAC'18), Vancouver, BC, Canada, April 29-May 4, 2018*. (Vancouver, BC, Canada). International Particle Accelerator Conference 9. JACoW Publishing, 2018, pp. 1237–1239. DOI: doi:10.18429/JACoW-IPAC2018-TUZGBF3.

[64] F. M. Velotti et al. "Septum shadowing by means of a bent crystal to reduce slow extraction beam loss". In: *Phys. Rev. Accel. Beams* 22 (9 2019), p. 093502. DOI: 10.1103/PhysRevAccelBeams.22.093502.

[65] M. A. Fraser et al. "Demonstration of slow extraction loss reduction with the application of octupoles at the CERN Super Proton Synchrotron". In: *Phys. Rev. Accel. Beams* 22 (12 2019), p. 123501. DOI: 10.1103/PhysRevAccelBeams.22.123501.

[66] L.S. Stoel et al. "Phase Space Folding Studies for Beam Loss Reduction During Resonant Slow Extraction at the CERN SPS". In: *Proc. of International Particle Accelerator Conference (IPAC'17), Copenhagen, Denmark, 14-19 May, 2017*. (Copenhagen, Denmark). International Particle Accelerator Conference 8. JACoW, 2017, pp. 615–618. DOI: https://doi.org/10.18429/JACoW-IPAC2017-MOPIK046.

[67] L.S. Stoel, M. Benedikt, K.A. Brown, M.A. Fraser, and B. Goddard. "Investigating Beam Loss Reduction with Octupoles During Slow Extraction in the CERN SPS". In: *Proc. 9th International Particle Accelerator Conference (IPAC'18), Vancouver, BC, Canada, April 29-May 4, 2018*. (Vancouver, BC, Canada). International Particle Accelerator Conference 9. JACoW Publishing, 2018, pp. 822–825. DOI: doi:10.18429/JACoW-IPAC2018-TUPAF051.

[68] V. Nagaslaev, K. A. Brown, and M. Tomizawa. "Third integer resonance extraction with presence of higher multipoles". In: *Phys. Rev. Accel. Beams* 22 (4 2019), p. 043501. DOI: 10.1103/PhysRevAccelBeams.22.043501.

[69] Y. Baconnier, P. Faugeras, K. H. Kissler, B. de Raad, and W. Scandale. "Extraction from the CERN SPS". In: *IEEE Transactions on Nuclear Science* 24.3 (1977), pp. 1434–1436.

[70] K. H. Kissler, J. Riche, W. Scandale, and G. Schroder. "Fast Resonant Extraction from the CERN SPS". In: *IEEE Transactions on Nuclear Science* 26.3 (1979), pp. 3228–3230.

[71] M. Gyr and E. B. Vossenberg. "Half integer fast resonant extraction with quasirectangular spill". In: *Conf. Proc. C* 920324 (1992), pp. 1501–1503.

[72] M. Gyr, K. H. Kissler, J. D. Pahud, and S. Peraire. "Simultaneous Slow Resonant Extractions from the SPS with Horizontal Tune-Split". In: *IEEE Transactions on Nuclear Science* 32.5 (1985), pp. 2997–2999.

[73] M. Gyr. *Proposal for a new servo-spill system: power requirements for different configurations*. Tech. rep. CERN-SL-95-103-BT. Geneva: CERN, 1995. URL: http://cds.cern.ch/record/292415.

[74] W. Sliwinski. *CMW/JAPC*. Workshop on QPS Software Layer, August 2015, CERN, Geneva, Switzerland. URL: https://indico.cern.ch/event/404646/contributions/962341/.

[75] W. Sliwinski. *CMW & JMS: Status Report & Future Plans*. BE-CO Technical Meeting, 15 March 2018, CERN, Geneva, Switzerland. URL: https://indico.cern.ch/event/704903/contributions/2901282/.

[76] S. Deghaye and E. Fortescue-Beck. *Introduction to the BE/CO control system*. CERN. 2016 Edition - Part 1. CERN, Geneva, Switzerland.

[77] K. Cornelis. *Momentum cleaning in the LHC with resonant Extraction*. Tech. rep. LHC-PROJECT-NOTE-176. Geneva: CERN, 1998. URL: http://cds.cern.ch/record/692007.

[78] H. Weisberg. *Effective Spill Length Monitor*. Tech. rep. AGS Division Technical Note 163. Upton, New York: Brookhaven National Laboratory, 1980.

[79] H. Grote, G. Roy, F. Schmidt and L. Deniau (editor). *The MAD-X Program - User's reference Manual*. CERN. Geneva, Switzerland. URL: http://madx.web.cern.ch/madx/.

[80] A. Savitzky and M. J. E. Golay. "Smoothing and Differentiation of Data by Simplified Least Squares Procedures." In: *Analytical Chemistry* 36.8 (1964), pp. 1627–1639. DOI: 10.1021/ac60214a047.

[81] L.S. Esposito et al. "Crystal for Slow Extraction Loss-Reduction of the SPS Electrostatic Septum". In: *Proc. 10th International Particle Accelerator Conference (IPAC'19), Melbourne, Australia, 19-24 May 2019*. (Melbourne, Australia). International Particle Accelerator Conference 10. JACoW Publishing, 2019, pp. 2379–2382. DOI: doi:10.18429/JACoW-IPAC2019-WEPMP028.

[82] B. Goddard et al. "Reduction of 400 GeV/c slow extraction beam loss with a wire diffuser at the CERN Super Proton Synchrotron". In: *Phys. Rev. Accel. Beams* 23 (2 2020), p. 023501. DOI: 10.1103/PhysRevAccelBeams.23.023501.

[83] M. Pivi et al. "Overview and Status of the MedAustron Ion Therapy Center Accelerator". In: *8th International Particle Accelerator Conference*. 2017, THPVA076. DOI: 10.18429/JACoW-IPAC2017-THPVA076.

[84] M. Pullia et al. "Betatron Core Driven Slow Extraction at CNAO and MedAustron". In: *7th International Particle Accelerator Conference*. 2016, TUPMR037. DOI: 10.18429/JACoW-IPAC2016-TUPMR037.

[85] R. Muto et al. "Current Status of Slow Extraction from J-PARC Main Ring". In: *J. Phys. Conf. Ser.* 1350.1 (2019), p. 012105. DOI: 10.18429/JACoW-IPAC2019-WEPMP007.

[86] C. D. Moore, D.E. Johnson, J. Johnstone, T. Kobilarcik, and C.T. Murphy. "Slow Extraction from the Fermilab Main Injector". In: *Conf. Proc. C* 0106181 (2001). Ed. by P. W. Lucas and S. Webber, pp. 1559–1561.

[87] V. Nagaslaev et al. "Third Interger Resonance Slow Extraction Scheme for a Mu->E Experiment at Fermilab". In: *46th ICFA Advanced Beam Dynamics Workshop on High-Intensity and High-Brightness Hadron Beams*. Sept. 2010. arXiv: 1207.6621 [physics.acc-ph].

[88] Brown, K. *Slow Extraction and Spill Control from BNLs AGS and Booster*. ICFA Mini-Workshop on Slow Extraction, 2019, Fermilab, Batavia, IL, USA. 2019. URL: https://indico.fnal.gov/event/20260/contributions/56661/.

[89] Ondreka, D. *Overview of Slow Extraction at GSI and FAIR*. Slow Extraction Workshop 2017, CERN, Geneva, Switzerland. 2017. URL: https://indico.cern.ch/event/639766/contributions/2750884/.

[90] C. Krantz et al. "Slow Extraction Techniques at the Marburg Ion-Beam Therapy Centre". In: *9th International Particle Accelerator Conference*. 2018, TUPAL036. DOI: 10.18429/JACoW-IPAC2018-TUPAL036.

[91] J. Wenninger. *SPS Spill Quality*. Tech. rep. AB-Note-2004-0YY OP. Geneva: CERN, 2003.

[92] V. Rödel. "La dynamique de l'extraction lente du synchrotron à protons de 400 GeV du CERN en vue d'un asservissement". PhD thesis. Lausanne: Ecole Polytechnique Federale del Lausanne, 1983, p. 154. DOI: 10.5075/epfl-thesis-506. URL: http://infoscience.epfl.ch/record/30986.

[93] M. Gyr. *Low frequency fluctuations of spill rates during slow resonant extraction*. Tech. rep. SPS Divisional Reports 89-21 (ABT). 1989. URL: http://cds.cern.ch/record/92595.

[94] J. Prieto et al. "Beam Dynamics Simulations of the Effect of Power Converter Ripple on Slow Extraction at the CERN SPS". In: *Proc. 9th International Particle Accelerator Conference (IPAC'18), Vancouver, BC, Canada, April 29-May 4, 2018*. (Vancouver, BC, Canada). International Particle Accelerator Conference 9. JACoW Publishing, 2018, pp. 818–821. DOI: doi:10.18429/JACoW-IPAC2018-TUPAF050.

[95] K. Okamura et al. "A Consideration on the Transfer Function Between RQ Field and Slow Extraction Spill in the Main Ring of J-Parc". In: *Proc. 10th International Particle Accelerator Conference (IPAC'19), Melbourne, Australia, 19-24 May 2019*. (Melbourne, Australia). International Particle Accelerator Conference 10. JACoW Publishing, 2019, pp. 2315–2317. DOI: doi:10.18429/JACoW-IPAC2019-WEPMP008.

[96] D. Naito et al. "Real-time correction of betatron tune ripples on a slowly extracted beam". In: *Phys. Rev. Accel. Beams* 22 (7 2019), p. 072802. DOI: 10.1103/PhysRevAccelBeams.22.072802.

[97] S. Sorge, P. Forck, and R. Singh. "Measurements and Simulations of the Spill Quality of Slowly Extracted Beams from the SIS-18 Synchrotron". In: *Journal of Physics: Conference Series* 1067 (2018), p. 052003. DOI: 10.1088/1742-6596/1067/5/052003.

[98] R. Singh, P. Forck, and S. Sorge. *Smoothing of the slowly extracted coasting beam from a synchrotron*. 2019. arXiv: 1904.09195 [physics.acc-ph].

[99] M. Pari et al. "Model and Measurements of CERN-SPS Slow Extraction Spill Re-Shaping - the Burst Mode Slow Extraction". In: *Proc. 10th International Particle Accelerator Conference (IPAC'19), Melbourne, Australia, 19-24 May 2019*. (Melbourne, Australia). International Particle Accelerator Conference 10. JACoW Publishing, 2019, pp. 2406–2409. DOI: doi:10.18429/JACoW-IPAC2019-WEPMP035.

[100] R. Singh, P. Forck, P. Boutachkov, S. Sorge, and H. Welker. "Slow Extraction Spill Characterization From Micro to Milli-Second Scale". In: *Journal of Physics: Conference Series* 1067 (2018), p. 072002. DOI: 10.1088/1742-6596/1067/7/072002.

[101] R. Singh, P. Forck, and S. Sorge. "Reducing Fluctuations in Slow-Extraction Beam Spill Using Transit-Time-Dependent Tune Modulation". In: *Phys. Rev. Applied* 13 (4 2020), p. 044076. DOI: 10.1103/PhysRevApplied.13.044076.

[102] J. Laskar. "The chaotic motion of the solar system: A numerical estimate of the size of the chaotic zones". In: *Icarus* 88.2 (1990), pp. 266 –291. ISSN: 0019-1035. DOI: https://doi.org/10.1016/0019-1035(90)90084-M. URL: http://www.sciencedirect.com/science/article/pii/001910359090084M.

[103] J. Laskar, C. Froeschlé, and A. Celletti. "The measure of chaos by the numerical analysis of the fundamental frequencies. Application to the standard mapping". In: *Physica D: Nonlinear Phenomena* 56.2 (1992), pp. 253 –269. ISSN: 0167-2789. DOI: https://doi.org/10.1016/0167-2789(92)90028-L. URL: http://www.sciencedirect.com/science/article/pii/016727899290028L.

[104] H. Roder. "Amplitude, Phase, and Frequency Modulation". In: *Proceedings of the Institute of Radio Engineers* 19.12 (1931), pp. 2145–2176.

[105] M. Hénon. "A two-dimensional mapping with a strange attractor". In: *Comm. Math. Phys.* 50.1 (1976), pp. 69–77.

[106] S. van der Meer. *A directive device for charged particles and its use in an enhanced neutrino beam.* CERN Yellow Reports: Monographs. Geneva: CERN, 1961. DOI: 10.5170/CERN-1961-007. URL: https://cds.cern.ch/record/278088.

[107] M. Giesch et al. "Status of magnetic horn and neutrino beam". In: *Nuclear Instruments and Methods* 20 (1963), pp. 58 –65. ISSN: 0029-554X. DOI: https://doi.org/10.1016/0029-554X(63)90391-4.

[108] A. A. Aguilar-Arevalo et al. "Neutrino flux prediction at MiniBooNE". In: *Phys. Rev. D* 79 (7 2009), p. 072002. DOI: 10.1103/PhysRevD.79.072002.

[109] S. S. Gilardoni. "Study of particle production and capture for a Neutrino Factory". PhD thesis. Geneva: Université de Genève, 2004, p. 187. URL: "https://cds.cern.ch/record/811486".

[110] A. Longhin. "A new design for the CERN-Fréjus neutrino superbeam". In: *The European Physical Journal C* 71.1745 (9 2011). DOI: 10.1140/epjc/s10052-011-1745-8.

[111] A.K. Ichikawa. "Design concept of the magnetic horn system for the T2K neutrino beam". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 690 (2012), pp. 27 –33. ISSN: 0168-9002. DOI: https://doi.org/10.1016/j.nima.2012.06.045.

[112] A. Liu, A. Bross, and D. Neuffer. "Optimization of the magnetic horn for the nuSTORM non-conventional neutrino beam using the genetic algorithm". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 794 (2015), pp. 200 –205. ISSN: 0168-9002. DOI: https://doi.org/10.1016/j.nima.2015.05.035.

[113] E. Baussan et al. "Neutrino super beam based on a superconducting proton linac". In: *Phys. Rev. ST Accel. Beams* 17 (2014), p. 031001. DOI: 10.1103/PhysRevSTAB.17.031001. arXiv: 1212.0732 [physics.acc-ph].

[114] E. Baussan et al. "Study of the pulse power supply unit for the four-horn system of the CERN to Fréjus neutrino super beam". In: *Journal of Instrumentation* 8.07 (2013), T07006–T07006. DOI: 10.1088/1748-0221/8/07/t07006.

[115] S. Agostinelli et al. "Geant4—a simulation toolkit". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 506.3 (2003), pp. 250 –303. ISSN: 0168-9002. DOI: https://doi.org/10.1016/S0168-9002(03)01368-8.

[116] J. Allison et al. "Geant4 developments and applications". In: *IEEE Transactions on Nuclear Science* 53.1 (2006), pp. 270–278.

[117] J. Allison et al. "Recent developments in Geant4". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 835 (2016), pp. 186 –225. ISSN: 0168-9002. DOI: https://doi.org/10.1016/j.nima.2016.06.125.

[118] A. Ferrari, Paola R. Sala, A. Fassò, and J. Ranft. *FLUKA: A multi-particle transport code (program version 2005)*. CERN Yellow Reports: Monographs. Geneva: CERN, 2005. DOI: 10.5170/CERN-2005-010. URL: http://cds.cern.ch/record/898301.

[119] T.T. Böhlen et al. "The FLUKA Code: Developments and Challenges for High Energy and Medical Applications". In: *Nuclear Data Sheets* 120 (2014), pp. 211 –214. ISSN: 0090-3752. DOI: https://doi.org/10.1016/j.nds.2014.07.049.

[120] R.B. Palmer. "Magnetic Fingers". In: *Informal Conference on Experimental Neutrino Physics*. Jan. 1965, p. 141. DOI: 10.5170/CERN-1965-032.141.

[121] D. Adey et al. "nuSTORM - Neutrinos from STORed Muons: Proposal to the Fermilab PAC". In: (July 2013). arXiv: 1308.6822 [physics.acc-ph].

[122] R. Brun and F. Rademakers. "ROOT — An object oriented data analysis framework". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 389.1 (1997). See also ROOT [software], Release v6.14/02 27/07/2018, pp. 81 –86. ISSN: 0168-9002. DOI: https://doi.org/10.1016/S0168-9002(97)00048-X.

[123] M. J. Kochenderfer and T. A. Wheeler. *Algorithms for Optimization*. The MIT Press, 2019. ISBN: 0262039427.

[124] E. Zitzler and L. Thiele. "Multiobjective evolutionary algorithms: a comparative case study and the strength Pareto approach". In: *IEEE Transactions on Evolutionary Computation* 3.4 (1999), pp. 257–271.

[125] C. Audet and W. Hare. *Derivative-Free and Blackbox Optimization*. Springer, 2017. DOI: https://doi.org/10.1007/978-3-319-68913-5.

[126] M. D. McKay, R. J. Beckman, and W. J. Conover. "A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code". In: *Technometrics* 21.2 (1979), pp. 239–245. ISSN: 00401706. URL: http://www.jstor.org/stable/1268522.

[127] *Centre de Calcule de l'IN2P3*. Calcule scientific et traitement de donneés. Lyon, France. URL: https://cc.in2p3.fr.