Johan Lundquist
Course: FYSM60 60HP Master Thesis VT21 and HT21 Lund University

Supervisors: Sverker Werin and Natalia Milas

Tuning of the ESS Drift Tube Linac using Machine Learning

February 23, 2022

LUNDS UNIVERSITET

EUROPEAN SPALLATION SOURCE

Acknowledgements

I would like to thank, first and foremost, my supervisor Natalia Milas. Without her drive and inspiration this project would not have become what it is, and I certainly would not have enjoyed my time with it as much as I have. Furthermore, I am very grateful to Sverker Werin for his guiding hand in all the writings and communications involved in this work.

I am also very grateful of the rest of the ESS beam physics section: Mamad Eshraqi, Ryoichi Miyamoto, Yngve Levinsen, Ben Folsom and Daniel Noll, as well as the former members Emelie Nilsson and Neven Blaskovic. Everyone named has been an inspiration or helping hand during this project.

External to the writing and work of the project, I must also thank my family and my dear Luisa, who have all helped me through any trying or stressful times throughout the year.

Contents

Abstract

The European Spallation Source, currently under construction in Lund, Sweden, will be the world's brightest neutron source. It is driven by a linear accelerator designed to accelerate a beam of protons with 62.5 mA, 2.86 ms long pulses, working at 14 Hz. The final section of its normal-conducting front-end consists of a 39 m long drift tube linac divided into five tanks, designed to accelerate the proton beam from 3.6 MeV to 90 MeV. The high beam current and power impose challenges to the design and tuning of the machine. In order to keep the beam quality throughout the accelerator and beam losses at a minimum the radio frequency amplitude and phase within the accelerating components have to be set within 1% and 1° of the design values.

One of the usual methods used to define the radio frequency set-point is called signature matching, which can be a challenging process, and new techniques to meet the growing complexity of accelerator facilities are highly desirable. Machine learning is a rapidly growing field which has found applications in a wide range of scenarios, accelerators being no exception, but the tuning of RF fields using machine learning has yet to be tried. This project explores the possibility of applying machine learning in this area of accelerator physics, comparing this novel technique with the established signature matching and introducing a new possibility for faster tuning using a different data structure.

For this purpose, simulations of the first tank in the drift tube linac section of the ESS linac were used to produce large amounts of tuning data at varying setpoints for the machine. Data like this was then used with the signature matching and machine learning techniques to fit the necessary functions for the traditional technique and train artificial neural networks for the novel technique. Random machine errors were later introduced to test each method's generalized performance. This data was also restructured to allow for machine tuning in a single shot, while usually a parameter scan is necessary, and machine learning techniques were tried for tuning on this data.

Machine learning has been found to perform well in comparison with the established method, with some select advantages inherent to machine learning. This faster RF tuning technique only possible with machine learning, is found to perform well, although not quite within the given 1% and 1° limitations. As the rest of the results of this project are all on simulated data for the ESS, a comparison with results of these techniques on real data collected from the Spallation Neutron Source in the USA was performed. Future improvements could include workarounds for faulty network inputs and further tuning of the networks used.

1 Introduction

The European Spallation Source (ESS) is a state of the art neutron science facility under construction in Lund, Sweden [1]. The basic process used by the facility is spallation, wherein one impacts a high neutron material, in this case tungsten, with high energy protons, causing the target to output neutrons. The high energy protons are provided by the ESS linear accelerator (linac), which are brought to the tungsten neutron source in the target, and the neutrons are then guided to different scientific stations called instruments where many types of studies can be performed. Among the many important fields which neutron science contributes to are chemistry, soft matter studies, material science and geoscience. The work at the instruments may very well contribute to the future of energy production, medical research and environmental science [1].

The ESS linac is a 600 m long accelerator consisting of many different sections combining various accelerator technologies in order to raise the proton energy from the 75 keV source output to the final 2.0 GeV arriving on the target. These accelerating structures use Radio Frequency (RF) oscillating electromagnetic fields to provide energy to the passing beam. As the machine is expected to deliver beam of high current and power, a primary concern is to avoid slow beam losses, as these lead to radiation activation of surrounding equipment. Radiation activation, beyond being harmful to staff or equipment, can also severely cost the facility in terms of beamtime, time the scientists at the beamlines can actually use the neutrons, as any work requiring direct access to the accelerator must wait for activation to reach safer levels. In order to avoid such losses, proper and careful tuning of the RF fields is crucial. As a result the requirement for accuracy of the RF setpoint is to be within 1% in RF amplitude and 1° in phase of the design setpoint for all accelerating structures [1]. In order to achieve this type of accuracy, much work has been performed in the last decades to develop new techniques to meet the growing scale and complexity of hadron facilities [2, 3, 4]. A crucial part of the machine is the 39 m long Drift Tube Linac (DTL) divided into five tanks, designed to accelerate the proton beam from 3.6 MeV to 90 MeV. This is one of the earlier parts of the accelerator, and the accuracy of its setpoint will influence the performance of subsequent accelerating sections.

Machine Learning (ML) is a very quickly growing field within computer science, of which new applications are rapidly being discovered. It is becoming widely used in image processing and data analysis, and many applications can be found in accelerator science in subjects such as virtual diagnostics and early error detection [5]. The process of tuning the RF fields of a DTL and the format of the data used suits an application of ML very well, and while ML as applied to accelerators has been developed in other areas, RF tuning remains a novel idea. The ESS specifically allows for new strides and challenges in this area, as its upcoming commissioning phase includes some new limitations on earlier established tuning techniques and the machine is state of the art, involving more diagnostic output than its older counterparts. This project will study the possibilities of using ML in this area, both as an enhancement of more established techniques, as well as for some applications unique to this expanding field, the ultimate goal being to use machine learning techniques to find the setpoint of the ESS DTL within an error margin of 1% and 1° in RF amplitude and phase.

The thesis will begin with a general introduction to hadron linear accelerators and descriptions of the two specific machines studied. The following chapter discusses the longitudinal dynamics of accelerating beams, the physics which underlay all sim-

ulations performed for this project. Next is a description of the tuning methodology and the data used for the calibration of the DTL. In chapter 5, ML is expanded on as a basis for the following method section. The results are then summarised and discussed before an outlook on potential future developments. Finally, many aspects of this project will not be directly discussed in the thesis, but it is worth mentioning that the results were presented at conferences as a form of paper and a poster (see annexes A and B). In addition, as part of the training, I also had the opportunity to participate in the doctoral school US Particle Accelerator School (USPAS), see annex C for further information.

2 Hadron Linear Accelerators

Acceleration of a charged particle in an electric field along a linear path can be traced as far back as early experiments by Norwegian physicist Rolf Wideröe in 1928 [6], but the applications of accelerated beams in other fields of science was not truly realised until after the revolution in microwave technology during World War II. Very early accelerators employed a constant electric field to achieve a gradient used to accelerate the charged particles, but in order to reach high energies a massive voltage is required and one quickly runs into electric breakdown at a few tens of megavolts. This has led to the application of RF fields, time-varying electric fields, in such a way that it maintains a phase relation with the beam to ensure continued acceleration as the field oscillates. These fields are housed within a cavity, with some form of entrance and exit so the beam can move straight through while gaining energy from the fields inside. The particle's velocity increases thus throughout the linear accelerator and is often described as its ratio with respect to the speed of light, as $\beta = \frac{v}{c}$. The most basic form of this type of "RF structure" is a pillbox cavity, a cylindrical space with openings for entrance and exit. An idealized model of this type of cavity and its fields can be seen in Fig. 1 below.

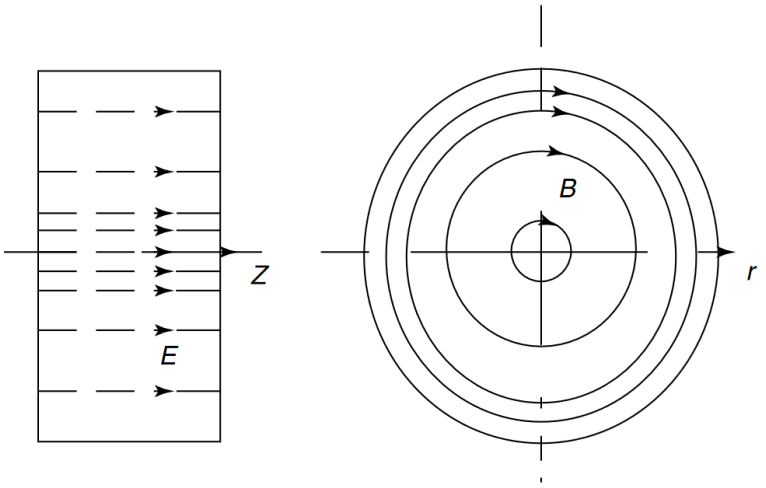

Figure 1: Idealized model of a pillbox cavity, housing electric (E) and magnetic (B) field components of a TM_{010} mode RF field. For proton acceleration, the beam would be moving along the z-direction [7].

Any type of charged particle could be accelerated in these machines, but a common separation within the field is between hadron and electron accelerators. Modern hadron linacs are in many ways comparable to their electron counterparts. Both utilize RF fields, commonly sourced in a modulator and amplified by a klystron, to produce the electric fields required for acceleration and use various techniques to ensure the particles only experience the accelerating part of such an oscillating field. The main difference is the β, which will increase far more rapidly in electron machines, due to the particle's low rest mass relative hadrons. This has many consequences for the design of these machines, as the effects of the relativistic velocities of the electrons will alter much of the underlying physics, and subsequently the design parameters for the accelerator.

Accelerating the beam is obviously the main function of a linac, but in order to achieve this, many more components than just the accelerating cavities are needed. The longitudinal behaviour, the energy and phase, of a hadron beam is mainly controlled by the RF systems, but it is vital to maintain good control and understanding of the transverse dynamics as well, the behaviour in an x-y plane, seeing the beam from head-on. For this, magnetic elements must be used. Transverse focusing can be achieved through quadrupole magnets, magnets with four poles, two south, two north. Such a component allows for focusing of the beam distribution, though only in one plane at a time. By combining multiple quadrupoles in different orientations, a net focusing effect can be achieved in both planes. These types of magnetic focusing elements, with alternating orientation, are installed throughout any accelerator, as the beam will rapidly be lost without them. A schematic of the cross-section of a quadrupole magnet can be seen in Fig. 2.

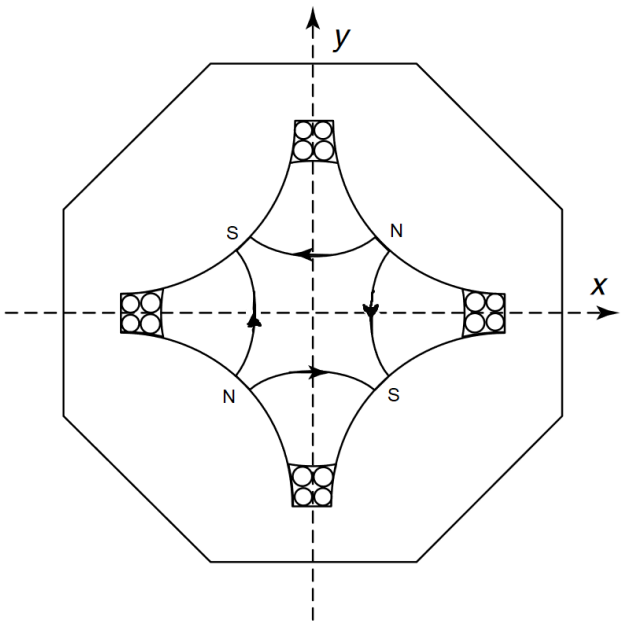

Figure 2: Schematic of the cross-section of a quadrupole magnet, with arrows indicating the directions of the magnetic fields [7].

The components which help to understand and diagnose the beam are most often called diagnostics. Perhaps the most important and commonly applied diagnostic is the Beam Position Monitor (BPM), consisting of four electrodes set within the walls of the vacuum chamber housing the beam. As the beam passes, the field excited by the beam on the vacuum chamber walls will result in a potential across the electrodes. A phase, relative to a central clock, can be measured, as well as the amplitude in each of the electrodes. These amplitudes can be used to calculate the beam's center of charge within the x-y plane while the relative phase has a dependency on the beam energy. Simple schematics of the longitudinal and transverse cross-sections of a BPM can be seen in Fig. 3.

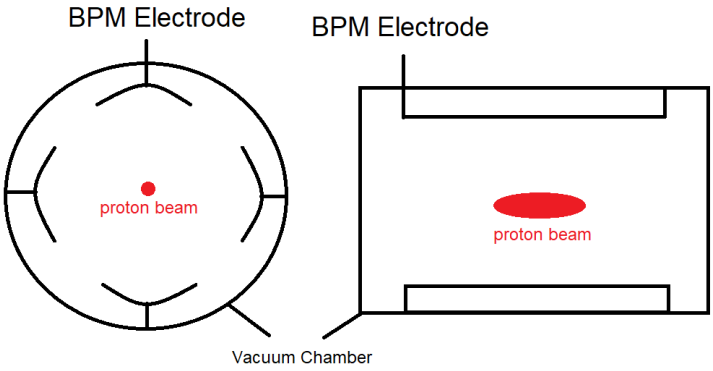

Figure 3: Schematic of the cross-section of a BPM, with longitudinal and transverse cross-sections, on the left and right respectively.

The following Sections, 2.1 and 2.2, will go into more detail on the different sections and components of the ESS and Spallation Neutron Source (SNS) linacs, as both of these machines are in someway investigated in this project. While this project is limited to specific sections, namely the DTLs of these machines, it is important to have an overview of the machines as a whole, as the entire systems may be effected by the tuning performed in the early DTLs.

2.1 The ESS Linac

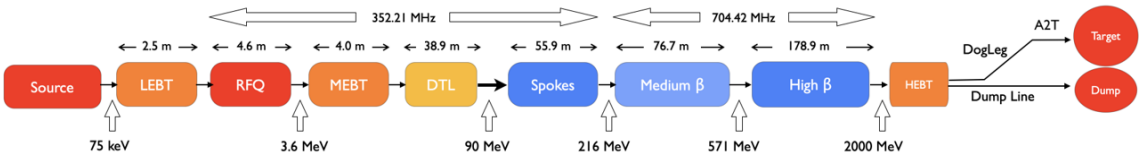

Figure 4: Simple schematic of the ESS linac layout [1].

In Fig. 4 a simplified model of the linac at ESS can be seen. Garoby *et. al* expresses the scientific requirements on the ESS facility as: "To deliver a time averaged flux of neutrons, which is comparable to the most intense continuous source in existence, namely ILL, and at a pulse repetition rate that is low enough to avoid loss of efficiency in the use of the high flux even for the slowest/coldest neutrons in large demand" [1]. This results in two targets for the linac design, that the final average beam power reach 5 MW, and a final beam energy of 357 kJ. With these goals in mind, the linac has been designed with 9 separate accelerating sections, 5 of which contribute to the final energy of the protons, 2 GeV on the tungsten target.

The protons are generated by a microwave-discharge ion source, designed to output a continuous beam with an energy of 75 keV. The particles then travel to the Low Energy Beam Transport (LEBT) section, which includes multiple diagnostics, magnetic focusing elements and a chopper that shapes the beam to the nominal 2.86 ms long pulse required in the later sections. The next element is the Radio-Frequency Quadrupole (RFQ). This is a RF structure, operating at a fundamental frequency of

352.21 MHz, which is designed to bunch, accelerate and focus the beam all at once. After this acceleration, the beam has reached the 3.6 MeV and enters the Medium Energy Beam Transport (MEBT) section. This serves a similar functionality to the LEBT, but also includes three buncher cavities to improve the beam input in the longitudinal plane to the following section. The Drift Tube Linac (DTL) then covers the next 40 m of the linac, raising the energy of the beam to 90 MeV, using the same frequency as the RFQ section. All the sections described thus far are part of the normal conducting front-end of the ESS linac, and subsequent sections are composed by accelerating cavities, which are cooled in cryomodules using superfluid helium at 2 K to become superconducting. This is needed in order to achieve very high accelerating gradients. The first of these superconducting structures are the spoke cavities, 13 separate cryomodules containing two spoke cavities each, raise the proton energy to 216 MeV. Transverse focusing is achieved between cryomodules, by what is called linac warm units, composed of two quadrupoles and extra diagnostics. After the spokes, the beam will travel through the medium and high β sections, consisting of 9 and 21 cryomodules respectively, each housing 4 cavities. These two last accelerating sections, operating at twice the RF frequency of previous ones, raise the beam energy to the final design value of 2 GeV. After there are 2 separate transport lines, which guide the beam to either a final beam dump if it is not to be used for neutron production, or through the DogLeg and Accelerator-to-Target (A2T) sections, to finally impact on the tungsten target to produce neutrons for the experimental stations [1].

2.2 The SNS Linac

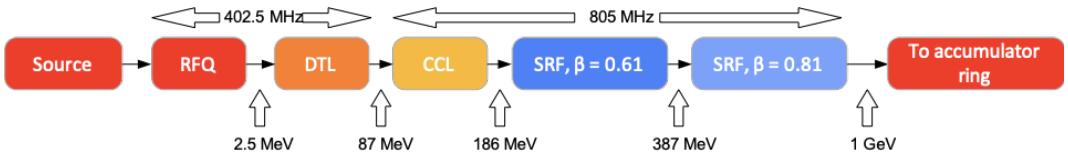

Figure 5: Simple schematic of the SNS linac layout.

The Spallation Neutron Source facility at the Oak Ridge National lab, USA, completed in 2006, is an older spallation neutron facility than ESS. The machines do have some similar components and sections, but SNS differentiates with the final energy of 1 GeV, the accelerated beam consisting of H-, the use of an accumulator ring before the beam arrives on target, as well as the peak current of 60 mA. After the ion source producing the H- there is a very short a low-energy beam transport, followed by a RFQ, raising the energy to 2.5 MeV, and a six tank DTL section increasing the energy to 87 MeV, both operating at a frequency of 402.5 MHz. Subsequent sections all operate at twice the RF frequency. The first of these sections is the Coupled Cavity Linac (CCL), with four modules raising the energy to 186 MeV. After this follows two superconducting sections, with 11 medium-β and 12 high-β niobium cavities, cooled to 2 K. These final sections raise the energy to the final 1 GeV before injection to the accumulator ring. During the injection process, the H- is stripped of both electrons, leaving bare protons in the ring. Here the pulse is shortened from 1 ms to 700 ns in order to increase the pulse intensity onto the target. The target consists of 50 tons of liquid mercury, upon which the entire stored pulse from the accumulator ring impacts and causes spallation to produce the final neutron pulses [8].

3 Longitudinal Dynamics

RF tuning consists of finding the correct values for the RF amplitude and phase with respect to the proton beam in order to provide the optimum acceleration in a given cavity. In order to be able to understand the main factors connected to tuning, it is important before that to understand the dynamics of the particles crossing an RF cavity and how they interact with the RF fields inside it. In this chapter the longitudinal dynamics of particles inside resonant structures will be discussed and the requirements for optimum acceleration presented.

3.1 Longitudinal Motion in Accelerators

As a beam of particles is accelerated through any linac structure, it will have a certain spread around what is called the synchronous particle. The synchronous particle arrives at some point z in the accelerator at time $t_s(z)$, and has energy W_s. It is called synchronous due to its relation to the accelerating time-varying fields. It is defined such that it always arrives with the field in the design amplitude when moving from gap to gap in an accelerating structure, thus experiencing the ideal acceleration. This is the particle which a linac structure is designed for, and yet no particle will be exactly synchronous and they can still remain stable and be accelerated. This is due to what is called longitudinal phase stability.

In hadron linacs, by maintaining the time relation between incoming particles and RF fields such that the particles arrive upon the rising or falling side of the crest, for positively and negatively charged particles respectively, a focusing effect in time can be achieved. This is illustrated for positively charged particles in Fig. 6. This is a somewhat intuitive process. We imagine a beam of particles moving through a series of subsequent accelerating gaps. The synchronous particle will maintain its position upon the rising crest as it moves through any gap, but a particle with less energy than the synchronous particle will arrive later since its has a lower velocity. This leads to this particle arriving closer to the crest of the time-altering field, and it will gain more energy than its earlier, more energetic counterparts. In a subsequent gap it will arrive earlier, closer to the synchronous particle. A similar logic can be used for particles more energetic than the synchronous one. On the other side of the crest we see this process working in reverse, particles more energetic than the synchronous particle actually gain more energy and move away from the synchronous particle, the same will be true for low energy particles which will gain less and less energy. This effect characterizes the two fixed points on the RF waves: the first is a stable fixed point and the second is an unstable fixed point. For ultra-relativistic particles in linacs the longitudinal phase stability is no longer valid since the bunch is effectively frozen. This is due to the fact that the differences in particle kinetic energy no longer effects the time a particle reaches a certain cavity in the accelerator.

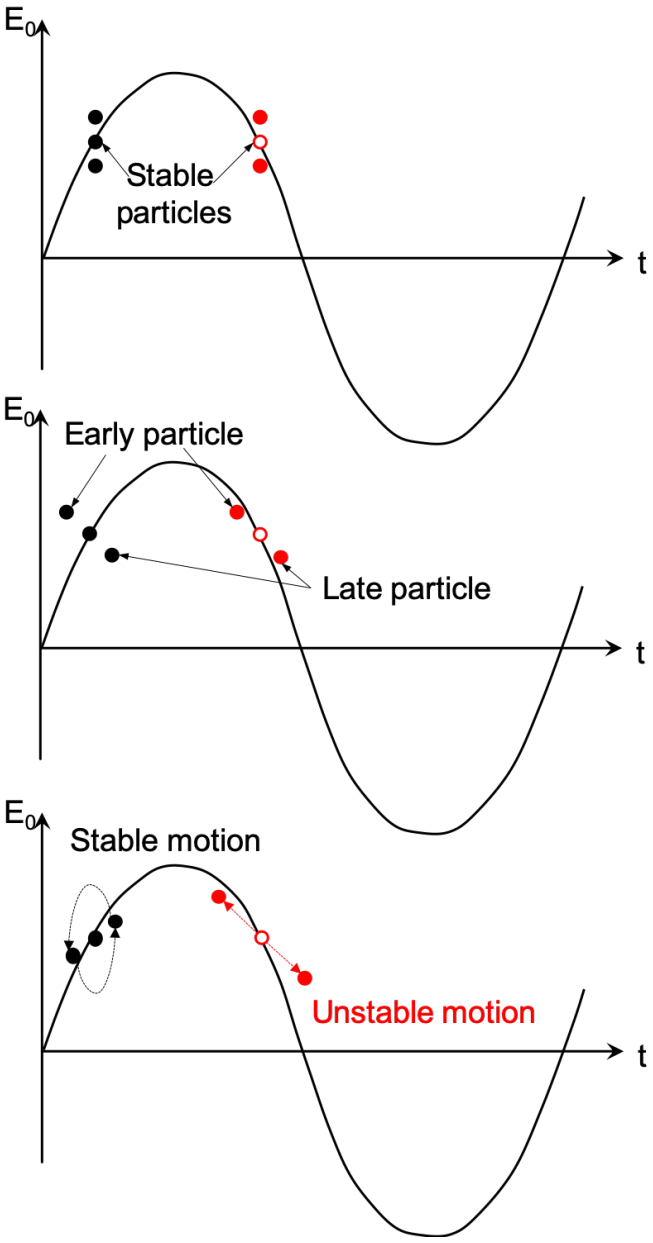

Figure 6: A sinusoidal electric field with particles arriving at different times.

The relation between a particle moving through an accelerating gap and the time-altering field inside can also be described using phase rather than time. The phase range of the field is bounded by 0° and 360°, and the phase of a particle is expressed using the difference in phase between when the particle arrives and when the field reaches its crest, its maximum magnitude. This may be called the phase of the particle or its phase advance, ϕ.

We consider a particle of charge q moving along the longitudinal axis through an alternating electric field $E(z,t)$, stored as a standing RF wave in a pillbox cavity of length L, see Fig. 1. The field experienced by this particle may be described as

$$E(z,t) = E(z)\cos[\omega t(z) + \phi]$$

where z is the longitudinal position, ω is the angular velocity of the RF field, ϕ is the phase advance between the RF field and the particle, and $t(z)$ is the time at which the

particle arrives at z. $\omega t(z) + \phi = 0$ is defined such that it occurs when the particle experiences the maximum amplitude of the oscillating electric field, i.e. when it is on crest. The kinetic energy gain of this particle can be defined as

$$\Delta W = q \int_{-L/2}^{L/2} E(z) \cos[\omega t(z) + \phi] dz$$
$$= q \int_{-L/2}^{L/2} E(z) \{\cos[\omega t(z)] \cos(\phi) - \sin[\omega t(z)] \sin(\phi)\} dz$$
$$\equiv q V_0 T \cos(\phi)$$

where we introduce

$$V_0 = \int_{-L/2}^{L/2} E(z) dz = E_0 L$$

and T, the transit-time factor, which we define as

$$T = \frac{\int_{-L/2}^{L/2} E(z) \cos[\omega t(z)] dz}{\int_{-L/2}^{L/2} E(z) dz}$$

which is the ratio between the energy gain of the real particle traversing the gap and a theoretical maximum energy gain. This factor is always less than 1, as the oscillating nature of the RF field, combined with the particle's real velocity $v < \infty$, will always lower its provided energy relative to a theoretical maximum. If the change of particle velocity as it transitions the RF gap can be considered comparatively small to the initial velocity as the particle enters the cavity we can use $\omega t \approx \omega z / v = 2\pi z / \beta\lambda$. Here $\beta = \frac{v}{c}$, and $\beta\lambda$ would be the distance covered by the particle during one RF period. From this we can express the transit-time factor as

$$T = \frac{\int_{-L/2}^{L/2} E(z) \cos(2\pi z / \beta\lambda) dz}{\int_{-L/2}^{L/2} E(z) dz}$$

If we go on to assume a square, uniformly oscillating electric field of magnitude $E(z) = E_g$, as in Fig. 7, which is zero outside the gap, we find a final simplified transit-time factor as given below [7].

$$T = \frac{\sin(\pi L / \beta\lambda)}{\pi L / \beta\lambda} \tag{1}$$

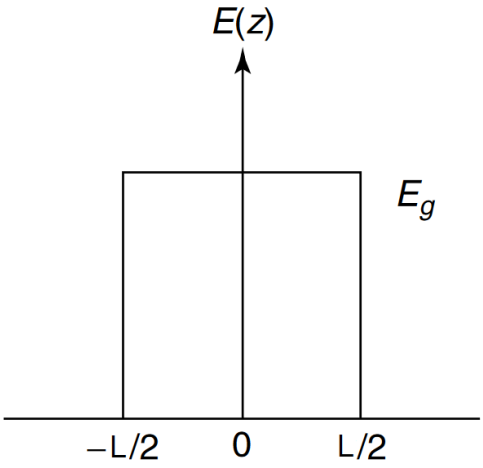

Figure 7: A square electric field of magnitude E_g.

The energy gain can then be simplified accordingly.

$$\Delta W = qE_0LT\cos(\phi) = qE_0\frac{\sin(\pi L/\beta\lambda)}{\pi/\beta\lambda}\cos(\phi) \tag{2}$$

This describes the energy gain of the beam while moving through a gap. Using Eq. 2 we can form difference equations describing the change to the beam when moving through sequential thin gaps indexed as n. We can look at the phase change of a particle with relation to the RF field when moving from gap $n-1$ to gap n as

$$\phi_n = \phi_{n-1} + \omega\frac{2l_{n-1}}{\beta_{n-1}c} \tag{3}$$

where

$$l_n = \beta_{s,n}\lambda/2$$

is half the distance between two gaps, or cells, defined after the synchronous particle. β_s is the velocity of the synchronous particle, which we can use to write a full cell length as

$$L_n = (\beta_{s,n-1} + \beta_{s,n})\lambda/2$$

which we can in turn use to describe the change of phase relative to the synchronous particle as below.

$$\Delta(\phi - \phi_s)_n = \Delta\phi_n - \Delta\phi_{n,s} = 2\pi\beta_{s,n-1}(\beta_{n-1}^{-1} - \beta_{s,n-1}^{-1}) \tag{4}$$

From here we Taylor expand in order to write

$$\beta^{-1} - \beta_s^{-1} = (\beta_s + \delta\beta)^{-1} - \beta_s^{-1} \approx -\frac{\delta\beta}{\beta_s^2}$$

which holds given $\delta\beta << 1$. We can relate this small change in velocity to a small change in energy as below.

$$\delta\beta = \frac{\delta W}{mc^2\gamma_s^3\beta_s} \tag{5}$$

We may now insert the relation from Eq. 5 into Eq. 4 and find the difference relation for phase relative to the synchronous particle as

$$\Delta(\phi - \phi_s)_n = \frac{-2\pi(W_{n-1} - W_{s,n-1})}{mc^2\gamma_s^3\beta_s^2} \tag{6}$$

and using this and Eq. 2 we can find a coupled difference equation for the energy relative to the synchronous particle.

$$\Delta(W - W_s)_n = qE_0L_nT\cos(\phi_n - \phi_{s,n}) \tag{7}$$

The two coupled difference equations, Eq. 6 and Eq. 7, describe the change of a particles energy and phase as it moves through thin sequential gaps, but in order to study the stability of the beam, the equations can be made continuous. Thus we rewrite these as coupled differential equations, modeling the behaviour of the particles moving through a continuous field rather than the discrete thin gaps. For this we treat n as a continuous variable and exchange it for the axial distance s as

$$n = \frac{s}{\beta_s\lambda}$$

which we use to find the differential equations

$$\gamma_s^3\beta_s^3\frac{d(\phi - \phi_s)}{ds} = \frac{-2\pi(W - W_s)}{mc^2\lambda} \tag{8}$$

and

$$\frac{d(W - W_s)}{ds} = qE_0T\cos(\phi - \phi_s) \tag{9}$$

Equations 8 and 9 describe the continuous motion of the particles within a space known as the *longitudinal phase space*. This is a space constructed from the individual particles' phase and energy, with coordinates as $(\Delta W, \phi)$. The motion of different particles within this space are shown in the central plot of Fig. 8. Here we see the trajectories of different particles experiencing the effect of a cosine field as plotted in the top graph. The synchronous particle with coordinates $(0, \phi_s)$ will remain stationary in this phase space throughout acceleration, as it by definition remains in phase with the accelerating fields. Note the similarity between the top graph in Fig. 8 and the scenario displayed in Fig. 6. Longitudinal phase stability allows the beam a certain amount of spread in the longitudinal phase space while still gaining energy and remaining stable, but this has specific limits. For a beam of positively charged particles, arriving on the falling crest is not stable. Here, logic of phase stability works in reverse, late and early particles are pushed away from the fixed point at $\phi = -\phi_s$. This is known as an unstable fixed point and it defines the rightmost boundary of the stable region in phase space. Coupled to this is the stability limitations on ΔW, particles arriving late and early in phase also being less and more energetic, leading to a spread of stable region also along the vertical axis of the phase space. The resulting limits of stability is often called the *separatrix*, within which is held the stable region of the phase space, commonly called the *bucket*. We can see the separatrix in the middle plot of Fig. 8, the stable region surrounding the synchronous particle which remains stationary in phase space. We can see one limit at $(0, -\phi_s)$, the unstable fixed point, and another at $(0, \phi_2)$, where the acceleration is not enough to maintain stability. This could be at

$E_z = 0$ in the top graph, but may be moved within the positive region of the field if other dispersive effects are accounted for, shrinking the separatrix.

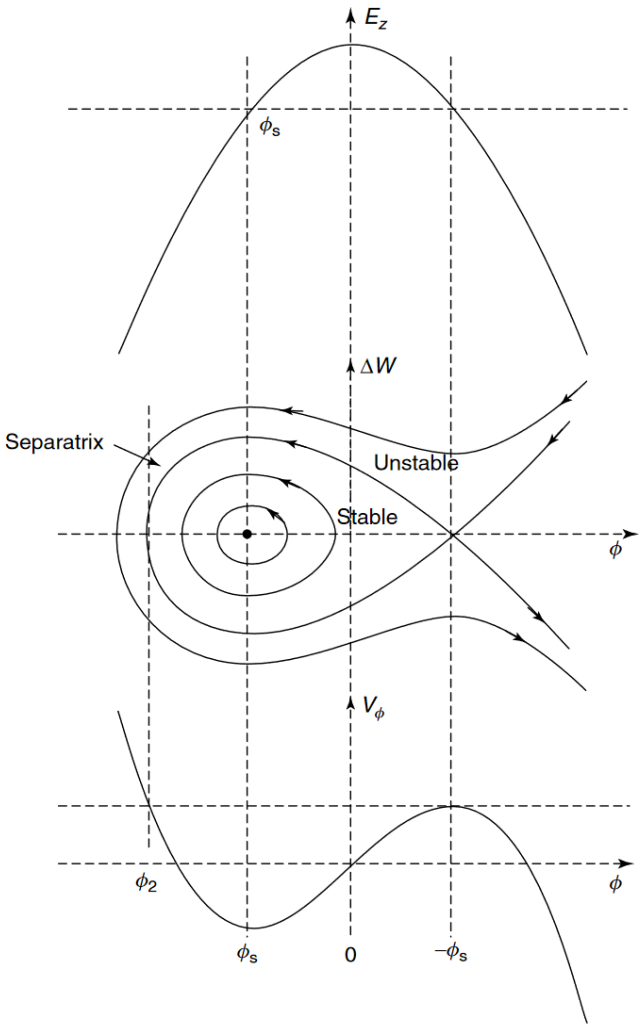

Figure 8: Three separate plots relating to the phase stability of the beam. At the top is an electric field component with a cosine relation to the phase. In the middle is the longitudinal phase space with the trajectories of different particles displayed. At the bottom is the longitudinal potential as a function of phase [7].

This separatrix is the reason the tuning of the DTL is so important. Acceleration through an insufficiently or inaccurately tuned machine leads to particles ending up outside the separatrix, and thus eventually being lost.

3.2 Drift Tube Linacs

The use of multiple simple pillbox cavities for hadron acceleration comes with a few major drawbacks. The housed RF fields lose a fair amount of power through dissipation on each cell wall, which would have to be numerous to reach higher powers. Furthermore, the need for transversely focusing magnetic elements and diagnostics would require spaces between cavities for the installation of quadrupoles, BPMs and other components. This would require a very long linac, and a solution with less cell

end walls and the possibility for components installed within the cavity is called for. One possibility is the Alvarez Drift Tube Linac (DTL), first tested in 1955 [9]. Fig. 9 shows a simple model of the construction.

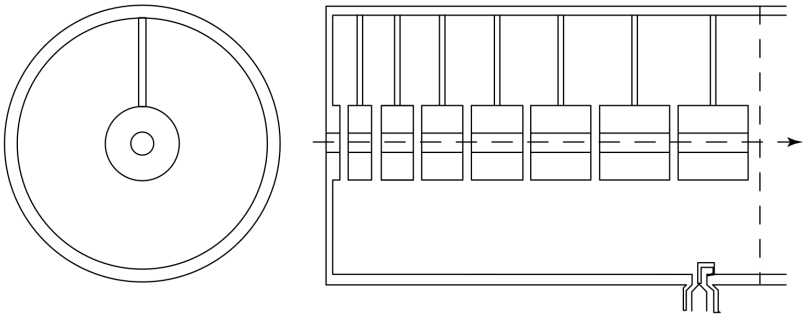

Figure 9: The Alvarez drift tube linac [7].

The DTL consists of a large tank which acts as the RF cavity, within which multiple metal tubes are suspended, known as drift tubes. This structure utilizes a long cavity, a tank, housing a TM_{010} mode field, long enough that the particle would experience multiple full oscillations of the RF field, and thus be decelerated, were it not for the drift tubes. The aim of these drift tubes is to effectively block out the RF field such that the particle will only see the accelerating part of the oscillating electric field. Using such a long tank removes much of the power dissipation into cell end walls, and the drift tubes allow for installation of focusing elements and diagnostics inside these regions without RF fields. For the particles to remain in phase with the RF field, the drift tubes must maintain a periodicity with a period of $\beta\lambda$. Thus the drift tubes, and the gaps between them, must extend in length as the particles accelerate. This will lead to a lowered transit-time factor and less efficient acceleration for high β values. This limits DTL usage to lower β applications, which is the reason DTLs are mainly utilized in hadron accelerators and not for electron machines, as an electron linac commonly reaches $\beta > 0.9$ only a few meters after the gun.

We can apply Faraday's law to the system of an Alvarez DTL to find

$$E_0 = -\frac{i\omega\Phi}{\beta\lambda} \tag{10}$$

where Φ represents the magnetic flux from the circulating magnetic field component of the TM_{010} field [7]. From this equation, and the increasing cell period with $\beta\lambda$, the only factor which must remain constant for a constant E_0 cell to cell is Φ. This is achieved in the precise construction of the DTL tank, tuning the exact B_θ component at the tank wall through perturbation techniques. With such a construction achieved to some acceptable degree we see a constant E_0 from cell to cell, and thus there is only one RF amplitude for the entire tank. There is also only a single setpoint for RF phase, which is the phase at the initial gap in the tank, as the remaining phases can be calculated using Eq. 3. We do see an increasing voltage, as $V_0 = E_0\beta\lambda$, which is part of the reason the DTL can not be utilized at higher β. There is also a lower β limit as when the drift tubes become too small there is no longer room for the vital magnetic quadrupoles for transverse focusing. There is a optimal working point for the DTL at around $0.05 < \beta < 0.4$. Within this region, the DTL accelerates efficiently and is the standard structure of choice in modern linacs.

4 RF Tuning

The ESS linac is a high power accelerator and any losses will have long-lasting repercussions. A sudden loss of the proton beam may result in damage to equipment and parts of the machine caused by the dissipated power. These types of losses can be catastrophic in their own right, but are often avoided using a machine protection system unrelated to this project. The type of losses relevant to the work of RF tuning are slow losses. Slow losses, from inefficient acceleration or mismatched optics, can result in activation of the surrounding components, meaning that any access necessary for ongoing operations or maintenance can be significantly delayed. These slow losses can be minimized by assuring the acceleration is optimized through proper tuning of the accelerating RF fields.

The RF fields within an accelerating cavity are generated by a klystron connected to an RF modulator. The exact field amplitude and phase experienced by the passing beam doesn't correspond exactly to the one set by the klystron control since the end value depends on losses in the wave guides and construction tolerances of the resonant structure the klystron is connected to. The following sections describe the most common methods for determining the actual RF field and phase experienced by the beam.

4.1 Acceptance Scan

In the process of RF tuning, a preliminary measurement to get a general region of interest is often performed. One such preparatory scan is an *Acceptance Scan*. The aim of this scan is to determine a good parameter range to use in secondary, more detailed scans. It uses a current measurement to record the transmission rate of particles through the RF cavity and thus narrows down the phase and amplitude range to be used. The measurement of transmitted particles can be done with, for example, a Faraday cup. Below in Fig. 10 two examples of the simulated results of such a scan can be seen. The upper plot was done prior to this project and tracks the transmission rate as mentioned. Here the different scans are labeled by the simulated difference in the RF amplitude, ΔE_{cav}, in relation to the ideal amplitude. The lower plot is a reconstruction using the simulation code used for this project, code which does not include the transmission rate of the beam. Thus, the final energy of particles leaving the tank is used instead, as it gives an indication of which RF phase is acceptable for acceleration. Here, as in many other figures, the scans are labeled with the corresponding amplitude A, given as a factor of the ideal amplitude A_0. Note the plateau of different phases which all return a transmission of 1, i.e. 100% transmission. A more detailed technique is required to resolve the ideal set-point within this plateau. Furthermore, these kinds of results are only seen in early cavities, or conversely for lower energy particles around a few MeV, such as in the first DTL tank at ESS; later accelerating cavities may not influence transmission as much and can have full transmission across the entire parameter space, even if the acceleration is inefficient [3].

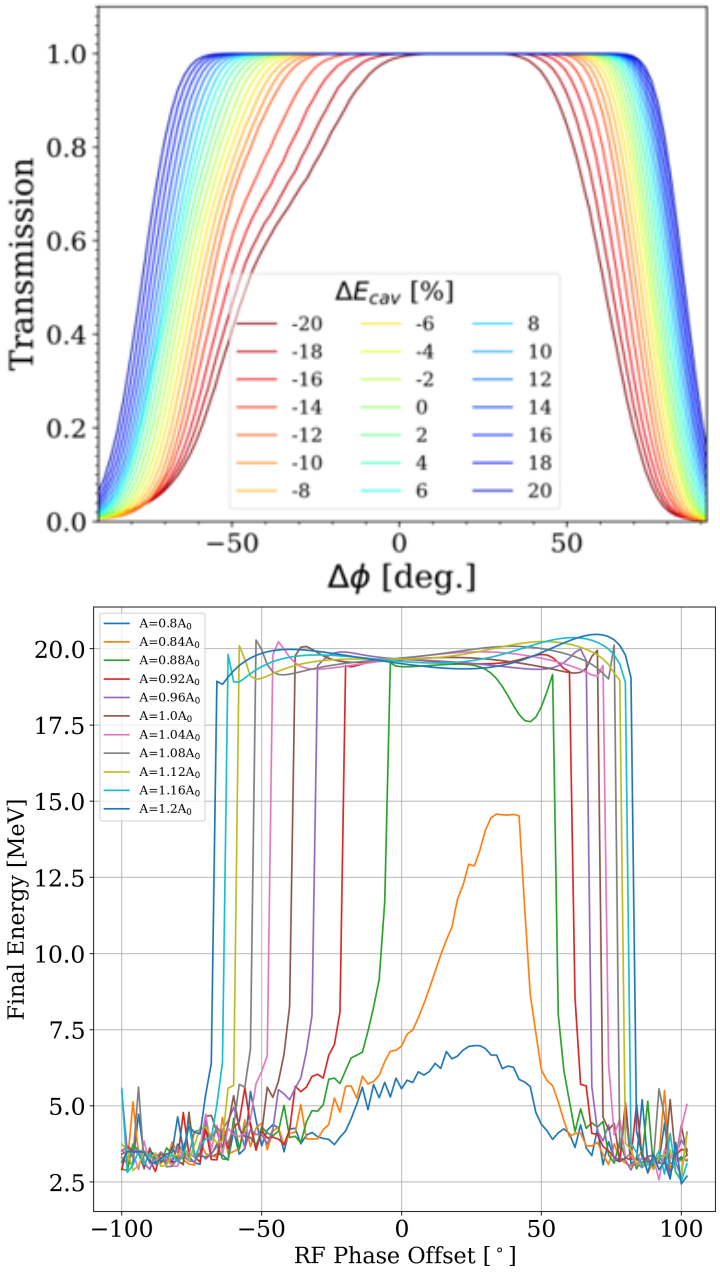

Figure 10: An example of results from an acceptance scan. In the upper plot, the y-axis shows the transmittance through a section of the accelerator, the x-axis the RF phase, and the many multicolored plots are different RF amplitude settings. The lower plot is a reconstruction using the simulation code used in this project, and as transmission is not used in these simulations, the y-axis instead shows the final energy of the particles leaving the tank. Data simulated for ESS DTL tank 1 [10].

4.2 RF Phase Scan

The acceptance scan described above may give a good region of transmission for early cavities, but this is not enough to get a modern accelerator working in optimal conditions. In order to have acceptable matching with later sections of the ESS linac, for example, the field in RF accelerating cavities must be set within 1% in amplitude and 1° in phase of the design setpoint. For this a more refined technique is required.

In order to be able to quantify how the beam responds to changes to the RF setpoint, another diagnostic, sensitive to the beam time of flight through the cavity, must be used. For those cases a Beam Position Monitor (BPM) can be used. As the beam passes a BPM both the amplitude and phase of the fields excited on the BPM sensor by the passing beam are recorded. Although this phase alone doesn't hold much information, by comparing two BPM phases we can get a fast measurement which is proportional to the time-of-flight, or looking with respect to acceleration in a RF cavity, the energy gain between the two devices. It is important to stress that this measurement is relative and that extracting the absolute values of the energy is not an easy task. For this technique using only the relative phase changes is proven to be enough.

The RF phase scan is done in a manner similar to the acceptance scan, accelerating structures after the one being tuned are turned off and used as drift space. The BPMs on those structures are then used to measure the phase difference as a function of the phase of the RF field in the accelerating cavity. The components involved can be seen in a simplified model in Fig. 11. As the BPM's measured phase is closely dependent on the energy of the beam, scanning RF amplitude and phase in a cavity and plotting out the resulting phase differences will give rise to different curves depending on the proximity to the ideal setpoint for the cavity. A few of these signature curves can be seen in Fig. 12, where the ideal setpoint can be found from the signature for the 0% amplitude deviation, which is the greener and the brown lines in the upper and lower plots respectively, and the -35° phase setpoint along the x-axis. Here, as with the acceptance scan in Figure 10, curves are labeled with their corresponding amplitude and input beam energy, given as a factor of the ideal amplitude A_0 and input beam energy E_0.

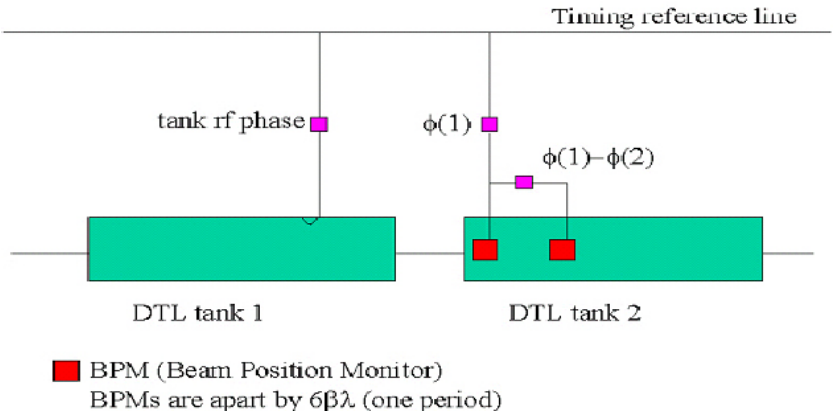

Figure 11: Model of the components in a DTL RF phase scan procedure [11].

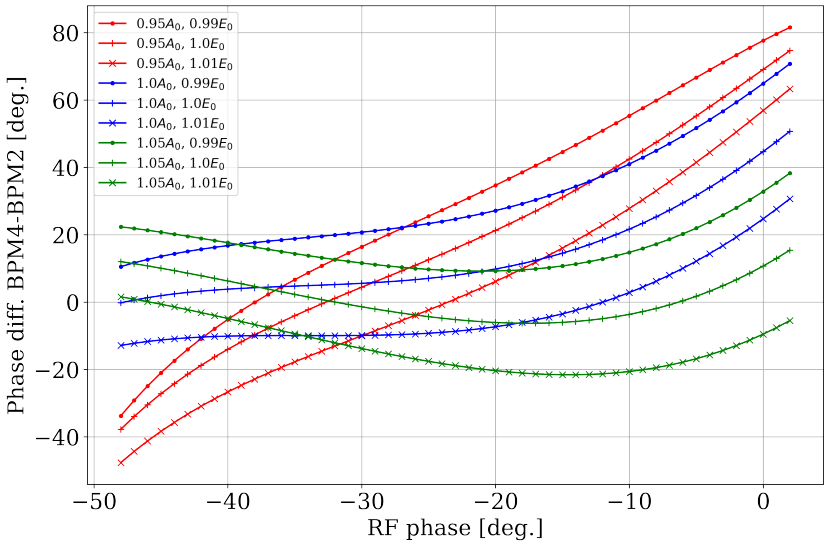

Figure 12: The phase curves for different RF amplitude setpoints. BPM phases simulated as comparison between two BPMs in the DTL tank 1 in the ESS linac.

Much work has been done to find efficient and fast methods for identifying these signatures. The sections below describe first the established technique of signature matching and then the novel technique of utilizing machine learning to achieve this goal.

4.3 Signature Matching

The phase scan curves displayed in Fig. 12 are smooth and can be differentiated by eye. This suggests they could be fitted by some function or model, and if this fitting function is dependent on the setpoints one wishes to calibrate, a perfect fit would return the setpoints producing a specific curve. Thus one could identify the curves using a fit to provided data. This method, which is widely used [2, 3, 4], is known as *Signature Matching*. One uses a set of simulated data (like the ones shown in Fig. 12) and fit each phase scan curve, for RF amplitude A and input beam energy ϵ, with a function or model. In our case, we used a simple polynomial,

$$f(\phi, A, \epsilon) = a_0(A, \epsilon) + a_1(A, \epsilon)\phi + a_2(A, \epsilon)\phi^2 + ... + a_n(A, \epsilon)\phi^n \tag{11}$$

and at this point we assume that an error in the input phase of the cavity acts as a simple offset on the variable ϕ of the whole curve. As a result from all the fittings, it is possible to construct a surface $a_n(A, \epsilon)$ for each coefficient of the fitted polynomial. This surface can then be approximated by another 2D polynomial function so that we have a continuum of coefficients for any given set-point in (A, ϵ).

With the parameters obtained from the fits described above, it is possible to investigate the variance, χ^2, between the simulated model and the corresponding measured data. This variance is defined as

$$\chi^2(\phi_0, A, \epsilon) = \frac{\sum_j^N (f(\phi_j - \phi_0, A, \epsilon) - W_j)^2}{N} \tag{12}$$

where f is our fit to the model prediction of the BPM phase difference (Eq. 11) and W_j is the measured BPM phase difference at some unknown RF phase, RF amplitude

and incoming beam energy. One then optimizes the values of ϕ_0, A and ϵ to minimize Eq. 12, thus determining to which set-point the signature curve W_j corresponds to. [3, 4]

Despite this method and others being impressive progressions in time and cost efficiency, the RF tuning of a large scale accelerator is still a major undertaking and one that must be performed multiple times during a machine lifetime, especially after long-term shutdowns for maintenance and upgrades. This leaves room for improvement in this area of accelerator physics and operations.

4.4 Machine Learning for RF Tuning

The examples of data shown in Fig. 12 may appear easily discernible by eye or by a χ^2-analysis as explained above, but this is not always the case. The BPM 4-BPM 2 phase difference displayed is a specifically discernible diagnostic channel, while others are far more similar despite adjusting the RF setpoint. Figure 13 shows the phase difference between the second and first BPMs of DTL 1, and we can see how a fitting function may have difficulty differentiating between some of these scans. This renders the fitting function $f(\phi, A, \epsilon)$ a crucial part, as it must be able to have distinct fits for very similar data. Furthermore, the accelerating and drift cavity scanning technique described in 4.2 is not possible in the case of the initial commissioning of ESS. This will be performed with only DTL tank 1 present in the machine, and as such no subsequent drift space can be used and the RF phase scans for tuning must be carried out with the BPMs within the accelerating cavity. When using BPMs in a deactivated cavity, there is no acceleration between the BPMs and comparing their phases gives a more reliable pseudo time-of-flight measurement, while allowing acceleration between the diagnostics complicates their relation in phase. This also puts further emphasis on the choice of BPM combination to use with the signature matching technique, as the particle velocity is changing throughout the tank. ML could potentially discern between scans without a given underlying function, leading to the potential of all scans being used at once, and thus no need to investigate the sensitivity of specific BPM combinations. If ML could be shown to outperform the established methodology, it could save ESS valuable resources and time during machine set up. Furthermore, ML allows for novel data structures to be used.

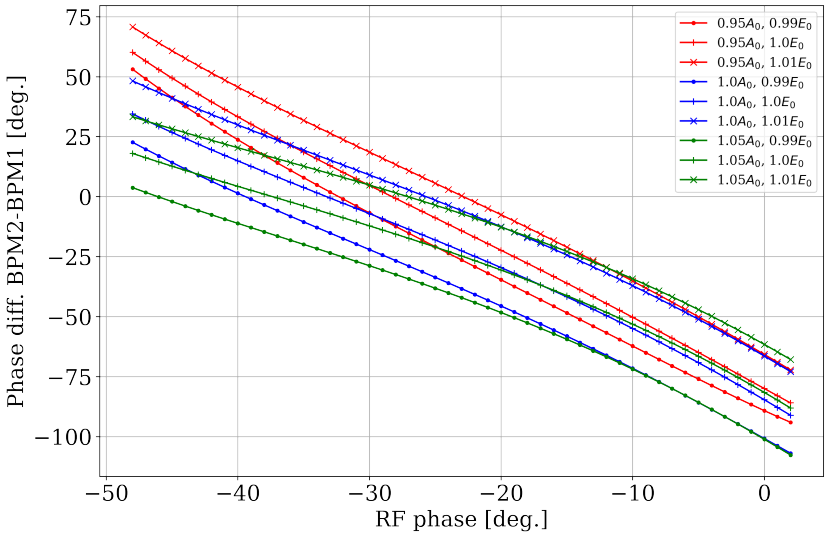

Figure 13: The phase curves for different RF amplitude setpoints. BPM phases simulated as comparison between the first two BPMs in the first DTL tank showing the similarity between different phase scans.

Traditional RF phase scans are an established and reliable method for extracting the information needed to achieve good tuning, and with a more limited diagnostic output it is the only option available. However, with the large number of BPMs within the ESS DTL section, a restructuring of the data can be done such that we can see distinct signatures for each cavity setpoint in amplitude, phase and beam input energy without the need of a full phase scan. We look at BPM phase differences, not against RF phase, but against each diagnostic output, the pairing of BPMs. Figure 14 shows an example of this type of plot, where each line represents a cavity setpoint and is measured in a single pass through the machine, without scanning any parameter. From here we encounter the same problem to be solved as with the phase scan data, needing to accurately identify these new signatures. The nature of the signatures in this data format leaves ML uniquely equipped for the task.

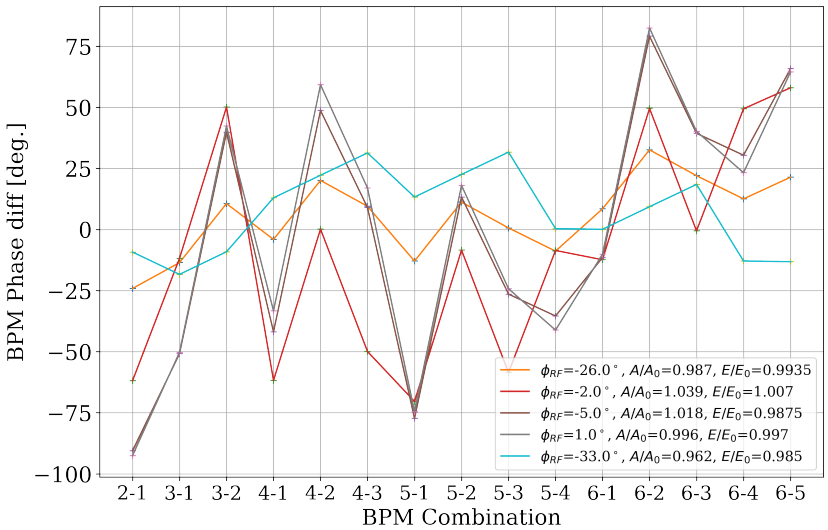

Figure 14: BPM phase differences for each possible BPM coupling, with the different plots each corresponding to a single cavity setpoint.

If accurate predictions can be made with this data structure, a few new advantages manifest. Being able to tune the machine acceptably with a single shot could cut down set up times substantially. One would also not require to determine a range for the scanned parameters as in more traditional RF tuning [2, 3] but would in principle reload the last machine state with good settings and run a single verification pulse. Thus, this type of data and method is henceforth referred to as *Single Shot*.

5 Machine Learning

Machine learning is a term used to describe a specific field of modern computer algorithms capable of learning from experience. Most often, the algorithm consists of what has been termed a neural network, a network of individual artificial neurons, where the weighted connections between the neurons can be trained to reach an ideal output. This training can be performed in different ways, but the most relevant to this project is the technique of supervised learning. Machine learning algorithms come in many forms and can solve many distinct problems using varying network structures, definitions of loss and learning algorithms. Some examples of general groups of problems are classification, regression and structured output. Classification asks a network to take many different input sequences and classify these into one of a select number of groups; regression asks the network to provide some number of continuous values for each input sequence; and structured output asks the network to detect and identify underlying connections between data points in the input [12].

For the purposes of the problem at hand in this thesis, regression network structures are most relevant. A brief walkthrough on the principles of machine learning as well as its application to accelerators in general will be presented. A brief walkthrough of decision tree models and XGBoost is also included, as well as an overview of the current state of machine learning in accelerators.

5.1 The Perceptron

The basic model of the biological neuron is to feed in a signal until the neuron reaches a certain threshold, after which it activates. In the artificial neuron the signal being fed in is the input data, or alternatively signals from other neurons, and the activation is defined by an activation function, the action on the input being performed by the neuron. This single neuron, or node, was the first machine learning algorithm, termed the *perceptron*. What allows the perceptron to learn is the weights it prescribes to each connected input. These weights are then updated according to a learning algorithm, such as supervised learning, that will be outlined in the next session.

The activation function of a node is the manipulation the node will perform on the input data. This could be any manipulation, but a few functions are most common, described here for input θ and output h. One could simply feed the data through without manipulation beyond the application of the weight factors, this being the linear activation function, $\theta = h$. This has the advantage of being continuous, but this also means it can only solve problems with a linear dependence between input and output, such as separating groups of data using only a straight line.

There is also the threshold function which most follows the model of a neuron, as output appears as a certain threshold is reached, i.e.

$$h(\theta) = \begin{cases} 1, \theta > 0 \\ 0, \theta < 0 \end{cases}$$

This can often be used acting on a summary of the signals from the rest of the network if a binary classification is the needed network output. We then have a softer threshold with the sigmoidal function and a modified linear function with the rectified linear

unit function, ReLU. The sigmoidal is defined as

$$h(\theta) = \frac{1}{1 + e^{-\theta}}$$

and ReLU as

$$h(\theta) = \begin{cases} \theta, \theta > 0 \\ 0, \theta < 0 \end{cases}$$

These are some of the most commonly used activation functions, the sigmoidal holding a distinct advantage in its non-linearity, as it can solve more general problems with fewer nodes. The ReLU is computationally cheap and much like the linear function, its simplicity allows application to many different problems [12].

These types of nodes, with a given function, are then set up in a network of weighted connections. This can be called a multi-layer perceptron, or otherwise an Artificial Neural Network (ANN), and a very basic layout is shown in Fig. 15. In this figure, the arrows would have designated weights w which would be updated through some form of learning. Applying the weight to an input channel to a perceptron provides the input, $\theta = W \cdot x$, and the output would be weighted as well, giving a total function for the network output as $y = w \cdot h(W \cdot x)$.

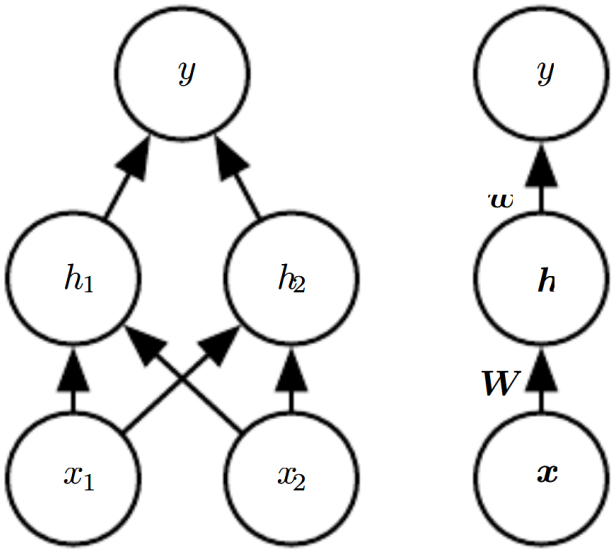

Figure 15: Simple model of an artificial neural network. To the right is a single perceptron with input and output, and to the left are two fully connected perceptrons [12].

5.2 Supervised Learning

In supervised learning, input data for the network comes with predefined labels, giving the correct output expected from an ideal network. The network will always produce the wrong output to begin with, as it uses randomized initial weights, i.e. connections between nodes. From this incorrect output, a loss is defined, say as the mean square difference between the networks output and the given labels. This loss will then influence an optimizer which updates the weights throughout the network. As

this change to the weights depends on the magnitude of the loss, the weights will in theory settle once the network is able to accurately predict the given labels.

In order to utilize this form of learning algorithm, we must first define the loss. For the purposes of the task addressed here, mean absolute error is a satisfactory loss function and is defined as

$$L = \frac{\sum_i^N |y_{lab}^i - y_{pred}^i|}{N} \tag{13}$$

where y_{lab} are the labels, y_{pred} are the networks predictions and N is the number of data points. This error allows for improvement on continuous outputs, as compared to classification errors which can be binary.

This type of error is then used to update the weights throughout the network through some form of optimization algorithm. The most simple version of such an algorithm is gradient descent, which can be defined as

$$w_n = w_{n+1} - \eta \nabla L \tag{14}$$

where w_n is a weight at iteration n, and η is the learning rate, a factor set as a meta parameter before training begins. In a multi-layered neural network, this type of update rule moves backwards from the output and updates all weights throughout the system, before the network tries again, hopefully outputting a lower L in the next iteration. Gradient descent is a fast and reliable method, but it is rarely used in modern machine learning applications, as it can easily get "stuck" in local minima, as it will see a positive gradient in every direction in the parameter space. To avoid this, randomness is introduced to the steps taken by the optimizer. Simply allowing gradient descent to randomly take steps in the "wrong" direction produces the optimizer known as stochastic gradient descent, which is employed more frequently. However, if one continues to introduce more elaborate methods, such as a momentum factor and bias correction, one can arrive at ADAM, named after adaptive moment estimation. This is the most commonly applied optimizer for ML, and the one employed for all networks trained in this project [13].

The data is rarely presented in full or only once to the network, but rather introduced in chunks and many times over. The chunks of data are known as batches and a full run over all data is known as an epoch. A network can take thousands of epochs to discern all the patterns in a dataset, depending on the complexity of the problem at hand.

An important measure of quality of a neural network is how well the network can perform predictions on unprecedented data, i.e. the generalized performance. For the purpose of measuring this, input datasets are commonly split into parts, where some part is withheld from the network during the application of the training algorithm. This withheld part is referred to as the validation dataset, and the part used for training is called the training dataset. At the end of each training epoch, the validation data is introduced to the network to get an idea of how it performs on unprecedented data. This dual performance measurement is exemplified in Figure 16, where the blue dashed line represents the performance on training data, and the green line shows the generalization error, the performance on a validation dataset. These are plotted against the epochs the network trains. As weights are optimized to the training data, the network may lose performance on data not part of this set as training runs longer and longer and the network loses its variance. This is known as *overfitting*, and when optimizing a network it is always a balancing act between lowering losses and keeping

the network's generalized performance high. The point of optimal capacity is marked in Figure 16, along with the zones of overfitting and *underfitting*, i.e. where the performance of the network is still improving [12].

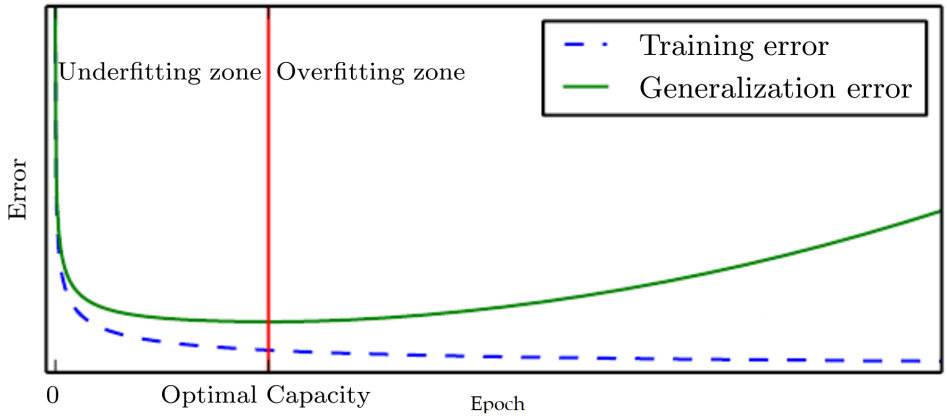

Figure 16: The performance of a network on training and validation data as it changes with the training epochs, with the point of optimal performance marked, before which the network is underfitted and after which the network is overfitted [12].

5.3 Decision Trees and XGBoost

The discussion thus far has covered only the ML method of an artificial neural network, but this is not the only way of structuring an ML system. The modern ML system of XGBoost (eXtreme Gradient Boosting) is an open source tree boosting model, which has proven extremely powerful for solving varied, nonlinear problems [14]. Tree boosting models such as XGBoost are based on the decision tree model of ML structures, with a regularized objective function. This means the loss function punishes system complexity, i.e. the number of trees and branches. In a decision tree ML system, the parameters adjusted in training are not the weighted connections within a network of nodes, but rather the branching criteria in a large decision tree. A gradient boosting ML system uses an ensemble of many decision trees in order to improve the final predictions, and commonly a regularized loss function which penalises increasing complexity of the model as well as the usual error of predictions. This regularized loss is then applied to the ensemble of trees iteratively to improve the output by training the branching criteria.

5.4 Machine Learning in Modern Accelerators

Currently, the field of ML is actively breaking through into the world of accelerator science. The complexity of modern accelerators and the magnitude of data created at a frequency too high for even expert comprehension, lends the field well for many different applications of ML. For example, simulations are an important part of working on accelerator. Online modeling of the accelerator helps during operation, and outside operation simulations are vital for optimizing and setting up the machine, as in this project. As mentioned, accelerators are becoming more and more complicated, and consequently, simulations are taking longer and becoming more computationally

expensive. This can be circumvented somewhat using *surrogate modeling*, where an artificial neural network may observe the output of a heavier physical model and attempt to mimic the behaviour whilst cutting down on time and resources. Such a model was tested at the FAST facility to good results in 2018 [15].

Diagnostics, such as the BPMs employed for this project, are a vital tool for accelerator physics and operations. These devices and other forms of diagnostics can be problematic in a number of ways. They can become unreliable or break down with time, they may need to disrupt the beam in order to collect data, or the construction parameters of the accelerator may not allow for diagnostics in important areas. In all these cases, a *virtual diagnostic* could be employed. The idea is to use ML methods to predict the output of non-functioning or lacking diagnostics, training with output from simulations or the physical diagnostic one is hoping to replace. Examples include longitudinal phase space predictions from SLAC [5] and a virtual instrument developed at LANSCE [16].

Applying ML to the task of optimising a system such as the ESS DTL is quite a natural fit, especially given the high amount of diagnostics within the DTL tanks. The type of data shown in Fig. 12 and even in Fig. 13 should be easily distinguishable by some type of ANN. Yet this type of tuning with the use of ML has not been attempted in the past, and this project will hopefully help to break in ML into accelerator physics to the advantage of both fields.

6 Method

6.1 The ESS DTL

The ESS DTL section consists of five accelerating tanks, each 8 m long, which raise the energy of the protons from 3.6 MeV to 90 MeV. This means the beam enters the DTL with $\beta = 0.087$, within the most efficient span for an Alvarez DTL. The entire section has an operating RF frequency of 352.2 MHz. For this project, the first of these five tanks was simulated in the accelerator environment OpenXAL [17, 18]. This first tank, hereafter DTL 1, has 61 individual accelerating cells, or gaps, and an accelerating gradient $E_0 = 3.00$ MV/m. DTL 1 raises the energy from 3.6 MeV to 21.29 MeV. The transverse focusing required to maintain the beam is done using permanent magnet quadrupoles installed within the drift tubes [1]. The diagnostics used for extracting the RF phase scan data are the six BPMs situated within the tank, also installed in the drift tubes.

6.2 OpenXAL Simulations

OpenXAL is an open source development environment written in Java for accelerator physics. It has been developed in a collaboration between many different accelerator facilities, ESS and SNS among them. OpenXAL was used in this project to simulate the first tank of the ESS DTL during acceleration and to reproduce the signals from the six BPMs inside DTL 1 [17, 18]. As phase difference is the data of interest, this results in 15 different BPM combinations, each combination producing one data point for each cavity setpoint, in RF amplitude and phase and input beam energy. ML requires large amounts of data for training networks and for this purpose an error free dataset was used. This consisted of 110 different amplitude setpoints, with a variation of $\pm 5.5\%$ around the design RF cavity amplitude $A_0 = 6.89$ MV, 60 different input beam energy setpoints, with a variation of $\pm 1.5\%$ around the design input energy $E_0 = 3.6$ MeV, and 55 different phase setpoints, spread evenly around the -35° design setpoint.

To this perfect machine four different types of errors were then applied. BPM longitudinal position within the machine was adjusted, potentially caused by installation and construction, as well as the phase readout from these BPMs, produced by electronic limitations. There are also errors arising from production limitations when constructing the cavities. Such limitations could give rise to errors in both RF amplitude and phase gap-to-gap. The errors in the RF parameters and of the BPM longitudinal placement are "static", in that, once the machine is installed they will have a given reproducible value, as compared to the BPM readout error, which will be random each time the diagnostic is used. The different types of errors and their magnitudes are summarized in Table 1.

Initial simulations and error estimates were made using the design requirements for the DTL. From these parameters, the lattice used within OpenXAL for all the simulations was set up. However, in late August 2021, final assembly and exact measurements were finished on the DTL 1 tank. This meant there were exact measurements of the previously roughly estimated values of the gap-to-gap amplitude and phase errors of the tank. A new lattice was constructed using these values, with little variation from the "design lattice", but the errors produced were now included, and as such the previously pessimistic errors shown in Table 1 could be reduced. A factor 10

Table 1: The different types of errors used in simulations and their corresponding magnitude.

Error	Design Magnitude	As-Built Magnitude	Error Type
BPM $\Delta\phi$	$\pm 1^\circ$	$\pm 1^\circ$	Dynamic
BPM Δs	$\pm 100\,\mu$m	$\pm 100\,\mu$m	Static
RF Amplitude	$\pm 2\%$	$\pm 0.2\%$	Static
RF Phase	$\pm 0.5^\circ$	$\pm 0.05^\circ$	Static

reduction of the RF phase and amplitude errors was used. These new errors would thus not be due to construction limitations, but rather due to uncertainties in these final measurements. This new lattice is henceforth termed the "as-built" lattice, and the previous one with no errors included is termed the "design" lattice. Methods were trained on an error-free lattice and then tested with the corresponding errors added. In the results, two cases are presented, Case A with the design lattice and Case B with the as-built lattice. To summarise:

- **Case A:**
  - Training with data from design lattice.
  - Testing with data from design lattice + design errors.
- **Case B:**
  - Training with data from as-built lattice.
  - Testing with data from as-built lattice + as-built errors.

The SNS linac DTL section was also simulated in OpenXAL using a separate lattice file. The provided measured data from the SNS DTL section included 5 separate phase scans of the first DTL tank, while the simulations done included 10000 separate RF amplitude setpoints, each containing 23 values of RF phase surrounding the optimum phase point. Note that input beam energy was not scanned as in the ESS simulations, as this was not included as an output in the final ML network.

6.3 Signature Matching

As covered in Section 4.3, signature matching requires a fitting function $f(\phi, A, \epsilon)$. There were early discussions on which form this function should take, as the RF phase relation to the BPM phase was unknown. Observing the output from the OpenXAL simulations as in Fig. 12, using the BPM combinations which had the most variance between setpoints, these appeared to resemble high-order polynomials. Trials started with the order of this fit. Different polynomial orders, ranging from third order to ninth order, against the χ^2 of the fits can be seen in Fig. 17. The χ^2 represented in this figure is the mean of the fits of all BPM combinations. Here we see that sixth order polynomials appear to achieve good results, and going to higher order may improve the initial predictions but risks overfitting. With an order of polynomial established, fits were performed to the perfect training data. This constructed a basis of coefficients $a_n(A, \epsilon)$ which resembled a 2D polynomial of some higher order. In order to generalize

performance so that coefficients could be picked outside this discrete surface, each coefficients surface was fitted with a fourth order 2D polynomial. An example of this type of surface, its fit, and the difference between the two can be seen in Fig. 18. As can be seen in the difference surface, there appears a structural difference between the data and its 2D fit, although it is around a factor 0.01 of the actual data. With these fitted surfaces provided, predictions on new data with errors could be tested with this methodology.

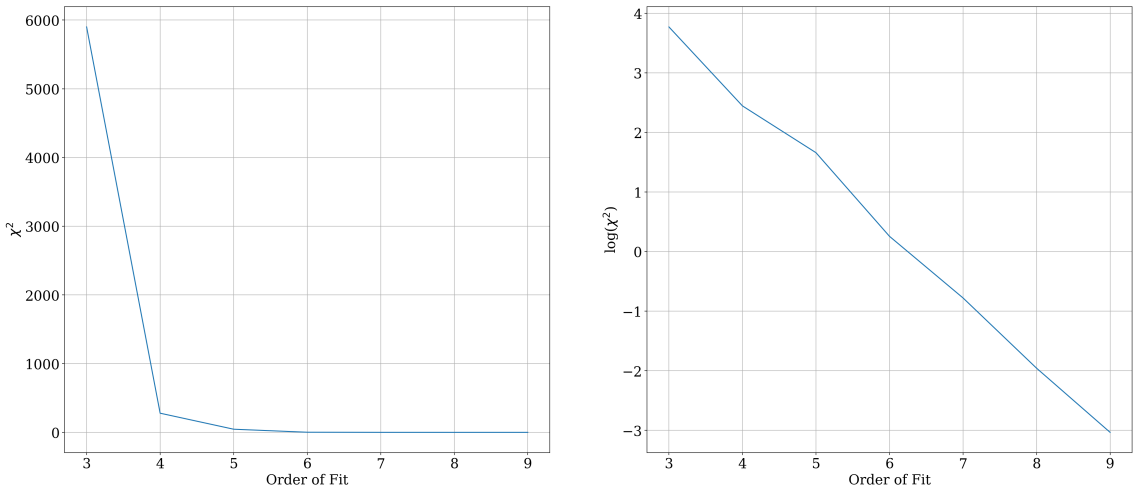

Figure 17: χ^2 of different orders of polynomial fits of the phase scan data, ranging from third order to ninth order, presented both linear and logarithmic. χ^2 is taken as an average of all BPM combinations.

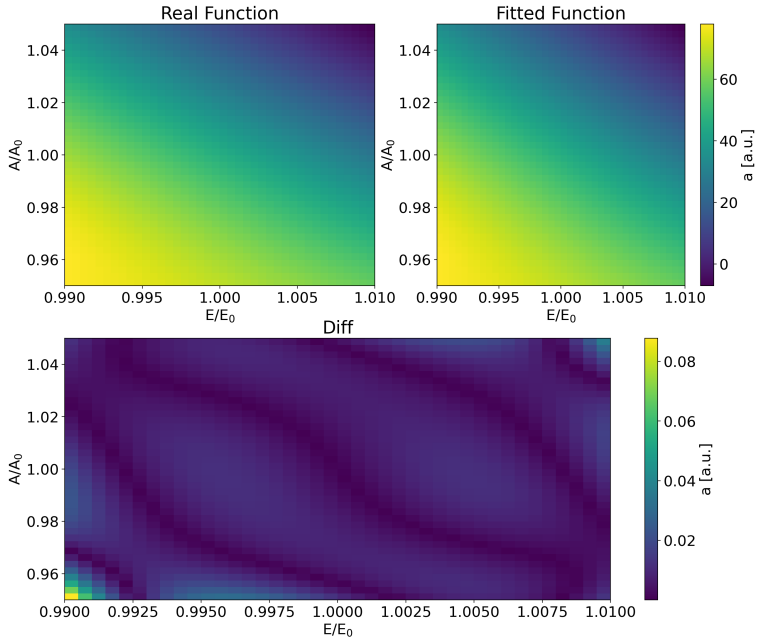

Figure 18: The 2D surfaces of the first parameter $a_0(A, E)$, top left showing the points given from the fits of the signature curves, top right showing the 2D fit of this surface and bottom showing the difference between the two.

6.4 Machine Learning Optimization

Initial optimization of the network was performed on the OpenXAL simulated dataset, collected from the six BPMs of the first DTL tank. The first dataset used to train all networks included no errors, and generalized performance was then tested on datasets including all errors, for both cases A and B. As the phase scan method requires combining BPMs together, there was a total of 15 phase scan plots produced for every amplitude and input energy. An early problem was that introducing phase scans one after another as would seem natural meant showing the network many different data signatures for the same RF amplitude, one for each BPM combination. This made training the network unreliable as the weights could not converge given the varied input/output relation. The data was restructured, combining all BPM combinations into a single large vector. To this vector was also added the RF phase being scanned over, with an offset, to allow this as a label for the network to train on as well. The final network task was then to take these phase scan vectors with lengths of the individual amount of data points in the phase scans multiplied by the amount of BPM combinations, plus the points in phase added to the normalized vector and output a three-vector including the RF amplitude and input beam energy calibration factors, as well as the RF phase.

The hyperparameters of the neural networks being tuned were mainly the learning rate, batch size and the structure of the network, i.e. number of layers and nodes per layer. There are other hyperparameters one could fine tune in the ADAM optimizer, but no issues arose from leaving these on the default values. In order to minimize overfitting, the number of epochs to run was handled automatically by running checkpoints during learning and saving the weights at the epoch minimizing validation loss. For all the final trained ANN models, 20000 epochs were used, the one with the best generalized performance saved. The learning rate, as η in Eq. 14, was varied between 0.001 and 0.00001 during different training runs, mainly effecting the pace at which the network converged on the best epoch, and how much overfitting followed. For the final networks used, the best generalized performance was found using the lower learning rates at 0.00001. Batch sizes were similarly varied between 100 and 1000 phase scans. Using lower batch sizes lead to fast fluctuations of the loss between epochs. After optimization we arrived at 100 or 500 for the training of the final networks using phase scan data and single shot data respectively. A flowchart of the optimization method for learning rate and batch size can be seen in Figure 19.

The specifics of the structure of the network went through some larger variations. Starting from 5 layers of 20 fully connected nodes, it iteratively developed to a final 5 layers with 80 nodes in each for the phase scan version of the data. For the single shot method, a more complex structure of 160-160-80-80-40-40-80-80-160-160 into a final 3 node output layer was found to perform well. A similar network to this was used for the SNS predictions, although with only 2 output channels as input beam energy was not predicted.

The XGBoost decision tree ensembles similarly had a few hyperparameters to tune. These were mainly the number trees to train, similar to choosing the epoch number for the ANN; the depth each tree would be allowed to run, i.e. how many decisions to each tree; and learning rate. XGBoost is also limited to a single scalar output, so separate networks were used for each output channel. The number of trees was often automated by an early stopping system, where if generalized performance did

not improve within a given number of trees, training would stop. This was however not always utilized, as training until such convergence could be very computationally heavy and the workstation performing the training often ran out of working memory during the process. In the end, an ensemble network consisting of $5 \cdot 10^4$ trees was used for predictions of RF amplitude and phase, and one with 1000 trees was used for the input beam energy predictions. The trees in the ensembles predicting RF phase and amplitude were allowed a maximum of 10 decisions, whilst 15 were allowed for energy predictions. These ensembles were all trained with a learning rate of 0.00001.

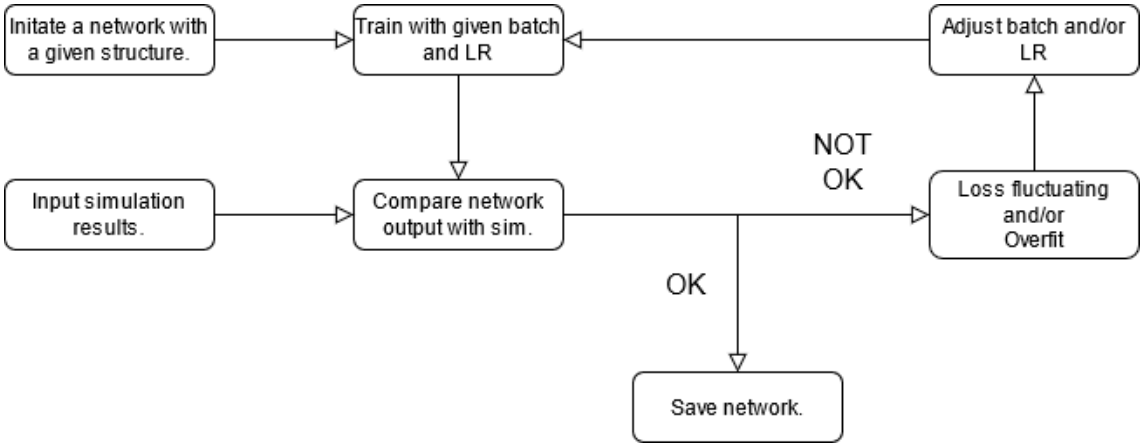

Figure 19: Flow diagram of the hyperparameter optimisation for the batch size and learning rate (LR).

7 Results

7.1 ESS DTL Tuning: RF Phase Scan

Section 7.1.3 shows the final results achieved using the RF phase scan method. Sections 7.1.1 and 7.1.2 below will demonstrate how these values are reached, showing raw predictions and histograms of prediction errors.

7.1.1 Machine Learning

Below in Fig. 20 the final predictions of the RF amplitude and phase using the ML network with phase scan data can be seen. These predictions were done using case B. These figures show the ML predictions against the correct values from the dataset simulated using the as-built lattice, predictions for RF amplitude on the left and phase on the right. With these axes, the perfect model would produce a simple slope one line, and any spread from such a line, which is plotted along with the predictions as dots, are thus errors. Also plotted is the limits set on accuracy, the 1%/1° limit, in dashed green lines. We can see a larger spread in amplitude, but both predictions appear to stay well within the given limits.

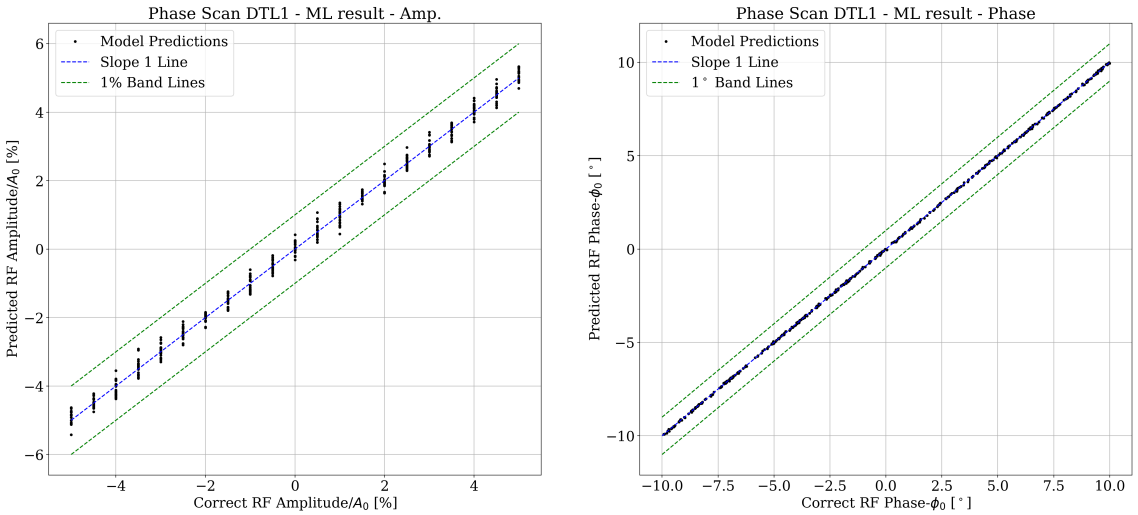

Figure 20: Predictions by the ML model for phase scans plotted against the correct values for RF amplitude and phase, along with the perfect prediction in form of a slope one line and the limit band lines.

Numerical values for the spread in the plots in 20 are reached by subtracting from each prediction of the model the correct setpoint given in simulation, thus resulting in the error of each prediction. These errors, for RF amplitude and phase, are plotted in histograms in Fig. 21 below. A normal distribution can be fitted to these, as is plotted in black, and the 3σ of this distribution must then fall within the given limits. In the figure, the 3σ is shown with green vertical lines and the limits are shown in red.

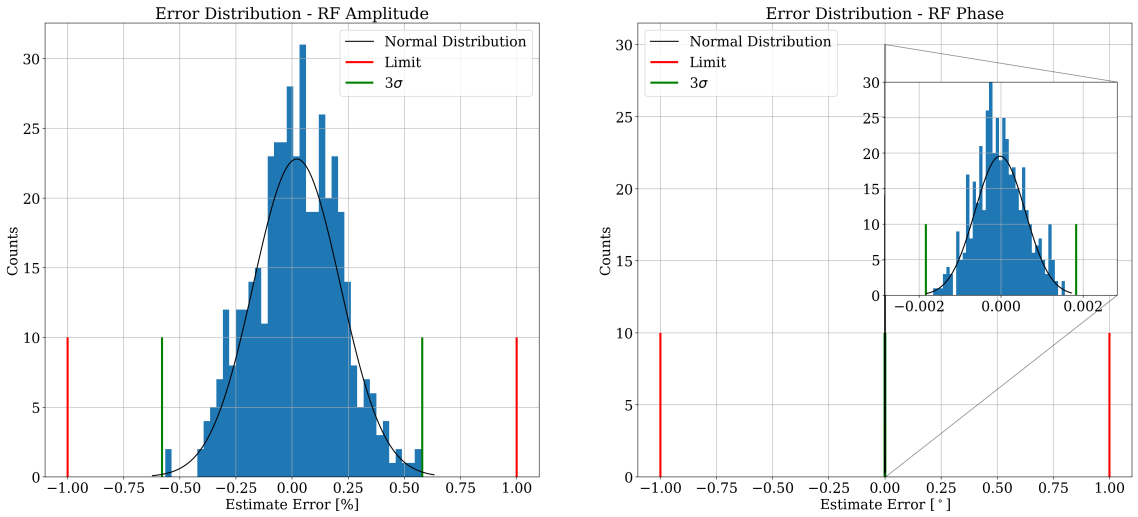

Figure 21: Histograms of the prediction errors of the RF amplitude and phase predictions using ML, with amplitude predictions on the left and phase predictions on the right. Along with these are also plotted a fitted normal distribution, the 3σ of this distribution, and the given limits on accuracy.

7.1.2 Signature Matching

Figure 22 below shows all the different standard deviations in prediction errors arising from the SM methodology, produced using case B. This figure is included mainly to motivate the choice of BPM combinations to contribute to the final results. Here we can see how the σ is reduced in the BPM combinations 4-3, 4-2, and 3-1, potentially caused by the resulting phase curves from these comparisons being best fitted by the polynomial $f(\phi, A, \epsilon)$ used. The mean value of these three is then used as the final results. This is an empirically motivated sampling, which may need revising for operational use.

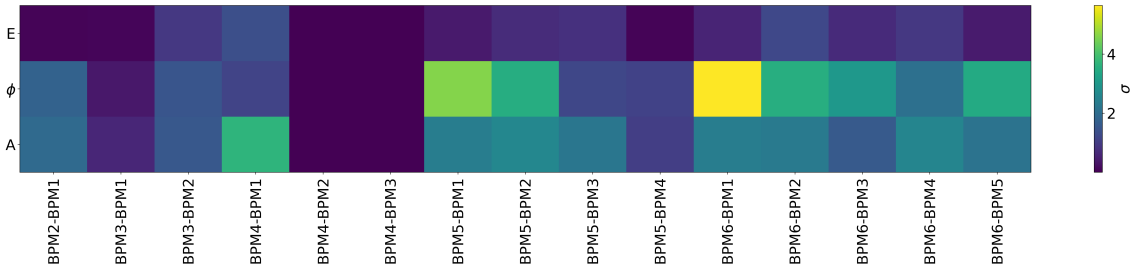

Figure 22: Heatmap of the σ of amplitude, phase and energy estimates for each BPM combination using SM.

To exemplify the behaviour of the signature matching method we used the results from BPM combination 4-3 to produce plots similar to Figures 20 and 21, also using case B for the simulations and testing. These example results are shown in Figures 23 and 24. In Fig. 23, the slope one line once again represents the perfect prediction. We see an increase in accuracy on the RF amplitude and a decrease in accuracy on the phase as compared with Fig. 20. The histograms in Fig. 24 were produced much the same, and show similar results, although the decrease in phase prediction accuracy is much clearer here.

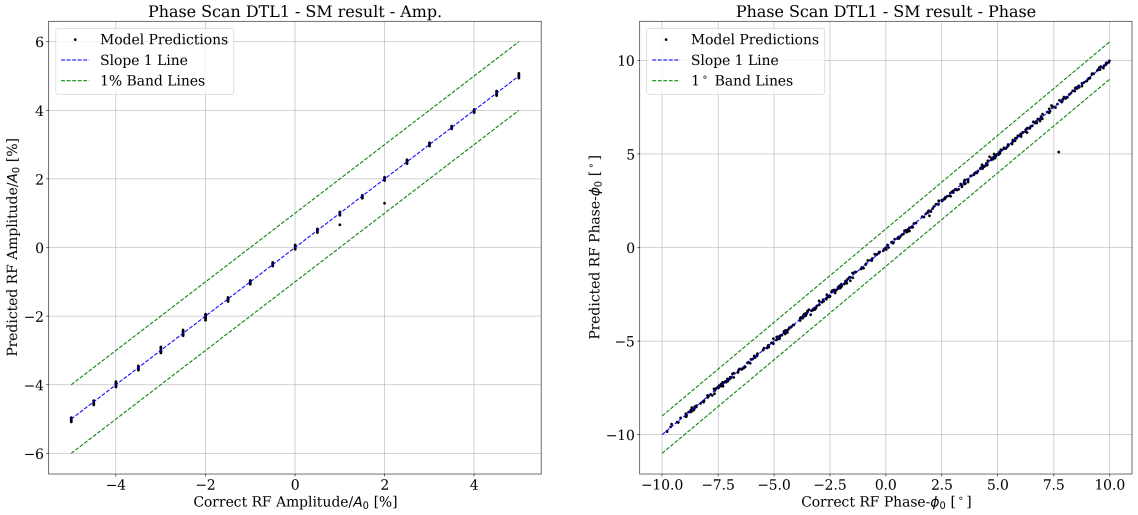

Figure 23: Predictions of the SM method plotted against the correct values for RF amplitude and phase, along with the perfect prediction in form of a slope one line and the limit band lines.

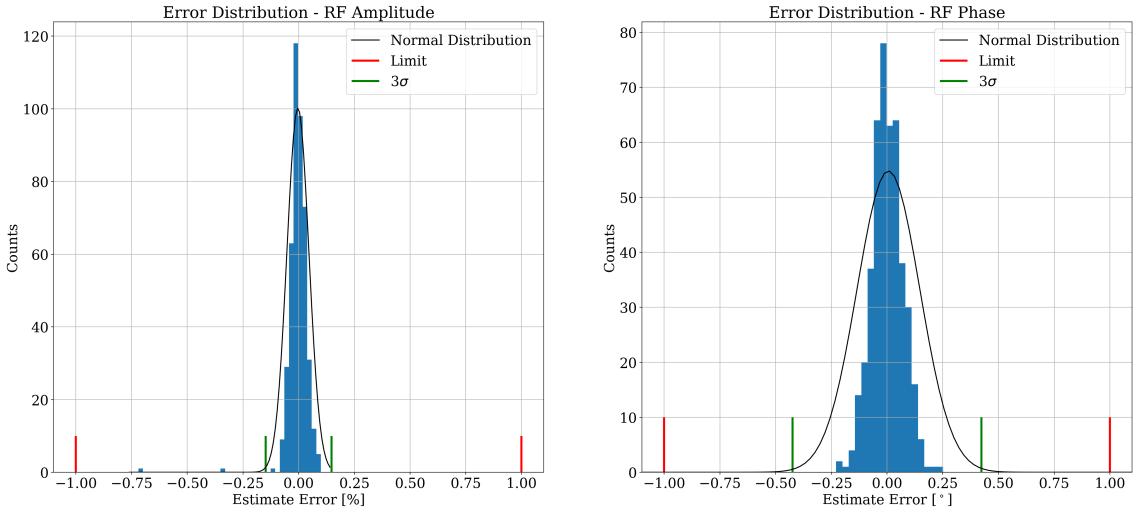

Figure 24: Histograms of the prediction errors of the RF amplitude and phase predictions using SM, with amplitude predictions on the left and phase predictions on the right. Along with these are also plotted a fitted normal distribution, the 3σ of this distribution, and the given limits on accuracy.

7.1.3 Summary and Discussion

Below in Tables 2 and 3 the results from the RF phase scan tuning techniques are shown, Table 2 showing the results on the original design lattice with errors (case A), and Table 3 showing the results on the as-built lattice with errors (case B). The different errors applied can be seen in Table 1 in Section 6.2. These results are presented as 3σ to ensure final estimates fall with a 99.7% certainty within the 1%/1° limit. These are final results achieved using the phase scan methodology and data. In both tables we can see how ML performs nearly perfectly on the training dataset without errors and SM still has some inaccuracy even here. This is to be expected, as ML is based on

iteratively improving this row of results until near perfection, and it is the networks more generalized performance which should be tested. SM will have some inaccuracy in predictions even on the dataset without errors, as it must use a general fitting function $f(\phi, A, \epsilon)$ which may not always differentiate well enough between phase scans. The low mean of the errors in predictions, μ, in all rows shows there is little to no systematic offset to the predictions.

In the last two rows of Table 2 we can see the most relevant results for generalized performance using the initial error estimates, i.e. case A. For ML, we can see that phase predictions are still well within the given limit, but the RF amplitude and input beam energy predictions have rather poor accuracy. It appears the design requirements gave a too wide span of potential errors for these methods to handle, as suggested by the substantial improvement shown in the case B results in Table 3. This new lattice renders the old results somewhat redundant, as the errors of the tank are now well known, but the previous results still serve as an indication of the limitations of the methods. In the first two rows of Table 3 we see similar results for the design lattice, but in the last two rows we see both methods performing within the given limits. This achieves the goals of this project. While SM is still somewhat more accurate, which is impressive considering the use of an operating tank, we must consider the advantage covered in chapter 4.4, that ML requires no specific choice of BPM combination. Thus it could be more flexible to changes in the machine state. Furthermore, both its training time and the time for a trained network to perform predictions outperforms SM.

Table 2: Final results for case A. Three standard deviations in the difference between predicted and correct values for the RF amplitude and phase and the input energy. Results shown for both the ML and SM methodology.

Data Set		$3\sigma_A$ [%]	$3\sigma_\phi$ [°]	$3\sigma_E$ [%]	μ_A [%]	μ_ϕ [°]	μ_E [%]
No Errors	ML	0.039	0.000	0.018	0.001	0.000	0.002
	SM	0.543	0.462	0.111	0.008	0.001	0.001
All Errors	ML	1.815	0.006	1.530	0.022	0.000	0.013
	SM	0.879	1.671	0.456	-0.004	0.010	0.001

Table 3: Final results for case B. Three standard deviations in the difference between predicted and correct values for the RF amplitude and phase and the input energy. Results shown for both the ML and SM methodology.

Data Set		$3\sigma_A$ [%]	$3\sigma_\phi$ [°]	$3\sigma_E$ [%]	μ_A [%]	μ_ϕ [°]	μ_E [%]
No Errors	ML	0.035	0.000	0.020	0.005	0.000	0.002
	SM	0.123	0.443	0.549	0.000	0.000	-0.004
All Errors	ML	0.581	0.002	0.493	0.022	0.000	0.019
	SM	0.118	0.264	0.112	0.001	0.001	-0.001

7.2 ESS DTL Tuning: Single Shot Measurement

The following sections will show the final results of the single shot tuning methodology, both using the Artificial Neural Network and eXtreme Gradient Boosting ML

methods. The structure is much the same as 7.1, 7.2.1 and 7.2.2 showing raw predictions and error histograms, followed by summarizing tables and discussion in 7.2.3. Results for figures were produced using case B.

7.2.1 Artificial Neural Network

In Fig. 25 the predictions of RF amplitude and phase performed by the ANN using the single shot data can be seen, spreading from the perfect slope one line. Figure 25, as well as Fig. 26, were produced using case B. Here we can already see some spread outside the given limits, which will also be reflected in the error histogram below.

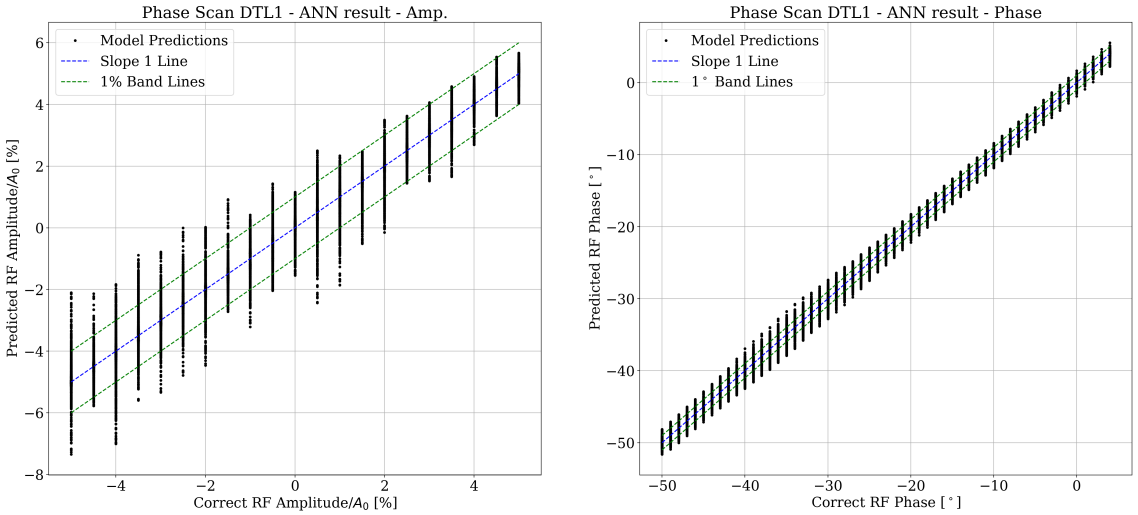

Figure 25: Predictions of the ANN method on single shot data plotted against the correct values for RF amplitude and phase, along with the perfect prediction in form of a slope one line and the limit band lines.

Figure 26 shows the distribution of prediction errors for RF amplitude and phase arising from the ANN. One can see the 3σ lines in both cases remain outside the red $1\%/1^\circ$ limits.

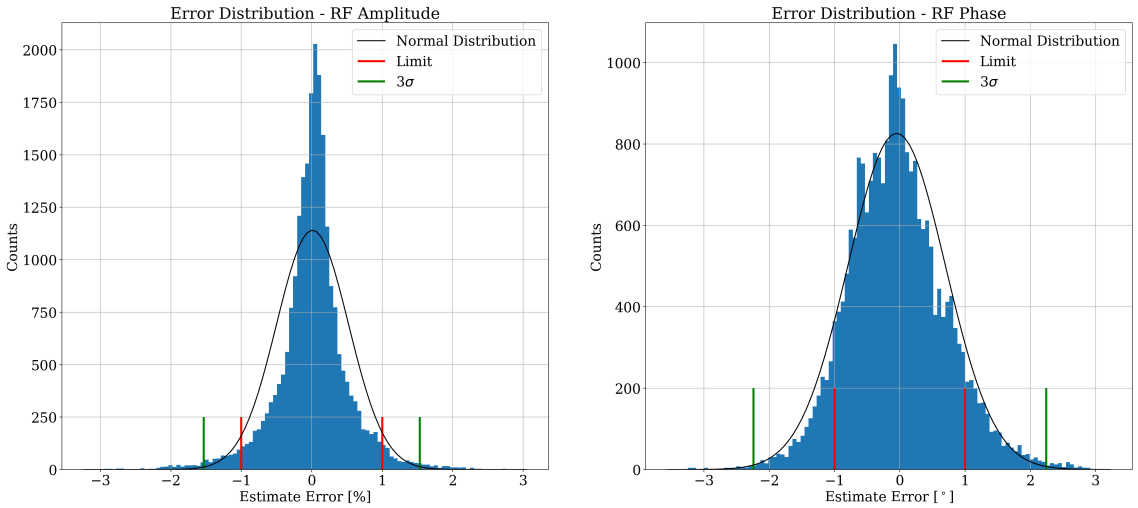

Figure 26: Histograms of the prediction errors of the RF amplitude and phase predictions using the ANN, with amplitude predictions on the left and phase predictions on the right. Along with these are also plotted a fitted normal distribution, the 3σ of this distribution, and the given limits on accuracy.

7.2.2 Extreme Gradient Boosting

Below in Fig. 27 we can see the predictions of RF amplitude and phase performed by the XGBoost ML method on the single shot data. These predictions were also performed on case B. We can see a smaller spread as compared to the ANN predictions in Fig. 25, although one may notice the systematic error in the amplitude predictions. This is clearer in the histogram and tables to follow.

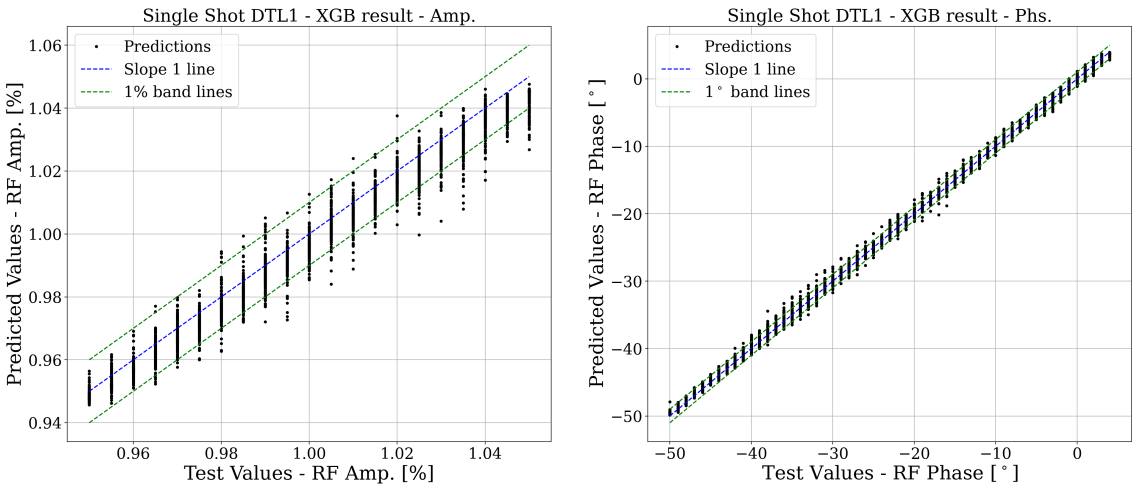

Figure 27: Predictions of the XGBoost method on single shot data plotted against the correct values for RF amplitude and phase, along with the perfect prediction in form of a slope one line and the limit band lines.

In Fig. 28 we can see the prediction errors present in the XGBoost results. Once again, the 3σ lines fall just outside the 1%/1° limits. Here the systematic offset can also be clearly seen in the RF amplitude predictions, shifting the histogram to more positive values.

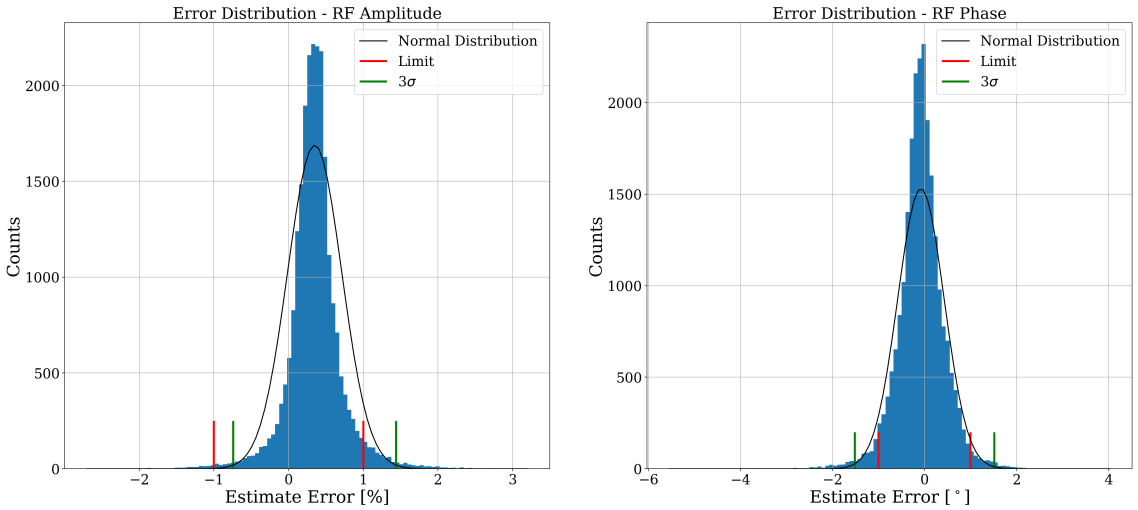

Figure 28: Histograms of the prediction errors of the RF amplitude and phase predictions using the XGBoost method, with amplitude predictions on the left and phase predictions on the right. Along with these are also plotted a fitted normal distribution, the 3σ of this distribution, and the given limits on accuracy.

7.2.3 Summary and Discussion

In Tables 4 and 5 the final results of the novel single shot methodology can be seen, Table 4 showing the results of case A and Table 5 showing the results of case B. In both tables we have the 3σ values for each of the three predictions, as well as the mean μ, for both ML methods, on data with no errors and all errors. Table 2 shows some rather poor predictions, from both methods. In the last two rows we can see that both methods are well outside the 1%/1° limits. The deviations between the different single shot signatures are simply too large, and probably too effected by the inclusion of errors, for these specific models to handle. The high errors in the XGBoost training results suggest this ensemble may have needed further training, but as the ANN can not handle the errors despite near perfect performance on error free data, it could indicate the design lattice errors are simply too significant for ML to manage predictions within the given limitations.

Table 4: Final results for case A. Three standard deviations in the difference between predicted and correct values for the RF amplitude and phase and the input energy. Results shown for both ANN and XGBoost ML methods.

Data Set		$3\sigma_A$ [%]	$3\sigma_\phi$ [°]	$3\sigma_E$ [%]	μ_A [%]	μ_ϕ [°]	μ_E [%]
No Errors	ANN	0.075	0.051	0.045	0.002	0.000	0.002
	XGB	0.891	1.755	0.153	0.025	-0.013	0.002
All Errors	ANN	4.002	5.568	0.804	-0.013	-0.068	-0.009
	XGB	3.171	5.217	0.750	0.022	0.038	0.016

Table 5 however shows somewhat more optimistic results for case B. Here we seen an increase in accuracy from both methods as the included errors have decreased. While they still fail to fall within the 1%/1° limits, they are not far off, XGBoost performing especially well. There is however a systematic offset μ to the XGBoost pre-

dictions, both in RF amplitude and beam input energy. The offset to the beam input energy prediction appears only once errors are included, suggesting one or multiple of the errors is the origin of this offset. The RF amplitude prediction offset is there already in the training on data without errors, suggesting this offset is inherent to the trained ensemble. Thus some fundamental change to the hyperparameters and re-training would be necessary to eliminate this offset. Attaining accuracy within the 1%/1° limits could perhaps also be accomplished with further training, or with some decrease to the inherent errors included in the data.

Table 5: Final results for case B. Three standard deviations in the difference between predicted and correct values for the RF amplitude and phase and the input energy. Results shown for both ANN and XGBoost ML methods.

Data Set		$3\sigma_A$ [%]	$3\sigma_\phi$ [°]	$3\sigma_E$ [%]	μ_A [%]	μ_ϕ [°]	μ_E [%]
No Errors	ANN	0.102	0.063	0.071	-0.002	0.000	0.000
	XGB	0.369	0.492	0.153	0.339	-0.055	0.002
All Errors	ANN	1.535	2.242	0.421	-0.007	-0.049	0.011
	XGB	1.091	1.514	0.581	0.349	-0.079	0.153

7.3 SNS DTL Tuning

Figure 29 displays two figures showing data from SNS, the left one showing the simulated data of phase scans, the right one showing the measured data provided. Along with the five datasets provided were the respective setpoints to the klystron. The machine learning method developed in this project was trained using the simulated data and then used to predict the setpoints in RF amplitude for each provided klystron setpoint. The relation between each RF amplitude and its corresponding klystron setpoint was used to calculate a calibration coefficient which could be used to set the RF amplitude at the design value or some other sought value. This calibration factor was calculated to be 5.7±0.1 kV/m, and the network also predicted an ideal phase of -106±1°. This matches somewhat well with values provided later form SNS, which using their own techniques reached a calibration factor of 6.15 kV/m and an ideal phase of -115.3° [19].

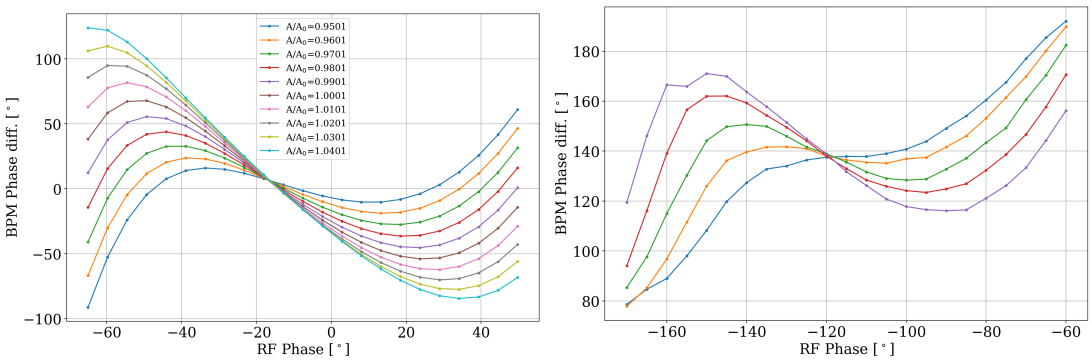

Figure 29: Simulated and real data from the SNS, the left one showing the simulated data of phase scans, the right one showing the measured data provided from SNS.

8 Outlook

In this project the use of machine learning for the tuning of the ESS DTL was investigated. Simulations using the accelerator development environment OpenXAL were used to produce datasets both with and without errors. These have been used to train artificial neural networks for the task of tuning the ESS DTL1. These networks have shown to be able to perform within the given limits on accuracy for an acceptable RF tuning method, performing well as compared with established methodology in the form of signature matching. Since machine learning showed comparable accuracy to signature matching in tuning the RF the natural next step is to integrate it for use in the control room on a more routine basis. With this achieved, it would be possible to implement machine learning to improve the tuning process at ESS, saving time and resources for the work at the neutron instruments. A brief study was performed of machine learning's performance in tuning a separate, working machine with data provided from SNS, and the method was also found to perform well in this new system. Beyond these results, a new method has also been tested, which could allow for the tuning process to be performed using data produced from a single pass of the beam through the machine. While the two machine learning methods tested for this type of data were not able to perform within given limits, it still shows promising results for future developments.

The methods using RF phase scans is where both methods tested performed within the limits with the case B lattice errors. While this was achieved, both methods could be improved as future work. Signature matching is heavily dependent on its basic fitting function $f(A, \epsilon, \phi)$, and in the case of this project, this was set to a high-order polynomial. This fitting function was chosen empirically and has shown to perform well, but it is still a non-physical basis for the method. One clear avenue of improvement could be to investigate analytically the relation between the setpoints (A, ϵ, ϕ) and the final BPM phase. If one would arrive at a somewhat simple function with variable coefficients, this analytical model could be used as the underlying function for the signature matching method. This could improve fittings such as seen in Fig. 18, and perhaps also the final accuracy of the method.

On the machine learning side, while something quite novel has been achieved with this tuning, the current method still includes some drawbacks common to machine learning. The developed network is rather limited to the exact data format used in simulations, that is, using all BPMs in DTL1 and scanning exactly 55 points in RF phase about the optimal point, which will not be known during tuning. Lacking any of the BPMs or scanning on fewer data points will not allow the network to function. In fact, this is partly what motivated the development of the single shot tuning. These issues, while acceptable in this proof-of-concept scenario, can easily become problematic in an operating situation. The need for an exact structure of the input data can be solved in a few different ways. One is to fill in missing data points through swift simulation, or, as the phase scan curves are commonly smooth, simply averaging neighboring points. Investigations should be done as to how fast simulations could fill in a scan or how well the network performs with averaged data points. Otherwise, a new network could simply be trained with whatever data format will become the standard for real tuning scans. As for the issue of missing diagnostics, BPMs can breakdown or become unreliable with time. This is not as easily handled by simulation or averaging, but once more, a new network is always an option.

As for the single shot tuning, the accuracy is still not quite within the given limits. Extensive training was performed, but machine learning is an expansive and rapidly growing field, and it is very possible that a more fitting solution to this problem may exist, or appear. This method does not suffer any issue with input depending on how data is collected in the control room, but can in reality be run continuously as soon as beam is moving through DTL1. Problematic or missing diagnostics is of course still an issue, but could now more simply be solved using simulation, as only a few data points will be missing with each network input, rather than entire scans. This method could also be easier to develop to include more of the entire ESS DTL, as no scans are performed of individual cavities. One would simply append more BPM phase differences to each input and append more output channels to the network or ensemble for the new cavities setpoints.

A future project will be necessary to expand the machine learning methods to the rest of the ESS linac, wherever highly accurate RF tuning is necessary. This would include setting up new networks for the remaining accelerating tanks of the ESS DTL, and other sections.

This concludes this project. As mentioned, the goals in accuracy was achieved machine learning method, and novel techniques using machine learning were investigated. These results and their connected methods could prove very important for the future of the ESS project, and may lead to further development of the tuning techniques at accelerators in general.

List of Acronyms

(A)NN (Artificial) Neural Network.

BPM Beam Position Monitor.

DTL Drift Tube Linac.

ESS European Spallation Source.

Linac Linear Accelerator.

ML Machine Learning.

RF Radio Frequency.

SM Signature Matching.

SNS Spallation Neutron Source.

XGBoost eXtreme Gradient Boosting.

References

[1] Roland Garoby et al. "The European Spallation Source Design". In: *Physica Scripta* 93.1 (Dec. 2017), p. 014001. DOI: 10.1088/1402-4896/aa9bff. URL: https://doi.org/10.1088/1402-4896/aa9bff.

[2] T. L. Owens et al. "Phase scan signature matching for linac tuning". In: *Part. Accel.* 48 (1994), pp. 169–179.

[3] Shen Guo-Bao and Masanori Ikegami. "Tuning of RF amplitude and phase for the drift tube linac in J-PARC". In: *Chinese Physics C* 33.7 (July 2009), pp. 577–582. DOI: 10.1088/1674-1137/33/7/014. URL: https://doi.org/10.1088/1674-1137/33/7/014.

[4] M. Ikegami et al. "RF Amplitude and Phase Tuning of J-PARC DTL". In: *Conf. Proc. C* 070625 (2007). Ed. by C. Petit-Jean-Genaz, p. 1481. DOI: 10.1109/PAC.2007.4440796.

[5] C. Emma et al. "Machine learning-based longitudinal phase space prediction of particle accelerators". In: *Phys. Rev. Accel. Beams* 21 (11 Nov. 2018), p. 112802. DOI: 10.1103/PhysRevAccelBeams.21.112802. URL: https://link.aps.org/doi/10.1103/PhysRevAccelBeams.21.112802.

[6] R. Wideröe. "Über ein neues Prinzip zur Herstellung hoher Spannungen". In: *Archiv f. Elektrotechnik* 21 (1928), pp. 387–406. DOI: 10.1007/BF01656341. URL: https://link.springer.com/article/10.1007/BF01656341.

[7] T.P. Wangler. *RF Linear Accelerators*. John Wiley & Sons, Ltd, 2008. DOI: 10.1002/9783527623426.

[8] Norbert R Holtkamp. "STATUS OF THE SNS* LINAC: AN OVERVIEW". In: 2004.

[9] Luis W. Alvarez et al. "Berkeley Proton Linear Accelerator". In: *Review of Scientific Instruments* 26.2 (1955), pp. 111–133. DOI: 10.1063/1.1771253. eprint: https://doi.org/10.1063/1.1771253. URL: https://doi.org/10.1063/1.1771253.

[10] *Internal Communication, Emelie Nilsson, ESS*.

[11] Dong-o Jeon. "Comparison of Techniques for Longitudinal Tuning of the SNS Drift Tube Linac". In: *Proceedings of the 2005 Particle Accelerator Conference* (2005), pp. 2616–2618.

[12] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. http://www.deeplearningbook.org. MIT Press, 2016.

[13] Diederik P. Kingma and Jimmy Ba. *Adam: A Method for Stochastic Optimization*. 2017. arXiv: 1412.6980 [cs.LG].

[14] Tianqi Chen and Carlos Guestrin. "XGBoost". In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (Aug. 2016). DOI: 10.1145/2939672.2939785. URL: http://dx.doi.org/10.1145/2939672.2939785.

[15] Auralee Edelen et al. "Neural Networks for Modeling and Control of Particle Accelerators". In: *IEEE Transactions on Nuclear Science* 63 (Apr. 2016), pp. 878–897. DOI: 10.1109/TNS.2016.2543203.

[16] Xiaoying Pang. "Advances in Proton Linac Online Modeling". In: *6th International Particle Accelerator Conference*. 2015, WEXC2. DOI: 10.18429/JACoW-IPAC2015-WEXC2.

[17] J. Galambos et al. "XAL Application Programming Structure". In: *Proceedings of the 2005 Particle Accelerator Conference*. Knoxville, TN, USA: IEEE, 2005, pp. 79–83. ISBN: 9780780388598. DOI: 10.1109/PAC.2005.1590365. URL: http://ieeexplore.ieee.org/document/1590365/ (visited on 09/18/2019).

[18] *Open XAL*. 2003. URL: http://openxal.github.io/ (visited on 09/18/2019).

[19] *Personal Communication, Andrei Shishlo, SNS*.

A IPAC'21 Paper

The results of the comparison between the early performance of ML as applied to the RF phase scans and the traditional signature matching method were presented at the 2021 International Particle Accelerator Conference (IPAC'21), held remotely from Brazil in May. The proceedings paper submitted to the conference is included below.

ESS DTL TUNING USING MACHINE LEARNING METHODS

J. S. Lundquist*, S. Werin, Lund University, Lund, Sweden
N. Milas, E. Nilsson, ESS, Lund, Sweden

Abstract

The European Spallation Source, currently under construction in Lund, Sweden, will be the world's most powerful neutron source. It is driven by a proton linac with a current of 62.5 mA, 2.86 ms long pulses at 14 Hz. The final section of its normal-conducting front-end consists of a 39 m long drift tube linac (DTL) divided into five tanks, designed to accelerate the proton beam from 3.6 MeV to 90 MeV. The high beam current and power impose challenges to the design and tuning of the machine and the RF amplitude and phase have to be set within 1% and 1° of the design values. The usual method used to define the RF set-point is signature matching, which can be a time consuming and challenging process, and new techniques to meet the growing complexity of accelerator facilities are highly desirable. In this paper we study the usage of Machine Learning to determine the RF optimum amplitude and phase. The data from a simulated phase scan is fed into an artificial neural network in order to identify the needed changes to achieve the best tuning. Our test for the ESS DTL1 shows promising results, and further development of the method will be outlined.

INTRODUCTION

The European Spallation Source (ESS) is a state of the art neutron science facility under construction in Lund, Sweden [1]. The basic process used by the facility is spallation, wherein one impinges a high neutron material, in this case Tungsten, with high energy protons, causing the target to shed excess neutrons. The high energy protons are provided by the ESS linear accelerator (linac), a 600 m long accelerator consisting of many different sections utilizing varied accelerator technologies in order to raise the proton energy from the 75 keV source output to the final 2.0 GeV arriving on the target. A crucial part of this machine is the 39 m long drift tube linac (DTL) divided into five tanks, designed to accelerate the proton beam from 3.6 MeV to 90 MeV. As the machine is expected to deliver beam of high current and power, a primary concern is to avoid slow beam losses, as these lead to radiation activation of surrounding equipment. In order to avoid such losses, proper and careful tuning of the RF fields is crucial. As a result the requirement for accuracy of the RF set point is to be within 1% in RF amplitude and 1° in phase [1]. In order to achieve this type of accuracy, much work has been performed in the last decades to develop new techniques to meet the growing scale and complexity of facilities [2–4]. Within this paper we will investigate how Machine Learning (ML) may serve this purpose. This paper presents our current strides in the development of a tuning technique using ML and a comparison of our current results with those of the Signature Matching (SM) technique, a more established methodology for RF tuning [1–4].

* johan.lundquist@ess.eu

RF PHASE SCAN

In order to be able to quantify how the beam responds to changes to the RF set-point, a diagnostic sensitive to the beam time of flight through the cavity must be used. For those cases a Beam Position Monitor (BPM) can be used. As the beam passes a BPM, not only is the transverse position measured, but also the amplitude and phase of the fields excited by the passing beam on the BPM sensor. Although this phase alone doesn't hold much information, by comparing two BPM phases we can get a fast measurement which is proportional to the time-of-flight, or looking with respect to acceleration in a RF cavity, the energy gain between the two devices. It is important to stress that this measurement is relative and that extracting the absolute values of the energy is not an easy task. For this technique, using only the relative phase changes has proven to be enough.

In order for clear signatures to emerge, it is not uncommon to use subsequent unpowered accelerating structures as drift space for the beam, and to extract the relative phase change from BPMs in these locations [2]. It simplifies tuning when the energy is not varying between BPMs, as in a non-accelerating cavity, but it is not a requirement. Initial commissioning of the ESS DTL will be done with only the first DTL tank, and as such it is required to reach the RF field requirements using only the BPMs inside this tank for tuning. Thus the relative phase change needed to be extracted from internal BPM pairs within the accelerating structure for the results in this article.

The BPMs on this structure are then used to measure the energy gain (or time-of flight) as a function of the set points in the accelerating cavity. As the BPM's measured phase is closely dependent on the energy of the beam, scanning RF amplitude and phase in a cavity and plotting out the resulting phase differences will give rise to different curves depending on the proximity to the ideal set point for the cavity. A few of these signature curves can be seen in Fig. 1, where the ideal set point can be found from the signature for the ideal amplitude A_0 = 6.89 kV, the ideal input beam energy E_0 = 3.62 MeV and the -35° offset phase set point.

Much work has been done to find efficient and fast methods for extracting and identifying this ideal signature.

Simulations

OpenXAL was used to simulate the first tank of the ESS DTL during acceleration and to reproduce the signals from the six BPMs [5, 6]. As phase difference is the data of interest, this results in 15 different BPM combinations, each producing a phase scan for each amplitude set point of the

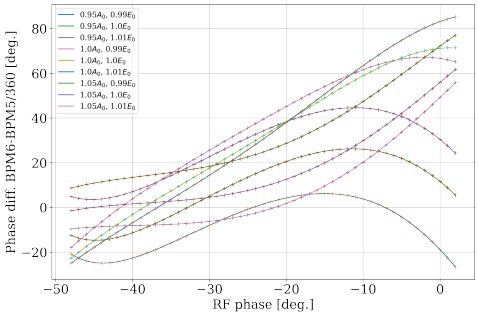

Figure 1: The phase curves for different RF amplitude and input energy set points. BPM phases simulated as comparison between two BPMs in the first DTL tank in the ESS linac.

cavity, which was varied in RF amplitude and input beam energy. Both techniques outlined below require training or fitting of some type and for this purpose a large data set was produced for a machine free from errors. This consisted of 110 different amplitude set points, with a variation of $\pm 5.5\%$ around the design RF cavity amplitude A_0, and 60 different input beam energy set points, with a variation of $\pm 1.5\%$ around the design input energy E_0. Each phase scan consisted of 55 phase points, spread evenly around the -35° set point.

To this perfect machine four different types of errors were then applied. BPM longitudinal position within the machine was adjusted, potentially caused by installation and construction, as well as the phase readout from these BPMs, produced by electronic limitations. There are also errors arising from production limitations when constructing the cavities. Such limitations could give rise to errors in both RF amplitude and phase gap-to-gap. The different types of errors and their magnitudes are summarized in Table 1.

Table 1: The Different Types of Errors Used in Simulations and Their Corresponding Magnitude

Error Type	Magnitude
BPM Δs	$\pm$ 100 µm
BPM $\Delta\phi$	$\pm 1°$
RF Amplitude	$\pm 2\%$
RF Phase	$\pm 0.5°$

SIGNATURE MATCHING

One method, which is widely used [2–4], is to use a set of simulated data (like the ones shown in Fig. 1) and fit each phase scan set with RF amplitude A and input beam energy ϵ with a polynomial

$$f(\phi, A, \epsilon) = a_0(A, \epsilon) + a_1(A, \epsilon)\phi + ... + a_n(A, \epsilon)\phi^n \quad (1)$$

and at this point we assume that an error in the input phase of the cavity acts as a simple offset on the variable ϕ of the whole curve. As a result from all the fittings, it is possible to construct a surface $a_n(A, \epsilon)$ for each coefficient of the fitted polynomial. This surface can then be approximated by another 2D polynomial surface so that we have a continuum of coefficients for any given set-point in (A, ϵ).

With the parameters obtained from the fits described above, it is possible to investigate the variance, χ^2, between the simulated model and the corresponding measured data. This variance is defined as

$$\chi^2(\phi_0, A, \epsilon) = \frac{\sum_j^N (f(\phi_j - \phi_0, A, \epsilon) - W_j)^2}{N} \quad (2)$$

wherein f is our fit to the model prediction of the BPM phase difference (Eq. (1)) and W_j is the measured BPM phase difference at some unknown RF phase, RF amplitude and incoming beam energy. One then optimizes the values of ϕ_0, A and ϵ to minimize Eq. (2), thus determining to which set-point the signature curve W_j corresponds to. This method is known as Signature Matching [2–4].

MACHINE LEARNING

Machine learning is a term used to describe a specific field of modern computer algorithms capable of learning from experience. Most often, the algorithm refers to a neural network, a network of individual artificial neurons, wherein the weighted connections between the neurons can be trained to reach an ideal output. This training can be performed in different ways, but the most relevant to this project is the technique of supervised learning. Machine learning algorithms come in many forms and can solve many distinct problems using varying network structures, definitions of loss and optimization algorithms. The problem we are looking at in this project involves reducing larger scans of data down to three dependent variables, RF amplitude, RF phase and input beam energy. This leads us to the use of a network and loss function fitting for linear regression. This network is trained using a mean squared error function as loss, and the ADAM optimization algorithm [7, 8].

This network was defined using the python library Keras. The library comes with predefined versions of our loss function, mean squared error, and our optimization algorithm, ADAM. ADAM has different coefficients which may be tuned to improve the networks performance, although the learning rate is most relevant. Optimization of the network structure and training parameters was done iteratively, looking at generalized performance as the figure of merit. This was quantified as the loss on a subset of data separated during training. Through this process we arrived at a 10-layer structure with 80 neurons in each layer, and a final output layer of three neurons. This was trained for 20 000 epochs with a learning rate of 0.00001. This network was used to produce the results presented in the following section [8, 9].

RESULTS

Figure 2 shows the results for both methods when attempting to produce the correct set point in amplitude. The x-axis

Table 2: Difference between the Predicted and Expected Value for the RF Amplitude and Phase and the Input Energy

Data Set		σ_A [%]	σ_ϕ [°]	σ_E [%]	μ_A [%]	μ_ϕ [°]	μ_E [%]
No Errors	ML	0.013	0.000	0.006	0.001	0.000	0.002
	SM	0.181	0.154	0.037	0.008	0.001	0.001
All Errors	ML	0.605	0.002	0.510	0.022	0.000	0.013
	SM	0.293	0.557	0.152	-0.004	0.010	0.001

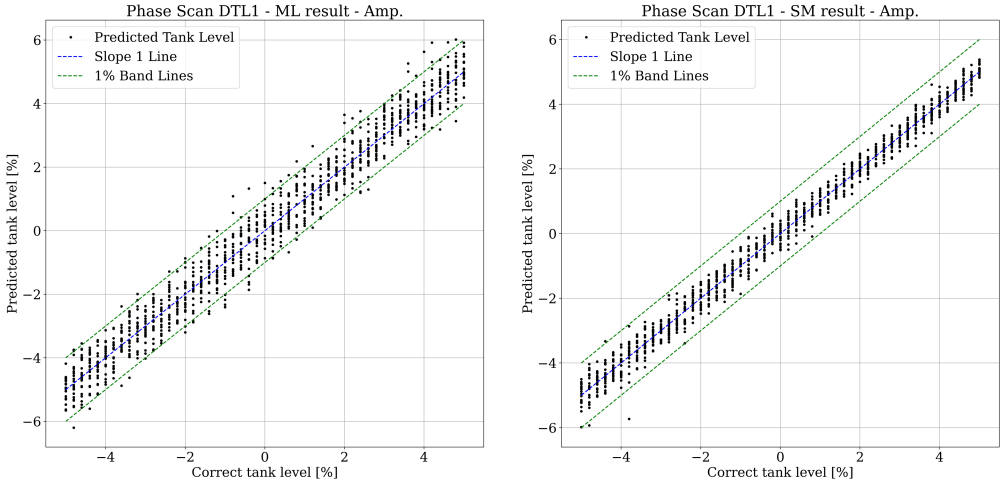

Figure 2: The predicted RF amplitude set points from both methods plotted against the correct amplitudes, along with a slope one line and 1% limits.

represents the correct set points and the y-axis the predictions of the methods. The blue line in Fig. 2 is a slope 1 line which then represents the perfect prediction, and the two green lines represent the $\pm 1\%$ band lines, the minimum amplitude accuracy required for the ESS cavities. Here we can see the spread around the ideal predictions visually, though it is numerically summarized in Table 2.

Table 2 shows the standard deviation (σ) and the mean (μ) of the difference between the predicted and expected value for the RF Amplitude (A) and phase (ϕ), as well as for the input energy (E). The low mean in all rows shows there is little to no systematic offset to the predictions. We see higher accuracy from ML on the training set. If we look at the results with all errors present we see SM performing better on amplitude and energy, while ML performs much better on the phase prediction. It should be noted that the SM predictions take 60 times as long to be calculated as the comparable ML results. Due to the BPM data being extracted from an accelerating cavity, SM is also quite sensitive to which BPM combination is used as input. For the results presented here, the three best combinations (BPM3-BPM1, BPM4-BPM2 and BPM4-BPM3, indexed in order of longitudinal placement) were chosen and averaged.

OUTLOOK

We have managed to produce promising results with a somewhat rudimentary neural network and short training. This suggests that the problem of RF tuning, even with the limitation of using internal BPMs within the accelerating structure, is solvable using ML. In the near future this project will move on to constructing and training more complex neural networks, expanding the scope to include the rest of the DTL, and possibly importing data from other facilities to test this new methodology on real data. The possibility of a hybrid method has also been discussed, using ML to perform the fitting required for SM. This could potentially conserve the advantages of both methods while avoiding their respective drawbacks.

ACKNOWLEDGEMENTS

This study has only been possible with the help of the entire Beam Physics team at ESS, as well as that of other teams at the facility.

REFERENCES

[1] R. Garoby *et al.*, "The European Spallation Source Design", *Physica Scripta*, vol. 93, p. 014001, 2018. doi:10.1088/1402-4896/aa9bff

[2] G. Shen and M. Ikegami, "Tuning of RF amplitude and phase for the drift tube linac in J-PARC", *Chin. Phys C*, vol. 33, p. 577, 2009. doi:10.1088/1674-1137/33/7/014

[3] T. L. Owens, K. L. Junck, T. Kroc, E. S. McCrory, and M. Popovicacute, "Phase Scan Signature Matching for Linac Tuning", in *Proc. 1994 Linear Accelerator Conf. (LINAC'94)*, Tsukuba, Japan, Aug. 1994, paper TH–80, pp. 893–895.

[4] G. B. Shen, H. Sako, and S. Sato, "RF Amplitude and Phase Tuning of J-PARC SDTL", in *Proc. 22nd Particle Accelerator Conf. (PAC'07)*, Albuquerque, NM, USA, Jun. 2007, paper TUPAN062, pp. 1529–1531.

[5] J. Galambos, A. V. Aleksandrov, C. Deibele, and S. Henderson, "PASTA An RF Phase and Amplitude Scan and Tuning Application", in *Proc. 21st Particle Accelerator Conf. (PAC'05)*, Knoxville, TN, USA, May 2005, paper FPAT016, pp. 1529-1531.

[6] OpenXAL, https://openxal.github.io/.

[7] A. Edelen *et al.*, "Opportunities in Machine Learning for Particle Accelerators", 2018. arXiv:1811.03172

[8] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization", 2017. arXiv:1412.6980

[9] Keras, https://keras.io

B IBIC'21 Paper and Poster

The results of the single shot tuning methods, both using an ANN and XGBoost, were presented at the 2021 International Beam Instrumentation Conference (IBIC'21), held remotely from South Korea in September. The proceedings paper submitted to the conference is included below, as well as the poster used for presentation.

MACHINE LEARNING METHODS FOR SINGLE SHOT RF TUNING

J. S. Lundquist*, S. Werin, Lund University, Lund, Sweden
N. Milas, ESS, Lund, Sweden

Abstract

The European Spallation Source, currently under construction in Lund, Sweden, will be the world's most powerful neutron source. It is driven by a proton linac with a current of 62.5 mA, 2.86 ms long pulses at 14 Hz. The final section of its normal-conducting front-end consists of a 39 m long drift tube linac (DTL) divided into five tanks, designed to accelerate the proton beam from 3.6 MeV to 90 MeV. The high beam current and power impose challenges to the design and tuning of the machine and the RF amplitude and phase have to be set within 1% and 1° of the design values. The usual method used to define the RF set-point is signature matching, which can be a challenging process, and new techniques to meet the growing complexity of accelerator facilities are highly desirable. In this paper we study the use of ML to determine the RF optimum amplitude and phase, using a single pass of the beam through the ESS DTL1 tank. This novel method is compared with the more established methods using scans over RF phase, providing similar results in terms of accuracy for simulated data with errors. We also discuss the results and future extension of the method to the whole ESS DTL.

INTRODUCTION

The European Spallation Source (ESS) is a state of the art neutron science facility under construction in Lund, Sweden [1]. The basic process used by the facility is spallation, wherein one impinges a high neutron material, in this case Tungsten, with high energy protons, causing the target to shed excess neutrons. The high energy protons are provided by the ESS linear accelerator (linac), a 600 m long accelerator consisting of many different sections utilizing varied accelerator technologies in order to raise the proton energy from the 75 keV source output to the final 2.0 GeV arriving on the target. A crucial part of this machine is the 39 m long drift tube linac (DTL) divided into five tanks, designed to accelerate the proton beam from 3.6 MeV to 90 MeV. As the machine is expected to deliver beam of high current and power, a primary concern is to avoid slow beam losses, as these lead to radiation activation of surrounding equipment. In order to avoid such losses, proper and careful tuning of the RF fields is crucial. As a result the requirement for accuracy of the RF set point is to be within 1% in RF amplitude and 1° in phase [1]. In order to achieve this type of accuracy, much work has been performed in the last decades to develop new techniques to meet the growing scale and complexity of facilities [2–4]. Within this paper we will investigate how Machine Learning (ML) may serve this purpose. This paper presents our current strides in the development of a tuning technique using ML, with simulated data used in such a way that a single pass through the untuned cavity could be sufficient for setting it up.

* johan.lundquist@ess.eu

RF TUNING

RF Phase Scan

In order to be able to quantify how the beam responds to changes to the RF set-point, a diagnostic sensitive to the beam time of flight through the cavity must be used. For those cases a Beam Position Monitor (BPM) can be used. As the beam passes a BPM both the amplitude and phase of the fields excited on the BPM sensor by the passing beam are recorded. Although this phase alone doesn't hold much usable information for cavity tuning, by comparing two BPM phases we can get a fast measurement which is proportional to the time-of-flight, or looking with respect to acceleration in a RF cavity, the energy gain between the two devices. It is important to stress that this measurement is relative and that extracting the absolute values of the energy is not an easy task. For this technique, using only the relative phase changes has proven to be enough [2–4].

The BPMs are used to measure the energy gain (or time-of flight) as a function of the set points in the accelerating cavity. As the BPM's measured phase is closely dependent on the energy of the beam, scanning RF amplitude and phase in a cavity and plotting out the resulting phase differences will give rise to different curves depending on the proximity to the ideal set point for the cavity. A few of these signature curves can be seen in Fig. 1, where the ideal set point can be found from the signature for the ideal amplitude A_0 = 6.89 kV, the ideal input beam energy E_0 = 3.62 MeV and the -35° phase set point.

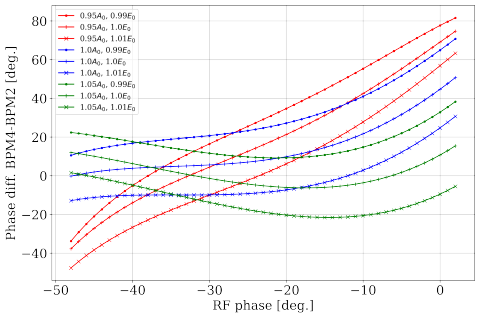

Figure 1: The phase curves for different RF amplitude and input energy set points. BPM phases simulated as comparison between two BPMs in the first DTL tank in the ESS linac.

Identifying these types of signatures is the basis of most established techniques for cavity tuning [2–4]. We are suggesting a new type of identifiable signature for this type of tuning, which could be measured in a single pass through the cavity, rather than requiring scanning over a parameter, described in more detail in the following section.

Single Shot Measurement

RF phase scans are an established and reliable method for extracting the information needed to achieve good tuning, and with a more limited diagnostic output it is the only option available for now. However, with the large number of BPMs within the ESS DTL section, a restructuring of the data can be done such that we can see distinct signatures for each cavity setpoint in amplitude, phase and beam input energy. We look at BPM phase differences, not against RF phase, but against each diagnostic output, the pairing of BPMs. Figure 2 shows an example of this type of plot, where each line represents a cavity set point and is measured in a single pass through the machine, without scanning any parameter. From here we encounter the same problem to be solved as with the phase scan data, needing to accurately identify these new signatures. The nature of the signatures in this data format leaves ML uniquely equipped for the task.

If accurate predictions can be made with this data format, a few new advantages manifest. Being able to tune the machine acceptably with a single shot could cut down set up times. One would also not require to determine a range for the scanned parameters as in more traditional RF tuning [2, 3] but would in principle reload the last machine state with good settings and run a single verification pulse.

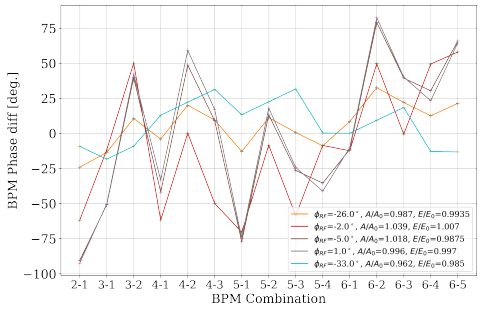

Figure 2: BPM phase differences for each possible BPM coupling, with the different plots each corresponding to a single cavity set point.

Simulations

OpenXAL was used to simulate the first tank of the ESS DTL during acceleration and to reproduce the signals from the six BPMs inside the DTL tank 1 [5, 6]. As phase difference is the data of interest, this results in 15 different BPM combinations, each combination producing one data point for each cavity set point, in RF amplitude and phase and input beam energy. ML requires large amounts of data for training networks and for this purpose an error free dataset was used. This consisted of 110 different amplitude set points, with a variation of ±5.5% around the design RF cavity amplitude A_0, 60 different input beam energy set points, with a variation of ±1.5% around the design input energy E_0, and 55 different phase set points, spread evenly around the -35° design set point.

To this perfect machine four different types of errors were then applied. BPM longitudinal position within the machine was adjusted, potentially caused by installation and construction, as well as the phase readout from these BPMs, produced by electronic limitations. There are also errors arising from production limitations when constructing the cavities. Such limitations could give rise to errors in both RF amplitude and phase gap-to-gap. The different types of errors and their magnitudes are summarized in Table 1.

Table 1: The Different Types of Errors Used in Simulations and Their Corresponding Magnitude

Error Type	**Magnitude**
BPM Δs	±100 µm
BPM $\Delta\phi$	±1°
RF Amplitude	±2%
RF Phase	±0.5°

MACHINE LEARNING

Machine learning algorithms come in many forms and can solve many distinct problems using varying network structures, definitions of loss and optimization algorithms [7]. The problem we are looking at in this project involves reducing larger scans of data down to three dependent variables, RF amplitude, RF phase and input beam energy. We compare two types of network, a traditional linear regression structure, and a newer decision tree boosting model called XGBoost.

Linear Regression Network

This network was defined using the python library Keras [8]. The library comes with predefined versions of our loss function, mean squared error, and our optimization algorithm, ADAM. ADAM has different coefficients which may be tuned to improve the networks performance, although the learning rate is most relevant [9]. Optimization of the network structure and training parameters was done iteratively, looking at generalized performance as the figure of merit. This was quantified as the loss on a subset of data separated during training. Through this process we arrived at a 10-layer structure with an 160-160-80-80-40 symmetrical neural layout, and a final output layer of three neurons. This was trained for 20 000 epochs with a learning rate of 0.00001. This network was used to produce the results presented in the following section.

Table 2: Three standard deviations in difference between predicted and correct values for the RF Amplitude and phase and the input energy. Results shown for both linear regression (LR) and XGBoost (XGB) network structures.

Data Set		$3\sigma_A$ [%]	$3\sigma_\phi$ [°]	$3\sigma_E$ [%]	μ_A [%]	μ_ϕ [°]	μ_E [%]
No Errors	LR	0.075	0.051	0.045	0.002	0.000	0.002
	XGB	0.891	1.755	0.153	0.025	-0.013	0.002
All Errors	LR	4.002	5.568	0.804	-0.013	-0.068	-0.009
	XGB	3.171	5.217	0.750	0.022	0.038	0.016

XGBoost

The modern ML system of XGBoost (eXtreme Gradient Boosting) is an open source gradient boosting model, which has proven extremely powerful for solving varied, nonlinear problems [10]. Gradient boosting tree models such as this are based on the decision tree model of network structures, with a regularized objective function. In a decision tree ML system, the parameters adjusted in training are not the weighted connections within a network of nodes, but rather the branching criteria in a large decision tree. A gradient boosting ML system uses an ensemble of many decision trees in order to improve the final predictions, and commonly a regularized loss function which penalises increasing complexity of the model as well as the usual error of predictions. This regularized loss is then applied to the ensemble of trees iteratively to improve the output by training the branching criteria.

For the results produced here an ensemble of 10000 trees was used, each with a max allowed depth, the amount of branching criteria, of 20. A learning rate factor of 0.0001 was applied and an early stopping system was also used during training, forcing the training to halt if the generalized performance of the ensemble network did not improve over a period of 500 iterations.

RESULTS

Table 2 shows three standard deviations (3σ) and the mean (μ) of the difference between the predicted and expected value for the RF Amplitude (A) and phase (ϕ), as well as for the input energy (E). The low mean in all rows shows there is little to no systematic offset to the predictions. We see both types of network performing within the given limits on the training data set, although we do see higher standard deviation from the XGBoost training. However, it is generalized performance on the realistic error data set which presents the more relevant figure of merit. Here we see good performance in the energy predictions, but both methods fail to produce the sought results in both phase and amplitude. XGBoost only slightly outperforms the more traditional linear regressor, but remains far outside the limit in the phase prediction. The variation in the single shot signatures as a function of phase is quite small, so networks struggling to distinguish between these is understandable.

OUTLOOK

While both methods may fail the limits for operation at this stage of investigation, there are still many factors which could prove this method more reliable than suggested by these results. The error data set produced may have been pessimistic in the predictions of one or many of the factors included. Further optimization of the meta parameters for the training of the networks could reveal better results in the future. New data sets could be produced with more distinguished patterns arising from difference in RF phase.

Furthermore, even if this technique would prove unable to produce the sought results for initial cavity tuning, the 1° and 1% error in RF phase and amplitude predictions, it does not render it useless. The single shot nature of the data could allow for updated tuning information during operation of the machine, as well as long term tracking of drifts on the RF parameters. Also, the data available for each set point increases rapidly with added diagnostics. Applying this single shot data to the full ESS DTL section would include more BPM combinations, and the increase in available data could perhaps be sufficient to improve the method to within the restrictions. Further applications of this online tuning information could be developed in the future, for use in the ESS control room or elsewhere.

ACKNOWLEDGEMENTS

This study has only been possible with the help of the entire Beam Physics team at ESS, as well as that of other teams at the facility.

REFERENCES

[1] R. Garoby *et. al.*,"The European Spallation Source Design", *Physica Scripta*, vol 93, p. 014001, 2018.

[2] G. Shen and M. Ikegami, "Tuning of RF amplitude and phase for the drift tube linac in J-PARC", *Chin. Phys C*, vol. 33, p. 577, 2009.

[3] T.L. Owens, M. B. Popovic, E. S. McCrory, C. W. Schmidt and L.J. Allen, "Phase Scan Signature Matching for Linac Tuning", *Particle Accelerators*, vol. 8, no 1, pp. 169-179, 1994.

[4] G. B. Shen, H. Sako, and S. Sato, "RF Amplitude and Phase Tuning of J-PARC SDTL", in *Proc. 22nd Particle Accelerator Conf. (PAC'07)*, Albuquerque, NM, USA, Jun. 2007, paper TUPAN062, pp. 1529–1531.

[5] J. Galambos, A. V. Aleksandrov, C. Deibele, and S. Henderson, "PASTA An RF Phase and Amplitude Scan and Tun-

ing Application", in *Proc. 21st Particle Accelerator Conf. (PAC'05)*, Knoxville, TN, USA, May 2005, paper FPAT016, pp. 1529-1531.

[6] OpenXAL, `https://openxal.github.io/`.

[7] A. Edelen *et al.*, "Opportunities in Machine Learning for Particle Accelerators", **1811.03172**, arXiv, 2018.

[8] Keras, `https://keras.io/`.

[9] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization", **1412.6980**, arXiv, 2017.

[10] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System", **1603.02754**, arXiv, 2016.

Machine Learning Methods for Single Shot RF Tuning

J. S. Lundquist[1,2], S. Werin[2] and N. Milas[1], [1]ESS and [2]Lund University, Lund, Sweden

IBIC'21 September 13 - 17, 2021

IBIC 2021

Abstract

The European Spallation Source, currently under construction in Lund, Sweden, will be the world's most powerful neutron source. It is driven by a proton linac with a current of 62.5 mA, 2.86 ms long pulses at 14 Hz. The final section of its normal-conducting front-end consists of a 39 m long drift tube linac (DTL) divided into five tanks, designed to accelerate the proton beam from 3.6 MeV to 90 MeV. The high beam current and power impose challenges to the design and tuning of the machine and the RF amplitude and phase have to be set within 1% and 1° of the design values. The usual method used to define the RF set-point is signature matching, which can be a challenging process, and new techniques to meet the growing complexity of accelerator facilities are highly desirable. In this paper we study the use of ML to determine the RF optimum amplitude and phase, using a single pass of the beam through the ESS DTL1 tank. This novel method is compared with the more established methods using scans over RF phase, providing similar results in terms of accuracy for simulated data with errors. We also discuss the results and future extension of the method to the whole ESS DTL.

Introduction

- The European Spallation Source (ESS) is a state of the art neutron science facility under construction in Lund, Sweden [1].
- As the machine is expected to deliver beam of high current and power, a primary concern is to avoid slow beam losses, as these lead to radiation activation of surrounding equipment.
- A simple model of the ESS linac can be seen in Fig. 1.
- **The requirement for accuracy of the RF set point in the DTL is to be within 1% in RF amplitude and 1° in RF phase [1].**

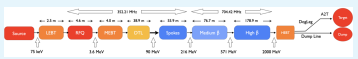

Figure 1: Simplified figure of the ESS linac.

Simulations

- Simulations of DTL Tank 1 done in OpenXAL [5].
- Envelope simulations were used, rather than trajectory tracking, which has proved to be sufficient information for RF tuning.
- ML Networks trained with perfect data sets, with no errors deviating from the lattice design files.
- Networks were then tested with new datasets with the errors shown in Table 2.

Table 1: The different types of errors used in simulations and their corresponding magnitude.

Error Type	Magnitude
BPM Δs	$\pm 100\,\mu$m
BPM $\Delta\phi$	$\pm 1°$
RF Amplitude	$\pm 2\%$
RF Phase	$\pm 0.5°$

RF Phase Scan

- By comparing two BPM phases we can get a fast measurement which is proportional to the time-of-flight.
- It is important to stress that this measurement is relative and that extracting the absolute values of the energy is not an easy task. For this technique, using only the relative phase changes has proven to be enough [2, 3, 4].
- As the BPM's measured phase is closely dependent on the energy of the beam, scanning RF amplitude and phase in a cavity and plotting out the resulting phase differences will give rise to different curves depending on the proximity to the ideal set point for the cavity.
- RF phase was scanned to produce signature curves shown in Fig. 2.

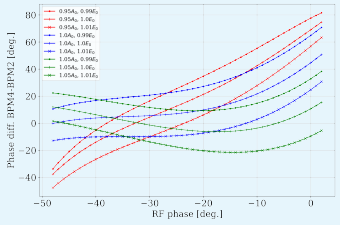

Figure 2: The phase curves for different RF amplitude and input energy set points. BPM phases simulated as comparison between two BPMs in the first DTL tank in the ESS linac.

- Identifying these types of signatures is the basis of most established techniques for cavity tuning [2, 3, 4].

Single Shot Measurement

- With the large number of BPMs within the ESS DTL section, a restructuring of the data can be done such that we can see distinct signatures for each cavity setpoint in amplitude, phase and beam input energy.
- We look at BPM phase differences, not against RF phase, but against each diagnostic output, the pairing of BPMs. Figure 3 shows an example of this type of plot, where each line represents a cavity set point and is measured in a single pass through the machine, without scanning any parameter.

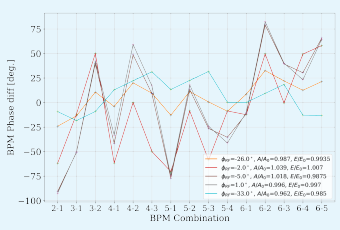

Figure 3: BPM phase differences for each possible BPM coupling, with the different plots each corresponding to a single cavity set point.

- From here we encounter the same problem to be solved as with the phase scan data, needing to accurately identify these new signatures. The nature of the signatures in this data format leaves ML uniquely equipped for the task.
- This new data format was used to produce the results presented here.

Machine Learning

- Machine learning algorithms come in many forms and can solve many distinct problems using varying network structures, definitions of loss and optimization algorithms [6].
- We compare two types of network, a traditional linear regression structure, and a newer decision tree boosting model called XGBoost.

Linear Regression

- This network was defined using the python library Keras [7].
- Our loss function was mean squared error and our optimization algorithm was ADAM [8].
- Optimization of the network structure and training parameters was done iteratively, looking at generalized performance as the figure of merit.
- Through this process we arrived at a 10-layer structure with an 160-160-80-80-40 symmetrical neural layout, and a final output layer of three neurons, which was trained for 20 000 epochs with a learning rate of 0.00001.

XGBoost

- The modern ML system of XGBoost (eXtreme Gradient Boosting) is an open source gradient boosting model, which has proven extremely powerful for solving varied, nonlinear problems [9].
- A gradient boosting ML system uses an ensemble of many decision trees in order to improve the final predictions, and commonly a regularized loss function which penalises increasing complexity of the model as well as the usual error of predictions.
- Figure 4 shows a simplified model of the principle of decision tree ensembles.

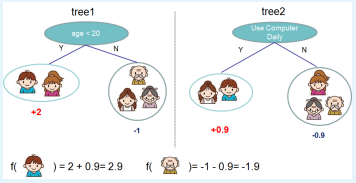

Figure 4: Simplified figure of the working principle of a tree ensamble network. In this case, the final prediction of the ensamble is a summation of the predictions of individual trees [9].

- For the results produced here an ensemble of 10000 trees was used, each with a max allowed depth, the amount of branching criteria, of 20. A learning rate factor of 0.0001 was applied and an early stopping system was also used during training, forcing the training to halt if the generalized performance of the ensemble network did not improve over a period of 500 iterations.

Results

- Table 2 shows three standard deviations (3σ) and the mean (μ) of the difference between the predicted and expected value for the RF Amplitude (A) and phase (ϕ), as well as for the input energy (E).
- We see good performance in the energy predictions, but both methods fail to produce the sought results in both phase and amplitude.
- The variation in the single shot signatures as a function of phase is quite small, so networks struggling to distinguish between these is understandable.
- XGBoost only slightly outperforms the more traditional linear regressor, but remains far outside the limit in the phase prediction.

Conclusions

- While both methods may fail the limits for operation at this stage of investigation, there are still many factors which could prove this method more reliable than suggested by these results.
- The error data set produced may have been pessimistic in the predictions of one or many of the factors included and further optimization of the meta parameters for the training of the networks could reveal better results in the future.
- The single shot nature of the data could allow for updated tuning information during operation of the machine, as well as long term tracking of drifts on the RF parameters.
- Further applications of this online tuning information could be developed in the future, for use in the ESS control room or elsewhere.

References

[1] R. Garoby *et. al.*, "The European Spallation Source Design", Physica Scripta, vol 93, 014001 (2018).

[2] G. Shen and M. Ikegami, Chin. Phys C, **33**, 577 (2009).

[3] T.L. Owens, M. B. Popovic, E. S. McCrory, C. W. Schmidt and L.J. Allen, "Phase Scan Signature Matching for Linac Tuning", Particle Accelerators, vol 8, no 1, pp 169-179 (1994).

[4] G. B. Shen, H. Sako, and S. Sato, "RF Amplitude and Phase Tuning of J-PARC SDTL", in *Proc. 22nd Particle Accelerator Conf. (PAC'07)*, Albuquerque, NM, USA, Jun. 2007, paper TUPAN062, pp. 1529–1531.

[5] OpenXAL, `https://openxal.github.io/`.

[6] A. Edelen et. al., "Opportunities in Machine Learning for Particle Accelerators", **1811.03172**, arXiv (2018).

[7] Keras, `https://keras.io`.

[8] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization", **1412.6980**, arXiv (2017).

[9] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System", **1603.02754**, arXiv (2016).

Table 2: Three standard deviations in difference between predicted and correct values for the RF Amplitude and phase and the input energy. Results shown for both linear regression (LR) and XGBoost (XGB) network structures.

Data Set		$3\sigma_A$ [%]	$3\sigma_\phi$ [°]	$3\sigma_E$ [%]	μ_A [%]	μ_ϕ [°]	μ_E [%]
No Errors	LR	0.075	0.051	0.045	0.002	0.000	0.002
	XGB	0.891	1.755	0.153	0.025	-0.013	0.002
All Errors	LR	4.002	5.568	0.804	-0.013	-0.068	-0.009
	XGB	3.171	5.217	0.750	0.022	0.038	0.016

C Courses Attended

Also as part of this project, two online courses were attended: The LENS machine learning school hosted by the ISIS neutron and muon source in the UK, and the US Particle Accelerator School graduate course Accelerator Physics, hosted by Stony Brook University. The LENS school was not graded, but the USPAS course was attended on a scholarship and was completed with the highest grade.