Iterative method for emittance measurements in matching sections

Nino Čutić
Lund University

An iterative version of the multiple–quadrupole method to measure the beam emittance and the Twiss parameters is presented. The version is programmed into an algorithm that runs fully automated and finds the parameters and their variances for Gaussian–like beams in about twenty shots in both transverse planes from measurements of beam sizes. The method is explained and then demonstrated by simulating such measurements on a matching section of the MAX–IV linac consisting of straights, quadrupoles and a fluorescent screen. Kalman filter is used for the state estimation in the algorithm and except the energy no other information is needed about the beam. The determined beam parameters are shown to be correct for close–to–Gaussian beams with noise and saturation of the fluorescent screen present on the screen. Performance for non–Gaussian beams is discussed.

Introduction

Motivation

This paper is a result of work on the matching sections of the MAX IV linear electron accelerator (linac) in Lund. The 3 GeV linac, currently in construction, has two bunch compressors each of them preceded and followed by a matching section. For proper operation of the bunch compressors measurements of the Twiss parameters and emittance (in further text, the beam parameters) are needed. The measurements are planned in the matching sections using an iterative method presented here which is based on the multiple–quadrupole method [1]. The measurement process is described and demonstrated using simulations and it is planned to be performed on the linac once it is operational. The goal is to measure the beam parameters in the horizontal and vertical transverse planes at the entrance to a straight matching section. This is possible using one downstream fluorescent screen and multiple quadrupoles upstream of the screen. Figure 1 illustrates one such matching section.

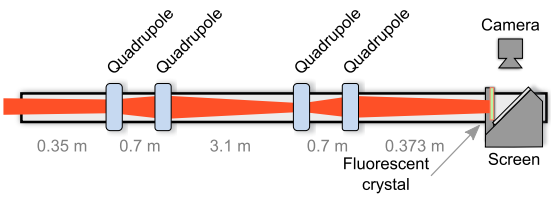

Figure 1
Illustration of the straight matching section consisting of four quadrupoles and a fluorescent screen. The emittance and the beam parameters need to be determined by changing the quadrupole strengths and measuring the beam size at the screen. The lengths given underneath each drift corresponds to drift lengths in the first matching section of MAX–IV linac. This matching section is used for simulations to demonstrate the method later in the paper.

The beam parameters should be remeasured occa-

sionally or during the startup in short time which corresponds to performing a complete measurement in several dozen shots. Thus a fully automatic fast method was needed. Except the energy, no previous knowledge about the beam parameters is needed. Other techniques for measuring emittance of course exist, for example pepper-pot [2, 3] or emittance-meter [4], each with its drawbacks and advantages. Here, only multiple quadrupoles separated by straight drifts and followed by a fluorescent screen are considered. The main advantage of quadrupole scans is that they do not require any extra specialized equipment.

In a simple *single–quadrupole* scan the emittance of a beam is determined using a fluorescent screen and only one quadrupole. The beam size is measured on the screen while the strength of the quadrupole is changed. For accuracy it is required that the scan is performed so that the minimum beam size (focus) is reached. The quadrupole strengths of interest are around a strength that gives focused beam. Such a scan results in a parabola as illustrated in Figure 2. After a least–squares fit the emittance can be determined from the parameters of the parabola. The drawback of this method is that close to the focus the beam often causes a saturation on the screen, while far from the focus the beam size becomes too large. An over–sized beam is either outside of the screen or too weak compared to the noise for reliable beam size measurement. Saturation and an over–sized beam both lead to wrong estimates of the beam size and wrong parabola parameters and thus a misestimate of the emittance.

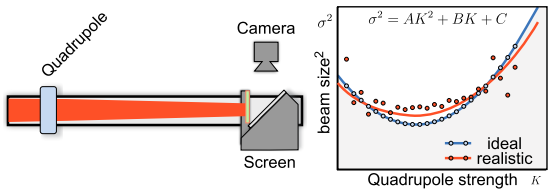

Figure 2
Single–quadrupole scan suffers from saturation close to the focus of the beam (beam size is over estimated) and from poor signal to noise for very large beam size (measurements are imprecise and inaccurate). Emittance and beam parameters can be wrongly estimated since they are calculated from the fit to a parabola that is badly determined.

One way around the saturation problems faced by the single–quadrupole scan is to expand the beam in the orthogonal transverse plane (in which the measurement is not performed) by using at least one extra upstream quadrupole. A better and more general method is, having several quadrupoles at disposal, to perform the *multiple–quadrupole scan*. This is done by keeping the beam size within a certain range, somewhere between the focus size and the over-sized beam in both transverse planes while changing the strengths of the quadrupoles. The beam sizes are measured for each set of quadrupole values and used in a least square estimate of the beam parameters. It will be later discussed which characteristics a set of measurements needs to posses in order to determine the beam parameters with acceptable accuracy, and also which quadrupole strengths can be considered different enough from previous ones to constitute a new measurement. The multiple–quadrupole method of course has its own weaknesses like the hysteresis of the magnets or detrimental chromatic effects for beams with large energy spread [1, 5]. MAX IV linac beam will have negligible energy spread. All these drawbacks of the multiple–quadrupole scan apply also to the *iterative multiple–quadrupole scan*, presented here. The method's clear advantages are price and availability since it can be performed without requiring any specialized or extra equipment. The method is destructive since the screen is inserted into the beam path. The screen is not a necessary part for the method, any measurement of the beam size after the quadrupoles is sufficient.

In this paper a basic definition of the RMS emittance is given and then the iterative method is presented in detail. Main parts of the method are described in their corresponding paragraph. The simulations of three measurements for MAX–IV matching section are shown and discussed. The first measurement shows the method's performance for Gaussian beams while the following ones investigate influence of non-Gaussian profiles on the final results.

Definitions

Let vector X represent the transverse coordinates, x and x' in horizontal plane and y and y' in vertical plane $X = (x, x', y, y')$ for a particle. A relativistic bunch consisting of N particles will be described by N such vectors. By switching to an index based notation in which $(x, x', y, y') = (x_1, x_2, x_3, x_4)$ The central moment C_{mnlk} of the bunch is defined as:

$$C_{mnlk} = \frac{1}{N} \sum_{\text{all N particles}} (x_1 - < x_1 >)^m (x_2 - < x_2 >)^n (x_3 - < x_3 >)^l (x_4 - < x_4 >)^k \quad (1)$$

where $< x_i >$ represents the mean value of the coordinate x_i over all particles. The symmetric variance–covariance matrix ${}^4\mathbf{\Sigma}$ is given by:

$${}^4\mathbf{\Sigma} = \begin{bmatrix} C_{2000} & C_{1100} & C_{1010} & C_{1001} \\ C_{1100} & C_{0200} & C_{0110} & C_{0101} \\ C_{1010} & C_{0110} & C_{0020} & C_{0011} \\ C_{1001} & C_{0101} & C_{0011} & C_{0002} \end{bmatrix} \quad (2)$$

This matrix describes the transverse profile of the beam up to a second moment and the correlations between the components. With this type of description any features of the bunch distribution

not describable by the second order moments have been implicitly neglected i.e. they have been assumed not important. Another assumption is that the transverse planes are uncoupled and thus ${}^4\mathbf{\Sigma}$ is a block–diagonal matrix. There are no correlations between components of vector X if the components belong to different planes. This splits ${}^4\mathbf{\Sigma}$ into two ${}^2\mathbf{\Sigma}$ matrices

$$ {}^4\mathbf{\Sigma} = \begin{bmatrix} {}^2\mathbf{\Sigma}_h & 0 \\ 0 & {}^2\mathbf{\Sigma}_v \end{bmatrix}. \tag{3} $$

The horizontal matrix ${}^2\mathbf{\Sigma}_h$ and the vertical ${}^2\mathbf{\Sigma}_v$ are independent of each other. Horizontal and vertical RMS non–normalized emittances are defined by:

$$ \epsilon_{h,v}^2 = \det\left({}^2\mathbf{\Sigma}_{h,v}\right) \tag{4} $$

The structure of the variance–covariance matrix guarantees that the emittance will never be imaginary since any variance–covariance matrix is positive–semidefinite. Thus its principal minors are non–negative and the determinant in the definition of the emittance is always non–negative. With the definition of the emittance the ${}^2\mathbf{\Sigma}$ can be rewritten in a way that explicitly contains the Twiss parameters α, β and γ:

$$ \mathbf{\Sigma}_i^h = \begin{bmatrix} C_{2000} & C_{1100} \\ C_{1100} & C_{0200} \end{bmatrix} = \begin{bmatrix} \epsilon_h\beta_{h,i} & -\epsilon_h\alpha_{h,i} \\ -\epsilon_h\alpha_{h,i} & \epsilon_h\gamma_{h,i} \end{bmatrix} \tag{5} $$

$$ \mathbf{\Sigma}_i^v = \begin{bmatrix} C_{0020} & C_{0011} \\ C_{0011} & C_{0002} \end{bmatrix} = \begin{bmatrix} \epsilon_v\beta_{v,i} & -\epsilon_v\alpha_{v,i} \\ -\epsilon_v\alpha_{v,i} & \epsilon_v\gamma_{v,i} \end{bmatrix} \tag{6} $$

New subscript i is added to describe the position along the transport (matching section), i will take the values 1 or 2 depending on the position in the transport; 1 is upstream at the entrance of the matching section while 2 is at the screen. The emittance is position independent since it is assumed that it does not change during the transport through the section (no non-linear effects, no radiation, the particles are highly relativistic). If the transport is given by matrix $\mathbf{M}$ so that the coordinates of each particle transform from vector X_1 to vector X_2 according to:

$$ X_2 = \mathbf{M}X_1 \tag{7} $$

then the $\mathbf{\Sigma}_1$ will transform to $\mathbf{\Sigma}_2$ according to the congruence:

$$ \mathbf{\Sigma}_2 = \mathbf{M}\mathbf{\Sigma}_1\mathbf{M}^\intercal \tag{8} $$

These transformation rules are valid in general for second order central moments of distributions. From (4) and (6) it is easy to see that $\beta\gamma - \alpha^2 = 1$ and thus all the information is already contained in only α, β and ϵ.

For each plane separately the transport matrix $\mathbf{M}$ from point 1 to the point 2 can be decomposed into a product of three matrices [6]:

$$ \mathbf{M}_{1,2} = \begin{bmatrix} \sqrt{\beta_2} & 0 \\ -\frac{\alpha_2}{\sqrt{\beta_2}} & \frac{1}{\sqrt{\beta_2}} \end{bmatrix} \begin{bmatrix} \cos\Psi & \sin\Psi \\ -\sin\Psi & \cos\Psi \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{\beta_1}} & 0 \\ \frac{\alpha_1}{\sqrt{\beta_1}} & \sqrt{\beta_1} \end{bmatrix} \tag{9} $$

$$ = \begin{bmatrix} \sqrt{\frac{\beta_2}{\beta_1}}(\cos\Psi + \alpha_1\sin\Psi) & \sqrt{\beta_1\beta_2}\sin\Psi \\ -\frac{1+\alpha_1\alpha_2}{\sqrt{\beta_1\beta_2}}\sin\Psi + \frac{\alpha_1-\alpha_2}{\sqrt{\beta_1\beta_2}}\cos\Psi & \sqrt{\frac{\beta_1}{\beta_2}}(\cos\Psi - \alpha_2\sin\Psi) \end{bmatrix} \tag{10} $$

where Ψ is known as the phase–advance. This decomposition is not unique, since different beam parameters will lead to different exit beam parameters even if the matrix $\mathbf{M}$ is fixed. Fixed in the sense that the quadrupole strengths might be set already and the transport matrix $\mathbf{M}$ is just a product of matrices of individual components. The decomposition will become unique once the entrance alpha and beta (α_1 and β_1) are known together with the transport matrix $\mathbf{M}$. In other words, once the input beam and the transport matrix are known, downstream alpha and beta are uniquely determined as well as the phase advance.

If α_1 and β_1 are fixed and if matrix $\mathbf{M}$ is also fixed (does not contain any free parameters, it could be only a drift of certain length for example), then there is no free choice of α_2 and β_2. If the transport contains a controllable quadrupole then it is possible to change β_2. Two quadrupoles give the possibility to freely select both α_2 and β_2 (up to a certain extent, aperture limitations). In both planes four quadrupoles are needed to have the freedom of selecting fixed $\beta_{2,h}$ and $\beta_{2,v}$ and a possibility of changing $\alpha_{2,h}$ and $\alpha_{2,v}$.

Iterative method

The iterative method could be described as a multiple-quadrupole scan divided into small pieces consisting of S measurements. In each iteration, S measurements of beam sizes $\sigma_{2,h}$ and $\sigma_{2,v}$ are performed and the estimate of the beam parameters is produced. The quadrupole strengths used in these S measurements are calculated based on the last beam parameter estimates (see Figure 4). The strengths are selected to produce a constant desired beam sizes $\sigma_{2,h}^*$ and $\sigma_{2,v}^*$. This was suggested as a possible way of performing the emittance measurement in [1]. The method's flow diagram is presented in the Figure 3 and each block of the diagram is explained in the further text. Figure 4 illustrates the differences between the quadrupole scan methods. The desired beam size for the measurement (not saturated, not too large) is provided to the measurement script by the machine operator. The operator also enters acceptable deviation from this beam size, e.g. 10%. The iterative method tries to keep

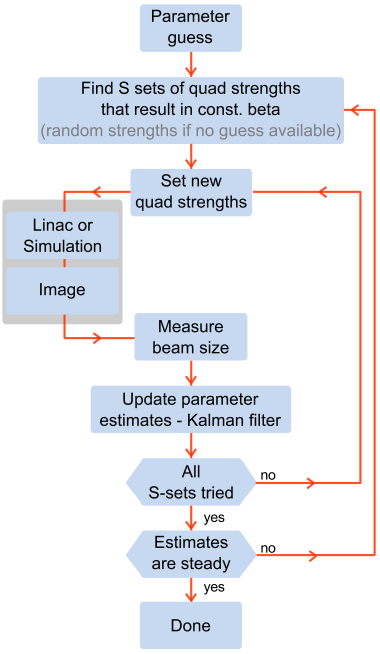

Figure 3
The flow of the iterative routine, described in detail in the text. Starting from the initial guess the routine is remeasuring the beam size and changing quadrupole strengths. The quadrupole strengths are selected in a way that would maintain the beam size constant based on the current estimate of the beam parameters. This estimate in each iteration improves and in the end the beam size is constant even though quadrupoles change significantly in strength.

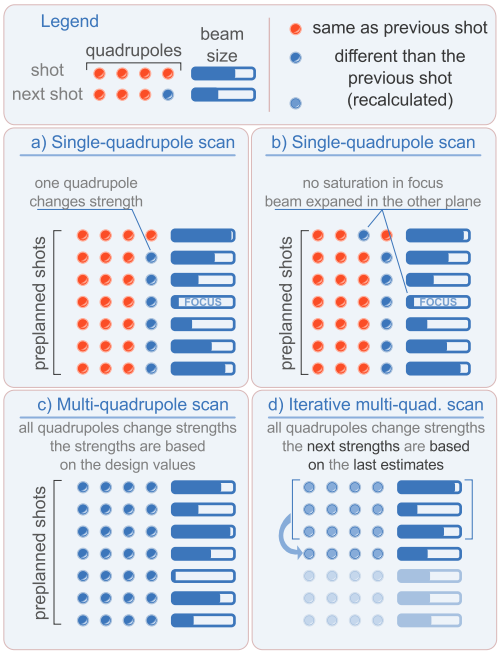

Figure 4
Differences between quadrupole scan methods. a) One quadrupole is used for scans and the beamsize is measured. Saturation in focus and signal to noise issues for too large beams are present. b) To avoid the saturation the beam is expanded in orthogonal plane using an upstream quadrupole first, so that it is not saturated even when in focus. This on the other hand can enhance signal to noise problem when the beam size is large. c) A scan that uses all available quadrupoles is planned so that the beam size is constant and moderate. Without correct input parameters of the beam this results in a changing beam size (as described later, and shown in Figure 6). Too large or too small beam sizes need to be either discarded or taken into consideration with larger uncertainty. d) Iterative version generates the next values of strengths only after processing the previous measurements. Most recent estimate is used for that purpose. With more measurements the beam size becomes more stable and closer to wanted size. All measurements can be used. This scan method converges faster.

the beam at the desired size. However, these attempts to maintain certain beam size and change quadrupole strengths will only be successful once the real (true) beam parameters are known, once the estimates of beam parameters based on measurements start approaching real values. As the estimates approach real values the beam size will stabilize. Before that, the wrong estimates of input beam parameters result in the calculation of required quadrupole strengths that will not keep the beam size constant.

Parameter guess. The method starts from an initial guess of the upstream parameters ϵ, α_1, β_1 (for each plane separately) which are coming from the design parameters or values that were previously measured. If such a guess is not available (first time measurement) the method can proceed with random quadrupole strengths. To obtain a reasonable first guess it takes approximately 10 iterations (shots). Measurements in both transverse planes are performed simultaneously.

Finding quadrupole strengths. Given the guess of the upstream parameters the script needs to find which other quadrupole strengths produce a beam with the desired beam size $\sigma_{2,x} = \sigma^*_{2,x}$, $\sigma_{2,y} = \sigma^*_{2,y}$ but will not create too large beam size along the matching section. A discussion on the properties of transport matrices that have the desired property will reveal an important aspect of these scans.
If entrance parameters of the beam α_1 and β_1 are available, what are the properties of a transport matrix $\mathbf{M}$ that preserves some chosen β_2? From the transformation (8), after dividing by the appropriate emittance, it follows that:

$$\begin{aligned} \beta_2^h &= \beta_1^h M_{11}^2 - 2\alpha_1^h M_{11} M_{12} + \gamma_1^h M_{12}^2 \\ \beta_2^v &= \beta_1^v M_{33}^2 - 2\alpha_1^v M_{33} M_{34} + \gamma_1^v M_{33}^2 \end{aligned} \tag{11}$$

where M_{jk} are the elements of the matrix $\mathbf{M}$. Dividing these equations with β_2^h and β_2^v respectively,

two equations of rotated ellipses follow, one for each plane as illustrated in Figure 5:

$$\begin{aligned} A_h M_{11}^2 + B_h M_{11} M_{12} + C_h M_{12}^2 &= 1 \\ A_v M_{33}^2 + B_v M_{33} M_{34} + C_v M_{34}^2 &= 1 \end{aligned} \tag{12}$$

If the estimates of α_1 and β_1 are very close to the real

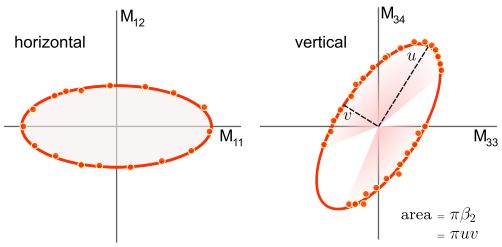

Figure 5
The ellipses defined in equations (12). The transport matrices **M** for which the element pairs (M_{11}, M_{12}) and (M_{33}, M_{34}) lie on these ellipses will produce a beam of constant size but different divergence and phase advances for certain α_1 and β_1. The orientations and eccentricities of these ellipses depend only on input Twiss parameters in each plane and their sizes depend on output betas. If the entrance alpha is zero the ellipse will not be rotated as illustrated for horizontal plane. Due to limitations in beam size along the matching section or due to constraints on magnet strengths some transport matrices cannot be used/generated and thus certain regions of the ellipses might be unavailable (illustrated for the vertical plane).

values the transport matrices whose elements lie on these ellipses will keep the beam size constant. This is the case only later in the whole iterative process. Otherwise, if the estimates differ from the real parameters the beam will vary in size, as illustrated in Figure 6. Red ellipses portray the transport matrices that are supposed to keep the beam size constant, but since red ellipses are generated based on estimates that are not correct the points that lie on the red ellipses actually belong to blue ellipses of varying beam size. The blue ellipses correspond to the transport matrices that would actually keep the beam size constant.
If an ellipse is given by its implicit form $Ax^2 + Bxy + Cy^2 = 1$ then it can be described by an associated symmetric bilinear form, the matrix **W**:

$$\mathbf{W} = \begin{bmatrix} A & \frac{B}{2} \\ \frac{B}{2} & C \end{bmatrix} \tag{13}$$

The ellipse will have the semi–major and semi–minor axes u and v directed along the eigenvectors of **W**. The lengths of the semi–axes are $\lambda_i^{-1/2}$ where λ_i are eigenvalues of **W**. With a bit of algebra it is possible to see that the surface of this ellipse will be $\pi u v = \pi \beta_2$ while its rotation angle and eccentricity depend only on α_1 and β_1. This is the key to the method since knowing the ellipse means knowing α_1, β_1, β_2 and subsequently α_2 and emittance. Output beta combined with the beam size leads to the

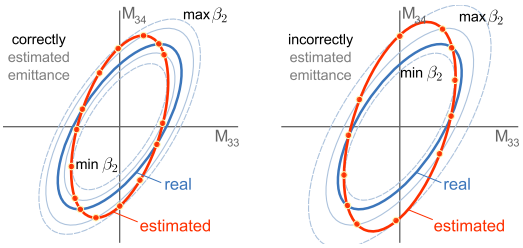

Figure 6
Influence of the estimate of the input parameters on the beam size, illustrated for the vertical plane; the horizontal would be analogous. During the measurement a wrong estimate of Twiss parameters but correct estimate of emittance will result in beam size variation between maximum and a minimum. The size of those extremes are is illustrated by blue ellipses which have correct (unknown) input beam parameters but differing β_2. On the right an illustration of a case with underestimated emittance is shown. The red ellipse (estimated parameters) will have a different size since the larger β_2 compensates the underestimate of the emittance.

emittance, and thus all three parameters at the entrance to the matching section are determined. The points on the ellipse need to be far enough from each other so that the reconstruction of the ellipse parameters is reliable given the precision in measuring the beam size and the noise levels present, as well as quadrupole strengths (real performance of the magnet vs. control system values). This gives the idea of which transport matrices, which quadrupole values, are sufficiently distinct or "far enough from previous measurement" when changing to another quadrupole strengths. The more of the ellipse is covered, the better.
Which quadrupoles strengths should be applied in the scan to generate transport matrix elements that lie on ellipses given in (11) and (12)? This is a non-trivial problem which is solved numerically. At this point information about the matching section, like quadrupole lengths and drift lengths, are used.
In the case of a matching section with four quadrupoles, the complete general transport matrix will be a result of multiplication of at least four quadrupole matrices and four drifts matrices. The elements of the general matrix will non–linearly depend on the quadrupole strengths K_q. There are also constraints present, the strengths of physical quadrupoles are limited and the beam size is limited by the vacuum pipe at the locations of the quadrupoles. The problem can now be reformulated so that it does not involve ellipses, but in essence it is still these ellipses that are constructed: "Given a general transport matrix which is a product of transport matrices of the matching section components and which depends only on constrained quadrupole strengths K_q, find *all* or sufficiently many clearly distinct sets of quadrupole strengths that result in a constant horizontal and vertical beam size, σ_2^h *and* σ_2^v, if the entrance beam pa-

rameters are given, and beam sizes are constrained along the transport." In the implementation used in this paper MATLAB routines *MultiStart*, *createOptimProblem* and *run* are used to solve this problem. A proper fitness functional $\mathscr{F}$ for the optimisation also needs to be defined and it can be as simple as:

$$\mathscr{F} = \left(\frac{\sigma^{h}_{2,\text{found}} - \sigma^{h}_{2,\text{desired}}}{\sigma^{h}_{2,\text{desired}}} \right)^2 + \left(\frac{\sigma^{v}_{2,\text{found}} - \sigma^{v}_{2,\text{desired}}}{\sigma^{v}_{2,\text{desired}}} \right)^2 + F^* \tag{14}$$

where the F^* part is penalizing excessive beam sizes along the transport and $\sigma^{h,v}_{2,\text{found}}$ are exit beam sizes for current quadrupole strengths that constrained optimization algorithm is testing out.

Measurement and image processing. After the quadrupole strengths have been found and set to magnets an image of the screen is obtained. The image usually contains various types of noise. Efforts to reduce the noise either prior to the measurement or afterwards in processing are very important since the whole method relies on beam size measurements. The measured beam size is then forwarded to the estimation part of the method. If the beam size on the screen is not the desired one or even within the acceptable range, it can still be used. All the measurements are taken into consideration but the method will weigh their contribution to the estimate of the real parameters based on how far off they are from the desired size.
In the simulations used in this paper several types of noise are generated to simulate realistic conditions and afterwards the script that is performing the measurement is filtering the noise. The filtering part does not have any information about the types and levels of noise the simulation part is using. Since the beam size is calculated as variance of what is on the screen it is very important to remove noise that is far away from the center of the beam, such noise can significantly increase the measured beam size. 2D median filter is used to find and remove the hot pixels. In case of the Gaussian beams the profile on the screen will be ellipse–like. A preliminary principal component analysis (PCA) will give the parameters of that ellipse. These values can then be used to find the Mahalanobis distance of all the pixels in the image ("how far they are from the beam"). Mahalanobis distance takes into consideration correlation between variables and can be used to clean the image of everything larger than the beam. In second pass all the points further away than certain Mahalanobis threshold are set to zero intensity. The PCA and Mahalanobis filtration can be done several times, but once or twice proved to be sufficient. This method proved to be very successful when measuring beam size for Gaussian beams even with large amounts of noise but, as is shown later, reduced the reliability of the method for non-Gaussian beams.

Estimates - Kalman filter. In the iterative method a discrete Kalman filter (KF) is implemented to produce the estimates of the beam parameters in each iteration [7, 8, 9, 10]. Although itself the KF offers much more, everything it does for this iterative method could be replaced by a weighted least squares method. The Kalman filter however has a very easy implementation and by design contains information about the variance of the estimate. Each transverse plane will have its separate KF.
Let x be a state vector in one plane that is to be estimated. As an example, in horizontal plane x will be a triplet of $(x_1, x_2, x_3) = (C_{2000}, C_{1100}, C_{0200})$. KF is able to estimate a state vector x that in each iteration changes according to $x = \mathbf{A}x + \mathbf{B}u + v$ based on observations of quantity (or a vector) z which is connected with state vector x through $z = \mathbf{H}x + \mathbf{D}u + w$ where:

- the matrix $\mathbf{A}$ is a state transition matrix. For this iterative method it is assumed that the input beam parameters do not change with each shot to new parameters based on previous parameters. Thus this matrix is the identity matrix.
- the matrix $\mathbf{B}$ is describing influence of the control u onto the state. Since there is no controller of the beam parameters this additive term can be ignored (u is constantly zero).
- v is a zero–centered Gaussian noise with variance Q,
- the matrix $\mathbf{D}$ is similarly to matrix $\mathbf{B}$ zero.
- the matrix $\mathbf{H}$ is the observation matrix (more about it in later text) and
- w is a zero–centered Gaussian noise with variance R.

The KF in three steps produces an estimate of x and its covariance $\mathbf{P}$:

- Prediction:
  $x \leftarrow \mathbf{A}x + \mathbf{B}u$
  $\mathbf{P} \leftarrow \mathbf{APA}^\intercal + Q$
- Optimal Kalman gain:
  $\mathbf{K} \leftarrow \mathbf{PH}^\intercal(\mathbf{HPH}^\intercal + R)^{-1}$
- Updates of estimates:
  $x \leftarrow x + \mathbf{K}(z - \mathbf{H}x)$
  $\mathbf{P} \leftarrow \mathbf{P} - \mathbf{KHP}$

The real implementation of this code might differ slightly for numerical stability reasons. Whether the measured beam size is in the acceptable range

can be easily communicated to the Kalman filter by specifying different R in each iteration. Higher R for certain z means less trustworthy measurement and such measurements will have very small influence on the estimates of beam parameters, and vice versa, small R means trustworthy measurement and larger influence on current estimate. Figure 7 shows qualitatively the R dependence based on the measured beam size and acceptable beam size range. The observation matrix $\mathbf{H}$ is describing what part of the state vector is being measured. Only the beam size is observed, and its square is sent to the KF ($z = \sigma^2_{\text{beam}}$). Since our state vector x is $(\epsilon\beta, -\epsilon\alpha, \epsilon\gamma)$, comparing with equations (11) it follows that $\mathbf{H}$ is in this case a row vector $(M_{11}^2, 2M_{11}M_{12}, M_{12}^2)$. The observation matrix changes for each measurement, but the state does not change. Each measurement we are getting a new view of the state.

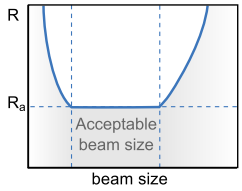

Figure 7
Variance of the beam size measurement R that is used by the Kalman filter is made to depend on the measured beam size. This effectively makes the KF sensitivity to measurements outside of the acceptable range artificially even lower. Impact of this is visible only in the beginning of the scan when possibly the beam parameters are not known very well and beam sizes are outside of the acceptable range. Since no better measurements are available the KF will use what it has but when the beam sizes are in the acceptable range the measurements that were outside will be suppressed. R_a is the variance of the measurements for beam sizes inside the acceptable range.

Steady estimates. As the scan progresses the estimates of the beam parameters converge to true parameters and the beam size (up to the noise and the precision of measurement) stabilizes despite significant changes in quadrupole strengths. At this point the measurement is finished.

Simulations

Simulations in MATLAB were performed for matching section 1 (MS1), see Figure 1, of the MAX IV linac.

Simulation results. Simulated scan for 35 shots is shown in Figures (8–11). Figure 8 demonstrates that the beam parameters are found quite fast, even though initial guess was quite different from the true values. Figure 9 shows determined emittance and a reduction in its uncertainty with the number of scans progressing. The standard deviation of emittance σ_ϵ is calculated from the covariance matrix $\mathbf{P}$ of the respective Kalman filter using:

$$\sigma_\epsilon^2 = \mathbf{J}\mathbf{P}\mathbf{J}^\intercal \tag{15}$$

where Jacobian $\mathbf{J}$ only contains the beam parameters $\mathbf{J} = \left(\frac{\partial\epsilon}{\partial x_1}, \frac{\partial\epsilon}{\partial x_2}, \frac{\partial\epsilon}{\partial x_3}\right) = \left(\frac{\gamma}{2}, \alpha, \frac{\beta}{2}\right)$. Figure 10 shows

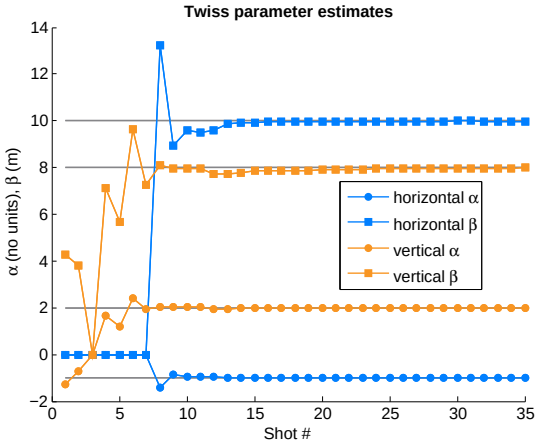

Figure 8
α (no unit) and β (in meters) in the horizontal and the vertical plane as estimated by the measurement versus the shot number. Horizontal lines show true values. The values start at design values (at a 'shot zero', not shown). The final values are given in the Table 1.

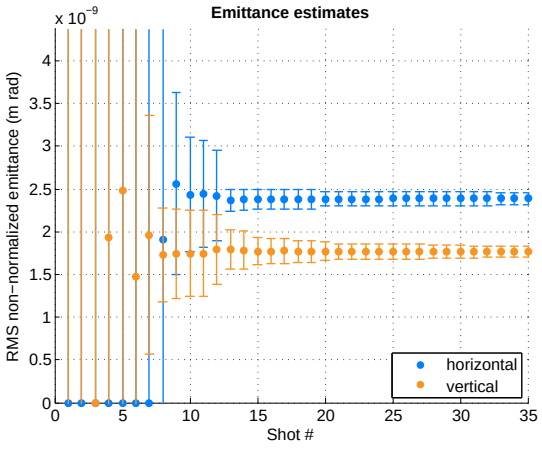

Figure 9
Emittance measured in both planes. The starting, true and final values are given in Table 1. Uncertainties are propagated from the covariance matrix P in Kalman filters. Error bars show $\pm 3\,\sigma$.

how the beam size was changing during the measurement. The beam size stabilizes as intended once the beam parameters are well known. Beam size measurements were clearly slightly biased in

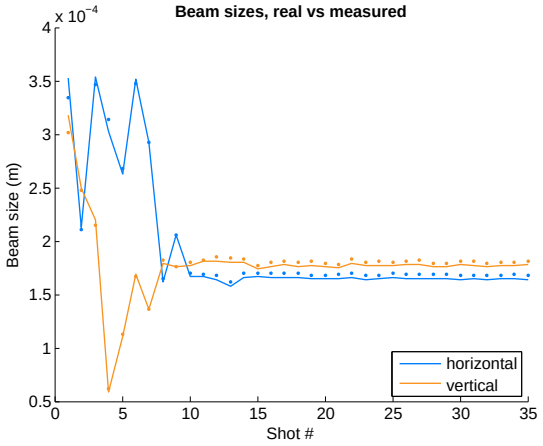

Figure 10
Measured and real beam size during the measurement. The beam size stabilizes around the ideal beam size as the scan progresses. Lines show the real size while dots show the size that was measured by the measurement script after it filtered out the noise. Beam size measurements are slightly biased due to various effects of image processing.

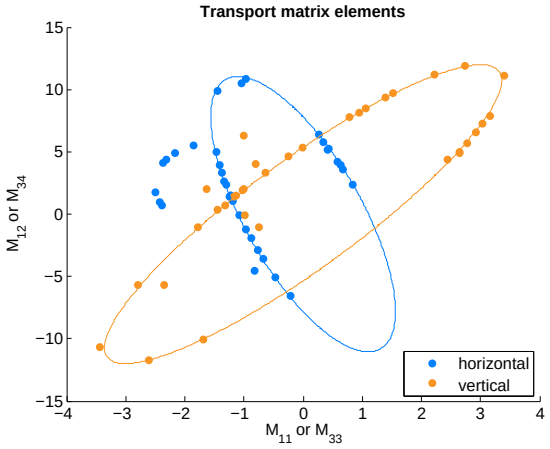

Figure 11
Matrix elements M_{12} vs M_{11} (horizontal) and M_{34} vs M_{33} (vertical) of the transport matrix that were used in the scan with 35 shots. To generate these transports quadrupole strengths varied greatly as shown in Figure 12. The ellipses are generated using the found beam parameters and equations (12).

both planes. This a compounded effect of many factors like noise, image processing, camera resolution and pixel size calibration. It is interesting to compare the Figure 10 with Figure 11 which shows transport matrix elements for each plane and ellipses constructed with determined beam parameters. Points that lie on ellipses in Figure 11 are the points that have ideal beam size in Figure 10. Eight points in horizontal plane that are outside the ellipse are the measurements with beam size larger than ideal size. Three points inside the vertical ellipse produced the beam size smaller than the ideal $\sigma^*_{2,y}$. Although reliable measurements of beam parameters and emittance can be done with just about 20 shots it should be said that the quadrupole strengths need to change significantly between each shot, so these shots most likely do not correspond to consecutive shots on a high repetition machine since one needs to wait for the quadrupole strengths to change. In these simulations after each measurement the new strengths of quadrupoles were searched, i.e. number S in Figure 3 is set to 1. For the shots where current estimates are unrealistic (zero emittance estimates) the last non zero–emittance estimate is used for quadrupole strength search.

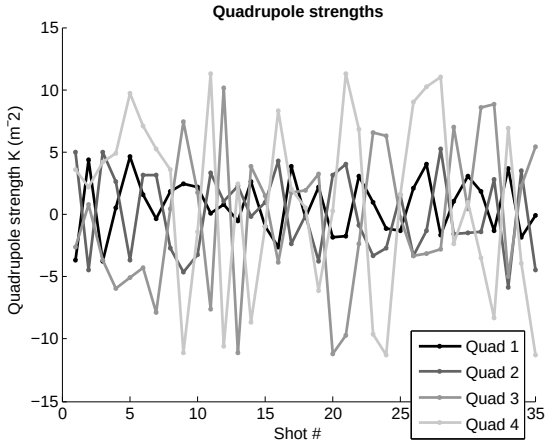

Figure 12
Quadrupole strengths during the scan. Despite such changes the beam size stays constant because the beam parameters are known well enough.

Design	Horizontal	Vertical
Emittance (10^{-9} mrad)	1.93	1.93
α	−0.42	−4.76
β (m)	2.64	15.4

Simulation (true)	Horizontal	Vertical
Emittance (10^{-9} mrad)	2.3	1.7
α	−1	2
β (m)	10	8

Simulation (found)	Horizontal	Vertical
Emittance (10^{-9} mrad)	2.388	1.764
α	−0.998	2.000
β (m)	9.985	7.983

Table 1
The beam parameters at the entrance to the matching section 1 according to the linac design is shown on top. The measurement is initialized with these values. In simulation the values are on purpose set different than the design to demonstrate that the method will converge to them. These are shown as Simulation (true). The values determined after 35 shots, which ideally should be identical to the true values, are shown in Simulation (found).

Simulation details. The simulations are split into two parts, one part simulating the linac and the second part, the measurement script.

The linac part is marked grey in Figure 3, it includes particle tracking and apertures of quadrupoles, screen saturation, noise on the screen and camera, camera resolution, point spread function and dynamic range as well as differences between set quadrupole strengths and applied quadrupole strengths. The particles are generated at the entrance of the matching section as multivariate normal distribution (*mvnrnd* function) and propagated through the matching section. A 2D histogram is applied (corresponding to the camera pixels) and charge densities calculated. Saturation on the screen is simulated. Poisson noise and salt and pepper noise is added. Only the final noisy image is given to the measurement script as input. The linac simulation script adds certain error to the quadrupole set values when propagating particles. This is an effort to simulate hysteresis and control system granularity (DAC resolution) i.e. the true strengths are not necessarily the set strengths.

The measurement part communicates with the linac part only through inputs and outputs that would be available on a real machine (image as input and quadrupole strengths as output). The measurement script then filters the image, measures the beam size, produces the estimates of the beam parameters, and assigns values of quadrupole strengths. The quadrupoles of MS1 are 20 cm long and their range of the K–parameter is from 11.3 to 11.3 m^{-2}. These limits are one of the constraints for the function that is searching for quadrupole strengths to be applied. The design beam parameters at the entrance to the matching section are given in the Table 1 and they the initial guess in the measurement. The true values are shown in the middle, and the values finally determined are shown on the bottom of the table. Ideal beam sizes are set to $\sigma^*_{2,x} = 170\ \mu$m and $\sigma^*_{2,y} = 180\ \mu$m. These values are within the acceptable range. Acceptable beam size is set to be between 70 μm and 290 μm for both planes because in that range saturation is not expected and the negative effects of a too large beam are avoided too. The beam size measurement is assumed to be 10% ($p = 0.1$) accurate (99.7% of measurements will deviate from the true value at most 10%). This gives an estimate of the variance of beam size measurements when the size is in the acceptable range. Since KF input is a square of the beam size ($z_i = \sigma_i^2$) the variance R_a of z is estimated to be $R_a \approx \frac{4}{9} p^2 \bar{\sigma}^4$ where $\bar{\sigma}$ is the ideal beam size in appropriate plane. The value of R_a is artificially increased outside of the acceptable range as in Figure 7.

Non–Gaussian beams It is interesting to look into the performance of this method for beams where the approximation of a Gaussian beam does not hold. For that purpose the simulation part is modified to create a distribution where only 70% of the particles is generated with requested parameters and the remaining 30% with five times larger emittance thus creating a background halo. The parameters used and measured are given in Table 2. The

70% of particles	Horizontal	Vertical
Emittance (10^{-9} mrad)	1.93	1.93
α	−1	1.5
β (m)	10	5
30% of particles	**Horizontal**	**Vertical**
Emittance (10^{-9} mrad)	9.65	9.65
α	−1	1.5
β (m)	10	5
Total (combined)	**Horizontal**	**Vertical**
Emittance (10^{-9} mrad)	4.25	4.25
α	−1	1.5
β (m)	10	5
Simulation (found)	**Horizontal**	**Vertical**
Emittance (10^{-9} mrad)	3.54	3.5
α	−1.01	1.49
β (m)	10.16	4.97

Table 2
Simulated measurement of beam parameters for a beam consisting of two Gaussian beams with different emittances but equal beam parameters (case I). The values determined after 40 shots are shown at the bottom. The emittance is underestimated (3.5 vs 4.25) because the beam size was systematically underestimated. The halo created by the beam part with larger emittance influenced the noise filtering algorithm in such a way that the beam size was reduced.

performance of the measurement method is shown as case I in Figure 13. The measurement of the RMS beam size (dots) was systematically lower than the real value (lines) as shown at the top left subfigure because of the noise filtering algorithm which expects a Gaussian–like beam and is not necessarily method's fault. The bias lead to an underestimated emittance for both planes (top right), but did not have a noticeable effect on the Twiss parameters which are still determined properly (top center), as expected.
In the following example, case II, the 30% of the particles have different Twiss parameters and emittance. One does not expect correct determination of either the Twiss parameters or the emittance. Table 3 lists the values for this case.
The two parts of the beam, 70% and 30% combined, using the same definitions given before have the emittance and the beam parameters different than each part separately does. These values are not determined entirely correctly by the measurement (bottom subfigures of Figure 13). The bottom left

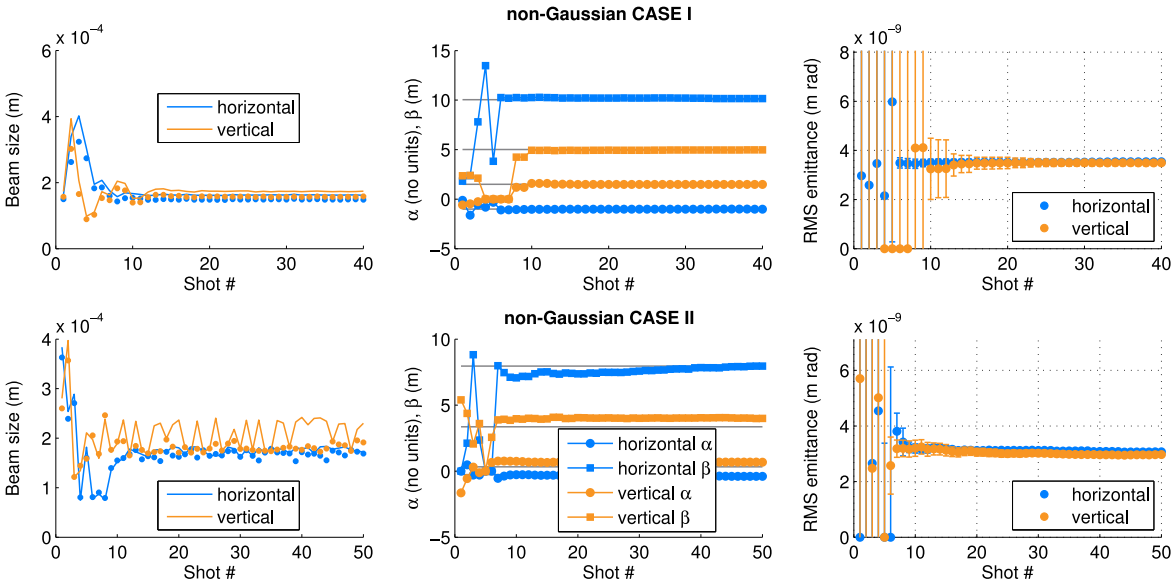

Figure 13
Performance of the measurement script for two cases of non-Gaussian beams (description in the text). Many of the problems arise due to incorrect beam size measurement for such beams. This is because the noise filtering implemented in this script assumes Gaussian–like type of beam. The values for top figures are given in Table 2 and for the bottom subfigures in the Table 3.

70% of particles	Horizontal	Vertical
Emittance (10^{-9} mrad)	2.3	1.7
α	−1	2
β (m)	10	8
30% of particles	**Horizontal**	**Vertical**
Emittance (10^{-9} mrad)	2.53	1.83
α	1	−2
β (m)	13	5.6
Total (combined)	**Horizontal**	**Vertical**
Emittance (10^{-9} mrad)	3.27	3.77
α	−0.26	0.33
β (m)	7.9	3.35
Simulation (found)	**Horizontal**	**Vertical**
Emittance (10^{-9} mrad)	3.07	2.97
α	−0.39	0.68
β (m)	7.95	3.99

Table 3
Simulated measurement of beam parameters for a beam consisting of two Gaussian beams with different beam parameters (case II). The values determined after 50 shots are shown at the bottom.

subfigure for vertical plane shows that the deviation between the real and the measured is changing but constantly underestimating the beam size. Looking at the image on the screen (before noise filtering and processing) these changes in beam size are obvious but after the filtering most of them get filtered out and the beam size is measured as almost constant (not shown). This results in wrong estimates of total, combined, beam parameters as well as underestimated emittance. Similar simulations with different combinations of Twiss parameters lead to the conclusion that the method will perform better the closer the beam is to the Gaussian (multivariate normal distribution). If the measured beam size stabilizes during the measurement that is not a guarantee that the emittance is correctly measured since the measurement of the beam size itself could be biased as is the case for the vertical plane in case II or for both planes in case I. However, stabilized beam profile (as seen on the screen) and a stabilized beam size measurement is a good indicator that the alpha and beta of the beam are correct, also in such cases it might be better to reduce the noise filtering.

Conclusion

References

[1] Tenenbaum P. *Emittance measurements in CTF 2 drive beam* (1997).

[2] Yamazaki Y, Kurihara T, Kobayashi H, Sato I, and Asami A. *High–precision pepper–pot technique for a low–emittance electron beam.* Nuclear Instruments and Methods in Physics Research A, **322**:139–145 (1992).

[3] Zhang M. *Emittance formula for slits and pepper–pot measurement.* Fermilab-TM-1988 (1988).

[4] Catani L, Chiadroni E, Cianchi A, Tazzari S, Boscolo M et al. *Design and characterization of a movable emittance meter for low-energy*

electron beams. Review of Scientific Instruments, **77**(9):093301–093301–7 (2006). doi: 10.1063/1.2336763.

[5] Mostacci A, Bellaveglia M, Chiadroni E, Cianchi A, Ferrario M, Filippetto D, Gatti G, and Ronsivalle C. *Chromatic effects in quadrupole scan emittance measurements*. Physical Review Special Topics – Accelerators and Beams, **15**:082802 (2012). doi: 10.1103/PhysRevSTAB.15.082802.

[6] Lee SY. *Accelerator Physics*. World Scientific Publishing Co. Pte. Ltd. (2004). ISBN 9–789812–562005.

[7] Faragher R. *Understanding the basis of the Kalman filter via a simple and intuitive derivation*. IEEE Signal Processing Magazine, **128** (2012). doi:10.1109/MSP.2012.2203621.

[8] Johansson R. *System Modeling and Identification*. 2nd edition.

[9] du Piessis RM. *Poor man's explanation of kalman filtering or how i stopped worrying and learned to love matrix inversion*. Technical report, North American Rockwell Electronics Group (1967).

[10] Maybeck PS. *Stochastic models, estimation, and control*, volume 141 of *Mathematics in Science and Engineering* (1979).