Free-electron-laser (FEL) machines generate high power, coherent and tunable radiation in the X-ray range. They are used to study molecular structural and dynamical processes at the angstrom-femtosecond spatio-temporal scale. The Swiss Free Electron Laser (Swiss-FEL) facility—under development at the Paul Scherrer Institut and scheduled to begin operation in 2017—will generate intense and short pulses of X-ray light enabling researchers to observe extremely fast processes, such as the creation of new molecules in chemical reactions, or to study the detailed structure of proteins. This article introduces FEL technology from a control systems perspective and provides insight into the types of control problems that arise in pulsed-mode accelerators and beams. Although developed here for the SwissFEL machine, and we expect that the control approaches are transferable to other systems as well.

The SwissFEL machine, shown in Figure 1, is composed of three main sections: an injector with an electron source, a linear accelerator (linac), and magnetic undulators. The injector system, triggered by an optical laser, generates an electron bunch with an energy of 330 MeV. Even though the electron bunch is short in time and space, it is common to refer to it as a beam. The electron bunch is compressed and then accelerated by a pulsed radio frequency (RF) driven linac. The resulting relativistic electron beam can achieve an energy of up to 5.8 GeV and will be directed to one of two undulator beam lines—periodic arrays of magnets—applying a transverse periodic acceleration to the electron beam. The undulation process transfers a part of the electron beam energy into a coherent photon pulse resulting in an extremely fast, high-intensity X-ray pulse. The motor-controlled undulators are tunable and produce coherent optical pulses with wavelengths ranging from 1 to 7 Ångstrom in one undulator beam line and 7 to 70 Ångstrom in the other. More detail on linacs and the FEL mechanism is given in “A Brief Introduction to FEL Machines.”

The SwissFEL machine is operated in a pulsed mode...
with the RF accelerating structures only being energized for 1 to 3 $\mu$s at a time. When stably energized either a single electron bunch, or two bunches 28 ns apart (in “two-bunch” mode), are fired. The entire procedure is repeated at a rate of 100 Hz. There are stringent requirements on the repeatability of the electron beam (and consequently the X-ray pulse) properties, both at the 100 Hz repetition rate, and for two-bunch operation, for both of the 28 ns spaced electron bunches.

These repeatability and precision requirements pose an interesting set of control challenges. Many of the actuation systems operate in a cascade structure giving rise to strongly interacting multi-input, multi-output (MIMO) plants with significant uncertainty. The dominant control approaches combine both iterative and robust MIMO methods and these are the focus of this paper. Much of this work has appeared in more detailed individual technical publications. Our objective in summarizing this work here is to give the reader an appreciation of the technology behind pulsed accelerator beams, the key control challenges that arise, and the most appropriate control design methods for beam operation.

**Control Challenges in SwissFEL**

To achieve a high quality X-ray pulse, all system components must operate stably and reliably. Thermal drifts and high-voltage klystron breakdowns are inevitable in such a large facility and achieving a stable and repeatable X-ray laser beam, with an X-ray pulse duration of the order of femtoseconds, is a challenging control problem.

The injector and linac RF drives operate in a pulsed mode at the rate of 100 Hz; every 10 milliseconds an electron bunch is produced and injected into the machine to be accelerated. The electron bunch travels at relativistic speeds and so electron beam control can only be applied in a pulse-to-pulse manner—measuring the beam properties (or the RF field) during one RF energy pulse and applying a correction to the next RF energy pulse. The pulsed mode of operation gives a compact and efficient physical design but introduces higher requirements on the stability and repeatability of the control system, particularly for the injector and linac RF drives. The control is also more difficult than it would be for a continuous wave operation mode.

The characteristics of many components of the system depend very strongly on temperature, and within the accelerating structures temperature is controlled to within $\pm 0.01$ °C. Although technologically challenging, the temperature control methods are standard and will not be detailed here. Changes in RF power and external disturbances give rise to temperature transients and these remain a significant source of disturbances to the electron beam properties. Precision control of the both the RF fields and the electron beam is needed to meet the operational beam stability (repeatability) requirements.

The SwissFEL injector and linac consist of 113 accelerating structures, driven by 33 high power (approximately 50 MW peak power) klystron amplifiers. Each klystron and associated accelerating structures are collectively referred to as one RF station. A more complete description of an RF station is given in “Radio Frequency Station System.” For each RF station a local control system is used to ensure a precise and stable RF voltage at the accelerating structures. As the overall beam characteristics depend on the entire RF-driven injector/linac chain, a “beam-based feedback” loop is required to coordinate between the RF station controllers. A centralized MPC-based supervisory control architecture has been developed [1], but this will not be the focus of the work discussed here.

The pulsed operation of the SwissFEL lends itself to
A Brief Introduction to FEL Machines

A linear accelerator (linac) is a machine used to accelerate charged particles along a linear path by subjecting the particles to a series of alternating electromagnetic fields. This method of acceleration was patented by R. Wideröe in 1928 [44]. Wideröe showed that electrons can be accelerated through a tube by applying a radio frequency (RF) voltage across the gap between the separated sections of the tube so that the electrons feel an accelerating electric field when they pass the gap [45]. If the system is designed in such a way that the electrons arrive at the next gap at the right phase of the RF voltage, they will again be accelerated. The linear accelerator extends the idea of Wideröe to a long linear array of accelerating “cells” powered by RF voltage sources in the power range of Megawatts and in the frequency range of Gigahertz. In addition to adjusting the consecutive cells so that each one is longer than the one before to account for the increasing particle speed, there are other engineering subtleties about matching the particle speed and the relative phase of the electromagnetic wave in the accelerator. In current linear accelerators the RF cavity of the accelerator is designed in such a way that the phase velocity of the electromagnetic wave matches the particle speed at the acceleration points.

The concept of free-electron laser (FEL) was introduced by Madey in 1971 [46]. In the FEL mechanism, the accelerated electrons are passed through an array of permanent magnets and experience transverse acceleration which results in the release of photons. This array of magnets is called an undulator, because it forces the electrons to wiggle transversely with respect to the beam axis. The underlying principle of generating intensive X-ray pulses lies in the principle of self-amplified spontaneous emission (SASE), which leads to microbunching. All electrons are initially distributed evenly and they emit spontaneous radiation incoherently. During the interaction of the electrons’ oscillations and the spontaneous radiation, they move into microbunches separated by one radiation wavelength. As a result, all electrons begin emitting coherent radiation in phase. This mechanism leads to an exponential increase of emitted radiation power which brings high beam intensities [47]. FELs based on the SASE principle are considered as the most promising light sources for extremely coherent light with wavelengths down to 1 Å (0.1 nm). Another key feature of FELs is that—by adjusting the undulator driving frequency—the wavelength of the X-ray pulse is adjustable over a significant frequency range. Some of the FEL facilities operating on the SASE principle are the European X-ray free electron laser (XFEL) and FLASH in Hamburg, the Linac Coherent Light Source (LCLS) at the SLAC National Accelerator Laboratory in the United States, the SPring-8 Angstrom Compact free electron laser (SACLA) in Japan and the SwissFEL at the Paul Scherrer Institute in Switzerland.

One of the required components in FEL machines is the electron accelerator. These accelerators are normally powered by high power klystrons, which require high voltage power supplies. The electron beam must be kept in vacuum during the acceleration which requires the use of numerous vacuum pumps along the beam line. Even though these devices make FEL machines large and expensive, the performance of FEL machines in generating high peak beam power and their tunability makes them highly desirable in a wide range of disciplines, including physics, chemistry, biology and medical diagnosis.

REFERENCES


Iterative Learning Control (ILC) techniques and an application, involving local control of the RF pulse shaping, is described. The RF pulse duration is short (1–3 μs), and no intra-pulse digital feedback loop is used within the RF pulse duration. Because the initial conditions for each RF pulse are essentially identical, and a feedforward calculation for the next pulse is required, ILC is preferred over repetitive control. Repetitive control is more applicable to the case where a periodic disturbance must be rejected in continuous-mode operation.

ILC can be applied as the appropriate shaping for one RF voltage pulse can be based on the information obtained from pulse shaping measurements taken from prior pulses. In the “two-bunch” mode successive electron bunches, separated by 28 ns, are accelerated within one RF pulse and then directed to separate undulator lines to generate separate X-ray pulses in each beam line. In this operation mode the two bunches must experience the same energy as they pass through the linac’s energized RF cavities. This problem can be addressed by an iterative control method that modifies the RF pulse shape. Because of the electron beam’s relativistic speed and extremely short duration, beam control actuators can be viewed as having no dynamics. Local actuator control can handle local actuation dynamics so that the actuators’ affect on the beam depends only on the actuation value at the instant when the beam passes. Beam control problems can therefore model the plant as a static map.

However, many actuators affect the beam’s properties and there is often significant uncertainty in the model of these effects. This naturally leads to the consideration of robust MIMO static control design problems and this approach is described here for the control of the electron beam energy, bunch compression and arrival time, or
equivalently and respectively, the wavelength, the peak intensity and the timing of the laser light. To achieve repeatable beam properties in the presence of slow parameter drifts an integrating control approach is used. The design of the beam injector control system is used to illustrate these methods.

Static MIMO control problems also arise in the design of the beam’s trajectory (or “orbit”) control. These problems also have the additional complication of actuator saturation. The more commonly used approach for orbit correction is based on approximately inverting the response matrix, with the possible refinement of regulating or filtering the singular values. The high uncertainty in the plant model cannot be handled in this way; the uncertain plant perturbations must be more explicitly accounted for. To improve the performance the shape and size of the robust perturbation set can be adapted as a function of prior beam orbit measurements [2].

This control approach is also applied to an optical pulse problem, that of laser pulse stacking. Mechanical rotation of a series of birefringent crystals creates an optical system that uses partial reflections to build a longer duration laser pulse from a short duration one. This is an interesting technical problem and the robust MIMO techniques give an automated crystal tuning method that can produce uniform extended laser pulses.

This paper presents only a selection of interesting and representative accelerator beam control problems using these methods. Further examples and more technical detail can be found in [3].

**CONTROL METHODS**

Several control design methods that are particularly applicable to the types of regulation and disturbance rejection problems that arise in the SwissFEL are described: iterative learning control, and two variants of an optimization-based robust static MIMO compensation design. A wider variety of methods are applied—for example PI control of the many temperature compensation loops within the RF structures—and while these are technologically challenging they are less specific to FEL problems and so are not described in detail here.

**Iterative learning control**

ILC is a method for iteratively generating a feed-forward (or combined feedforward/feedback) signal to track references or reject repeating disturbances in systems that operate in a repetitive, or trial-to-trial, mode. The method was originally introduced in [4] for robotic manipulation and remains an active area of research. See for example [5, 6] and the surveys [7, 8]. ILC has been applied to a wide range of applications including robot manipulators [9, 10], CNC machine tools [11], automotive control [12], nanopositioning [13], and aerial vehicles [14]. The approach described here is a variant of what is known as Norm-optimal Iterative Learning Control (NOILC) [6] which has been previously applied in on a beam accelerator machine [15, 16]. In contrast, the SwissFEL configuration precludes the use of digital intrapulse feedback and a completely feedforward approach is required.

The objective is the control of the plant output—in this case the measured RF structure field—over a finite time interval, \( y(k) \in \mathbb{R}^n, k = 1, \ldots, N \). The actuation \( u(k) \in \mathbb{R}^m \) is also applied over a finite interval and the input-output relationship is modeled by the linear relationship,

\[
y = Gu,
\]

where both \( y \in \mathbb{R}^{nN} \) and \( u \in \mathbb{R}^{mN} \) are vectors formed by stacking the \( N \) instances of \( y(k) \) and \( u(k) \) from each time interval. The assumption that \( u \) has \( N \) time points is true in the SwissFEL case but could be relaxed. If the system is causal \( G \in \mathbb{R}^{nN \times mN} \) is a block lower-triangular matrix. If it is also time-invariant then \( G \) is the block Toeplitz matrix of the impulse response matrices.

The idea behind ILC is to iteratively update the entire input trajectory \( u_i \in \mathbb{R}^{mN} \) to obtain the desired output trajectory \( y_{ref} \in \mathbb{R}^{nN} \). The subscript \( i \) denotes the iteration index of the ILC process. The calculation of the entire length-\( N \) input signal, \( u_i(k), k = 1, \ldots, N \), is performed prior to it being applied in the next iteration of the ILC process. This process is not causal with respect to \( k \), but as \( u_i \) is a function of \( u_{i-1}, y_{i-1}, \) etc., it is causal with respect to the iteration index \( i \).

The reference tracking objective is specified using a weighted norm criterion,

\[
\|y_{ref} - y\|_Y := \left( (y_{ref} - y)^TY(y_{ref} - y) \right)^{1/2},
\]

with the weighting matrix \( Y = Y^T > 0 \). Note that this formulation allows time-dependent weighting of the error \( y_{ref}(k) - y(k) \). The output is assumed to be corrupted by an additive disturbance \( d \in \mathbb{R}^{nN} \), and the measurement is corrupted by zero-mean noise \( n \in \mathbb{R}^{nN} \). At the \( i \)-th iteration, the measured output is

\[
y_i = Gu_i + d + n_i,
\]

The disturbance, \( d_i \), is assumed to be unknown but not varying from iteration to iteration. Under this assumption

\[
d = E\{y_i\} - Gu_i,
\]

where \( E\{\} \) denotes the expectation operator. At the \( i \)-th iteration the application of \( u_i \) is repeated over \( L \) pulses and \( E\{y_i\} \) is estimated (with \( \hat{y}_i \) denoting the estimate) in a batch manner,

\[
\hat{y}_i = \frac{1}{L} \sum_{j=1}^{L} y_{i,j}
\]

where \( y_{i,j} \) denotes the \( j \)-th repetition of the measurement with input \( u_i \). The choice of \( L \) is a trade-off between reducing the effect of noise and speeding up the ILC process. From \( \hat{y}_i \) the disturbance is estimated as

\[
\hat{d}_i = \hat{y}_i - Gu_i.
\]
The SwissFEL injector and linac consist of 113 accelerating structures, driven by 33 high power klystrons. Each klystron and associated accelerating structures are collectively referred to as one RF station. For each RF station a control system is designed to ensure a precise and stable RF voltage at the accelerating structures. Figure S1 illustrates a simplified layout of an RF station. Depending on the operating frequency, S-, X- or C-band, some components differ. For C-band RF stations, a pulse compressor is placed after the klystron to achieve high power RF at the output to power four accelerating structure [48, 49, 50]. The pulse compressor can increase the peak power by more than six times by shortening the RF pulse length. The details of the pulse compressor are discussed in “Klystron, Pulse Compressor and RF Gun.”

The RF sinusoid signal source is generated by the master oscillator. The waveforms of the in-phase, $I$, and quadrature, $Q$, components of the RF signal are up-converted to the carrier frequency through the vector modulator. The discrete $I$ and $Q$ waveforms each contain 2048 samples with a sampling time of $T_s = 4.2$ ns. The RF system operates in a pulsed-mode, that is, the short RF pulse of length 1 to 3 $\mu$s repeats every 10 milliseconds. The RF signal is amplified by the pre-amplifier to nearly 400 W to drive the klystron (a power amplifier) which delivers high RF power of the order of 50 MW at its output. This high power is produced by charging up capacitors, in the capacitor bank of the klystron high voltage modulator, and then discharging them through a low resistance load to deliver high voltage pulses to the cathode of the klystron. A high voltage power supply in the klystron modulator system determines the charged voltage on the capacitors. The Low Level Radio Frequency (LLRF) system can also directly change the value of the high voltage power supply (HVPS) through the EtherCAT interface.

The high RF power feeds the accelerating structures to build up the electric field inside the cavities for accelerating the beam. The RF voltage, at any node, is measured by directional couplers, and mixed down to the intermediate frequency (IF) of 39.6 MHz. The resulting signal is then sampled at the rate of 238 MHz by the ADCs, followed by an IQ-demodulation algorithm to obtain the discrete sequences of $I$ and $Q$, each containing 2048 samples. The sampling frequencies of the ADCs and DACs are identical in the LLRF system. These waveforms are also converted to the amplitude and phase waveforms to be more informative for the machine users. These measured values can be used as a feedback signal to update the $I$ and $Q$ waveforms to the DACs. More details about the LLRF design and architecture are reported in [51, 52].

REFERENCES
The control input for the \((i+1)\)-th iteration is the solution of an optimization problem,

\[
u_{i+1} = \arg \min_u J(u), \text{ subject to } \hat{y}_i = G u + \hat{d}_i,
\]

where \(\hat{y}_i\) and \(\hat{d}_i\) are given by (2) and (3) respectively. The optimality criterion is defined as

\[
J(u)_i = \| y_{\text{ref}} - \hat{y}_i \|^2_Y + \| u - u_i \|^2_U,
\]

where \(U\) is chosen to include a norm penalty on changing the input signal. This makes it possible to tune the convergence speed of the ILC procedure.

The solution to the optimization in (4) can be written explicitly as,

\[
u_{i+1} = u_i + (U + G^T Y G)^{-1} G^T Y (y_{\text{ref}} - \hat{y}_i).
\]

To reduce complexity, the weighting matrices \(U\) and \(Y\) can be taken as constant (that is, trial-invariant), and so the matrix inversion need only be calculated once. Input saturation limits can be directly included in the optimization formulation in (4), in which case it becomes necessary to solve (4) online in order to implement the controller.

NOILC has also been applied to the control of RF fields on the DESY FEL accelerator in Hamburg [16], but, because of the configuration and operational details, the DESY application is able to use both feedforward and feedback. The SwissFEL configuration cannot use feedforward feedback but has a longer time available for optimization back but has a longer time available for optimization formulation in (4), in which case it becomes necessary to solve (4) online in order to implement the controller.

Robust structured control design

Many of the electron beam properties are determined by the actions of multiple actuators. The dynamics in these cases are either handled by faster local loops or are only slowly changing, and these systems can be modeled by MIMO nonlinear static maps,

\[
y = f(u), \quad u \in \mathbb{R}^m, \quad y \in \mathbb{R}^n.
\]

The system can be linearized around the operating point \(y_0\) with the associated input \(u_0\),

\[
y - y_0 \approx R(u - u_0),
\]

where \(R \in \mathbb{R}^{n \times m}\) is commonly referred to as the system response matrix. The key features in this problem are the significant cross-coupling and significant uncertainty in \(R\).

For notational simplicity \(y\) and \(u\) are the linearized variables, and the nominal response matrix is \(R_{\text{nom}}\). Uncertainty in the system is captured via a set of models,

\[
y = Ru,
\]

\[
R \in \left\{ R_{\text{nom}}(I_m + W \Delta) \mid \Delta \in \mathbb{R}^{m \times m}, \| \Delta \|_2 \leq 1 \right\},
\]

where \(\| \|_2\) is the maximum singular value. The perturbation matrix \(\Delta\) is assumed to be unknown, but bounded in norm. The matrix \(W \in \mathbb{R}^{n \times m}\) is a weighting matrix specifying the level of uncertainty in \(R\). This particular formulation is well-suited to capturing uncertainty in the actuation which dominates in SwissFEL problems.

The uncertain model here is similar to those typically used in robust control theory, although here it is restricted to static system models. The problem of designing a static feedback, with a guaranteed level of performance for all models in the set is solved in [17]. This approach is extended to be able to specify an integral control structure using a delayed measurement. This control problem matches the measurement constraints in the SwissFEL case and gives the system the ability to track slow drifts in operating conditions.

The closed-loop schematic of the weighted control system design is illustrated in Figure 2. The dashed box indicates the controller which is composed of an integrator on each error channel and a constant matrix gain, \(K\). This structure introduces dynamics into the optimal control design problem, although the plant itself is modeled as a linear static system. There is also a single-sample delay in the measurement feedback loop. The problem is effectively a weighted sensitivity design with the weighting function,

\[
W_e(z) = \frac{\alpha}{z - z_0} I_{n \times n},
\]

with \(\alpha\) and \(z_0\) as the weight parameters.

This formulation can be transformed into the fractional interconnection shown in Figure 3 and described by the equation,

\[
\begin{bmatrix}
\dot{e}_{k+1} \\
x_{k+1} \\
\dot{e}_k \\
w_k
\end{bmatrix} =
\begin{bmatrix}
0 & -\alpha R K & 0 & -\alpha R \\
\alpha I_n & 0 & I_n & R \\
0 & 0 & WK & 0 \\
0 & 0 & 0 & 0
\end{bmatrix}
\]

\[
\times
\begin{bmatrix}
\dot{e}_k \\
x_k \\
r_{k-1} \\
v_k
\end{bmatrix}_K,
\]

together with those closing the loops,

\[
v_k = \Delta w_k,
\]

and

\[
\dot{e}_k = [z^{-1} I_n, 0, z^{-1} I_n] [\dot{e}_{k+1}].
\]

The partitioning in (6) corresponds to the signals shown in Figure 3.

The robust performance criteria is the \(H_{\infty}\) norm minimization of the closed-loop transfer function from the reference \(r\) to the weighted error \(\hat{e}\), which minimizes the closed-loop error of the worst-case plant,

\[
\beta_{\text{opt}} := \inf_{K \text{ stabilizing}} \sup_{\| \Delta \|_2 \leq 1} \sigma \left(N(K, \hat{\Delta})\right),
\]
where $N(K, \Delta)$ denotes the closed-loop transfer function from $r$ to $\hat{e}$. The augmented perturbation, $\Delta$ includes both the perturbation $\Delta$ and the $z$-transform variable,

$$\hat{\Delta} := \text{diag}(z^{-1}I_{2n}, \Delta).$$

See [18] or [19] for details of this robust control design framework.

The robust performance criteria can be reformulated with the upper bound [20, 17]

$$\beta_{\text{opt}} \leq \inf_{K} \| D M(K) D^{-1} \|_{\infty},$$

with $\mathcal{D}$ defined as the set of matrices,

$$\mathcal{D} = \{ \text{diag}(D, I_n, d_{\Delta}I_m) \mid D \in \mathbb{C}^{2n \times 2n}, D = D^* > 0, d_{\Delta} > 0 \}.$$

If $\beta_{\text{opt}} < 1$ then the closed-loop is stable and satisfies the weighted error performance criterion, $\| \hat{e} \|_2 \leq \| r \|_2$, for all $\| \Delta \|_2 \leq 1$. In the purely static robust design control problem solved in [17] the optimization problem in (9) is convex. However, the dynamics introduced by the weight $W_e$ and the integrators give an optimization problem that is bilinear with respect to $K$ and $D$ (and thus non-convex in general). A solution can still be obtained through an iterative scheme alternating between analysis and synthesis. In the analysis step, the control gain, $K$, is held constant, and the optimal scaling, $D \in \mathcal{D}$, is determined by minimizing the objective over $D$. In the synthesis step, the scaling $D$ is fixed and the control gain is synthesized by minimizing over $K$. This procedure is a special case of the robust control synthesis method known as $D$-$K$ iteration [20, 18]. As $M(K)$ is linear in $K$, each step of this procedure is convex. However the iteration scheme is not jointly convex in $K$ and $D$ and so is not guaranteed to converge to the global minimum. For a detailed analysis of the controller design see [21].

**Robust-adaptive control design**

If there is significant uncertainty in the response matrix $R$ then the perturbation set $R_{\text{nom}}(I + W \Delta)$ is large and the achievable performance may be too low. To overcome this problem a similar perturbation formulation, using an adaptation scheme which estimates the perturbation $\Delta$ from the past data, is developed. The static response matrix $R \in \mathbb{R}^{n \times m}$ is modeled as a member of a set of perturbed models in a similar manner to the previous section,

$$R = R_{\text{nom}}(I_m + \Delta), \quad \Delta \in \mathbb{R}^{n \times m}, \quad \| \Delta \|_2 < \gamma < 1.$$

The assumption that $\| \Delta \|_2 < 1$ implies that all $R$ in the set of perturbed models have the same rank. This is not restrictive in the applications considered here.

By imposing additional structure on $\Delta$ knowledge about the dominant sources of uncertainty can be exploited. Denote the set of structured block diagonal perturbations by $\Delta$. A more accurate representation of actuator uncertainty is modeled by,

$$\Delta \in \Delta, \quad \Delta = \{ \Delta \mid \Delta = \text{diag}(\delta_1, \ldots, \delta_m) \}.$$

FIGURE 2: Closed-loop schematic of the robust control design. The control block consists of integrators and a constant matrix gain $K$. Although the plant model is static, dynamics enter the design problem via the control input integrators, feedback delay, and tracking performance weight, $W_e$. 

FIGURE 3: Linear fractional representation of the closed-loop transfer function, $N(K, \Delta)$. The matrix $M(K)$ is linear in $K$ and given in (6).
In the robust adaptive scheme $\Delta \in \Delta$ is estimated via an optimization problem using past data. The estimate of $\Delta$ is then used in the calculation of the actuation for the next time instant, $u_{k+1}$. Each step of this control approach therefore requires the solution of two optimization problems. These problems are not difficult from a computational point of view, but the computation requirement is significantly higher than that for the robust design of the previous section.

The algorithm is split in three steps. In the first step, referred to as “initialization”, the input is determined by an optimization problem which minimizes the norm of the error for the worst case uncertainty in the model. Once this input is applied to the system and the measurements are taken, another optimization is performed to estimate the perturbation matrix, $\Delta$. This is discussed in the step “uncertainty estimation”. The third step involves finding the optimal input for the next control actuation. The estimation and control optimization give an actuation input which is applied to the system, and this is repeated until convergence to the reference is achieved.

Initialization

The initial input $u_0$ can be determined by minimizing the tracking error for the worst case perturbation,

$$u_0 = \arg\min_u \max_{\Delta \in \Delta} \|y^{ref} - R(I_m + \Delta) u\|_2$$

subject to $\|u\|_2 \leq u_{max}$

(11)

The cost function defined in (11) has an upper bound,

$$\|y^{ref} - R(I_m + \Delta) u\|_2 \leq \|y^{ref} - Ru\|_2 + \gamma \|R\|_2 \|u\|_2.$$ 

Note that without the $\Delta \in \Delta$ constraint there exists a $\Delta$ such that this limit is achieved. In either case this is used to calculate a conservative initial input

$$u_0 = \arg\min_u \|y^{ref} - R_{nom} u + \gamma \|R_{nom}\|_2 \|u\|_2$$

subject to $\|u\|_2 \leq u_{max}.$

(12)

The input $u_0$ is applied to the system and the output $y_0$ is measured.

Perturbation estimation

For iteration $k \geq 0$, the perturbation $\Delta_k$ can be estimated iteratively using input-output data. The objective is to determine the $\Delta_k \in \Delta$ that minimizes the discrepancy between the predicted and observed output,

$$\Delta_k = \arg\min_{\Delta \in \Delta} \|y_k - R_{nom} (I_m + \Delta) u_k\|_2$$

subject to $\|\Delta\|_2 \leq \gamma.$

(13)

In the uncertainty estimation (13), only the last input-output datum is used and the earlier measurements and inputs are not considered. This choice is made in order to better estimate the system around the last operating point. However, the solution to (13) is in some sense optimistic and also may not be unique. A more advanced uncertainty estimation and control method, developed in [3, Chapter 6], estimates the system response using the last most informative observation sets, and determines the control inputs in such a way that the parameter space is explored as much as possible.

Calculating the optimal input

The next step is to determine the optimal control input for $k \geq 1$ using the specific perturbation $\Delta_k$ estimated from the prior step.

$$u_{k+1} = \arg\min_u \|y^{ref} - R_{nom} (I + \Delta_k) u\|_2$$

subject to $\|u\|_2 < u_{max}$

$$\|u - u_k\|_2 < \delta u_{max}.$$ 

A bound on the allowable change in input, $\delta u_{max}$, limits the convergence speed of the method. The optimal solution of (14) is then applied to the system. The $\Delta_k$ and $u_k$ calculation steps are repeated iteratively until convergence is achieved.

The overall algorithm starts from a conservative input and then proceeds in a “confident” manner, taking only the last estimated perturbation $\Delta_k$ for use in the model of the plant in the next input calculation. This can be made more conservative in the iteration steps by instead using multiple prior $\Delta_k$ estimates to refine $\gamma$ and/or $R_{nom}$. Calculating $u_{k+1}$ now requires solving (11) which is potentially computationally expensive. This is similar in concept to the approach in [22].

**RF PULSE SHAPING BASED ON ITERATIVE LEARNING CONTROL**

In the two-bunch operation mode of the SwissFEL, two electron bunches, spaced 28 ns apart, are produced and accelerated within an RF pulse of length 1–3 $\mu$s. For pulsed RF machines, such as the SwissFEL, this means that the amplitude and phase of the RF pulses feeding the accelerating structures must be kept constant over the pulse length so that the two bunches experience the same energy gain.

This operation is repeated at a rate of 100 Hz and as the desired trajectory and the initial conditions are the same for each repetition this is an ideal ILC application. The objective is to produce flat-topped (or generally any desired shape) RF pulses, by modulating both the in-phase ($I$) and quadrature ($Q$) components of the input RF signal. See [23, 24, 25] for additional details. The ILC method has been applied on three different high power RF systems: the klystron, the pulse compressor and the RF Gun, which are described in more detail in “Klystron, Pulse Compressor and RF Gun.”

There are $N$ sampled measurements in the time interval of the pulse to be controlled. The ILC method de-
Klystron, Pulse Compressor and RF Gun

Klystron
The klystron amplifier is composed of a klystron and a high voltage modulator and provides the high power RF pulses for the beam acceleration. The RF power output can reach up to 50 MW, but only for a short duration of the order of microseconds. The klystron, shown in Figure S2, operates in a similar manner to a small linear accelerator. The high voltage modulator provides the electrical potential to accelerate electrons generated in the cathode towards the collector. The input RF signal, feeding the buncher resonator, modulates the velocity of the non-relativistic electrons. This velocity modulation results in a growing electron density modulation in the drift space. At the other end the catcher cavity provides an amplified RF signal by absorbing the electron beam energy. Figure S3 shows the klystron and its high voltage modulator for one RF station.

Figure S4 illustrates typical RF waveforms (amplitude and phase), measured at the klystron output. The ILC algorithm applies only over the shaded region (the “flat-topped region”) of the graph.

RF pulse compressor
The pulse compressor is a passive device used to store the RF energy and release it under certain conditions [53]. It essentially converts a long RF pulse to a short one with much higher peak magnitude. The SwissFEL pulse compressor (shown in Figure S5) is single Barrel Open Cavity (BOC) design which has a high quality factor resulting in a relatively long filling time and significant energy storage capacity [50]. In the original operation of RF pulse compressors, commonly referred to as the “phase jump” regime, the input phase is flipped by $180^\circ$, generating a reflected wave into the acceleration structures. This high power transient decays slowly giving the cavities time to build up an accelerating gradient higher than the klystron alone. With this operation of the pulse compressor, the peak power increases by a factor of up to six times.

The relationship between the klystron and pulse compressor voltage is [53]

$$\alpha V_c(t) = V_c(t) + \tau \dot{V}_c(t),$$

(S1)

where $V_c(t)$ and $V_c(t)$ are respectively the pulse compressor and klystron voltage phasors—expressed here as complex values to indicate both gain and phase at the modulation frequency. The constants $\alpha$ and $\tau$ are determined by the physical design of the BOC and the frequency of the RF wave, $\omega_0$, and in this case the values are $\omega_0 = 2\pi \times 5.712$ GHz., $\alpha = 1.82$ and $\tau = 1.086 \mu$s. To derive (S1), it has been assumed that the unloaded resonant quality factor is sufficiently high and that $V_p(t)$ is constant or only smoothly changing.

![Figure S3: SwissFEL klystron including high voltage modulator. The peak output power is approximately 50 MW.](image)

![Figure S4: Typical klystron pulse output. The measurement window consists of 2048 samples with sampling time of $T_s = 4.2$ ns. The electron-bunches are fired in the shaded region (after the filling time of the accelerator structures), which is referred to as the flat-topped region. The amplitude is normalized with respect to the saturation level (with a 5% headroom to the maximum power).](image)
Figure S5: Pulse compressor. This is based on a single Barrel Open Cavity (BOC), the circular device in the image. The $RF_{in}$ connection is the input from klystron, and the $RF_{out}$ connection directs reflected radio frequency power to accelerator structures.

For the case where the RF wave frequency and the pulse compressor resonant frequency are different, Equation (S1) is replaced by

$$\alpha V_c(t) = V_c(t)(1 + j\tau\Delta\omega) + \tau V_c(t),$$

where $\Delta\omega = \omega_0 - \omega_c$, with $\omega_c$ being the nominal angular resonant frequency of the pulse compressor. Choosing $\Delta\omega \neq 0$ is referred to as “detuning” and here $\Delta\omega = 2\pi \times 250$ kHz. The reflected wave from the pulse compressor, which is fed to the accelerating structure, is the output voltage of the pulse compressor. This reflected voltage, $V_c(t) = V_c(t) - V_c(t)$, is the quantity of interest in the control problem. Discretizing (S2) with Euler backward method (with sampling time $T_s \ll \tau$) and taking the Z-transform, lead us to the transfer function relating the klystron voltage to the output voltage of the BOC,

$$G_{BOC}(z) = \frac{V_p(z)}{V_p(z)} = \frac{T_z(\alpha - 1) - \tau - jT_z\tau\Delta\omega + \tau z^{-1}}{T_z + \tau + jT_z\tau\Delta\omega - \tau z^{-1}},$$

where $T_z = 4.2$ ns. The RF drive chain, including the vector modulator and pre-amplifier, is modeled as a first-order low pass system with a bandwidth determined by $\gamma$ and a scalar gain $K$ that can be complex to capture the loop phase. Therefore, the overall transfer function from the DAC inputs to the measured voltage at the BOC output is modeled as

$$G(z) = K \frac{1 - \gamma}{1 - \gamma z^{-1}} G_{BOC}(z),$$

and the drive chain bandwidth is

$$f_{3dB} \approx \frac{1}{2\pi T_s} \ln \gamma,$$

and here $\gamma = 0.7$ giving $f_{3dB} \approx 13.5$ MHz. Note that to derive (S3), it has been assumed that the input-output delay has already been taken into account in the measurements.

RF Gun

The particle source of SwissFEL is a photocathode located in the first RF station, referred to as the “RF Gun” [55]. In the RF Gun the electron bunches are generated by photo emission of a short laser pulse at the photocathode which is placed in a high-field RF cavity. The timing of the laser pulse is synchronized to the RF, so that electron bunches emerge when the energizing RF field on the cathode reaches an optimum value [56]. The RF photo-injector is a 2.5 cell S-band (3 GHz) standing wave cavity (see Figure S6). The layout of the actuation and measurement is depicted in Figure S7. Since a standing wave structure is used, the reflected power from the structure is relatively high. Thus, an RF circulator is placed between the klystron and the cavities to isolate the klystron from reflected power. The RF circulator is a three port waveguide system that dumps any reflected power from the cavities into a load.

The cavity is modeled as a driven LCR circuit [57],

$$\ddot{V}(t) + \frac{\omega_c}{Q_L} \dot{V}(t) + \omega_c^2 V(t) = \omega_c U(t),$$

where,

$$V(t) = V(t)e^{j\omega_0 t},$$

$$U(t) = U(t)e^{j\omega_0 t},$$

are respectively the output and input voltages with $V(t)$ and $U(t)$ as the phasors and $\omega_0 = 2\pi \times 2.997$ GHz, is the RF angular frequency and $\omega_c$ is the resonant frequency of the cavity.

Figure S6: SwissFEL 2.5 cell radio frequency (RF) gun. RF power from the klystron and isolating circulator enters at the top port. Not visible here is the photocathode that generates the electron bunch. The electron bunch exits the gun on the horizontal axis to be accelerated 700 m. through the linac and undulators and reach the target as an X-ray pulse.
RF digital processing

**Figure S7:** Radio frequency (RF) gun station schematic. The RF station control system for the RF gun is similar to those for the linac accelerating structures. The main differences are the inclusion of the circulator and the replacement of the accelerating structures by the RF gun.

Since the loaded quality factor is relatively high ($Q_L = 7591 \pm 147$) and the input varies only smoothly (i.e., $\dot{U}(t) \ll \omega$), the first time-derivative of the inputs and the second time-derivative of the outputs are neglected. The system equation, discretized with Euler backward method, is

$$ G_{\text{struct}}(z) = \frac{\mathcal{K}}{1 + T_s \tau - z^{-1} + j T_s \Delta \omega}, $$

where $\tau = \frac{2Q_L}{\omega_0}$ denotes the cavity filling time, $\Delta \omega = \omega_0 - \omega_c$ is the detuning frequency, and $\mathcal{K}$ is a complex gain factor. For SwissFEL the RF gun station has a sampling time of $T_s = 8$ ns. and the model parameters are $\tau = 0.8057$ µs. and $\Delta \omega = 30$ kHz.

**REFERENCES**


**Experimental ILC results on RF pulse profile control**

The experiments have been conducted at the SwissFEL test facility using a single C-band RF station. For a more complete description of the experimental system see “Radio Frequency Station System.” To illustrate the procedure only the ILC results of the RF pulse compressor are presented here. For the experimental results of the model-based and model-free ILC application to the klystron and RF Gun the interested reader is referred to [25]. The number of samples in the region of interest is \( N = 120 \), and for every iteration of the ILC algorithm, \( L = 10 \) waveforms are captured and averaged to reduce the measurement noise. The update rate of the input waveforms is relatively slow (approximately 1 Hz). This is because during the initial iterations, where the input waveform, and hence the energy dissipated in the structures, changes significantly, it is necessary to wait between iterations until the cooling system stabilizes the pulse compressor temperature.

Figure 4 illustrates the output waveforms of the pulse compressor after 20 iterations and the comparison with the initial waveforms. The ILC is applied only over the flat-topped region; the colored area in the figure. The initial output waveforms are generated by flipping the input phase by \( 180^\circ \), which results in a large step in amplitude at the BOC output. Figure 5 compares the initial input amplitude and phase waveforms with the final ILC-based generated waveforms. Figure 6 illustrates the standard deviation of the flat-topped region, as a measure of the pulse flatness, for amplitude and phase trajectories as the iteration proceeds. The iteration index “0” corresponds to the initial waveforms.

The convergence of the ILC method is not monotonic. This is due to the large input changes in the initial steps significantly changing the operating point and therefore the actual response matrix. In this case the ILC is compensating for errors in \( R \) (multiplied by \( u \)) as well as repeated transients and repeated disturbances.

ILC reduces the relative standard deviation of the amplitude to less than 0.004, and the flatness of the phase waveform is improved by a factor of approximately two. These results are significant in the context of this application.

**ROBUST CONTROL DESIGN FOR THE BEAM INJECTOR**

The control of the electron beam properties is a strongly coupled MIMO problem and because of the pulsed nature of the machine and the local control at the RF stations, the plant model can be considered as a static response matrix \( R \) in (5)). This plant model has a high condition number, significant cross-coupling, and a large amount of uncertainty. The robust structured control approach above is applied to address these issues.

The experimental configuration is illustrated in Figure 7 and shows the section of SwissFEL known as the...
To have a stable beam profile, several beam quantities need to be simultaneously controlled. These are: the beam energy \( E \), the energy spread \( \sigma_E \), the bunch length \( \ell \) and the charge \( Q \). The charge of the electron bunch is measured in the magnetic chicane for every bunch. For the purposes of this research, the three other quantities are measured in a destructive way using a spectrometer camera to take an image of the bunch. In the future, the SwissFEL injector will be equipped with nondestructive diagnostic devices for these beam measurements.

In order to estimate the bunch length, a transverse deflecting cavity is used to shear the electron bunch. The deflected beam is then diverted to the spectrometer monitor to measure the beam energy, energy spread and bunch length. Figure 8 shows the energy versus longitudinal position of a typical electron bunch measured at the spectrometer monitor. The Y and X axes correspond to the longitudinal position and energy within a bunch respectively. This beam image is referred to as the “bunch profile”. A Gaussian curve is fit to the data in each direction to give an estimate of the mean and standard deviation of the bunch profile in the X and Y axes.

The beam energy is proportional to the mean position of the bunch in X coordinate in the dispersive section of the spectrometer beam line,

\[
\Delta E \propto \mu_x,
\]

where \( \Delta E \) denotes the change in the beam energy in MeV, and where \( \mu_x \) is the mean of the beam position on X-axis.

A similar relationship applies to the energy spread \( \sigma_E \),

\[
\sigma_E \propto \sigma_x,
\]

where \( \sigma_x \) is the standard deviation of the beam position on X axis. The bunch length, \( \ell \), can be estimated from the standard deviation of the charge distribution in Y axis,

\[
\ell \propto \sigma_y,
\]

where \( \sigma_y \) is the standard deviation of the beam position on Y axis.

**Response matrix characterization and control design**

The four beam profile quantities to control make up the measurement vector,

\[
y = \begin{bmatrix} Q & \mu_x & \sigma_x & \sigma_y \end{bmatrix}^T.
\]

Four actuation variables are used to control the beam,

\[
u = \begin{bmatrix} \theta & a_s & \varphi_s & \varphi_x \end{bmatrix}^T,
\]

where, \( \theta \) is the laser intensity that acts on the cathode to extract the electrons, and \( a_s, \varphi_s, \varphi_x \) are, respectively, the amplitude and phase of the S-band and phase of the X-band cavities.

The response matrix is measured by perturbing the actuators around their nominal operating point and measuring the output vector [26, 27]. Figure 9 illustrates the typical experimental data used to estimate the response matrix model, \( R_{nom} \), for the SwissFEL injector test facility.

A simple and commonly used way of controlling the vector \( y \) is to use an inverse-based controller,

\[
u_{k+1} = u_k + \gamma R_{nom}^{-1} (y_{ref} - y_k),
\]

where \( \gamma \) is a constant gain chosen to give a desired convergence rate. Since the condition number of \( R \) is of the order of \( 10^4 \), this inverse-based control is potentially sensitive to uncertainties in the model which may result in poor performance. Prior work has used a singular value decomposition (SVD) method to invert the matrix, followed by filtering the singular values [28]. The robust structured design is experimentally compared to the more standard inverse-based design.

The controller obtained by the robust structured method in (9) is denoted by \( K_{rob} \). For comparison an inverse based controller of the form,

\[
K_{inv} = \gamma R_{nom}^{-1}
\]

is designed. The robustness of each of these designs is evaluated by structured singular value analysis. This amounts to determining the smallest perturbation, \( \Delta \) in Figure 2, that would cause the robust performance specification to fail. For the inverse controller \( K_{inv} \), the smallest such \( \Delta \) has size \( \| \Delta \|_2 = 0.12 \). The robust controller \( K_{rob} \) satisfies the performance objective for all \( \Delta \) up to size \( \| \Delta \|_2 = 0.83 \). This suggests that the \( K_{inv} \) may lead to very large transient errors if there is even a small discrepancy between \( R_{nom} \) and the actual value of \( R \).
**Figure 7:** A simplified schematic of the SwissFEL injector test facility. The electron beam is produced by a photocathode laser shining on the cathode (via the photoelectric phenomena). The electrons emitted experience a very high electric field in the radio frequency injector (RF Gun) so that they are accelerated as quickly as possible. The electrons are then passed through several S-band (3 GHz) cavities with an alternating electric field to gain more energy. After the X-band (12 GHz) cavities, the electron bunch is compressed by passing through a magnetic chicane that acts as a bunch compressor. In order to measure the bunch length, a transverse deflecting cavity is used to shear the beam and a spectrometer camera captures the beam profile.

**Experimental evaluation of the robust beam profile control**

The control algorithms ($K_{rob}$ and $K_{inv}$) were implemented on the SwissFEL injector test facility using MATLAB with machine communication via an EPICS control system network. The feedback rate is approximately 2 Hz, which is limited mainly by receiving and processing the beam image. Figure 10 illustrates the experimental results of closed-loop performance for the robust and inverse-based control schemes. In this experiment, the value of $\gamma$ (see (15)) was set to $\gamma = 0.005$ to slow down the convergence for clarity. The inverse-based control, $K_{inv}$, has a poor response, even with a low gain $\gamma$, and loses control of the beam after approximately 90 samples. The reference values are set as $r = [150 \text{ pC} \ -6 \text{ mm} \ 0.4 \text{ mm} \ 4.5 \text{ mm}]^T$.

The robust controller’s ability to track reference steps was tested in the experiment shown in Figure 11. The step reference was $r = [180 \text{ pC} \ -4 \text{ mm} \ 0.2 \text{ mm} \ 3.5 \text{ mm}]^T$, which represents a small step in each of the four beam profile parameters. The dashed lines denote the simulated response using the response matrix as the model. The difference between the model response and the measurement data comes from the model-plant mismatch.

The experiments demonstrate the better performance and robustness of the robust structured design method in this control problem. Inverse-based methods may fail or have poor performance when the response matrix has a large condition number. Although the robust control method is more computationally expensive in the design stage, the implementation is of similar complexity to the inverse based method. However, the robust control scheme gives a controller that can simultaneously control multiple beam quantities over a wide range of operating points.

**Figure 8:** Electron bunch profile. The image is obtained by the spectrometer camera. The beam energy, energy spread and bunch length are estimated by fitting Gaussian functions (shown as thin orange lines) to the image. The x-axis corresponds to the energy and y-axis to the bunch length.
**Figure 9:** Experimentally determined response matrix. A response matrix models the static linear system. The slopes of the linear fits (in red) are the entries of the estimated $R_{\text{nom}}$ matrix. For each experimental (blue) point shown the machine was brought to equilibrium and 50 measurements were averaged. The blue error bars indicate the sample standard deviation.

**Figure 10:** Closed-loop beam profile performance comparison. The inverse-based controller, $K_{\text{inv}}$, (in grey) lost control of the beam after approximately 90 samples, whereas the robust controller, $K_{\text{rob}}$, (in color) maintained good beam profile control. The experiments were carried out on the SwissFEL injector test facility using a feedback rate of 2 Hz.

**Figure 11:** Closed-loop beam profile step response. The robust controller, $K_{\text{rob}}$, is capable of quickly tracking steps in the beam profile parameters without overshoot. The dashed lines show a simulated response to the same step command references.
CONTROL OF THE ELECTRON BEAM ORBIT

To illustrate the application of the adaptive perturbation estimation methods described earlier the static beam position control problem is considered. In accelerator machines (such as synchrotrons and linear accelerators), the beam trajectory is controlled by a set of magnetic coils and beam position sensors, to keep the beam on the desired “orbit”—a circular path in synchrotrons, or a linear path in linacs. Figure 12 depicts a schematic of the orbit correction system in a linear accelerator. The beam position (in the transverse plane of x and y) is measured by the Beam Position Monitors (BPMs) as the beam passes through. The BPMs are distributed between several corrector magnets. These correctors are used to influence the beam trajectory in x and y axes through the currents $I_x$ and $I_y$ in the magnets.

Many orbit correction algorithms have been devised for synchrotron light sources. The task of the orbit correction is to determine a suitable set of currents in the corrector magnets in order to keep the beam on a specified trajectory. A static response matrix maps the changes in corrector currents to the changes in the equilibrium position of the beam at each BPM. Orbit correction methods usually attempt to invert the response matrix to translate the beam positions into the actuation on each of the corrector magnets. Procedures based on the singular value decomposition (SVD) are already in use at many accelerator laboratories throughout the world [29, 30, 31, 32, 33, 34, 35, 36]. In some cases, an adaptive approach has been proposed to estimate the response matrix [37], and some have also taken the corrector’s limits into account in the SVD algorithm [38].

Orbit correction algorithms can be classified as static or dynamic [39]. The objective of the static orbit correction is to minimize the steady-state orbit deviation from the desired trajectory. The process runs slowly and the feedback rate is normally approximately 1 Hz. In dynamic orbit correction the objective is to minimize the beam motion on some frequency range. The process rate is of the order of several kHz and may take the corrector magnets’ dynamics into account. For SwissFEL the plant is modeled by a static response matrix $R$, however, the control approach uses a robust-adaptive control scheme that captures some of the uncertainty of the response matrix, and treats the input actuation limits of the system appropriately [2]. In addition, the system response matrix can be estimated on-the-fly, and therefore the approach can be applied to time-varying systems as well [3, Chapter 6].

The response matrix of the system is measured by perturbing the currents in the magnets and measuring the effects on the BPMs. For BPM $i$-th’s x-position change the model is of the form,

$$\delta x_i = \sum_{j=1}^{N} r_{ix,jx} \delta I_{jx} + r_{ix,jy} \delta I_{jy},$$

where $N$ denotes the number of correctors, which is assumed for simplicity to be equal to the number of BPMs. This gives an estimate of the linearization of the non-linear static relationship between the corrector magnets and BPMs (see (5)). In the general case the model includes coupling between the vertical correction and the horizontal orbit. From (16), the full response matrix relating the current changes to the beam position changes for the x and y directions is

$$\begin{bmatrix}
\delta x_1 \\
\delta y_1 \\
\vdots \\
\delta x_N \\
\delta y_N 
\end{bmatrix} =
\begin{bmatrix}
r_{1x,1x} & r_{1x,1y} & \cdots & r_{1x,Nx} & r_{1x,Ny} \\
r_{1y,1x} & r_{1y,1y} & \cdots & r_{1y,Nx} & r_{1y,Ny} \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
r_{Nx,1x} & r_{Nx,1y} & \cdots & r_{Nx,Nx} & r_{Nx,Ny} \\
r_{Ny,1x} & r_{Ny,1y} & \cdots & r_{Ny,Nx} & r_{Ny,Ny}
\end{bmatrix}
\begin{bmatrix}
\delta I_{1x} \\
\delta I_{1y} \\
\vdots \\
\delta I_{Nx} \\
\delta I_{Ny}
\end{bmatrix},$$
or, by defining the appropriate incremental input and output vectors, $y = Ra$, in the same manner as (5).

Figure 13 depicts the measured response matrix to be used as the nominal model $R_{\text{nom}}$ in the SwissFEL injector test facility. For this experiment, 20 BPM sensors and 20 corrector magnet actuators were used. The matrix has a checkerboard pattern as the x and y directions are decoupled.

**Experimental robust-adaptive beam orbit control results**

The robust-adaptive optimization was experimentally tested on the SwissFEL beam orbit correction problem. Current saturation in the correction magnets is a significant limitation and so actuation limits are applied in the control algorithm optimization in (14). The control input optimization is augmented with a penalty function to help avoid the strong magnetization non-linearities as the current approaches its limit. The modified optimization problem is,

$$u_{k+1} = \arg \min_{u} \| y - R_{\text{nom}} (I + \Delta_k) u \|_2 + p(u)$$
subject to $\| u \|_2 < u_{\text{max}}$

$$\| u - u_k \|_2 < \delta u_{\text{max}}.$$ (17)

In this experiment, the penalty function was chosen as

$$p(u) = \alpha \sum_{i=1}^{m} \left( \frac{u_i}{u_{\text{max}}} \right)^{10},$$

where $u_i$ denotes the $i$-th component of the $u$ vector and $\alpha$ is a constant. Since the orbit system response is in a triangular form, any small perturbation in the corrector currents affect all downstream BPMs. Therefore, uncertainty is modeled by perturbations on the actuators, that is, on the currents in the corrector magnets. To achieve this effect the perturbation matrix structure, $\Delta$, is constrained to be diagonal (as in (10)) and post-multiplies $R_{\text{nom}}$ in (17).

The algorithms were implemented in MATLAB scripts using the CVX convex optimization solver [40, 41]. The communication to the machine has a control protocol with a round-trip communication time of approximately...
Figure 12: Linear accelerator orbit correction system. The beam position monitors (BPMs) are the beam position sensors in the x and y axes. The corrector magnets are the actuators and drive the transverse components of the beam’s trajectory.

Figure 13: Experimentally determined response matrix $R$ for the SwissFEL injector test facility beam. The lower triangular form arises from the fact that corrector magnets cannot affect the upstream beam orbit. The individual magnets have variable strength and spatial influence. The condition number of the matrix is of the order of $1.5 \times 10^4$.

1 second and the optimization problems are easily solved within this time limit. Figure 14 illustrates the experimental results of robust orbit correction at the SwissFEL injector test facility using 10 BPMs. The beam is brought to the origin after approximately 20 seconds.

The beam position is initially off-axis over the whole BPM chain. In this experiment all BPMs are equally weighted, however different weights can easily be assigned to different locations along the beam line. The final beam position RMS error is approximately 0.025 mm which is 25 times less than the initial error. Figure 15 shows the beam position in x and y axes for different numbers of iterations. Correction magnets 9 and 10 (close to the downstream end of the beam line) are operated at close (within 10% of range) to their saturation limit.

LASER PULSE-STACKING AUTOMATION

Another example of the robust-adaptive control optimization method is the calibration of the laser pulse

Figure 14: Closed-loop robust orbit control over time. The robust control design, with online estimation of $\Delta_k$, reduces the orbit error within 20 samples (20 s.). The beam position in x and y axes, are measured at 10 beam position monitors.

Figure 15: Closed-loop robust orbit control illustrated spatially. A large initial offset, and significant variation along the beam, is reduced by the orbit control algorithm. The beam position in x and y directions is measured at 10 beam position monitors (BPMs) for selected iterations.
stacking system. This is an unusual application of feedback technology to a combined optical-RF calibration problem that is normally tuned “by-hand.” The calibration system to be designed is described in detail in “Laser Pulse-Stacking Mechanism.” The production of high-brightness electron beams is challenging as it introduces high demands such as a flat-topped pulse shape on the driving laser pulse applied to a copper photocathode.

The objective of the laser control is to generate a desired beam current profile by acting on the motor-driven crystals. To measure the current profile (charge distribution over the electron bunch length), the beam is deflected by a transverse deflecting cavity and a camera takes an image of the longitudinal profile which is then processed to extract the charge distribution profile [42]. The total duration of the profile is of the order of 5 ps. which again precludes the use of any intrapulse control. The pulse shape is influenced mainly by the RF accelerating cavities and the laser stacker system. In this control scheme, the pulse shape is modified by manipulating the crystals’ angles, which in turn, changes the relative intensities of the stacked pulses in the laser profile. To generate a flat-topped current pulse, the flat-topped region of the profile is divided temporally into several consecutive intervals (or “cells”) [2]. The beam current is averaged over each cell to give an estimate of flatness error which is then used to steer the crystals. In the SwissFEL test injector machine, the four crystals in the laser stacker system are motorized so the angles can be set remotely. Dividing the flat-topped region into more cells gives more information about the pulse shape, but as since the number of actuators is limited, adding more than a certain number of cells does not improve the system performance. Therefore the number of cells in the flat-topped region has been determined empirically. Because the electron beam is measured for feedback, the control system compensates for non-ideal behavior in both the laser stacker and the RF gun.

**Experimental evaluation of robust-adaptive laser crystal stacking**

In this experiment, the flat-topped interval of the pulse was split into 16 cells. The current at each cell is averaged and subtracted from the mean of the flat-topped region. This gives an output vector of length 16 representing the error in the beam profile flatness.

The response matrix relating the changes in crystal angles to the changes in the beam intensity $I$ is estimated and used to initialize the control and uncertainty estimation algorithms. As 4 crystals are used to modify the pulse shape, the response matrix is an $R_{16\times4}^{16\times4}$ matrix. It is initially estimated by perturbing the crystal angles and measuring all 16 cells. In this case, significant cross-coupling is present and no specific form is assumed for the perturbation matrix; $\Delta$ is a full (or unstructured) matrix. The response matrix, $R$, is updated iteratively in an adaptive manner,

$$R_{k+1} = R_k + R_k \Delta_k,$$

where $\Delta_k$ is the perturbation matrix estimated via (13). The processing time of each iteration is approximately 2 seconds which is determined by the computation time and speed of the crystal motors.

Figure 16 compares the final pulse shape after 50 iterations to the initial one. To further quantify the initial and final variations, Figure 17 shows the initial and final error value estimated for each of individual cells.

To illustrate the convergence characteristics, Figure 18 plots the standard deviation of the current pulse over the flat-topped region to provide a measure of flatness as a function of the control iteration number. The relative standard deviation is improved by a factor of 3, starting from an arbitrary beam current profile. To quantify the degree to which the $\Delta$ adaptation tightens the characterization of $R$, Figure 18 also plots the 2-norm of $\Delta_k$ versus iteration index. The norm, $\|\Delta_k\|_2$, decreases to the level of the residual noise. The perturbation bound was set to $\gamma = 0.3$.

Figure 19 compares the longitudinal profile of the electron beam, captured on the transverse profile monitor of the deflecting cavity, before and after applying the algorithm. The final image is taken after 50 iterations and shows that the charge is more uniformly distributed over the bunch length. The maximum rotation angle of crystals (with respect to their initial state) is approximately 4 degrees. This control approach is extremely effective at tuning the crystal angles and providing a very good electron beam pulse.

**SUMMARY AND CONCLUSION**

This article illustrates some of the RF and beam control problems in an FEL machine that can be addressed by advanced control methods. Four case studies illustrate both the interesting technical control issues that arise in pulsed-mode accelerator beams, and effective methods for handling them. In most cases, the control is based on an optimization scheme. Controlling the RF voltage inherently controls the electron beam profile, which includes the wavelength and the intensity of the resulting X-ray pulse. Several methods have been developed to control the RF voltage at the accelerating modules, either with or without the beam information.

For a pulsed machine, such as the SwissFEL, repetitive disturbances on the RF pulse can be suppressed using an Iterative Learning Control approach to correct the RF excitation for subsequent pulses. ILC is particularly well suited to pulsed applications such as this as the reference and initial conditions are the same from pulse to pulse. The use of an optimization based ILC method opens up the possibility of applying additional constraints or objectives to suit specific aspects of the configuration or operation.
Laser Pulse-Stacking Mechanism

Triggering an electron bunch from the cathode requires a laser pulse of a specified duration. This pulse is created by passing a shorter duration Gaussian pulse through a laser pulse-stacker. This creates a longer quasi-flat-topped pulse in the temporal domain [58, 59].

A laser pulse-stacker is composed of $N$ birefringent crystals that transform a single Gaussian pulse into a temporally spaced sequence of $2^N$ Gaussian pulses that overlap to produce the long flat-topped pulse. Figure S8 illustrates the concept of laser pulse-stacking for two crystals. The initial single pulse, which is linearly polarized in the vertical direction, is passed through the first crystal with the optic axis tilted by some degrees relative to the vertical axis. The crystal splits the input pulse into two pulses separated in time by $\Delta t_1$. The delay between the two pulses is constant and depends on the characteristics of the crystal such as its length and refractive indices. The relative intensity of the two laser pulses can be adjusted by rotating the optic axis. The two resulting pulses from crystal 1 are oriented at different angles to the vertical. These two intermediate pulses then pass through the second crystal to be each divided into 2 pulses separated by $\Delta t_2$. The whole process is repeated by adding more crystals so that with $N$ crystals the initial single pulse is divided into $2^N$ overlapping pulses.

In the SwissFEL laser pulse-stacker four crystals are used giving $2^4 = 16$ pulses that overlap in the temporal domain to approximate a longer flat-topped pulse. The crystal angles act as actuators as they adjust the relative intensity of the pulses split by each crystal. Because all downstream intensities are also changed the response matrix has strong cross-coupling.

REFERENCES


Figure S8: The schematic diagram of laser pulse-stacking concept. The arrows represent the polarization axis. The initial single pulse is passed through 2 birefringent crystals to produce 4 stacked pulses. The first crystal with length $L_1$ splits the input pulse into two intermediate pulses, separated by $\Delta t_1$. The resulting pulses are then passed through the second crystal with length $L_2$ to be further split and delayed by $\Delta t_2$. The process is repeated to finally produce $2^N$ overlapping pulses by passing the initial pulse through $N$ crystals.

Figure 16: Injector beam current profile control. Iterative uncertainty estimation and optimal control design significantly improves the flatness of the beam current profile. The profile is measured using the transverse deflecting cavities. The flat-topped region of the pulse is divided into 16 intervals and each interval is averaged to estimate the pulse shape and flatness error.

Figure 17: Flatness error in the injector beam current profile. The control significantly reduces the initial error in flatness. The error is calculated by averaging over each interval (cell) and subtracting the mean of the entire flat-topped region.
FEL components are subject to slow variation over time and this makes model-based solutions potentially sensitive to the significant uncertainty in the plant. To cope with this drift and uncertainty in the system response, two approaches have been taken: a conservative robust one, which is based on the worst-case error minimization [21], and a robust-adaptive solution to effectively estimate the system response on-the-fly [2].

Switched-mode accelerator control raises many more control issues than the small sample discussed in this article. An example is the voltage control system within each RF station. See [43] and [3, Chapter 3] for more on this problem. With a large number of RF stations contributing to the beam quality, deciding between a centralized or decentralized control architecture is important. A Model Predictive Control approach, specifying the RF voltage setpoints over multiple stations in the beam line to control beam energy is described in [1].

The control strategies discussed here are of course applicable to a wider range of pulsed-beam accelerator machines and RF systems and our hope is that advanced control methods will see greater application in this interesting problem domain and others that share some of its characteristics.

ACKNOWLEDGMENTS

This research work was supported by Paul Scherrer Institut, Switzerland. The authors would like to thank Thomas Schietinger, the beam-time coordinator of the SwissFEL injector test facility, as well as Riccardo Zennaro and Alessandro Citterio for providing materials regarding the accelerating structures. Moreover, we would also like to express our gratitude to the Low Level RF team at the Paul Scherrer Institut, with special thanks to Roger Kalt for his help and support.

REFERENCES


