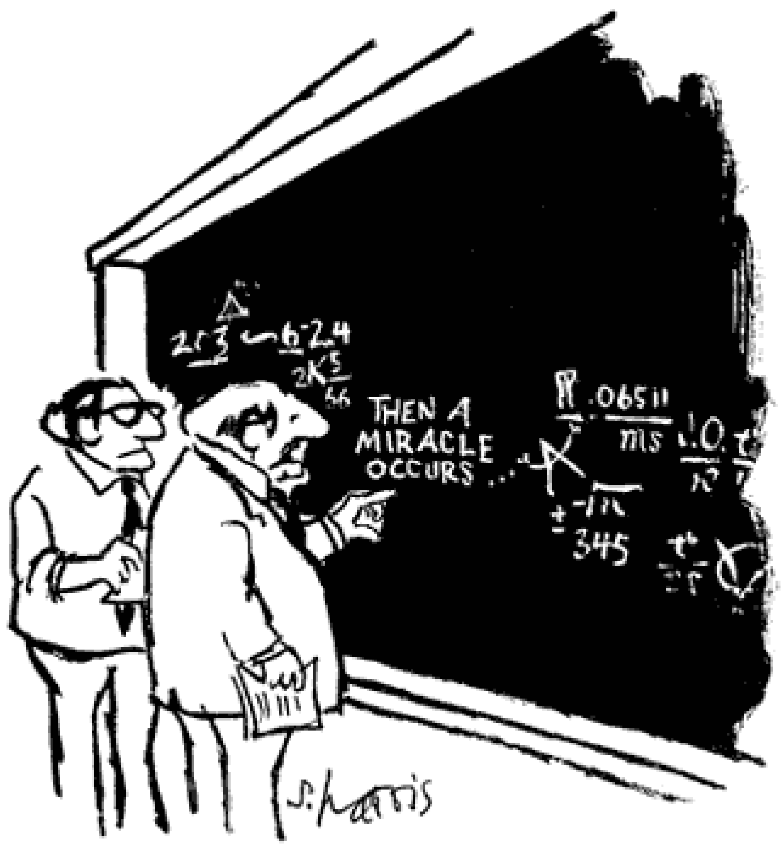
11th International Workshop on Personal Computers and Particle Accelerator Controls

PCaPAC 2016

October 25th – 28th, 2016
Campinas, Brazil

Feedback Control for Particle Accelerators



"I think you should be more explicit here in step two."

Ralph J. Steinhagen

Overview

- Part I – Introduction
  - Real-World examples, similarities across domains and its synergies
  - Classic 'control theory' recap: s-parameter, time- & frequency-domain definitions, terminology, etc.
    - Stability, Controllability, Observability
  - a practical but not-so-optimal PID tuning strategy

- Part II – Optimal Linear Multivariate- & MIMO-Controller Design
  - Space & Time Domain concepts
  - Trade-off between disturbance rejection & noise attenuation
  - Examples

- Part III – Optimal Non-Linear Controller Design
  - focus on latency/lag- & rate-limiter compensation (communication, digitization, GBW-limits, power-limits, etc.)
  - Inter-loop dependencies: cross-dependability and cross-constraints between feedback loops
  - Robustness and modelling errors
  - best practices: control room-level integration, system validation, improvement of model/feed-forward

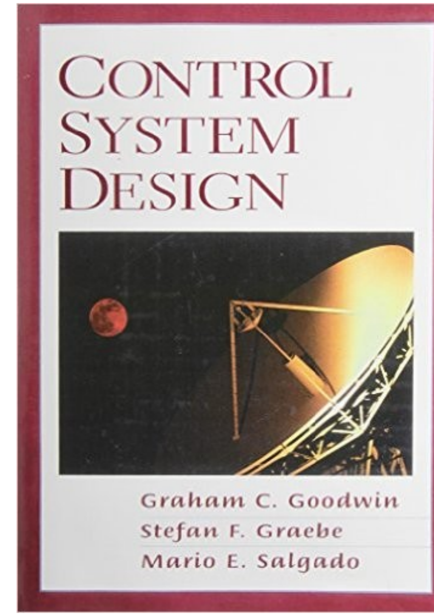
- Part IV – Discussion, Open-Round and more detailed Q&A

- Primary goal: provide a roadmap to avoid less obvious FB 'pot holes'

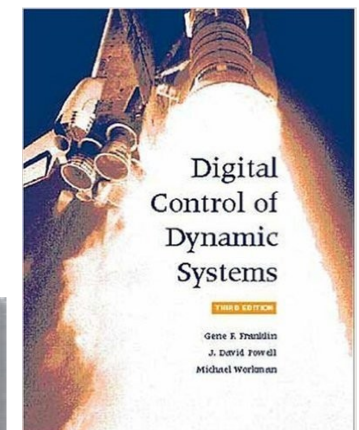
N.B. please feel free to interrupt me
in case you have pressing questions

Literature

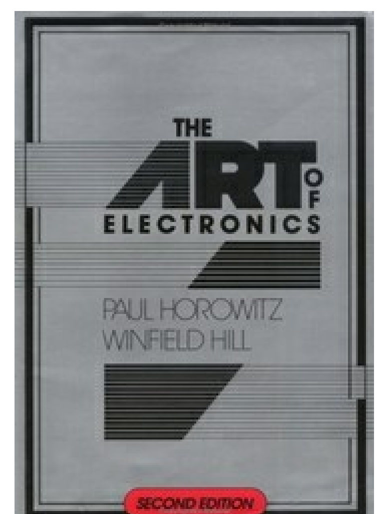
- M. Salgado G. Goodwin, S. Graebe,
  "Control System Design", Prentice Hall, 2000



- G. F. Franklin, J. D. Powell, Abbas Emami-Naeini,
  "Feedback Control of Dynamic Systems", 7th Edition,
  Pearson, 2014
- Gene F. Franklin, J. David Powell, Michael Workman,
  "Digital Control of Dynamic Systems", 3rd edition,
  Addison Wesley, 1997



- Paul Horowitz & Windfield Hill,
  "The Art of Electronics", Cambridge Uni.-Press

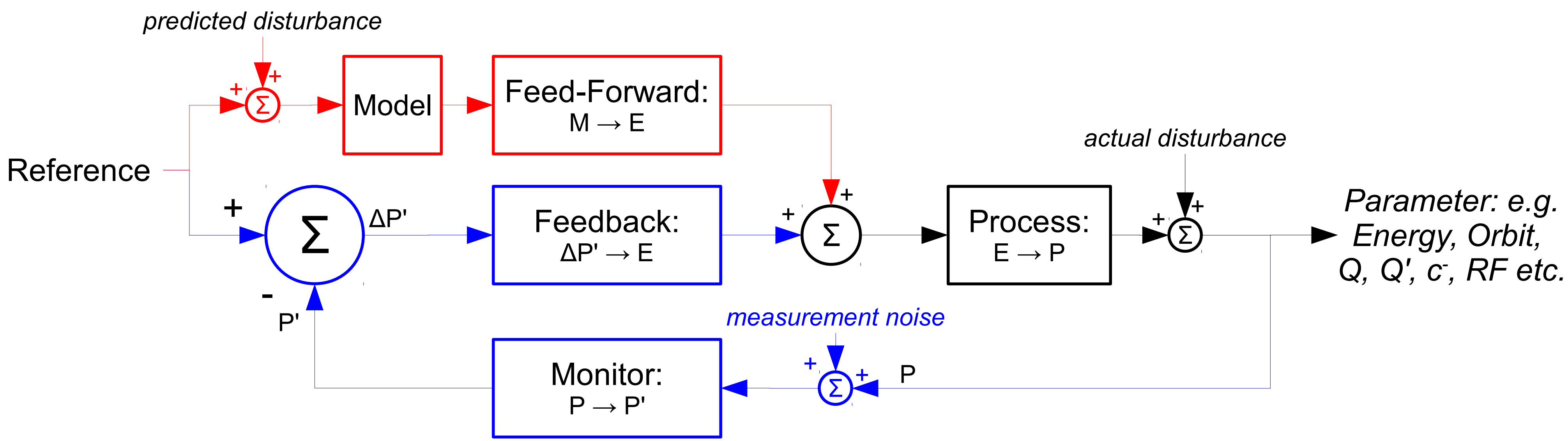


- IEEE Transactions and Journals, notably:
  - Instrumentation and Measurement
  - Microwave Theory and Techniques
  - Control Theory & Applications
- … and of course: https://www.microwaves101.com/

Control Paradigms I/II

Parameter control, either through...

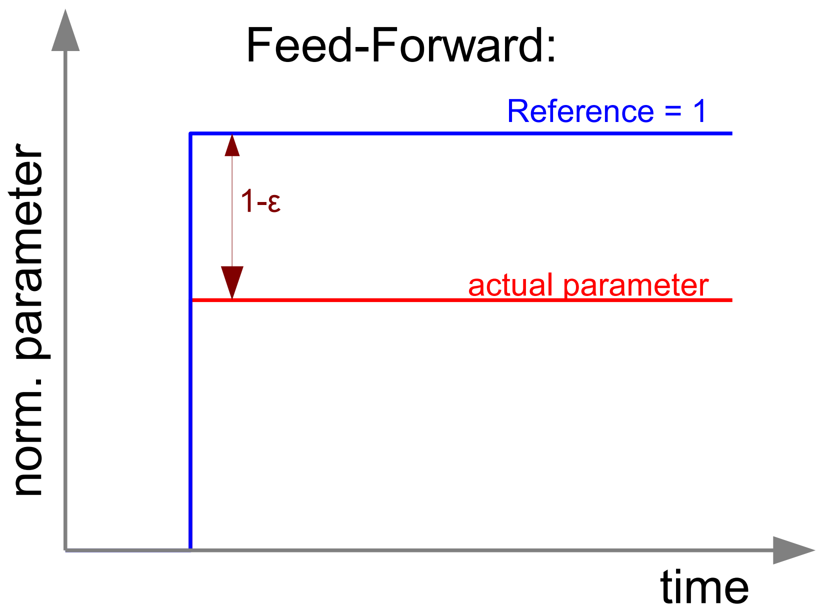
- Feed-Forward: (FF)
  – Steer parameter using precise process model and disturbance prediction
- Feedback: (FB)
  – Steering using <u>rough</u> process model and measurement of parameter
  – Two types: within-cycle (repetition Δt<<10 hours) or cycle-to-cycle (Δt>10 hours)

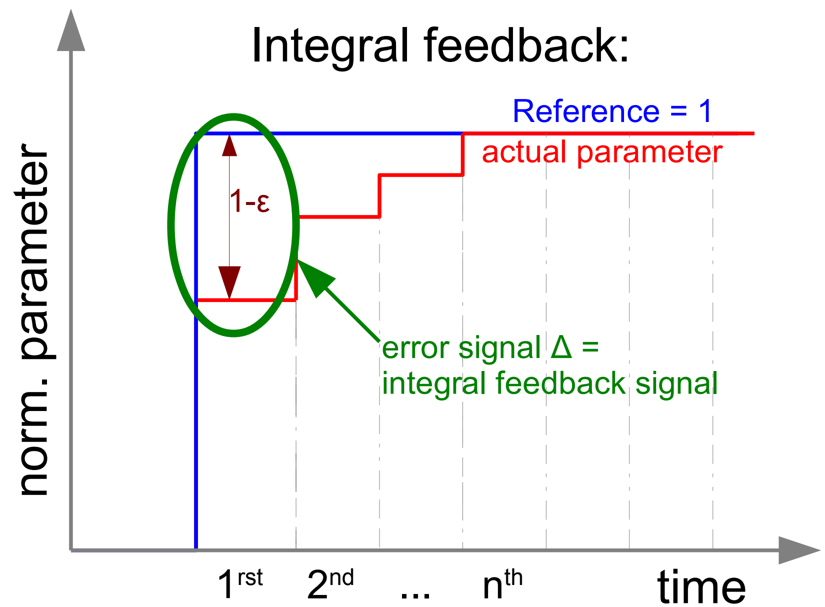


Control Paradigms II/II

- Machine imperfections cause steady-state offset ε_{ss} and scale error ε_{scale}:

$$\Delta x(s) = R_i(s) \cdot \delta_i \;\rightarrow\; \Delta x(s) = R_i(s) \cdot (\epsilon_{ss} + (1 + \epsilon_{scale}) \cdot \delta_i)$$





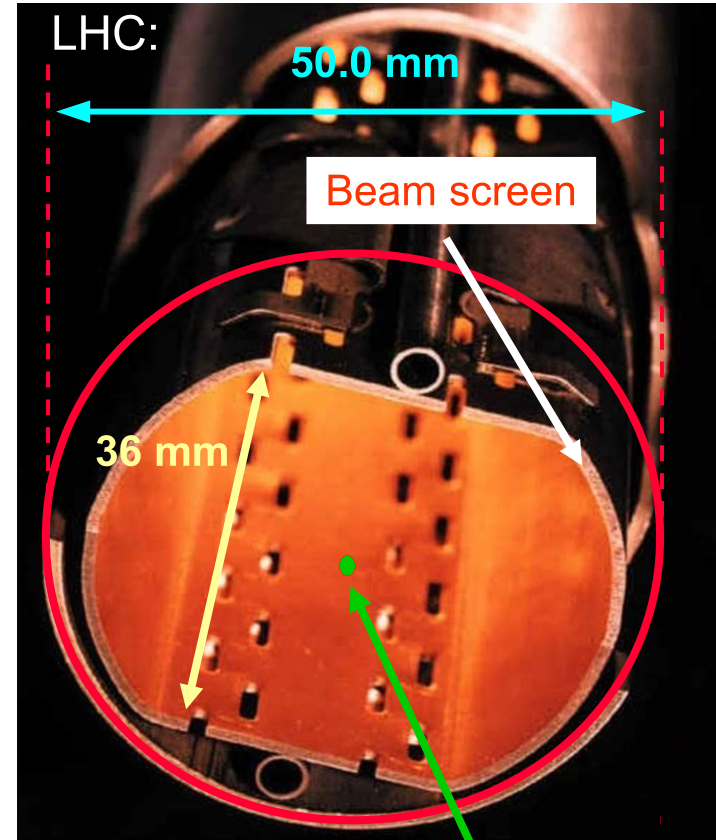
- Uncertainties and scale error of beam response function affects convergence speed (= feedback bandwidth) rather than achievable stability

Beam Parameter Stability in Particle Accelerators
… notably in Hadron Machines

- Traditional requirements on beam stability…

... to keep the beam in the pipe!

- LHC's increased stored intensity and energy much tighter requirements on beam stability:
  1. Capability to control particle losses
     - Machine protection (MP) & Collimation
     - Quench prevention
  2. Commissioning and operational efficiency



Beam 3 σ envel.
~ 1.8 mm @ 7 TeV

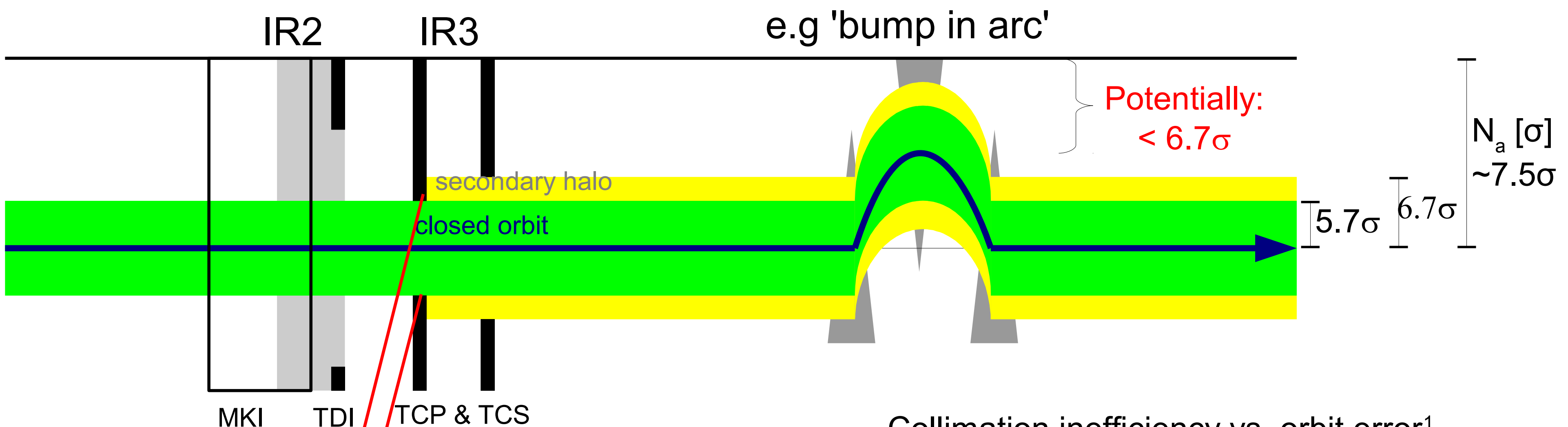
- FBs became a requirement for safe and reliable nominal LHC operation
  – implications on controller reliability, availability and system integration

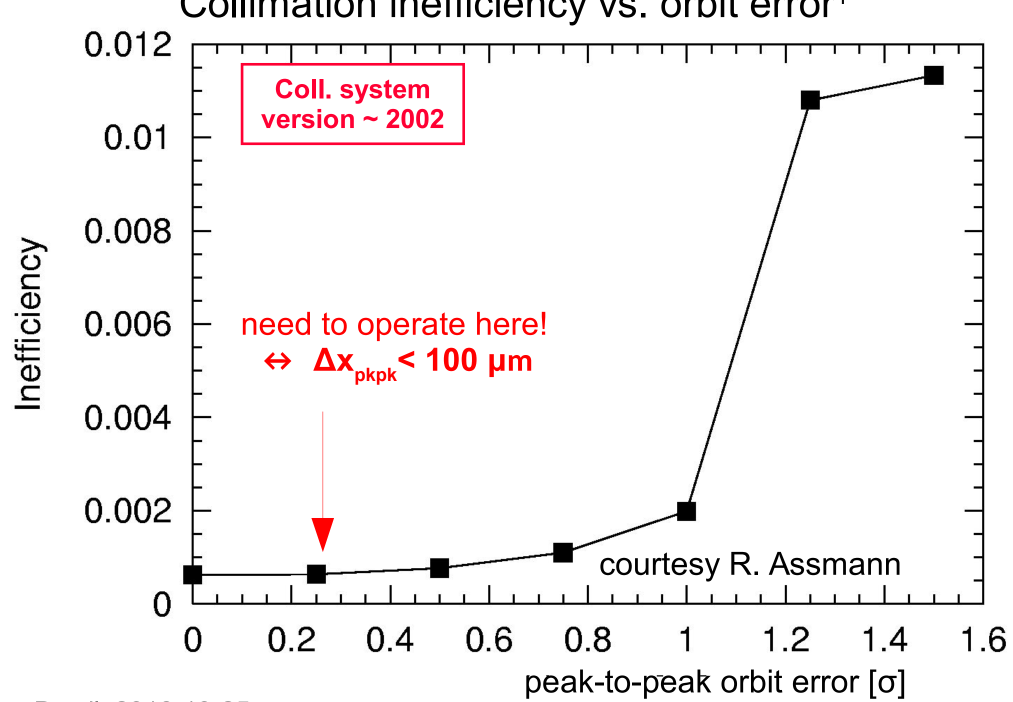
Beam Parameter Stability in Particle Accelerators

LHC Requirements on Orbit – Machine Protection

Combined failure[1]: Local orbit bump and collimation efficiency (/kicker failure)

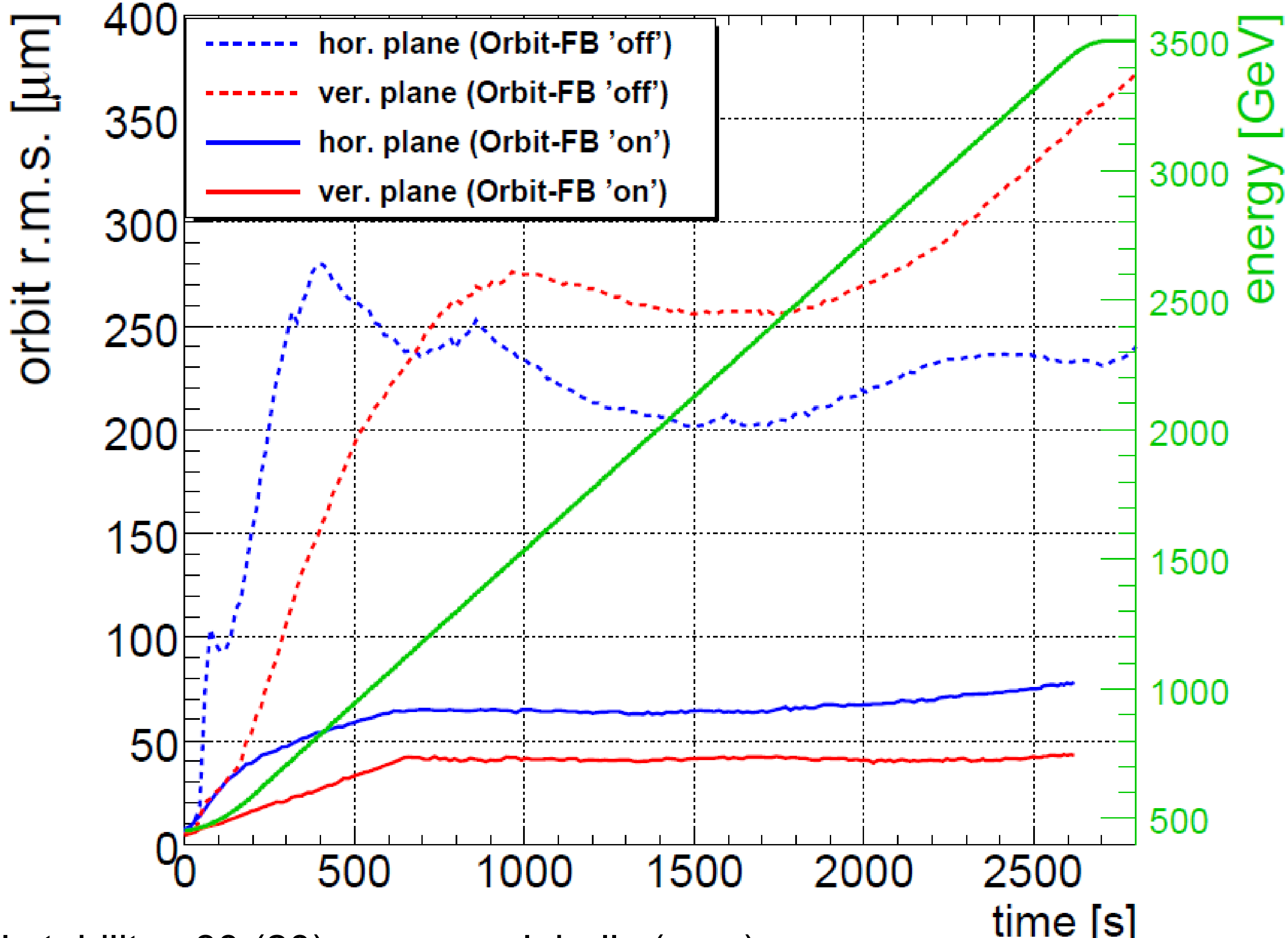
→ local orbit bumps may potentially compromise collimation function





LHC Feedback Operation – Example

- Orbit feedback used routinely and mandatory for nominal beam



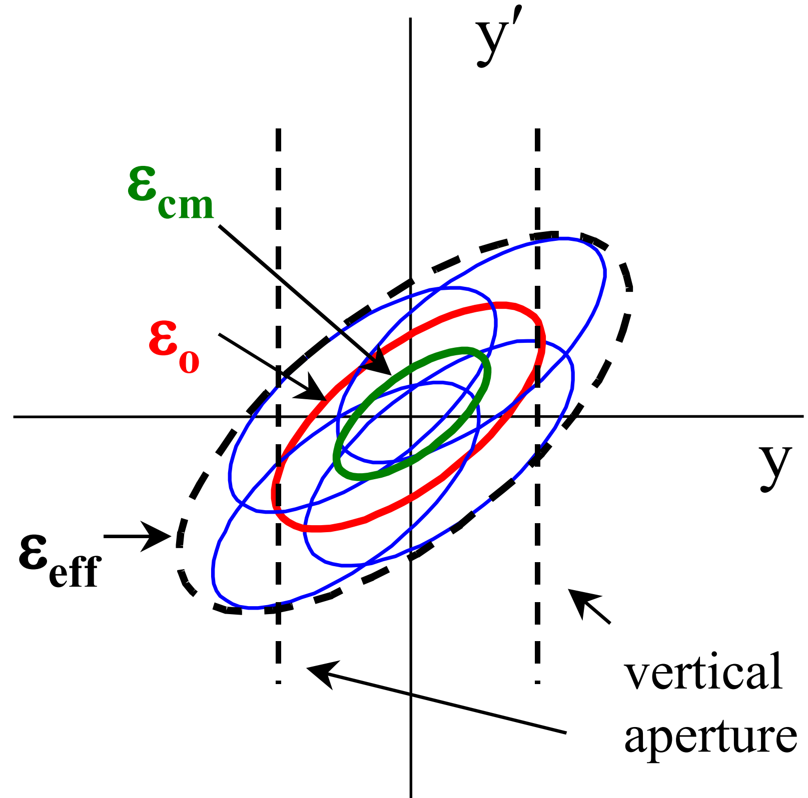
- Typical stability: 80 (20) µm rms. globally (arcs)
- Most perturbations due to Orbit-FB reference changes around experiments

Beam Parameter Stability in Lepton Machines
(e^+e^- Collider, Light Sources, ...)

- Main requirements for orbit stability[8]:
  - Effective emittance preservation
    (τ_d sampling/integration time, τ_f fluctuation time)

$$\tau_d \gg \tau_f: \quad \epsilon_{eff} = \epsilon_0 + \epsilon_{cm}$$
$$\tau_d \ll \tau_f: \quad \epsilon_{eff} \approx \epsilon_0 + 2\sqrt{\epsilon_0\,\epsilon_{cm}} + \epsilon_{cm}$$

    - Minimisation of coupling (vertical orbit in sextupoles)
    - Minimisation of spurious dispersion (vertical orbit in quadrupoles)
  - Collider Luminosity and collision point stability



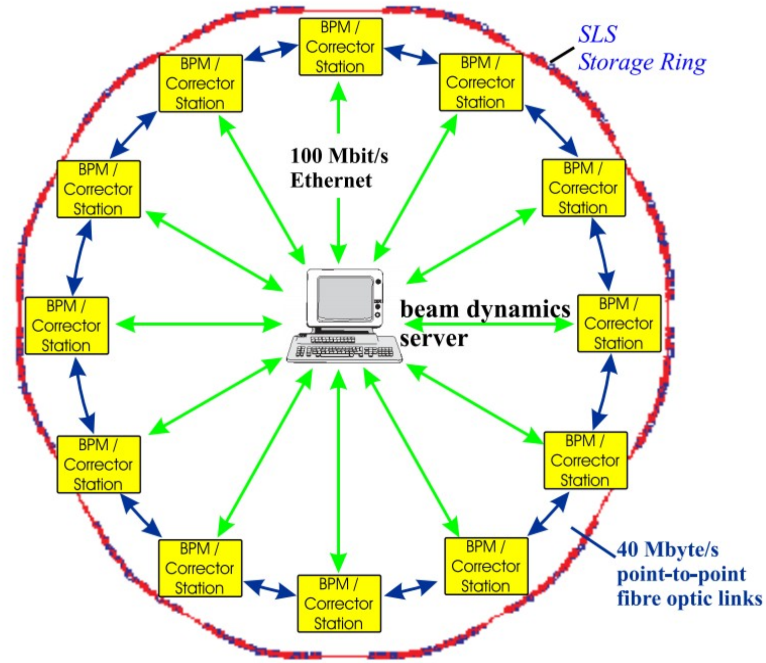
L.Farvacque, ESRF

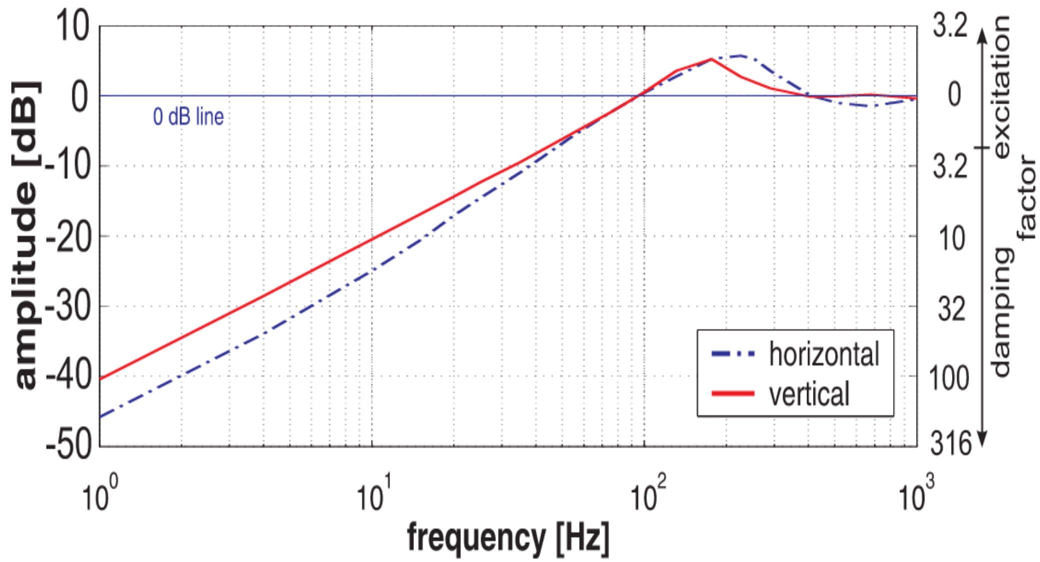
$$L = L_0 \cdot \exp\left\{\frac{(\bar{x}-x)^2}{2\sigma_x^2} + \frac{(\bar{y}-y)^2}{2\sigma_y^2}\right\} \cdot 1/\sqrt{1+\left(\frac{\theta_c\,\sigma_z}{2\,\sigma_{x/y}}\right)^2} \cdot \quad \ldots$$

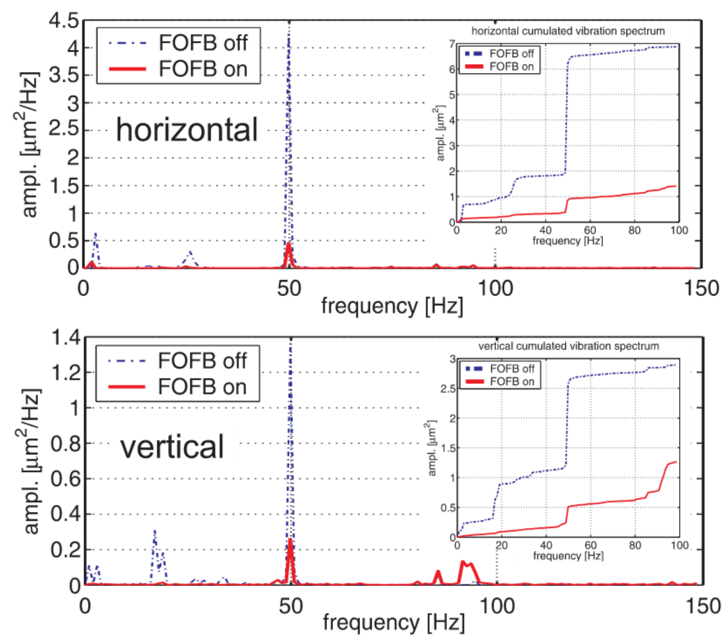
→ Nearly all 3rd generation light-sources deploy at least orbit/energy feedbacks[1-3]

Beam Parameter Stability in Lepton Machines
(e^+e^- Collider, Light Sources, …) here: Swiss-Light-Source, PSI

- Orbit-FB @ Swiss-Light-Source, PSI (SLS)
  - um-resolution orbit stability achieved during routine operation
- Organised IWBS'04: http://iwbs2004.web.psi.ch/
  - very good and well organised workshop!
  - validated, basis and jump-started Orbit-FB designs of many other synchrotron light sources & LHC to follow







courtesy T. Schilcher, M. Böge, B. Keil et al., PSI

Higher-Order Beam Parameter Stability

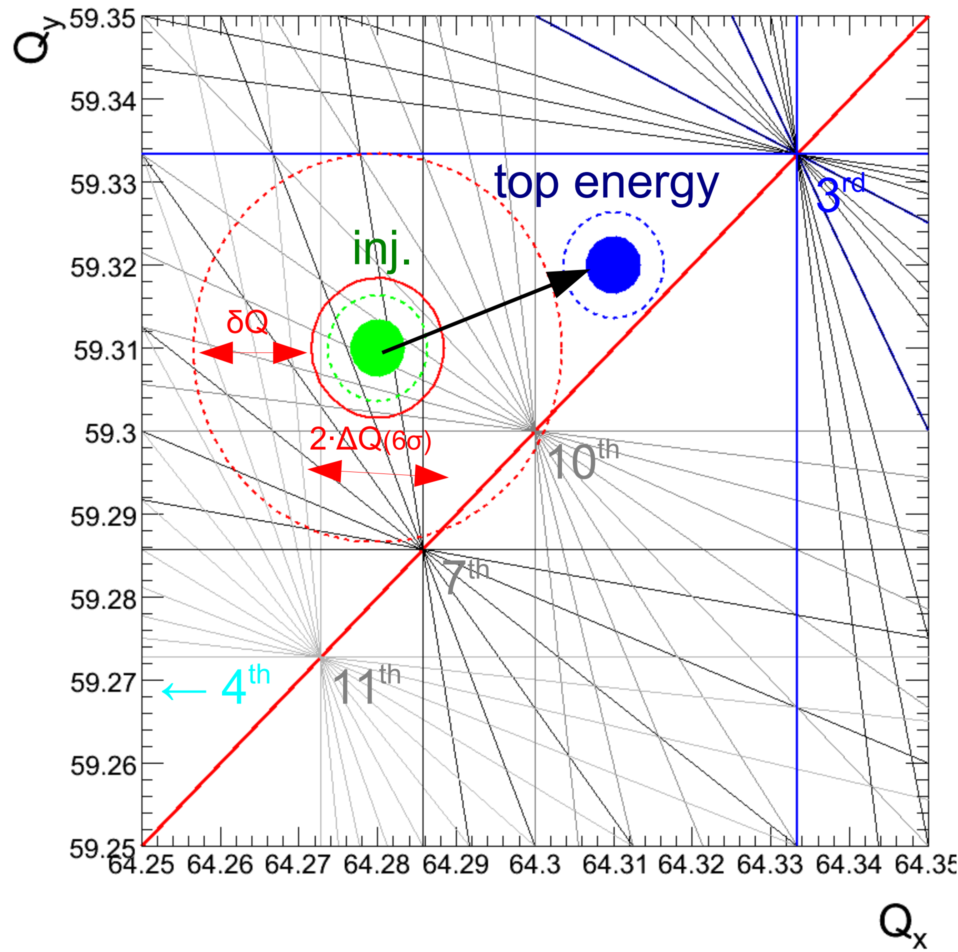
Beam-Based Feedbacks on Q, C^- & Q'

- Lepton machines: $\delta Q \sim 10^{-2} \dots 10^{-3}$
  - synchrotron-light damping
    → avoid up to ~3rd order resonance
- Hadron machines:
  - negligible synch. radiation damping
  - large tune footprints
  - avoid up to 12th order resonances
- Example LHC:
  - Tune spread (LHC) $\Delta Q|_{av} \approx 1.15\cdot10^{-2}$
    (fixed by available space in Q-diagram)
    → $\delta Q \leq 0.003 \dots 0.001$ (nominal)
  - Chromaticity (SPS: $\Delta p/p \approx 2.8\cdot10^{-4}$)
    - allowed max lin. chromaticity[14,15] (5-6 σ, 1st-order):

→ $Q'_{max} \approx (2 \rightarrow 10) \pm 1$ & $Q' > 0$

(expected drifts[1]: $\Delta Q' \approx 140$)

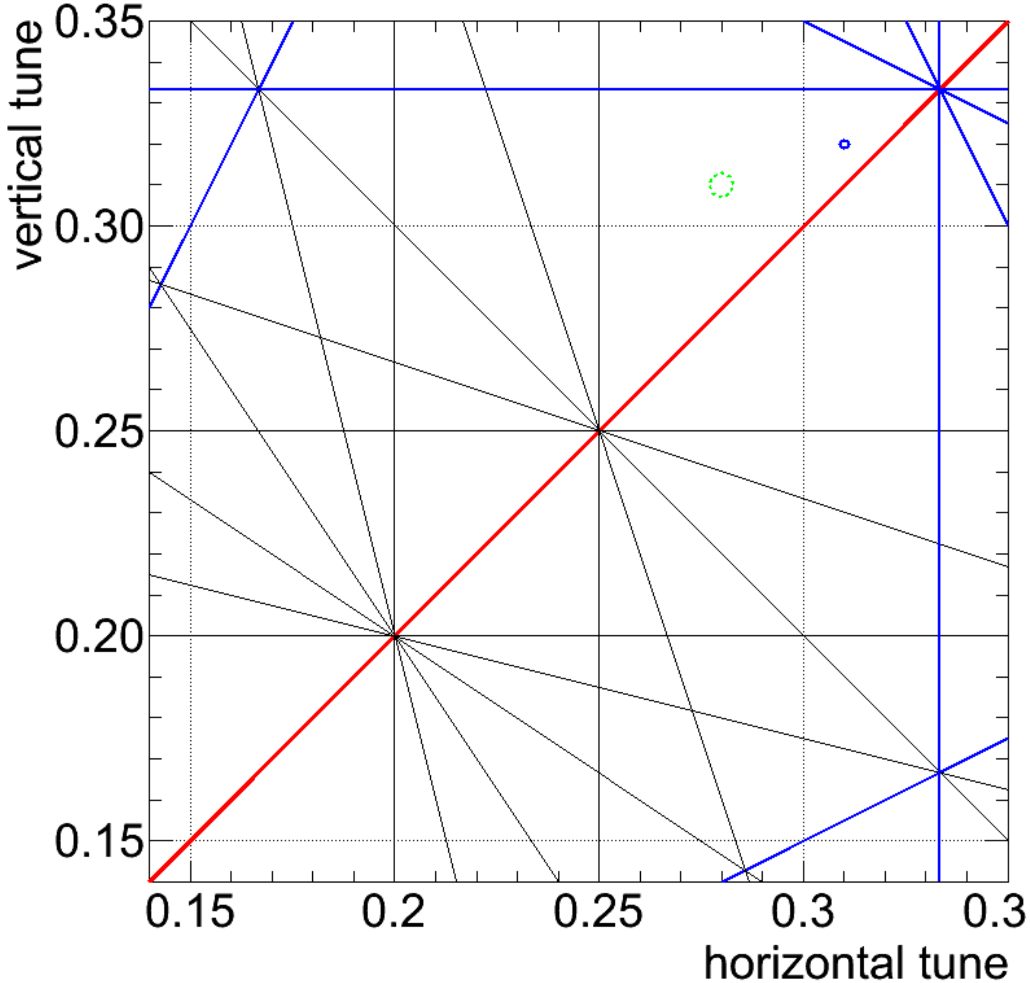
$$Q'_{max} \propto \frac{\Delta Q_{av}}{\Delta p/p}$$

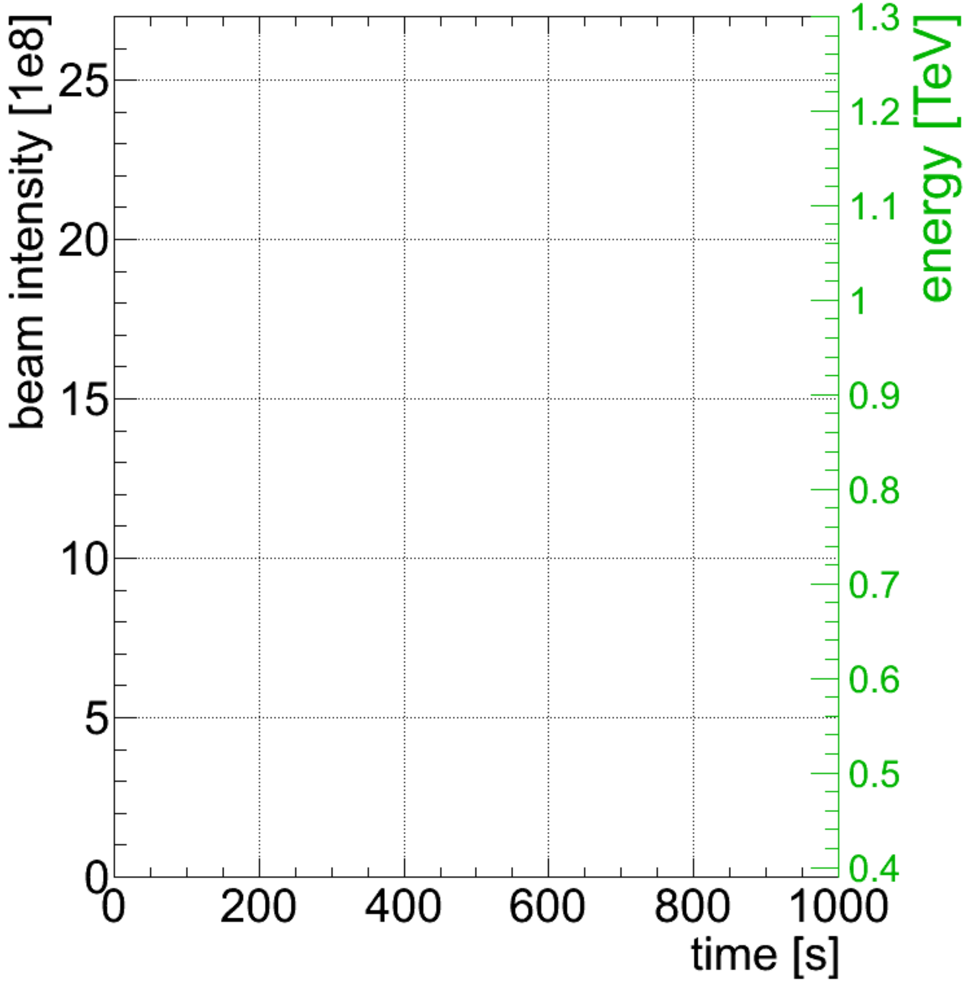


Original main focus:
chromaticity meas. & control
(very hard to measure reliably)

Higher-Order Beam Parameter Stability
Example: 2009 LHC Commissioning

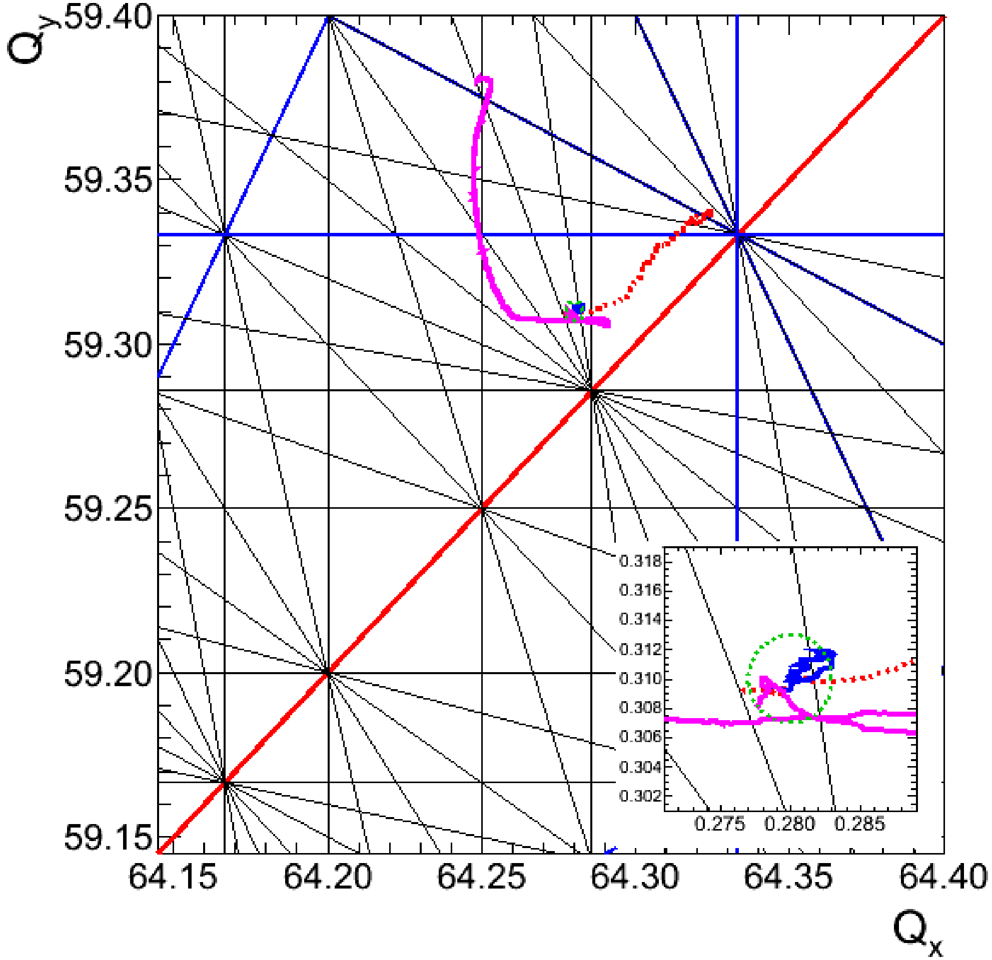
- … somewhat a surprise: 3rd ramp without Tune-Feedback

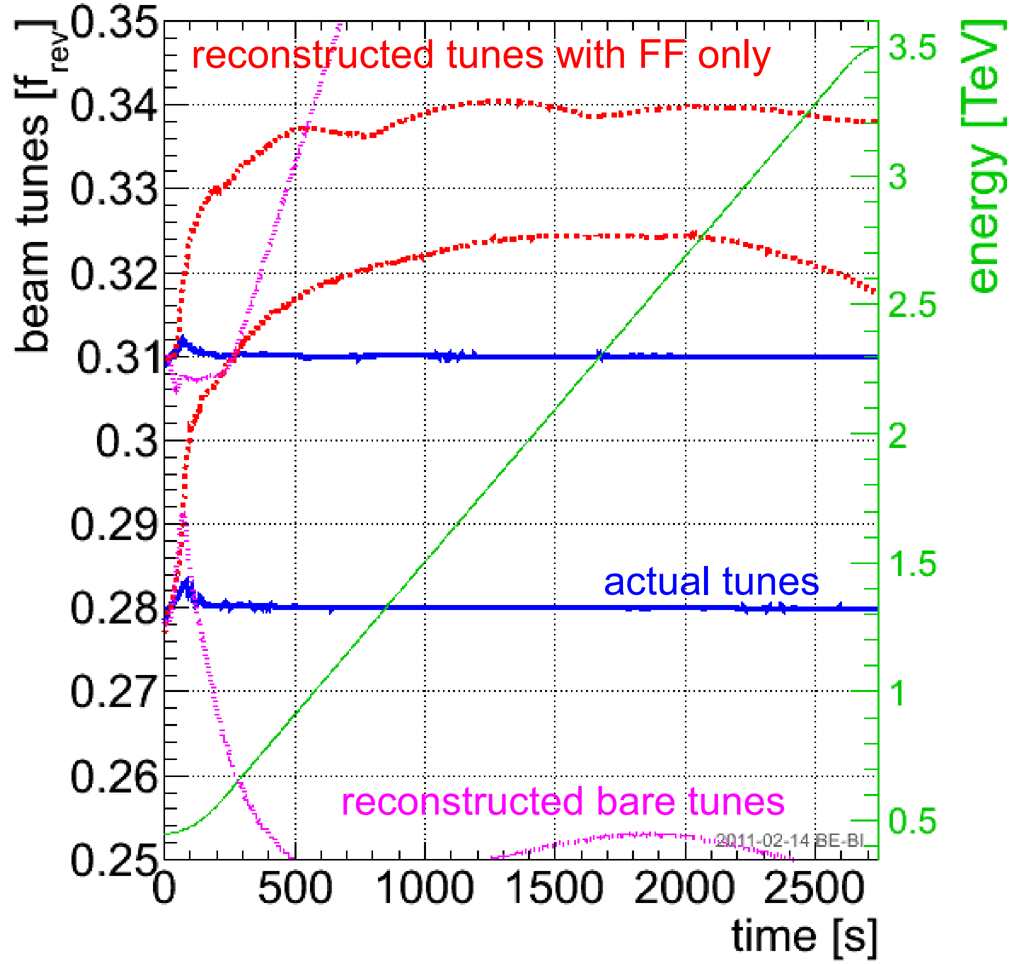




Higher-Order Beam Parameter Stability
Example: LHC Tune Feedback Operation

- Tune-FB driving and accelerating early commissioning in 2009-2011
  – Tunes kept stable to better than 10^{-3} for most part of the ramp and squeeze

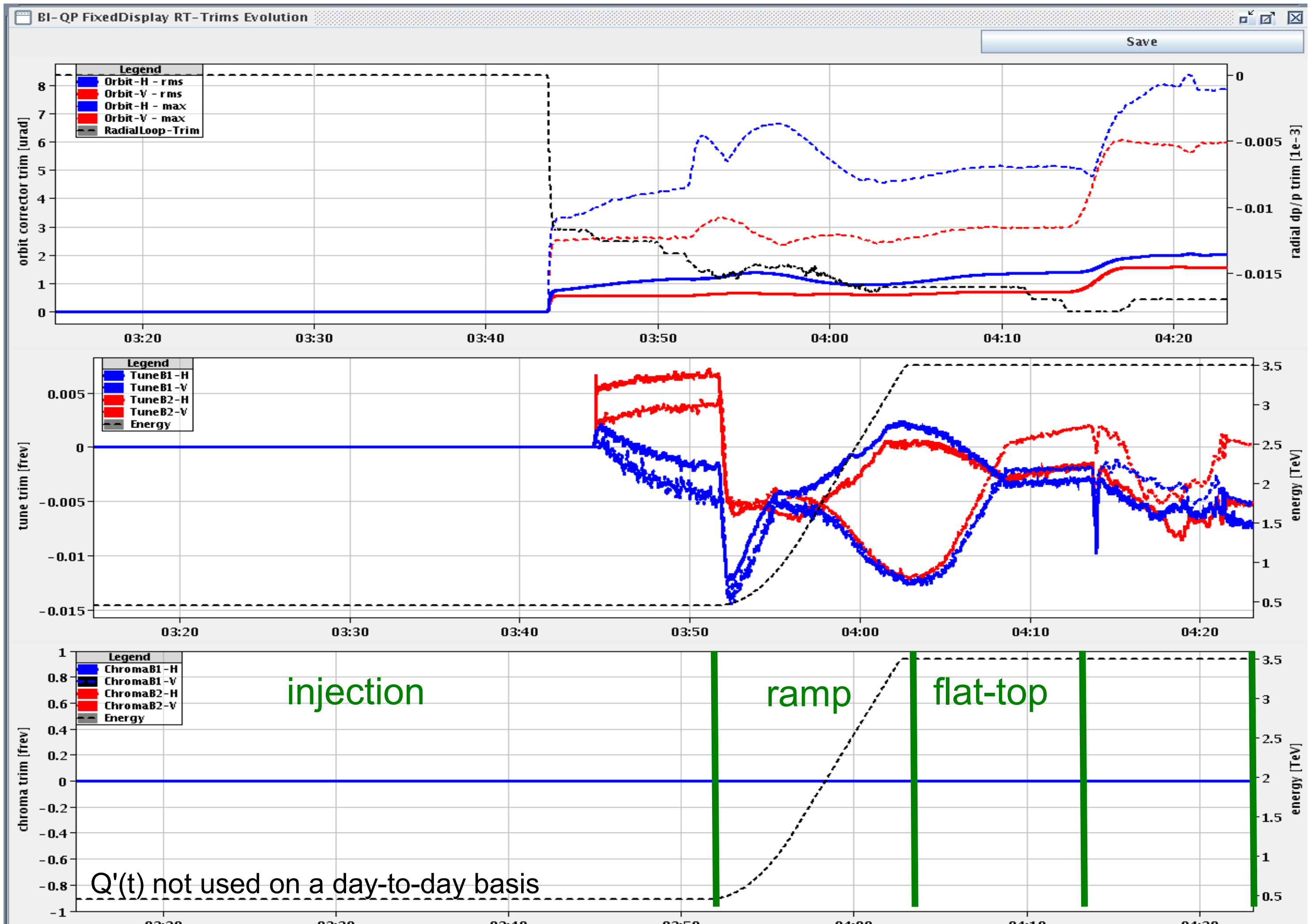




Higher-Order Beam Parameter Stability

Example: Feedback Integration and Operation at LHC

- Most accelerator facility: stability of actual observable became secondary
- Trims became de-facto standard to assess the FB and machine performance
- Common control-room question: “is … FB on” or “why is the FB off” (→ reliability/dependability)



Orbit-FB & Radial-Loop Trims (μrad)

Tune-FB trims

Q'(t)-FB trims

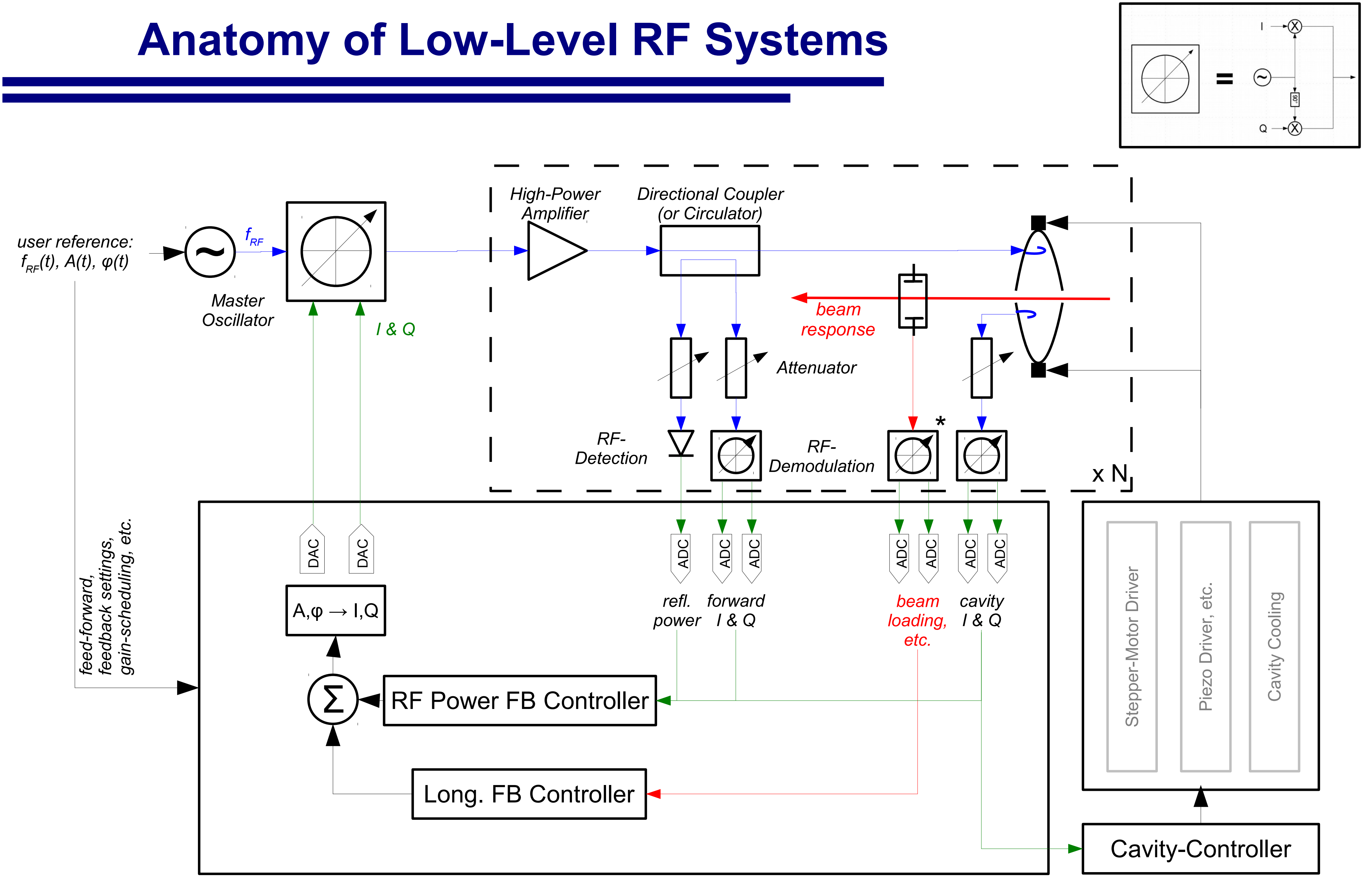
Energy (TeV)

β*-squeeze

Incomplete Feedback Overview Worldwide

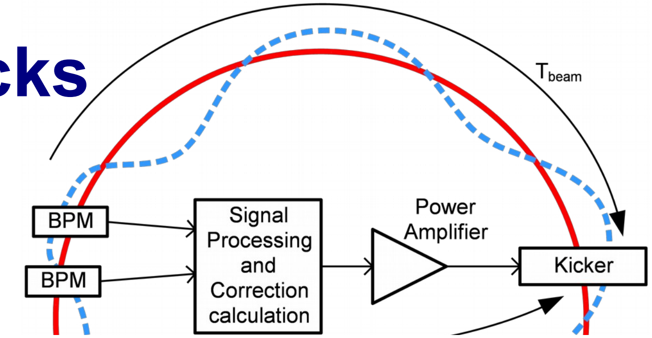
- Low-level hardware-focused systems – ubiquitous in nearly all accelerators (mostly SISO-like):
  - Magnet powering: converter voltage/current regulation → most w.r.t. quantity
  - Low-level Radio-Frequency (RF) control (f_{RF}, phase loop, radial loop*, synchronisation loop):
    - Low-level power – SISO-like but often non-linear (RF source working point dependent)
      - RF source: amplitude, frequency (synchro-loop), phase, compensation of drifts & noise
      - Cavity tuning: resonance frequency, quality factor (Q), …
    - Longitudinal RF feedbacks (bunch/batch arrival w.r.t. cavity):
      - one-turn-feedbacks/phase-loop: longitudinal shunt impedance, beam loading compensation (limited RF power & power drainage by beam)
- Fast transverse feedbacks (turn-by-turn → bunch-by-bunch → intra-bunch time-scales)
  - damping of injection oscillations, improving single/coupled bunch (in-)stability thresholds
- Beam-based feedback systems – higher complexity (usually MIMO, indirect parameter observation)
  - Light Sources: mostly orbit and energy feedback (radial steering) only
  - Lepton Collider: LEP[4], PEP-II[5], KEK-B – orbit and tune feedback (mostly during ramp)
  - Hadron Collider: Hera, LHC, RHIC, Tevatron – mostly slow orbit feedback, except:
    - Hera: Orbit, Tune
    - RHIC: Orbit, Tune[6]/Coupling, Chromaticity[7]
    - LHC: Orbit/Energy, Tune/Coupling, Chromaticity, …
  - Special case – pulsed accelerators: linacs, fast cycling circular machines (CERN, FAIR, GSI, ESS, PSI, SNS …)
    - pulse-by-pulse or cycle-to-cycle feedbacks

Anatomy of Low-Level RF Systems

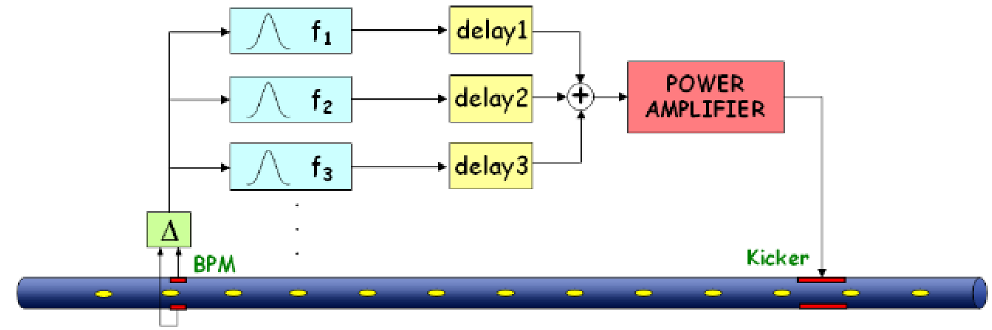


*not necessarily frequency- (demodulation) but also in time-domain (long. bunch-by-bunch feedbacks, synchro-loop, radial loop, ...

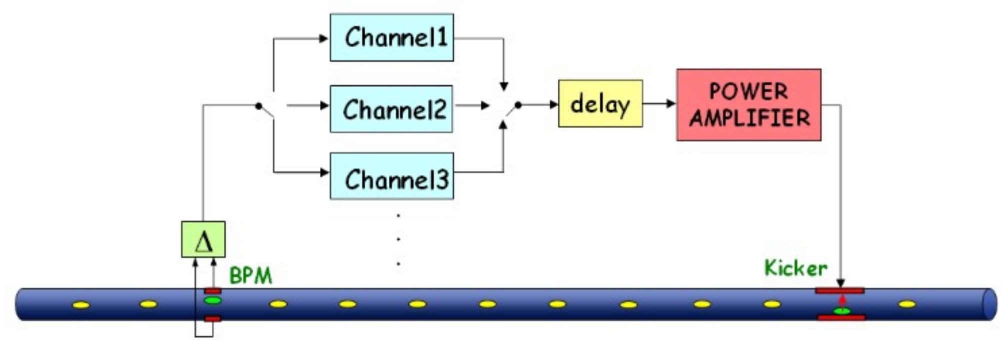
Anatomy of Fast Transverse Feedbacks



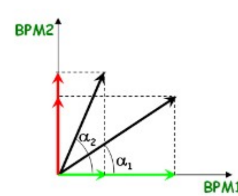
- "Simple" from a FB design point of view:
  - monitor → Hilbert-filter (0→ 180° phase adj.) → K_p(-only) control → actuator (RF kicker)

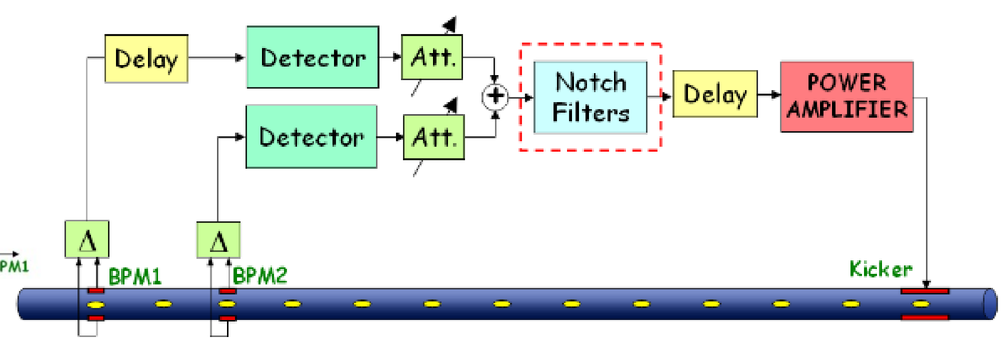


- Historic evolution of bandwidths:
  - 'turn-by-turn'
  - 'mode-by-mode' (frequency-domain)
  - 'bunch-by-bunch' (gated, time-domain)
  - hybrid: 'vector-sum' (Hilbert-filter lag reduction)
  - 'intra-bunch' (J-PARC, GSI/FAIR, CERN)
  - stochastic cooling



- primary challenges are by far w.r.t. technology used for the implementation
  - RF MHz → GHz bandwidths
    - pick-ups, RF kickers, processing
  - went full-circle from fully-analog
    → fully-digital
    → (hybrid-)analog designs



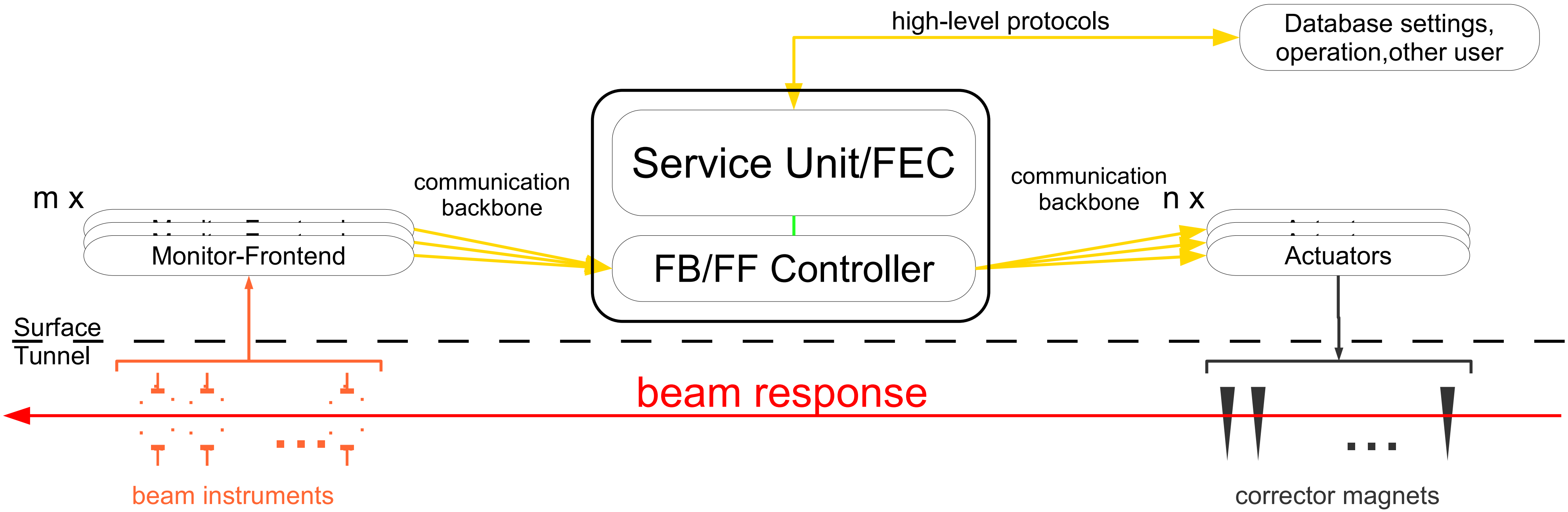


Courtesy W. Höfle, CERN
& Marco Lonza, Elettra

Anatomy of Beam-Based Feedbacks

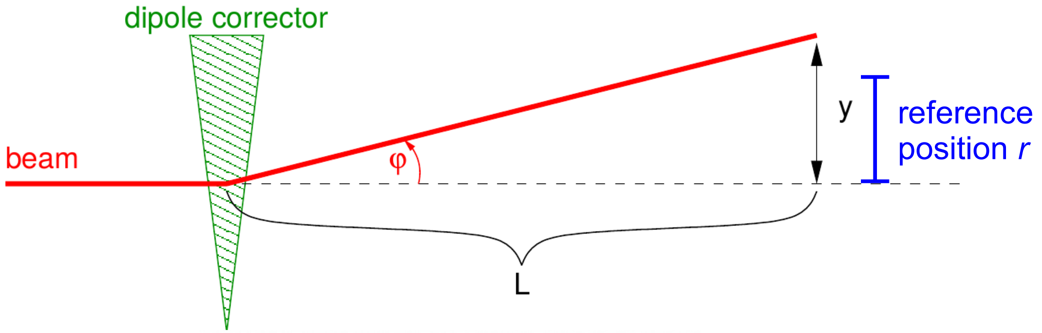
Common Control Layout & Implementation

- fairly generic, typically MIMO & often split into two sub-systems
  - Feedback Controller: actual feed-forward/feedback controller logic
    - specific implementation depends on the bandwidth requirement
  - Service Unit/FEC: Interface to control system/OP/the world
    - dominated by industrial PCs, less specific DSPs → PCaPAC!
- Overall strength depends on the knowledge/reliability of the weakest link in the chain
  - Sensor and timing/communication often overlooked



Feedback Basics I
Feed-Forward – static

- Simple example: beam steering in transfer-line
  - N.B. ideal world: perfect dipole transfer-function/magnet calibration



$$y = G(s) \cdot \varphi \approx L \cdot \varphi$$

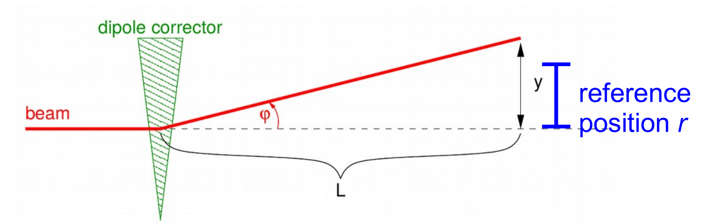
- 1st order control problem description:

"find control law that steers beam to position r"

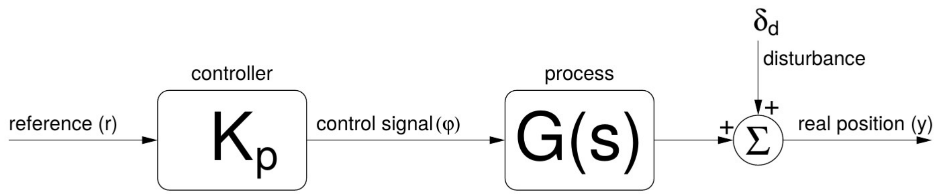
$$y = K_p G(s) \cdot r \stackrel{!}{=} r$$

  - still trivial solution: $$\boxed{\varphi \cdot r \rightarrow K_p \cdot r \quad \text{and} \quad K_p = \frac{1}{L}}$$

Feedback Basics II
Feed-Forward – disturbance rejection



- Formalise example:



- Some definition: open-loop (aka. 'feed-forward')

Transfer function ('response'): $T(s) := \frac{y}{r} = \underbrace{D(s)}_{K_p} \cdot G(s)$ → this example: equals '1'

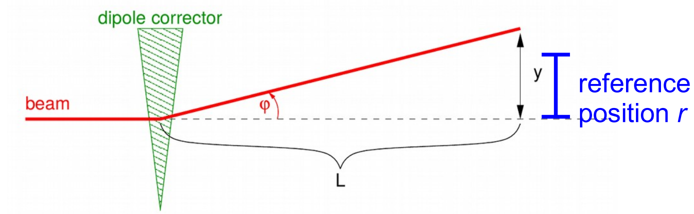
(or 'nominal complementary sensitivity')

Disturbance rejection: $S(s) := \frac{y}{\delta_d} = 1$

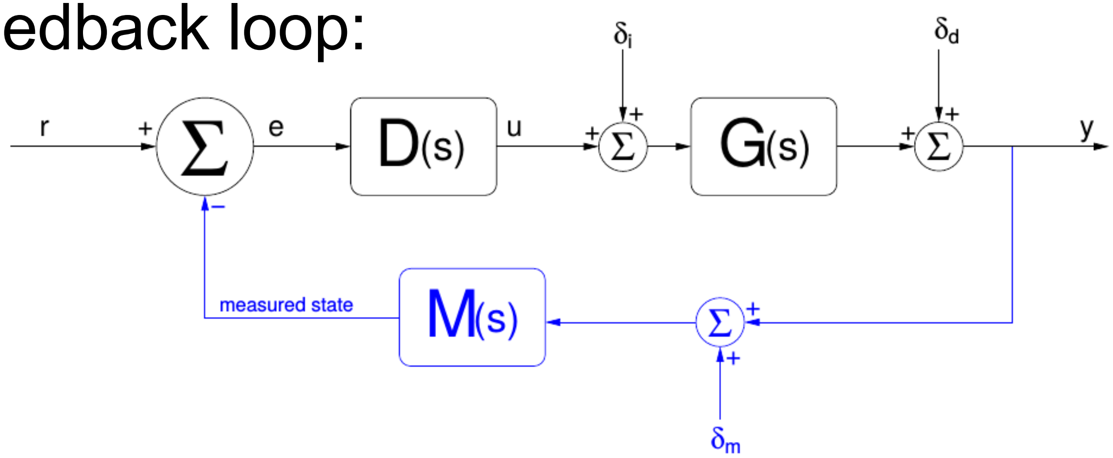
(or '*first sensitivity function*')

poor disturbance rejection ↔ basically for '$\delta_d > r$' position is determined external perturbations

Feedback Basics III
Feed-Back



- Simple feedback loop:



example:

$$\boxed{\begin{array}{l} D(s)=K_p=\frac{1}{L} \\ D(s)G(s)=1 \end{array}}$$

transfer function: $$T_0(s) := \frac{y}{r} = \frac{D(s)G(s)}{1+D(s)G(s)}$$ → $T_0(s) = 0.5$ – what?? *'steady-state error'*

nominal sensitivity: $$S_{d0}(s) := \frac{y}{\delta_d} = \frac{1}{1+D(s)G(s)}$$ → this example: 0.5

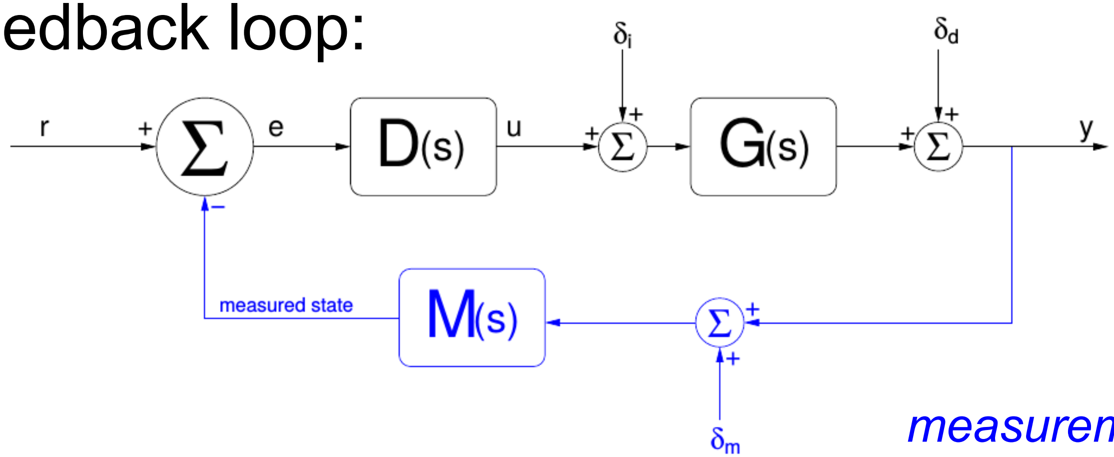
to note:

$$\boxed{T_0(s)+S_d(s)=1}$$

- However: $S_{d0}(s) = ½$ – good
  - can be further improved by increasing '$K_p \to \infty$' decrease '$S_{d0} \to 0$'
  - also improves but does not fully remove the steady state error
  - Our little example: everything has been invariant in time
    → real world systems are time-dependent → what we need is some integrator action

Feedback Basics IV
Feed-Back

- Simple feedback loop:



transfer function: $T_0(s) := \frac{y}{r} = \frac{D(s)G(s)}{1 + D(s)G(s)}$

disturbance rejection: $S_{d0}(s) := \frac{y}{\delta_d} = \frac{1}{1 + D(s)G(s)}$

input sensitivity: $S_{i0}(s) := \frac{y}{\delta_i} = \frac{G(s)}{1 + D(s)G(s)}$

control sensitivity: $S_{u0}(s) := \frac{u}{\delta_d} = \frac{D(s)}{1 + D(s)G(s)}$

measurement noise sensitivity:

$$\boxed{S_{n0}(s) := \frac{y}{\delta_n} = \frac{D(s)G(s)}{1 + D(s)G(s)}}$$

to note: Johnson noise ~ $\sqrt{\Delta f_{BW}}$

- Important take-away:

a) improving closed-loop (gain-) bandwidth $T_0(s)$ also improves disturbance rejection $S_d(s)$

b) improving closed-loop (gain-) bandwidth $T_0(s)$ also increases sensitivity to measurement noise

requires trade-off

Laplace Transform I/II
Definition

- … avoid "convoluted" math → use Laplace Transform:

$$\mathcal{L}(f(t)) = F(s) := \int_{0^-}^{+\infty} f(t)e^{-st}dt$$

most important features:

Linearity:	$\mathcal{L}(\alpha f_1(t)) = \alpha\mathcal{L}(f_1(t)) = \alpha F_1(s)$
Superposition:	$\mathcal{L}(\alpha f_1(t) + \beta f_2(t)) = \alpha F_1(s) + \beta F_2(s)$
Time Delay:	$\mathcal{L}(f(t' \to t - \lambda)) = e^{-s\lambda}F(s)$
Time Scaling:	$\mathcal{L}(f(t' \to at)) = \frac{1}{\lvert a\rvert}F(\frac{s}{a})$
Differentiation:	$\mathcal{L}(\dot{f}(t)) = -f(0^-) + sF(s)$
Integration:	$\mathcal{L}\left(\int f(t)dt\right) = \frac{1}{s}F(s)$
Convolution:	$\mathcal{L}(f_1(t) * f_2(t)) = F_1(s) \cdot F_2(s)$ and
	$\mathcal{L}(f_1(t) \cdot f_2(t)) = F_1(s) * F_2(s)$

- To note: similarity of Laplace and Fourier transform (s → '$i\omega$', and integration interval $]0^-,+\infty[$ → $]-\infty,+\infty[$)

Laplace Transform II/II
Example – Simple Harmonic Oscillator

- From differential equations:

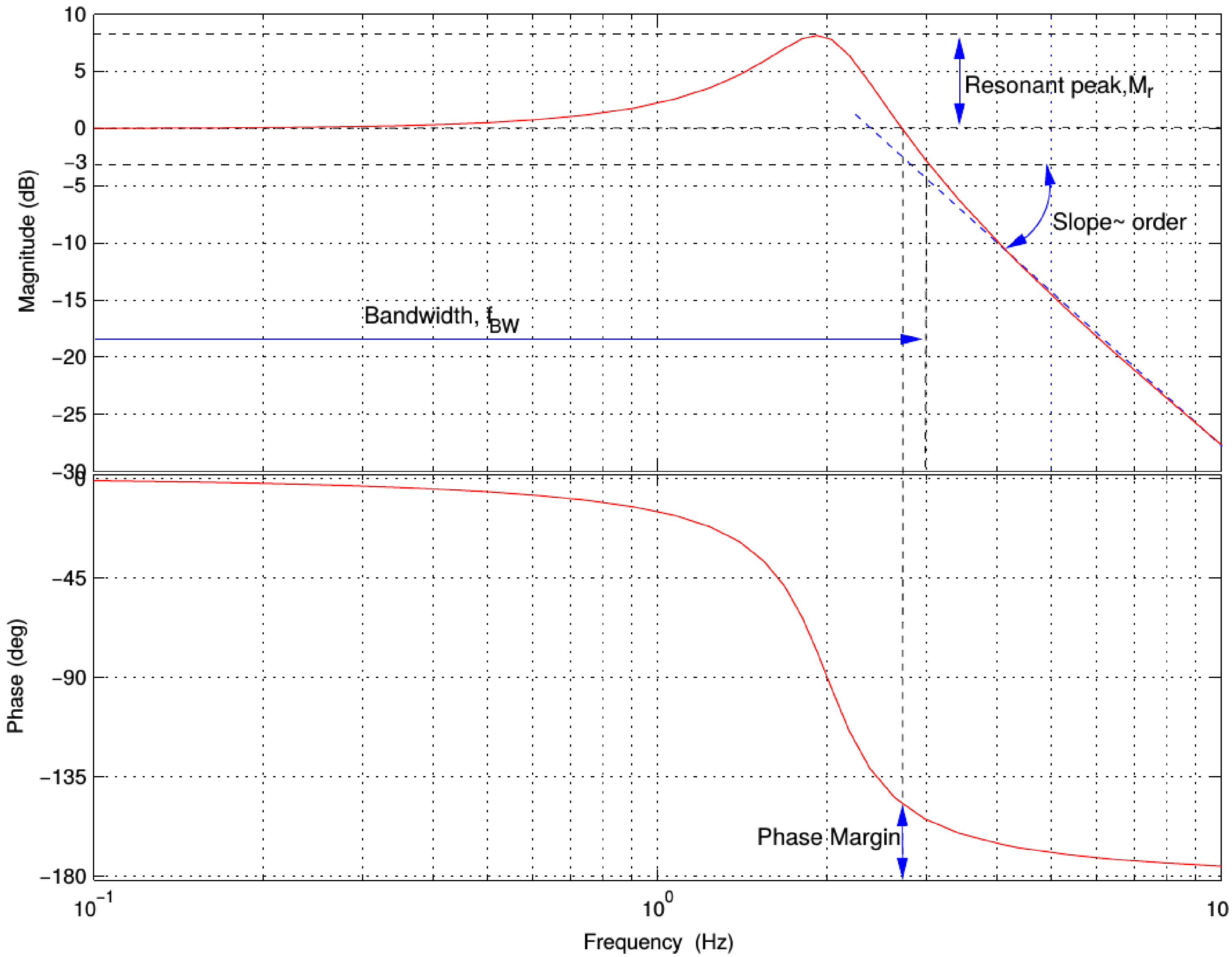
$$\begin{aligned}
\mathcal{L}\{\ddot{y}(t) + 2\zeta\omega_0 \cdot \dot{y}(t) + \omega_0^2 y(t)\} &= \mathcal{L}\{f(t)\} \\
s^2 Y(s) + 2\zeta\omega_0 s Y(s) + \omega_0^2 Y(s) &= F(s) \\
(s^2 + 2\zeta\omega_0 s + \omega_0^2) \cdot Y(s) &= F(s)
\end{aligned}$$

$$Y(s) = H(s) \cdot F(s) \quad \Longleftrightarrow \quad H(s) := \frac{Y(s)}{F(s)}$$

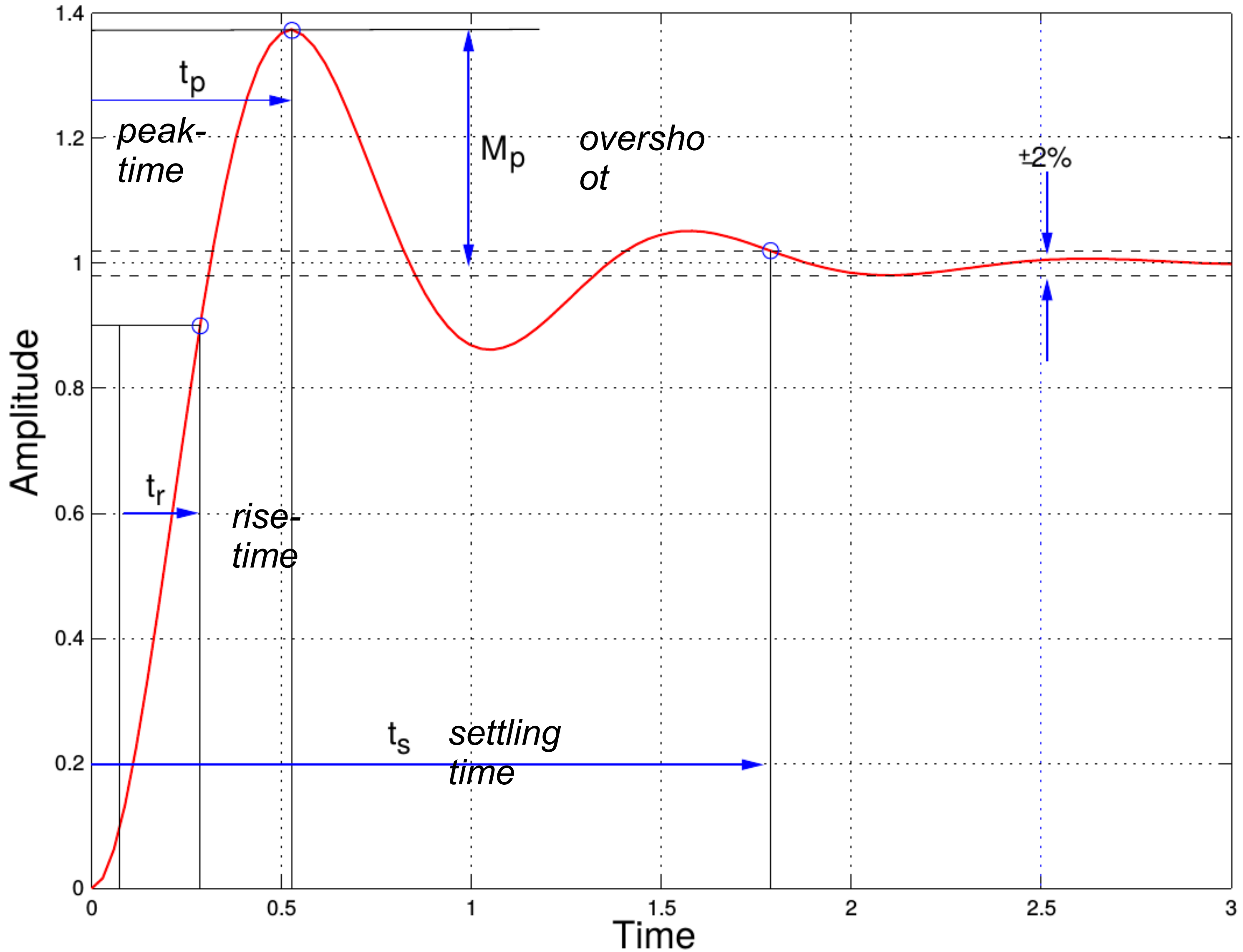
- To transfer function:

$$\begin{aligned}
H_{1^{st}}(s) &= \frac{K_0}{\tau \cdot s + 1} \\
H_{2^{nd}}(s) &= \frac{K_0 \omega_0^2}{s^2 + 2\zeta\omega_0 \cdot s + \omega_0^2}
\end{aligned}$$

Frequency Domain – Bode-type Plot
Direct Transfer Function $T_0(s)$



Time-Domain
reference step Response



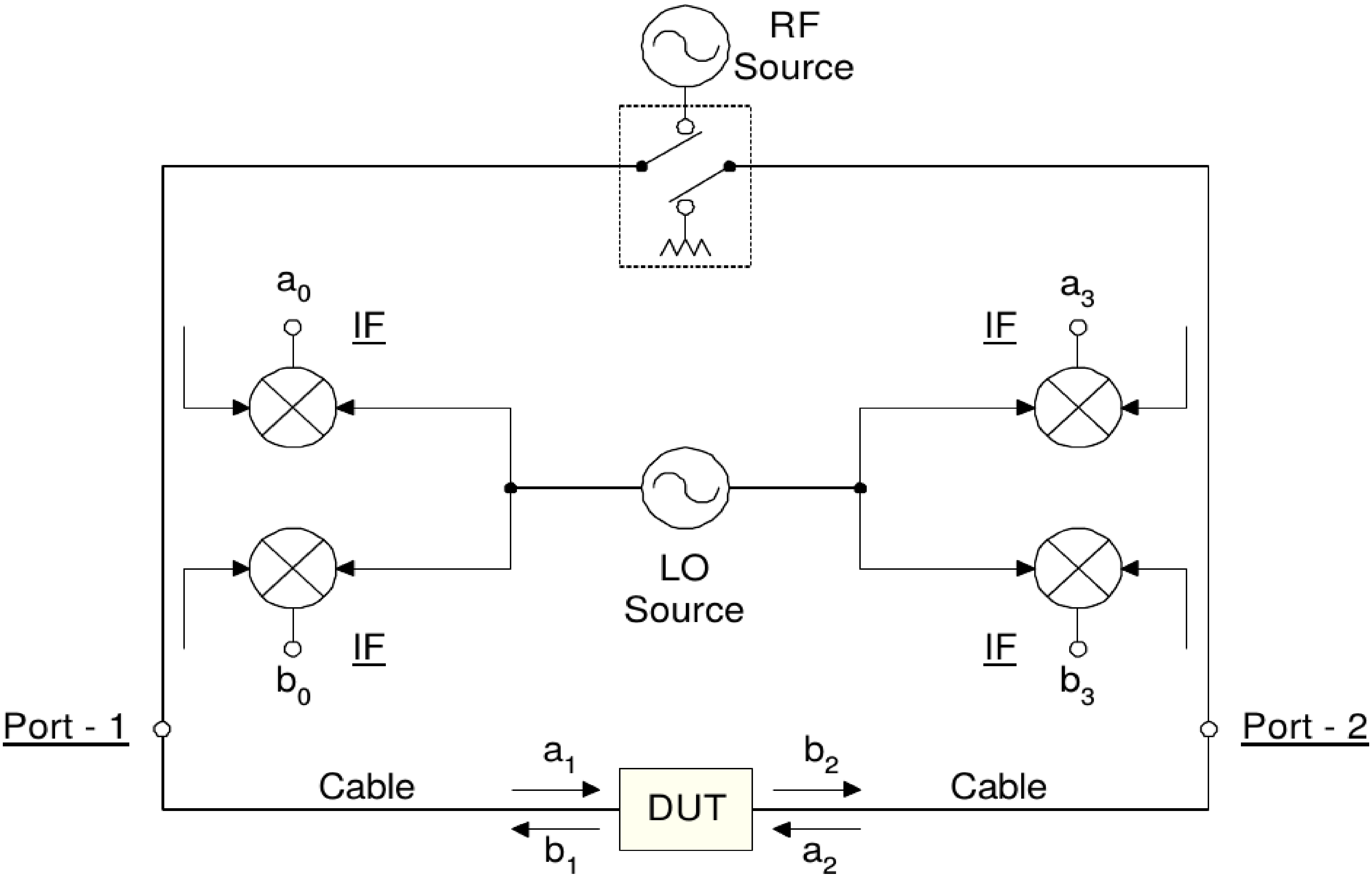
- Importance: closed-loop responses are often expressed with the same metrics

- Coaxial measurement line
  - old fashion method – no more in use but good for understanding of VSWR concept

- Network analyzer
  - Excites a network (circuit, antenna, amplifier or similar) at a given CW frequency and measures response in magnitude and phase → determines S-parameters
  - Covers a frequency range by measuring step-by-step at subsequent frequency points
  - Application: characterization of passive and active components, time domain reflectometry by Fourier transforming reflection response, etc.

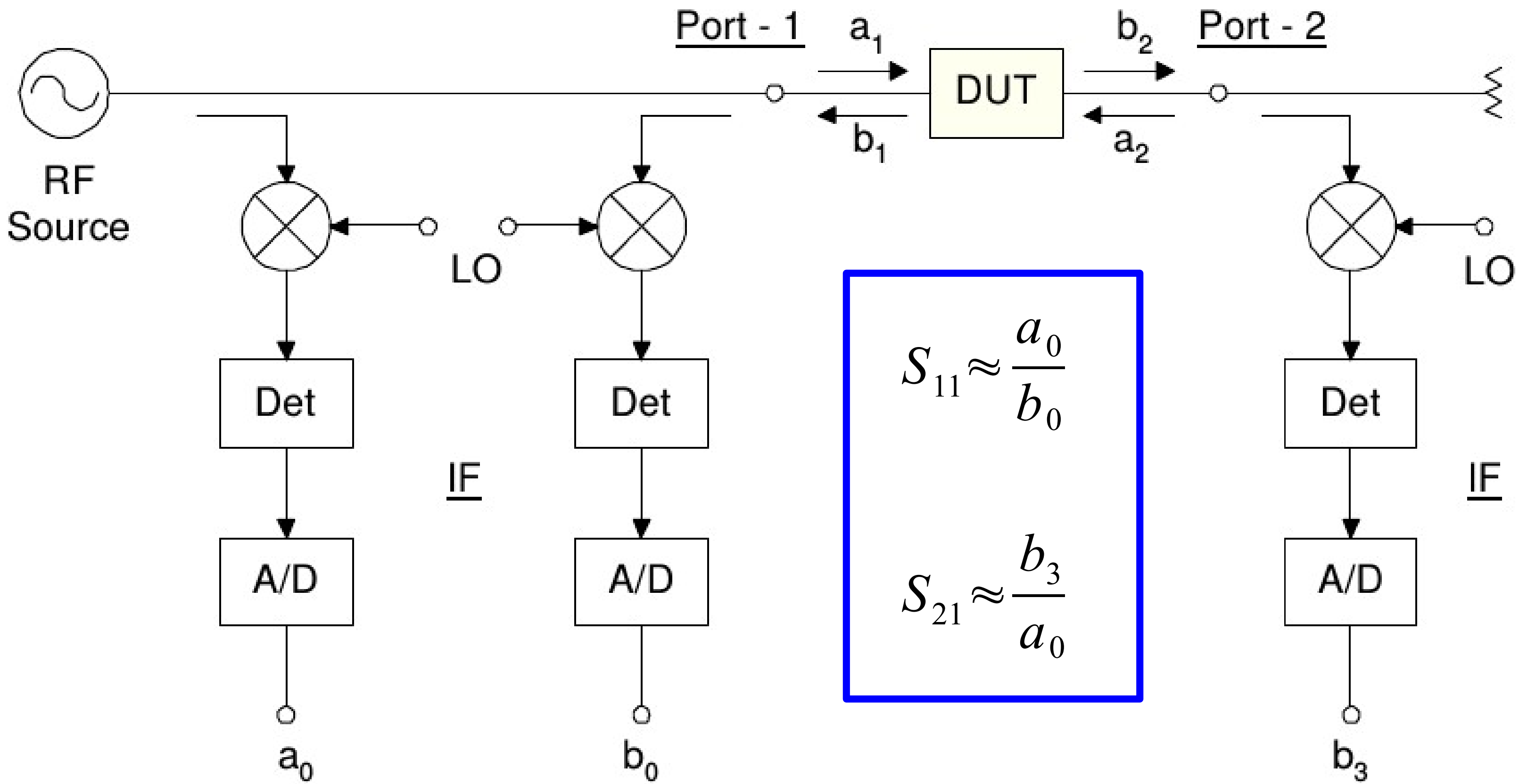
Calibration kit: – handle with great care!!
They are more worth than their weight in gold!

System Identification – RF Domain II/III
Vector-Network-Analyser Schematic



System Identification – RF Domain III/III
Vector-Network-Analyser Schematic

- Forward-direction only:



$$S_{11} \approx \frac{a_0}{b_0}$$

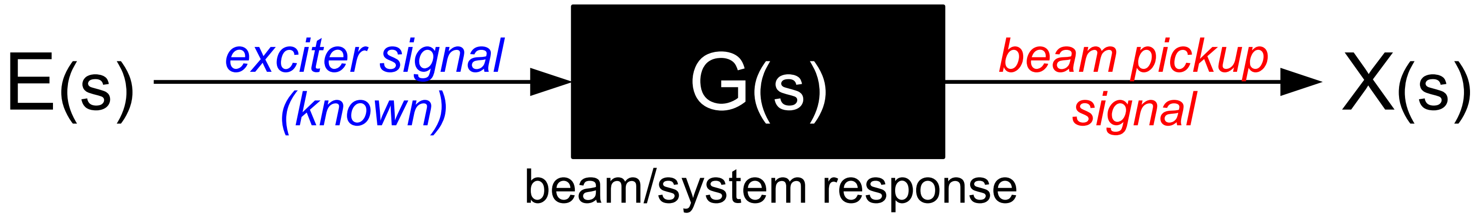
$$S_{21} \approx \frac{b_3}{a_0}$$

- VNAs are based on <u>relative</u> power level measurements
  → needs calibration to equalise $a_0=a_1$, $b_0=b_1$ and $b_3=b_2$
  → the importance of calibration standards

System Identification I/II

e.g. Tune Diagnostics Principle (↔ fast transverse feedbacks)

Control Theory → System Identification

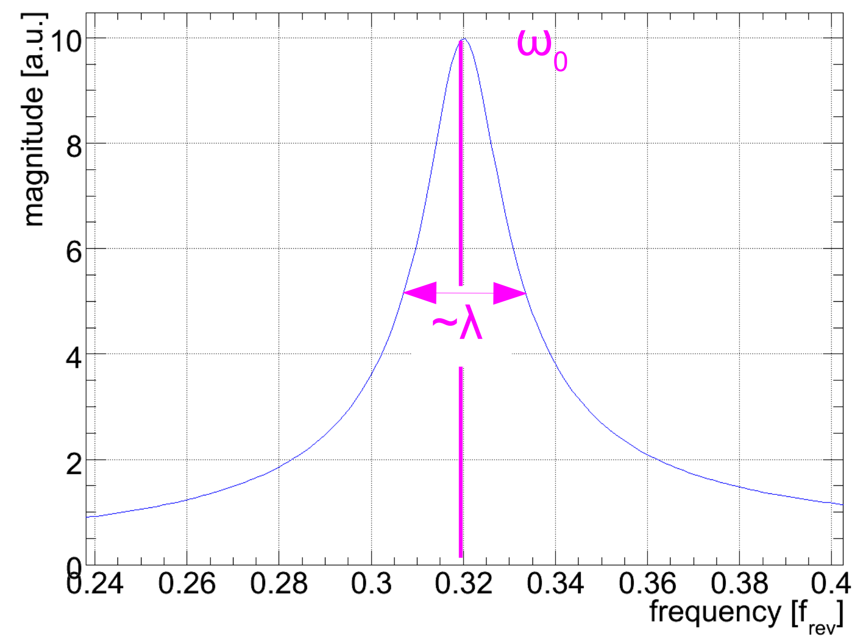


Example (first order) beam response ≈ damped harmonic oscillator resonance

(ω_0: resonant frequency (Q), λ: tune resonance width (σ_Q),

ω: driving frequency)

$$\left|G(\omega)\right| := \left|\frac{X(s)}{E(s)}\right| \approx \frac{\omega_0}{\sqrt{\left(\omega^2-\omega_0^2\right)^2+\left(2\lambda\omega_0\omega\right)^2}}$$



Excitation choices:

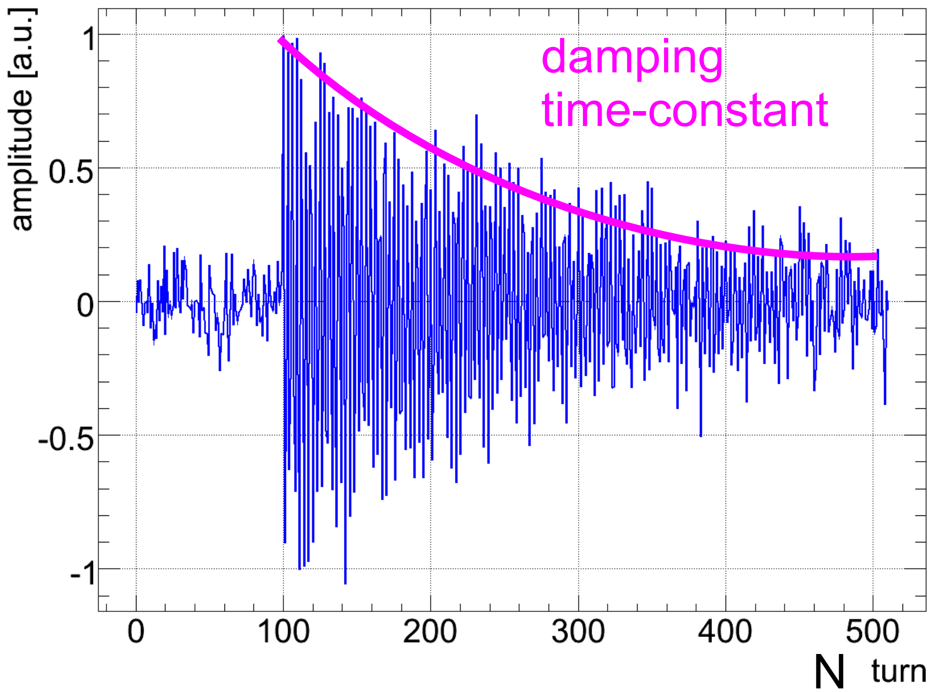
- White or remnant noise
  - no information on signal phase
- Single-turn transverse kick (classic step-respons
- Frequency Sweep aka. 'Chirp'
  - focuses excitation power on frequency range of interest → less ε-blow-up, constant power
- Phase-Locked-Loop Systems = resonant excitation on the Tune (↔ Vector-Network-Analyser principle)

Note: Exciter and pick-up have additional non-beam related responses!

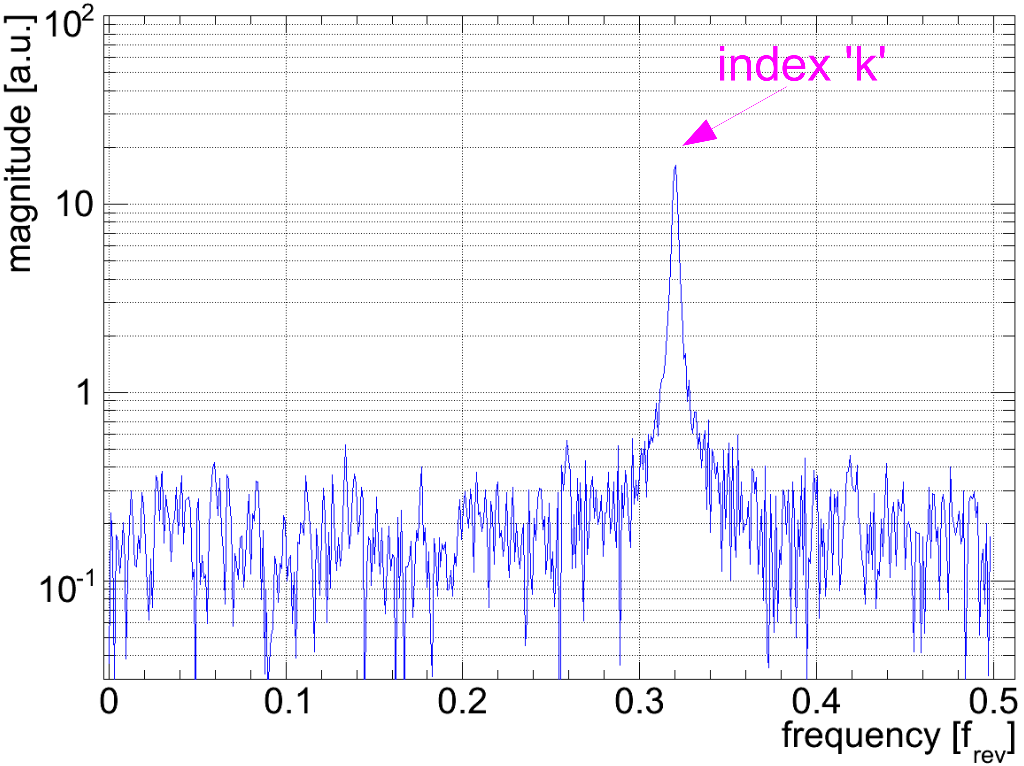
System Identification II/II

e.g. Tune Diagnostics Principle – step response (↔ fast transverse feedbacks)

.... how an kick-induced beam oscillation usually looks like (no sync. beating)



FFT



Fourier analysis of turn-by-turn data:

magnitude peaks at q_{frac}

$$q_{frac} \approx \frac{k}{N}$$

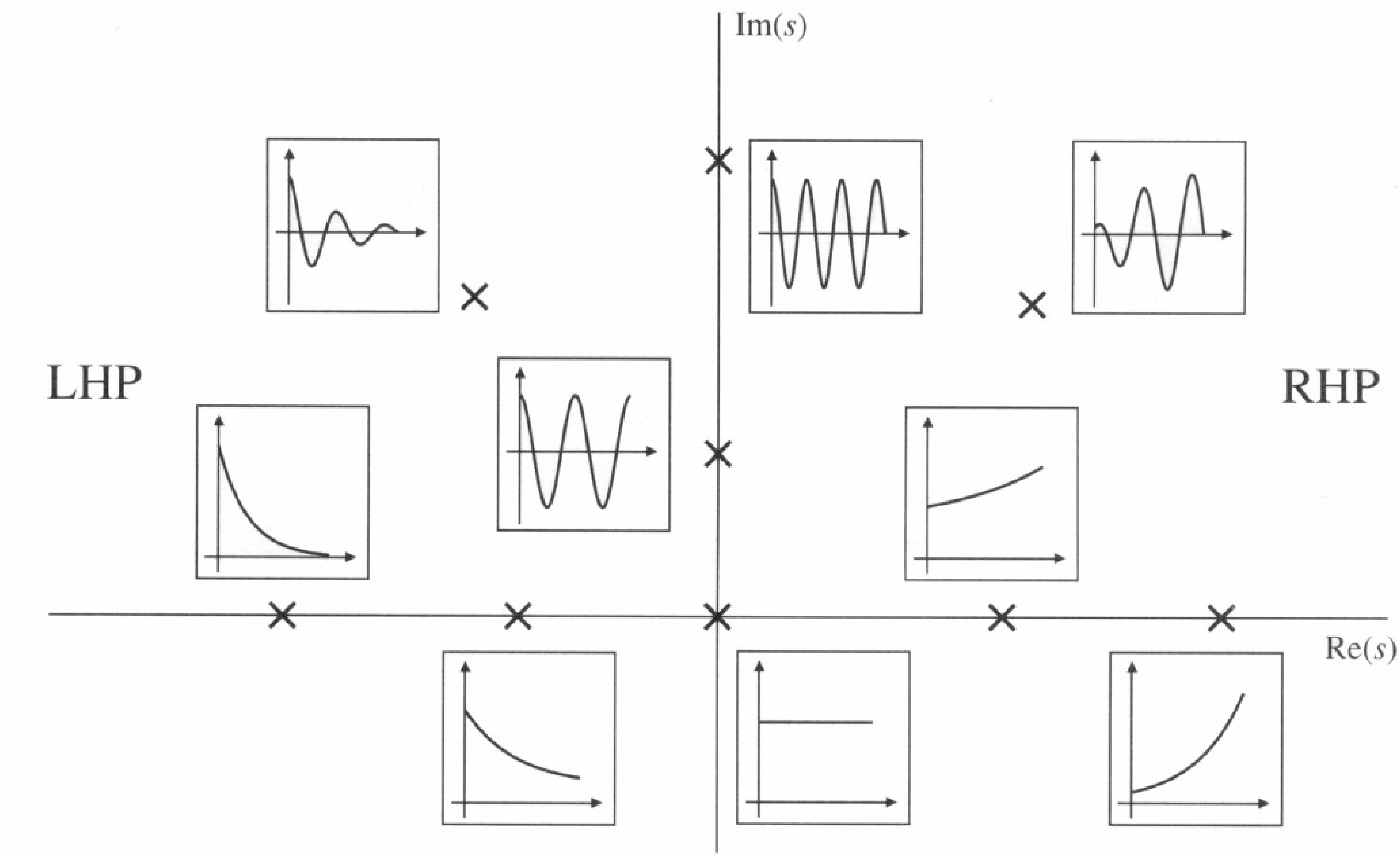
N.B. no information on Q_{int} !

Can improve resolution by fitting central bin width

Stability – Poles & Zeros

$$H(s) := \frac{\prod_{i=0}^{n}(s+b_i)}{\prod_{i=0}^{m}(s+a_i)}$$

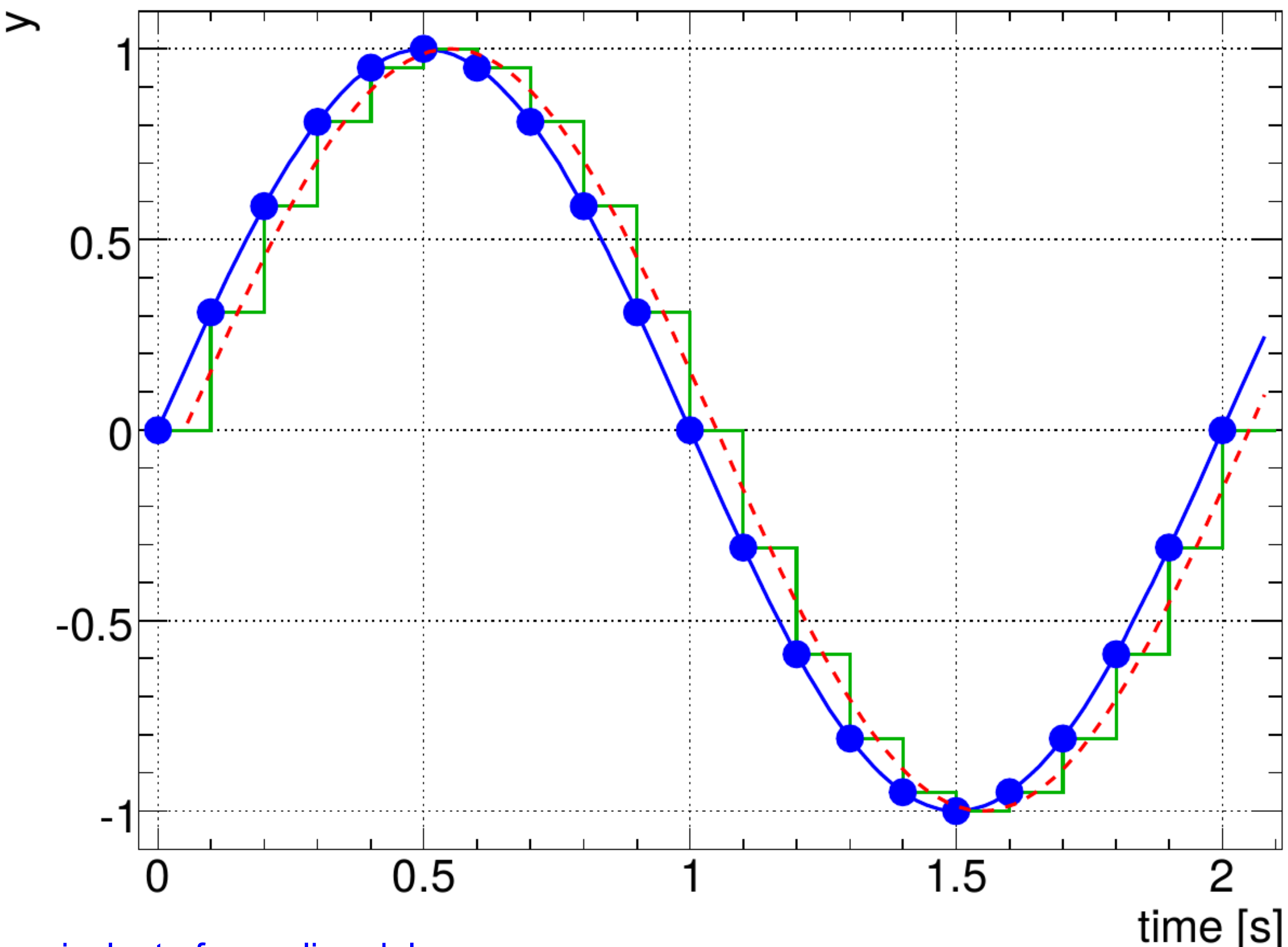
- 'Poles' ↔ '$(s-a_i)=0$' & 'Zeros' ↔ '$(s-b_i)=0$'



Minimum-phase:
invertible, causal and stable transfer function
↔ all poles & zeros on left-half-plane

Nice S-domain simulation tool:
http://web.mit.edu/6.302/www/pz/

- Discrete representation of analog signals:

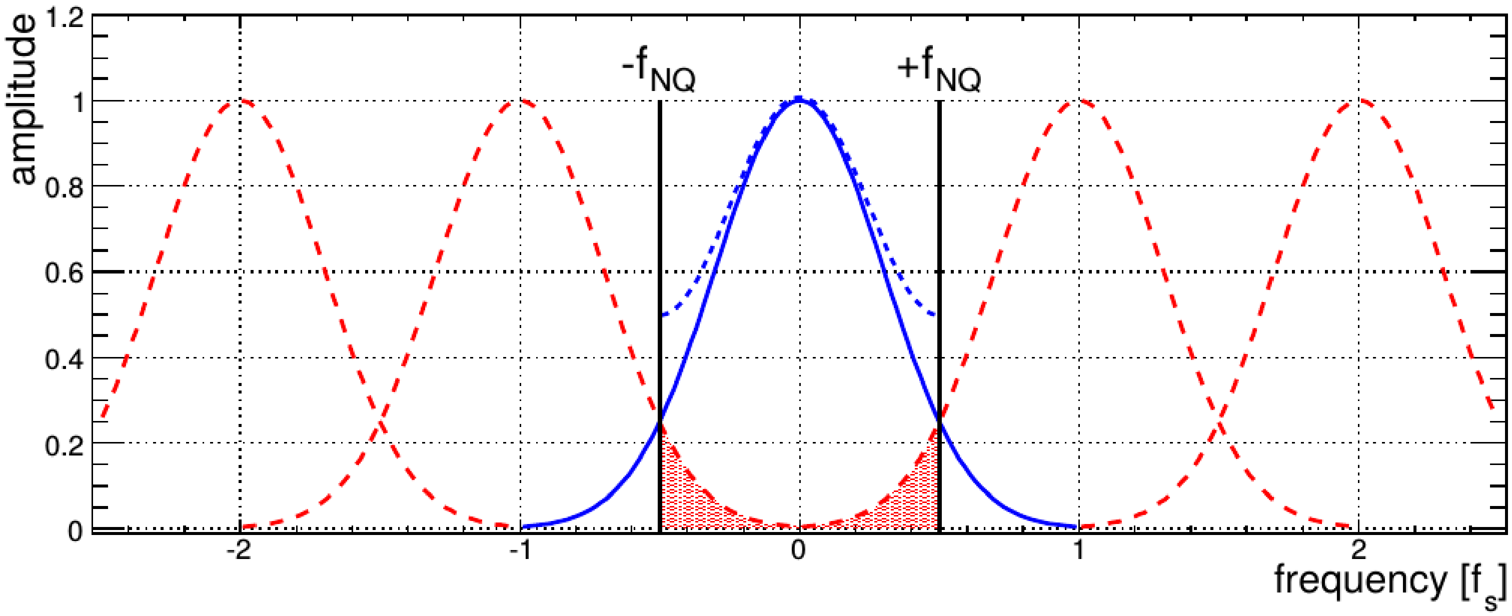


analog equivalent of sampling delay

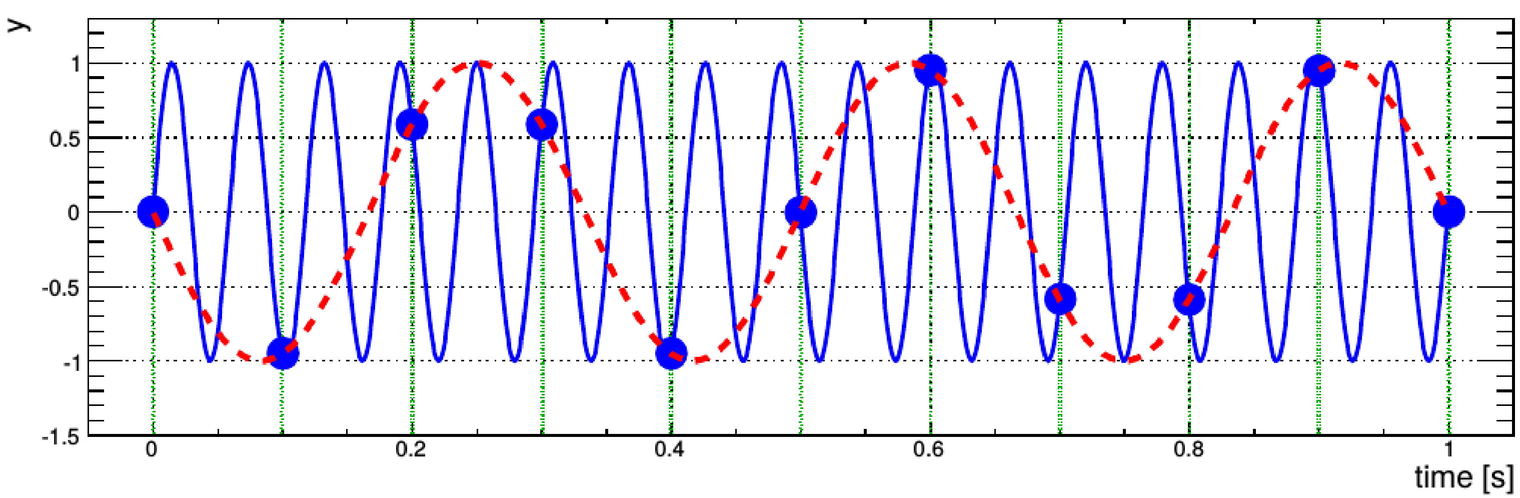
$$\lambda = \frac{T_s}{2} \qquad f(s) = e^{-\lambda s} f_0(s)$$

Digital Control & Sampling II/II
Aliasing

- Watch out for aliasing and Shannon-Nyquist criteria:



- Sometimes useful in RF for under-sampling, if not → real analog low-pass prior to ADC

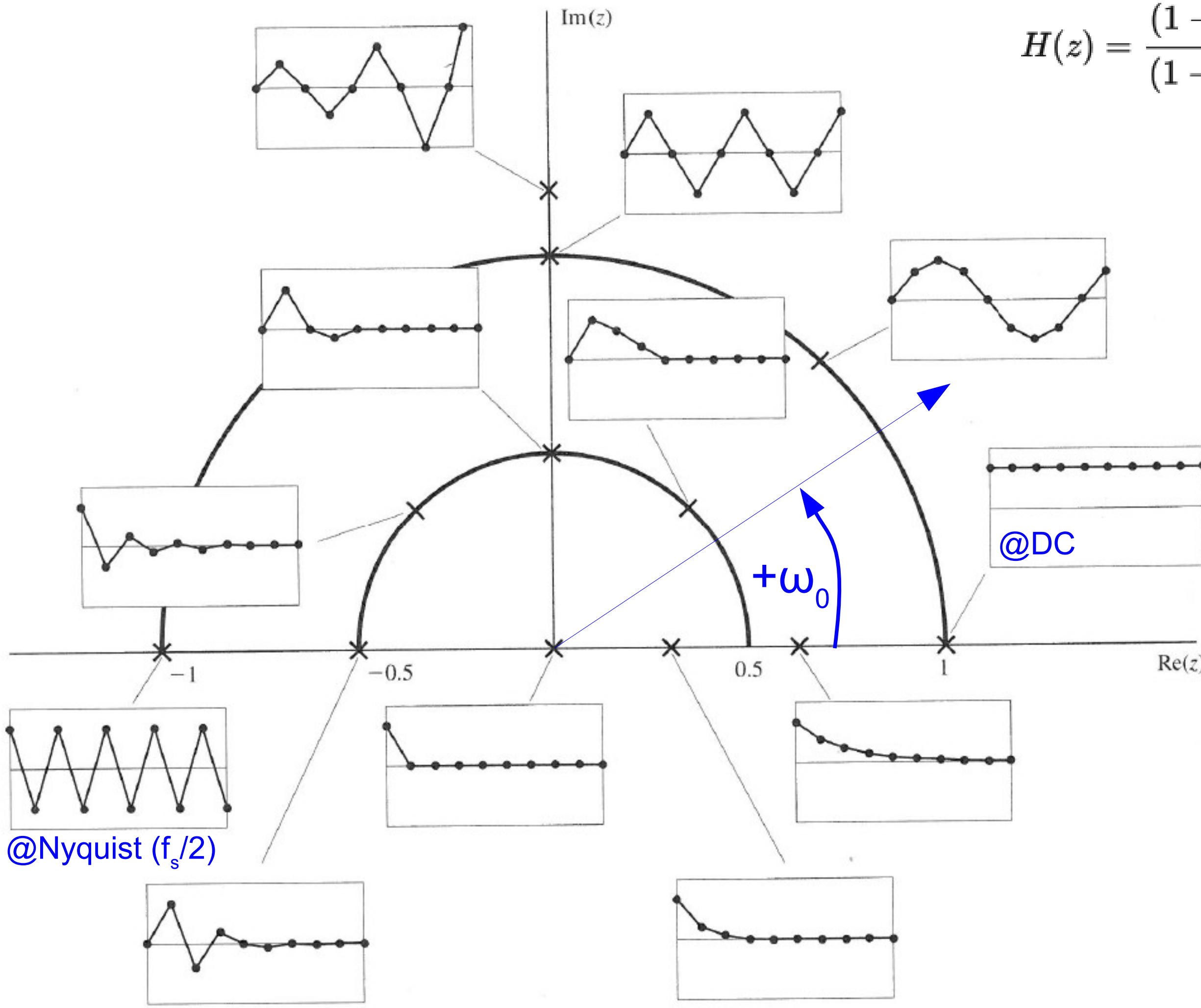


Z-Transform

$$X(z) = \mathcal{Z}\{x[n]\} = \sum_{n=0}^{\infty} x[n] z^{-n}.$$

- System response expressed in terms of sampling frequency $f_s/2$

$$H(z) = \frac{(1-q_1 z^{-1})(1-q_2 z^{-1})\cdots(1-q_M z^{-1})}{(1-p_1 z^{-1})(1-p_2 z^{-1})\cdots(1-p_N z^{-1})}$$



very handy:
bilinear transform
's' → 'z' domain

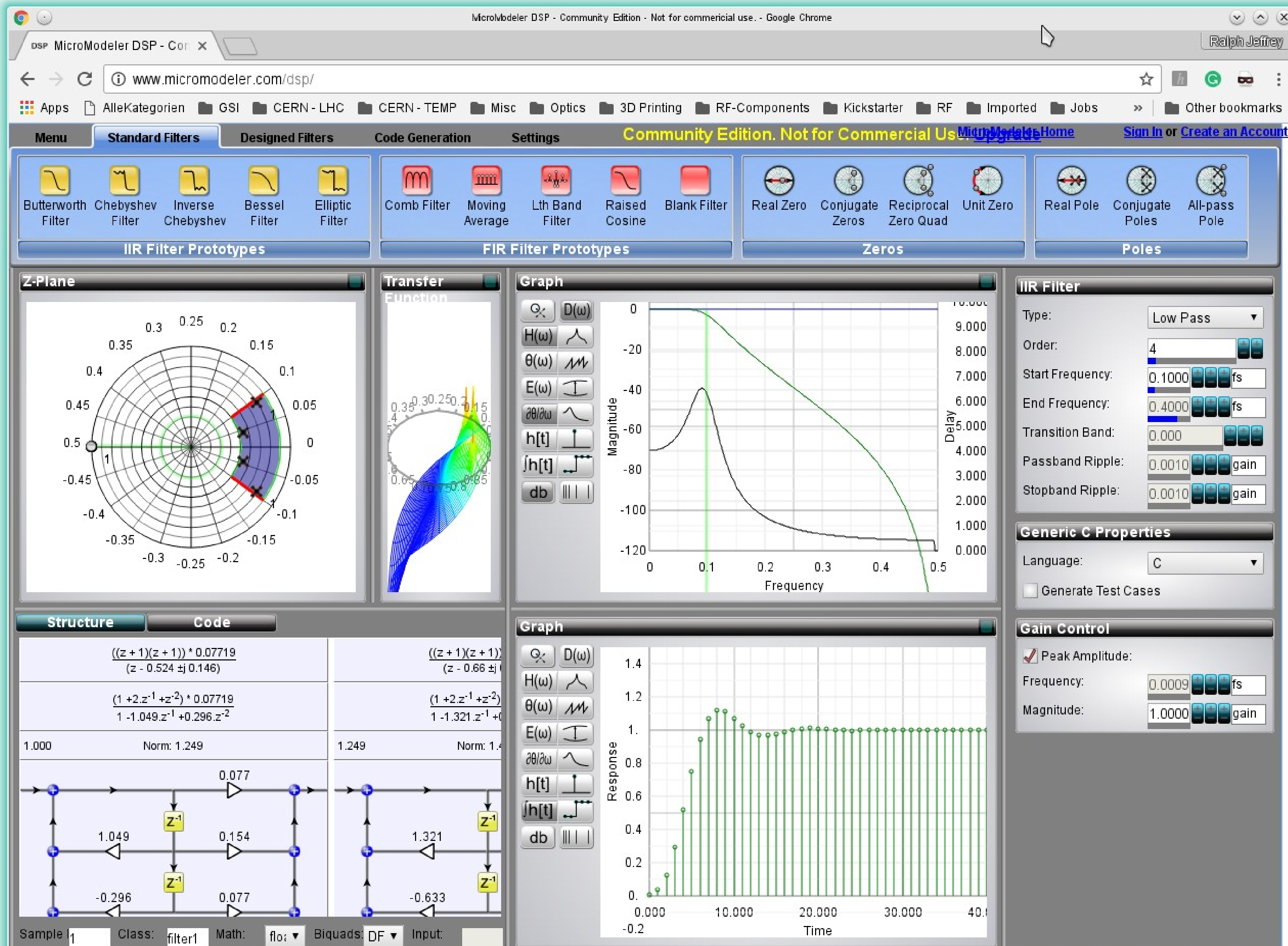
$$\boxed{s = \frac{2}{T}\frac{(z-1)}{(z+1)}}$$

Minimum-phase:
invertible, causal and stable transfer function
↔ all poles & zeros within unity circle

Figure from:
Franklin, Powell, Workman, “Digital Control of Dynamic Systems”, Addison Wesley, 3rd edition, 1997

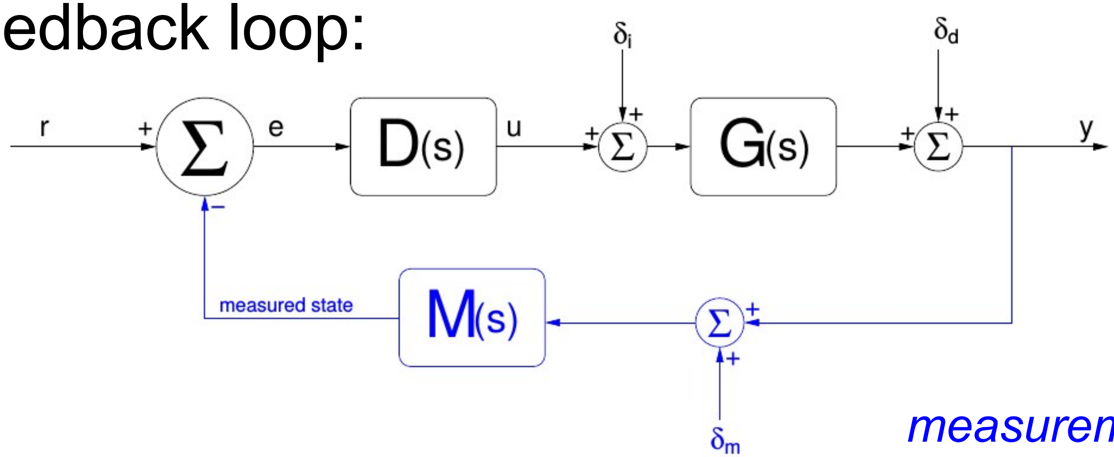
Good & Free Digital-Filter-Design Tool

http://www.micromodeler.com/dsp/



Feedback Basics – revisited

- Simple feedback loop:



transfer function: $$T_0(s) := \frac{y}{r} = \frac{D(s)G(s)}{1+D(s)G(s)}$$

measurement noise sensitivity:

$$\boxed{S_{n0}(s) := \frac{y}{\delta_n} = \frac{D(s)G(s)}{1+D(s)G(s)}}$$

disturbance rejection: $$S_{d0}(s) := \frac{y}{\delta_d} = \frac{1}{1+D(s)G(s)}$$

to note: Johnson noise ~ $\sqrt{\Delta f_{BW}}$

input sensitivity: $$S_{i0}(s) := \frac{y}{\delta_i} = \frac{G(s)}{1+D(s)G(s)}$$

control sensitivity: $$S_{u0}(s) := \frac{u}{\delta_d} = \frac{D(s)}{1+D(s)G(s)}$$

- Important take-away:
  a) improving closed-loop (gain-) bandwidth $T_0(s)$ also improves disturbance rejection $S_d(s)$
  b) improving closed-loop (gain-) bandwidth $T_0(s)$ also increases sensitivity to measurement noise

  requires trade-off

Fundamental FB Design Paradigms

Before you start designing/implementing your controller ...

Spend early-on some time on thorough and detailed analysis of...

- Stability: *“parameter should be ~ reproducible within the targeted FB bandwidth … “*
- Controllability: “... affine (not necessarily linear) dependence between observable and control actuator, ...”
- Observability: “... ability to measure it reliably (noise, systematics, MTBF, …), ...”
- Control System Integration: “... ability to use, pre-debug and re-tune the system by non control-theory experts during day-to-day operation ...”
  - appropriate parametrisation, definition, training, … integration into OP environment
  - logging/archiving, error handling, fault analysis/failure diagnosis, …
  - (Re-)validation of nominal system performance

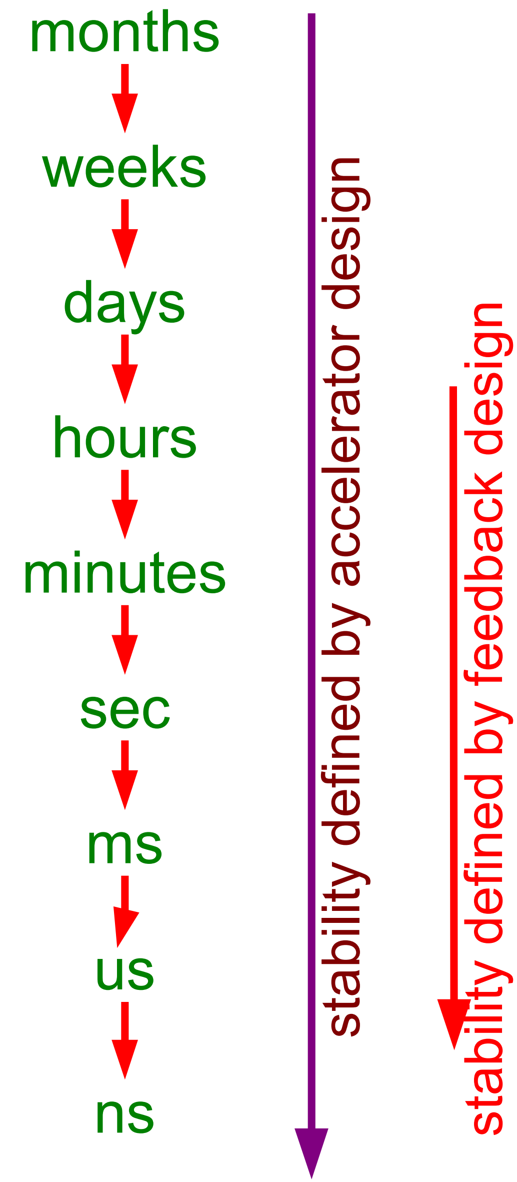
… in order to safe you time later-on w.r.t. debugging, retuning, etc.

FB Design Paradigms – Stability

Perturbation Sources or "Know your enemy"

...can be grouped into:

- Environmental sources:
  (mostly propagated through quadrupoles/girders)
  - temperature and pressure changes,
  - ground motion, tides,
  - 'cultural noise'
- Machine inherent sources:
  - decay and snap-back of magnetic multipoles,
  - cooling liquid flow, pumps/ventilation vibrations
  - eddy currents
  - changes of machine optics (feed-down effects)
  - machine impedance, trapped RF modes/wake-fields (RF)
  - Intensity-related and collective effects
- Machine element failures:
  - magnet quenches, power converter/RF trips, ...
  - corrector circuits (e.g. LHC: 1300++ circuits)



Observability I/II

Sensor technology choices → Know you Input Devices!!

- Feedbacks are as much about control laws as they are about choosing the right sensor (and actuators) for the job.

- Some common ~~mistakes~~ learning experiences:
  - assumption that instrument/sensors are perfect
    - ignores: noise, lag, limited bandwidth & dynamic range, ...
    - often optimised rather for BI than for FB constrains (i.e. lag ↔ noise)
      - e.g. massive low-pass applied to compensate for resolution → not ideal for feedback application (ADC lag + add. low-pass lag)
    - real-time: performance depends not only on correct result but when it is delivered
  - BI setup remains valid after initial commissioning
    - beam parameter changes as the machine performance improves
      - i.e. beam intensity, number of bunches, ...
    - machine modifications, addition of new insertion devices, ...
    - less-precisely known/new beam physics effects (e.g. collective effects)
    - Most accelerator R&D are moving targets → continuous improvement process

- Complexity & effort increases depending on type of parameter
  - 1^{st}- order: current, voltage, frequency, transmitted/reflected RF power, ...
  - 2^{nd} order: beam-current, beam-losses, wire-grids, screens, ...
  - complex dependence on 'diagnostic methods'*:
    - RF cavity Q-value & resonance frequency, tune, chromaticity, luminosity, …
    - Phase-detection → fast transverse FBs & Tune-PLLs

advise: think in terms of

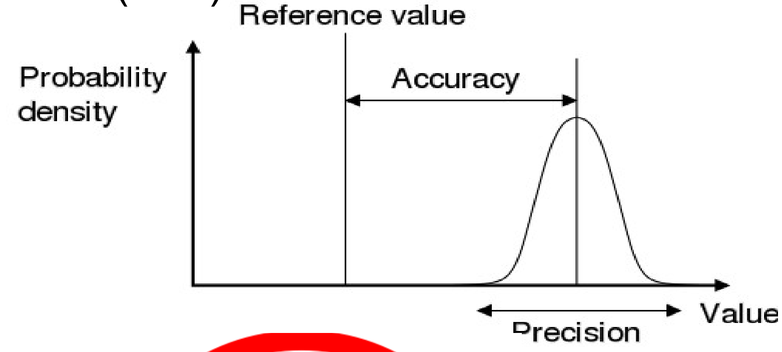
'reliability engineering' & FMECA

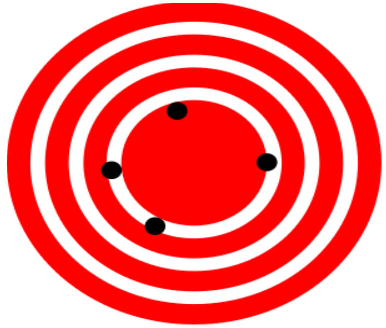
* diagnostics = the combination of instruments and measurement procedures

Observability II/II
Accuracy & Precision (also ISO 5725)

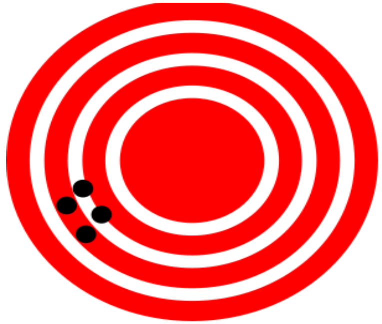
Good summary: http://en.wikipedia.org/wiki/Accuracy_and_precision

- **Accuracy:** "[..] closeness of measurements [..] to its actual (true) value"
- **Precision** (also: reproducibility or repeatability):
  "[..] degree to which repeated measurements under unchanged conditions show the same results."
- Example: "Target analogy" and the two extreme cases





High **accuracy**, but low **precision**
obtained through beam-based alignment



High **precision**, but low **accuracy**
we need this from the sensors

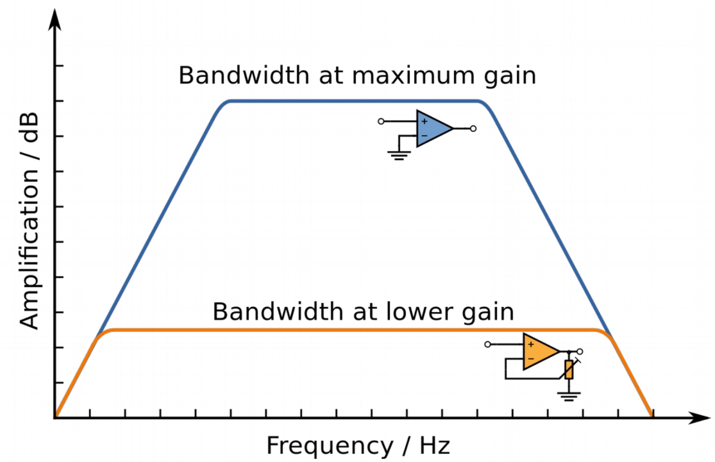
- **Resolution**: smallest change that produces a response in the measurement
- N.B. 'precision' is often sufficient for feedback operation

FB Design Paradigms – Stability

Perturbation Sources or "Know your enemy" → Actuator Choice

Cannot emphasize this enough: → it's worth to spend some studies on this:

- the justification whether and to what extend a fast/slow feedback is actually necessary:
  - e.g. bandwidth beyond technical control means → change machine design
    - improve e.g. magnet, power converter spec.
  - e.g. parameter more stable than what could be achieved via feedbacks ↔ required bandwidth vs. noise rejection
  - cost-benefit analysis



- definition of requirements on bandwidth, resolution, precision and accuracy → driver behind technology choice
  - required actuator gain-bandwidth product
  - other control parameters (divide/group systems together)

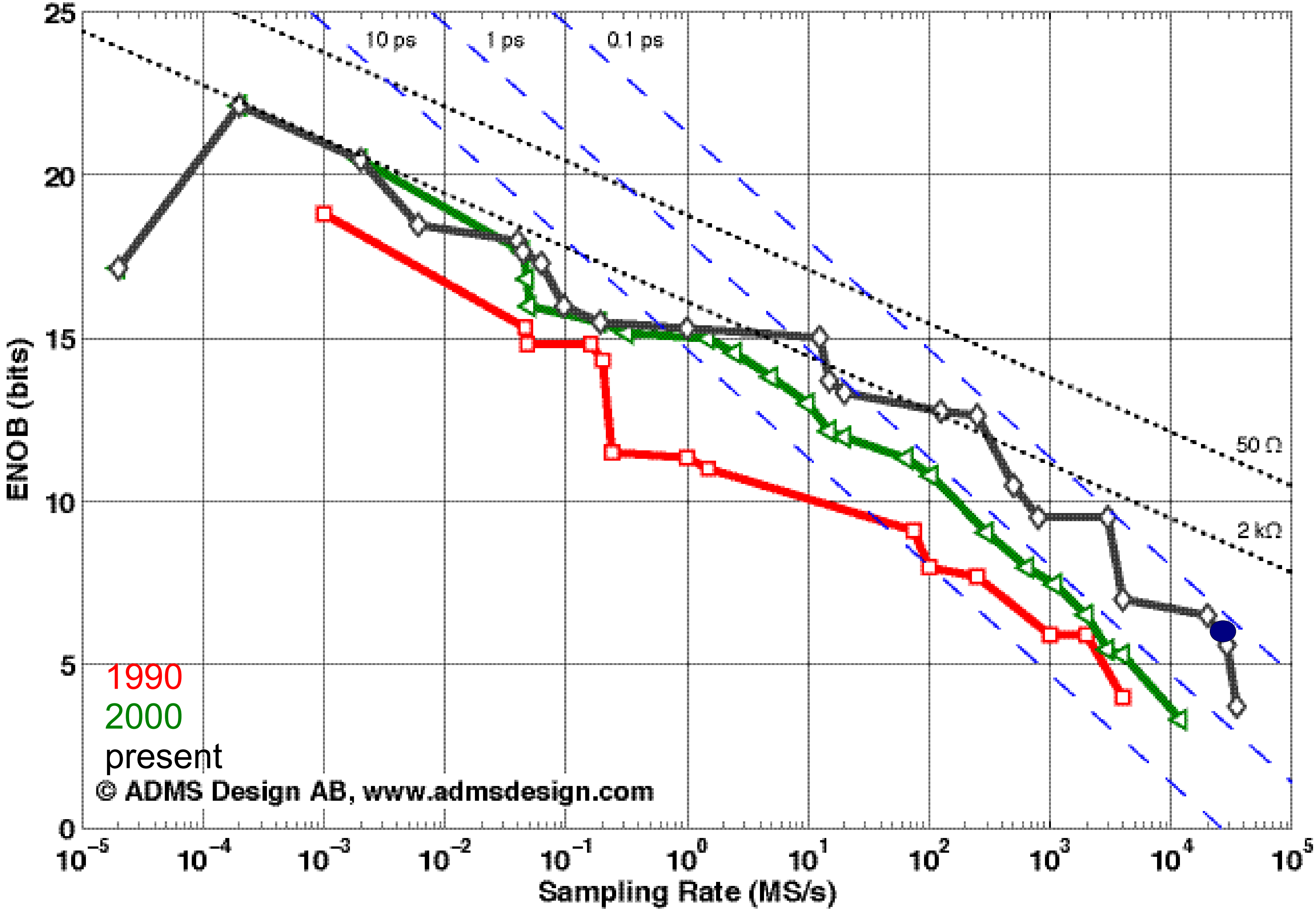
- primary decision point w.r.t. 'analog' vs. 'digital' determined by analog bandwidth
  - > ~ 250 MHz … GHz → mostly analog hybrid-designs
  - 10 MHz … < ~ 250 MHZ → mostly digital (DSP/FPGA)
  - … 10 MHz: → digital low-cost MCU & PC-based

- Looking forward to seeing some ideas, designs and implementation here at PCaPAC!

Digital vs. Analog Feedback Design

Limits of direct time-domain digitization

- ADCs' performance level out and approach fundamental physics limits



Digital vs. Analog Feedback Design

- Pro digital:
  - reproducibility: signals not subjected to temperature/environment changes or ageing
  - programmability/upgradable (start basic → upgrade during operation)
  - performance: possibility to implement algorithms not feasible in the analog domain
    - RF domain: direct digital down-conversion (superior phase/amplitude stability)
    - possibility to combine basic control algorithms and additional useful features like signal conditioning, saturation control, delay compensation, gain-scheduling, down-sampling, etc.;
  - implementation of diagnostic tools, used for both feedback commissioning and machine physics studies
  - easier and more efficient integration of the feedback in the accelerator control system, important for feedback set-up and tuning, fast data acquisition, easy and automated operations, etc.
- But also some disadvantages is the
  - higher delay of the feedback chain (due to ADC, digital processing, and DAC) with respect to equivalent analogue feedback (although with the use of FPGAs this delay is often reduced to acceptable values)
  - Dynamic range, bandwidth & digitization noise:
    - ADCs ENOB-vs-Sampling limitations (thermodynamics)
    - easier to make very broad-band, high-dynamic range, or low-noise analog systems
- The best choice is somewhere between: need a good AFE & high-level digital control

Continuous vs. Discrete Feedback Design

Design in s- or z-Parameter Space?

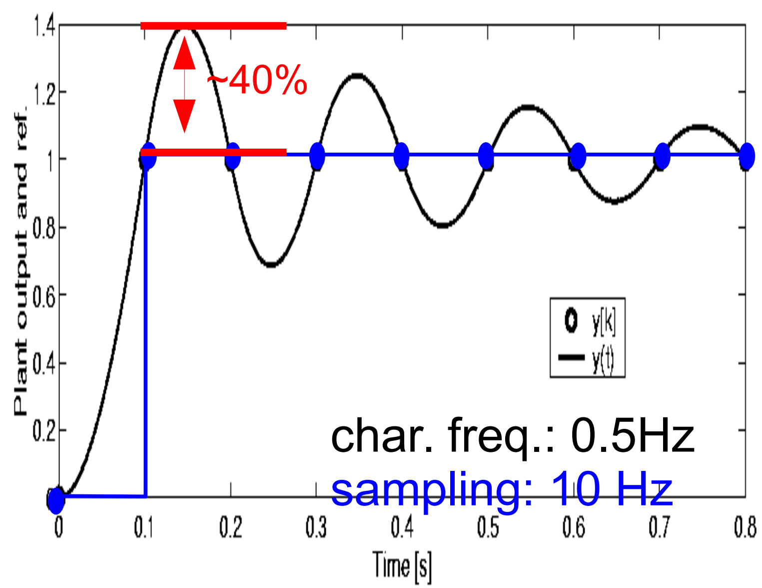
- Continuous time design which is discretized for implementation:
  - Continuous time signals and models during design → prior to implementation, the controller is replaced by an equivalent discrete time version (s → δ mapping, with δ being the delta operator)
  - assumption that the sampling rate is high enough to mask sampling effects
  - If the sampling period is chosen carefully, in particular with respect to the open and closed loop dynamics, then the results should be acceptable.
  - My personal preference:
    - a) allows decision of sampling frequency at the end
    - b) easier to model/design multi-rate feedbacks

- Discrete design based on a discretized process model → discrete controller
  - Caution must be exercised with so called inter-sample behaviour: the analysis is based entirely on the behaviour as observed at discrete points in time, but the process has a continuous behaviour also between sampling instances;
  - Problems can be avoided by refraining from designing solutions which appear feasible in a discrete time analysis, but are known to be unachievable in a continuous time analysis (such as removing non-minimum phase zeros from the closed loop!).

Loop Bandwidth versus Sampling Frequency I/II

Classic argument: Analogue vs. Digital Design

- Among many arguments (short-version):
  - Pro analogue: most process to be controlled are analogue (only thermal noise limit)
  - Pro digital: most controller are nowadays digital (thermal noise, clock, ENOB limits)
    - "Con-example": digital only controller design (inter-sample response)



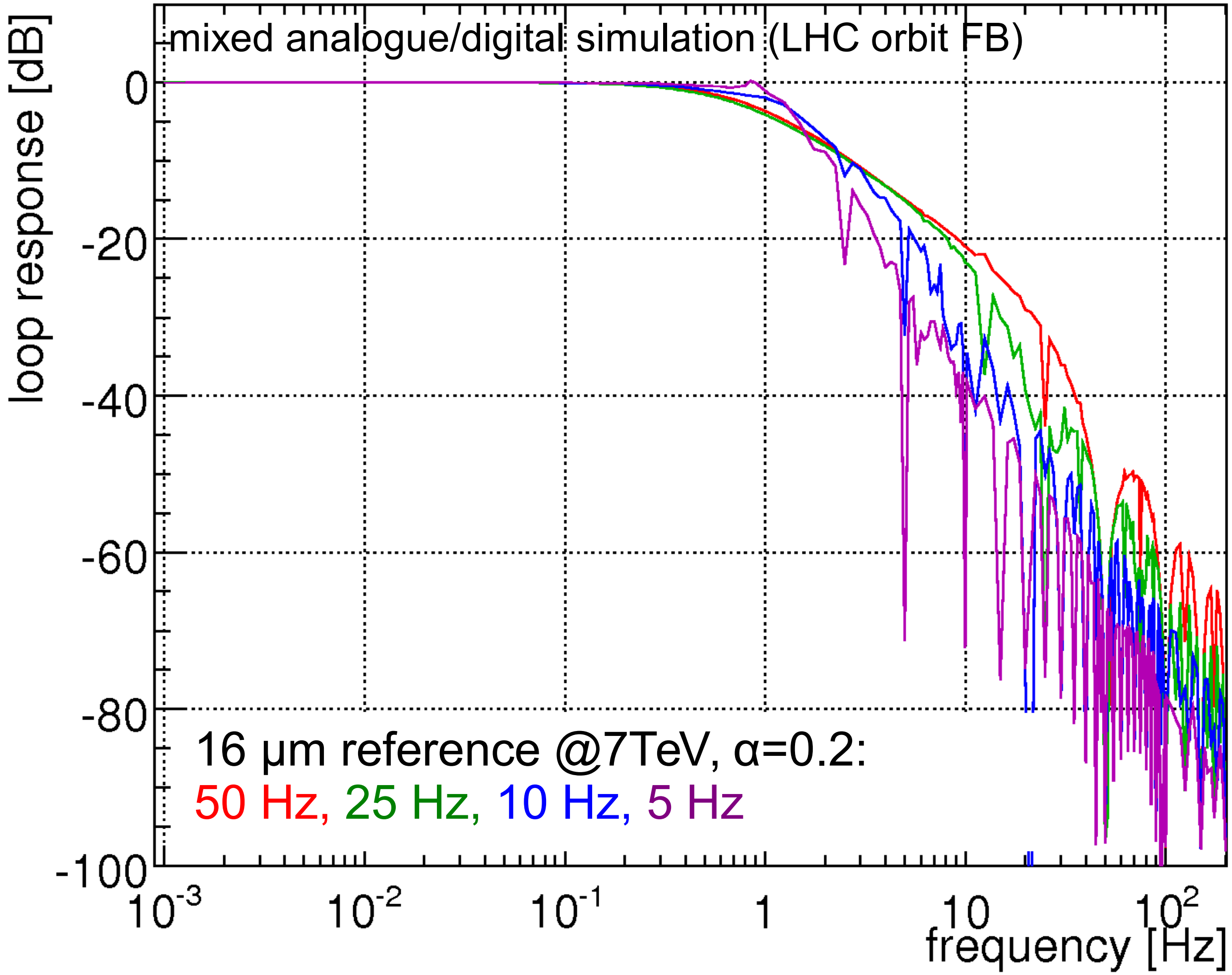
perfect digital response
but ~40% "analog" overshoot

- The following rules of thumb will help avoiding (inter-)sample problems
  - iterative design approach between analogue and digital domain
  - sample 10, better 20-40 times the desired closed loop bandwidth
    - improves inter-sample responses & phase-margin (↔ important for very fast FBs)
  - use simple anti-aliasing filters (low-order to avoid excessive phase shift)
  - never try to cancel or otherwise compensate for discrete sampling zeros!
  - always check the inter-sample response.

Loop Bandwidth versus Sampling Frequency II/II

Example: LHC orbit/Q/Q'/... feedback design

- ... 10Hz sampling to achieve a closed loop 1Hz bandwidth:



  - ... a theoretic limit assuming a perfect system (no noise, model errors)!
  - common sense/advise: f_s > 25 ...40 x desired closed-loop bandwidth f_{BW}

Digital Control System & 'Real-Time' I/III

Some common Misconceptions & Fallacies[1]

- As soon as your controller needs to do two or more things in parallel one runs into the domain of task scheduling and real-time constraints
  - e.g. primary controller function + monitoring of FB function, setting changes/gain scheduling, interlocks,
- What you often hear:

1. *"There is no science in real-time-system design"*
2. *"Advances in supercomputer hardware will take care of RT requirements."*
3. *"[..] is equivalent to fast computing."*
4. *"[..] research is performance engineering."*
5. *"[..] systems function in a static environment."*
6. *"[..] is assembly coding, priority IRQ programming, and device driver writing."*
7. *"[..] all been solved in other areas of computer science or operations research."*
8. *"It is not meaningful to talk about guaranteeing RT performance, because we cannot guarantee that the hardware will not fail and the software is bug free or that the actual operating conditions will not violate the specific design limits."*

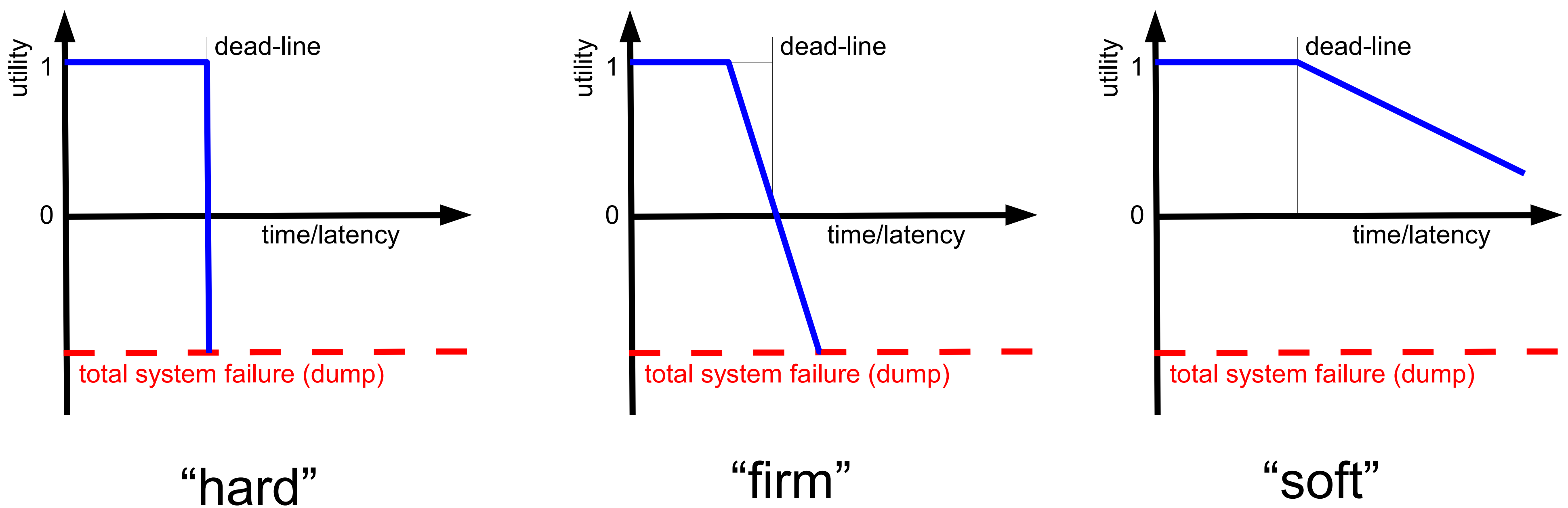
- Obviously, the above is wrong but seems to be sometimes forgotten when discussing the specific technical implications.

[1]John A. Stankovic, "Misconceptions about real-time computing: a serious problem for next-generation systems", IEEE Computer, Vol. 21 #10, 1988

Digital Control System & 'Real-Time' II/III

Definitions

- … "A system is said to be real-time if the total correctness of an operation depends not only upon its logical correctness, but also upon the time in which it is performed. [..] are classified by the consequence of missing a deadline:
  - Hard – Missing a deadline is a total system failure.
  - Firm – Infrequent deadline misses are tolerable, but may degrade the system's quality of service. The usefulness of a result is zero after its deadline.
  - Soft – The usefulness of a result degrades after its deadline, thereby degrading the system's quality of service."

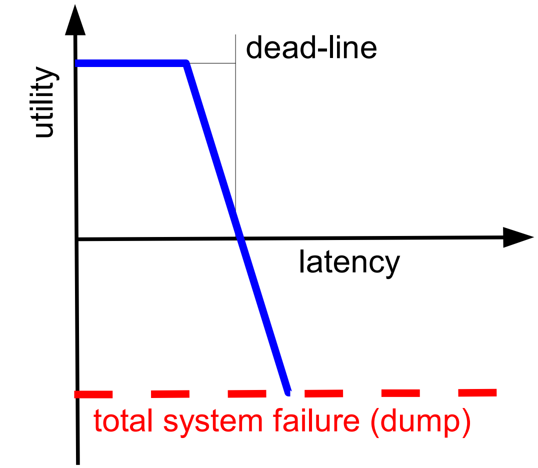


Digital Control System & 'Real-Time' III/III

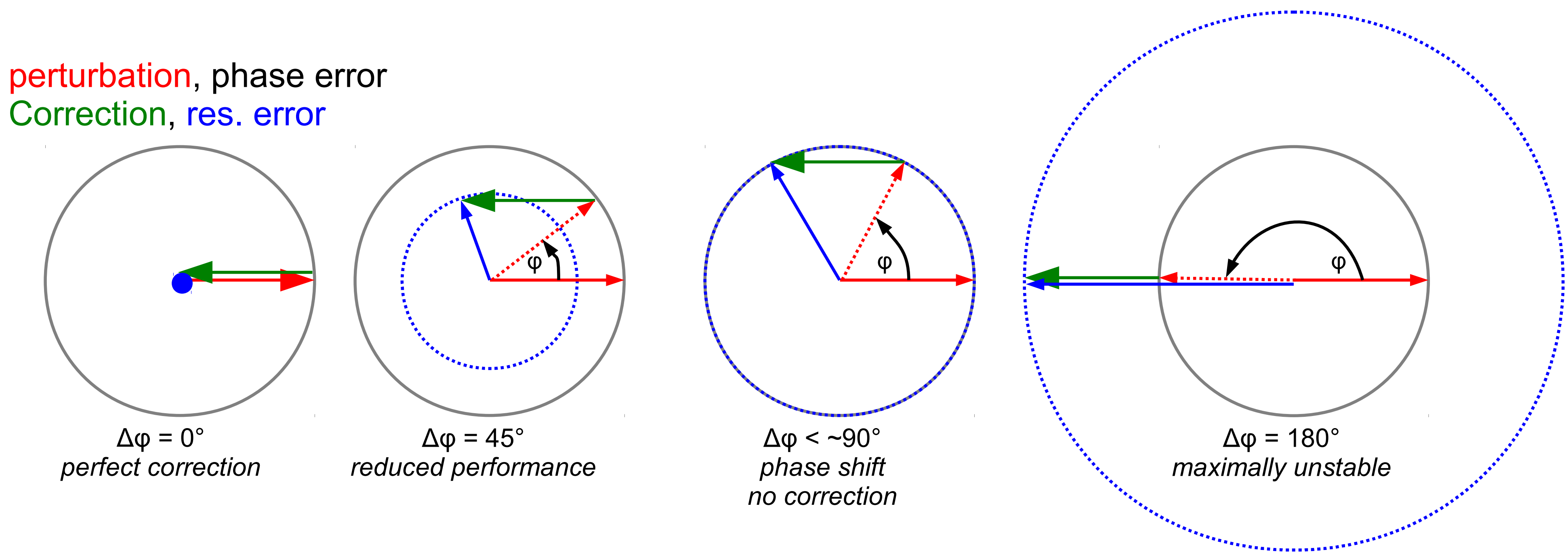
actual impact on feedback loops

- Most feedbacks in accelerator context are 'firm real-time systems'
  - some (limited) margin on occasional missing data
  - additional latencies are critical for loop stability

$$\Delta\phi = 2\pi f_{bw} \cdot \Delta t_{delay}$$



- “How much phase stability is required (i.e. @... MHz)?”



Δφ = 0°
perfect correction

Δφ = 45°
reduced performance

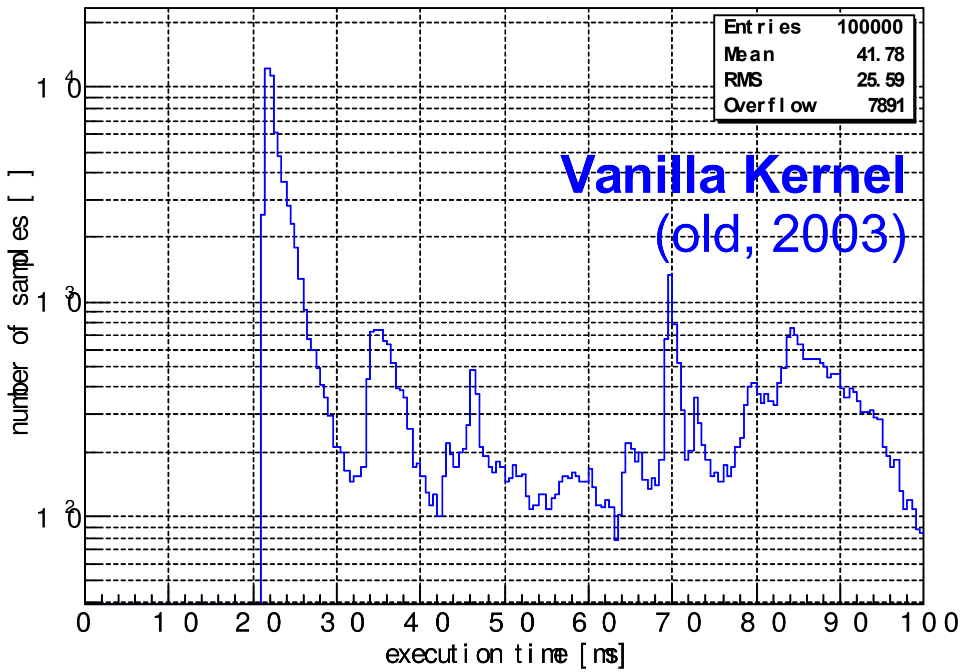
Δφ < ~90°
phase shift
no correction

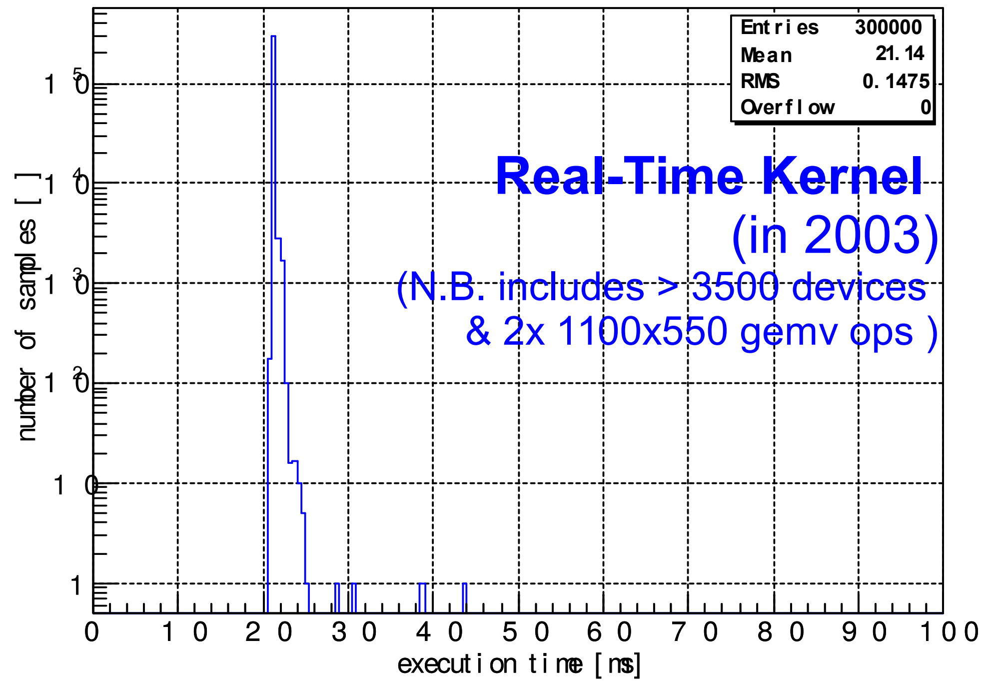
Δφ = 180°
maximally unstable

Real-Time Technology Choices

Real-World Example: real-world Real-time vs. Standard (Vanilla) Linux Kernel

- LHC OFC stress tests under IO, CPU and network load → complete loop latency:





- Worst-case latencies < 50 us (RaspPI) routinely & down-to 10 us possible (HW dep.)
- Some Tips:
  - get rid/disable non essential services (apm, IRQ balance, update-services, ...)
  - use IO & thread-CPU-affinity to shield RT-critical tasks from low-priority tasks (biggest gain)
    - consider running critical tasks on dedicated slave-MCU (chose-your-favourite-flavour)
  - analyse which threads/IO/RAM are actually needed → static allocation at programme start
  - analyse and verify (test) numerical complexity (big-O notation, avoid if/else, …)
  - https://rt.wiki.kernel.org/index.php/HOWTO:_Build_an_RT-application
  - https://rt.wiki.kernel.org/index.php/CPU_shielding_using_/proc_and_/dev/cpuset

Non-Optimal but Practical Feedback Controller Design

Imagine you are …

- … being called during the night, sleepy, drowsy
- … visiting some external accelerator laboratory
- … forgot your FB model design parameters
- … don't care for optimal design but some PID settings that just work
- … your life depends on getting a PLC PID going (the McGuyver scenario)

Who doesn't want to be like MacGyver?

- Don't worry there is a McGyver approach to FB controller design
  - often sufficient, especially for 1^{st}- and 2^{nd}-order system responses

Non-Optimal but Practical Feedback Controller Design
For reference: Historic Empirical PID Tuning Methods I/II

- Ziegler and Nichols[1], and later Cohen and Coon[2] proposed a generic tuning technique without knowing/doing the system modelling/analysis:
  – measure closed-loop response to step reference change, increase K_p (only) until system becomes unstable (↔ phase-margin = π), note gains and oscillation period when the process became unstable '$K_p \rightarrow K_c$' and P_c then use:

	Ziegler-Nichols Oscillation Method		
	K_p	K_i	K_d
P	$0.50 \cdot K_c$		
PI	$0.45 \cdot K_c$	$0.54 \frac{K_c}{P_c}$	
PID	$0.60 \cdot K_c$	$\frac{1.2 K_c}{P_c}$	$\frac{3 K_c P_c}{40}$

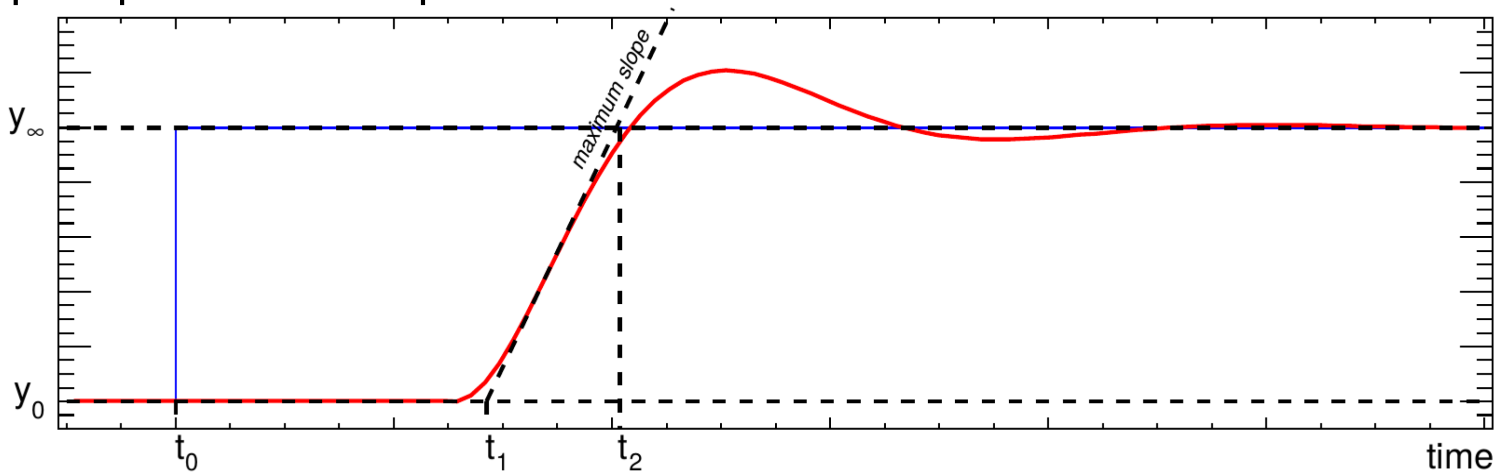
Provides a good base-line ↔ your optimal-controller design should "beat this"

[1] G. Ziegler and N. B. Nichols, "Optimum settings for automatic controllers". Trans. A.S.M.E., 64:759–765, 1942
[2] Coon and Cohen, "Theoretical consideration of retarded control", Trans. A.S.M.E., 75:827–834, 1953

Non-Optimal but Practical Feedback Controller Design

For reference: Historic Empirical PID Tuning Methods II/II

- If you feel uneasy (or may damage equipment) w.r.t. driving your system unstable use the open-loop response to a step:



One defines the following parameter

$$K_0 := \frac{y_{ss} - y_0}{u_{ss} - u_0}, \qquad \tau_0 = t_1 - t_0, \qquad \text{and} \qquad \nu_0 = t_2 - t_1$$

	Ziegler-Nichols		
	K_p	K_i	K_d
P	$\frac{\nu_0}{K_0\tau_0}$		
PI	$\frac{0.9\nu_0}{K_0\tau_0}$	$\frac{1}{3\tau_0} \cdot K_p$	
PID	$\frac{1.2\nu_0}{K_0\tau_0}$	$\frac{1}{2\tau_0} \cdot K_p$	$\frac{\tau_0}{2} \cdot K_p$

	Cohen-Coon		
	K_p	K_i	K_d
P	$\frac{\nu_0}{K_0\tau_0}\left[1 + \frac{\tau_0}{3\nu_0}\right]$		
PI	$\frac{\nu_0}{K_0\tau_0}\left[0.9 + \frac{\tau_0}{12\nu_0}\right]$	$\frac{9\nu_0 + 20\tau_0}{\tau_0\left[30\nu_0 + 3\tau_0\right]} \cdot K_p$	
PID	$\frac{\nu_0}{K_0\tau_0}\left[\frac{4}{3} + \frac{\tau_0}{4\nu_0}\right]$	$\frac{13\nu_0 + 8\tau_0}{\tau_0\left[32\nu_0 + 6\tau_0\right]} \cdot K_p$	$\frac{11\nu_0 + 2\tau_0}{4\tau_0\nu_0} \cdot K_p$

The Cohen-Coon method seems generally to produce solutions with less overshoot compared to the Ziegler-Nichols reactive curve method.

Intermediate Summary I

- Beam-based FBs are remedies for perturbations on slow/medium time scales
  - limited by thermal drifts, noise and systematics of involved devices

- In Accelerators, feedback optimal control problems are mostly frequency-scale invariant
  - allows to share design, parameters and concepts across different domains
  - provides some advantages w.r.t. FB operation and system integration
- Still, technology choice for implementation should be adapted to specific problem
  - Pre-requisite: systematic and thorough analysis of required 'Stability', 'Controllability' (↔ actuators), 'Observability' (↔ instrumentation) <u>& CO/OP integration</u> is essential!
  - Equivalence of continuous vs. discrete design → typically hybrid design

- Possible implementation options:
  - low bandwidths (< 10 MHz): → low-cost, rapid-prototyping, fast modifications
    - PC + MCUs (dealing with ADC/DACs) hybrid, notably software-defined-radios (SDR), …
  - medium bandwidth (10 … 250 MHz) → bit less flexible but easier RT requirements
    - purpose built DSP and FPGA-based processing boards
    - N.B. have a look at the latest generation of SDR (up to 160 MHz analog bandwidths)
  - high-bandwidth (> 250 MHz) → still mostly analog designs
    - provides easiest/best dynamic range, bandwidth, noise
    - N.B. usually with digital support w.r.t. providing monitoring/references/gain scheduling

- Ziegler-Nichols/Coohen-Coon PID tuning are outdated but sometimes still useful

Food for thought:
Controls engineering proverb concerning machine stability:
“Trust is good, control is better, … stable feedbacks are best!”

Break

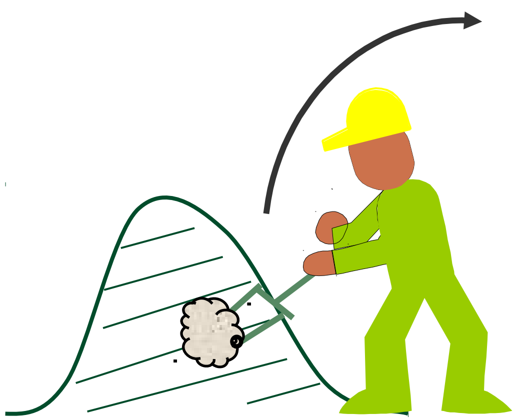
Controls engineering wisdom:

A) Most machines are stable with <u>negative</u> feedback
B) Humans work better (are stable?) with <u>positive</u> feedback

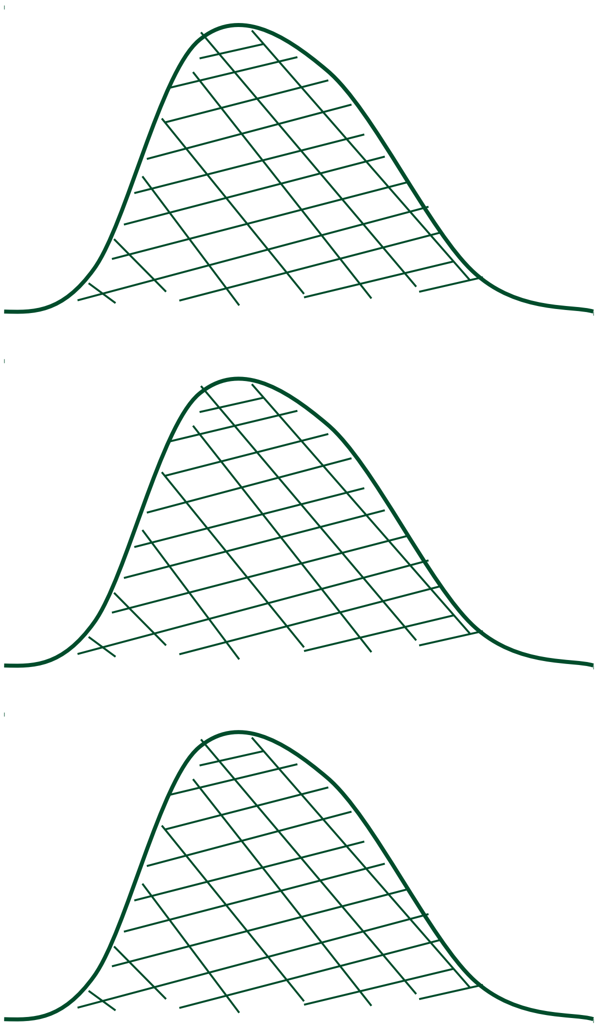
→ i.e. Humans are not like most machines, know the difference!

Optimal Linear Multiple-Input-Multiple-Output FB Design

Multivariable Case



Sensitivity dirt



Multiple piles

courtesy G. Goodwin

Control Problem Categories in Accelerators
(based on my personal experience)

A) Relatively simple loops for linear minimum-phase system
(e.g. 1st & 2nd-order systems, harmonic oscillators, or composites)
→ classic PID design gives often a very satisfactory solution

- examples: power converter, RF power/cavity controller, ...

~ 95 %

B) Slightly more complex, mild non-linear systems where an additional feature beyond classic PID yields significant performance advantages

- feed-forward control: improves measurement noise performance, FB dependability → reliability and fault sensitivity)
- Anti-windup scheme for rate-limited systems
- Smith predictor for significant time delays

C) Systems involving significant interactions but where some form of preliminary compensation essential converts the problem into separate non-interacting loops which then fall under the above categories

- adaptive gain-scheduling ↔ non-linear control, noise/bandwidth trade-off

~ 4 %

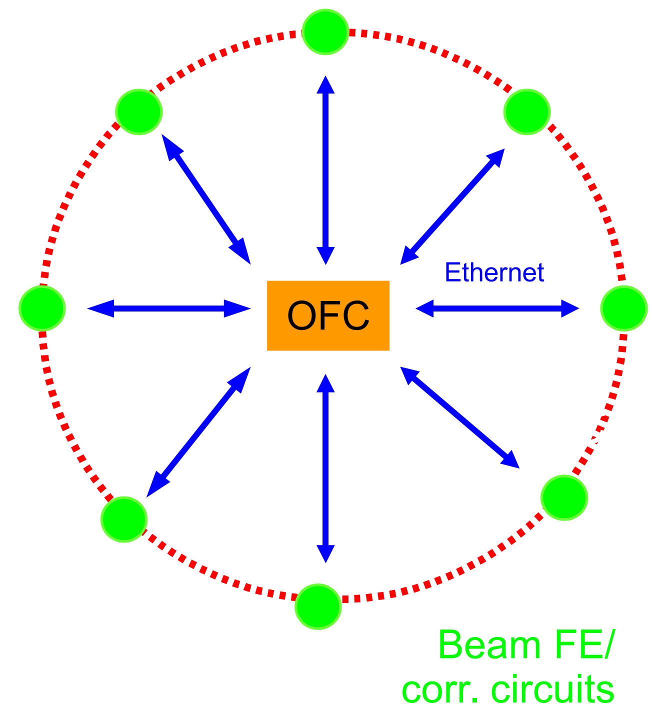
D) Exotic/difficult problems which require some form of numeric optimisation (e.g. non-linear, open-loop unstable, MIMOs)

- *hard* '*make*-or-*break*' problems → the joy of every hardcore control engineer but nightmare of day-to-day operation

~ 1 %

Orbit-Feedback as Prototype for all LHC Beam-Based Feedback Systems

- Orbit-Feedback is the largest and most complex LHC feedback:
  - 1088 BPMs → 2176+ readings @ 25 Hz from 68 front-end computers
  - 530 correction dipole magnets/plane, distributed over ~50 front-end computers
  - Total >3500 devices involved
- Specific requirements fairly distributed
  → opted for central global feedback system
- One central controller (OFC + hot spare):
  - higher numerical load
  - higher network load (↔ ~120 front-ends)
  - dependence of machine operation on single device
  - easier synchronisation between front-ends and FBs
  - flexible correction scheme changes and gain-scheduling
  - most efficient to handle cross-talk and (de-)coupling between FBs



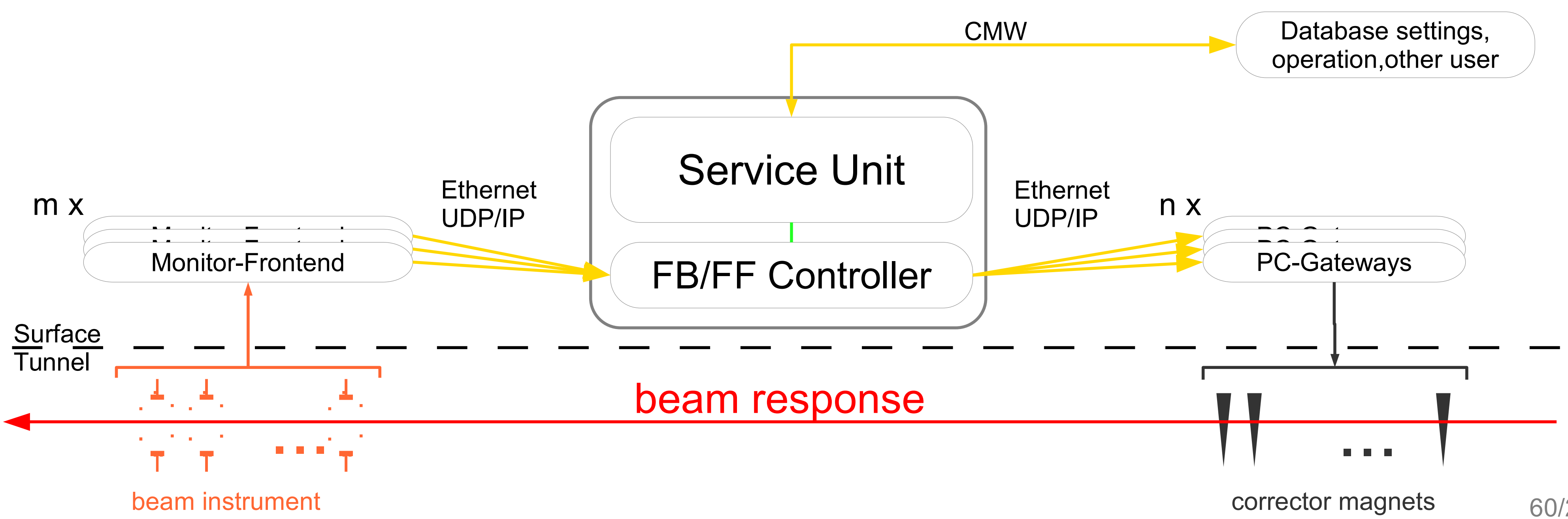
'Massive-Multiple-Input-Multiple-Output' (M-MIMO)
→ will use LHC's beam-based FBs to develop as design concepts

Disclaimer: this is not to express that other facilities have less-good or less-performing designs! Many FB aspects at CERN-LHC's designs are based on years of experience at many other synchrotron-light and collider facilities (notably: SLS, Diamond, Soleil, SLAC, BNL, …)
N.B. applicable technology choices may differ on required bandwidths and infrastructure

Common Feedback/Feed-forward Control Layout

- Feedback Controller (OFC) performing actual feedback controller logic
  - Simple streaming task (10% of total load)
  - Beam data quality checks and real-time filtering (80% of total load)
  - Server running Real-Time Linux OS with periodic constant load
    - multi-core, highly redundant – MTBF > 22 yrs (spec, 120 yrs meas.)
  - Technical Network as robust communication backbone

- Service Unit (OFSU): Interface to high-level software control and interlock systems
  - Proxies user requests, handles asynchronous non-RT tasks

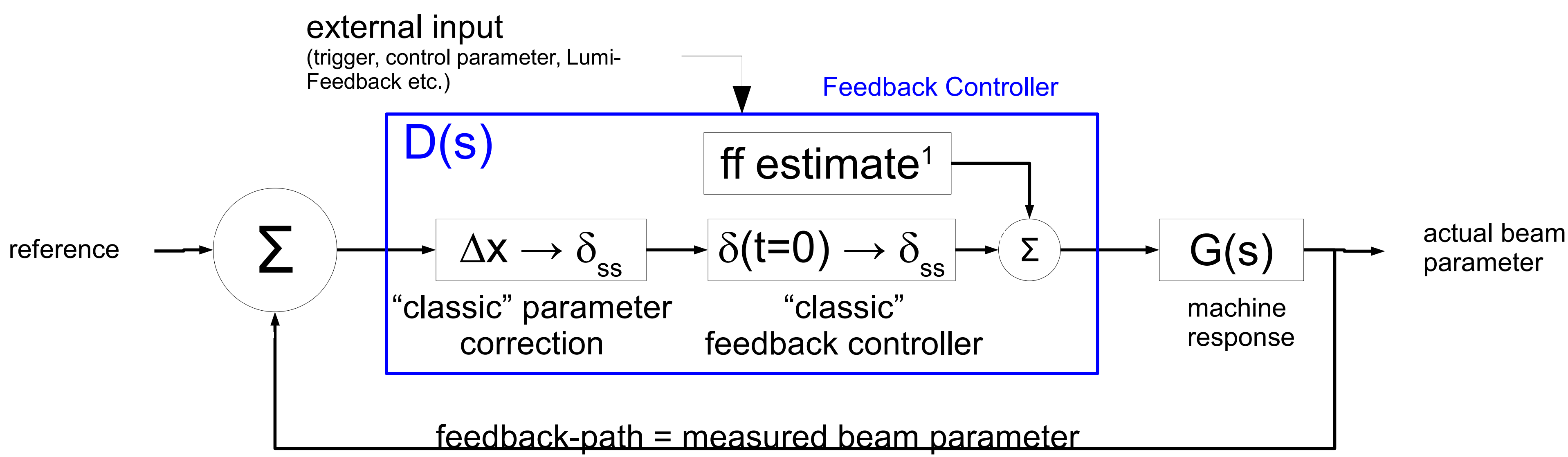


Multiple-Input-Multiple-Output (MIMO) Process Control

- 'Divide and Conquer' feedback controller design approach:
  1. Compute steady-state corrector settings $\vec{\delta}_{ss} = (\delta_{1,}\dots,\delta_n)$ based on measured parameter shift $\Delta x=(x_1,\dots, x_n)$ that will move the beam to its reference position for $t\rightarrow\infty$.
  2. Compute a $\vec{\delta}(t)$ that will enhance the transition $\vec{\delta}(t=0)\rightarrow\vec{\delta}_{ss}$
  3. Feed-forward: anticipate and add deflections $\vec{\delta}_{ff}$ to compensate changes of well known and properly described sources

space domain

time domain



(N.B. here G(s) contains the process and monitor response function)

Why the notion/split between 'space' and 'time' domain?

- Separates specific accelerator physics from specific control theory
  - can test the two domains independently
  - N.B. different/complementary control room expertise

- Multiple-Input-Multiple-Output (MIMO) in space-domain
  - Can modify correction algorithm without having to worry about whether overall loop remains stable
  - Maintains physical meaning of the individual control variables
  - In most cases need to maintain level of synchronisation to minimise inter-loop coupling and consistent solutions (e.g. closure of orbit bump)
  - Basically relying on inversion of response matrices → SVD

- Quasi-Single-Input-Single Output (SISO) in time-domain
  - Similar control problem/laws as e.g. for power converters
  - Time-domain controller identical for orbit, energy, Q/Q' vs. integrated/more complex 'Kalman' or 'Youla-Kucera-Klein'-based method

Space Domain: - No "black feedback magic"

- Effects on orbit, Energy, Tune, Q' and C⁻ but also RF power can essentially cast into matrices:

$$\boxed{\Delta\vec{x}(t)=\underline{R}\cdot\vec{\delta}(t)} \quad with \quad R_{ij}=\frac{\sqrt{\beta_i\beta_j}}{2\sin(\pi Q)}\cdot\cos(\Delta\mu_{ij}-\pi Q)+\frac{D_i D_j}{C(\alpha_c-1/\gamma^2)}$$

matrix multiplication

  - e.g. LHC matrices' dimensions:

$$\underline{R}_{orbit}\in\mathbb{R}^{1070\times 530} \qquad \underline{R}_{Q}\in\mathbb{R}^{2\times 16} \qquad \underline{R}_{Q'}\in\mathbb{R}^{2\times 32} \qquad \underline{R}_{C^-}\in\mathbb{R}^{2\times 10/12}$$

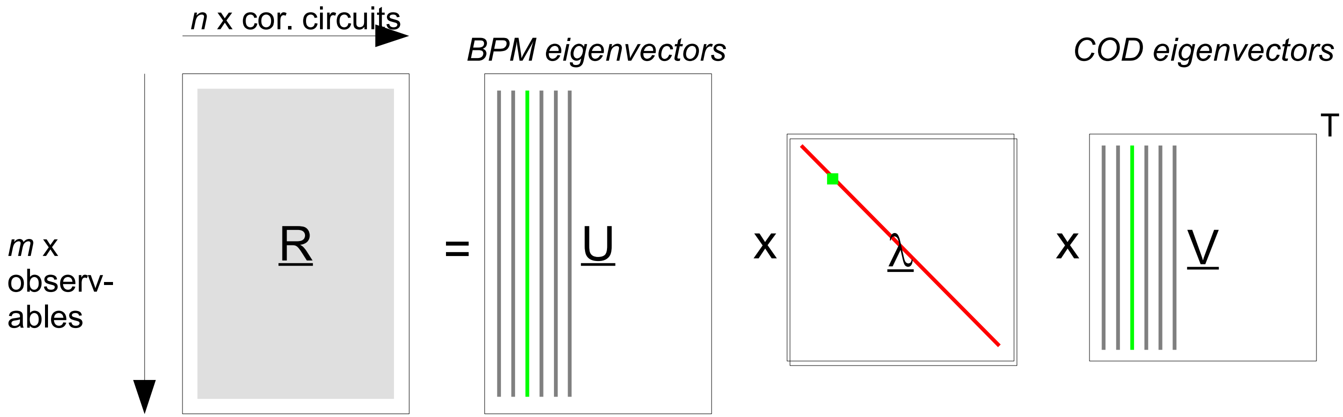
  - control consists essentially in inverting these matrices:

$$\boxed{\left\|\vec{x}_{ref}-\vec{x}_{actual}\right\|_2=\left\|\underline{R}\cdot\vec{\delta}_{ss}\right\|_2<\epsilon \;\rightarrow\; \vec{\delta}_{ss}=\tilde{R}^{-1}\Delta\vec{x}}$$

- Some potential complications:
  - Singularities = over/under-constraint matrices, noise, element failures, spurious BPM offsets, calibrations, …
  - Time dependence of total control loop → "The world goes SVD...."

Space-Domain: Singular Value Decomposition (SVD)
on a slide

Linear algebra theorem*:



$\underline{U}^T \underline{U} = 1$ $\qquad$ $V^T V = V V^T$

eigen-vector relation:

$$\lambda_i \vec{u}_i = \underline{R} \cdot \vec{v}_i$$
$$\lambda_i \vec{v}_i = \underline{R}^T \cdot \vec{u}_i$$

- though the SVD decomposition is numerically very complex, the final correction is a simple vector-matrix multiplication:

$$\boxed{\vec{\delta}_{ss} = \tilde{R}^{-1} \cdot \Delta \vec{x} \quad with \quad \tilde{R}^{-1} = \underline{V} \cdot \underline{\lambda}^{-1} \cdot \underline{U}^T} \Leftrightarrow \quad \vec{\delta}_{ss} = \sum_{i=0}^{n} \frac{a_i}{\lambda_i} \vec{v}_i \quad with \quad a_i = \vec{u}_i^T \Delta \vec{x}$$

- numerical robust, minimises parameter deviations Δx <u>and</u> circuit strengths δ
- Easy removal of singularities, (nearly) singular eigen-solutions have $\lambda_i \sim 0$

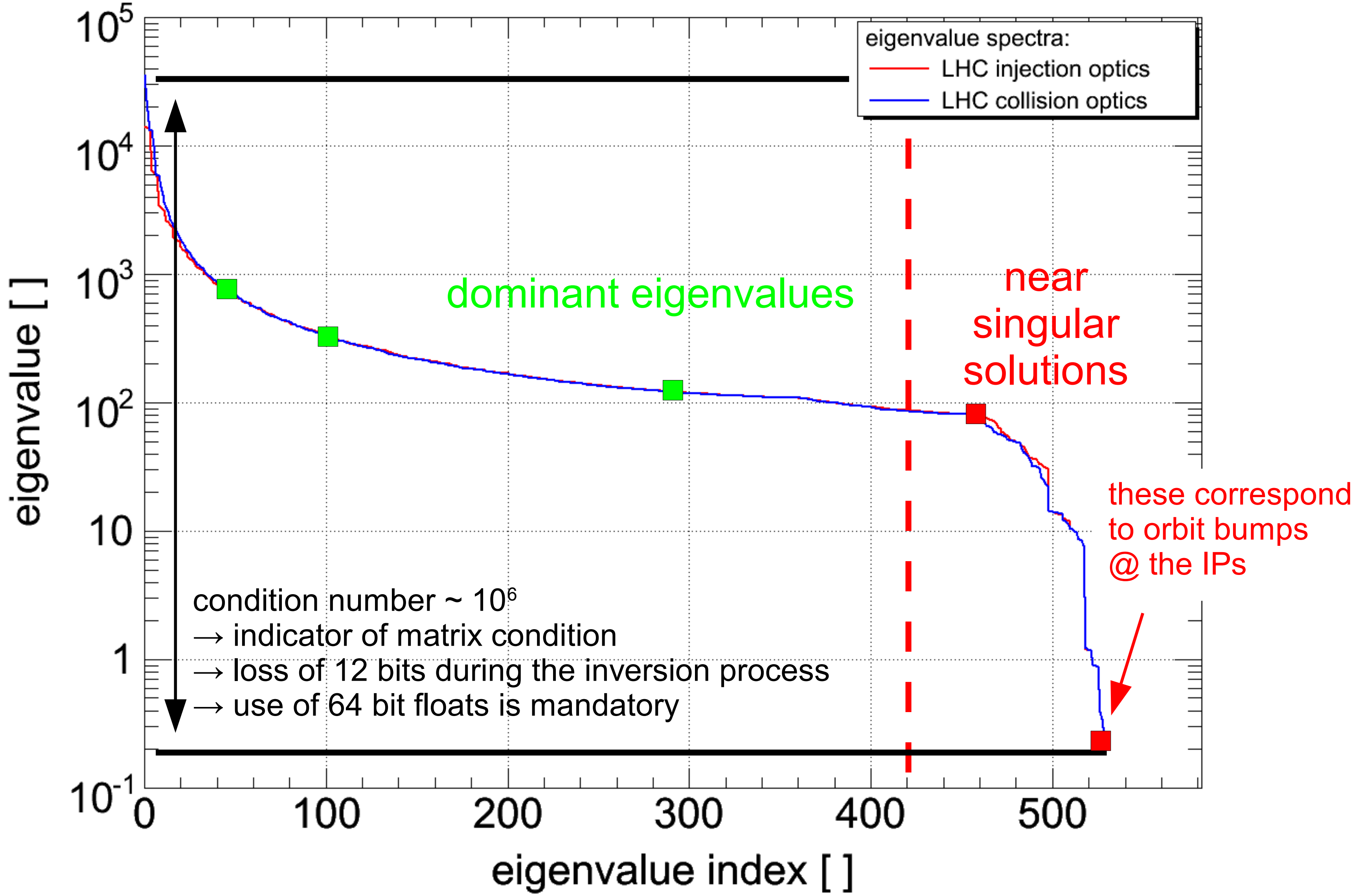
  to remove those solution: if $\lambda_i \approx 0 \rightarrow$ '$1/\lambda_i$:= 0'

  discarded eigenvalues corresponds to solution pattern unaffected by the FB

*G. Golub and C. Reinsch, "*Handbook for automatic computation II, Linear Algebra*", Springer, NY, 1971

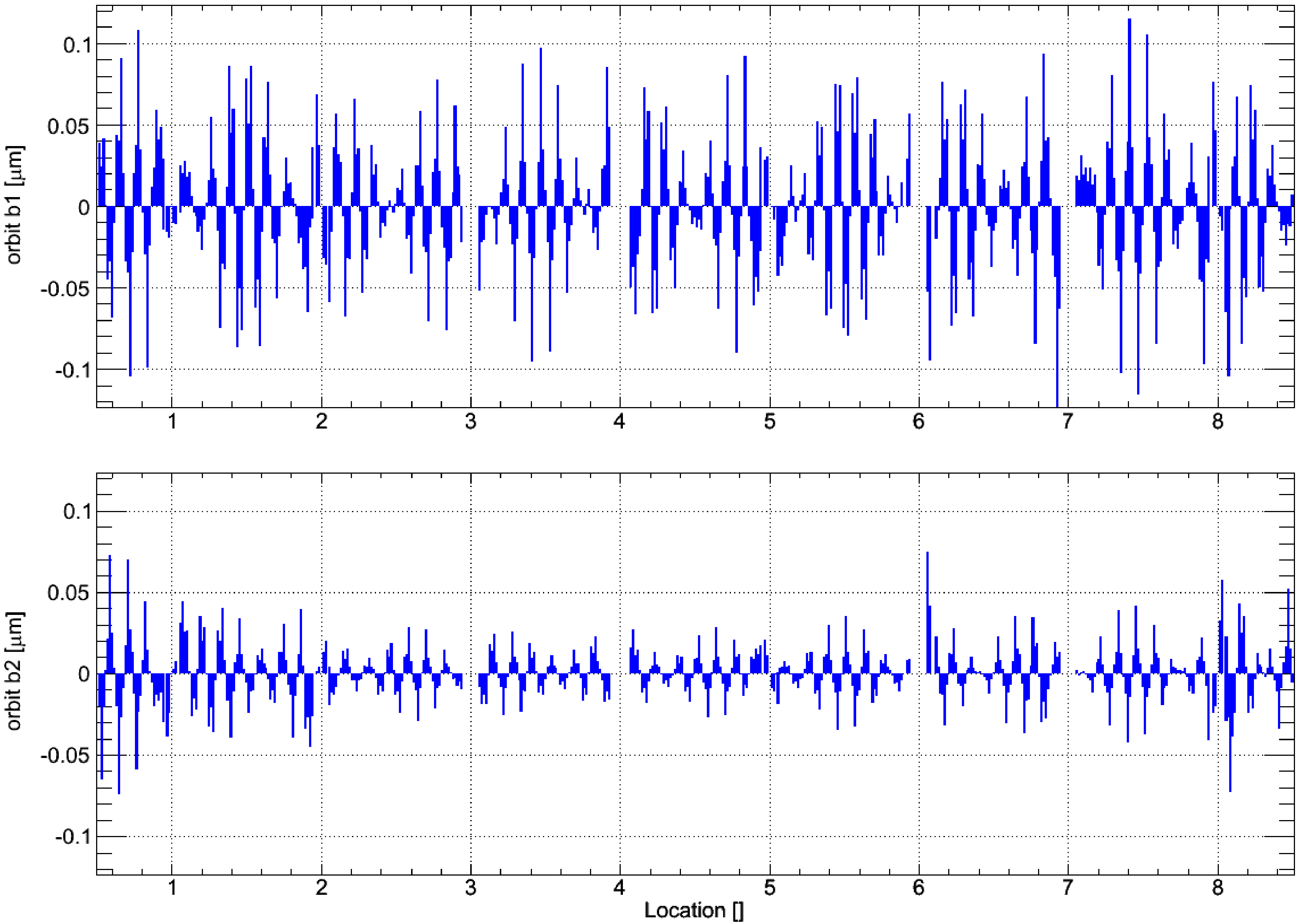
Space-Domain: SVD example
LHC eigenvalue spectrum

Eigenvalue spectra for vertical LHC response matrix using all BPMs and CODs:

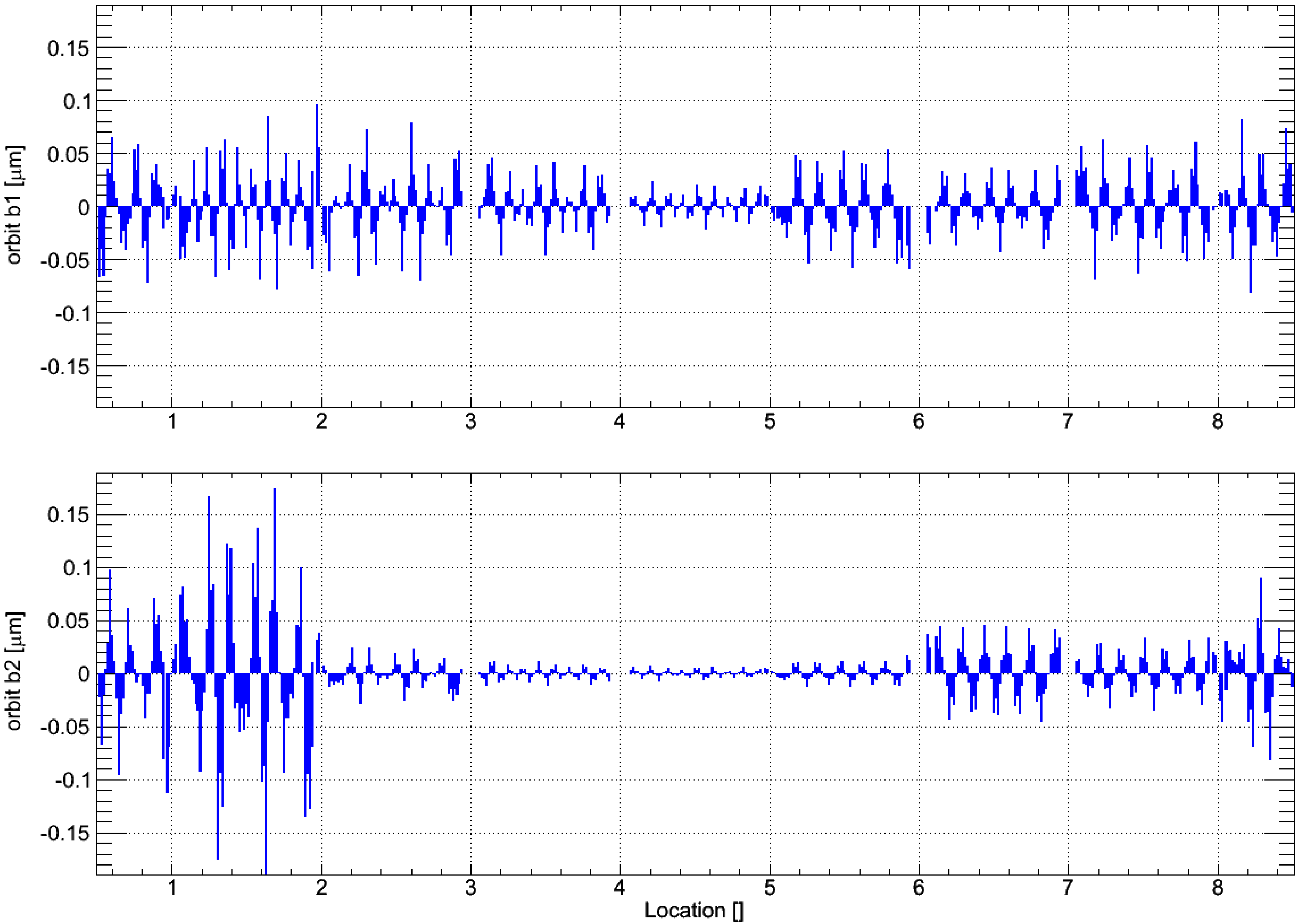


Space Domain:

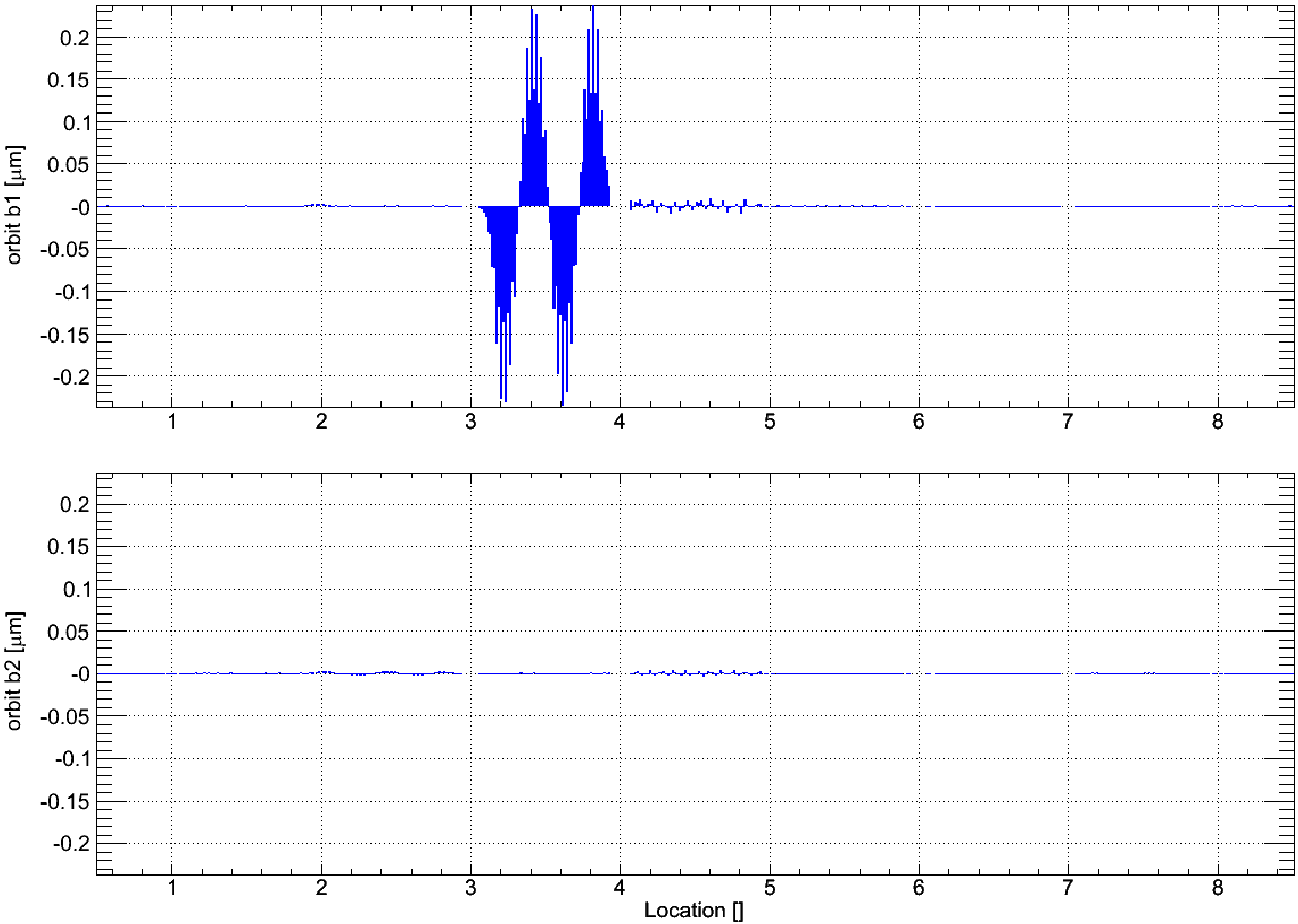
LHC BPM eigenvector #50 $\lambda_{50} = 6.69 \cdot 10^2$



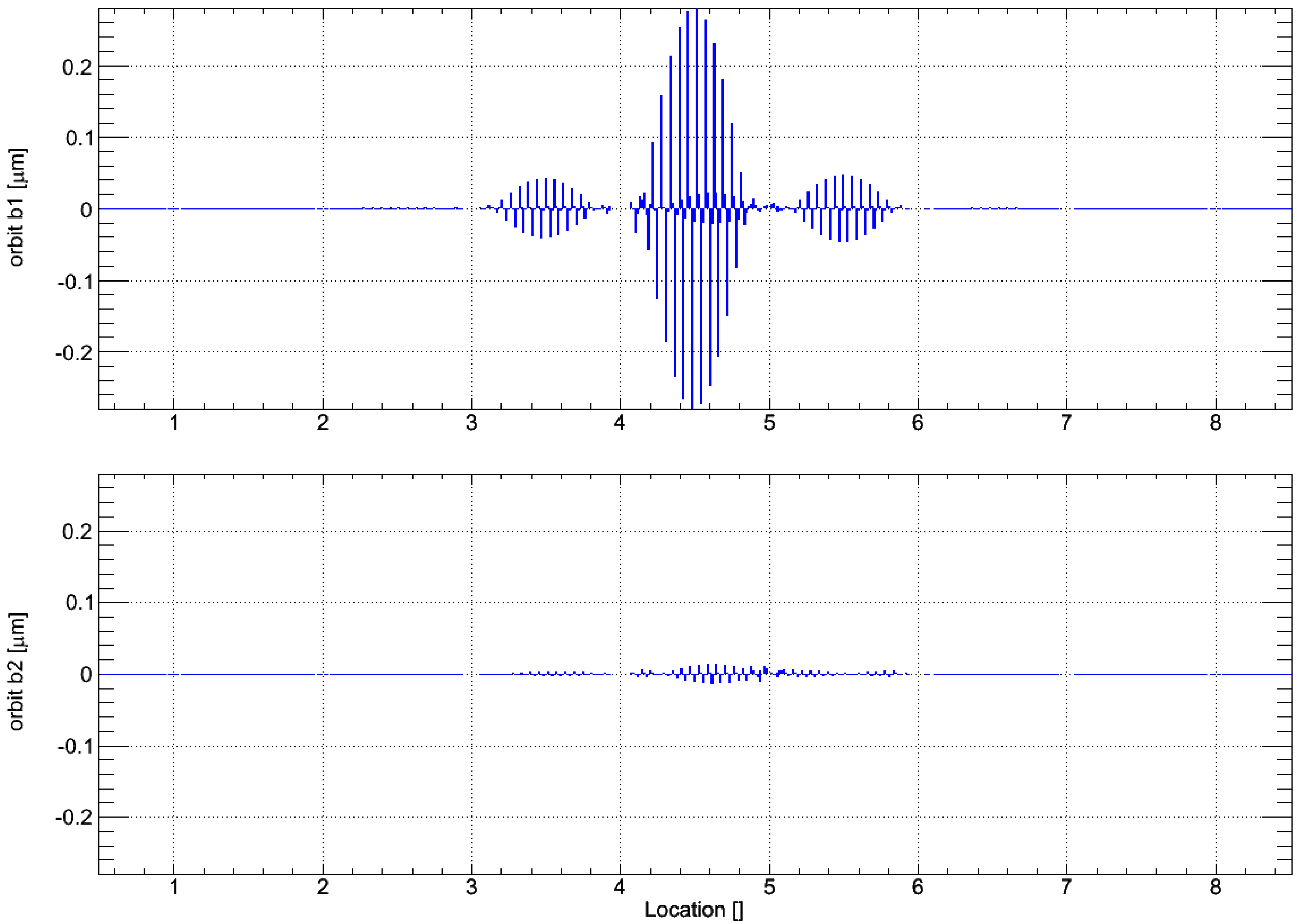
Space Domain:
LHC BPM eigenvector #100 λ_{100}= 3.38•10^2



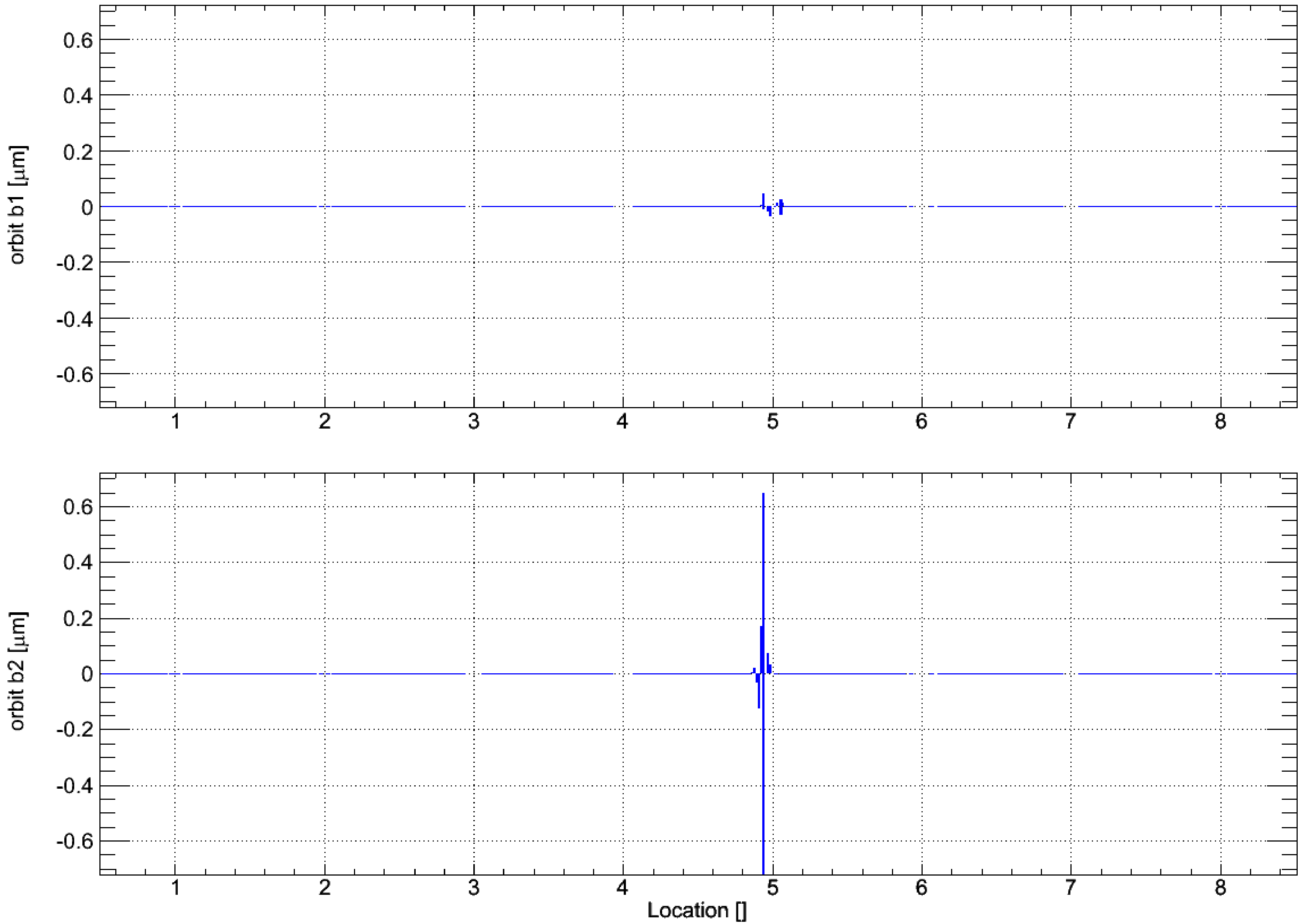
Space Domain:
LHC BPM eigenvector #291 $\lambda_{291} = 2.13 \cdot 10^2$



Space Domain:
LHC BPM eigenvector #449 λ_{449}= 8.17•10^1

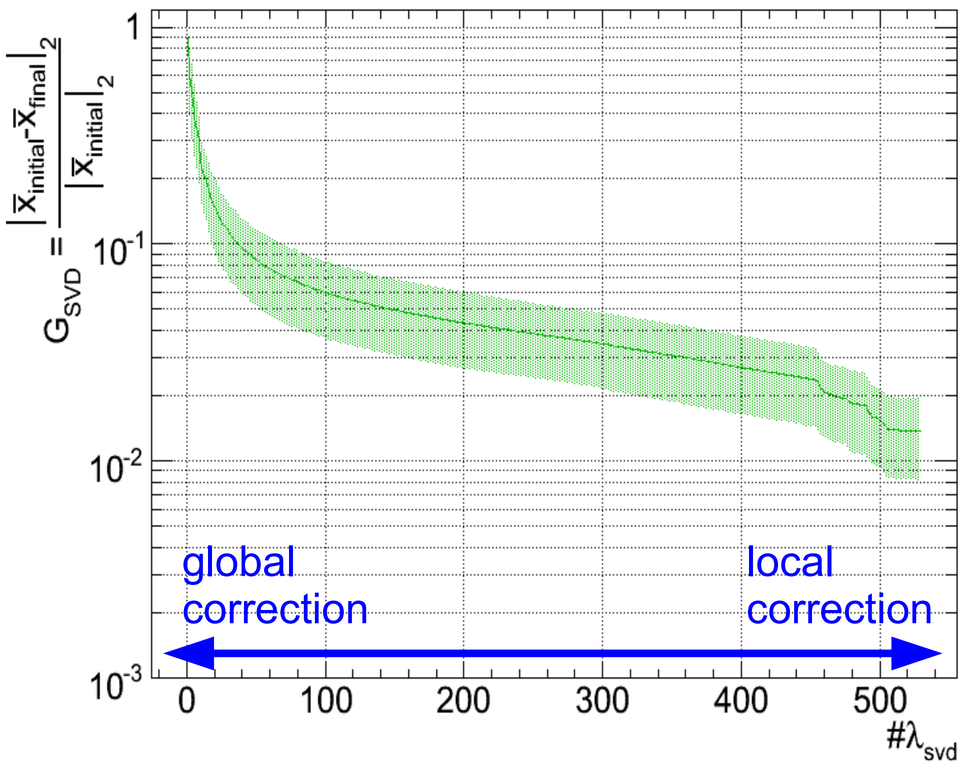


Space Domain:
LHC BPM eigenvector #521 $\lambda_{521} = 1.18 \cdot 10^0$

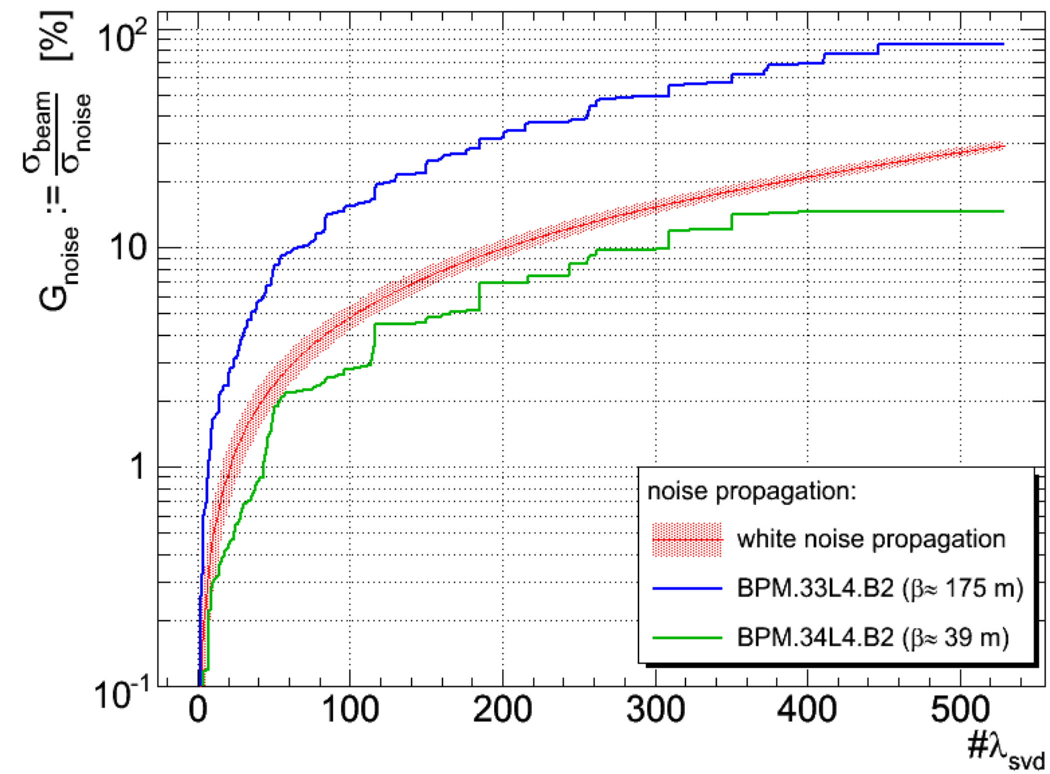


Space-Domain:
Orbit Attenuation Performance vs. Noise Propagation

Orbit attenuation



Sensitivity to BPM noise



Number of for the inversion used eigenvalues steers accuracy versus robustness of correction algorithm

Likewise applies for Tune, Chromaticity and Coupling correction

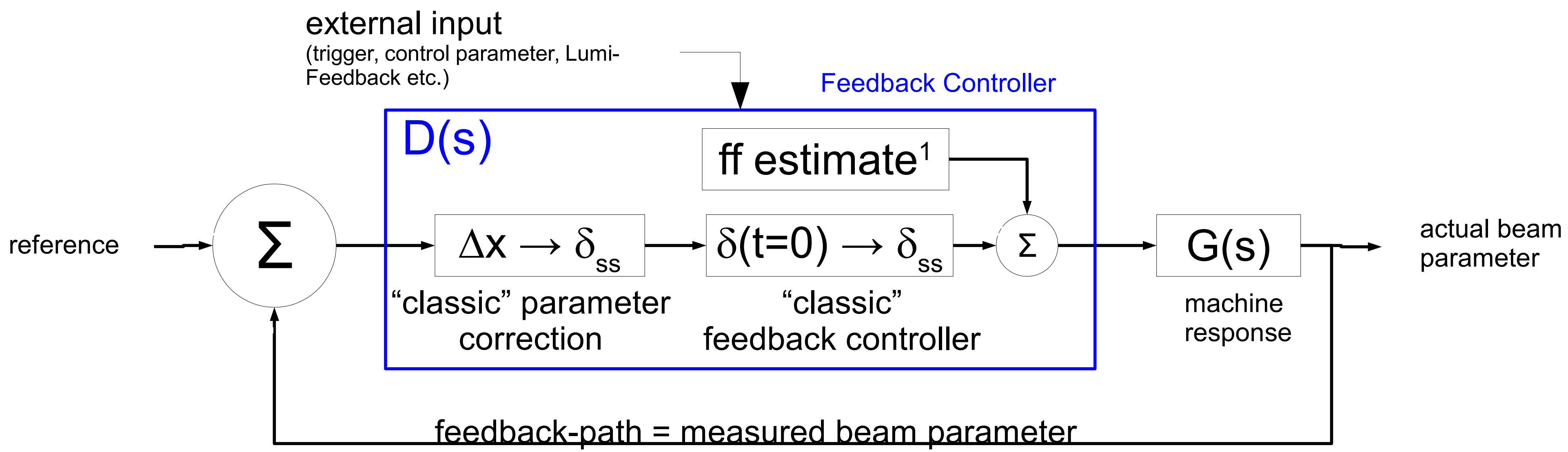
However: Only two out of '*n*' eigenvalues are non-singular

Multiple-Input-Multiple-Output (MIMO) Process Control

- 'Divide and Conquer' feedback controller design approach:
  1. Compute steady-state corrector settings $\vec{\delta}_{ss} = (\delta_{1,} \dots, \delta_n)$ based on measured parameter shift $\Delta x = (x_1, \dots, x_n)$ that will move the beam to its reference position for $t \rightarrow \infty$.
  2. Compute a $\vec{\delta}(t)$ that will enhance the transition $\vec{\delta}(t=0) \rightarrow \vec{\delta}_{ss}$
  3. Feed-forward: anticipate and add deflections $\vec{\delta}_{ff}$ to compensate changes of well known and properly described sources

space domain

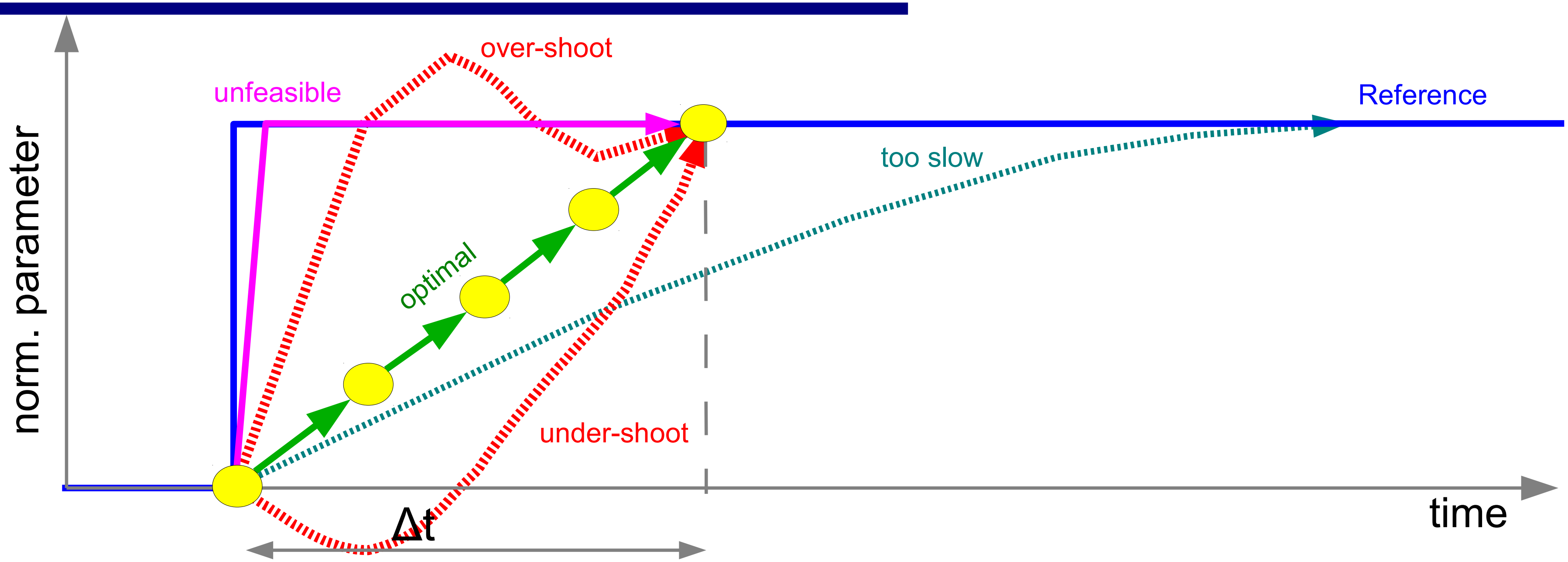
time domain



(N.B. here G(s) contains the process and monitor response function)

Time-Domain: Optimal Controller Design

Youla's affine parameterisation I/V – Cartoon



Optimal control [or design] ...

"... deals with the problem of finding a control law for a given system such that a given optimality criterion is achieved. A control problem includes a cost functional that is a function of state and control variables."

Common criteria: closed loop stability, minimum bandwidth, minimisation of action integral, power dissipation, ...

classic closed loop: $$T_0(s) = \frac{D(s)G(s)}{1+D(s)G(s)}$$ → "this tells me???"

Time-Domain: Optimal Controller Design
Youla's affine parameterisation II/V

- Youla's affine parameterisation for stable plants[1] - showed that all stable closed loop controllers D(s) can be written as:

$$\boxed{D(s)=\frac{Q(s)}{1-Q(s)G(s)}} \tag{1}$$

- Simplifies the form of the system transfer $T_0(s)$ and sensitivity function $S_0(s)$:

$$\begin{aligned} T_0(s) &= Q(s)G(s) && (2)\\ S_0(s) &= 1-Q(s)G(s) = 1-T_0(s) && (3) \end{aligned}$$

- Use following common *ansatz* for solving (1): $Q(s)=F_Q(s)G^i(s)$ (4)
- In case of a "perfect" inverse response function (no unstable poles/zeros) (2) (3) yields simply:

$$\begin{aligned} T'_0(s) &= F_Q(s)\\ S'_0(s) &= 1-F_Q(s) \end{aligned}$$

- → effective closed loop response can be deduced by construction of $F_Q(s)$

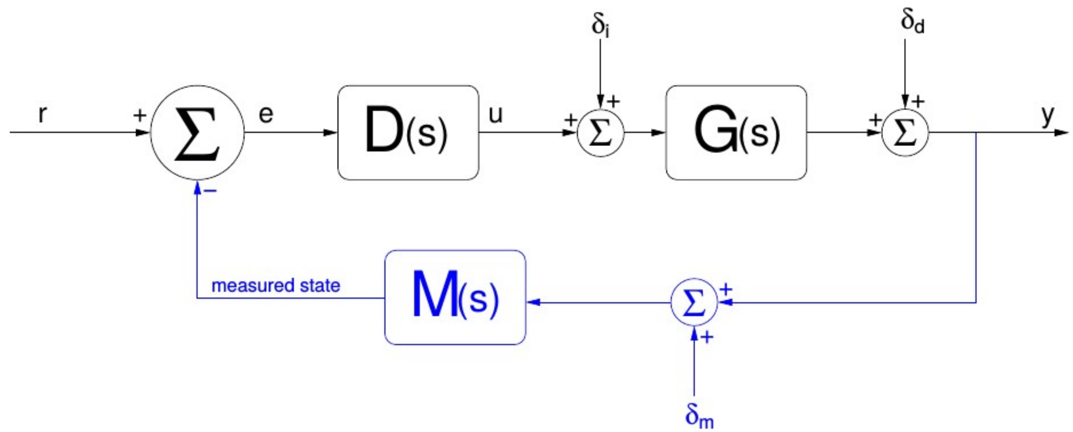
[1]D. C. Youla et al., *"Modern Wiener-Hopf Design of Optimal Controllers"*, IEEE Trans. on Automatic Control,1976, vol. 21-1,pp. 3-13 & 319-338

Time-Domain: Optimal Controller Design

Youla's affine parameterisation III/V

$$D(s)=\frac{Q(s)}{1-Q(s)G(s)} \qquad Q(s)=F_Q(s)G^i(s) \qquad (1)$$

- Using Youla's parameterisation: “design closed loop in a open loop style”



  – insert (1) into sensitivity functions defined earlier:

transfer function:	$T_0(s) := \frac{y}{r}$	$= \frac{D(s)G(s)}{1+D(s)G(s)}$	$= Q(S)G(S)$	$\overset{here:}{=}$	$F_Q(S)$
disturbance rejection:	$S_{d0}(s) := \frac{y}{\delta_d}$	$= \frac{1}{1+D(s)G(s)}$	$= 1-Q(s)G(S)$	$=$	$1-F_Q(S)$
input sensitivity:	$S_{i0}(s) := \frac{y}{\delta_i}$	$= \frac{G(S)}{1+D(s)G(s)}$	$= (1-Q(s)G(S))G(s)$	$=$	$(1-F_Q(S))G(S)$
control sensitivity:	$S_{u0}(s) := \frac{y}{\delta_u}$	$= \frac{D(S)}{1+D(s)G(s)}$	$= Q(s)$	$=$	$F_Q(S)G^i(S)$

- Some constraints on G'(s):
  – must not include zero-pole cancellation violating causality or other known time-domain limitations, e.g.
    • delay compensation (→ dealt with differently)
    • sampling zero cancellation, rate-limiter, saturation, …

Need an example?

Time-Domain: Optimal Controller Design

Youla's affine parameterisation IV/V

$$D(s)=\frac{Q(s)}{1-Q(s)G(s)}$$
$$Q(s)=F_Q(s)G^i(s) \quad (1)$$

- Example: first order system

$$G(s)=\frac{K_0}{\tau s+1} \quad \text{with } \tau \text{ being the circuit time constant} \quad (2)$$

- Using for example the following ansatz:

$$Q(s)=F_Q(s)G^i(s)=\frac{1}{\alpha s+1}\cdot\frac{\tau s+1}{K_0} \quad (3)$$

  - Response/optimality can be directly deduced by construction of $F_Q(s)$
  - $G^i(s)$ is the pseudo-inverse of the nominal plant G(s)

$$\rightarrow T_0(s)=\frac{1}{\alpha s+1}$$

- (1)+(2)+(3) yields the following controller which happens to be a PI controller:

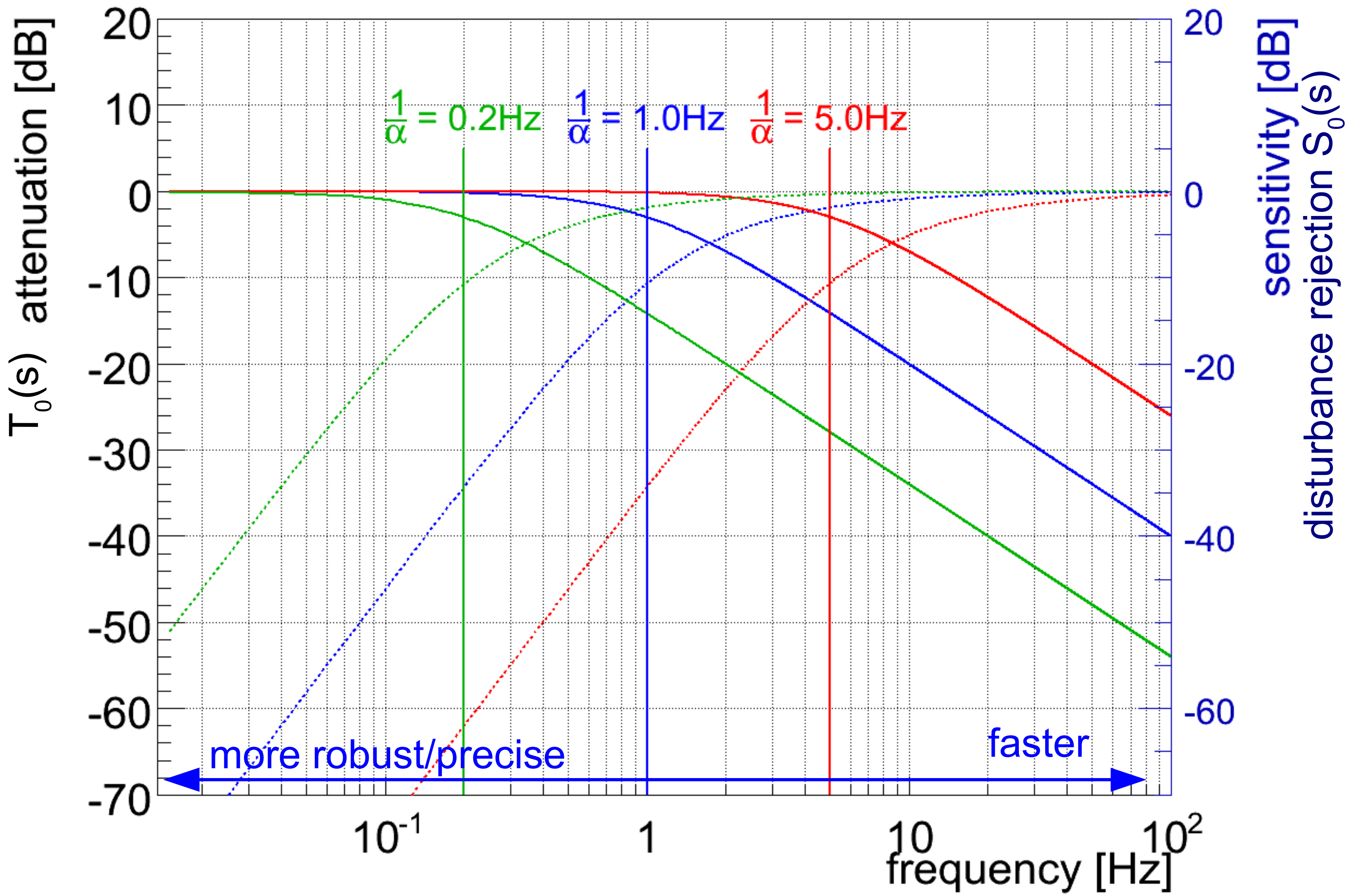
$$\boxed{D(s)=K_P+K_i\frac{1}{s} \quad \mathit{with} \quad K_p=K_0\frac{\tau}{\alpha} \wedge K_i=K_0\frac{1}{\alpha}}$$

Time-Domain: Optimal Controller Design

Example: Tune-PLL Closed Loop Controller - Bandwidth

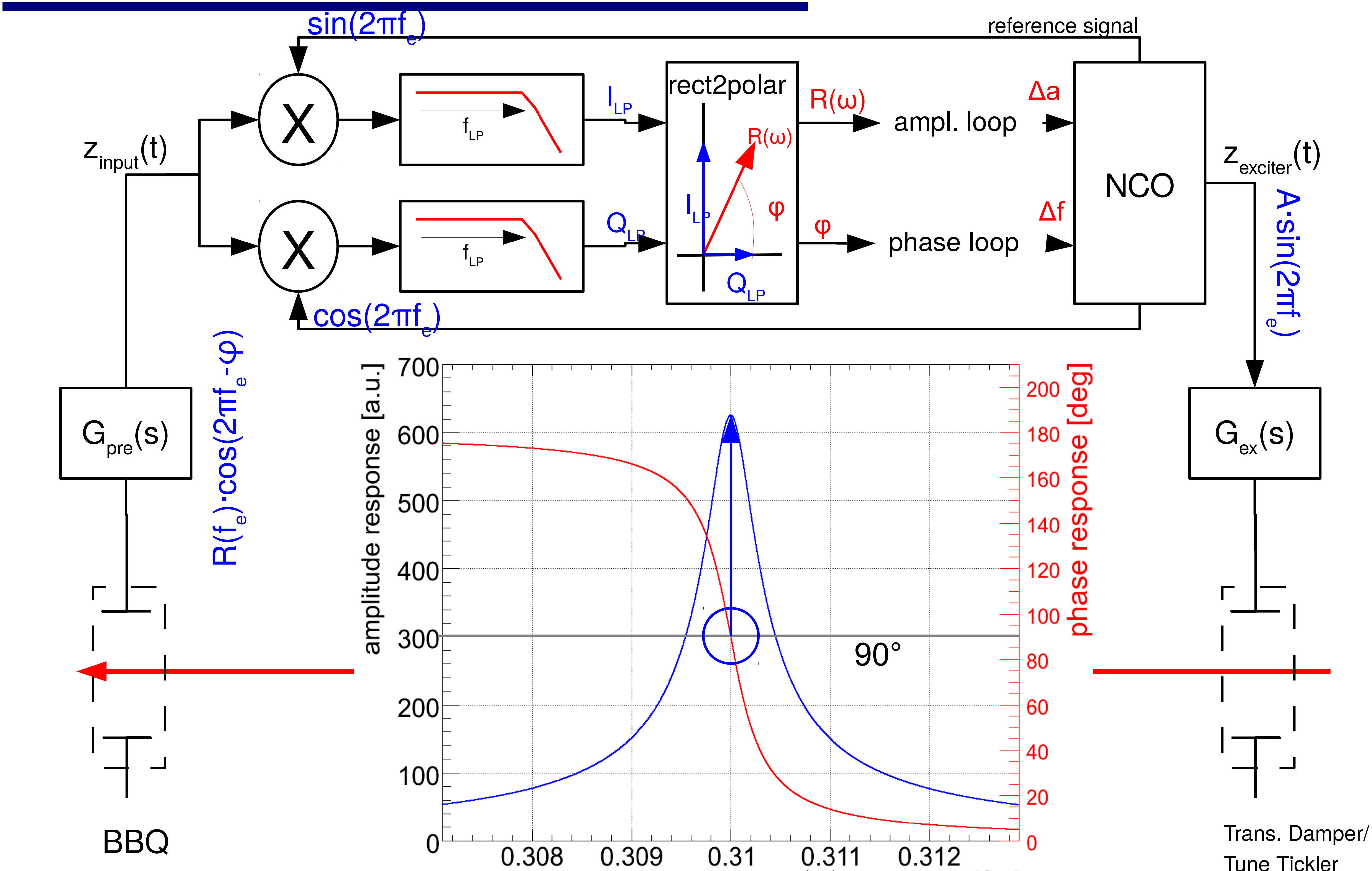
$$D(s)=\frac{Q(s)}{1-Q(s)G(s)}$$

- $\alpha > \tau \ldots \infty$ facilitates the trade-off between speed and robustness
  - operator has to deal with one parameter
    → enables simple adaptive gain-scheduling based on the operational scenario!



Time-Domain: Optimal Controller Design

Example: Simplified Phase-Locked-Loop Scheme (Q-Loop, cavity-Loop, Fast-FB, ...)

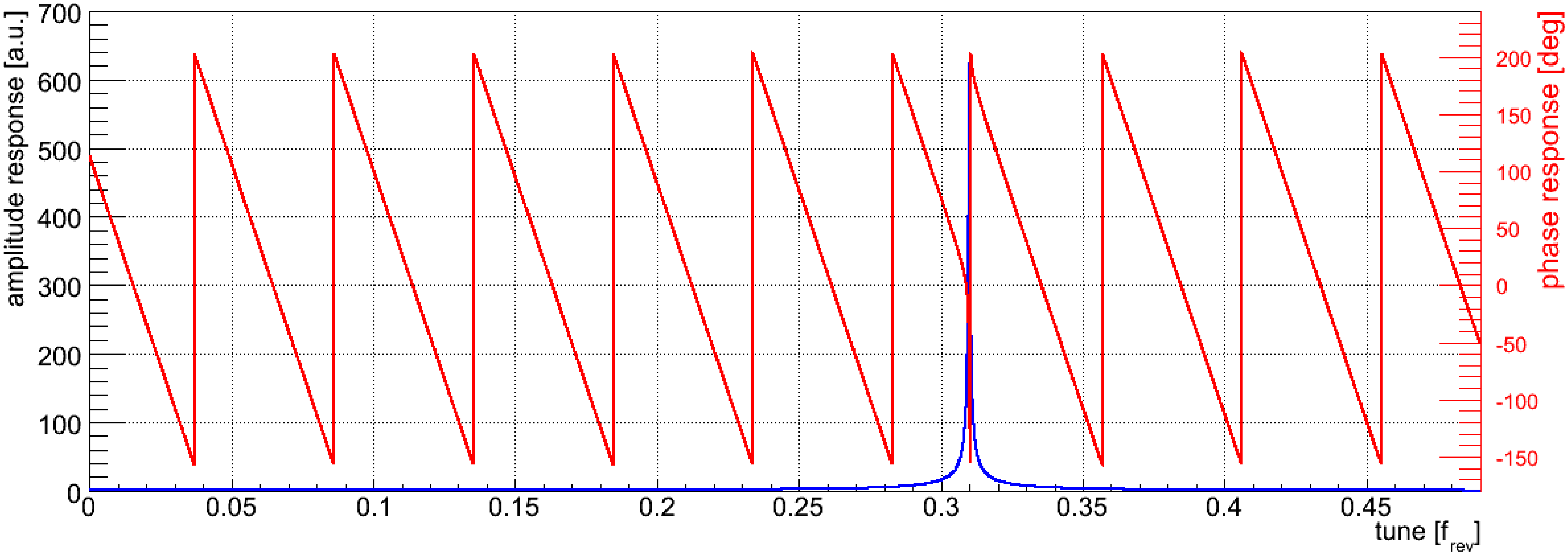


$$G_{Beam}(\omega) = R(\omega) \cdot e^{i\varphi(\omega)}$$

Time-Domain: Optimal Controller Design

Example: PLL Setup – Step I (HW lag compensation)

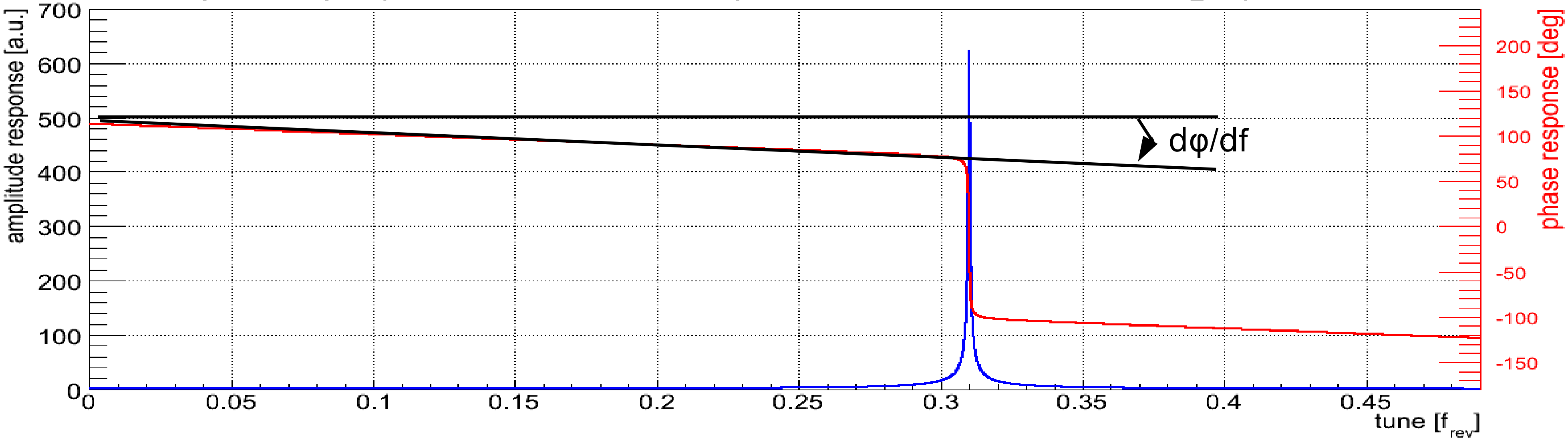
- BTF functions do not always look always as pretty as reports suggests or claim – an insider view on the real story:
- BTF and compensation consists of the adjustment of four parameters, preferably with stable beam condition ('chicken-egg' problem)
  - 1st step: verify necessary excitation amplitude and plane mapping (obvious?)
  - 2nd step: verify long sample delay (once per installation, constant)
    - full range BTF and count ±π wrap-around → number of delayed samples



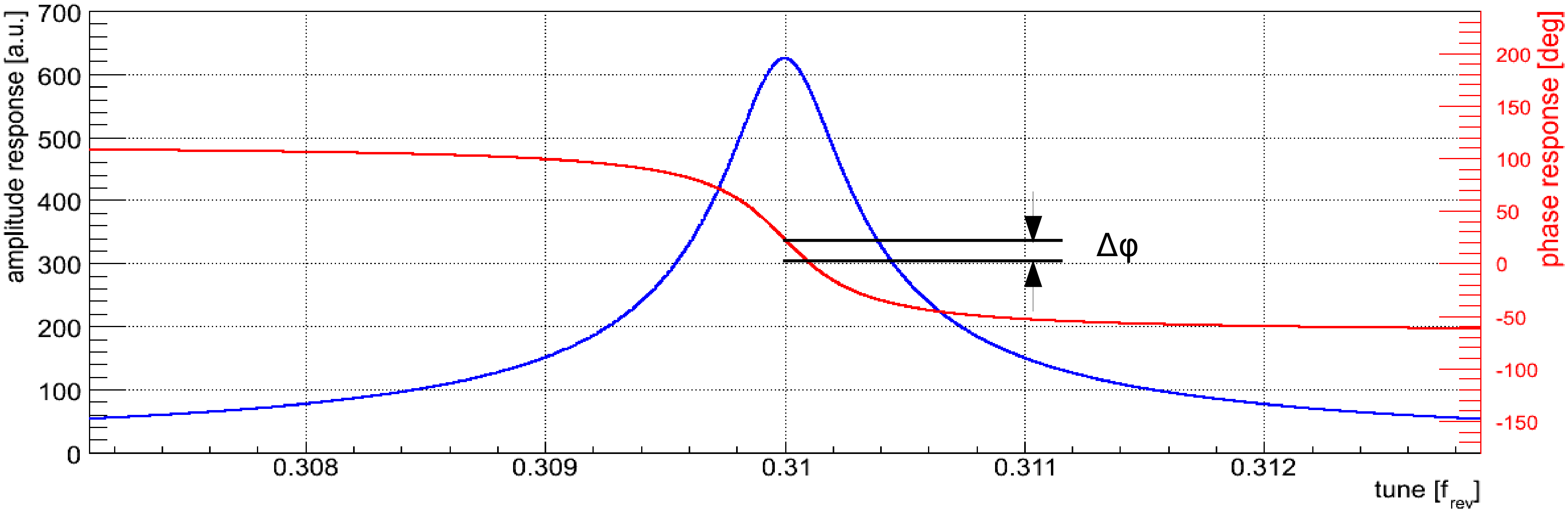
Time-Domain: Optimal Controller Design

Example: PLL Setup – Step II (beam phase compensation)

Measure dφ/df slope (~ front-end non-lin. phase and kicker cable length)



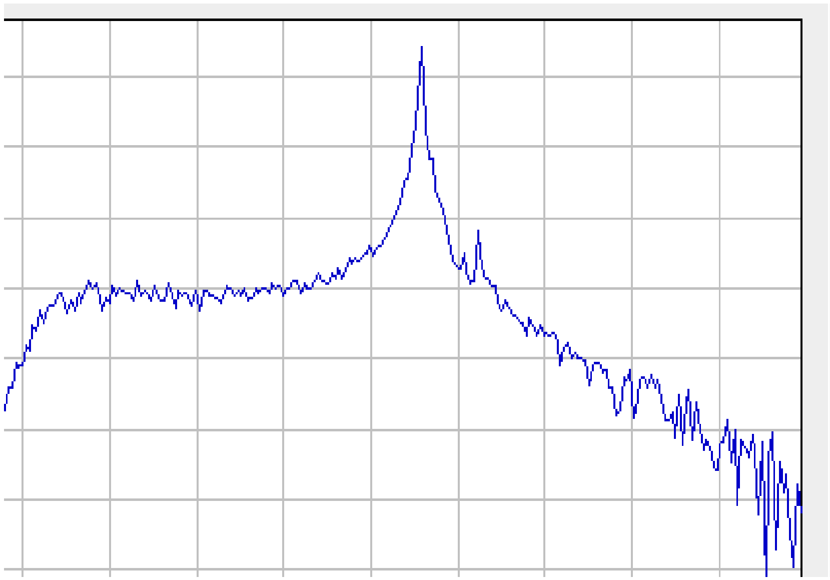
Adjustments of the locking phase (tune-peak – phase matching)



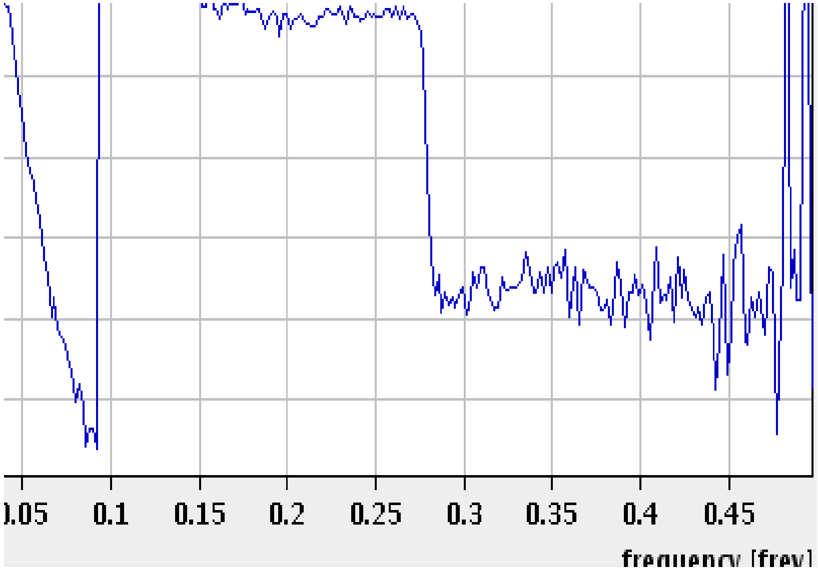
Time-Domain: Optimal Controller Design

Example: PLL Setup – Step III → Ready for Q/C⁻/Q' Tracking

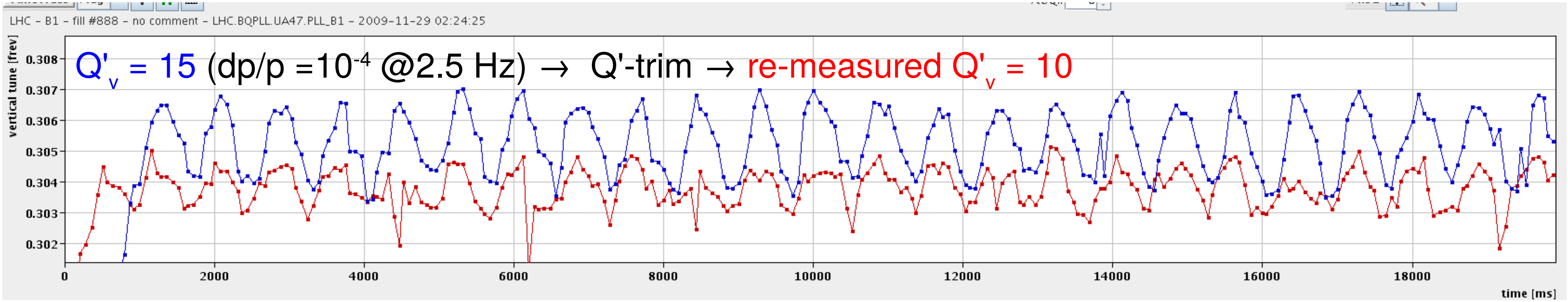
beam magnitude response:



beam phase response :

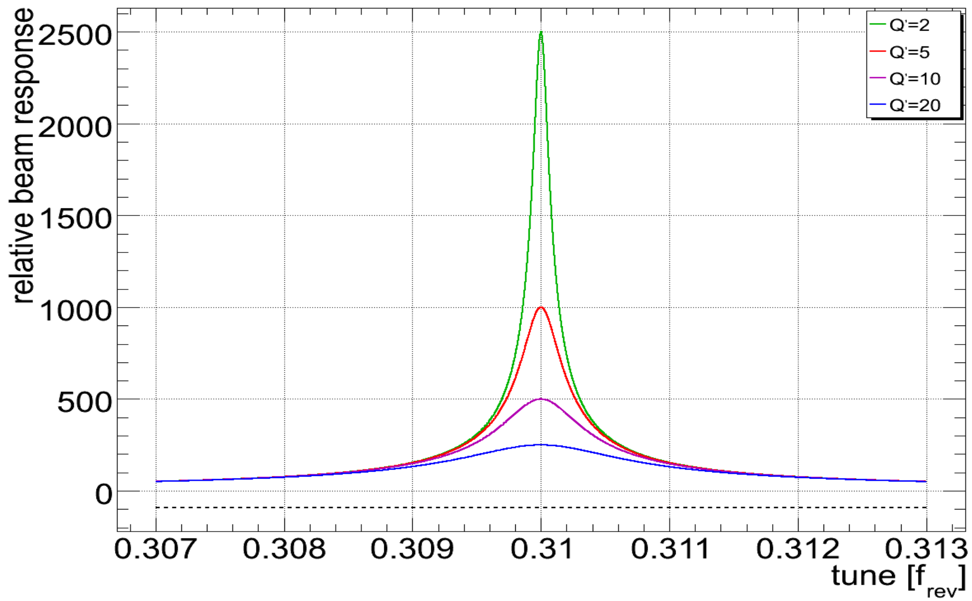


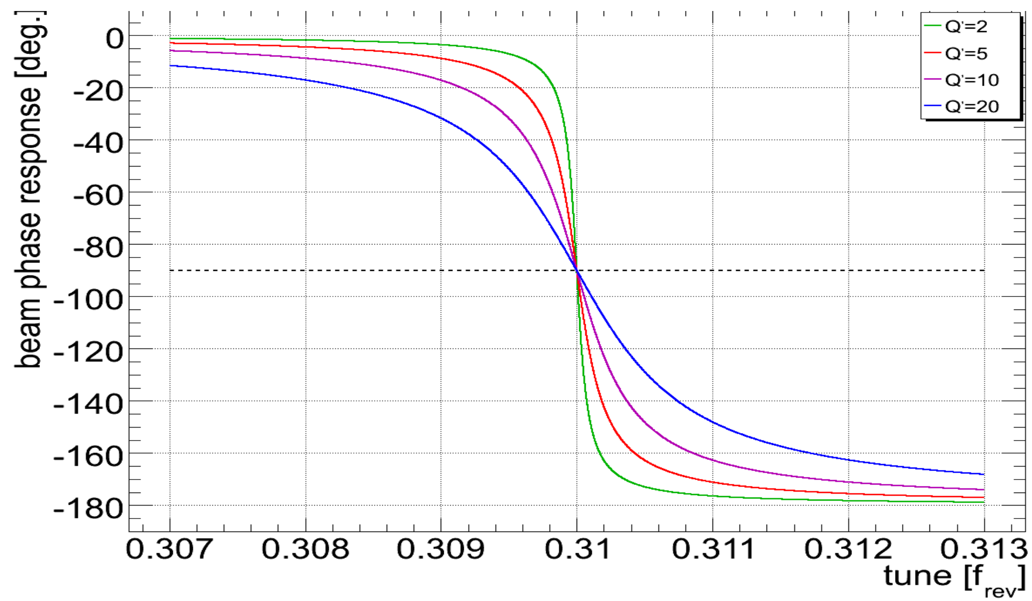
- PLL tracking in action:



Time-Domain: Optimal Controller Design

Example: Tune-PLL dependence on Q' & dynamic Gain-Control





- Beam response: open loop gain K_0 ~ phase response slope
- Common[4-6,19] (classic) PLL loop design: K_0 = const. & filter bandwidth = 1/τ
  → PLL low-pass:

$$G(s) = \frac{K_0}{\tau s + 1}$$

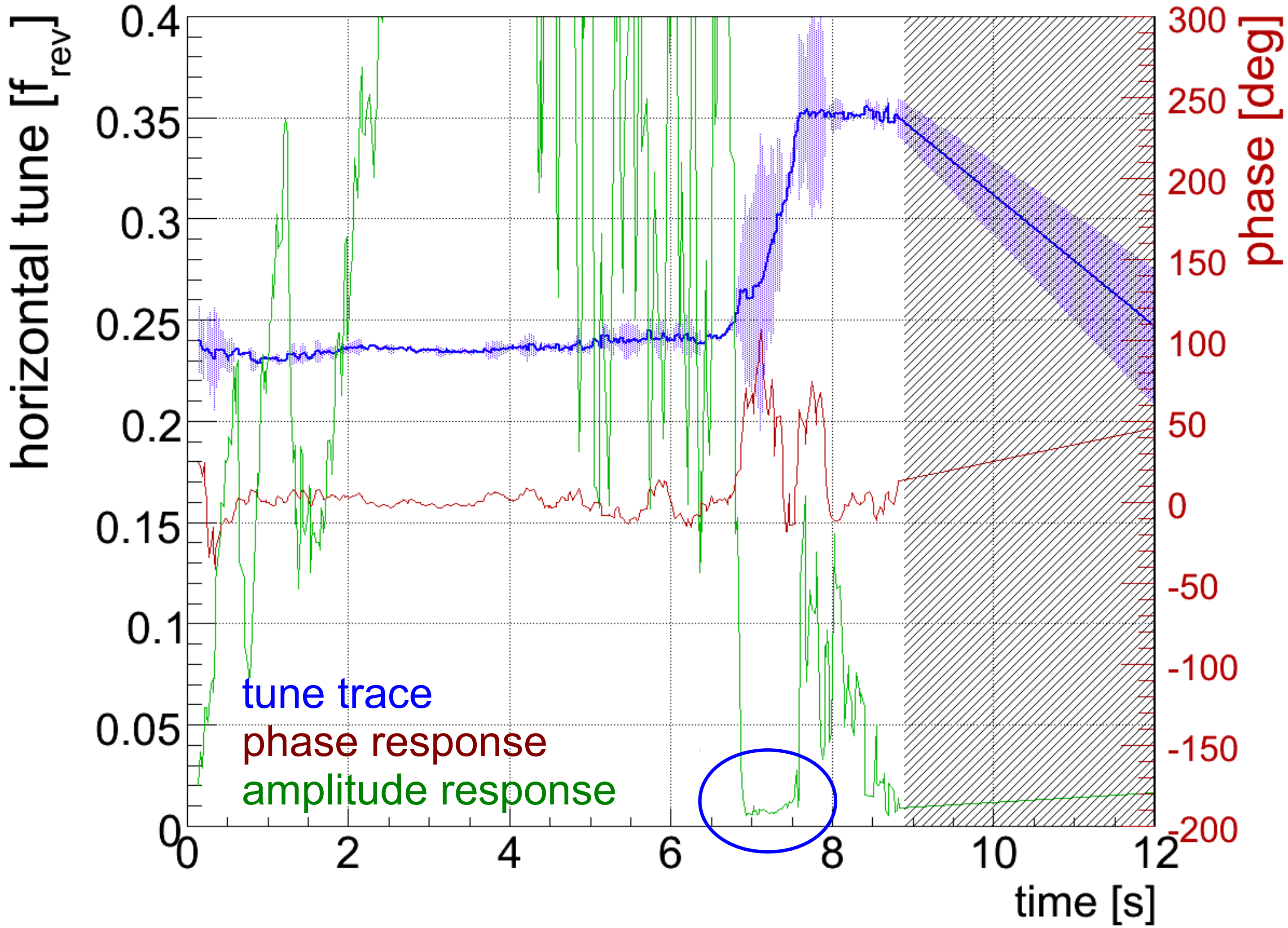
> Note:
> K_0 const. for |Δφ|≤ 60° (linear. regime)
> K_0 depends on Q' (non-linear. regime)

- Optimal tune PLL gain parameters depend on chromaticity[20,21]
  - Optimal PI for high Q' ↔ sensitivity to noise (unstable loop) for low Q'
  - Optimal PI for small Q' ↔ slow tracking speed for large Q'

Time-Domain: Optimal Controller Design

Example: CERN-SPS PLL Tune Tracking – fast tracking

Two domains of tracking, either slow and very precise (low loop bandwidth) or fast:

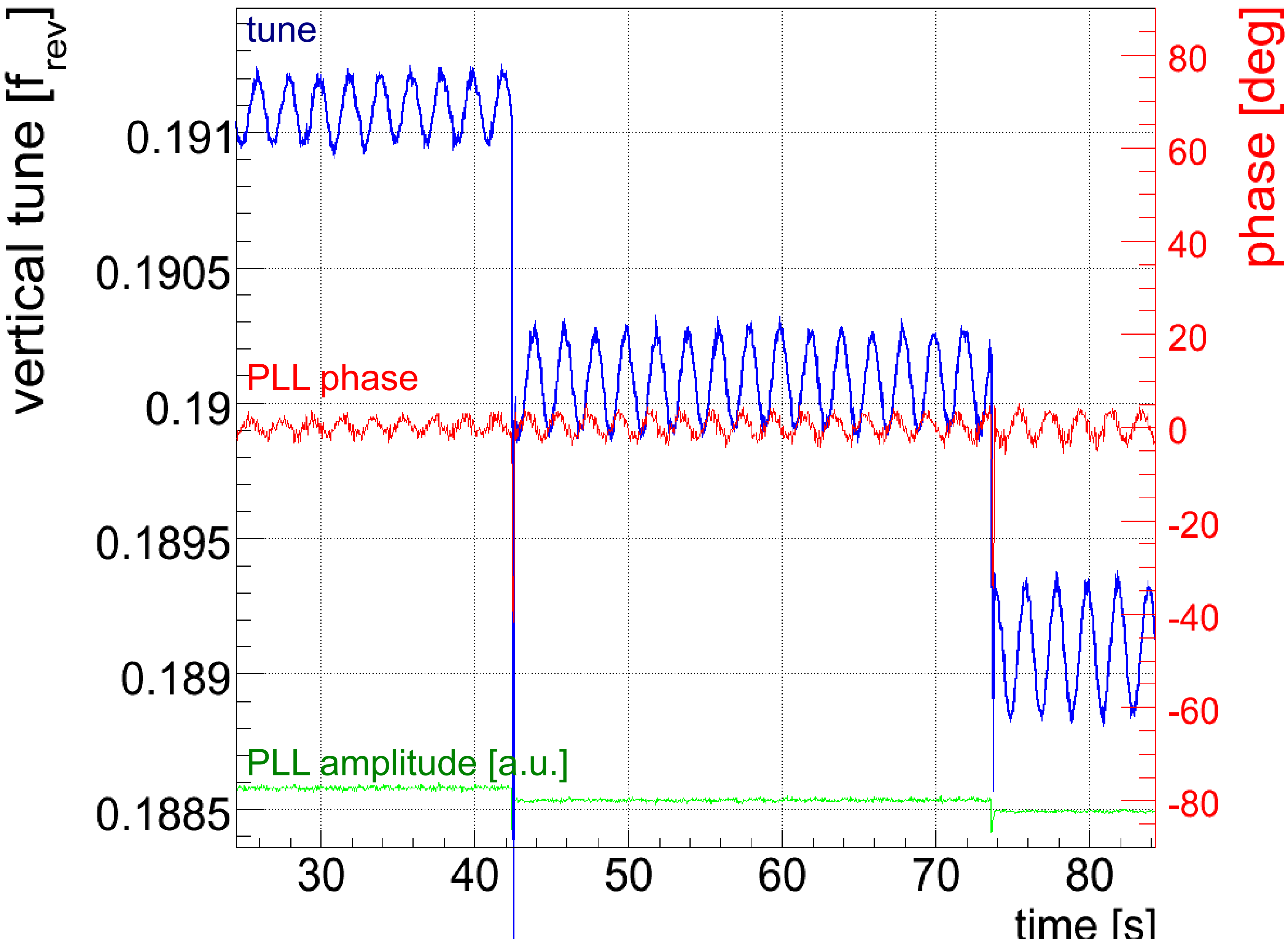


Phase error and non-vanishing amplitude indicates lock

here: $\Delta Q/\Delta t|_{max} \approx 0.3$ within 300 ms $\quad f_{rev} \approx 43$ kHz

Time-Domain: Optimal Controller Design

Example: CERN-SPS PLL Tune Tracking – precise tracking (Q', $\Delta p/p \approx 1.85 \cdot 10^{-5}$)



here: PLL-Tune resolution: $\Delta f_{res} \approx 10^{-6}$

Time-Domain: Optimal Controller Design
Youla's affine parameterisation V/V

$$D(s)=\frac{Q(s)}{1-Q(s)G(s)}\qquad Q(s)=F_Q(s)G^i(s) \tag{1}$$

- 2nd Example: classic 2nd order process:

$$G(s)=\frac{K_0\omega_0^2}{s^2+2\zeta\omega_0\ s+\omega_0^2}$$

K_0: open loop gain, ω_0: characteristic frequency
ζ_0: attenuation

- Using standard ansatz:

$$Q(s)=F_Q(s)\cdot G^i(s)=\frac{\omega_{cl}^2}{s^2+2\zeta_{cl}\omega_{cl}\ s+\omega_{cl}^2}\cdot G^i(s)$$

- yields classic PID controller (optimal gains):

practical real-life engineering: additional pole to suppress high-frequency noise that otherwise would be amplified by 's' and unnecessarily saturate u

$$D(s)\ =\ K_p+K_i\cdot\frac{1}{s}+K_d\cdot\frac{s}{\boxed{\tau_d s+1}}$$

with:

$$K_p=\frac{4\zeta_{cl}\zeta_0\omega_0\omega_{cl}-\omega_0^2}{4K_0\zeta_{cl}^2}\qquad K_i=\frac{\omega_0^2\omega_{cl}}{2K_0\zeta_{cl}}$$

$$K_d=\frac{4\zeta_{cl}^2\omega_{cl}^2-4\zeta_0\omega_0\zeta_{cl}+\omega_0^2}{8K_0\zeta_{cl}^3\omega_{cl}}\qquad \tau_d=\frac{1}{2\zeta_{cl}\omega_{cl}}$$

– further simplification: require critical damping $\rightarrow$ ζ_{cl}:=1
  - ω_{cl} ~ 'open loop bandwidth' is the remaining free parameter

Time-Domain: Optimal Controller Design

Youla's affine parameterisation – MIMO Controller

$$D(s)=\frac{Q(s)}{1-Q(s)G(s)}$$
$$Q(s)=F_Q(s)G^i(s)$$
SISO

- Similar to the SISO case, Youla's parameterisation[1] is also applicable for MIMO systems – all stable closed loop controllers **D**(s) (mxn matrix) can be written as:

$$\boxed{\boldsymbol{D}(s)=\boldsymbol{Q}(s)\left[\boldsymbol{I}-\boldsymbol{G}_0(s)\boldsymbol{Q}(s)\right]^{-1}}$$

- Simplifies the form of the system transfer $T_0(s)$ and sensitivity function $S_0(s)$:

$$\begin{aligned}\boldsymbol{T}_0(s) &= \boldsymbol{G}_0(s)\boldsymbol{Q}(s)\\ \boldsymbol{S}_0(s) &= \boldsymbol{I}-\boldsymbol{G}_0(s)\boldsymbol{Q}(s) = \boldsymbol{I}-\boldsymbol{T}_0(s)\end{aligned} \qquad (1)$$

- If required that '$\mathbf{T}_0$(s) = **I**' - use similar *ansatz* to SISO case shown earlier
  – again, use SVD for the pseudo-inverse response function solving (1):

$$\boldsymbol{Q}(s)=\underbrace{\hat{\boldsymbol{G}}^{-1}(s)}_{pseudo-inverse}\;\boldsymbol{F}_{\boldsymbol{Q}}(s)$$

→ robust inversion is the core issue in control system design

[1]D. C. Youla et al., *“Modern Wiener-Hopf Design of Optimal Controllers”*,
IEEE Trans. on Automatic Control,1976, vol. 21-1,pp. 3-13 & 319-338

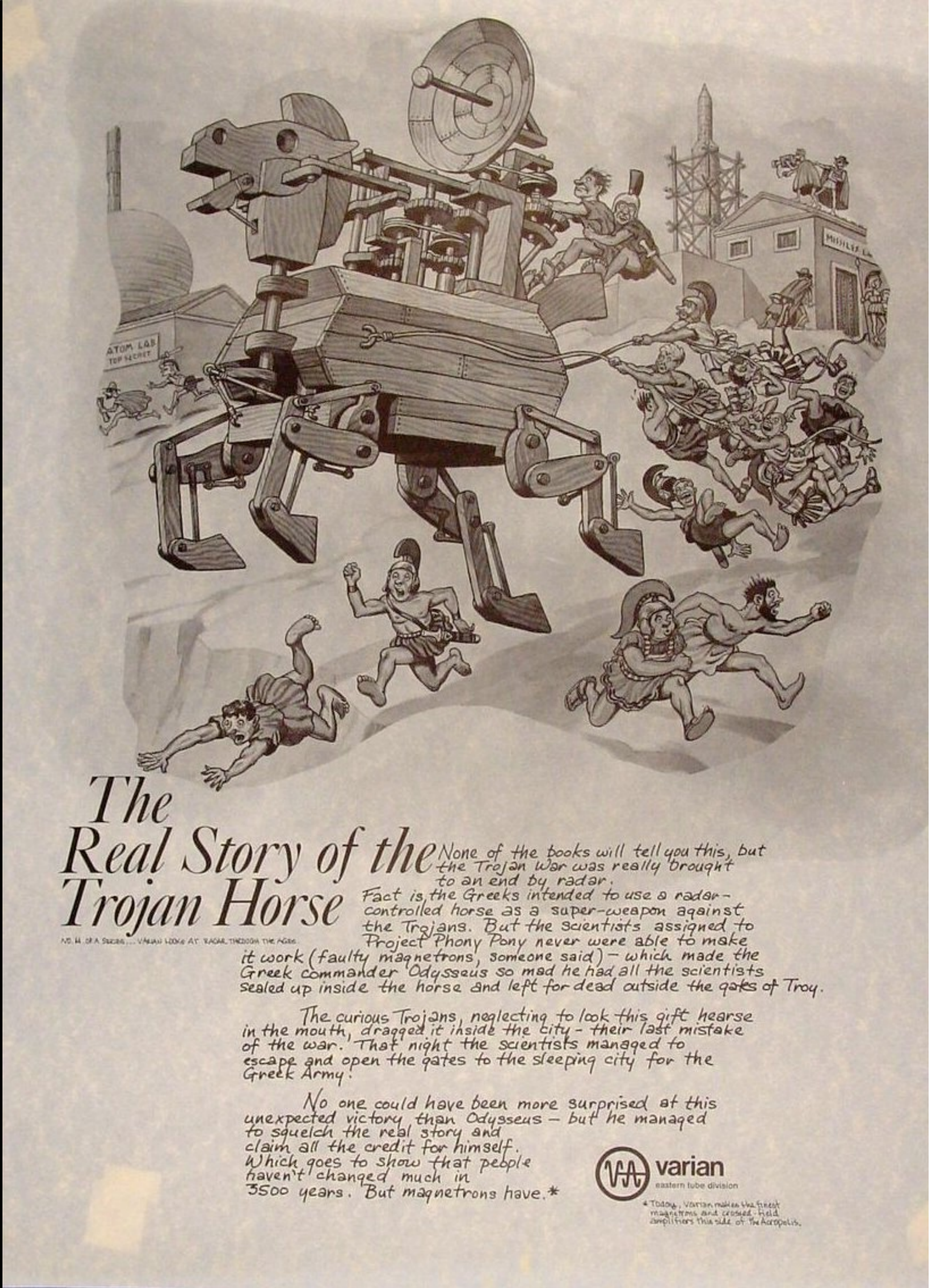
Intermediate Summary II

- Use of imperfect (design) beam response for SVD based FB systems:
  - does not affect the precision of the correction but reduces rather the effective bandwidth → favours higher feedback sampling frequencies

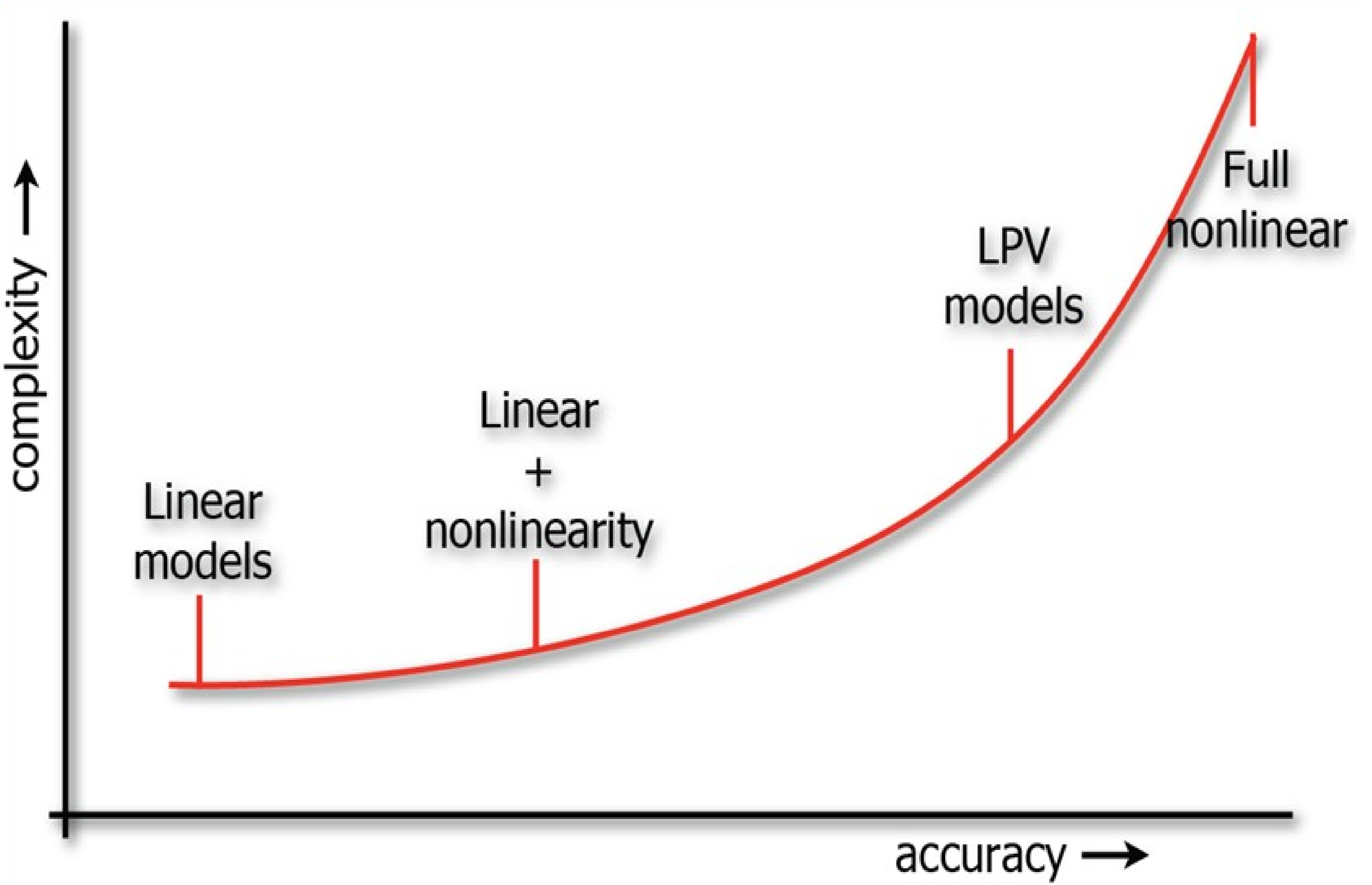
- Youla's affine parameterisation facilitates optimal adaptive (non-)linear control
  - enables gain-scheduling based on operational scenario

- Youla's parameterisation applies equally for SISO & MIMO systems, however – in the context of accelerators – it's suggested to split the problem into 'space-' and 'time-domain':
  - separates specific accelerator physics from specific control theory (N.B. different/complementary control-room-level expertise)
  - allows separate testing (accelerator physics vs. dynamics of actuators)
  - Space-domain:
    - maintains physical meaning of the individual control variables
    - often level of synchronisation required to minimise inter-loop coupling
    - MIMO control → basically relying on inversion of response matrices → SVD
      - numerically robust (often forgotten from a controls only perspective)
  - Time-domain:
    - more transparent optimal 'Wiener', 'Kalman', 'Youla-Kucera-Klein'-based filtering

Break!


ATOM LAB
The Real Story of the Trojan Horse
None of the books will tell you this, but the Trojan War was really brought to an end by radar.
Fact is, the Greeks intended to use a radar-controlled horse as a super-weapon against the Trojans. But the scientists assigned to Project Phony Pony never were able to make it work (faulty magnetrons, someone said) – which made the Greek commander Odysseus so mad he had all the scientists sealed up inside the horse and left for dead outside the gates of Troy.
The curious Trojans, neglecting to look this gift hearse in the mouth, dragged it inside the city – their last mistake of the war. That night the scientists managed to escape and open the gates to the sleeping city for the Greek Army.
No one could have been more surprised at this unexpected victory than Odysseus – but he managed to squelch the real story and claim all the credit for himself. Which goes to show that people haven't changed much in 3500 years. But magnetrons have.*
varian
eastern tube division
* Today, Varian makes the finest magnetrons and crossed-field amplifiers this side of the Acropolis.

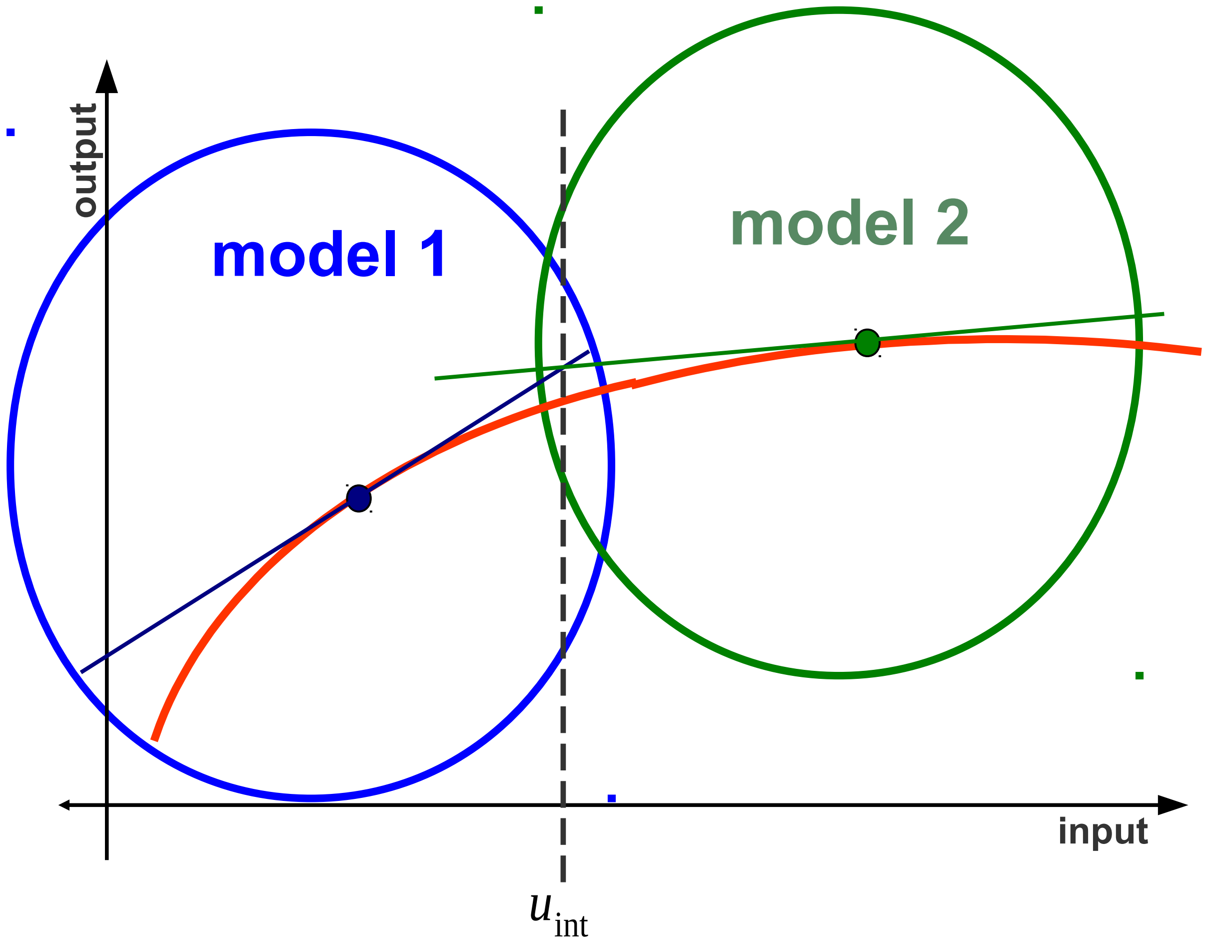
Non-Linearities



Time-Domain: Non-Linearities

Many systems are non-linear across wide operation range

- Option #1 – linearise around given working point and continue linear design



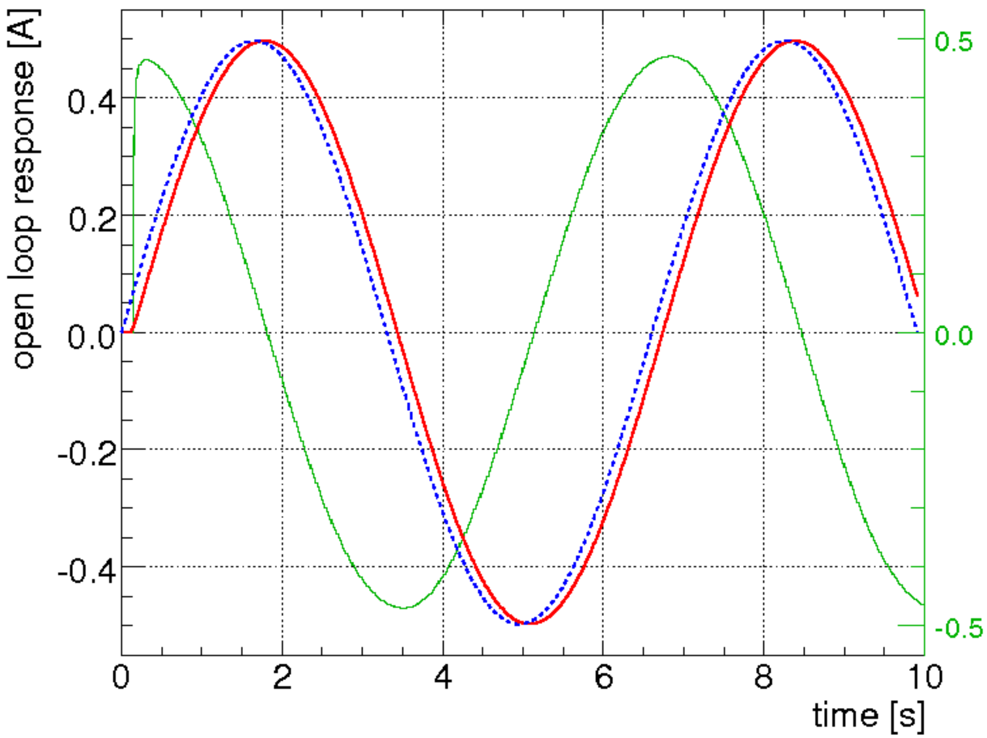
  - Common gain scheduling: use 'model 1' for tuning/set-up → shift to 'model 2' for routine operation
  - … but does not always work when the working point/operational range is a priori unknown

Time-Domain: Non-Linearities I/IV

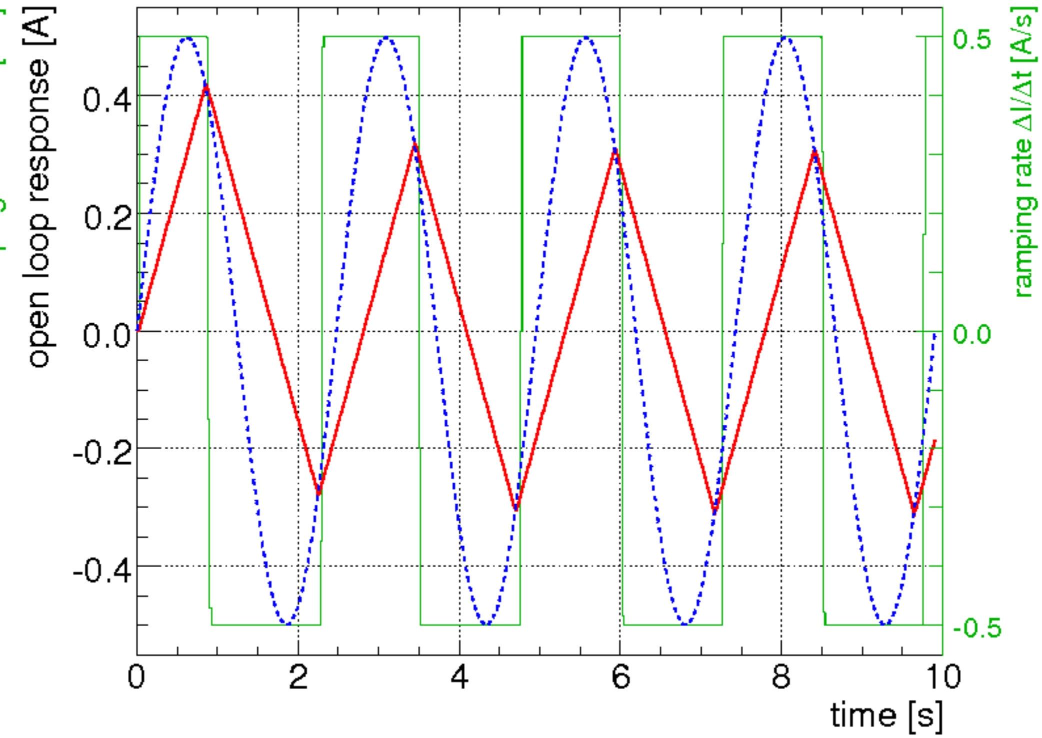
Two non-linear effects most common in accelerators that cannot be necessarily avoided by choice of working points:

- Delays: ADC sampling/pipe-line, computation, data transmission, dead-time, etc.
- Rate-Limiter: limited slew rate of corrector circuits (due to voltage limitations)

slow perturbation: perfect tracking



fast perturbation: saw-tooth

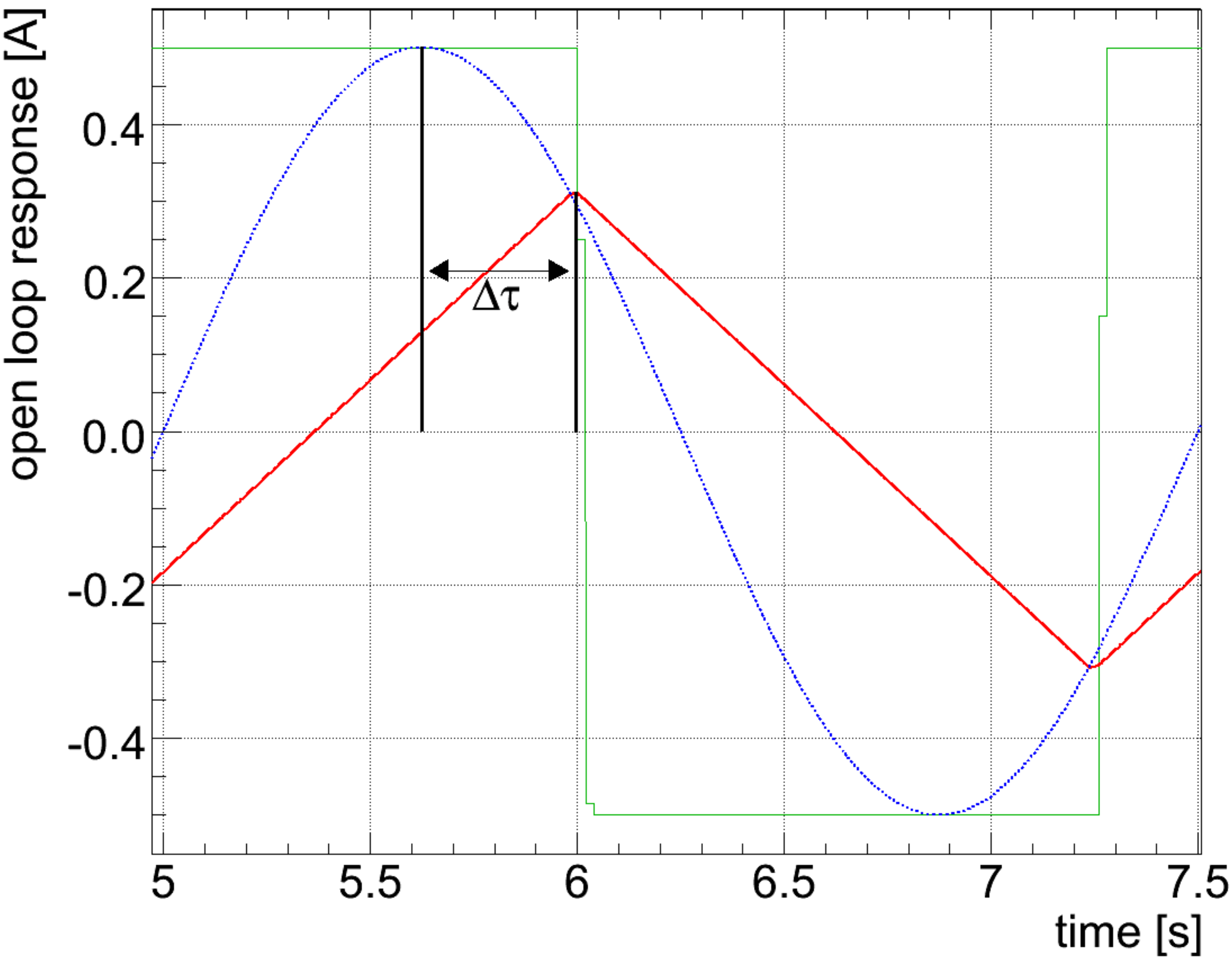


Time-Domain: Non-Linearities II/IV

Rate-limiter in a nut-shell:

- additional time-delay Δτ that depends on the signal/noise amplitude

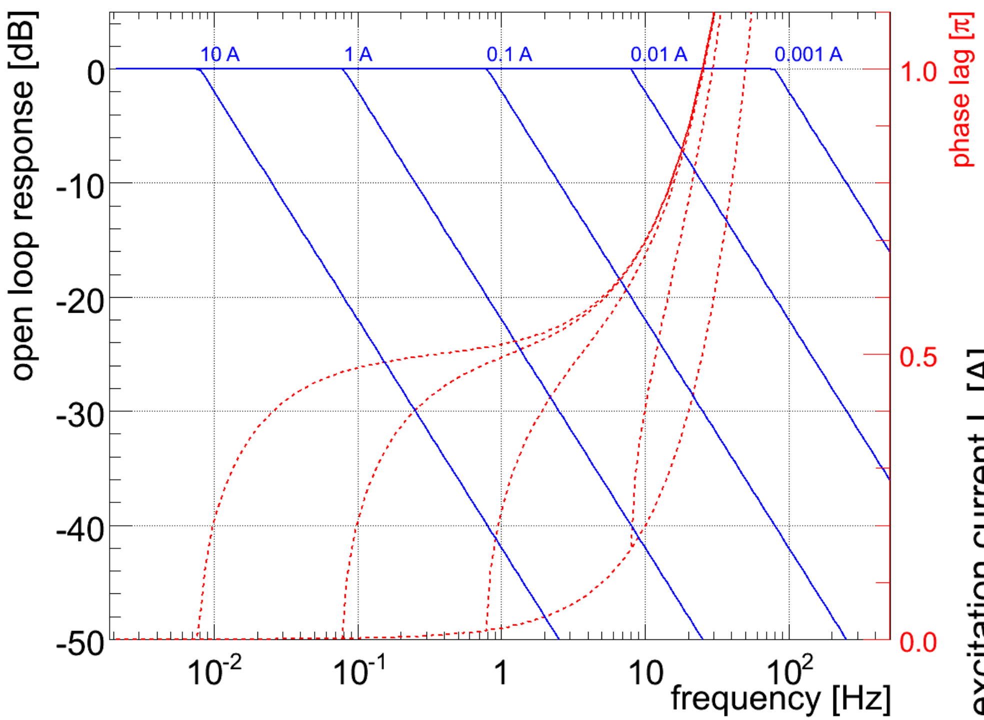
  (secondary: introduces harmonic distortions)



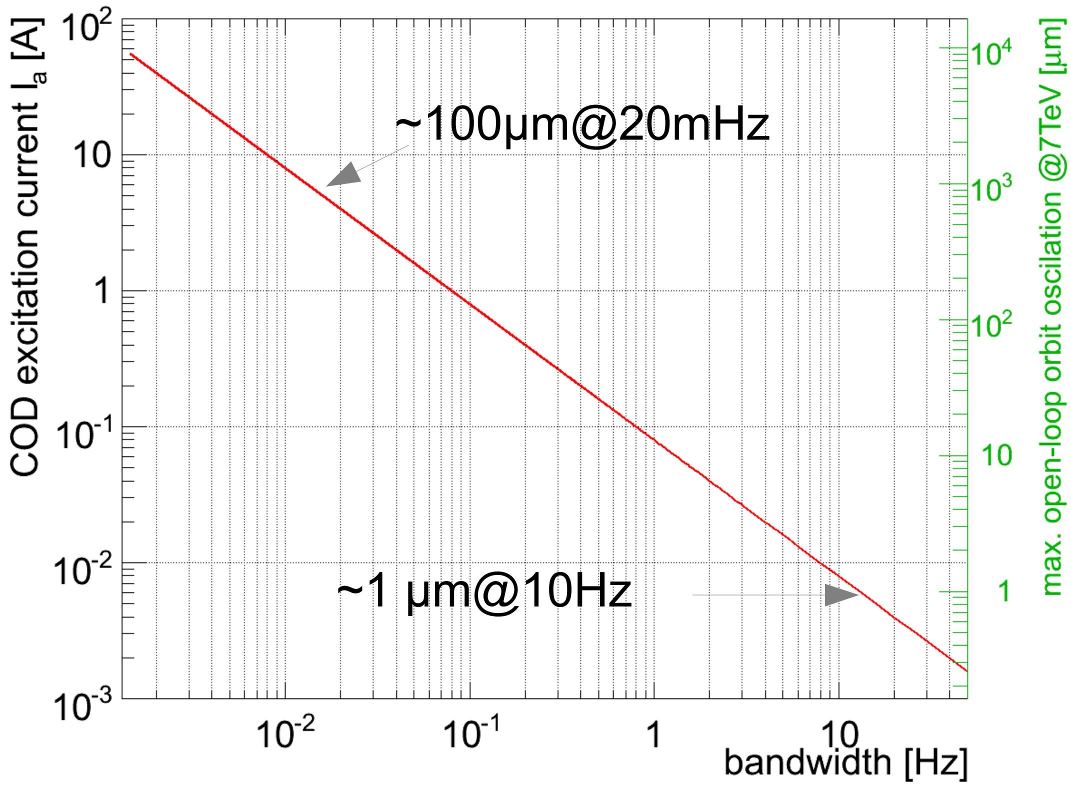
Time-Domain: Non-Linearities III/IV

- Open-loop circuit bandwidth depends on the excitation amplitude:

  + non-linear phase once rate-limiter is in action...



$\Delta I = 0.1A \leftrightarrow \Delta x \approx 16\ \mu m@\beta = 180m$



Time Domain: Non-Linearities IV/IV
Unstable Zeros/non-linearities and delays

$$D(s)=\frac{Q(s)}{1-Q(s)G(s)} \qquad Q(s)=F_Q(s)G^i(s) \qquad (1)$$

- ... cannot a priori be compensated.
  - however, their deteriorating effect on the loop response can be mitigated by taking them into account during the controller design.
- Example: process can be split into stable and unstable 'zeros'/components

$$G(s)=\frac{A_0(s)A_u(s)}{B(s)}=G_0(s)\cdot G_{NL}(s) \quad e.g. \quad G(s)=G_0(s)\cdot\underbrace{e^{-\lambda s}}_{\lambda\text{: delay}}$$

- Using the modified ansatz ($F_Q(s)$: desired closed-loop transfer function):

$$Q(s)=F_Q(s)\cdot G^i(s)=F_Q(s)\cdot G_0^{-1}(s)$$

- yields the following closed loop transfer function

$$\rightarrow \quad T(s)=Q(s)G(s)=F_Q(s)\cdot G_{NL}(s)\underbrace{=}_{here:}F_Q(s)\cdot e^{-\lambda s}$$

  - Controller design $F_Q(s)$ carried out as for the linear plant
  - Yields known classic predictor schemes:
    - delay → Smith Predictor
    - rate-limit → Anti-Windup Predictor

Time Domain: Non-Linearities

Example: LHC Feedbacks & Delays + Rate-limiter

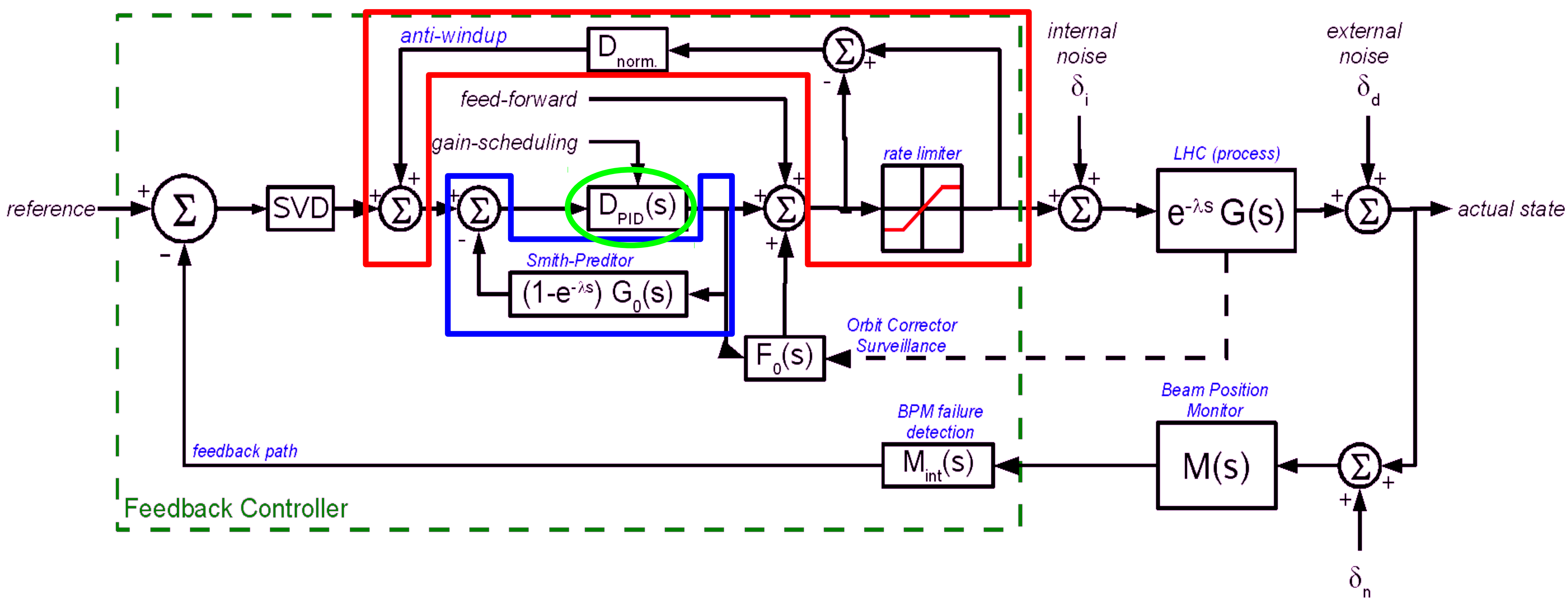
If G(s) contains e.g. delay λ & non-linearities $G_{NL}(s)$

$$G(s)=\frac{e^{-\lambda s}}{\tau s+1}G_{NL}(s) \qquad G^i(s)=\frac{\tau s+1}{1}$$

with τ the power converter time constant and

yields Smith-Predictor and Anti-Windup paths:

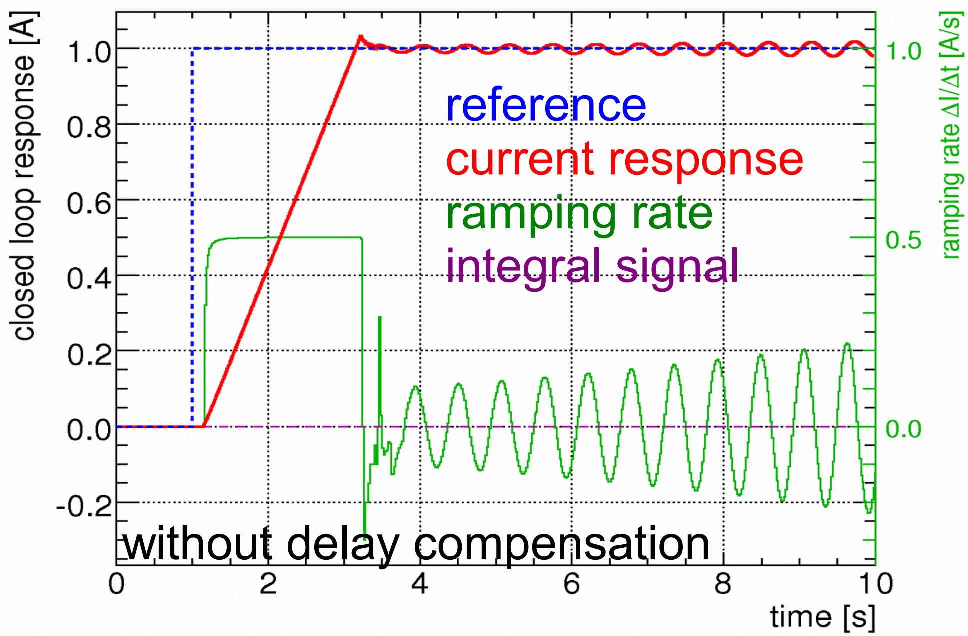
$$T(s)=F_Q(s)\cdot e^{-\lambda s}G_{NL}(s)$$

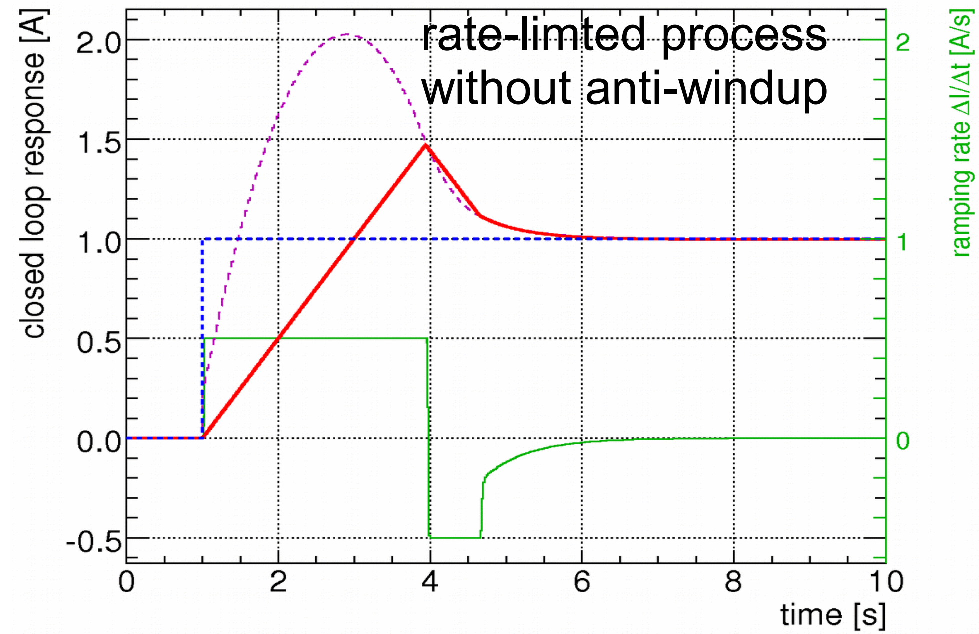


$D_{PID}(s)$ gains are independent on non-linearities and delays!!

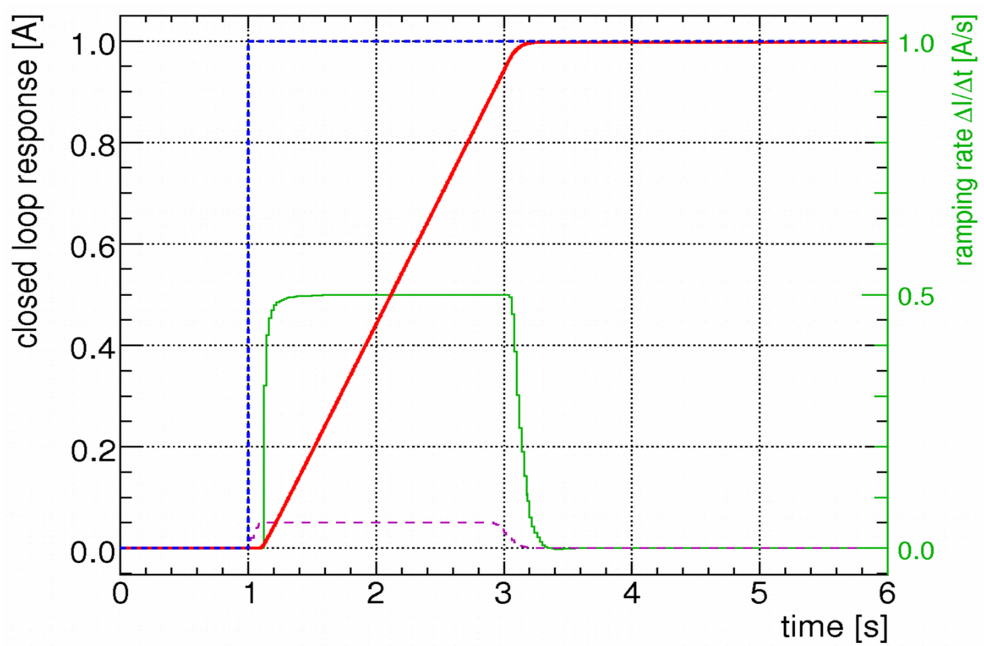
Time Domain: Non-Linearities

Example: LHC Feedbacks & Delays + Rate-limiter





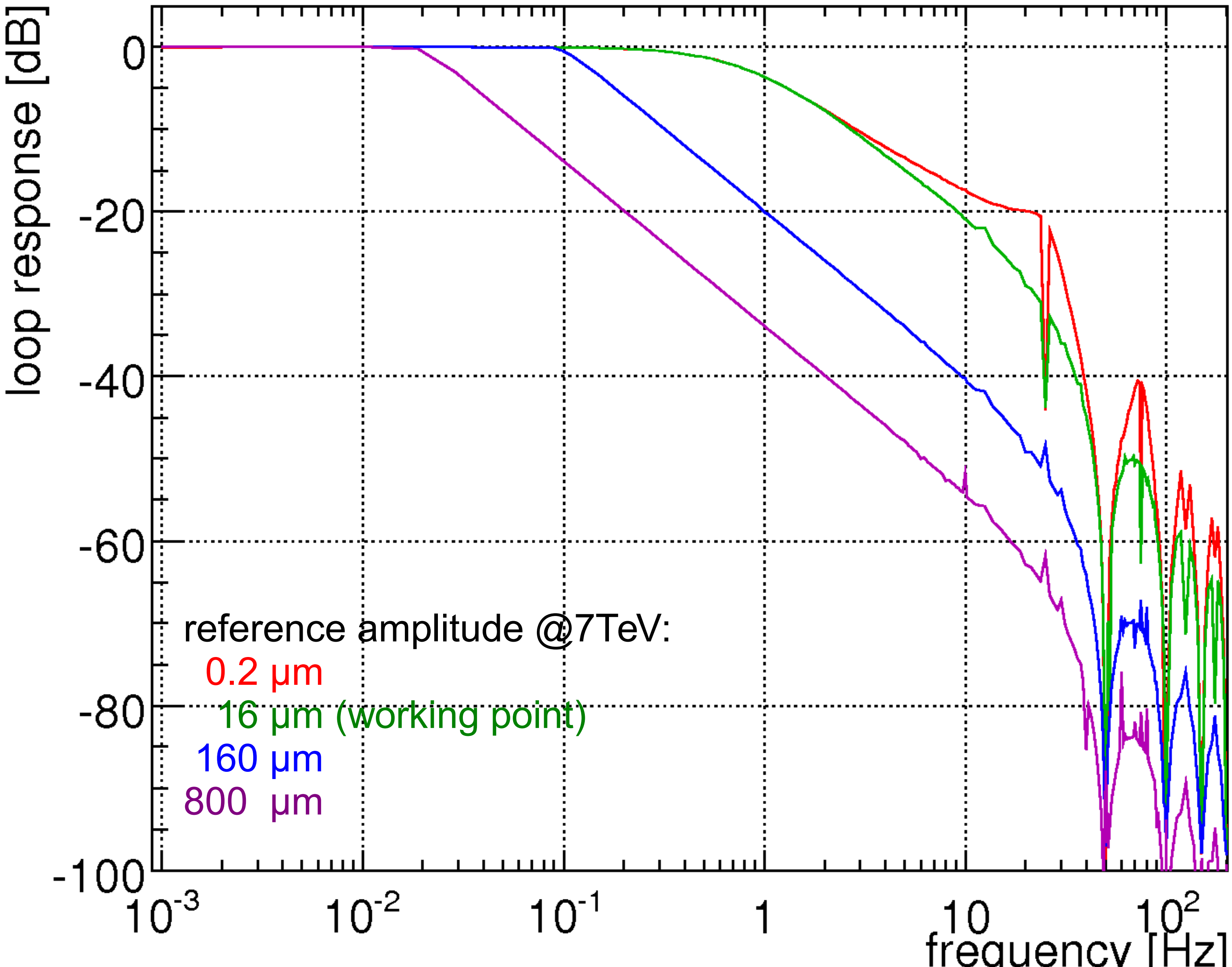
with full delay and windup
compensator scheme:



Time Domain: Non-Linearities

Example: LHC Feedbacks – Nominal Feedback Response T_0

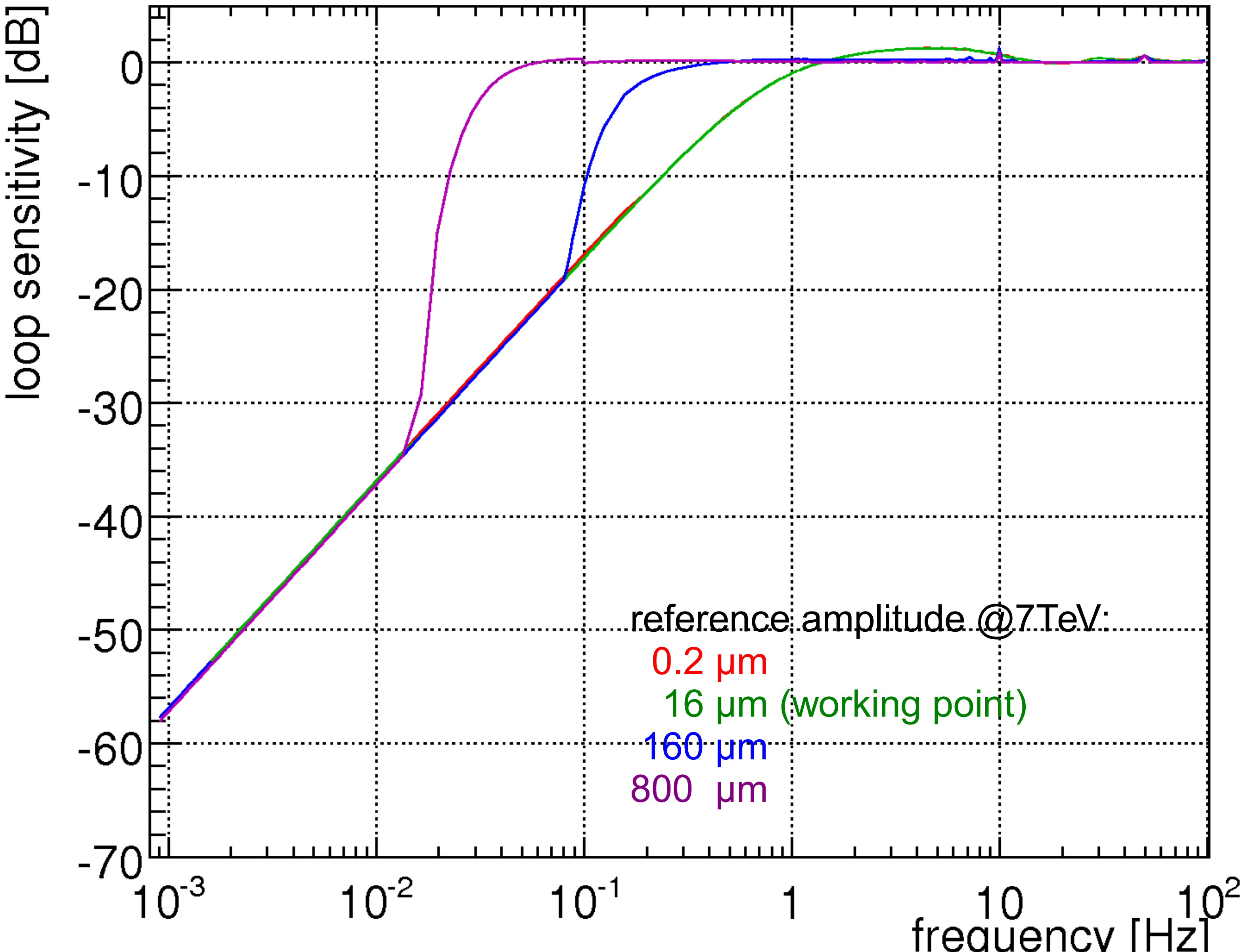
Full LHC orbit simulation @1KHz sampling, (BPM sampling: 25Hz)



Time Domain: Non-Linearities

Example: LHC Feedbacks – Nominal Feedback Disturbance Rejection S_{d0}

Full LHC orbit simulation @1KHz sampling, (BPM sampling: 25Hz)



Divide-and-Conquer Feedback Architecture

Divide:

FB zoo: Orbit, Tune, Chromaticity, β-Coupling, Energy, ..., Luminosity, (Beta-Beating)

develop/commission on a one-by-one basis

Feedback controller into:

Space Domain: $\Delta Q_{x/y} \rightarrow$ quadrupole circuits currents, etc.

classic parameter control – pre-requisite for any beam steering

Time Domain: compensate for dynamic behaviour

relaxed controller for commissioning (low-bandwidth)

Conquer:

Once feedback operation on a per-parameter basis is established, reintegrate and test/commission inter-loop coupling and other constraints.

LHC Feedback hierarchy:

Orbit (Energy) → Tune/Coupling PLL → Q' Tracker → Q/C^-/Q' feedback

Divide-and-Conquer Feedback Architecture

Inter-loop Cross-Dependencies

- Most accelerators rely on multiple feedback loops that simultaneous act on the beam:
  - Low-Level-RF: cavity control affected by RF source power loop, cavity tuner, synchro-loop, fast longitudinal feedbacks
  - beam-based feedbacks on: orbit, energy (radial loop), Tune-PLL, tune, chromaticity, coupling, luminosity, fast transverse feedback (damper), synchro-loop, ...

- Feedbacks on non-orthogonal/non-independent parameters can/will cause cross talks and even instabilities if not designed properly! Some choices:

  A) Decoupling of the parameter space:
     - Orbit FB (betatron-pertubations) vs. Energy FB (dispersion orbit)

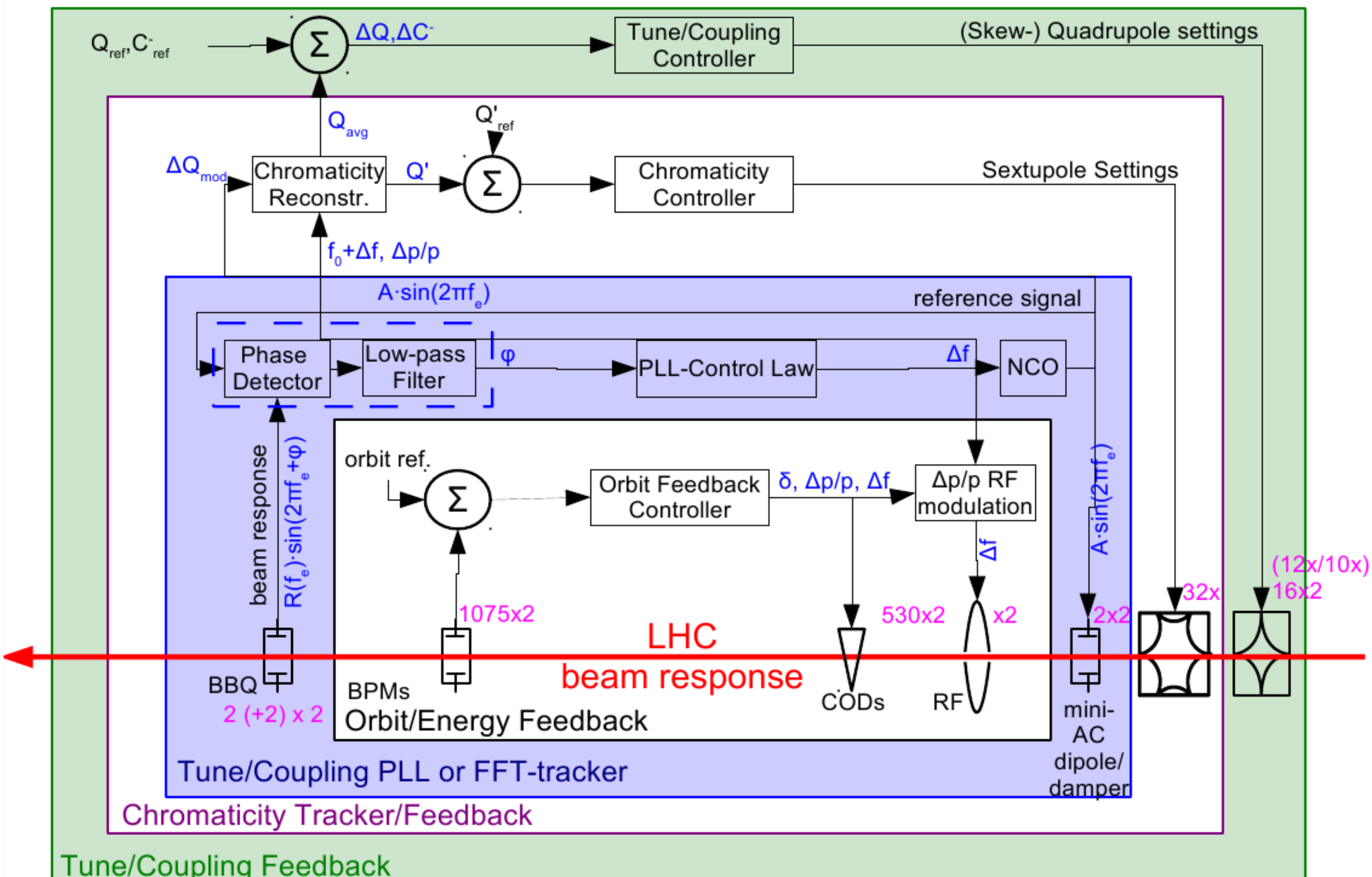
  B) Decoupling of operational ranges (either e.g. amplitude or time scales)
     - Q-PLL being faster than Q' tracker faster than actual Q loop
     - Q-PLL – transverse feedback cross-talk:
       - PLL operates within transverse feedback's "noise"
       - PLL operates on single bunch exempted from other fast Fbs.

  C) Introducing a Master-Slave Structure:
     - Energy FB & Q' Tracker sharing the same reference
     - Orbit FB being the slave to the luminosity FB, local bumps ...

Inter-loop Cross-Dependencies

example: LHC cascading to minimise inherent cross-talk between FB loops

- Main strategy: derive measurement from FB control variable
  - Q'-tracker using '$Q_{raw} = Q_{meas} - Q_{trim}$'
  - Sub. Δp/p-mod. from Radial-Loop & Orbit-FB reference



further: $f_{BW}(PLL) \gg f_{BW}(Q') \geq f_{BW}(Q, C^{-})$

Robustness & Modelling Errors

“In theory, 'theory' and 'praxis' are the same, in praxis they aren't”

- Real-world modeling errors are unavoidable:

$$G(s)=G_0(s)+G_\epsilon(s)$$

- resulting disturbance rejection:

$$\boxed{S(s)=\frac{S_0(s)}{\left(1+G_\epsilon(s)T_0(s)\right)}=\frac{1-Q(s)G_0(s)}{1+Q(s)G_\epsilon(s)}}$$

- Remedy: increase magnitude & phase margins to ensure robustness:
  - $|T_0(j\omega)|$ should rolls off before effects of modeling errors become significant
  - add appropriate poles in $F_Q(s)$.

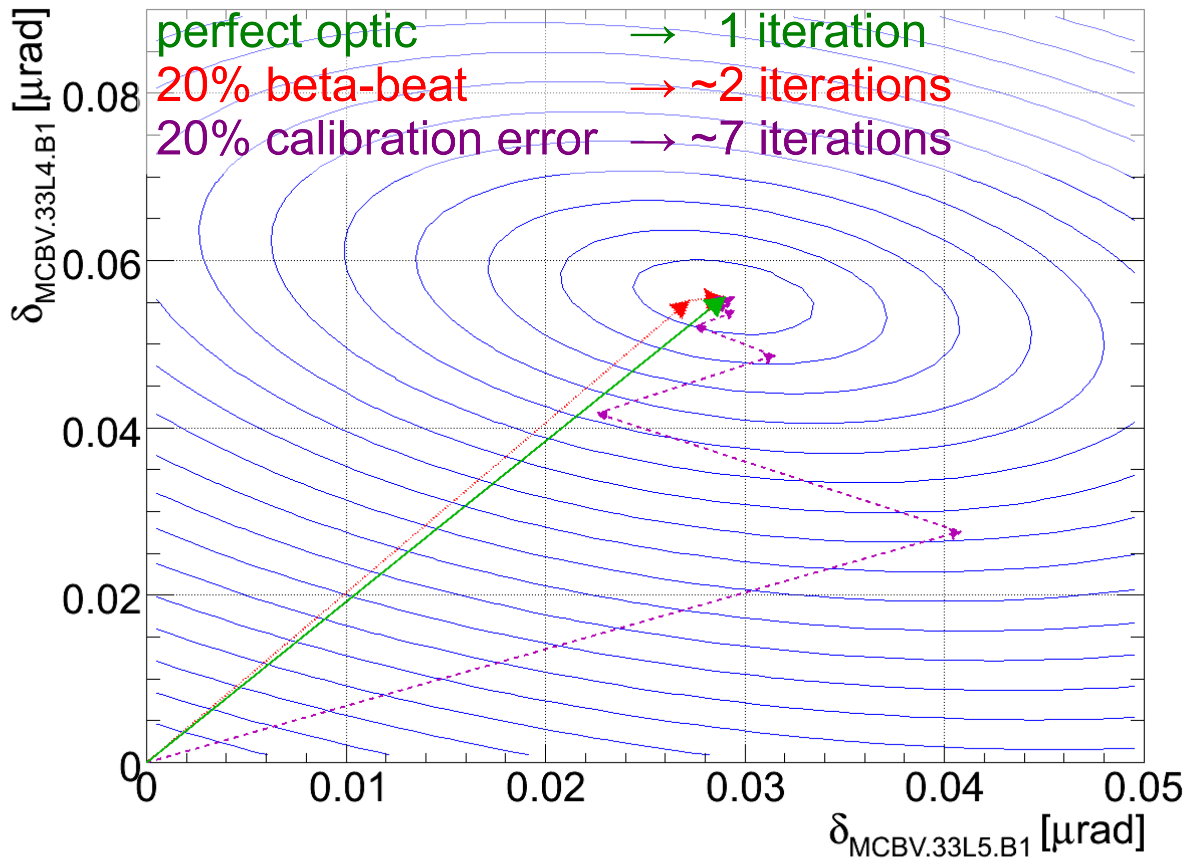
Robustness & Modelling Errors

Some additional constraints (stable open-loop poles)

- Non-minimum phase zeros:
  - internal stability precludes the cancellation of these zeros → must appear in $T_0(s)$ and gain of Q(s) reduced at these frequencies
- Relative degree:
  - excess poles in the model must necessarily appear as a lower bound for the relative degree of $T_0(s)$, since Q(s) must be proper to ensure that the controller C(s) is proper
- Disturbance trade-offs:
  - whenever we roll T_0 off to satisfy measurement noise rejection, we necessarily increase sensitivity to output disturbances at that frequency.
  - slow open-loop poles must either appear as poles of $S_{i0}(s)$ or as zeros of $S_0(s)$
    - in either case there is a performance penalty.
- Control energy: Most processes in accelerators are typically low pass
  - Q(s) being close to model's inverse → high-pass transfer function from $D_0(s)$ to U(s)
  - → may lead to large input signals and may lead to controller saturation
- Robustness:
  - modelling errors are usually more significant at high frequencies, and hence to retain robustness it is necessary to attenuate $T_0(s)$ and hence Q(s), at these frequencies.

Robustness & Modelling Errors
Example: Optics and Calibration Uncertainties

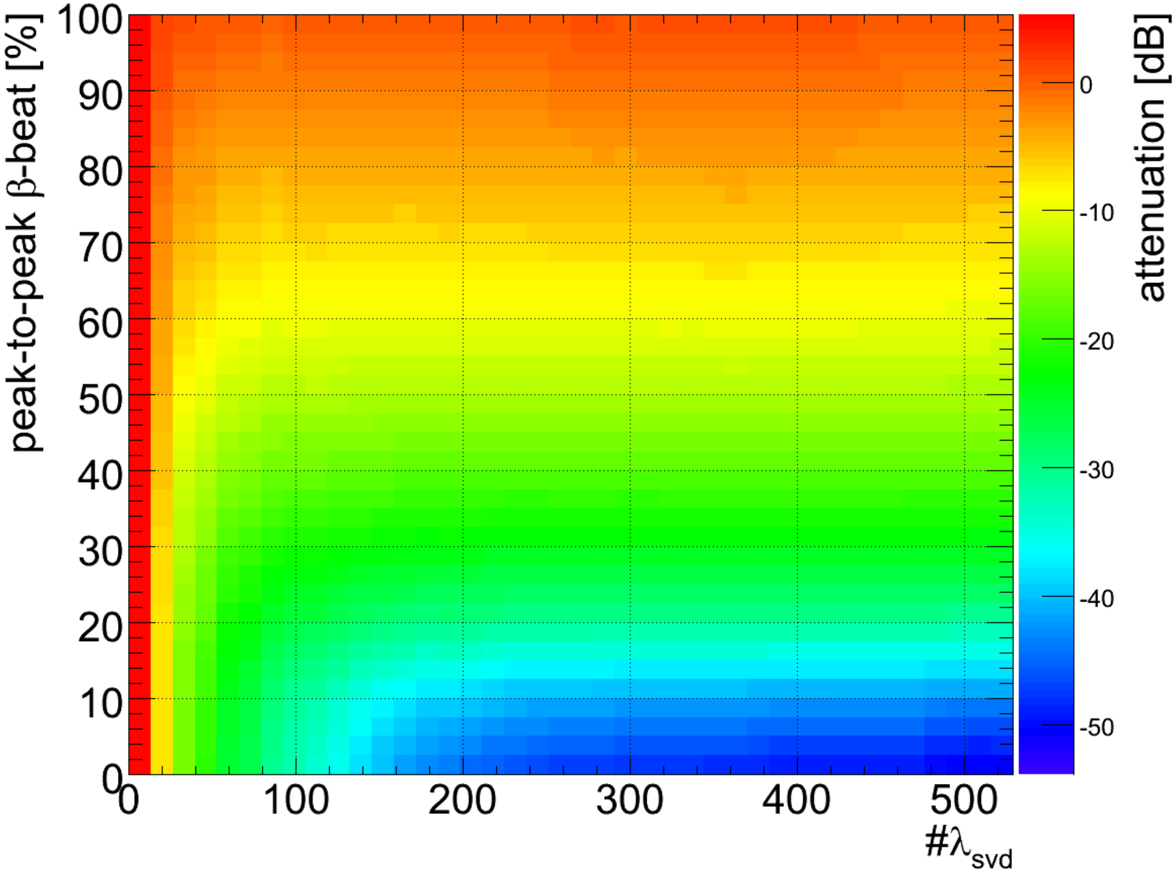
- Optics imperfections may deteriorate the convergence speed but do not affect absolute/steady-state convergence (response functions are 'monotonic')
- Example: 2-dim orbit error surface projection



- e.g. LHC orbit feedback is to 1st order practically insensitive to optics (= beta-beat) error.
  - However, pickup and corrector magnet polarities are crucial
  - Watch-out in time-domain: reduced convergence speed → reduces the closed-loop phase margin

Robustness & Modelling Errors
Example: LHC Orbit-FB Sensitivity to beta-beat

- Low sensitivity to optics uncertainties = high disturbance rejection:
  - LHC simulation: Inj. Optics B1&B2 corrected



$\#\lambda_{svd}$ controls correction precision

attenuation =

$$20 \cdot \log \left| \frac{orbit\ r.m.s.\ after}{orbit\ r.m.s.\ before} \right|_{ref}$$

- Robust Control: OFB can cope with up to about 100% β-beat!
  - Robustness comes at a price of a (significantly) reduced bandwidth!

Robustness & Modelling Errors

Space Domain: Number of used eigenvalues?

Gretchen Frage: "How many eigenvalues should one use?"

small number of eigenvalues:

- more coarse type of correction:
  - –use arc BPM/COD to steer in crossing Irs
  - –less sensitive to BPM noise
  - –less sensitive to single BPM faults/errors
  - –less sensitive to single COD/BPM faults/errors
- robust wrt. machine imperfections:
  - –beta-beat
  - –calibration errors
- easy to set up
- …
- poor correction convergence
- leakage of local perturbations/errors
- not fully closed bump affects all Irs
- squeeze in IR1&IR5 affects cleaning Irs
- ...

large number of eigenvalues:

- more local type of correction
  - –more precise
  - –less leakage of local sources onto the ring
  - –perturbations may be compensated at their location
- good correction convergence
- ...
- more prone to imperfections
  - –calibration errors more dominant
  - –instable for beta-beat > 70%
- more prone to false BPM reading
  - –errors & faults, reliability reduction
- ...

parameter stability requirement

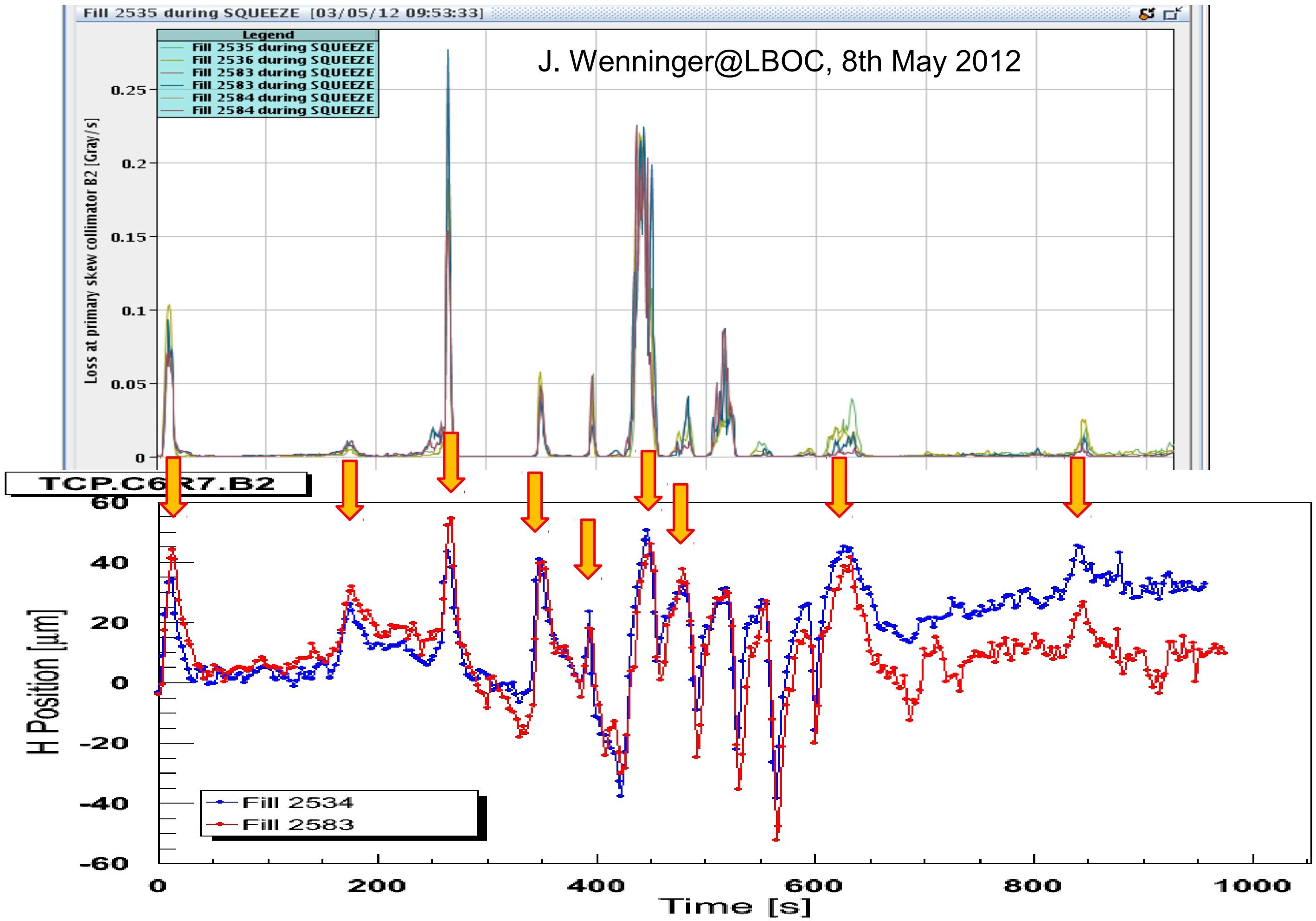
feedback stability requirement

Choice for Q, Q', C^- is much simpler: only two out of *n* non-vanishing eigenvalues!

Robustness & Modelling Errors

Example: Sensitivity to beta-beat & LHC Orbit Stability during β*-Squeeze

- Losses and orbit movement at H-TCP.C6R7.B2 well correlated

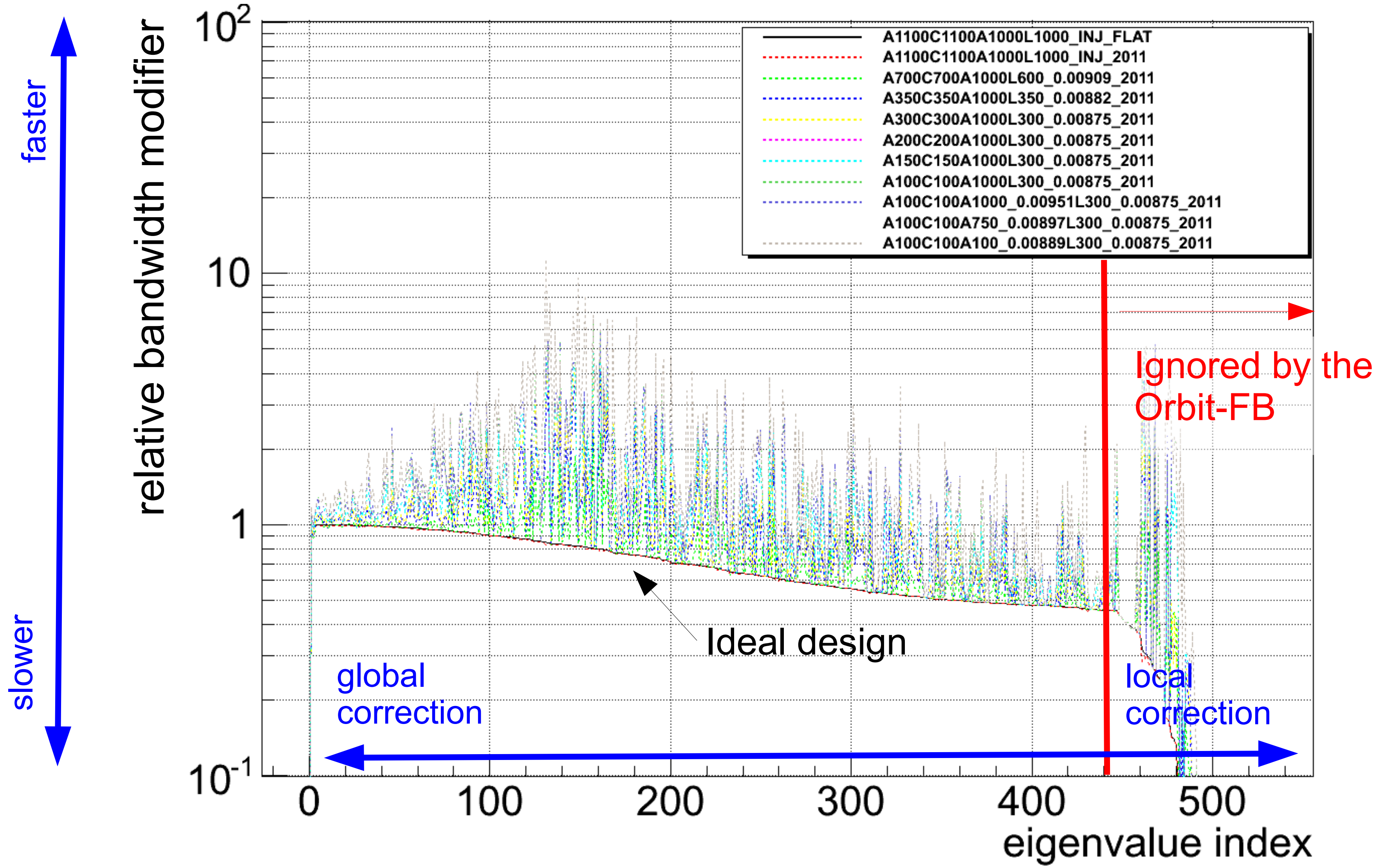


- Maximum drift rates of 40 um/s → (close to) limit of Orbit-FB at 4 TeV
  - Underpinned by FB instability observation for 5x bandwidth increase
- At this speed, OFC needs to operate with correct optics

Robustness & Modelling Errors

Example: Sensitivity to beta-beat & LHC Orbit Stability during β*-Squeeze

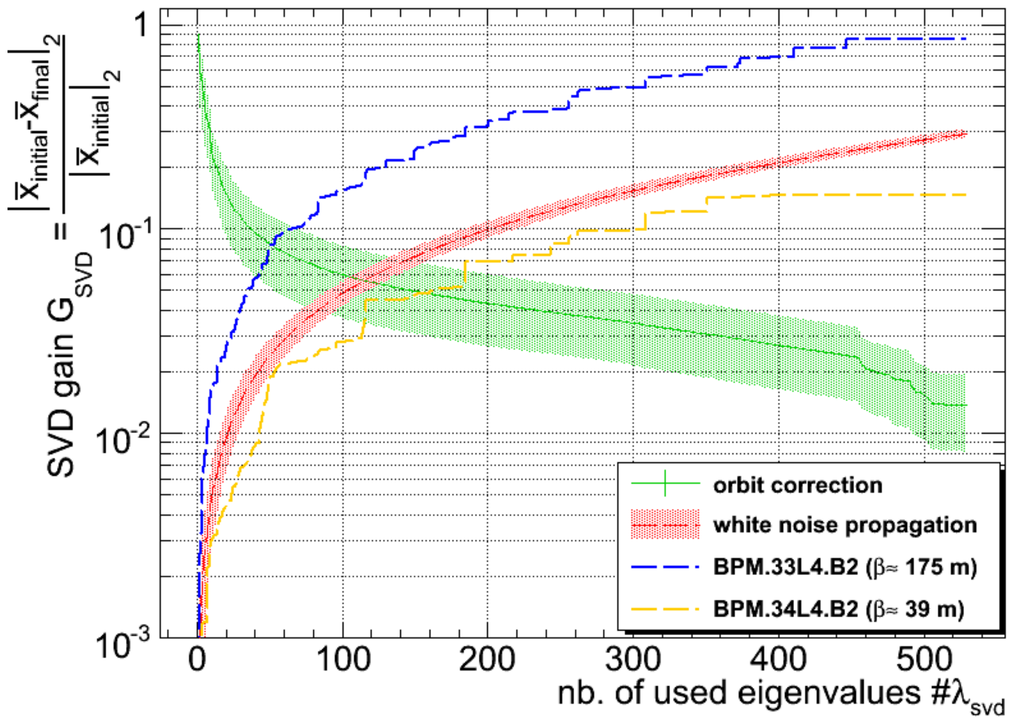
- Bandwidth modifier w.r.t. eigenvalue index (<1 more stable, >1 diminishes stability margin)

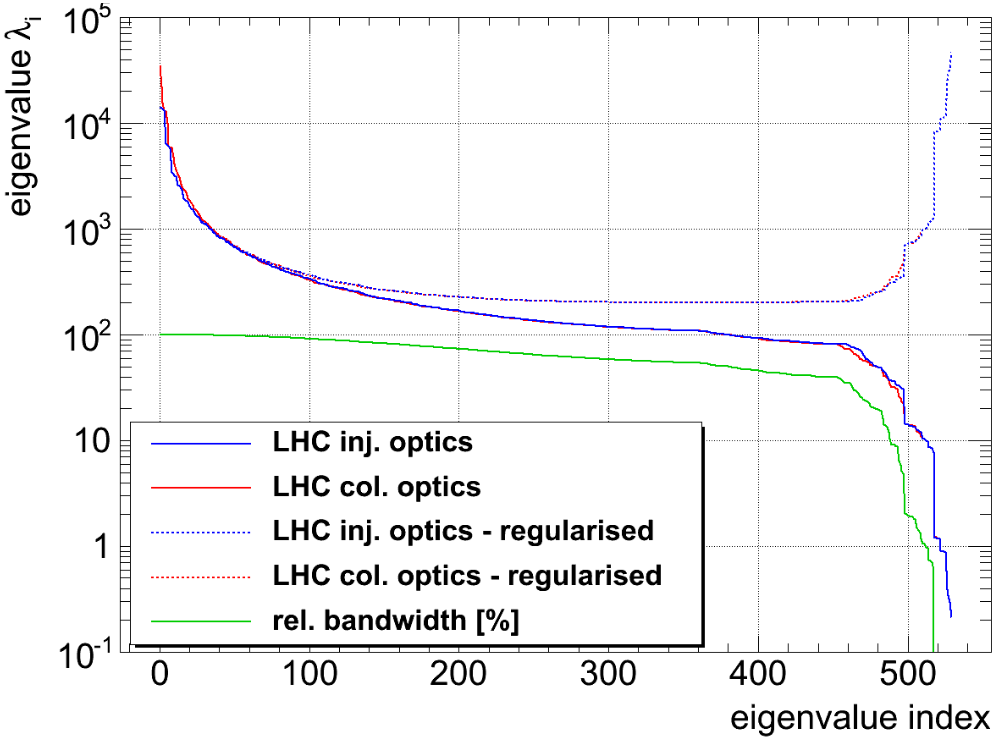


- Typ. opertional bandwidth <10% of maximum possible (sometimes too slow)

Robustness & Modelling Errors
Example: Sensitivity to beta-beat → Optimal Filter Design

- Initially: Truncated-SVD (set $\lambda_i^{-1}:= 0$, for i>N)
  - not without issues: removed λ_i allowed local bumps creeping in (e.g. collimation)
- Regularised-SVD (Tikhonov/opt. Wiener filter with $\lambda_i^{-1}:= \lambda_i/(\lambda_i^2 + \mu)$, $\mu>0$)
  - more robust w.r.t. optics errors and mitigation of BPM noise/errors → allowed re-using same ORM for injection, ramp and 10+ squeeze steps



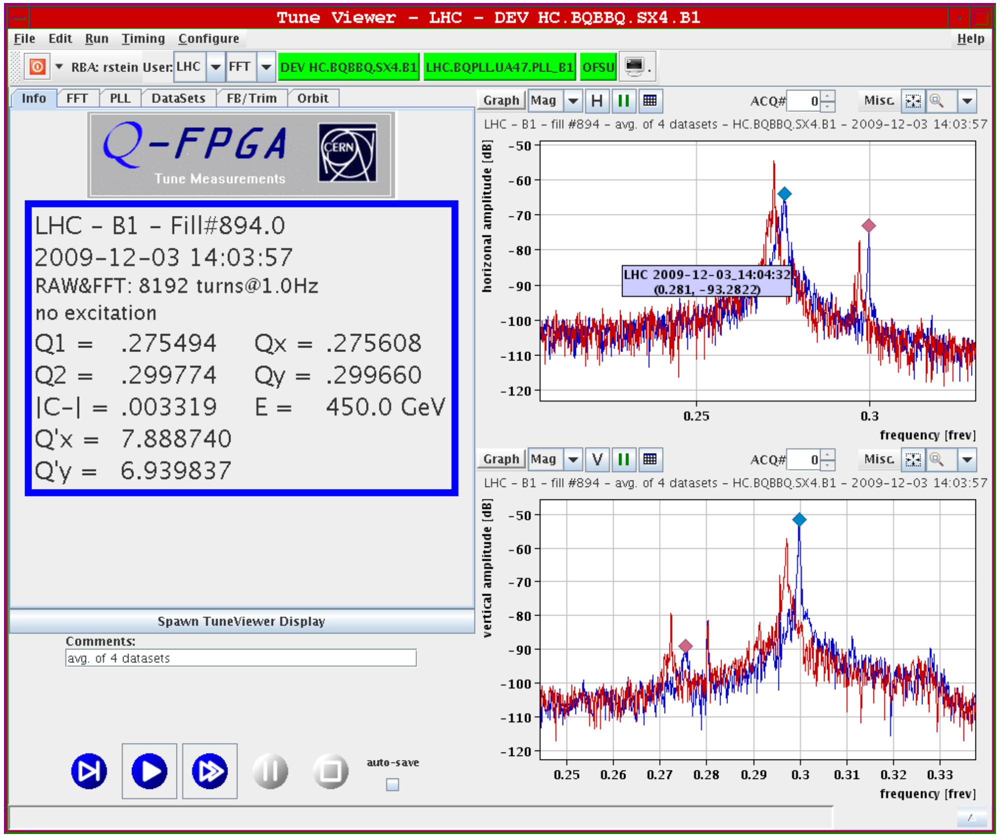


Robustness & Modelling Errors
Example: LHC Q/Q' Diagnostics and Residual Noise I/II

- Initial design assumption: no residual tune signatures on the beam (0 dB S/N)
  – Anticipated constant driving of the beam and – to limit the required excitation levels – the highly-sensitive BBQ system was developed

- Blessing/Curse after start-up:
  1. BBQ turn-by-turn res. < 30 nm
     - 30+dB more sensitivity than other LHC systems
       (e.g. ADT: 1µm, BPM: 50 µm)
  2. Ever-present Q oscillations
     few 100 nm to µm level

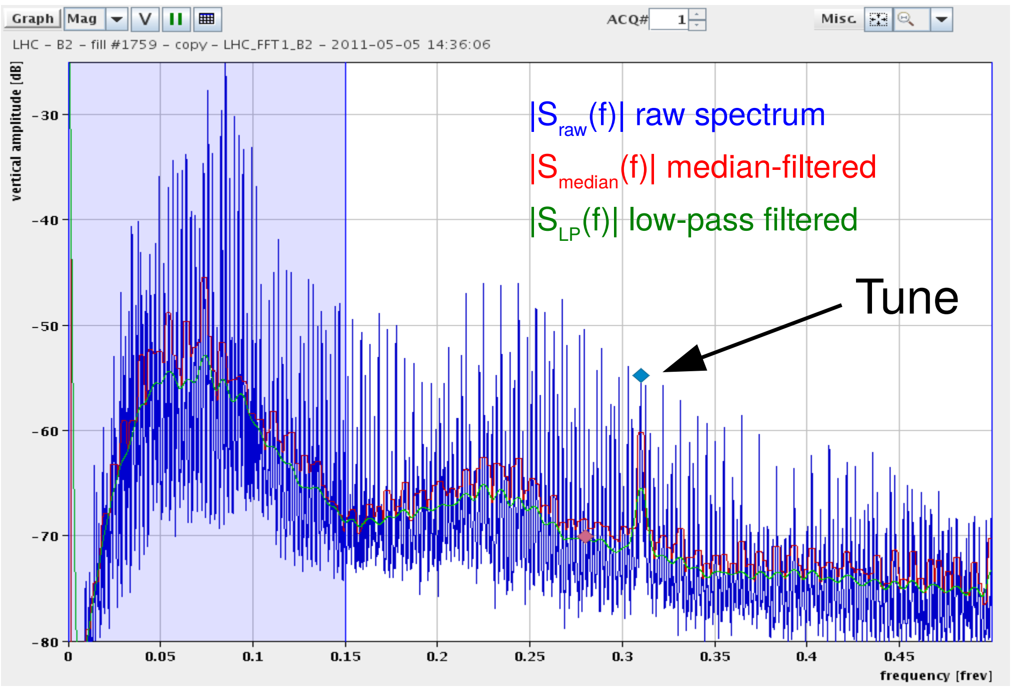
- Luxurious 30-40 dB S/N ratios enabled the passive monitoring, tracking and feedbacks without any additional excitation

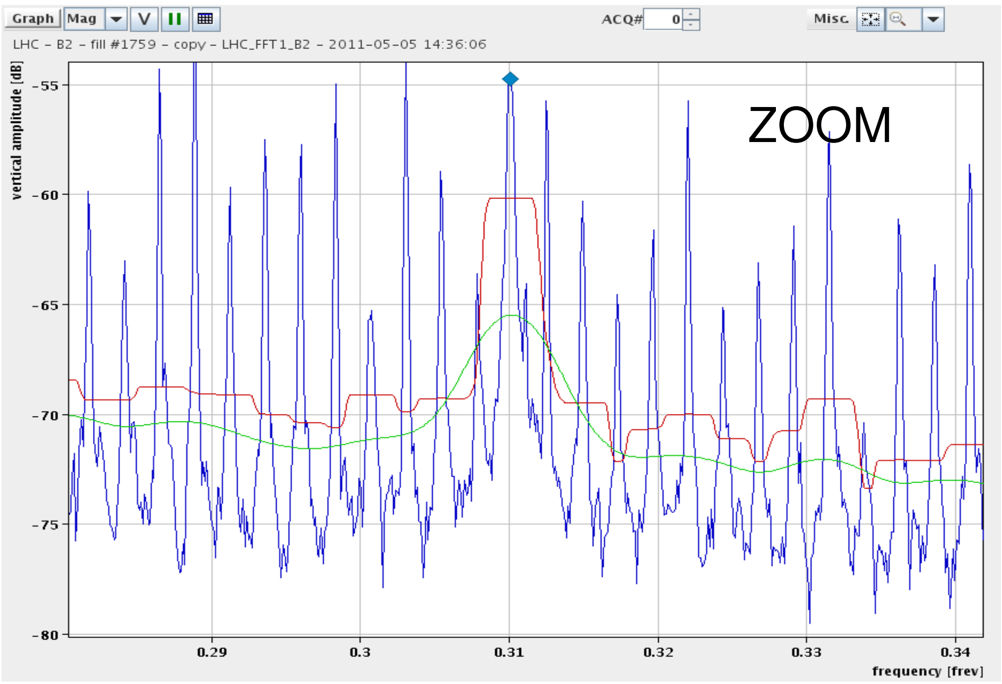


- However,made the Tune-PLL (driving the Tune-FB) de-facto obsolete
  – µm-level oscillations are incoherent "noise" from a Tune-PLL point of view
  – Need to excite ~30 dB above this "noise" to recover "passive" FFT performance
    → 10...100 µm oscillations vs. collimators gap < 200 µm

Robustness & Modelling Errors
Example: LHC Q/Q' Diagnostics and Residual Noise II/II

- Couple of months later with beyond than design intensities … less ideal!





- Q/Q' not a direct beam observable → highly non-trivial detection and tracking
  – strong dependence on beam intensity, filling pattern, particle species, RF settings, ADT, operational mode, …, many cross-links/interferences

- Improving Tune-FB stability implied resolving issues on the diagnostics side (sensors)

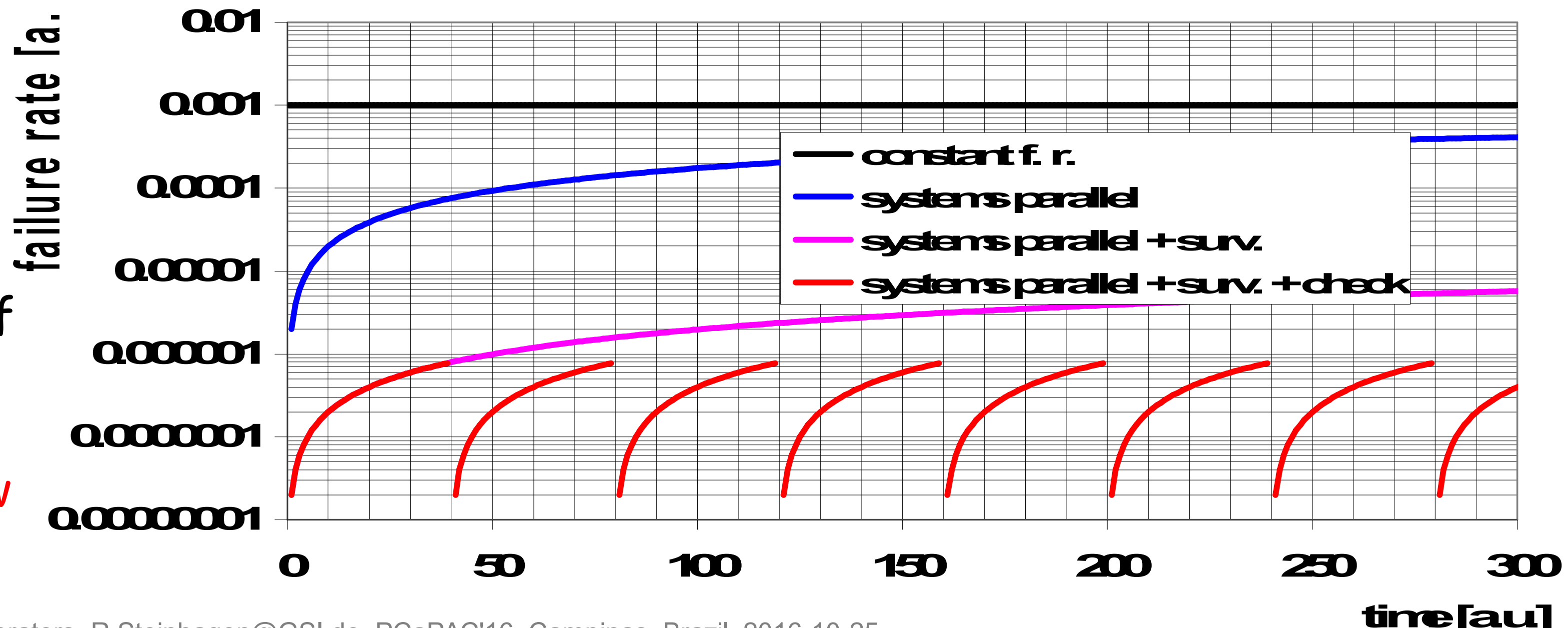
reference: proceedings of DIPAC'11 & CERN-BE-2011-016

Feedback System 'As good as New' Validation

Machine Protection becomes am important issue in many modern high-brightness, high-power and high-energy accelerators:

- FBs are designed to improve/ensure stability, but may equally drive instabilities and the machine into a unknown potentially dangerous state
- FB performance may deteriorate with time
- Actual beam/machine conditions may change w.r.t. conditions during the initial FB setup/tests
- How-to mitigate:
  - A) monitor, detect, intercept and report failures continuously and early
    - N.B. 80% of the LHC FB source code covers the detection of sub-system failures!
  - B) perform periodic checks of basic feedback functionality → verifies and mitigates failure rate early on before becoming an issue (reliability engineering)

Verification of functionality, system as good as new

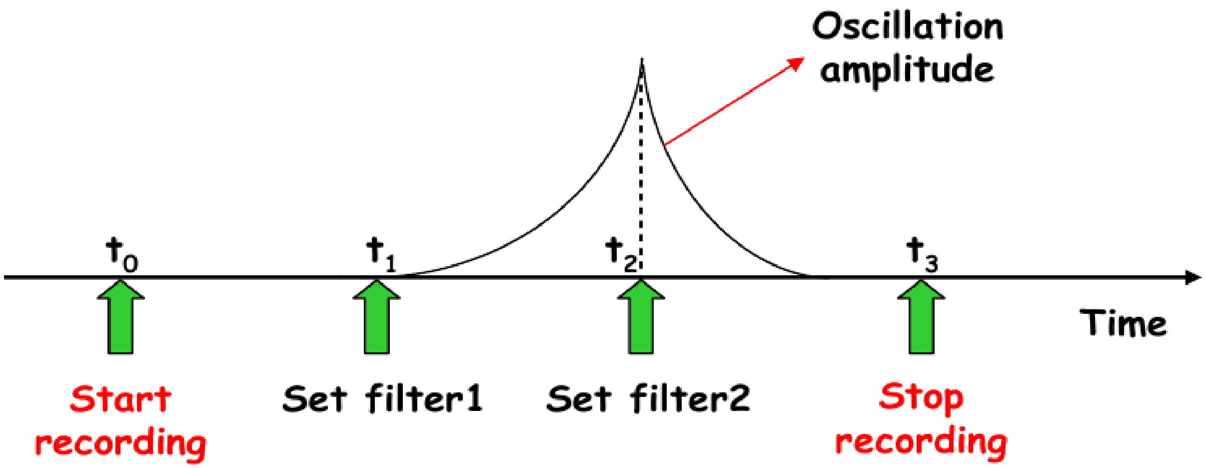


Feedback System 'As good as New' Validation

Fast Transverse Feedback: transient generation

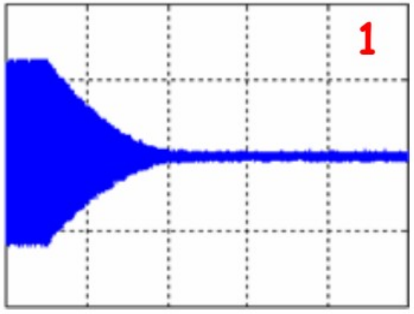
A powerful diagnostic application is the generation of transients.

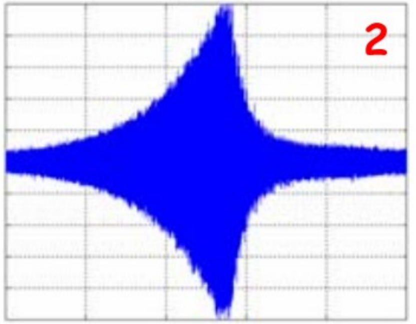
Transients can be generated by changing the filter coefficients accordingly to a predefined timing and by concurrently recording the oscillations of the bunches.

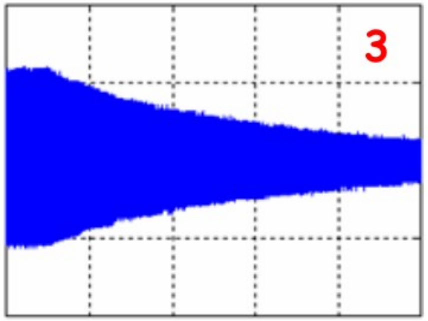


- Different types of transients can be generated, damping times and growth rates can be calculated by exponential fitting of the transients:

1. Constant multi-bunch oscillation → 'FB on': damping transient
2. 'FB on' → 'FB off' → 'FB on': grow/damp transient
3. Stable beam → positive 'FB on' (anti-damping) → 'FB off': natural damping transient



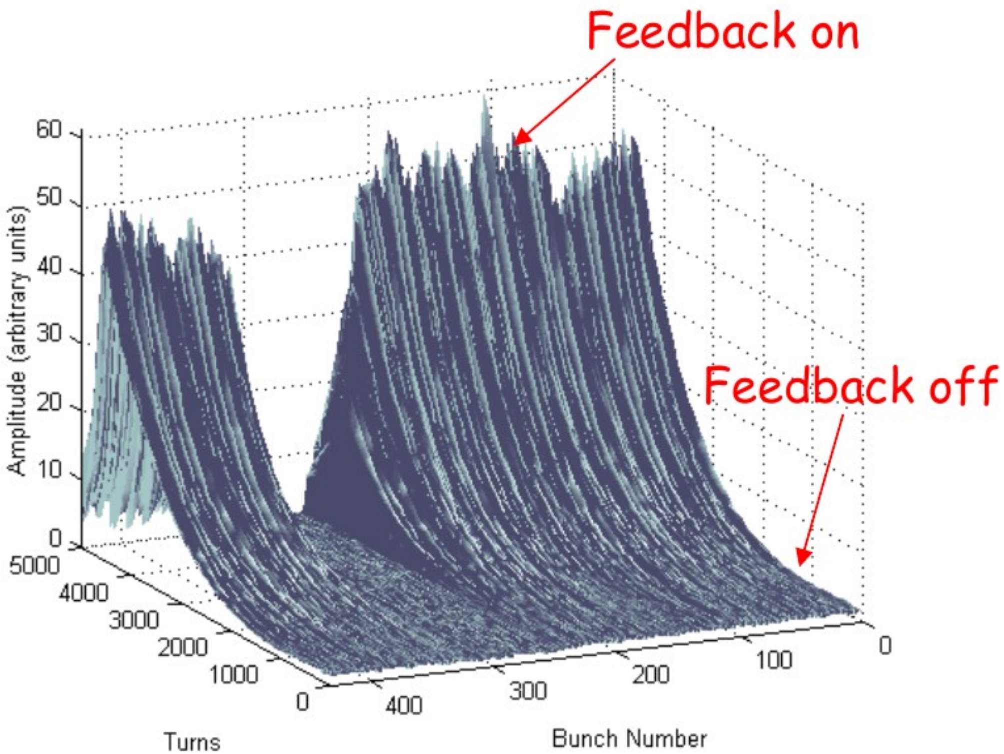




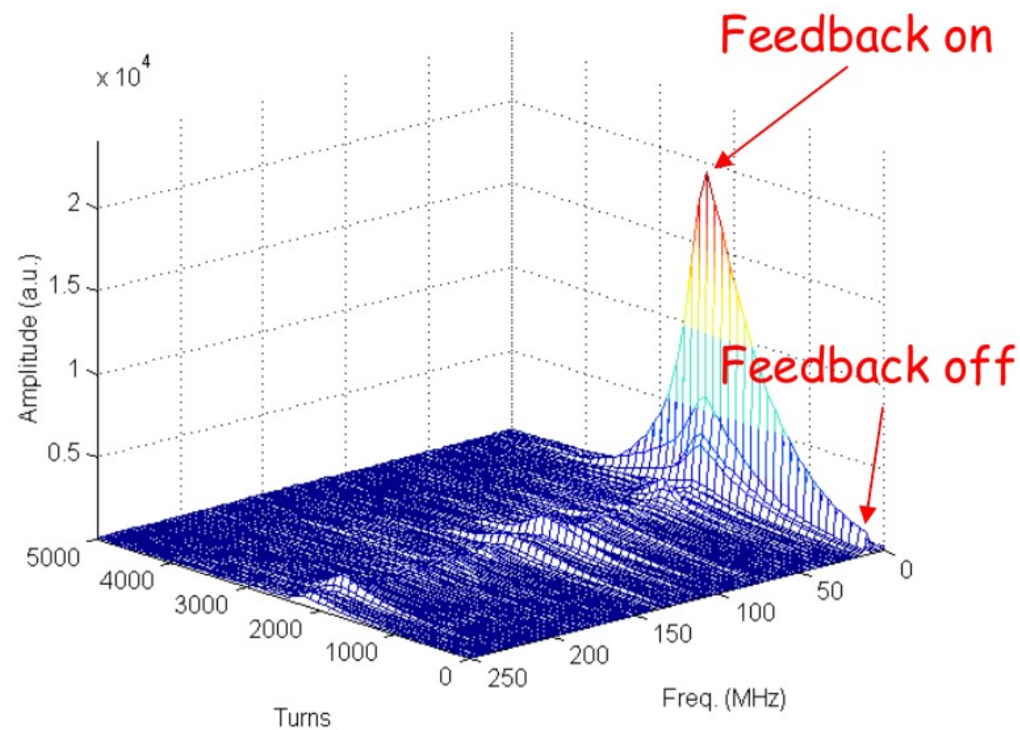
Courtesy Marco Lonza, Elettra

Feedback System 'As good as New' Validation

Grow/damp transients: 3-D graphs



Evolution of the bunches oscillation amplitude during a grow-damp transient



Evolution of coupled-bunch unstable modes during a grow-damp transient

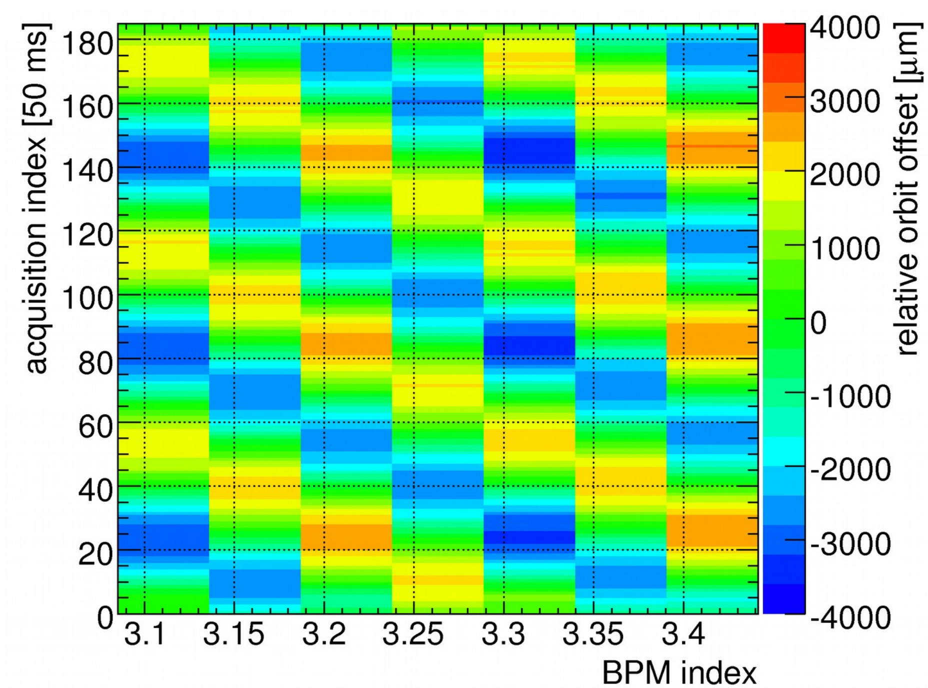
Courtesy Marco Lonza, Elettra

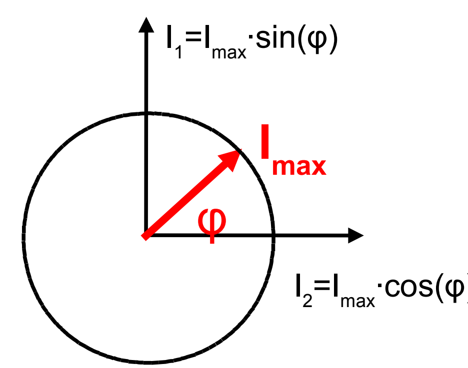
Feedback System 'As good as New' Validation

Orbit-FB: three main lines of defense against BPM, COD, …, errors and faults

1. Pre-checks without beam using the in-build calibration unit
   - eliminates open/closed circuits, dead BPMs
   - Compensates for large-scale temperature effects

2. Pre-checks with Pilot and Intermediate beams
   - Idea: “Every non-moving position reading indicates a dead BPM”
   - open-loop response: forced slow COD-driven betatron oscillation with rotating phase
     - Tests also calibration factors and/or rough optics estimate

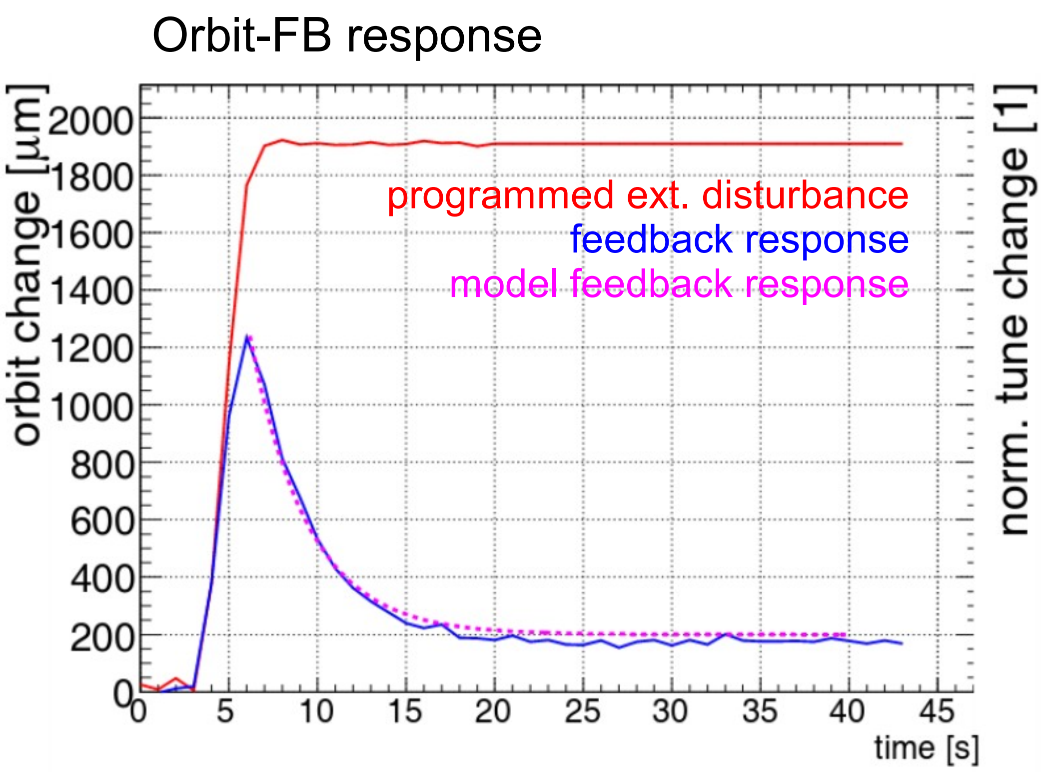


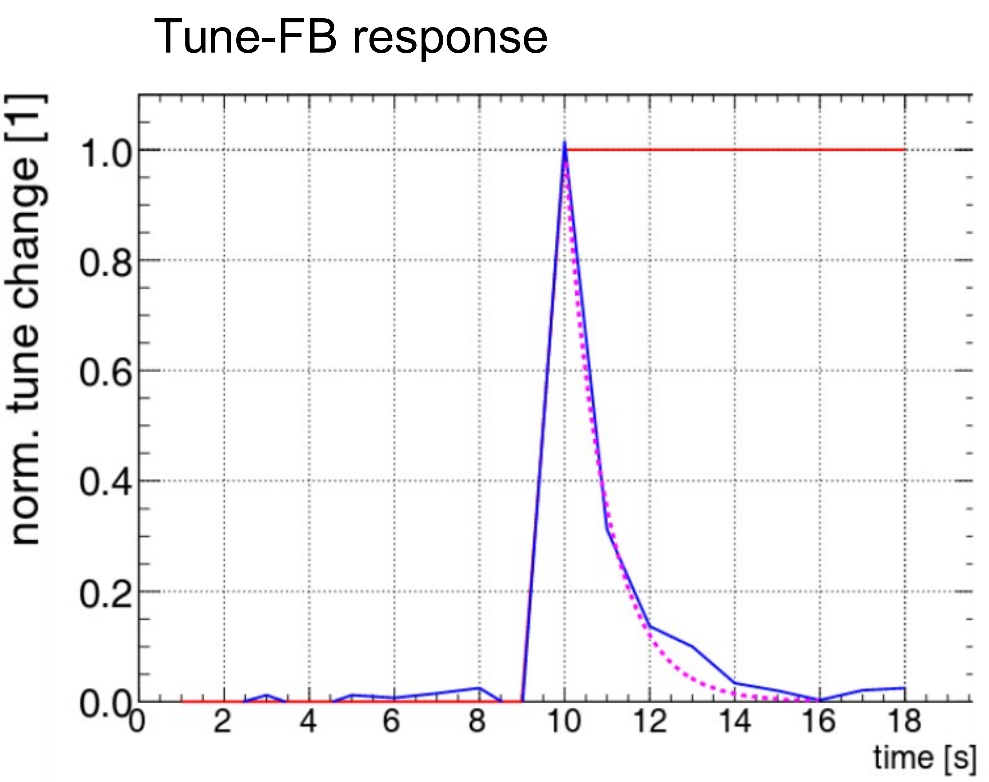


   - Closed-loop response measurement

3. Continuous data quality monitoring through Orbit Feedback
   - detects spikes, steps and BPMs that are under verge of failing

Feedback System 'As good as New' Validation
LHC Orbit-FB closed-loop verification





A Note on Dependence of Operation on Feedbacks
thoughts from LHC Feedback review:

- Could/should LHC run without Feedbacks: – NO

  1 More than 50% of fills would have probably been lost without FBs
    - mostly during or after of changing the mode-of-operation

  2 Even with perfect feed-forward, FBs provide a robustness to operation by mitigating “unforseen” or feed-down effects

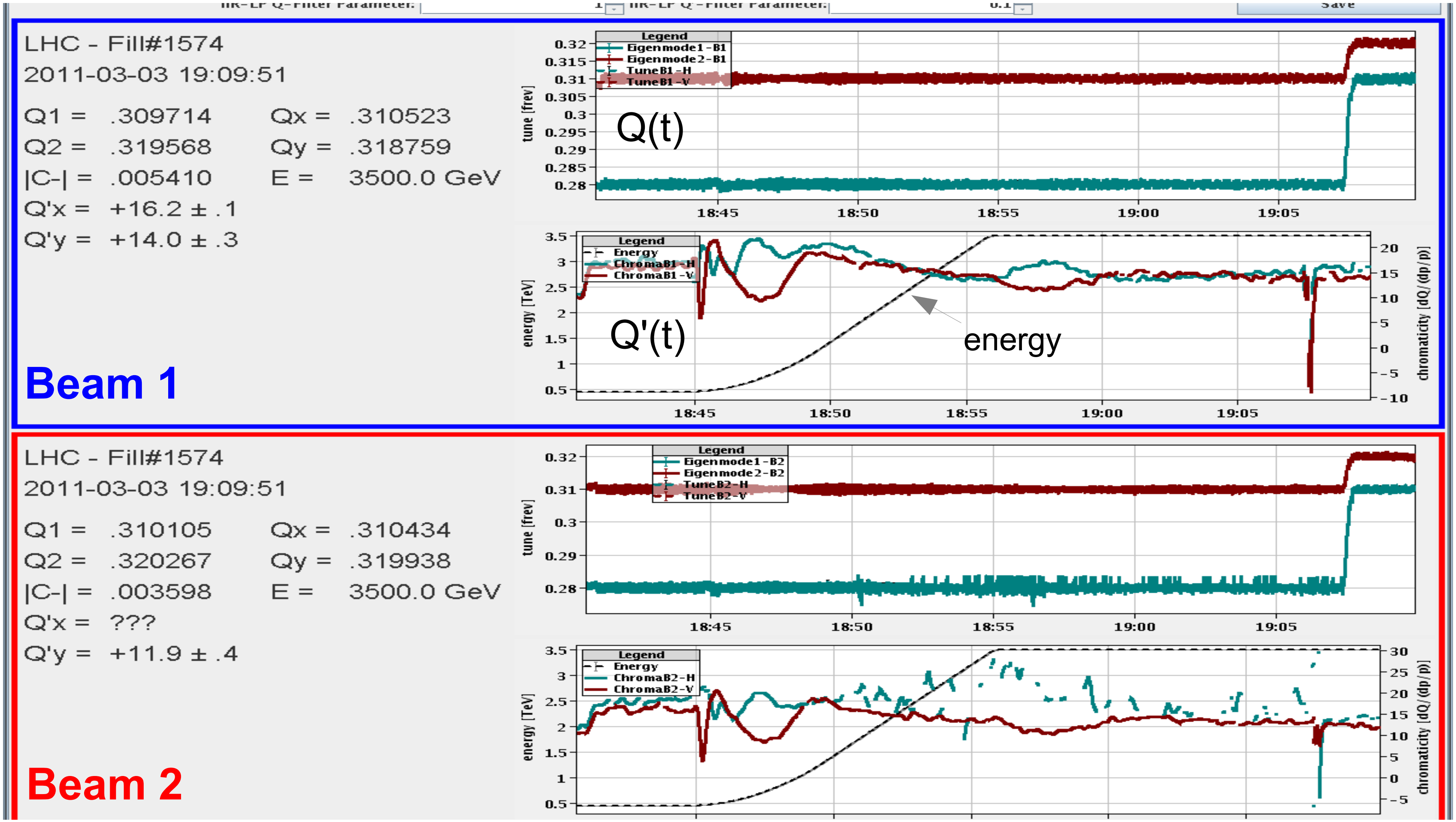
However:

“Having a car brake or ESP/ABS system does not justify reckless driving!”

- Feedbacks may and do shadow systematic machine problems
  → reduces additional safety margin and increases the dependence on them
  – acceptable to quickly advance and as temporary mitigation solution
  – logging of all feedback system actions used to
    - Improve static steady-state machine model
    - monitor and identify potential problems
    - facilitate feed-forwarding → reduce FB dependence/reliance (passive safety)

A Note on Dependence of Operation on Feedbacks

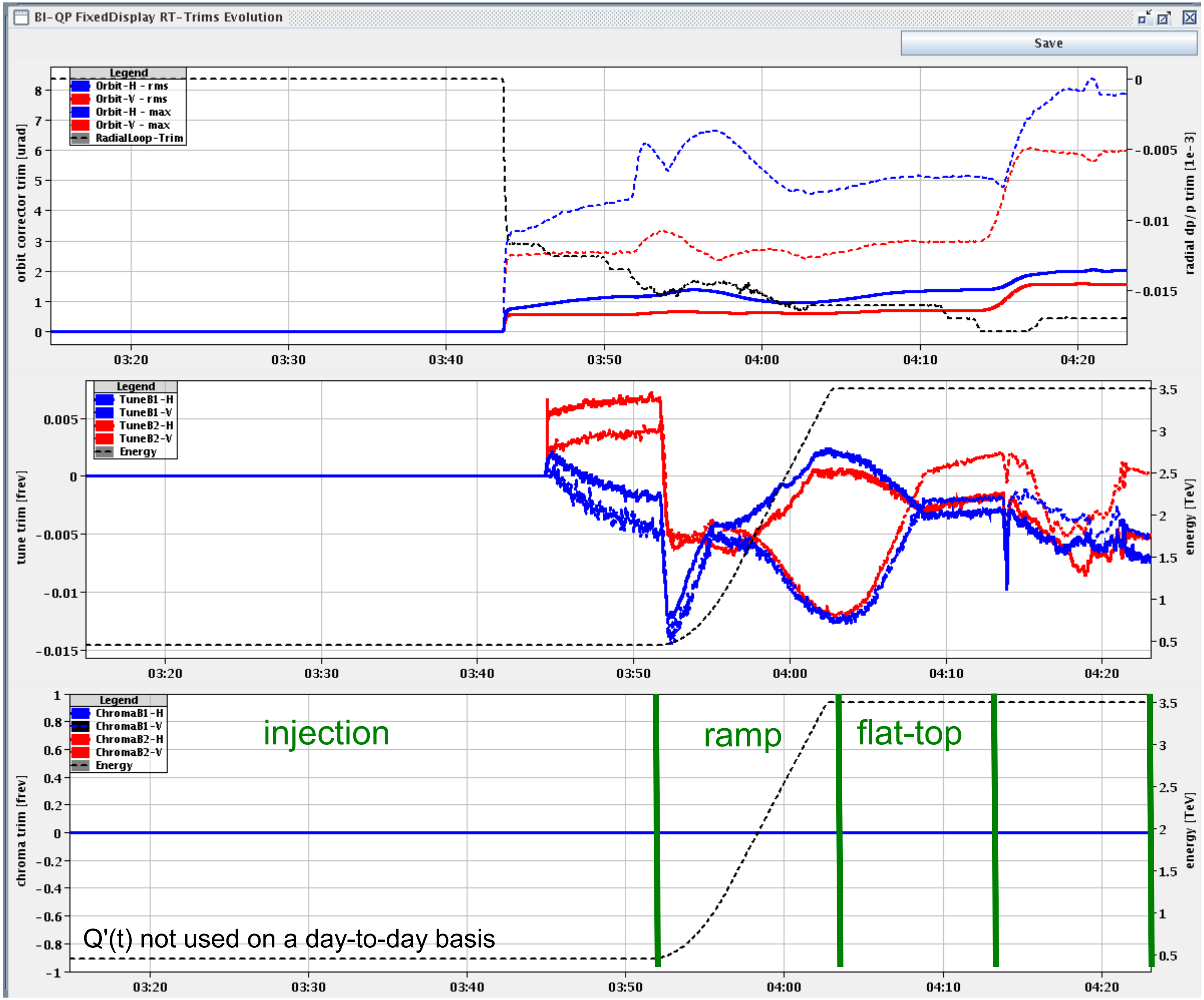
Example: Typical LHC Q/Q'(t) Control Room View



A Note on Dependence of Operation on Feedbacks

LHC Feedbacks in Action : Ramp & Squeeze

- Trims became de-facto standard to assess the FB and machine performance



β*-squeeze

Orbit-FB & Radial-Loop Trims (μrad)

Tune-FB trims

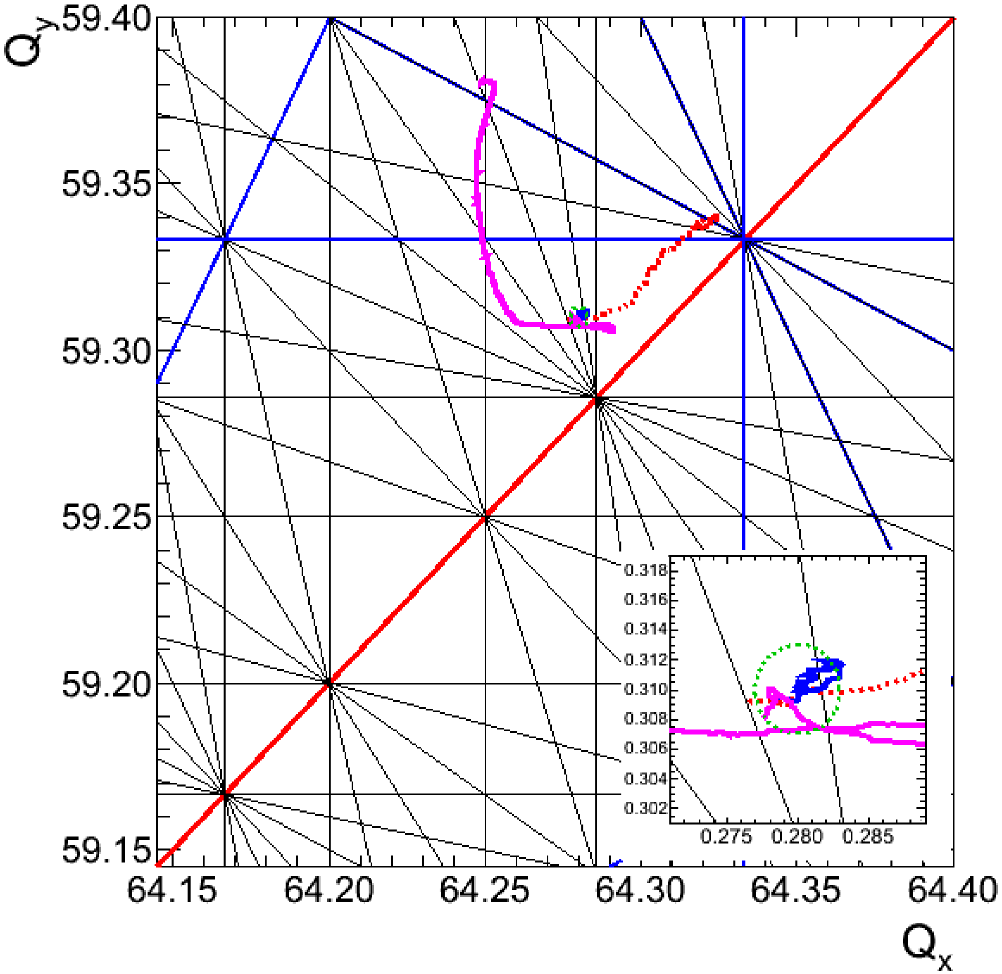
Q'(t)-FB trims

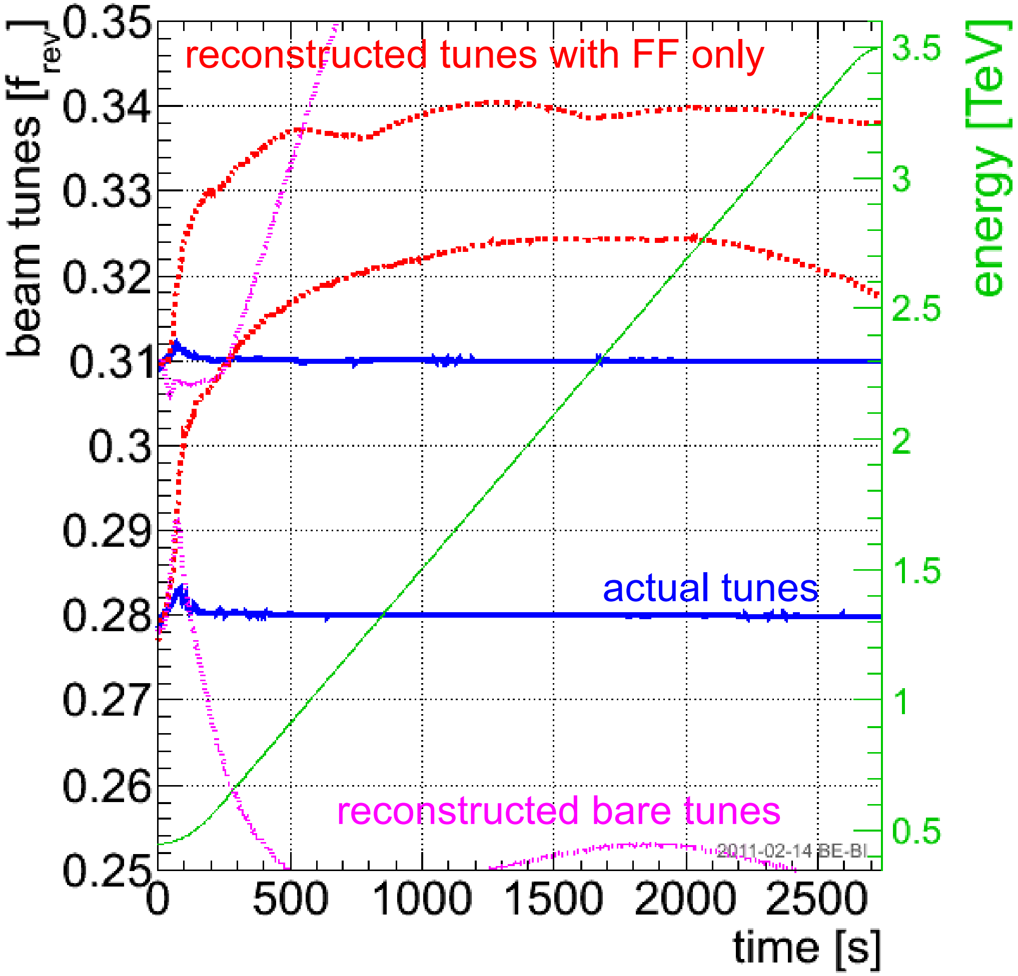
Energy (TeV)

A Note on Dependence of Operation on Feedbacks

'What-if-... Scenario' Analysis

- Tunes kept stable to better than 10^{-3} for most part of the ramp and squeeze



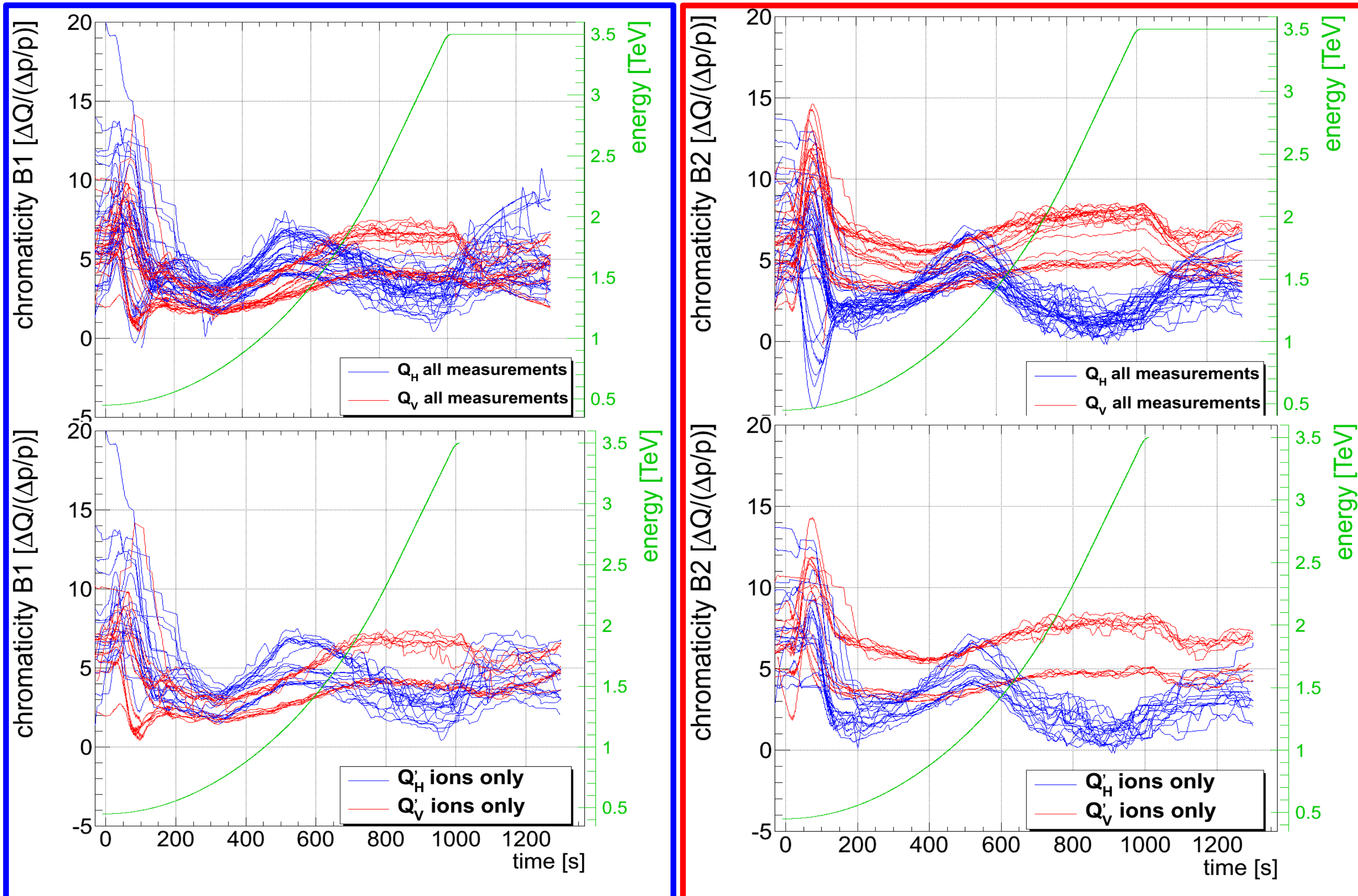


- Feed-forward errors during snap-back probably due to feed-down effects

A Note on Dependence of Operation on Feedbacks

Example LHC: Residual overall Chromaticity Stability

- Feed-forward of Q'(t)-Feedback signal for next fill turned out to be sufficient!
  - enforced by strict pre-cycling following physics, access or circuits 'off '...



Conclusion

- Beam-based FBs are remedies for perturbations on slow/medium time scales
  - limited by thermal drifts, noise and systematics of involved devices
  - Systematic and thorough analysis of involved beam instrumentation and corrector circuits is essential!
- Use of imperfect (design) beam response for SVD based FB systems:
  - does not affect the precision of the correction but reduces rather the effective bandwidth → favours higher feedback sampling frequencies
- Youla's affine parameterisation facilitates optimal adaptive non-linear control
  - enables gain-scheduling based on operational scenario
  - Ziegler-Nichols/Coohen-Coon PID tuning are outdated but sometimes still useful
- Beware of cross-constraints/coupling of simultaneous nested loops:
  - Feedbacks should be designed as an ensemble
- Feedback are designed to stabilise the beams but may equally drive instabilities in case of sub-system faults or changing beam/machine conditions
  - recommendation: consider adding system validation tests ('as good as new') as routine operation to your systems!

The author gratefully acknowledges contributions from many colleagues:

- A. Böhm (RWTH), J. Andersson, M. Anderson, A. Boccardi, T. Bohl, A. Butterworth, E. Calvo Giraldo, P. Cameron (BNL), R. Denz, K. Fuchsberger, M. Gasior, J. Fox (SLAC), W. Höfle, S. Jackson, L. Jensen, R. Jones, M. Jonker, J.M. Jouanigot, G. Kotzian, K. Kasinski (ex-BI), Q. King, K. Kostro, M. Lamont, T. LeFevre, S. Page, L. Ponce, V. Ranjbar (FNAL), G. Sivatskiy, S. Smith (SLAC), C.-Y. Tan (FNAL), E. Tedesco, D. Teytelman (Dimtel), J. Tückmantel, A. Verweij, J. Wenninger, D. Valuch, W. Venturini, T. Wijnands, M. Zuin (ex-IT), and many more
- Special thanks to our colleagues in the synchrotron-light-source-world colleagues: M. Böge, M. Boland, G. Decker, L. Nadolski G. Rehm, T. Schilcher, V. Schlott et al.

Thank you

for coming to PCaPAC2016 your interest in this feedback tutorial

Reserve Slides

References I/II

(1) Proceeding of “3rd International Workshop on Beam Orbit stabilization”, 2004, http://iwbs2004.web.psi.ch/

(2) M. Böge, “Achieving Sub-micron Stability in Light Sources”, EPAC'04, Lucerme, 2004

(3) G Decker, “Beam Stability in Synchrotron Light Sources”, DIPAC'05, Lyon, France, 2005

(4) O. Berrig et al., “The Q-Loop: A Function Driven Feedback System for the Betatron Tunes During the LEP Energy Ramp”, CERN SL-98-039, 1998

(5) A.Fisher et al., “Tune Feedback in PEP-II”,SLAC-PUB-10230, 2003

(6) P. Cameron et al., “Tune Feedback at RHIC”, PAC'01, Chicago, IL, 2001

(7) P. Cameron et al., “Measuring Chromaticity Along the Ramp Using the PLL Tune-Meter in RHIC”, EPAC'02, Paris, 2002

(8) R. Hettel, “Beam Stability Issues at Light Sources”, 25th ICFA Advanced Beam Dynamics Workshop, Shanghai, 2002

(9) Chamonix XIV, “Damage levels - Comparison of Experiment and simulation” and PAC'05

(10) R. Assmann, “Collimation and Cleaning: Could this limit the LHC Performance?”, Chamonix XII, 2003

(11) S. Redaelli, “LHC aperture and commissioning of the Collimation System”, Chamonix XIV, 2005

(13) R. Steinhagen, “Closed Orbit and Protection”, CERM, MPWG #53, 2005-12-16

(14) S. Fartoukh, O. Brüning, “Field Quality Specification for the LHC Main Dipole Magnets”, LHC Project Report 501

(15) S. Fartoukh, J.P. Koutchouk, “On the Measurement of the Tunes, [..] in LHC”, LHC-B-ES-0009, EDMS# 463763

(16) G. Golub and C. Reinsch, “*Handbook for automatic computation II, Linear Algebra”*, Springer, NY, 1971

(17) D. C. Youla et al., *“Modern Wiener-Hopf Design of Optimal Controllers”*, IEEE Trans. on Automatic Control,1976, Vol. 21-1,pp. 3-13 & 319-338

(18) B. Anderson, “From Youla-Kucera to Identification, Adaptive and Non-linear Control”, Automatica, 1998, Vol. 34, No. 12, pp. 1485-1506

(19) C.Y. Tan, “The Tevatron Tune Tracker PLL”, FERMILAB-TM-2275-AD

References II/II

(20) R.J. Steinhagen et al., “Influence of Varying Chromaticity on Robustness of the LHC Tune PLL and its Application for Continuous Chromaticity Measurement.”, PAC'07, 2007

(21) R.J. Steinhagen, “First Results of the PLL Tune Tracking in the SPS”, Tune Feedback Final Design Review, BNL, http://www.agsrhichome.bnl.gov/LARP/061024_TF_FDR/index.html

(22) R.J. Steinhagen et al., “The LHC Phase-Locked-Loop for Continuous Tune Measurement - Prototype tests at the CERN-SPS”, PAC, 2007

(23) M. Gasior, R. Jones, “The principle and first results of betatron tune measurement by direct diode detection”, CERN-LHC-Project-Report-853

(24) R. Jones et al., “Towards a Robust Phase Locked Loop Tune Feedback System”, DIPAC'05, Lyon, France, 2005

(25) T. Schilcher, “Commissioning and Operation of the SLS Fast Orbit Feedback”, EPAC'04, Lucerne/Switzerland, 2004

(26) K. Park, “Development of High Precision Magnet Power Supply using the DSP”, this proceedings

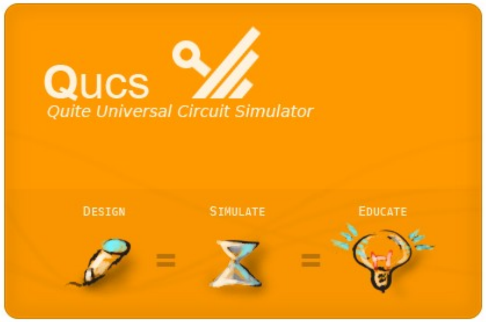
(27) Q. King, “The All-Digital Approach to LHC Power Converter Current Control”, ICALEPS'01, 2001

(23) R.J. Rushton, “Diamond Storage Ring Power Converters”, EPAC'06, Edinburgh, Scotland, 2006

(29) P. Gros, “The 3Hz Power Supplies of the SOLEIL Booster”, EPAC'06, Edinburgh, Scotland, 2006

Open Software

- Octave: http://www.gnu.org/software/octave/
- SciLab: http://www.scilab.org/scilab/gallery/xcos
- OpenModelica: https://www.openmodelica.org/
- QUCS – Quite universal circuit simulator: http://qucs.sourceforge.net/
  - similar to Spice™ & derivatives but open-source
  - DC, AC, S-parameter, harmonic balance analysis, noise analysis, RF structures, etc.

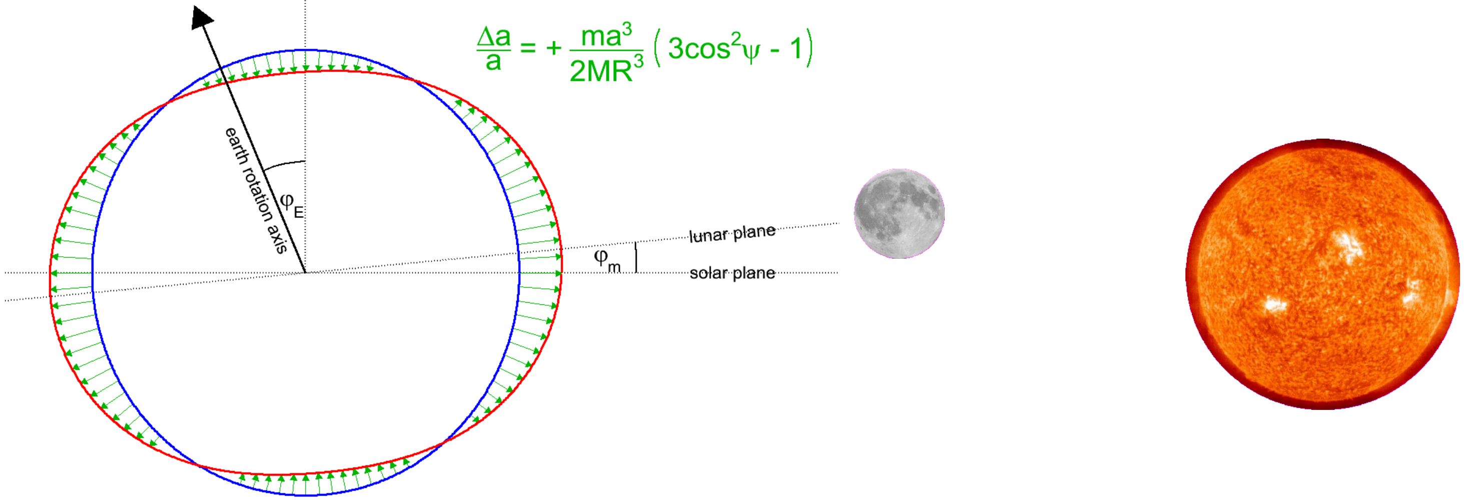


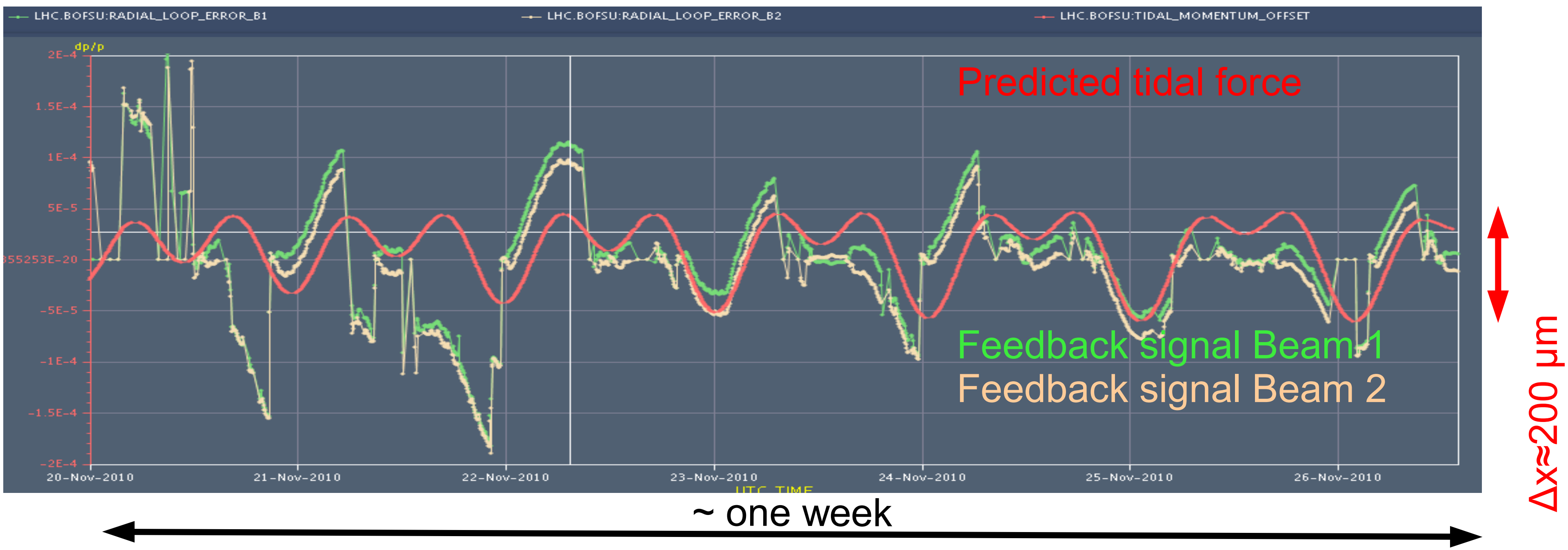
- Cool simple pole-zero simulation tools:
  - S-domain: http://web.mit.edu/6.302/www/pz/
  - Z-domain: http://www.micromodeler.com/dsp/
    - IIR/FIR filter design + C source code generation

FB Design Paradigms – Stability

Example: Earth Tides → Orbit Stability

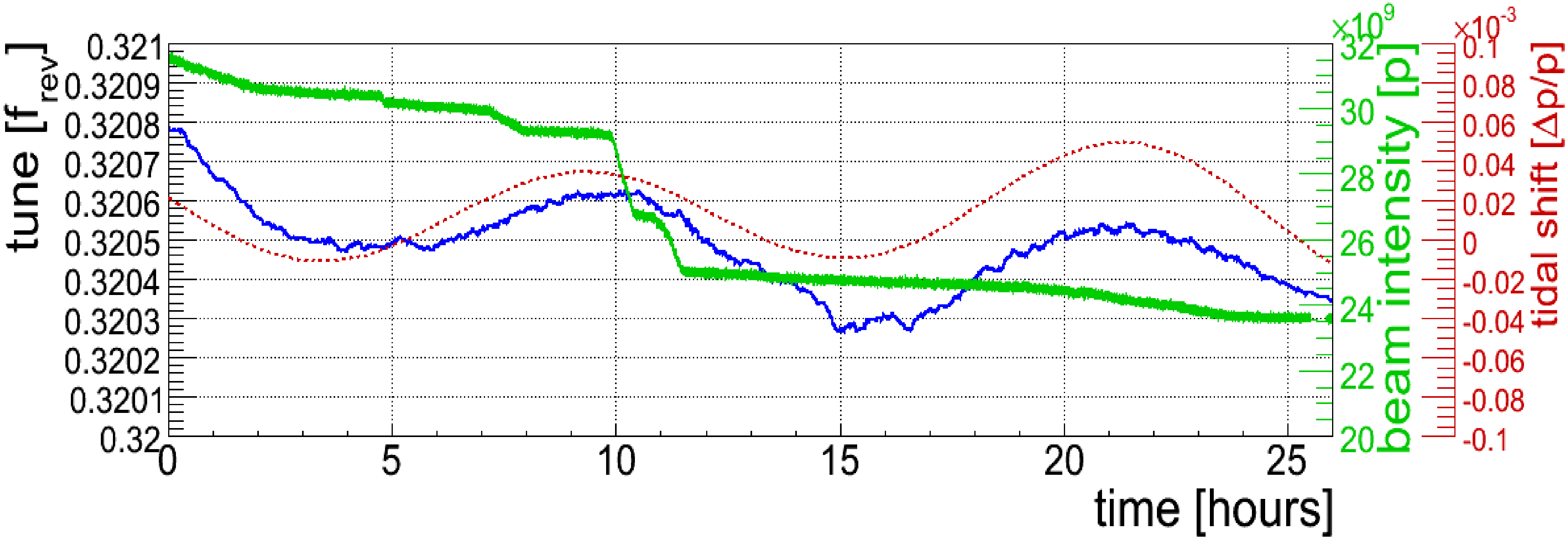
- Known effect from LEP → changes the machine circumference/energy





Quirky side effect:

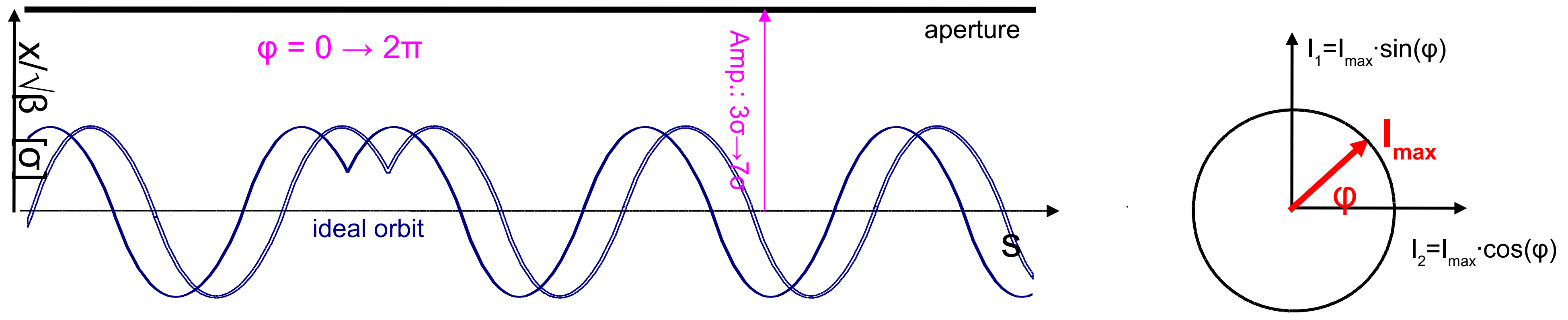
- Machine circumference changes are propagated via Q' also to



- Probably the slowest high-precision Q' measurement in the World
  - Short-Term Tune-Stability of ~10^{-6}!

Test of BPM Polarity, Mapping and Global Aperture

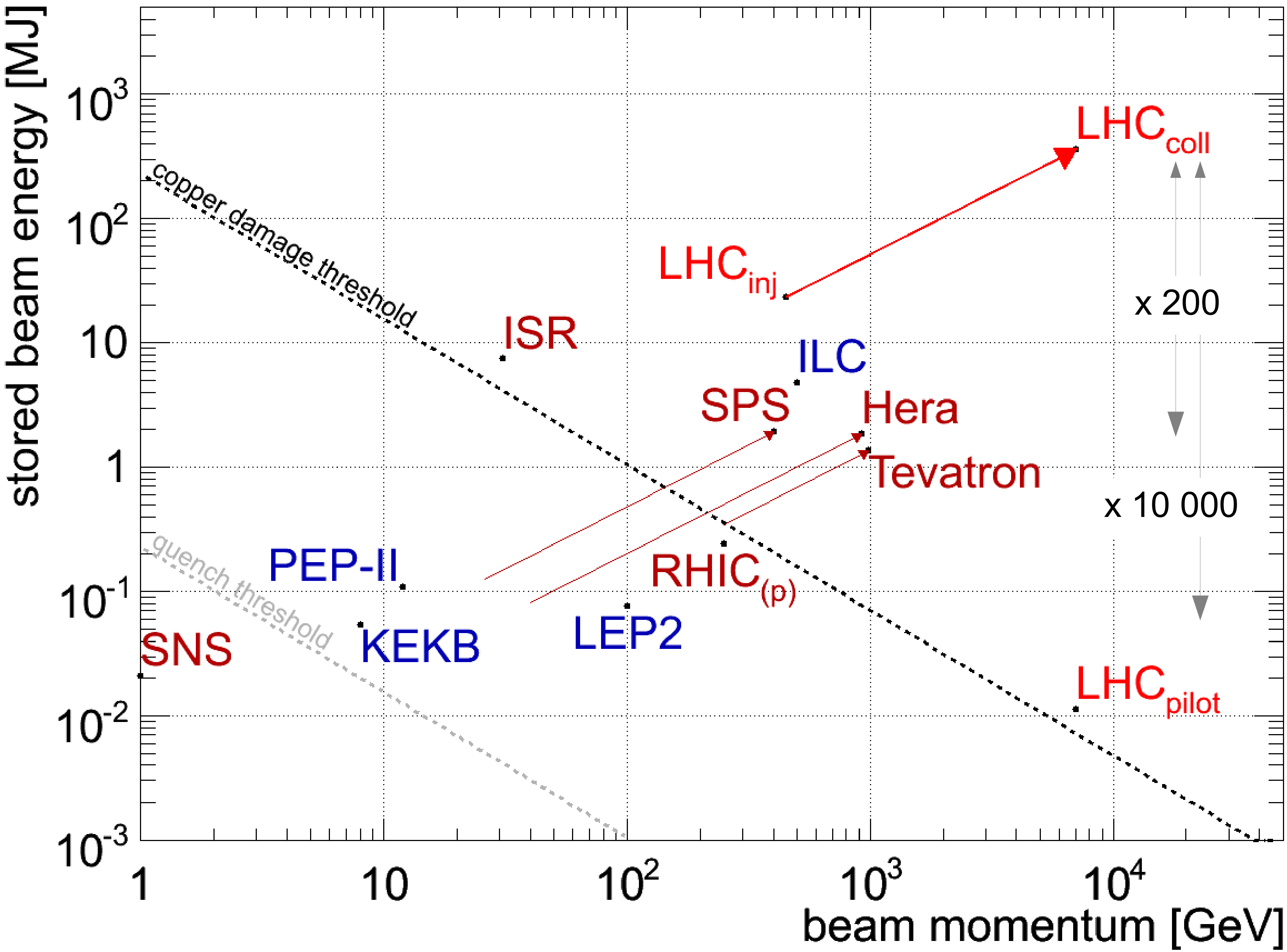
- Scan using two COD magnets (currents: I_1 & I_2) with π/2 phase advance:



  – Scan (assuming global aperture of ~ 7.5σ):
    - φ = 0→2π requires ~25 seconds @7σ, per transverse angle
    - propose to measure at four transverse angles: 0°, 45°, 90°, 125°
  – Increase amplitude (COD currents) till orbit shift ≈ 6.7σ
  – Loss does not exceed predefined BLM threshold if COD settings@ 6.7σ:
    - Yes: → mechanical aperture ≥ 6.7 s → orbit is safe
    - No: → mechanical aperture ≤ 6.7 s → orbit is un-safe
  – additional feature: compare measured with reference BPM step response (x_{co}= 0-3σ)

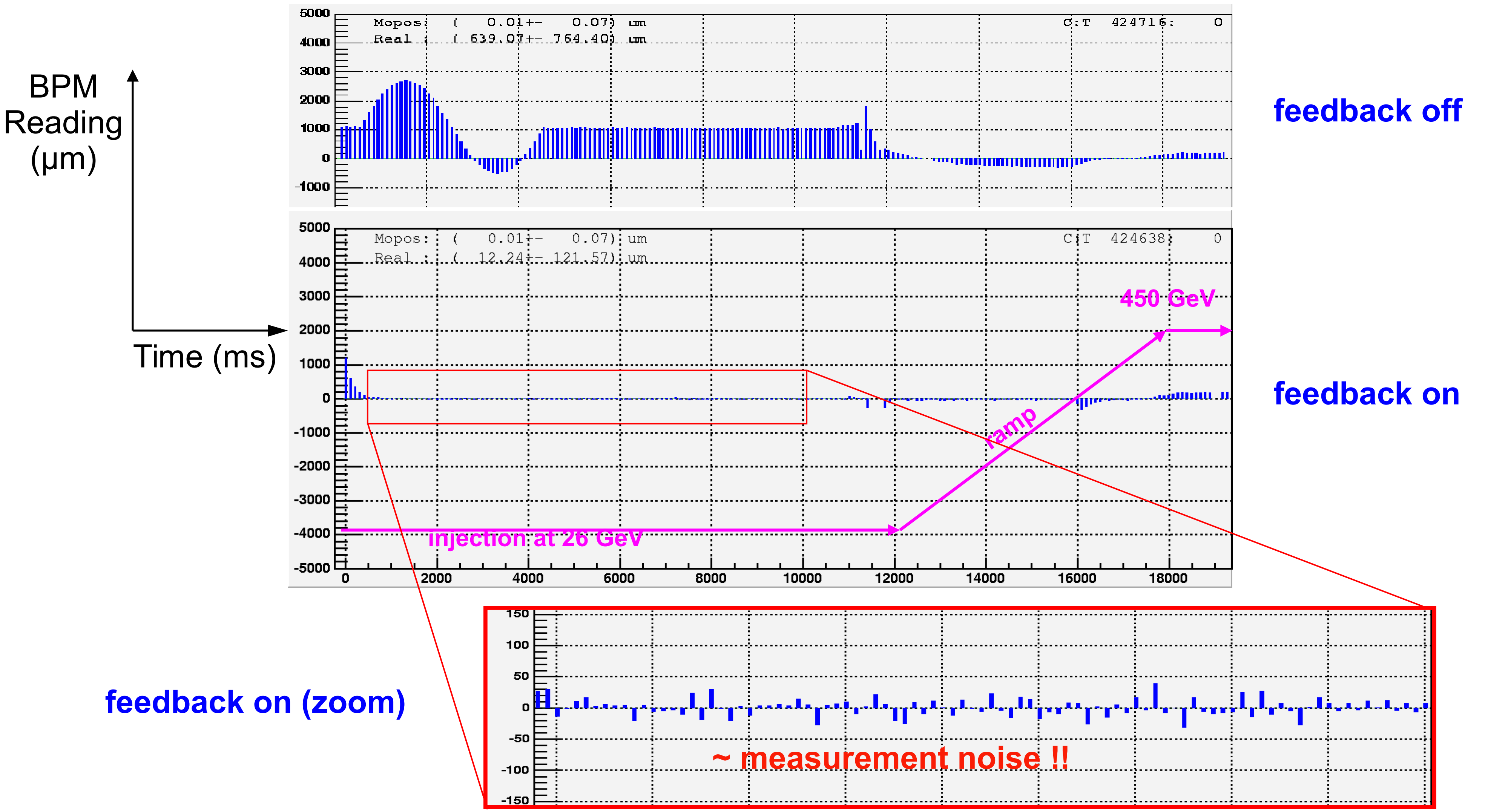
→ rough optics check (phase advance and beta-functions)

IWBS'04: “LHC is a pretty dangerous machine” Livingston Style plot



see Chamonix XIV: “Damage levels - Comparison of Experiment and simulation” and PAC'05 for details

LHC Orbit Feedback Test at the SPS I/II



Remaining Jitter Compensation: Fix Max Loop Delay

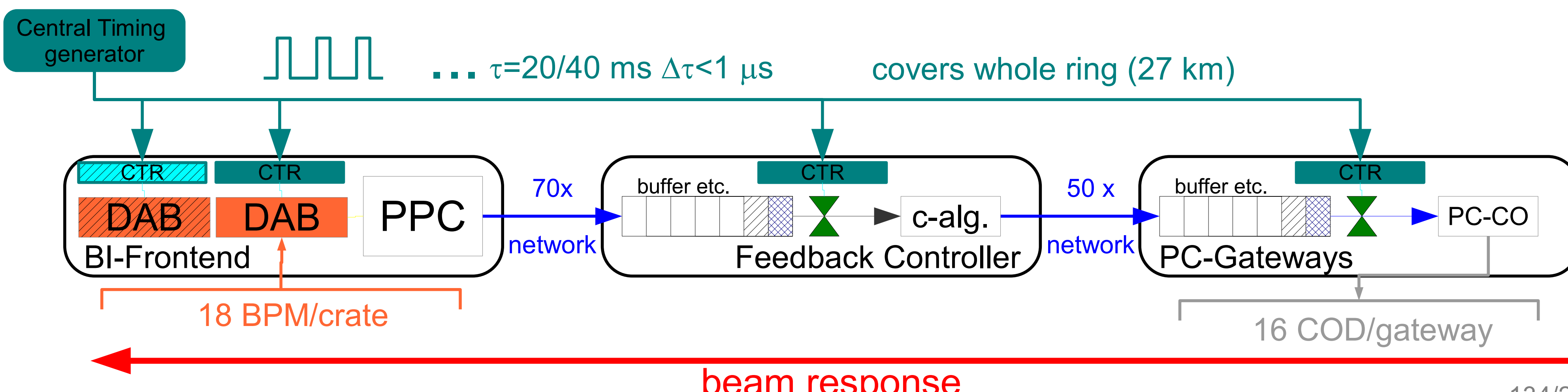
Two main strategies:

actual delay measurement and dynamic compensation in SP-branch:

- high numerical complexity, due to continuously changing branch transfer function
- only feasible for small systems

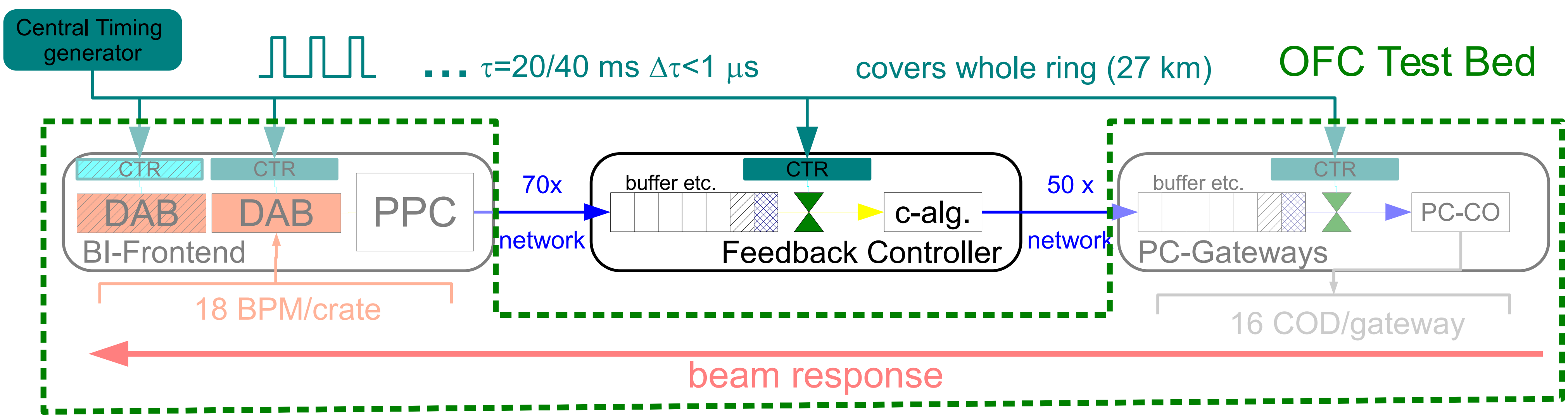
Jitter compensation using a periodic external signal:

- CERN wide synchronisation of events on sub ms scale that triggers:
  - Acquisition of BPM system, reading of receive buffers, processing and sending of data, time to apply in the power converter front-ends
- The total jitter, the sum of all worst case delays, must stay within "budget".
- Measured and anticipated delays and their jitter are well below 20 ms.
- feedback loop frequency of 50 Hz feasible for LHC, if required...



Commissioning the Orbit Feedback Controller – Test Bed

Test bed complementary to Feedback Controllers:
- Simulates the open loop and orbit response of COD→BEAM→BPM
  - Decay/Snap-back, ramp, squeeze, ground motion simulations, ...
  - Keeps/can test real-time constraints up to 1 kHz
- Same data delivery mechanism and timing as the front-ends
  - transparent for the FB controller
  - same code for real and simulated machine:
    - possible and meaningful "offline" debugging for the FB controller



Bottleneck I: Network in the high-level front-ends!

The front-end network interfaces are presently the bottleneck.

- e.g. feedback controller @ 50 Hz:

lots of in-/outbound connections:

- Two types of loads:
  - Real-Time: BPM and COD control data
    - Avg. bandwidth: ~13 Mbit/s
    - short bursts: full I/O load within few ms
      - resp. 1GBit/s, burst duration desired to be minimise the total loop delay)
  - Non-Real-Time:
    - transfer of new settings to OFC (matrix ~30 MB)
    - PID configuration etc.
    - relay of BPM and feedback data (monitoring/logging)
    - ...
- (Peak) load similar to high-end network servers
  - Nearly constant full load during certain operational phases

network interface should be scheduled on the device level to provide a Quality of Service (QoS) for real-time data

- One reserved FIFO queue for feedback data
- General purpose queue for other data

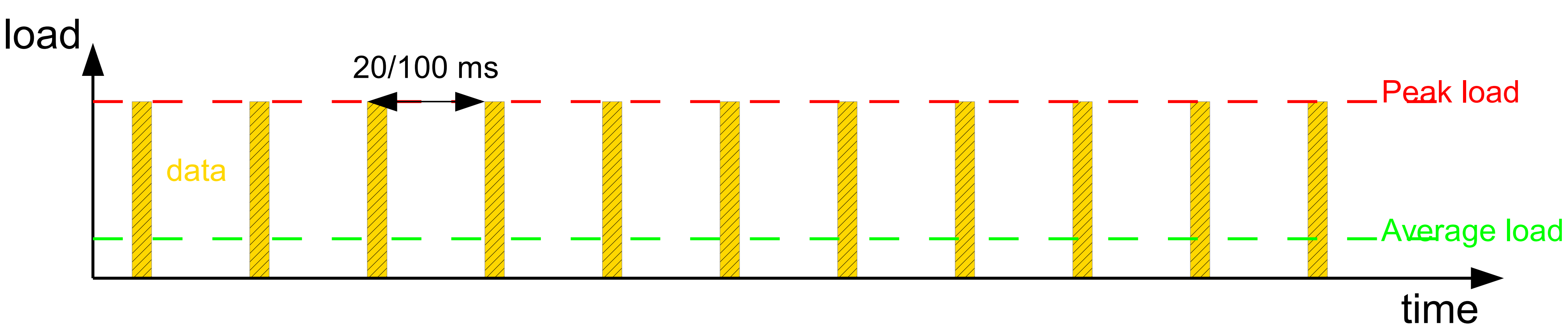
Bottleneck I: Network in Front-Ends: Data Rates

Hardware:
both rings covered by 1056 BPMs
Measure both planes (2112 readings)
Organised in front-end crates (PowerPC/VME) in surface buildings
- 18 BPMs (hor & vert) ⇔ 36 positions / VME crate
- 68 crates in total, 6-8 crates /IR

Data streams:
Average data rates per IR:

18 BPMs x 20 bytes+overhead	~	1500	bytes	/ sample / crate
1056 BPMs x 20 byte	~	94	kbytes	/ sample
@ 10 Hz:	~	7.7	Mbit/s	
@ 50 Hz:	~	38.4	Mbit/s	

Peak data rates (bursts): 100Mbit/s resp. 1Gbit/s (depending on Ethernet interface)



Context and Legacy of Earlier FB Reviews

- 2003: Initial Orbit-FB Prototype tests at SPS – main outcome:
  - Feasible for LHC established (tested up to f_s = 100 Hz) → to be deployed 2007
  - criticality of real-time latencies on the network and host operating system
  - Need for handling input & output errors (measurement data quality)
- 2003: Orbit Feedback Workshop → LTC: established architecture
- 2004: Stabilisation workshop in Grindlewald:
  LHC Orbit-FB more similar to those in SL-Sources
- 2005: Formalised Orbit-FB Specification (LHC OP Meeting #40)
- 2006: Chamonix XV (Spring): Architecture extended by Tune-FB & FBs on the roadmap for LHC commissioning
- 2006: LHC Commissioning WG: Review on FB Architecture
  - "[..] Biggest problem so far for LHC feedbacks: Human resources to implement the FB controller, service unit, GUIs, … [..]"
- 2006: Tune-FB Final Design Review (Autumn, CERN & US-LARP), OFSU
- 2007: LHC Commissioning WG: Status Update & Commissioning Plans
- 2007-10: LHC-CWG: Reviewed detection of LHC BPM errors and faults
- 2007-12: Ditanet WS on Q/Q' Diagnostics: … yet another review

Context and Legacy of Earlier FB Reviews – Cont.

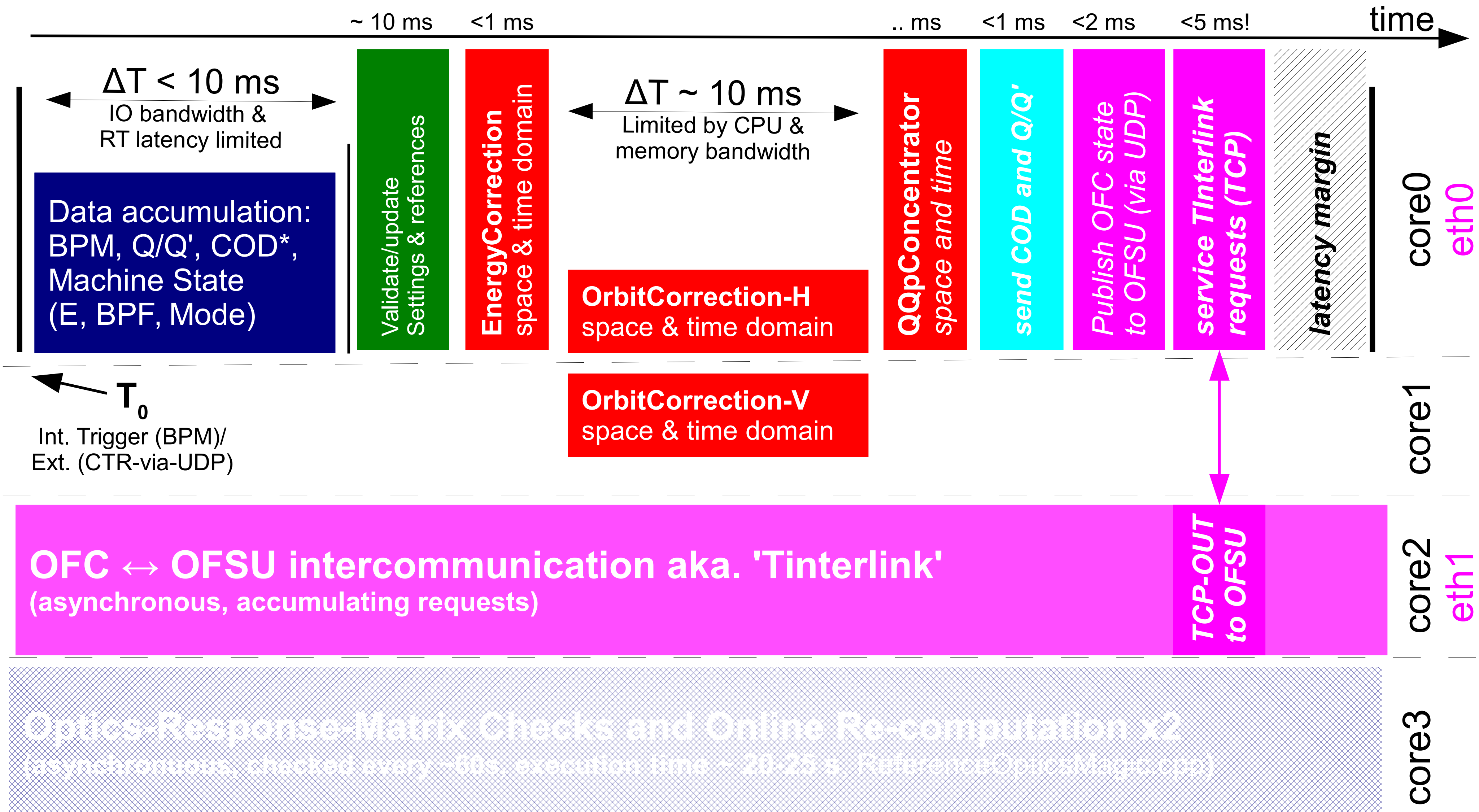
- 2008-03: LTC Summary & Review: LHC Q/Q' Diagnostics & FBs
- 2008-09: AB Seminar on LHC Feedbacks
  - for those who never heard of FBs (repeated in 2009)
- 2009-10: BI-Technical Board on LHC Feedbacks
- 2010-10: LHC First Tune-FB Ramps results
- 2010-06: MPS Review: Impact of FBs on Machine Protection
  - Identified previously not-handled issues (timing/energy telegrams, rogue packets, measurement quality, QPS cross-talk → solve non-FB specific issues at source)
- 2011-12: Internal BI review on OFC/OFSU software architecture
- 2012-03: LMC: Update on Orbit- & Tune-FB modifications
- 2013: MP Review: Experiences with FBs and foreseen Improvements for LS1

Some references:

- http://cern.ch/AB-seminar/talks/AB.Seminar.rst.pdf (CERN-AB-2007-049)
- http://lhccwg.web.cern.ch/lhccwg/Meetings/2007/2007.10.23/2007-10-23_LHCCWG-FAULTY_BPM.pdf
- http://accelconf.web.cern.ch/AccelConf/PAC2011/talks/weobn2_talk.pdf &
- http://accelconf.web.cern.ch/AccelConf/PAC2011/papers/weobn2.pdf
- LHC-BPM-ES-0004 rev. 2.0, EDMS #327557, 2002,
- svn+ssh://svn.cern.ch/reps/acco-co/trunk/lhc/lhc-feedbacks

Specific Orbit Feedback Controller (OFC) Structure III/IV

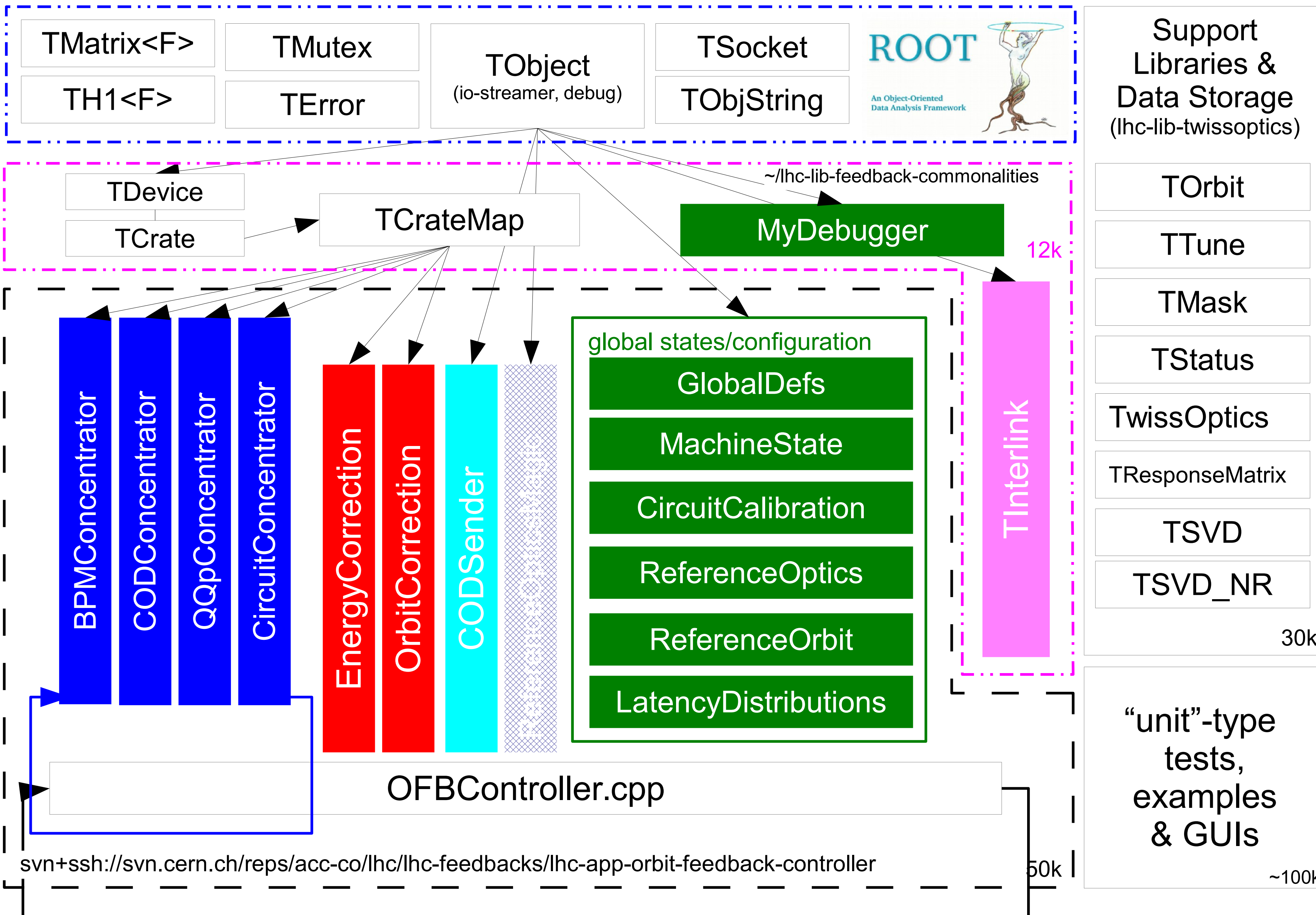
- Functional structure, timing diagram & core utilisation (CPU shielding):



- main performance limits: RT latencies, CPU/RAM & asynch. tasks

Specific Orbit Feedback Controller (OFC) Structure IV/IV

- Fairly flat C++ Class Hierarchy ↔ reflects io-streaming task:



Re-working, re-optimisation is inefficient and costly

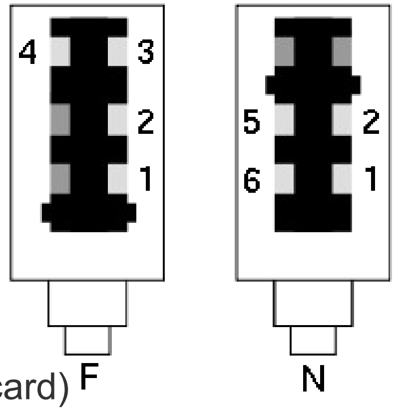
Follow a long-term strategy and 'lean principles':

- Continuous improvement
  - Right processes to produce right results and for getting it right the first time
    - *commissioning procedures as evolving operation standard*
    - system integration: definition of what, how and when (prioritisation) is needed
  - Prevention of inefficiencies, inconsistencies & waste by design
    - 'poka-yoke' or 'error proofing' principle – culture of stopping and fixing
      1. early, when and where they occur (at the source)
      2. with low-intensity beam rather than with high-intensity beam
      3. addressing first basic parameters before complex higher-order effects
    - Example #1: first fix injection, trajectory, orbit, Q/Q' before addressing space-charge or slow-extraction problems
    - Example #2: important losses for low-intensity beam have larger impact for high-intensity beam (↔ activation)

- Respect for people – "develop people, then build products"→ talk by S. Reimann
  - optimise operation ↔ smart tools & procedures, e.g. beam-based feedbacks, sequencer, …
    - automate routine task so that operator talents are utilised and focused on more important tasks
  - training, investment in and development of people – minimise overburden/strain of personnel

Poka-Yoke (ポ カ ヨ ケ) – 'Mistake-Proofing'

- Origin:
  - to avoid (yokeru) inadvertent errors (poka)
  - industrial processes designed to prevent human errors
    - Concept by Shigeo Shingo: 'Toyota Production System' (TPS, aka. 'lean' systems)
  - minimise common mistakes, procedural errors, etc. affecting machine performance and machine protection



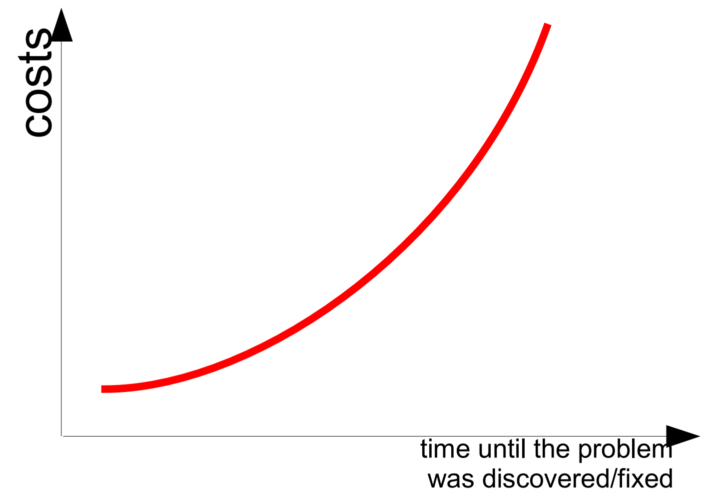
- Real-World Examples:
  - Polarity protection of electrical plugs (e.g. phone, Ethernet cable)
  - Procedures:
    - e.g. ATM machine: need to retrieve card before money is released (↔ prevents missing card)

- Respect for people – “develop people, then build systems”
  - optimise operation ↔ smart tools & procedures, e.g. beam-based feedbacks, sequencer, …
    - automate routine task so that operator talents are utilised and focused on more important tasks
  - training, investment in and development of people – minimise overburden/strain of personnel

Poka-Yoke (ポカヨケ) – 'Mistake-Proofing'
Reaction-Time and Cost → "fix" errors early

Fix problems early, when and where they occur

- Minimises procrastination of errors: "Safety starts with safe habits"!
  - big losses with big intensities → bad (activation)
  - large losses with small intensities → probably OK? … No!
    - requires paradigm change!
  - Interdependence between beam parameter & systems



- Early indication of developing/not-yet-critical faults:
  - Post-Mortem analysis ('as good as new' SIL assurance)
  - Preventative maintenance
  - fix "domino effect" problems at the source not its symptoms
    - e.g. fix problems with low-intensity beam rather than with high-intensity beam (avoids revalidation of loss patterns, MPS setup, …)
    - e.g. fix basic accelerator parameters before moving on to higher-order effect (e.g. extraction/injection energy/trajectory → orbit → tune → chromaticity → optic → … → driving term s