Transformers for field compensation

In the CERN SPS main magnets

Anton Lu

V. Kain, C. Petrone, V. Di Capua, M. Bonora, M. Schenk,
A. Bellelli, M. Taupadel, A. Ganesh

BE-CSS-DSB Data Science for Beam Operation
TE-MSC-MMM Magnetic Measurement and Modeling

Outline

- Hysteresis in accelerator magnets and impact on operation in CERN injectors
- Transformers for hysteresis prediction
  - Pre-training and transfer learning
- Validation, and Operational results in the CERN SPS
- Outlook and conclusions

Hysteresis in the accelerator magnets

And the need for reproducible fields

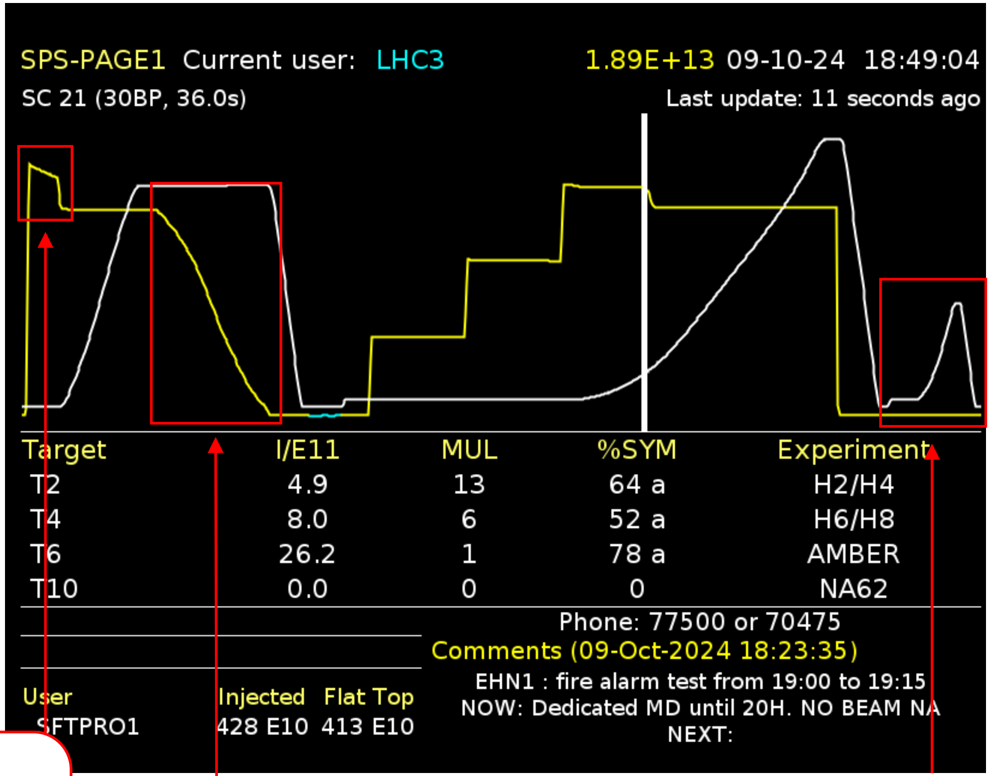
- The CERN SPS is a fast multi-cycling synchrotron
  - 3.6 - 24+ seconds long cycles, producing proton and ion beams at 30 GeV – 450 GeV
  - Cycle / magnetic sequences are often consistent …
  - … but manual changes to beam parameters are often required when sequences change

! Only CERN PS and PSB dipoles have field feedback

- In multi-cycling synchrotrons at CERN
  - Degaussing cycles, constrained cycle sequences **and** beam degradation due to hysteresis

- In general, operations treat hysteresis through symptom management without addressing the problem
  - Hysteresis is not a problem unique to CERN!

! Feed-back control requires online high-precision field measurements for each magnetic circuit



ion losses
dy current
cay
steresis

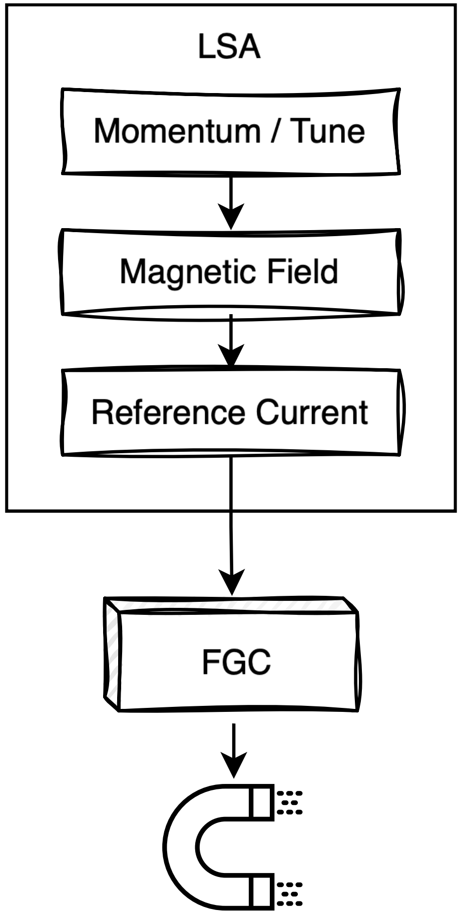
(Extracted) beam quality degradation
- Changing spill macrostructure

Waste of energy
- 5 kWh every cycle
- … for quasi-degauss
- … and for reproducible eddy current decays

What if we could have reproducible fields…

Through feed-forward field compensation

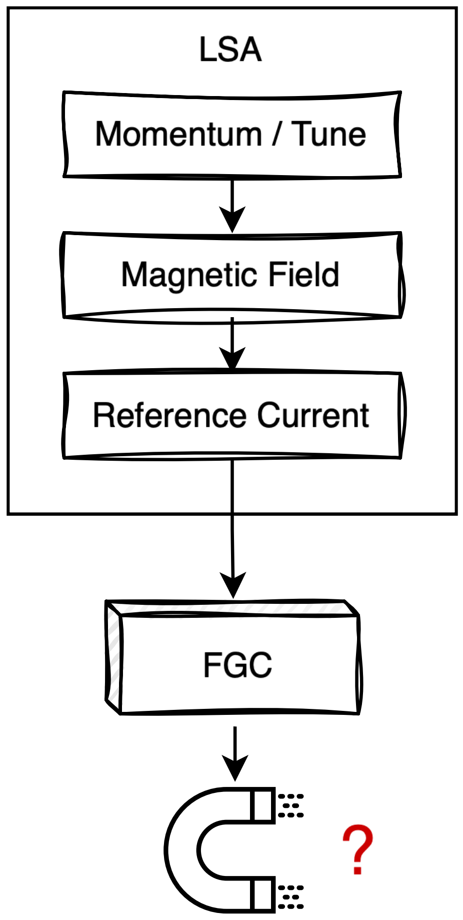
- Most magnetic circuits are controlled in current by translating momentum / tune / correction etc.



What if we could have reproducible fields…

Through feed-forward field compensation

- Most magnetic circuits are controlled in current by translating momentum / tune / correction etc.
  - Control system is agnostic to actual field response in the machine

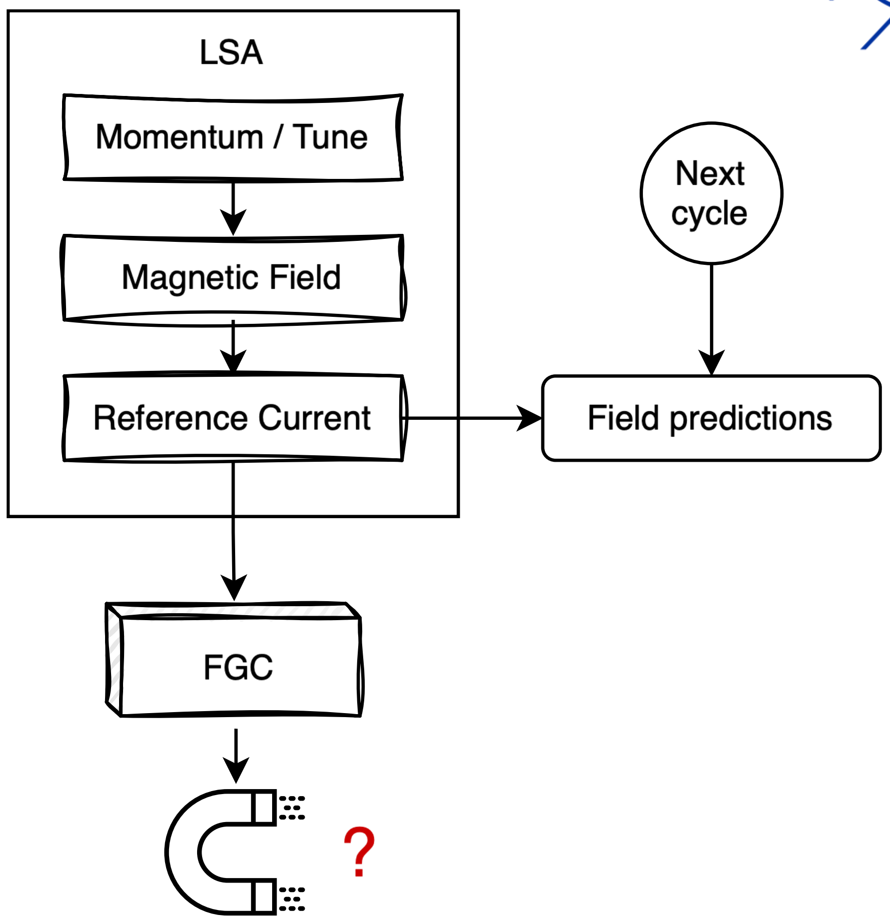


What if we could have reproducible fields…

Through feed-forward field compensation

- Most magnetic circuits are controlled in current by translating momentum / tune / correction etc.
  - Control system is agnostic to actual field response in the machine

- Instead: model magnetic field response $I \rightarrow B$ from magnetic field measurements

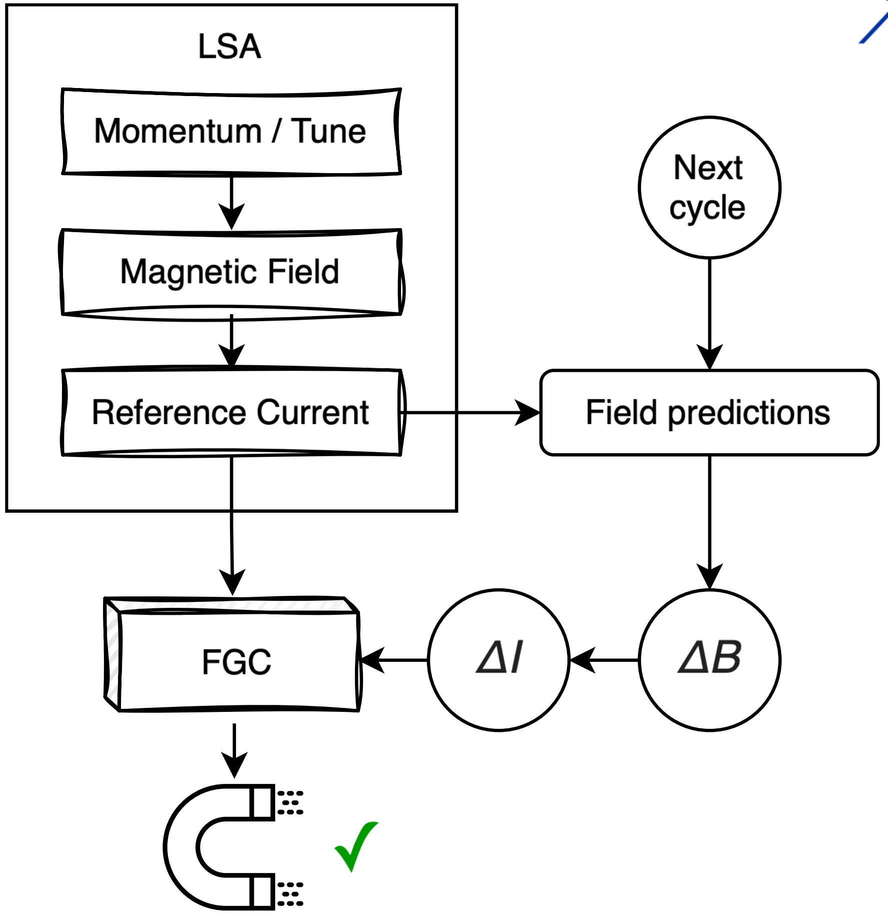


What if we could have reproducible fields…

Through feed-forward field compensation

- Most magnetic circuits are controlled in current by translating momentum / tune / correction etc.
  - Control system is agnostic to actual field response in the machine

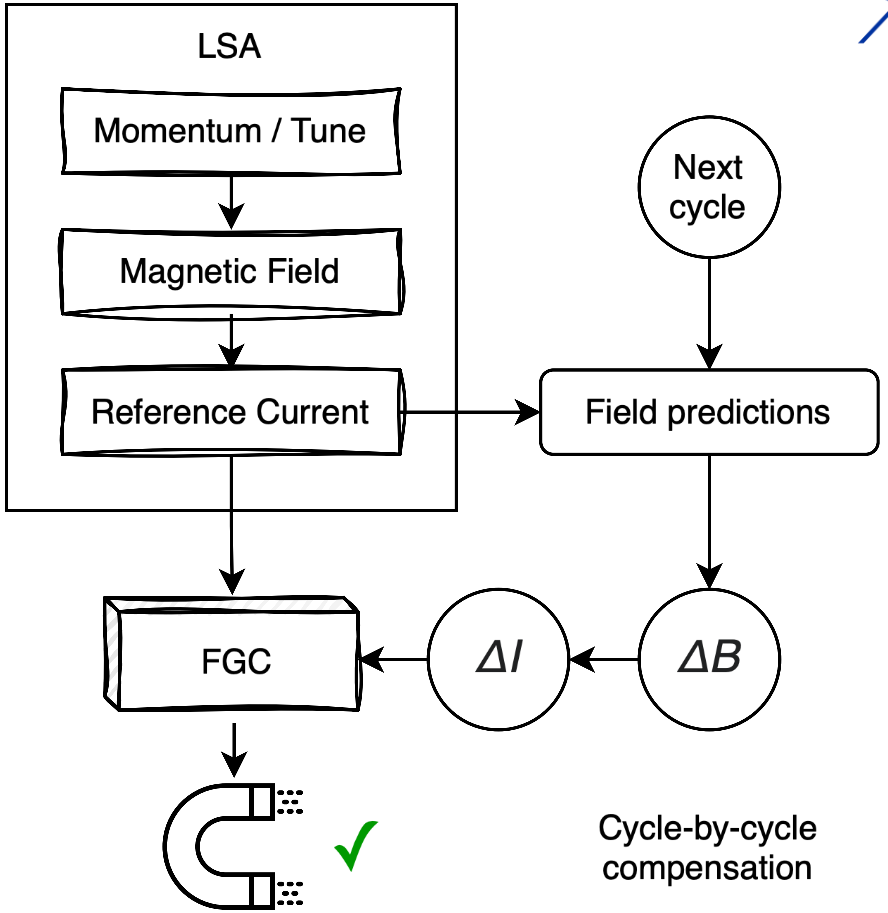
- Instead: model magnetic field response $I \rightarrow B$ from magnetic field measurements
  - Knowing next cycle to be played …
  - … feed-forward correct the field by applying a ΔI



What if we could have reproducible fields…

Through feed-forward field compensation

- Most magnetic circuits are controlled in current by translating momentum / tune / correction etc.
  - Control system is agnostic to actual field response in the machine
- Instead: model magnetic field response $I \rightarrow B$ from magnetic field measurements
  - Knowing next cycle to be played …
  - … feed-forward correct the field by applying a ΔI **for every cycle**
  - ⇒ We now can achieve reproducible fields
  - Control paradigm is transparent to set B / K

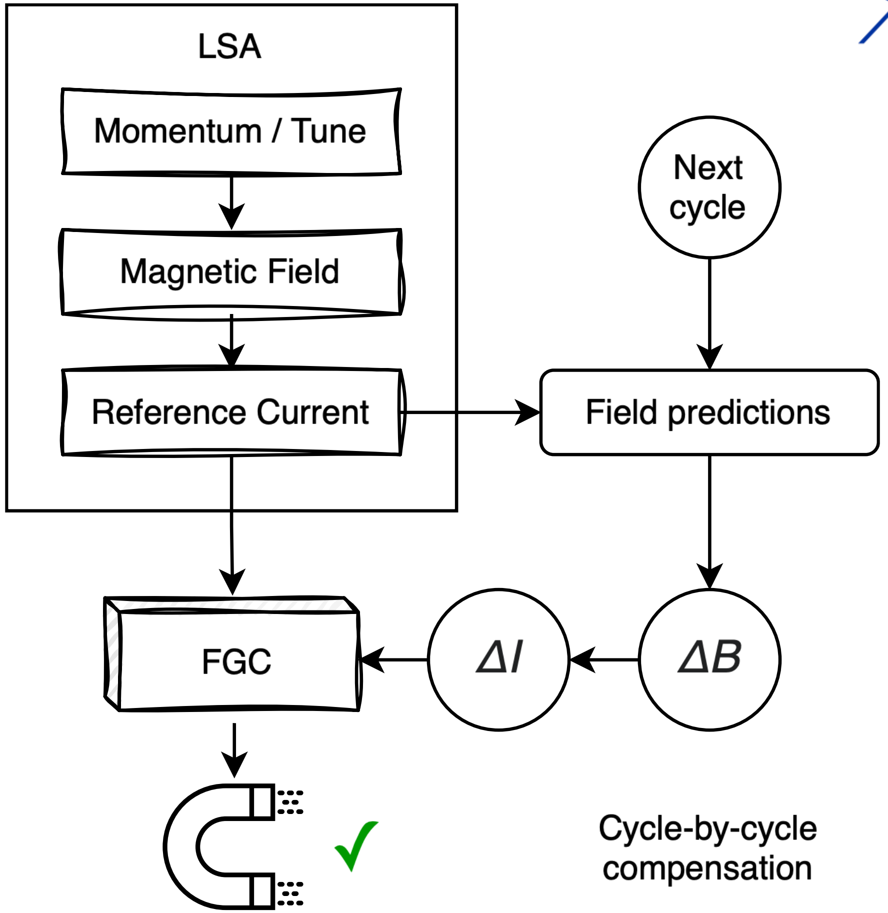


What if we could have reproducible fields…

Through feed-forward field compensation

- Most magnetic circuits are controlled in current by translating momentum / tune / correction etc.
  - Control system is agnostic to actual field response in the machine

- Instead: model magnetic field response $I \rightarrow B$ from magnetic field measurements
  - Knowing next cycle to be played …
  - … feed-forward correct the field by applying a ΔI **for every cycle**
  - ⇒ We now can achieve reproducible fields
  - Control paradigm is transparent to set B / K

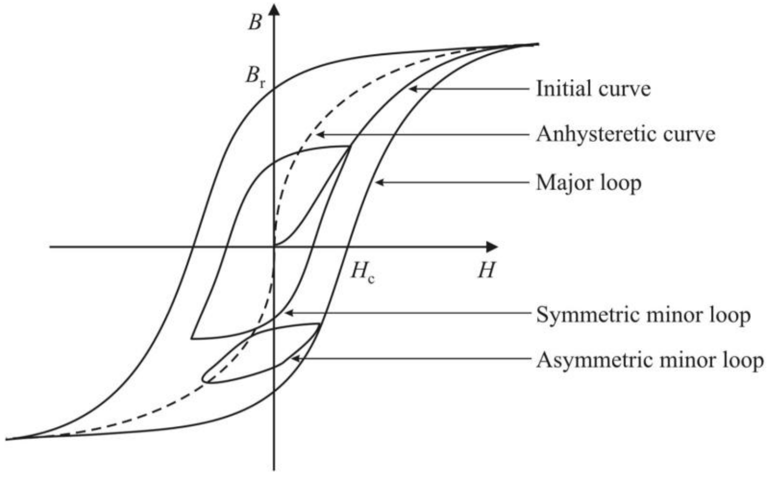


! N.B. Assume that all effects can be modeled by measuring field

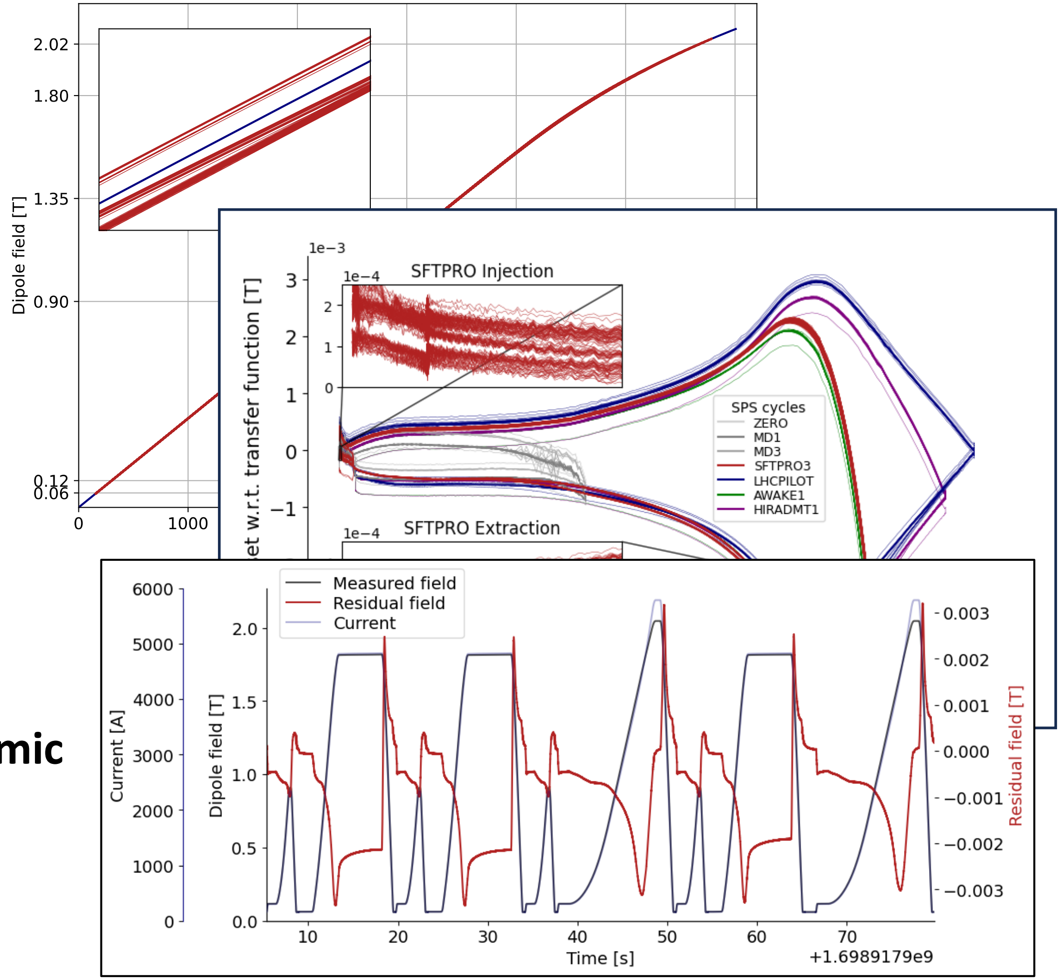
Hysteresis in the SPS main dipoles

Online measurements from the SPS

Expectations



Reality

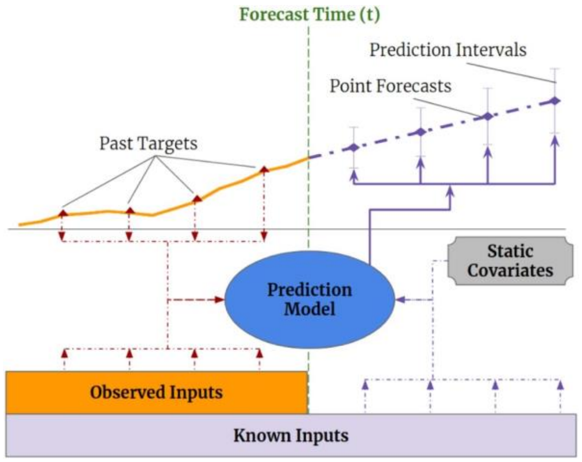


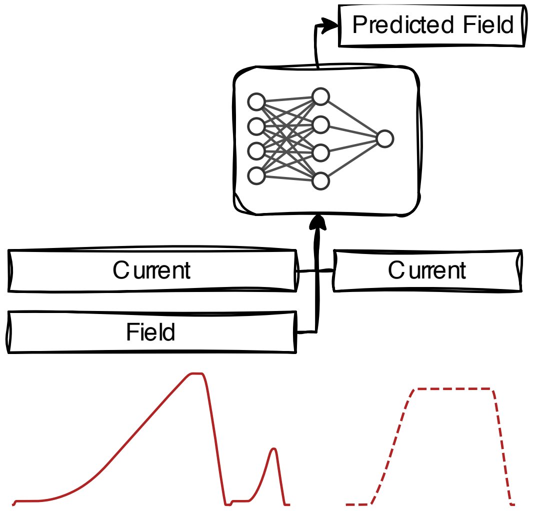
- Hysteresis and nonlinear effects are orders of magnitude smaller than anhysteretic magnetization
  - But measured field contains a mixture of **hysteresis, dynamic effects**, and measurement artifacts and errors
  - ⇒ fully data driven is the most straightforward way

Transformers for Field (Hysteresis) Prediction

Temporal Fusion Transformer – or simply LSTM + attention

- Encoder-Decoder style **time series models** are naturally suited for field-prediction task
  - › But long context and prediction windows (≈ 1000) prove challenging w.r.t. classic time series forecasting
  - › Models can be trained to predict every cycle in 1 pass (variable length prediction window)



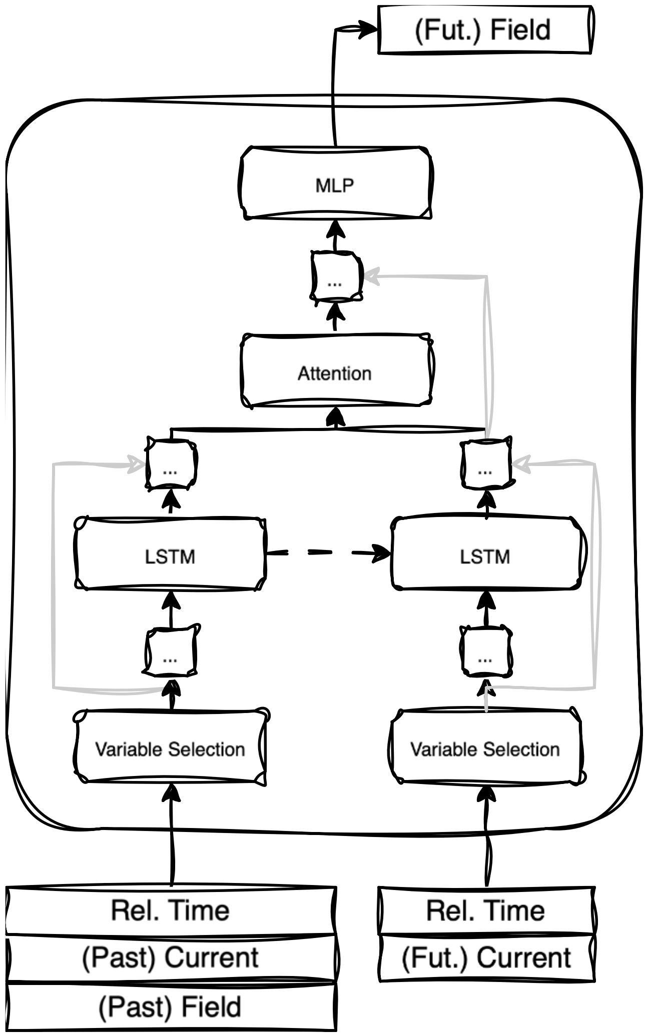


Transformers for Field (Hysteresis) Prediction

Temporal Fusion Transformer – or simply LSTM + attention

- Encoder-Decoder style time series models are naturally suited for field-prediction task
  - But long context and prediction windows (≈ 1000) prove challenging w.r.t. classic time series forecasting
  - Models can be trained to predict every cycle in 1 pass (variable length prediction window)

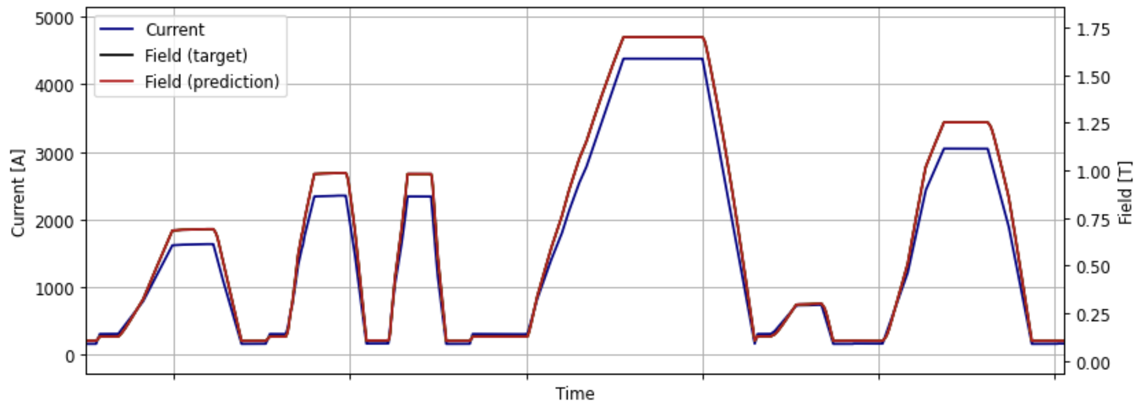
- Transformer-based models with attention are much more powerful than classic LSTMs for long-term dependencies
  - **Temporal Fusion Transformer** combines LSTM with self-attention, strong gating FC-NNs and residual connections …



Fitting transformers on simulated hysteresis

An interlude…

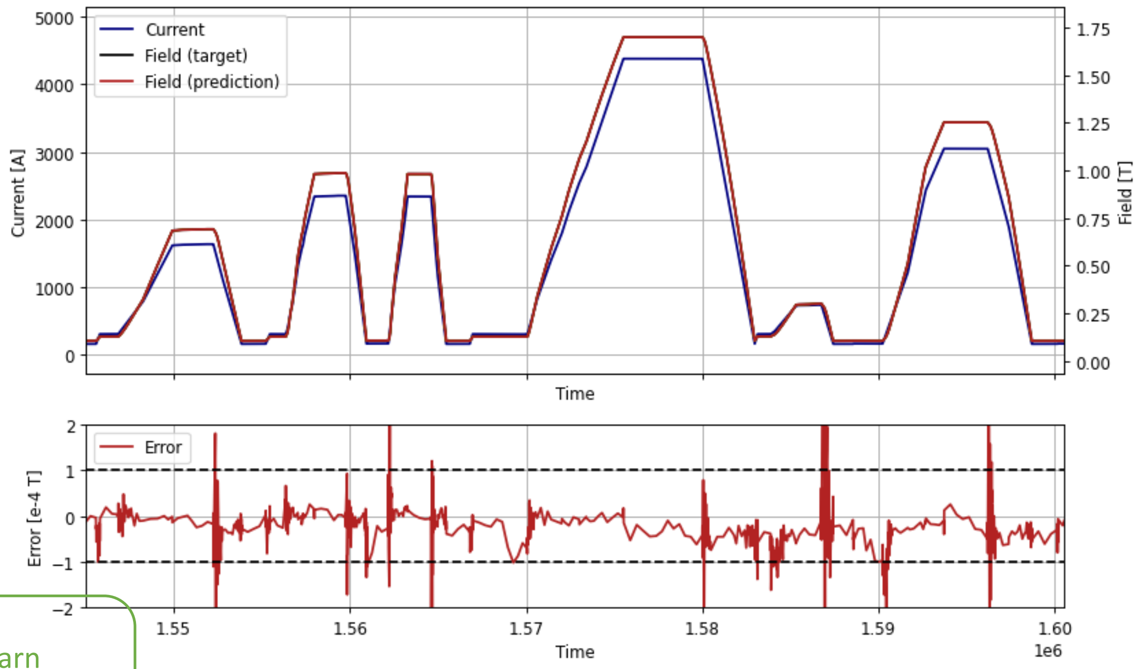
- Attention-based models need plenty of data to train and generalize well
  - Field measurements from operation can impossibly provide sufficient variety in data
- Large amounts of pseudo-realistic, pseudo-random waveforms can be generated and then simulated
  - Jiles-Atherton and Flatley ODEs have implementations in Python / MatLab
  - Pre-training learns common waveform shapes and generic hysteretic + dynamic behavior



Fitting transformers on simulated hysteresis

An interlude…

- Attention-based models need plenty of data to train and generalize well
    - Field measurements from operation can impossibly provide sufficient variety in data
- Large amounts of pseudo-realistic, pseudo-random waveforms can be generated and then simulated
    - Jiles-Atherton and Flatley ODEs have implementations in Python / MatLab
    - Pre-training learns common waveform shapes and generic hysteretic + dynamic behavior

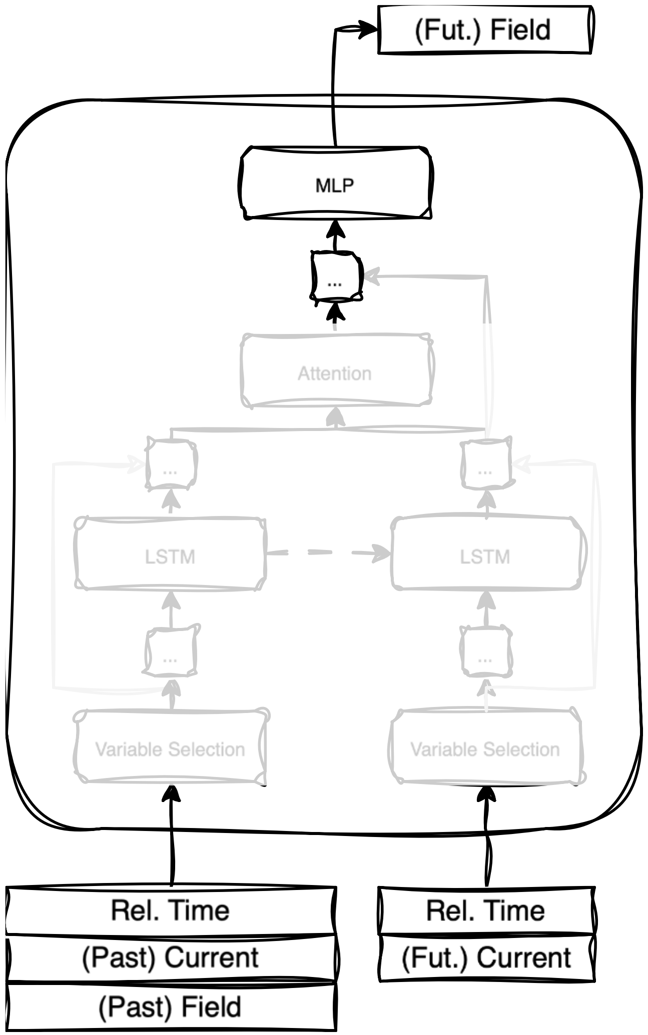


✓We can learn clean hysteretic data sufficiently well!

Transfer learning strategy for real hysteresis prediction

Making use of already trained models

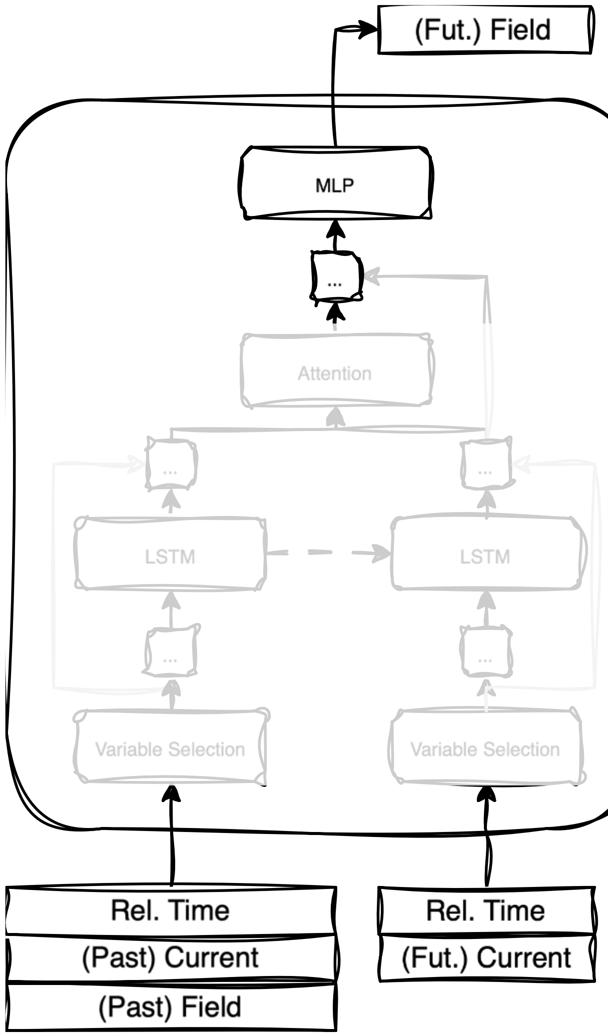
- Attention-based models need plenty of data to train and generalize well
  - Field measurements from operation can impossibly provide sufficient variety in data
- Large amounts of pseudo-realistic, pseudo-random waveforms can be generated and then simulated
  - Jiles-Atherton and Flatley ODEs have implementations in Python / MatLab
  - Pre-training learns common waveform shapes and generic hysteretic + dynamic behavior
- 2 step transfer learning
  1. Finetune the network head with small $lr \approx 10^{-5}$



Transfer learning strategy for real hysteresis prediction

Making use of already trained models

- Attention-based models need plenty of data to train and generalize well
  - Field measurements from operation can impossibly provide sufficient variety in data
- Large amounts of pseudo-realistic, pseudo-random waveforms can be generated and then simulated
  - Jiles-Atherton and Flatley ODEs have implementations in Python / MatLab
  - Pre-training learns common waveform shapes and generic hysteretic + dynamic behavior
- 2 step transfer learning
  1. Finetune the network head with small $lr \approx 10^{-5}$
  2. Unfreeze full model and train with higher lr

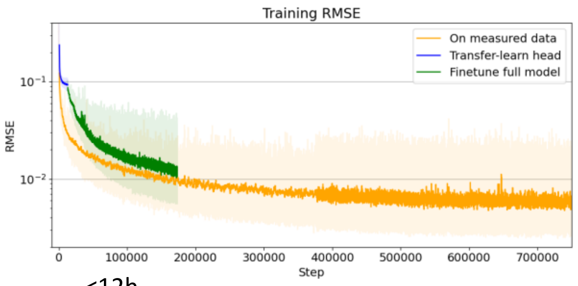


Finetuned transformer model predictions

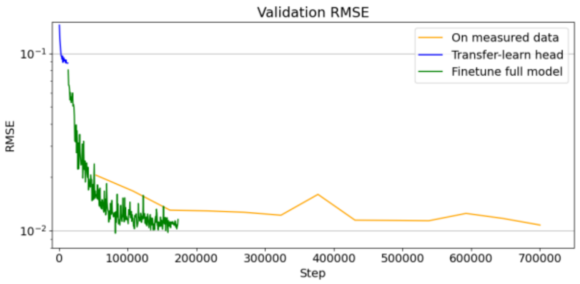
Results

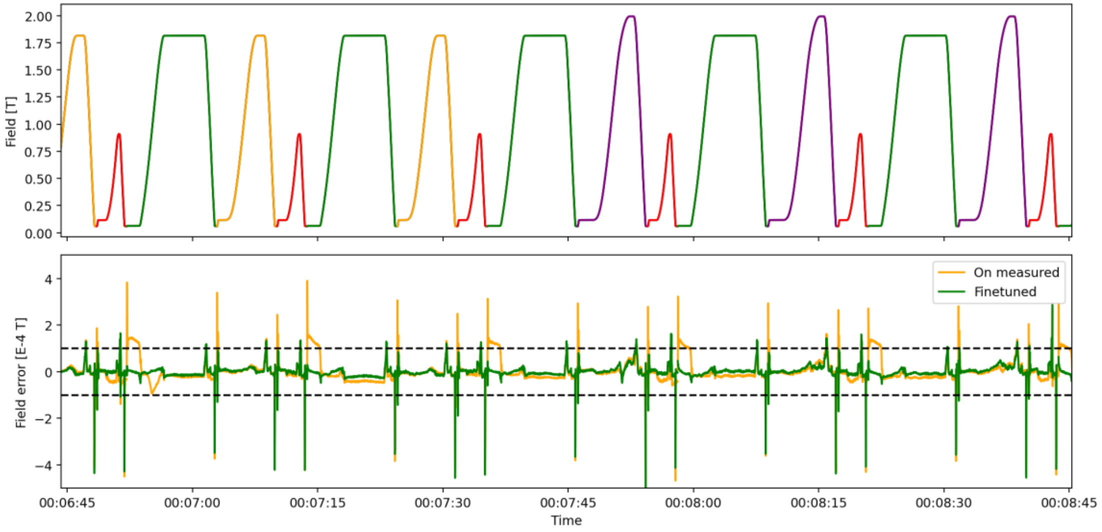
- Transfer learning from the pre-trained model is significantly faster than directly measured data
  - … the fine-tuned model marginally outperforms the model trained on only measured data in validation RMSE
  - … but significantly more robust in longer autoregressive predictions
  - Transfer learning opens new doors for transferring base model or finetuned model to other magnet families

RMSE [T]	On measured	Finetuned
Validation	6.57e-5	3.38e-5
Unseen cycles	9.39e-5	6.50e-5



<12h



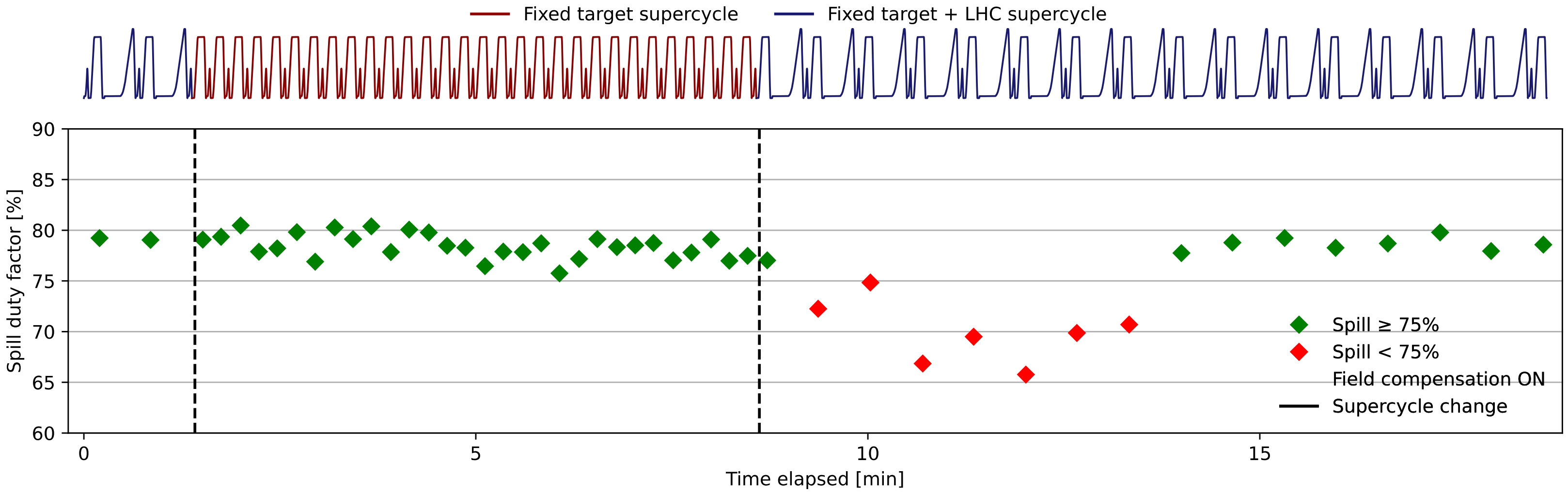


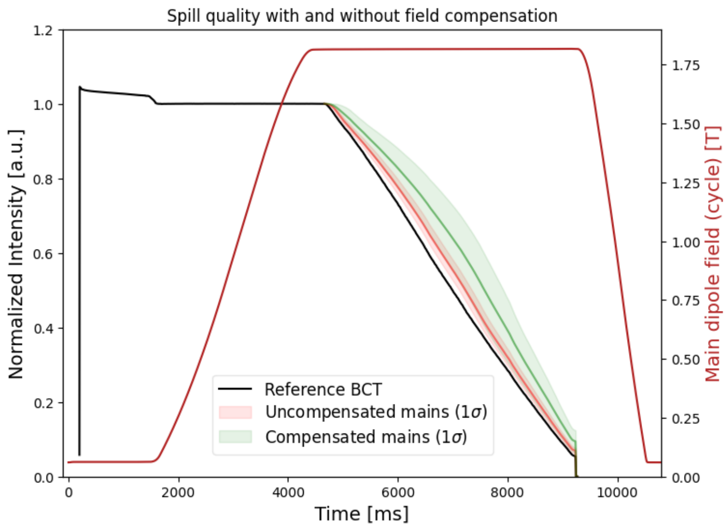
Operational results

Continuous field compensation on SPS fixed target cycle flat top

- Cycle-by-cycle compensation on SPS fixed-target compensates extraction field sufficiently ($\pm 3 \times 10^{-4}\ T$)
  - Keeps slow extracted spill stable, within margins due to quadrupole hysteresis

- Compensation stable runs for long periods (12+ hours) with measured field as input for prediction
  - … but **autoregressive predictions** and compensation runs stable for 1h+ test runs with changing cycle sequences, keeping extraction dipole field reproducible
  - 24/7 online predictions running since fall 2024

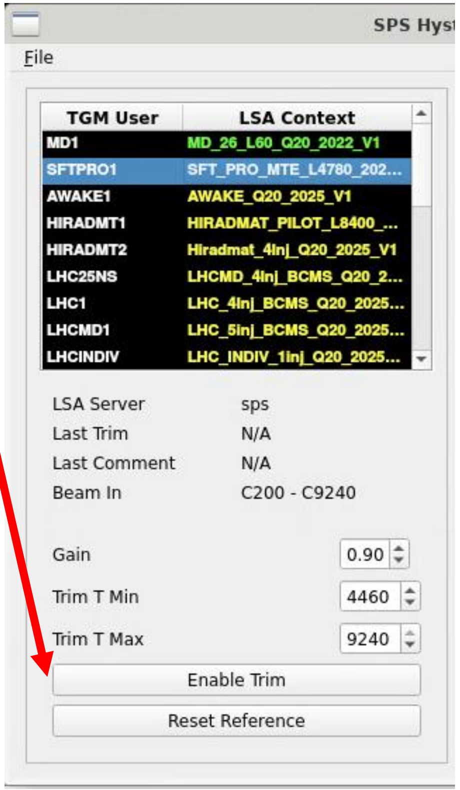




Outlook & conclusions

- High-accuracy, purely data-driven magnetic field predictions for hysteresis compensation shows operationally useful in the CERN SPS
  - Continuous efforts to improve field prediction accuracy and robustness
  - Robust autoregressive predictions will be important for magnets without state input
- Transfer learning strategy simplifies incorporating new data as it becomes available …
  - … as well as transferring the model to other magnet families
  - Q: can we train a “foundation model” for different hysteresis?

- But is it operational …?
  - Online field compensation being deployed to SPS main dipoles
  - SPS main quadrupoles to come when lab measurements arrive
  - Full-cycle and any cycle compensation, and removal of precycling cycles coming with improved model robustness and more measurements



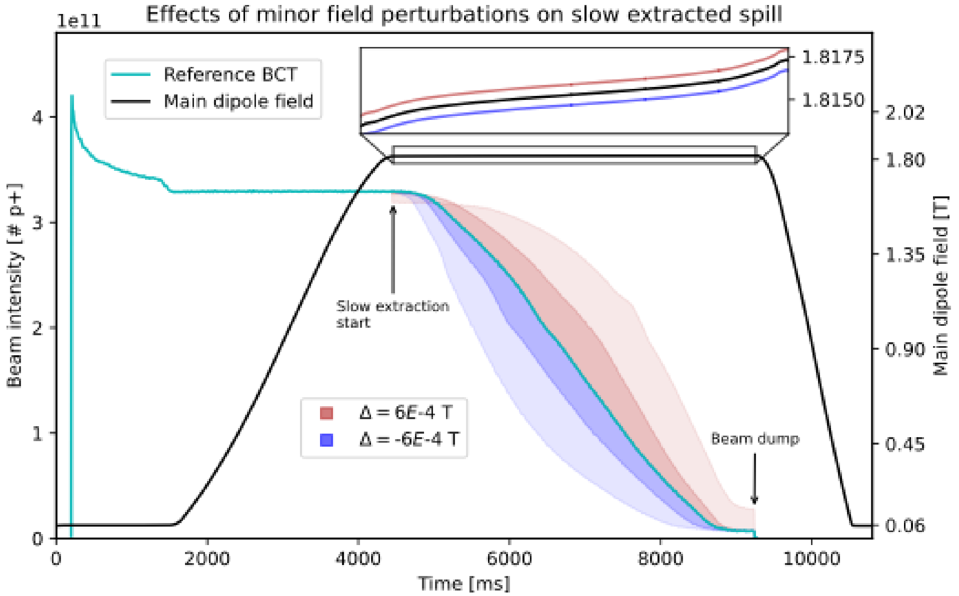
Extra slides

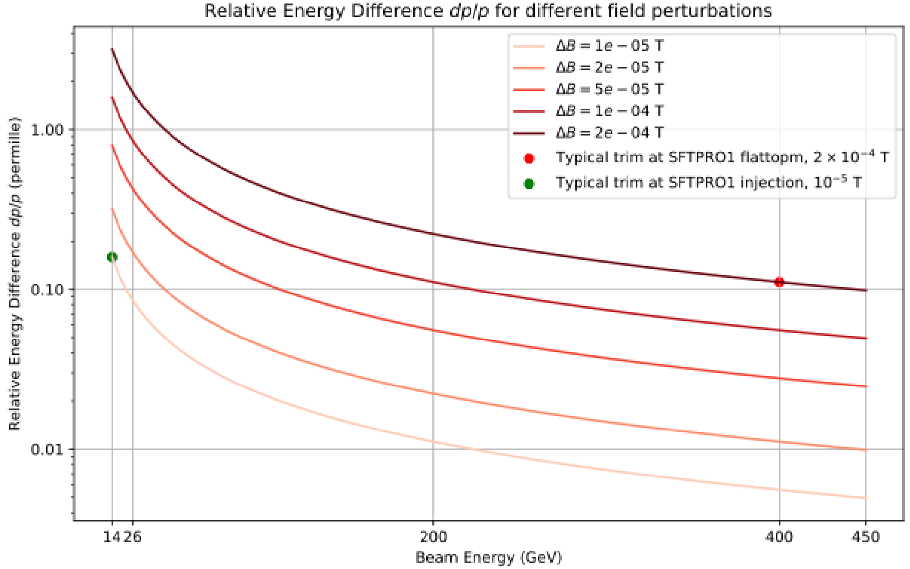
Analysis – Hysteresis in SPS main dipoles

Overview

- Field deviations due to hysteresis between $\pm 3 \times 10^{-3}$ T at most
  - But typically, below $3 \times 10^{-4}\ T$ cycle-to-cycle
  - Similar range for SPS QF/QD

- The field corrections depend on beam energy / field strength / tolerance
  - For SFTPRO slow extraction (400 Gev), tolerance is below 1×10^{-4} T
  - For SFTPRO injection (14 GeV) tolerance is $\approx 1 \times 10^{-5}$ T
  - For LHC-type injection (26 GeV) tolerance is $\approx 2 \times 10^{-5}$ T

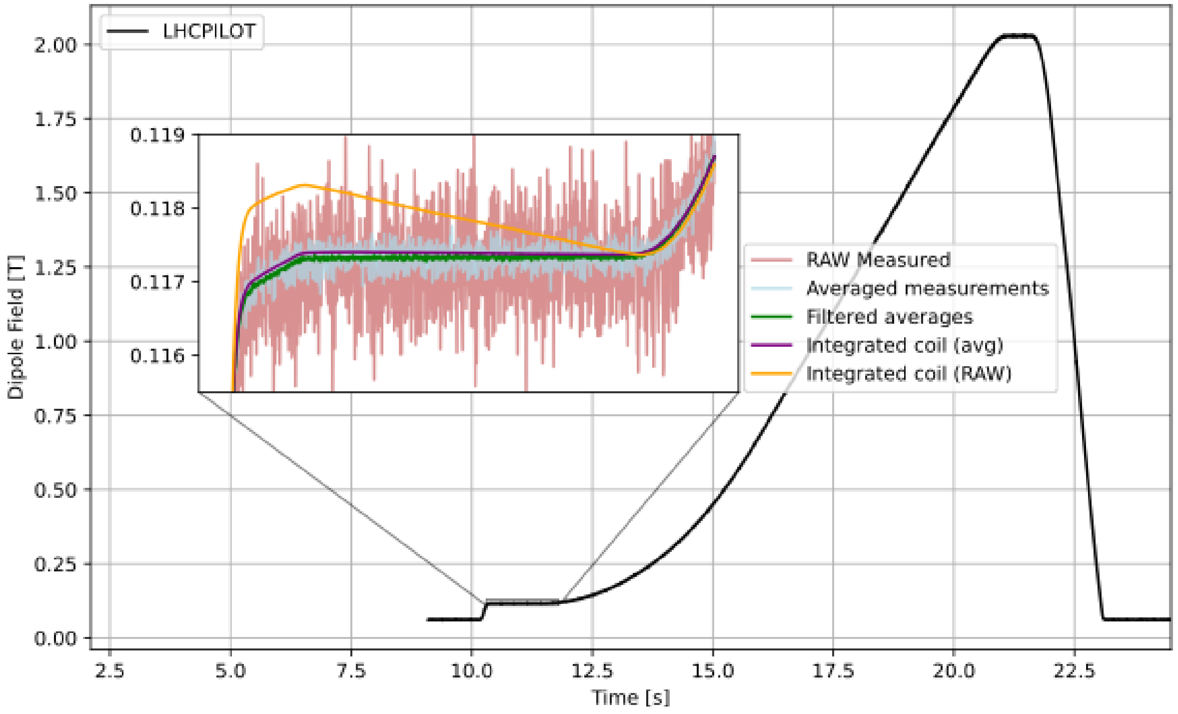




Magnetic measurements

Know

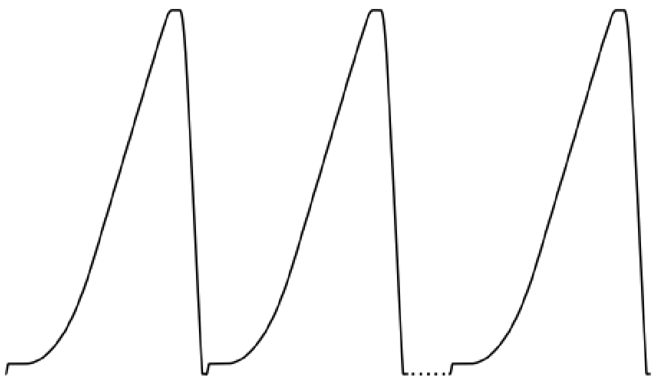
- Measure dipole magnetic response with online B-train
  - Problems with drift to reach desired accuracy, especially at SFTPRO FB
- Lab measurements for quadrupoles, sextupoles, octupules
  - Challenges to reach desired accuracy using induction coil (drift) or hall sensors (noise)
- Accurate pulsed measurements very challenging in the lab
- Power converter in lab is different from FGC
  - Lab not entirely representative of SPS conditions



Cycle-by-Cycle Online inference

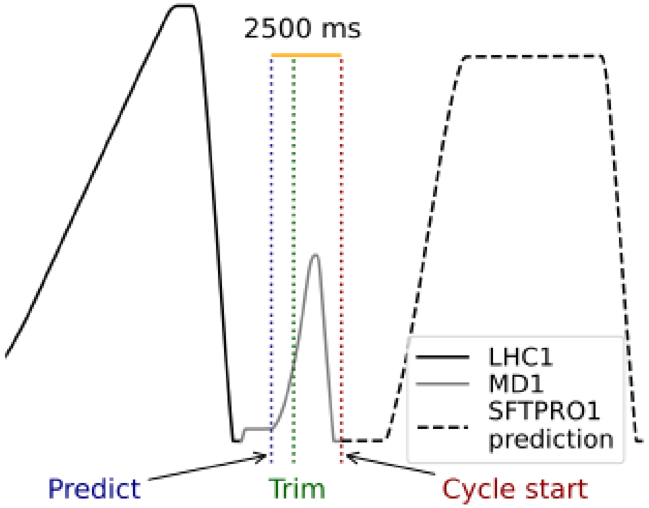
Predict

- Precycle the magnets to always start from the same field
  - For SPS MBIs: I_0, B_0 from B-Train
  - For the rest: I_0, B_0 from lab
- Then predict magnetic field for each cycle
  - 2500 ms before cycle start, predict $\hat{B}_{n+1}$ using programmed current I_{n+1}^{prog} and I_n, B_n for the next cycle
  - N.B. for magnets with B-train we can always use “true” past for future predictions



Precycle

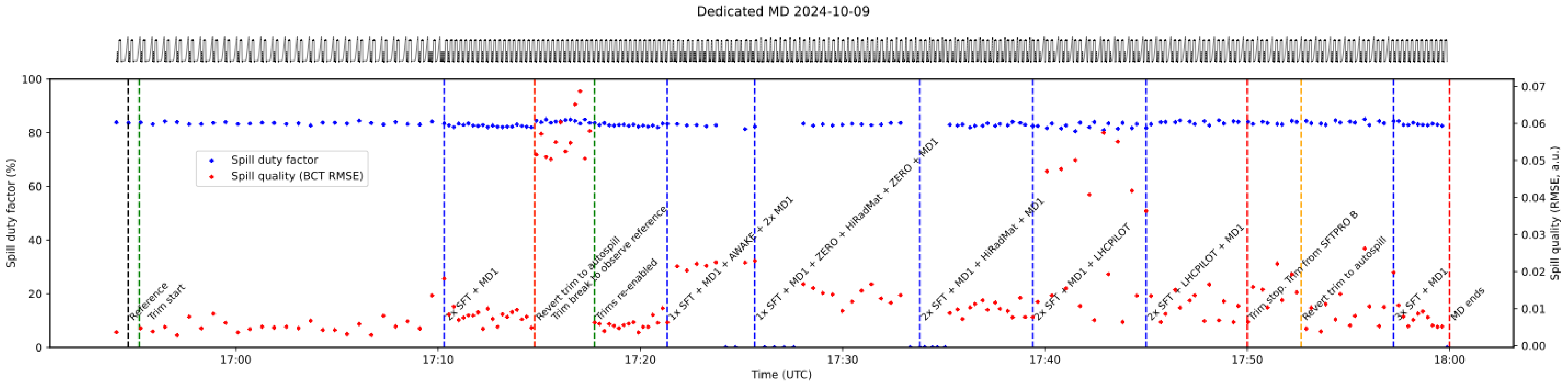
…



Significant and consistent improvements on spill quality

Results

- Evolution of spill quality over time, with autoregressive field predictions + compensation
  - › Corrections only on SFTPRO1 flat top, on **every cycle**
- Reference taken at the beginning and unchanged throughout the MD
  - › Spill duty factor remains largely unchanged
  - › … But RMSE between reference and measured BCT is significant when field is poorly / uncorrected



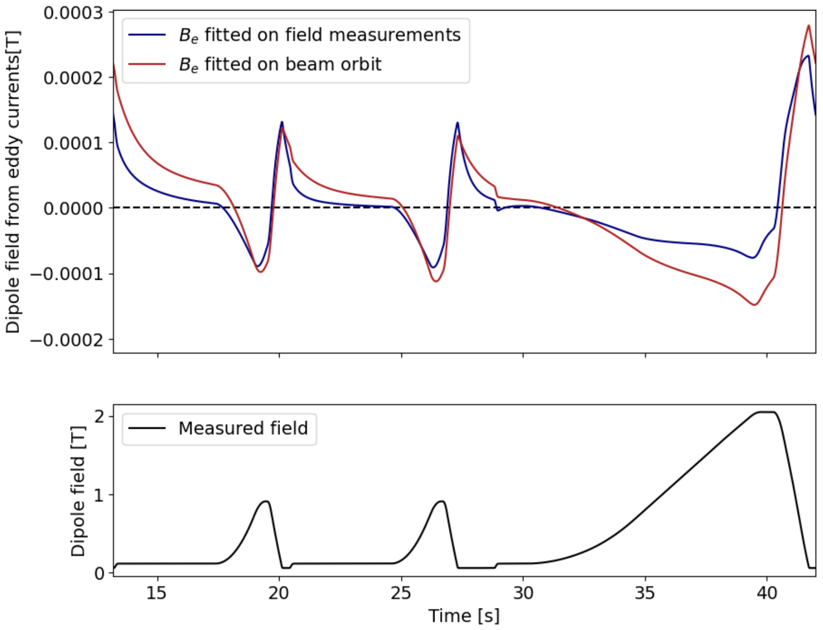
Dynamic effects will not need to be modeled

- Early tests in SPS showed that
  - Beam experiences dynamic effects that are incompatible with magnet design and field measurements
  - Deriving eddy current field contributions

$$B_e = a \sum_{t'=0}^{T} \dot{I} e^{-\frac{(t-t')}{\tau}}$$

  Eddy currents decays can be fitted from beam orbits directly, and therefore field prediction does **not need to model dynamics**

- Achieving $< 5 \times 10^{-5}\ T$ absolute accuracy, i.e., $>$ $99\ \%$ accurate predictions across the full prediction horizon (100s of points) is very challenging
  - Even when validation distribution overlaps with training)
  - Artifacts and “spikes” in predictions are not uncommon



Technical requirements for a data-driven model

- Must meet accuracy < 2e-5 T
- For all points in a forecasting horizon
- Must be able to capture dynamic effects as well as static effects
- Must be able to predict data infinitely far into the future (autoregressive predictions) without state observations

- Measured magnetic field contains
  - Measured magnetic field ($0 – 2\ T$)
    - Hysteretic component: ($\pm 3 \times 10^{-3}\ T$)
  - Eddy current decays (Up to $2 \times 10^{-4}\ T$)
  - Measurement noise ($5 \times 10^{-6}\ T$)
  - Integration drift (Normally up to $5 \times 10^{-4}\ T$)

- Phenomenological become too complex or impossible to use when static effects cannot be disentangled from measurement artifacts