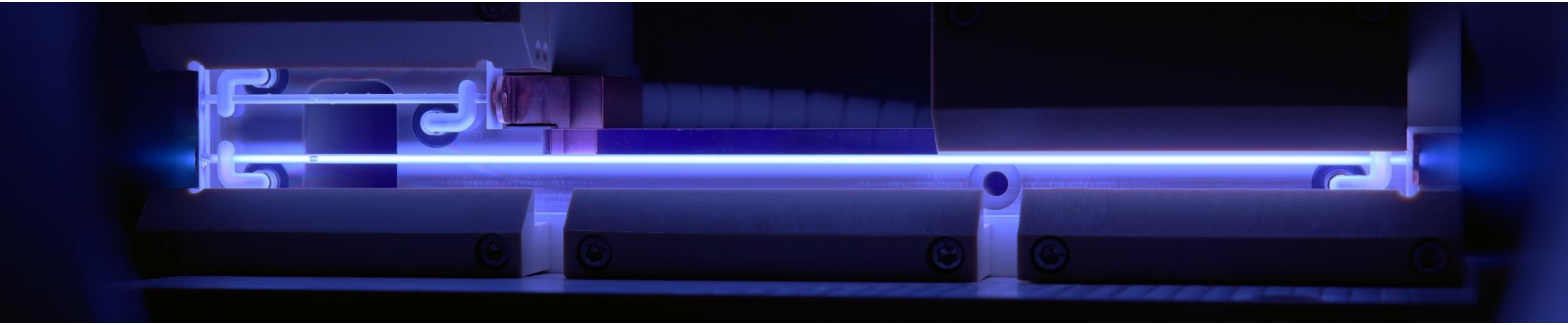
Advanced Controls and Machine Learning at FLASHForward



Lewis Boulton

Judita Beinortaite, Jonas Björklund Svensson, Philipp Burghart, Richard D'Arcy, Angel Ferran Pousa, Brian Foster, Pau Gonzalez Caminal, Maryam Huck, Harry Jones, Advait Kanekar, Carl A. Lindstrøm, Gregor Loisch, Tianyun Long, Andreas Maier, Mathis Mewes, Jens Osterhoff, Felipe Peña, Sarah Schroeder, Maxence Thévenet, Stephan Wesch, Matthew Wing, Jonathan Wood

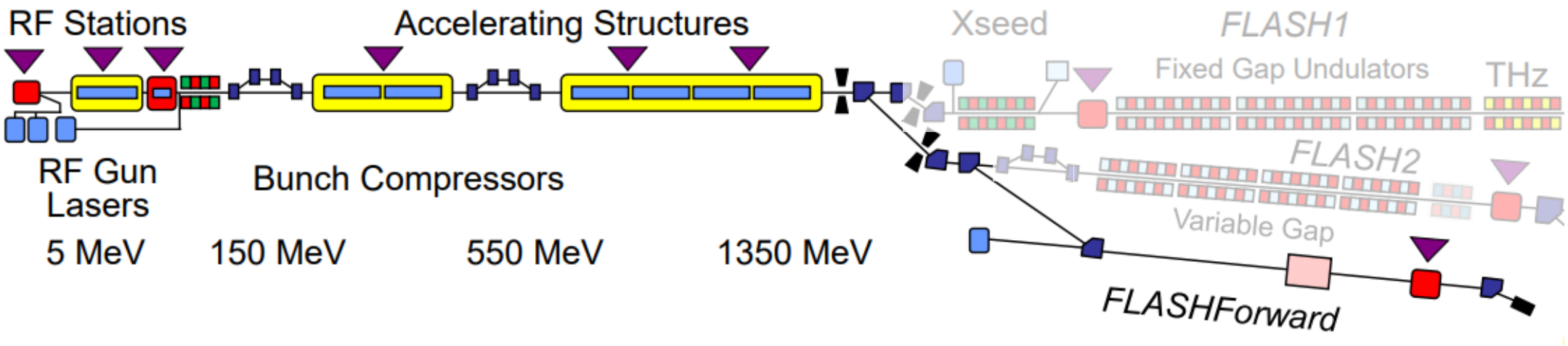
7th European Advanced Accelerator Concepts Workshop | September 21–27, 2025 | Elba, Italy

HELMHOLTZ RESEARCH FOR GRAND CHALLENGES

DESY.

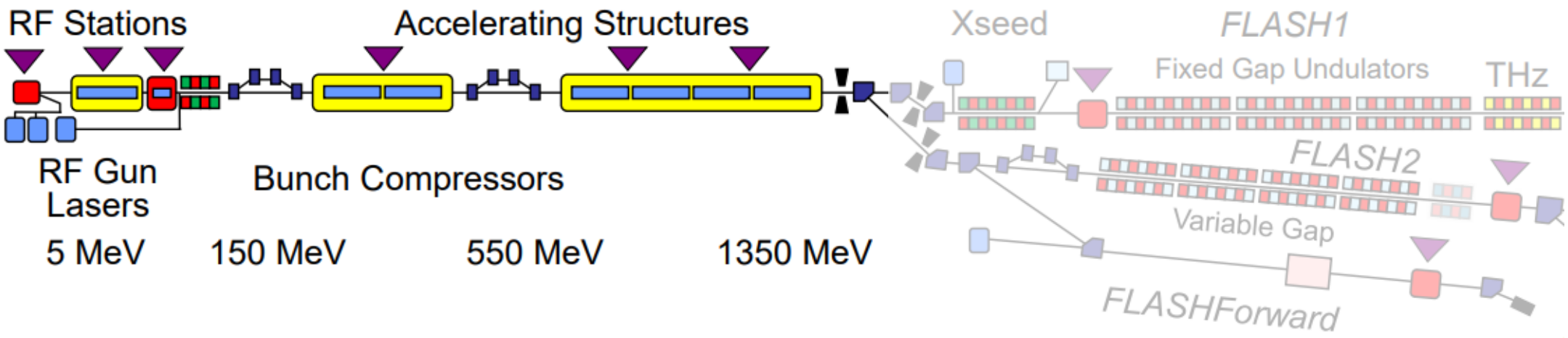
DESY.

FLASHForward



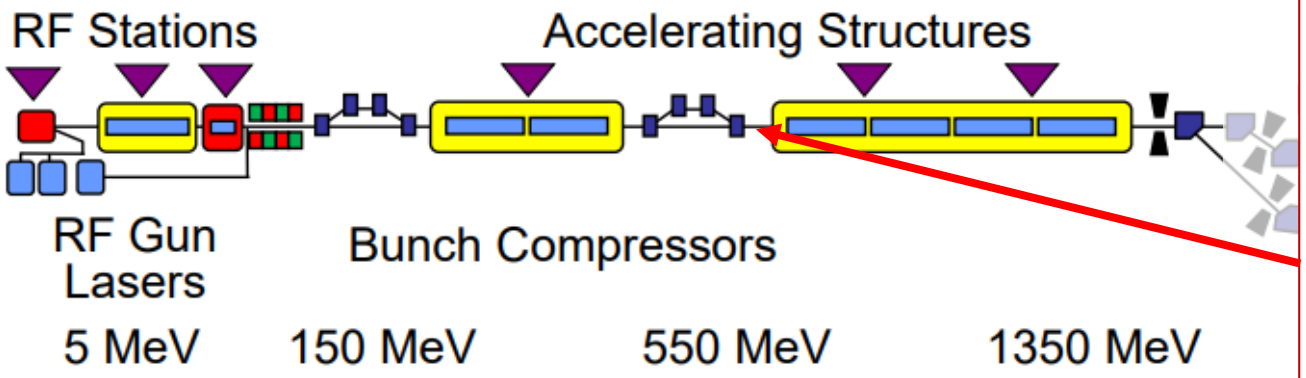
- Beam-driven plasma wakefield accelerator experiment at FLASH, DESY Hamburg
  - Aim: brightness preserved, high energy efficiency , high rep. rate PWFA
- Inherit the high-stability and diagnostic infrastructure from a FEL facility….
- Typically short on beam time… efficient setup is crucial!

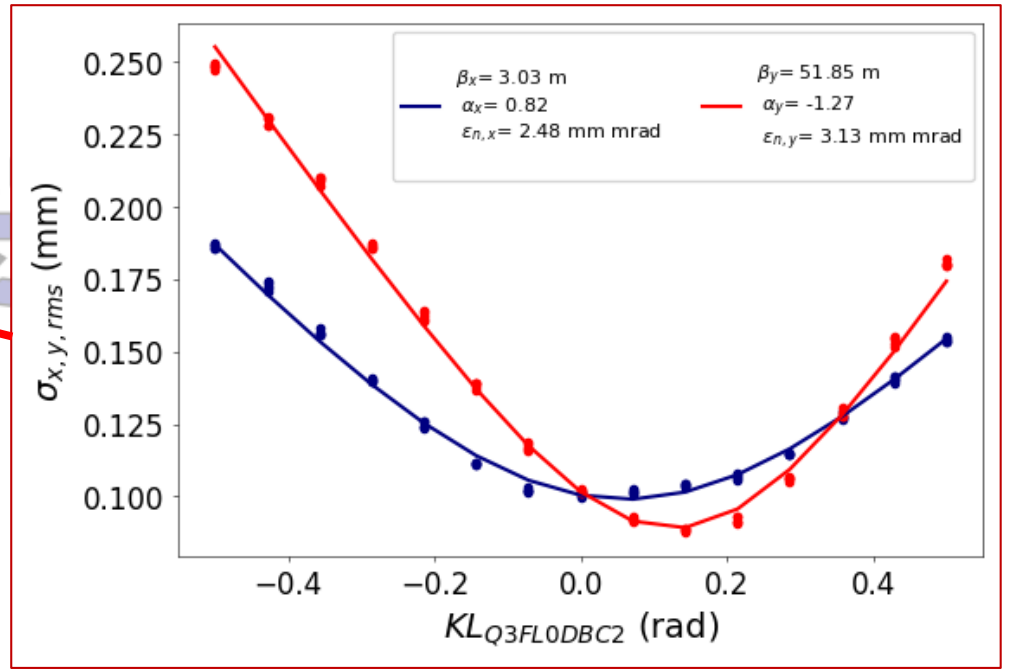
A typical linac setup…



- Linac setup: can take a long time, particularly in commissioning
- Any changes to the machine → need to repeat measurements, reoptimise etc…

A typical linac setup…

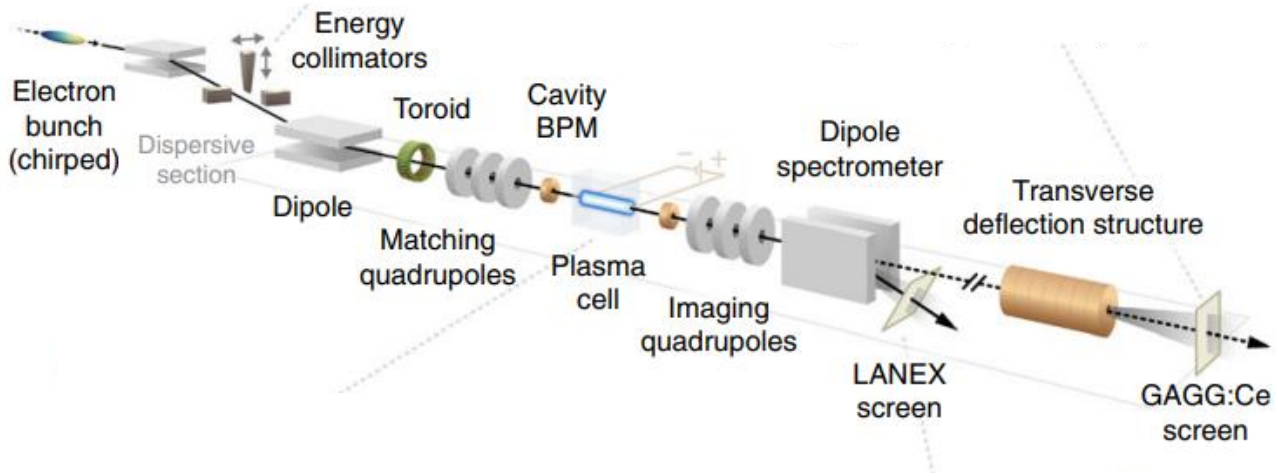




- Linac setup: can take a long time, particularly in commissioning
- Any changes to the machine → need to repeat measurements, reoptimise etc…
- E.g Matching after second bunch compressor
  - Analytical matching can be done, but only if the section of accelerator is well understood / collective effects are small
  - Measurements typically take ~5 mins each… many iterations make it a long procedure

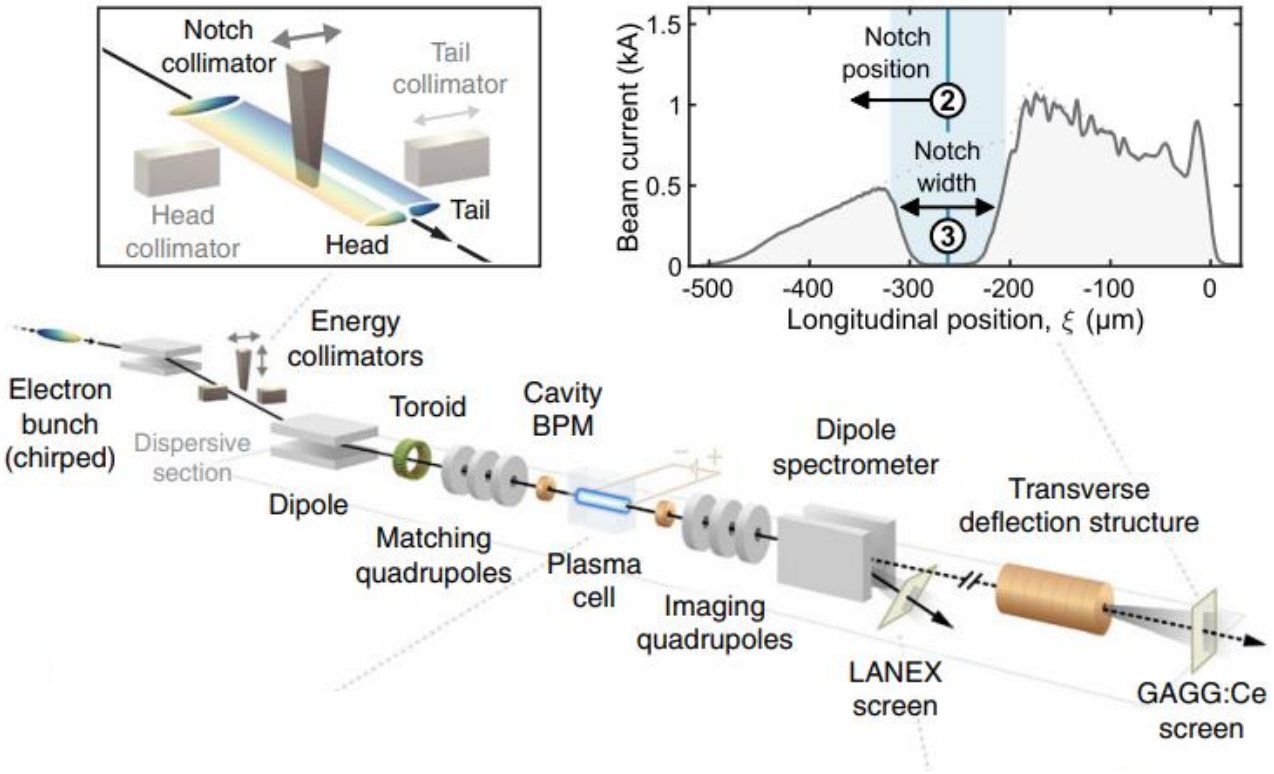
A typical PWFA setup…

C. A. Lindstrøm et al., Phys. Rev. Lett., vol. 126, p. 014801, Jan 2021



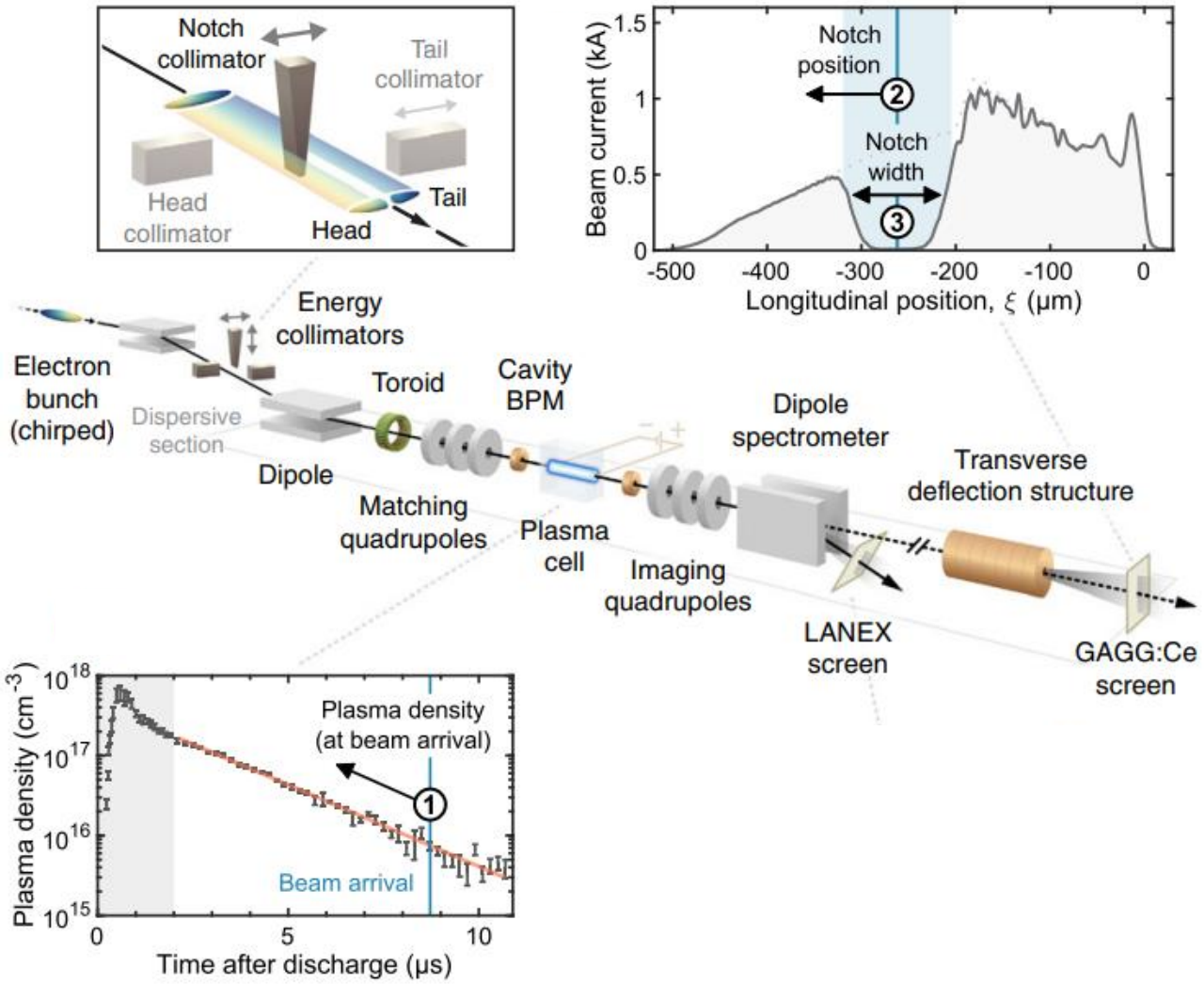
A typical PWFA setup…

C. A. Lindstrøm et al., Phys. Rev. Lett., vol. 126, p. 014801, Jan 2021



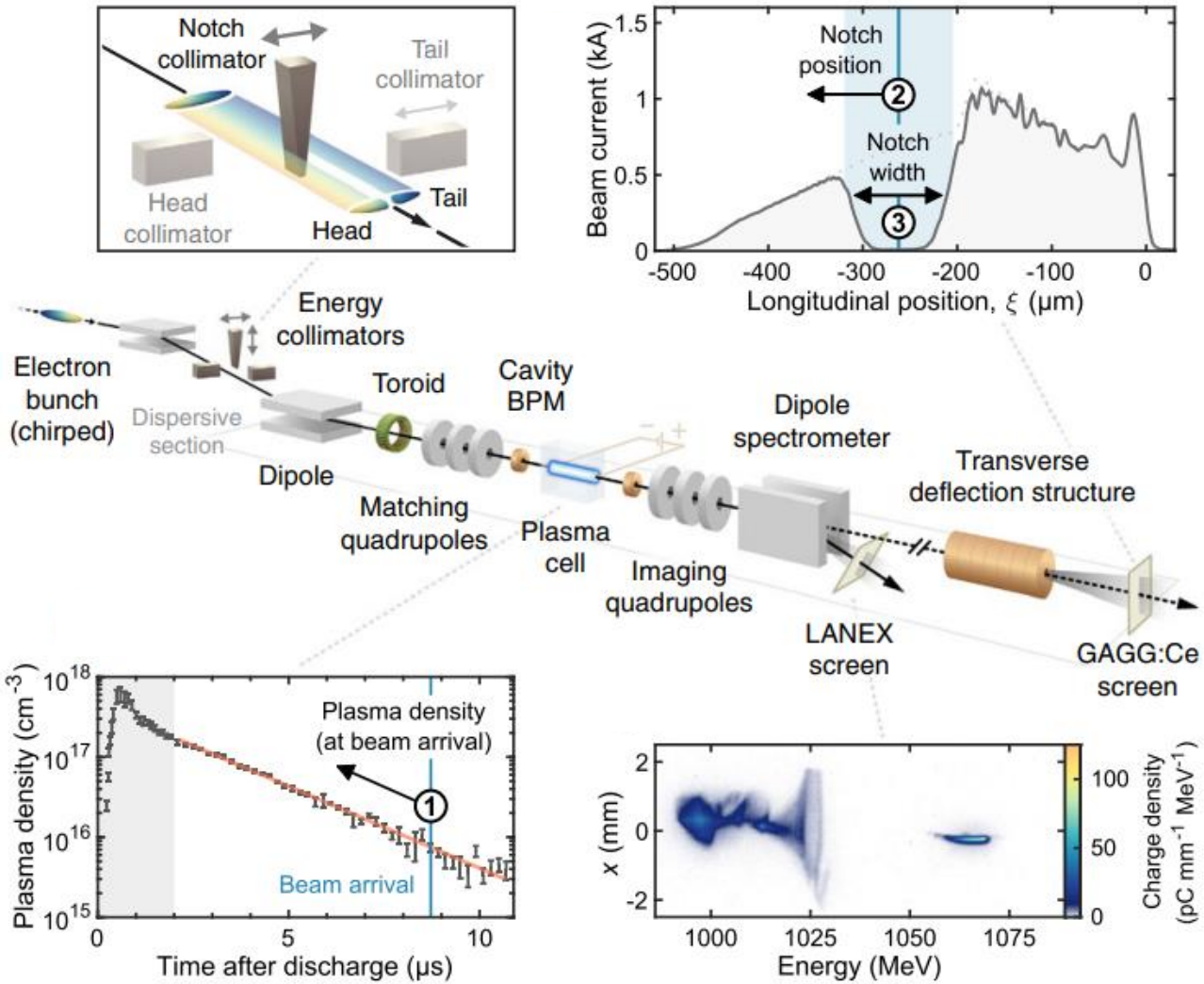
A typical PWFA setup…

C. A. Lindstrøm et al., Phys. Rev. Lett., vol. 126, p. 014801, Jan 2021



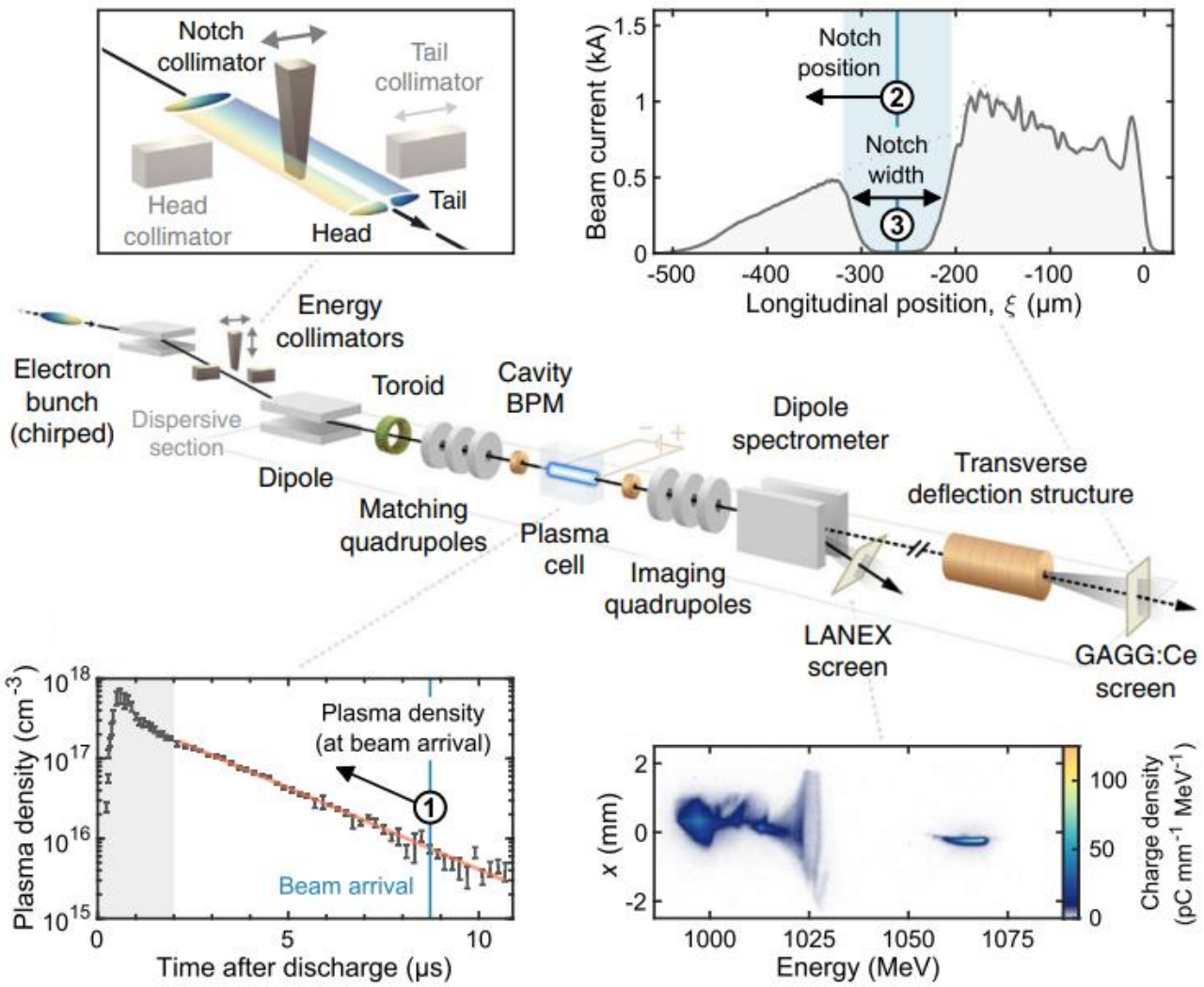
A typical PWFA setup…

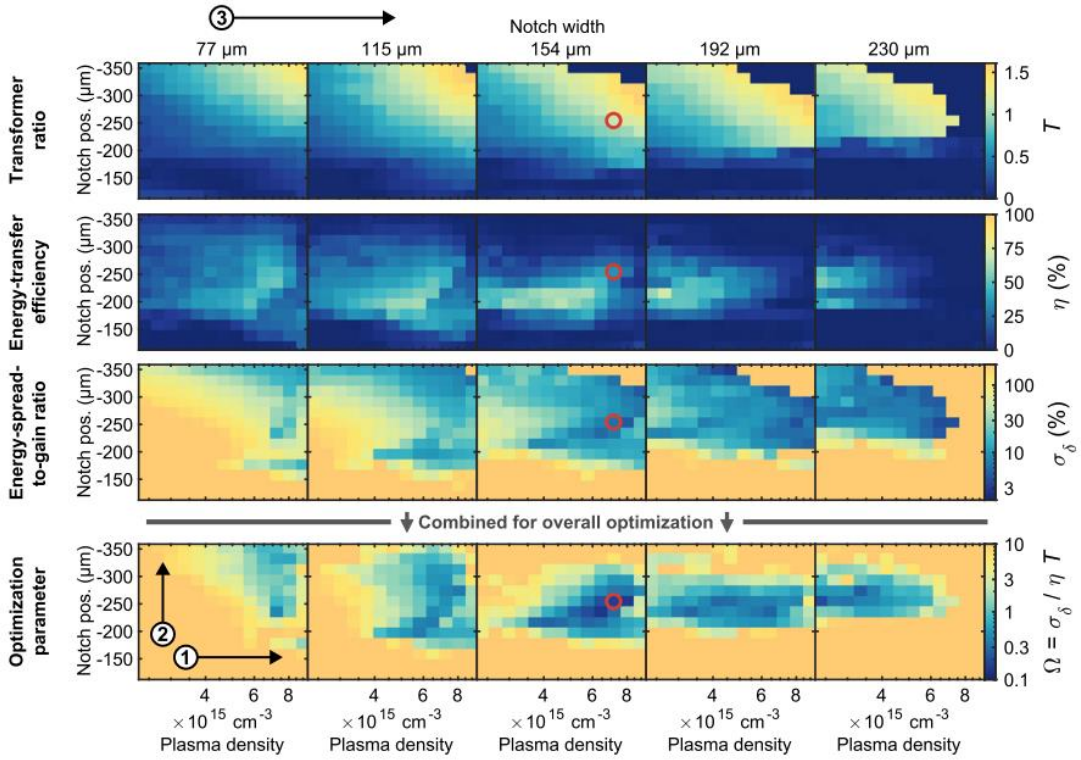
C. A. Lindstrøm et al., Phys. Rev. Lett., vol. 126, p. 014801, Jan 2021



A typical PWFA setup…

C. A. Lindstrøm et al., Phys. Rev. Lett., vol. 126, p. 014801, Jan 2021





> Optimise for high energy-transfer efficiency, low energy spread

> Baseline for further studies e.g high-rep rate, beam-quality preservation

> But is a very long and involved process…

Two Examples of Problems

- Matching:
  - *Relatively* stable environment…
  - But very time-consuming and complicated, multi-point measurements
- PWFA Setup:
  - Measurements are ‘cheap’: evaluate accelerated bunch spectra on single-shot basis…
  - But a very noisy environment, with a lot of tuning knobs available

Two Examples of Problems

- Matching:
  - *Relatively* stable environment…
  - But very time-consuming and complicated, multi-point measurements
- PWFA Setup:
  - Measurements are ‘cheap’: evaluate accelerated bunch spectra on single-shot basis…
  - But a very noisy environment, with a lot of tuning knobs available

How to do better…?

Bayesian optimisation for linac matching

- Matching in accelerators- ‘match’ the beam twiss to some ‘design’ twiss

$$\mathrm{B_{mag}} = \frac{1}{2}\left(\beta\gamma_D - 2\alpha\alpha_D + \gamma\beta_D\right)$$

$$\beta = \beta_D\,, \alpha = \alpha_D, \gamma = \gamma_D \;\Rightarrow B_{mag} = 1$$

$$\text{otherwise } B_{mag} > 1$$

Minty, M. G. & Zimmermann, F. (2003). Measurement and Control of Charged Particle Beams. Springer.

Bayesian optimisation for linac matching

optimas

Ferran Pousa, A. et al. (2023). PRAB, 26(8), 084601.

> Matching in accelerators- ‘match’ the beam twiss to some ‘design’ twiss

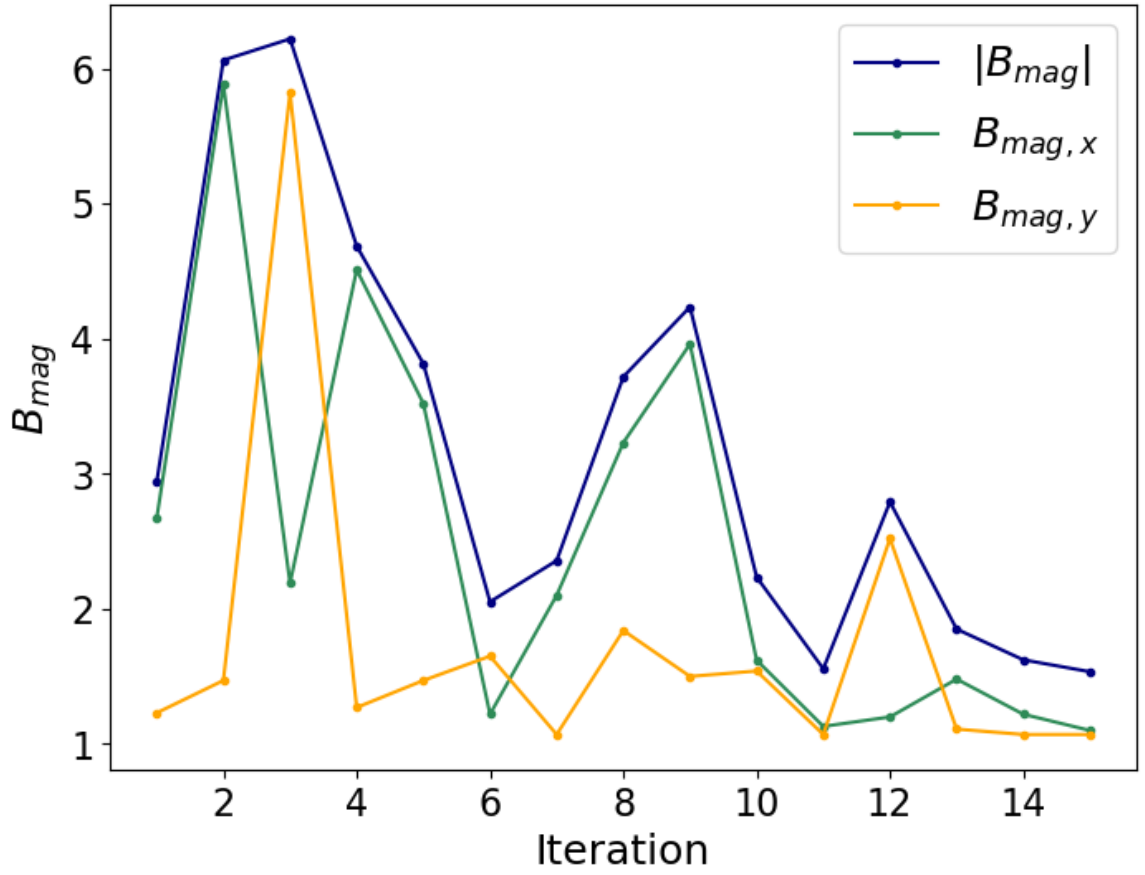
$$\mathrm{B_{mag}} = \frac{1}{2}\left(\beta\gamma_D - 2\alpha\alpha_D + \gamma\beta_D\right)$$

$$\beta = \beta_D\,, \alpha = \alpha_D, \gamma = \gamma_D \;\Rightarrow B_{mag} = 1$$

$$\text{otherwise } B_{mag} > 1$$

Minty, M. G. & Zimmermann, F. (2003). Measurement and Control of Charged Particle Beams. Springer.

> ‘Brute force’: Using BO, Minimise combined mismatch with ~5 matching quadrupoles



> Problems: still very time consuming and only as good as the diagnostic you use

> Emittance also comes out of this analysis: minimise for that in the future?

Bayesian optimisation of trailing bunch acceleration

> Goal function, maximise:

$$G = (\Delta E)^2 \left(\frac{\mathrm{d}Q}{\mathrm{d}E}\right)_{max} \sim$$

(High charge, high energy, low energy spread)

Energy gain

Peak spectral density

Bayesian optimisation of trailing bunch acceleration

> Goal function, maximise:

$$G = (\Delta E)^2 \left(\frac{\mathrm{d}Q}{\mathrm{d}E}\right)_{max} \sim$$

(High charge, high energy, low energy spread)

Energy gain

Peak spectral density

> Exposed 4 parameters to the optimiser:
- Chirp h (compression)
- Discharge delay t_d (plasma density)
- Quad current I_Q (beam tilt)
- Position of wedge x_{wedge} (driver/ witness charge distribution)

Bayesian optimisation of trailing bunch acceleration

> Goal function, maximise:

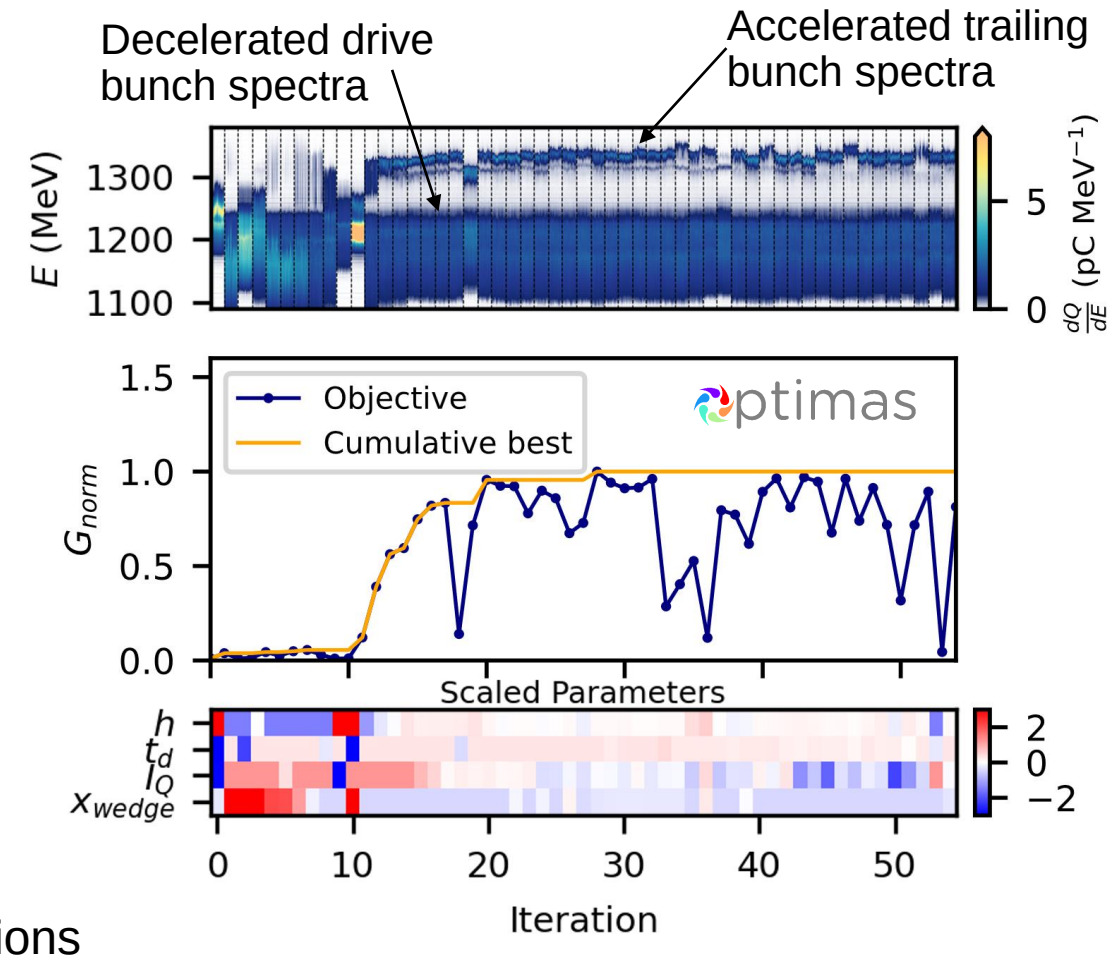
$$G = (\Delta E)^2 \left(\frac{\mathrm{d}Q}{\mathrm{d}E}\right)_{max} \sim$$

(High charge, high energy, low energy spread)

Energy gain

Peak spectral density

> Exposed 4 parameters to the optimiser:
- Chirp h (compression)
- Discharge delay t_d (plasma density)
- Quad current I_Q (beam tilt)
- Position of wedge x_{wedge} (driver/ witness charge distribution)



J. Wood et al., in Proc. IPAC'24, Nashville, TN, May 2024, pp. 541-544

> Starting from ~ noise, converged on a solution in 28 iterations to:
- (34 ± 2) pC accelerated charge , (2.4 ± 0.2) pC/MeV peak spec. density, (0.8 ± 0.1)% FWHM energy spread
- (103 ± 1) MeV energy gain in 50 mm (*2.1 GV/m*)
- Scan time: 37 minutes total, 19 to reach optimum

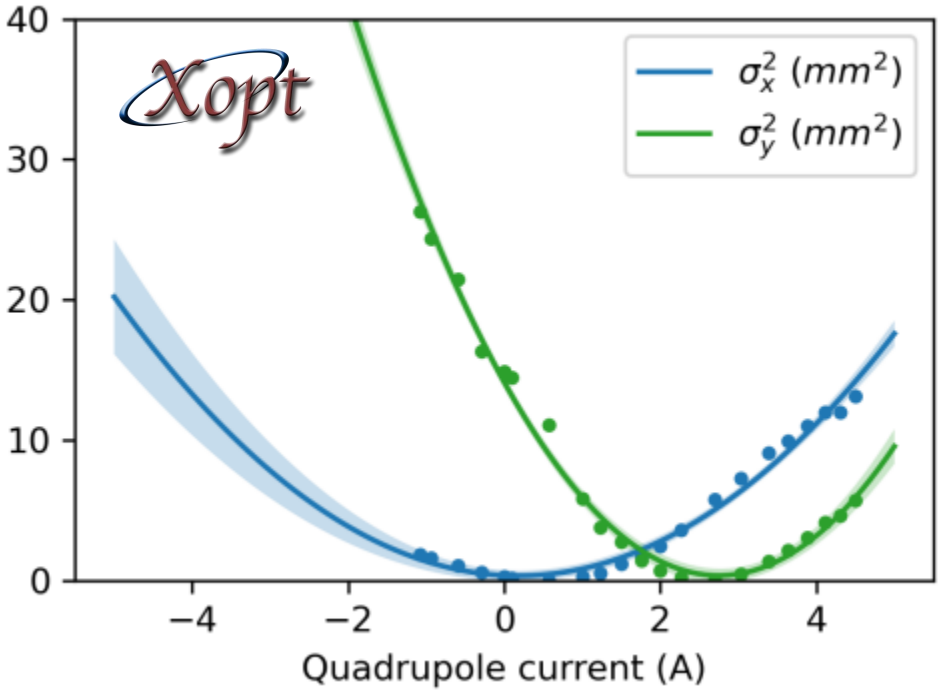
> Problems: combined goal function, poor scaling with more parameters, etc.

BO at FLASHForward: what next?

- Only tried ‘basic’ algorithms so far: clear improvements to make

- Faster, automated multi-point measurements for e.g optics:
  - Use Bayesian *Exploration* to efficiently sample the measurement parameter space → quicker iterations
  - Use constraints to avoid ‘bad’ measurements → more robust

- ‘Smarter’ PWFA optimisation:
  - How to deal with ‘slow’ actuators e.g wedge collimator width?
  - How much do drifts effect the optimiser? → time-dependent BO?



Roussel, R. et al. (2023). Instruments, 7(3), 29.

A Virtual Spectral Diagnostic for PWFA

Courtesy Phillip Burghart (FLASHForward Masters Student)

- During optimisation- collect a lot of data from non-invasive diagnostics → train a NN to predict accelerated (or decelerated) bunch properties (e.g energy spectrum) → a 'virtual diagnostic'
  - Examples:
    - Virtual LPS diagnostic: *Emma, C. et al. (2018). Phys. Rev. Accel. Beams, 21(11), 112802.*
    - Virtual photon pulse diagnostic (FELs): *El-Din, K. K. A., et al. (2024). Sci. Rep., 14, 7267.*

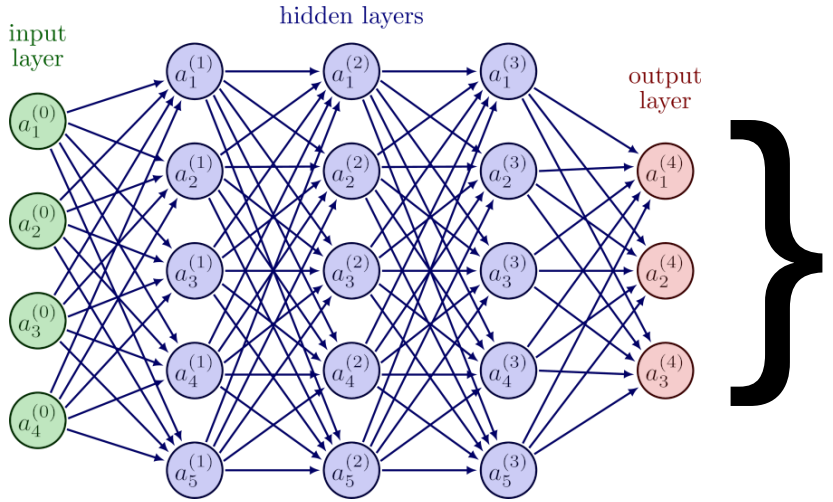
- Potential applications:
  - A useful approach for assessing sources of variation in the acceleration process e.g in *A. Maier et. Al., Phys. Rev. X 10, 031039, 2020*
  - Non-destructive: bunches could propagate to applications (or more acceleration stages)
  - Beamline diagnostics measure across MHz bunch trains → Predict spectra across bunch trains?

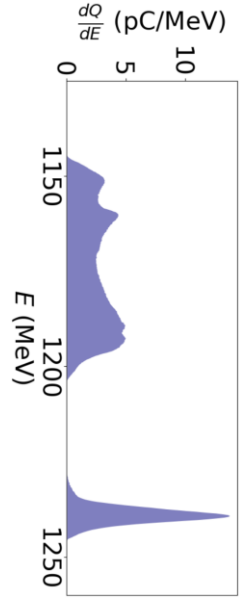
BPMs (orbit, charge)

BCMs (bunch length)

RF parameters

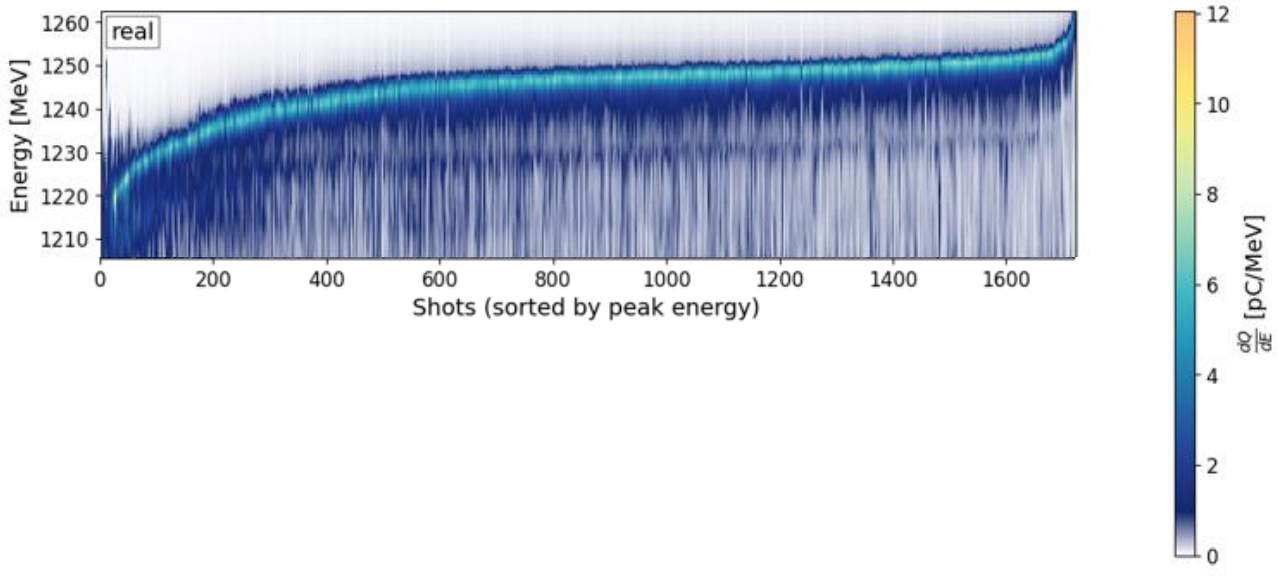
…





A Virtual Spectral Diagnostic for PWFA - Results

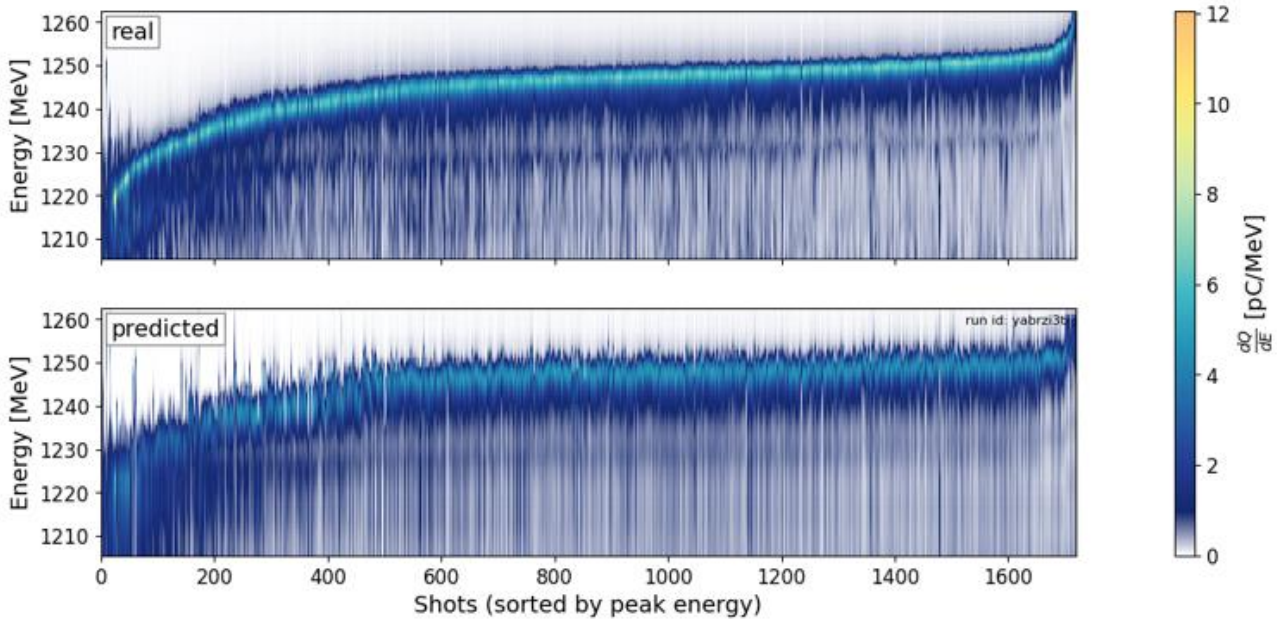
Courtesy Phillip Burghart (FLASHForward Masters Student)



- Predict PWFA trailing bunch spectra from ~12000 shot stats dataset, 80:20 train test split
- Simple Multi-layer perceptron network

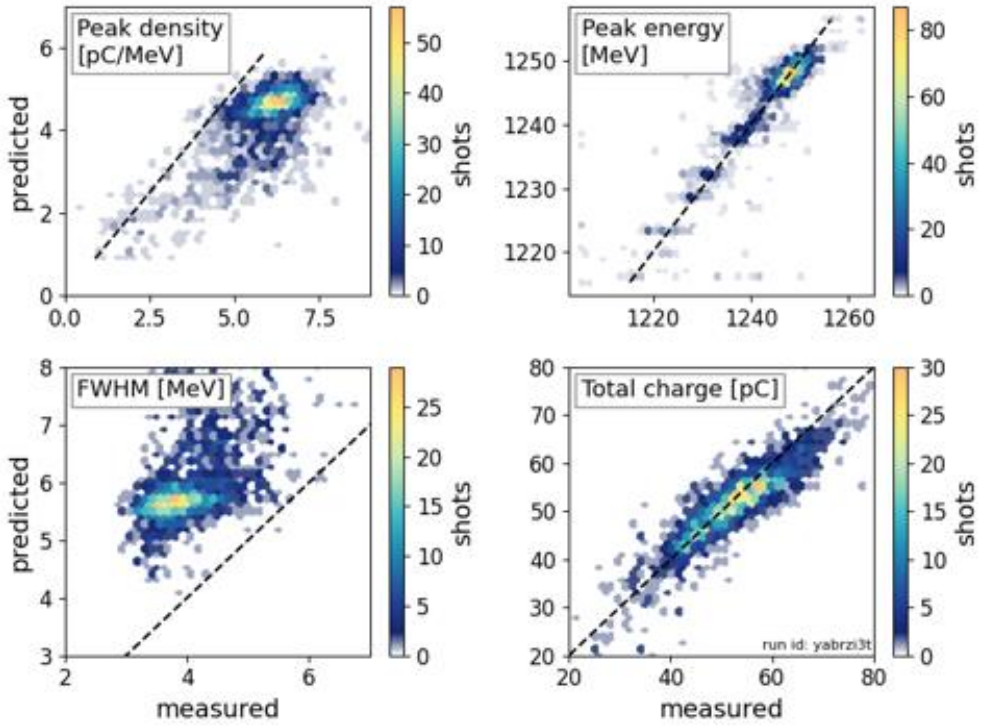
A Virtual Spectral Diagnostic for PWFA - Results

Courtesy Phillip Burghart (FLASHForward Masters Student)



- Peak energy and charge of accelerated bunches are quite well-predicted
  - Downstream BPMs have a lot of influence
- Energy spread and peak spectral density less so: why?
  - Beam loading is sensitive to the details of the wakefield and witness current profile: are we capturing that info shot-to-shot?

- Predict PWFA trailing bunch spectra from ~12000 shot stats dataset, 80:20 train test split
- Simple Multi-layer perceptron network

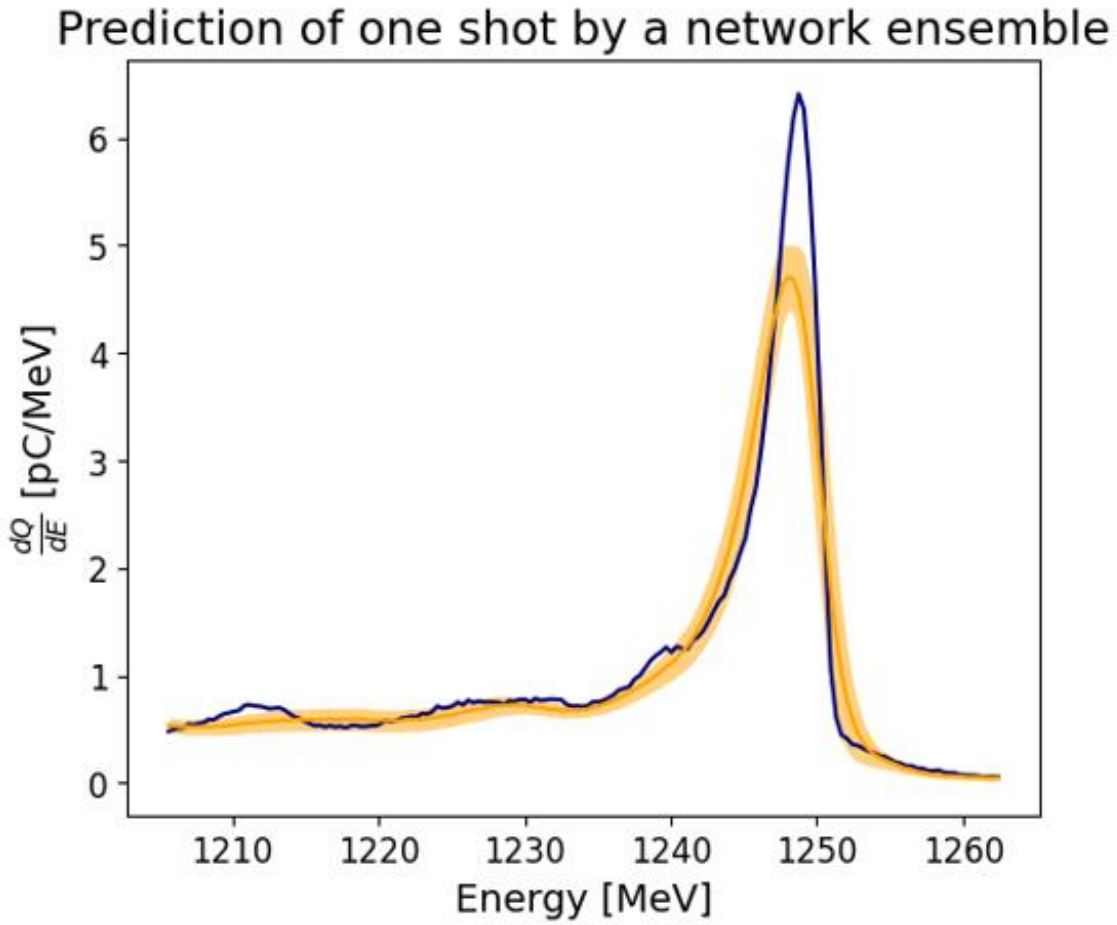


$R^2{}_{median} = 0.779$

A Virtual Spectral Diagnostic for PWFA- What next?

Courtesy Phillip Burghart (FLASHForward Masters Student)

- Different network architectures: VAEs
  - E.g *M. J. V. Streeter et al., High Power Laser Science and Engineering, 2023*
- Temporal stability:
  - How often do we have to retrain our networks?
- Uncertainty estimation:
  - Ensembling approach has been tested so far
    e.g Convery, O. et al. (2021). Phys. Rev. Accel. Beams, 24(7), 0746
  - How to include measurement errors from diagnostics?

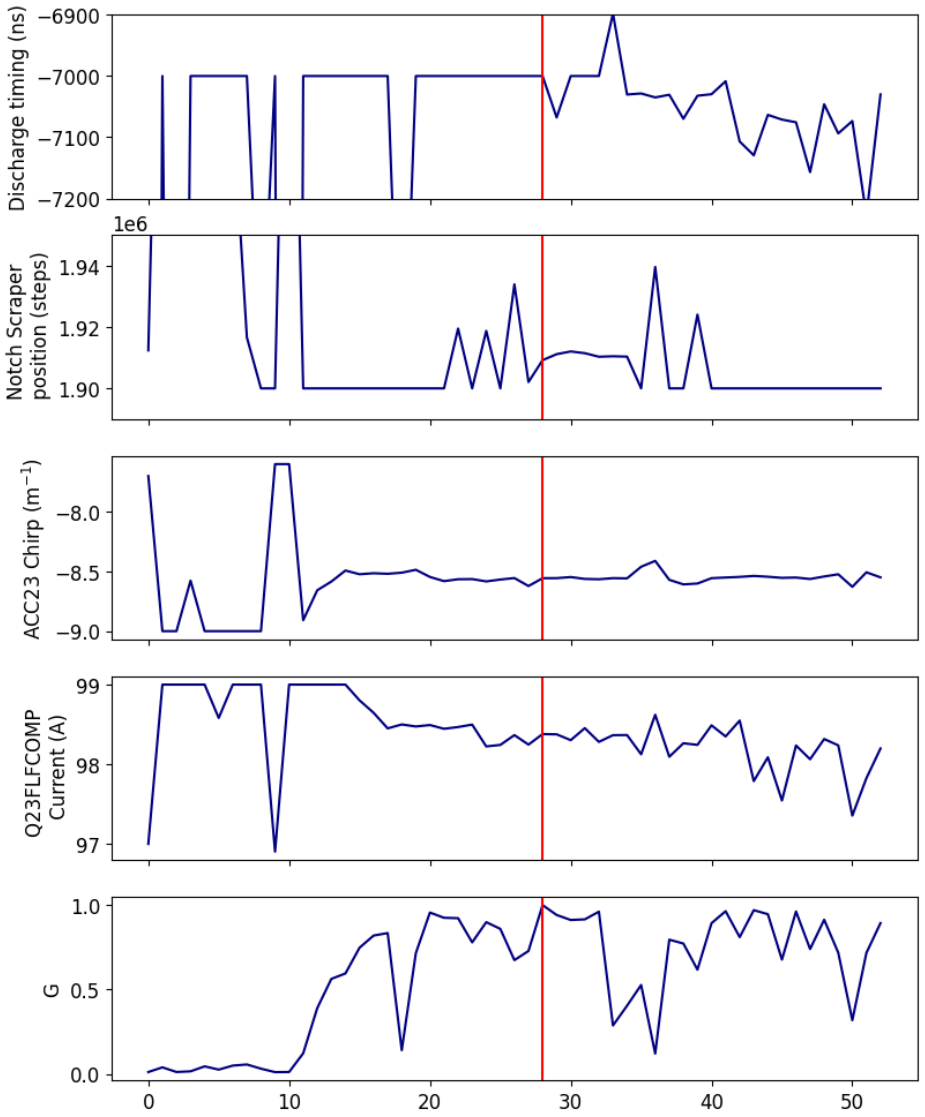


Summary

- Lots of examples of complicated setup tasks at FLASHForward:
  - Matching in the linac: a stable environment but time consuming measurements
  - PWFA optimisation: measurements are shot-to-shot but a lot of variables that are coupled
- Some initial success with Bayesian Optimisers:
  - Identifying the best goal functions and the best combinations of variables
  - Future: simpler goal functions, more advanced algorithms (e.g to address speed of evaluation)
- Development of Virtual spectral diagnostics for PWFA:
  - Good prediction so far of coarse features like energy gain and charge: energy spread not so well captured
  - Lots of things to try: new architectures, uncertainty estimation, temporal stability (more data??)
  - Applications: a non-invasive, MHz diagnostic for PWFA bunches?

Backup

Bayesian optimisation of trailing bunch acceleration



Goal function for witness acceleration

> Wakefield parameter, minimise:

$$\Omega = \frac{\sigma_\delta}{\eta T}$$

$$\eta = -\frac{\Delta\langle E\rangle_{\mathrm{acc}} Q_{\mathrm{acc}}}{\Delta\langle E\rangle_{\mathrm{dec}} Q_{\mathrm{dec}}}$$

Energy-transfer efficiency

$$\sigma_\delta = \frac{\sigma_{E,acc}}{\Delta\langle E\rangle_{acc}}$$

Energy spread-to-gain ratio

C. A. Lindstrøm et al., Phys. Rev. Lett., vol. 126, p. 014801, Jan 2021

> BO goal function, maximise:

$$G = (\Delta E)^2 \left(\frac{\mathrm{d}Q}{\mathrm{d}E}\right)_{max}$$

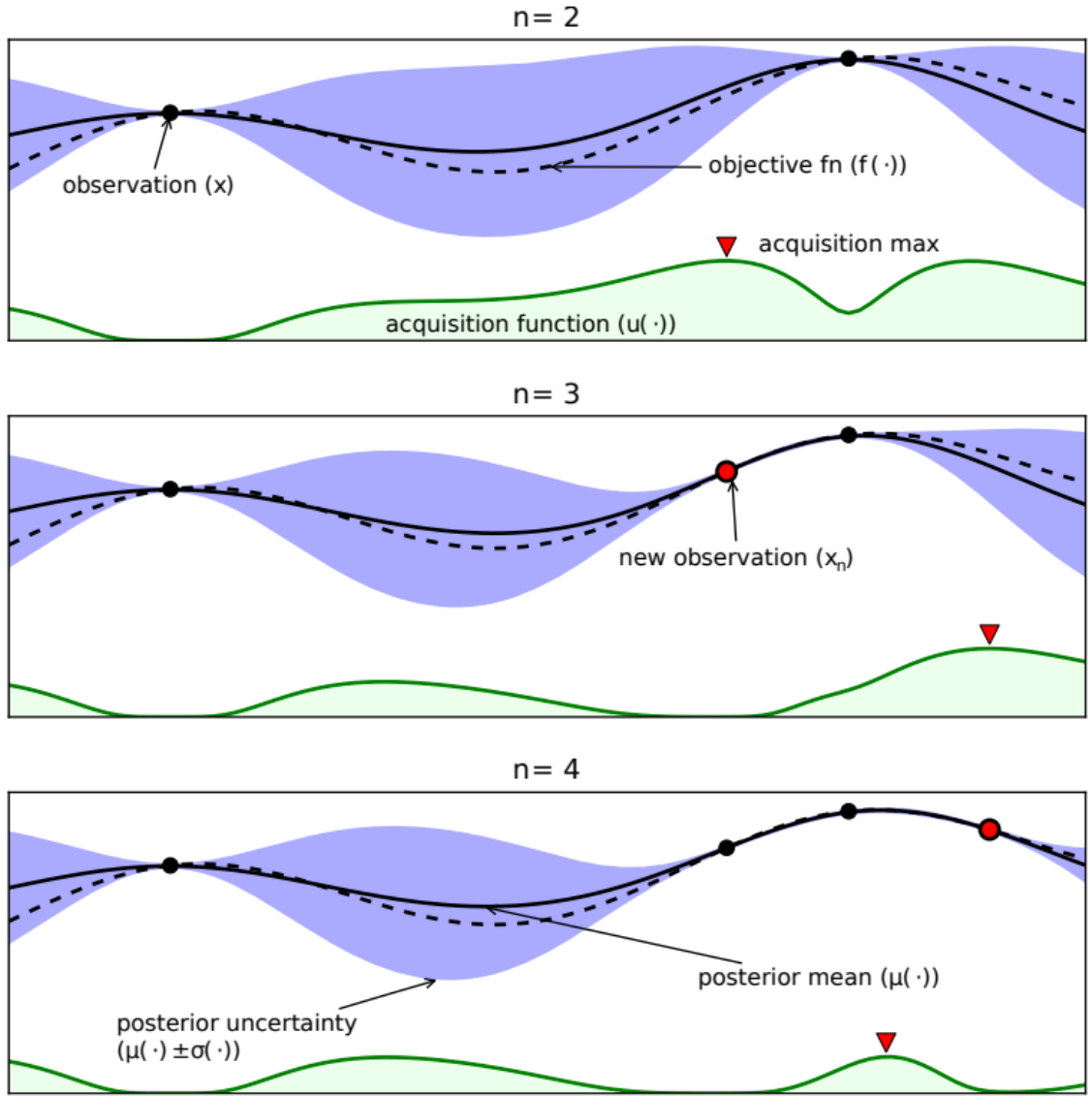
$$G \sim \frac{(\Delta\langle E\rangle_{acc})^2 Q_{acc}}{\sigma_{E,acc}}$$

Bayesian Optimisation

> Basic principle:

1. Fit gaussian process (GP) model to data → objective function value and uncertainties
2. Model → ‘Acquisition’ function- describes the ‘value’ sampling each point
3. Sample where the acquisition function is highest, update model, repeat

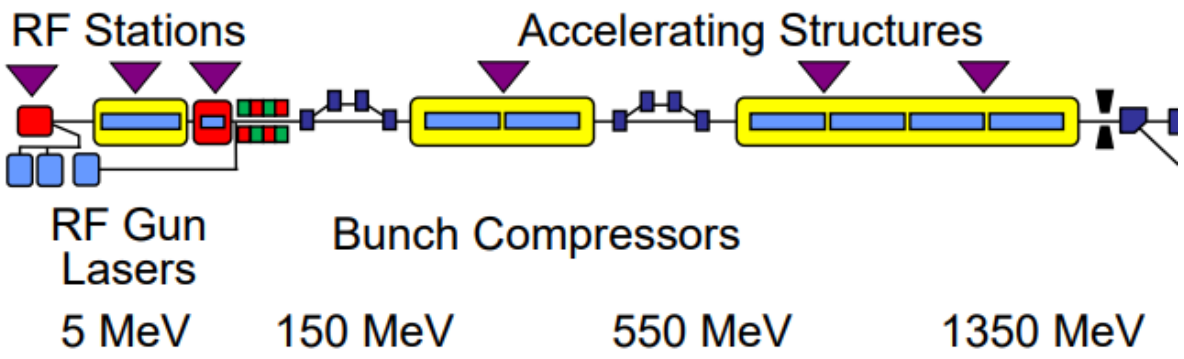
> Popular method for the sample-efficient optimisation of noisy, expensive-to evaluate objectives e.g in particle accelerators!



Shahriari, B. et al. (2016). Proc. IEEE, 104(1), 148–175.

A typical FLASHForward beamtime: Linac setup

FLASH linac: many examples of complicated, time-consuming setup tasks…



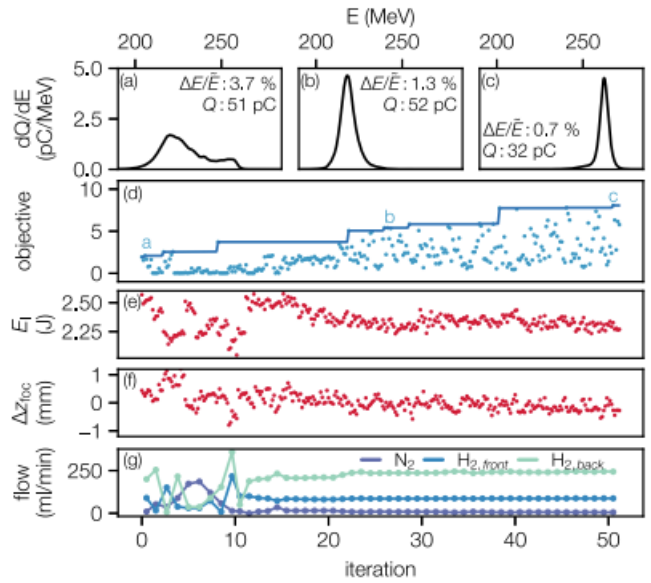
Injector setup:

- Highly non-linear dynamics (e.g from space charge)
- Many parameters to tune (injector laser, RF gun, optics etc.)
- Proper diagnosis requires multi-point measurements
- Twin bunch generation even more complicated, very little experience

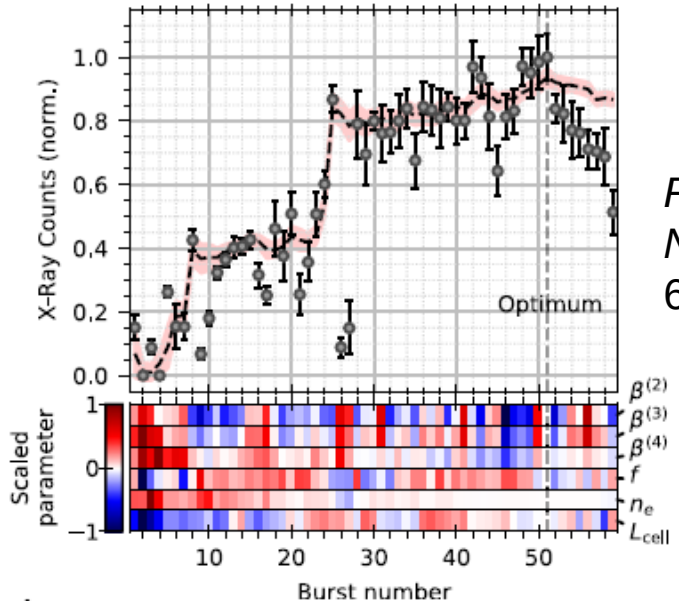
Compressor setup:

- PWFA requires high current bunches → coherent synchrotron radiation in compressors (complicated)
- Matching after BCs requires iteration loop of (multi-point) measurements and tuning (time-consuming)
- How to setup BCs optimally for plasma acceleration?

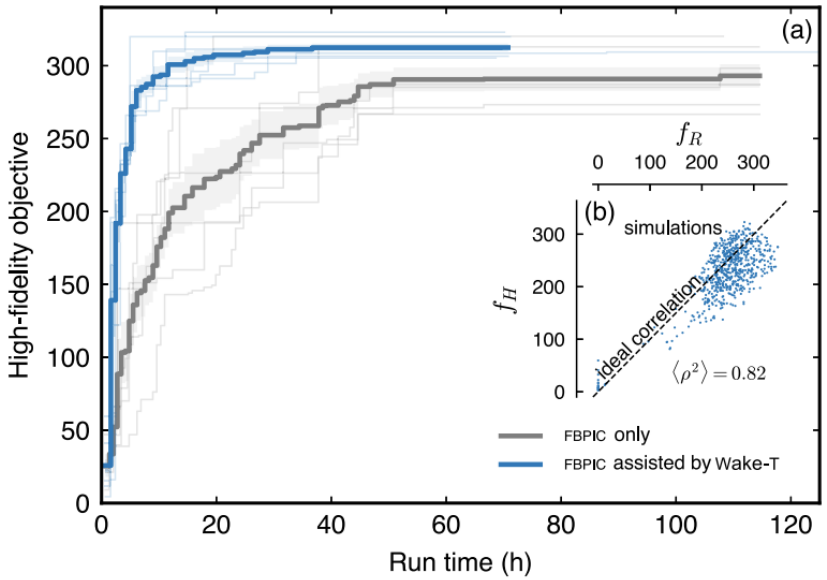
Examples already in (plasma) accelerator physics



S. Jalas et al., PRL, 126, 104801 (2021)



R. Shalloo et al., Nat. Comms, 11, 6355 (2020)



A. F. Pousa et al., PRAB, 26, 084601 (2023)

optimas

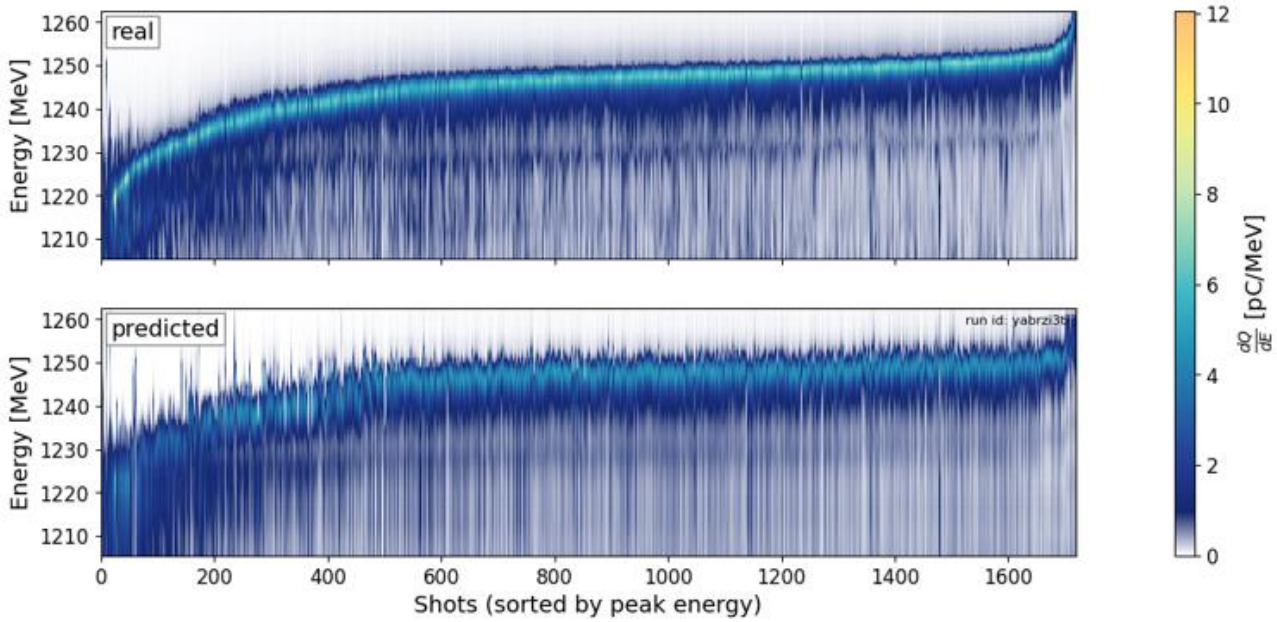
Ferran Pousa, A. et al. (2023). PRAB, 26(8), 084601.

Roussel, R., Mayes, C., Edelen, A., & Bartnik, A. (2023). IPAC2023, JACoW.

Xopt

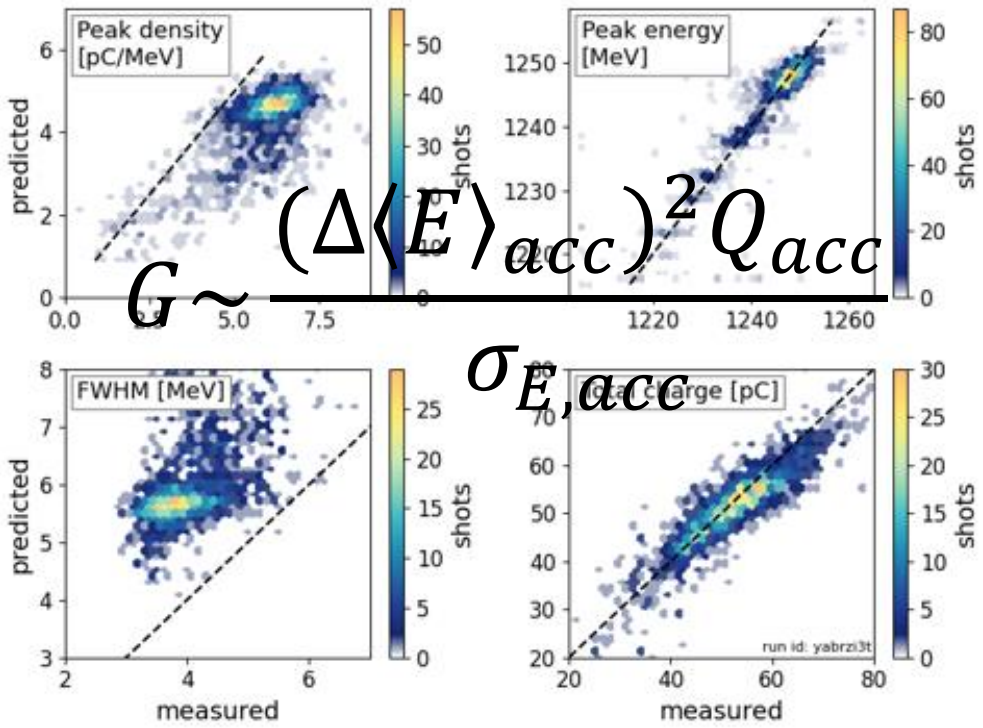
A Virtual Spectral Diagnostic for PWFA - Results

Courtesy Phillip Burghart (FLASHForward Masters Student)



- Peak energy and charge of accelerated bunches are quite well-predicted
  - Downstream BPMs have a lot of influence
- Energy spread and peak spectral density less so: why?
  - Beam loading is sensitive to the details of the wakefield and witness current profile: are we capturing that info shot-to-shot?

- Predict PWFA trailing bunch spectra from ~12000 shot stats dataset, 80:20 train test split
- Simple Multi-layer perceptron network: 3 hidden layers, 100 neurons wide, ReLU activation, MSELoss



$$G \sim \frac{(\Delta\langle E\rangle_{acc})^2 Q_{acc}}{\sigma_{E,acc}}$$