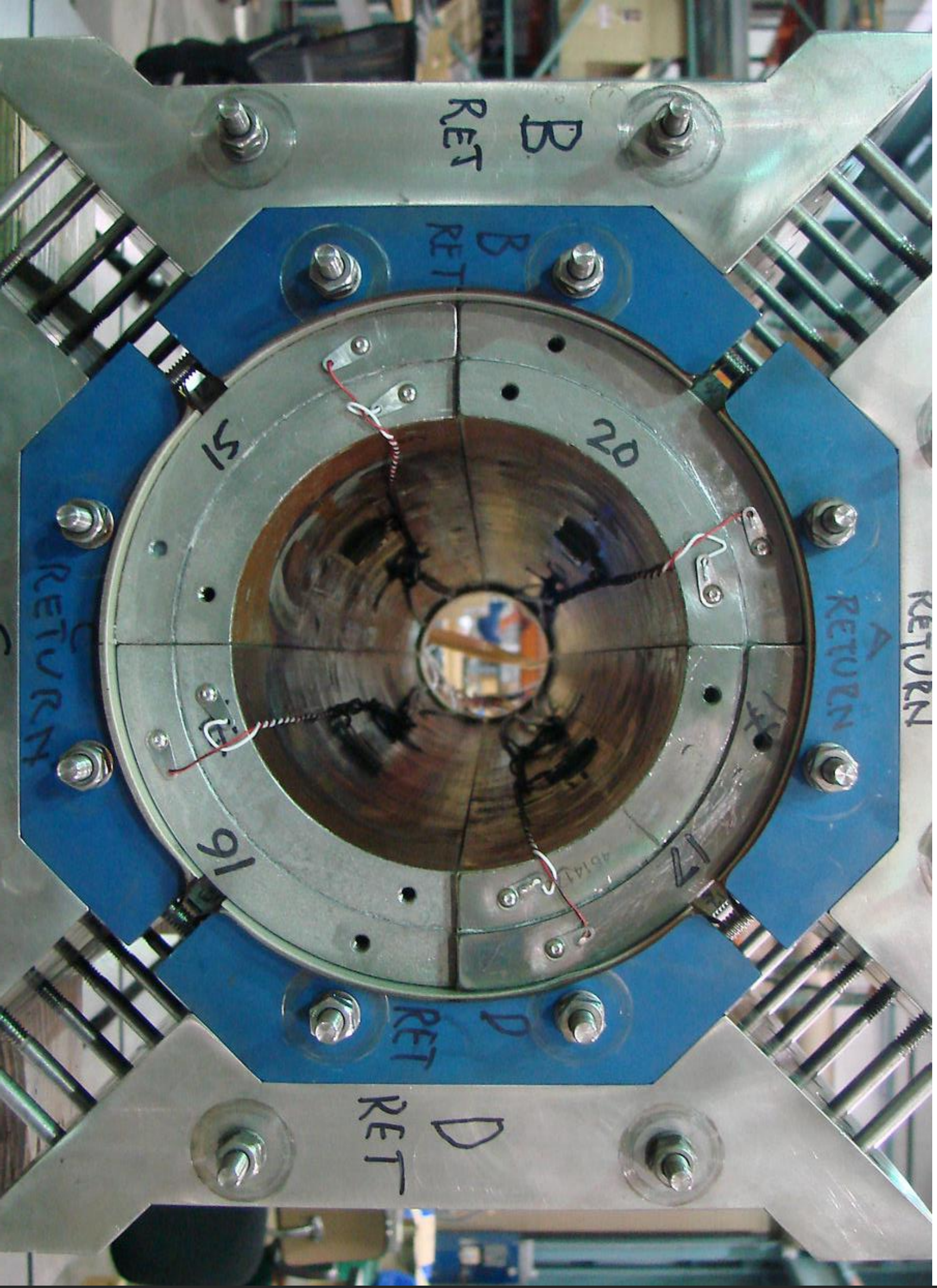


Preliminary report on

Dynamic Aperture Optimization of (proton) IOTA lattice v.8.4

with 12 sextupoles without space-charge and RF cavity

KILEAN HWANG, CHAD MICHELL, ROBERT RYNE

LBNL

9/28/2020

https://www.symmetrymagazine.org/article/july-2013/future-lhc-super-magnets-pass-muster

Introduction

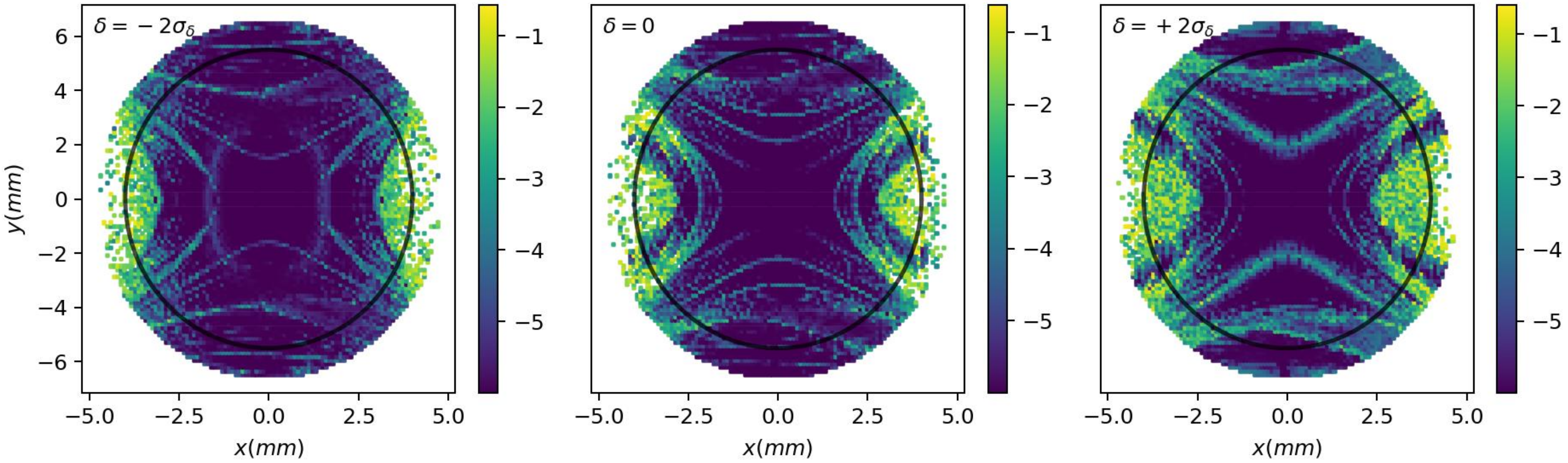
- Previous proton loss study showed that the sextupole settings of the IOTA v.8.4. lattice are historical:
  - Compared to the zero-strengh sextupoles, the existing settings resulted in about twice more particle losses
    - without space-charge and RF but
    - with beam energy spead, dipole and quadrupole fringe fields and nonlinear kinetic effects
- In this report we present our efforts to achieve
  - Chromatic integrablility
  - Dyamic Aperture (DA) Optimization (at NL (NonLinear insert) mid-point)
    - with realistic physical aperture

Previous

Proton Loss Study

SextOff

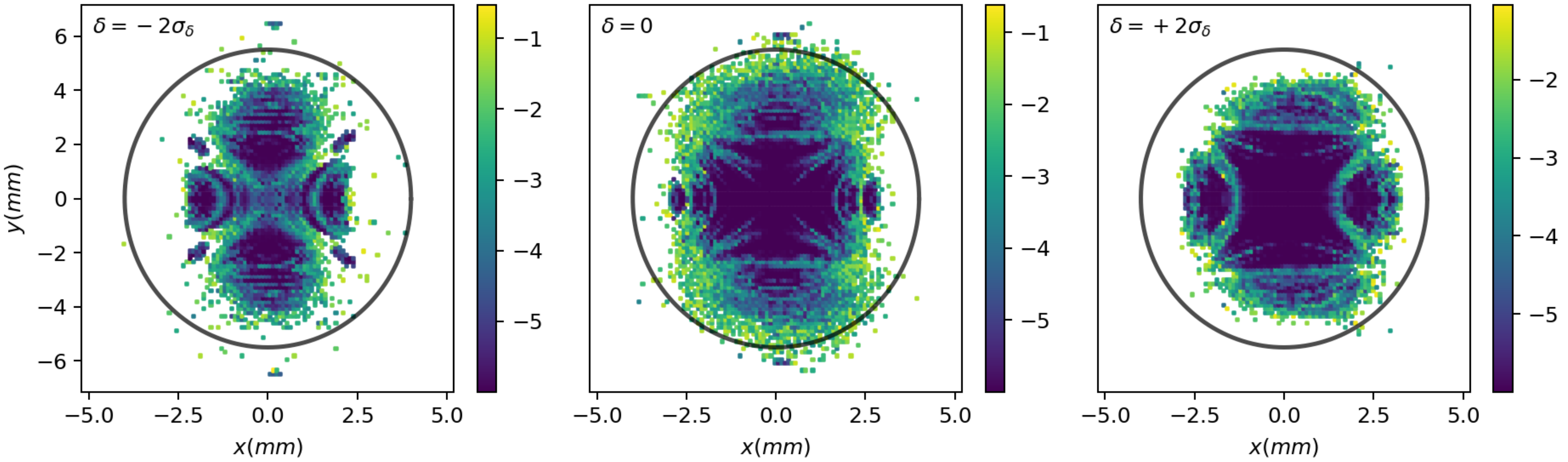
Dynamic Aperture plot using a chaos indicator at NL-mid point



- Black circle is the physical aperture
- Chaotic orbits near the singular point
- Recall that we saw less particle loss for lower energy particles when sextupoles turned off

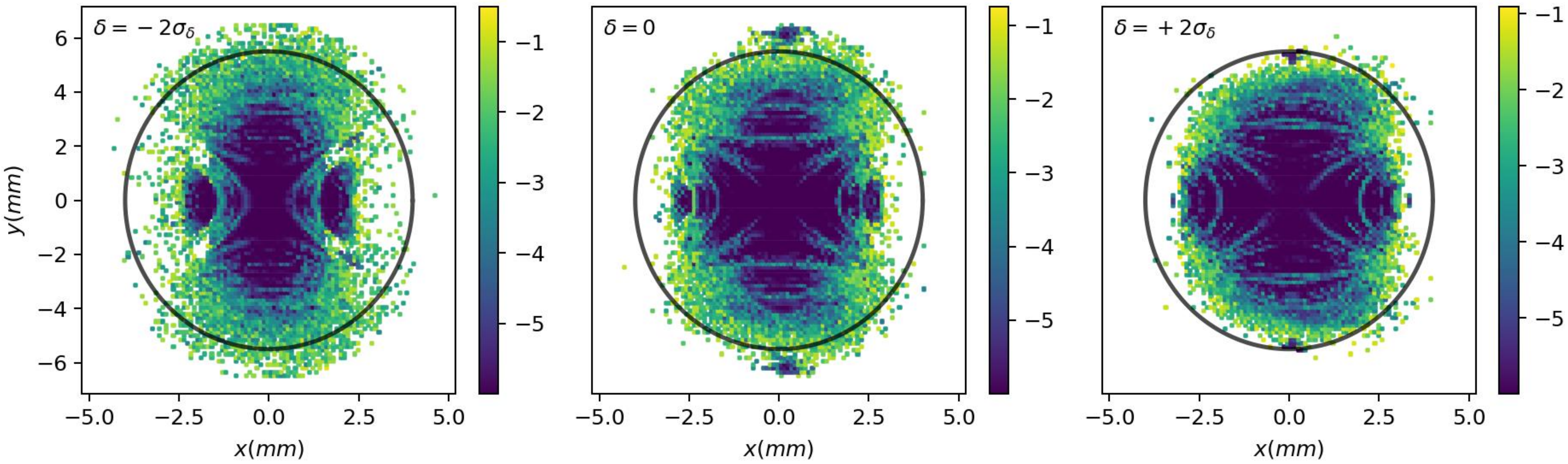
https://indico.fnal.gov/event/43326/contributions/185983/attachments/128391/155302/LossStudy4.pdf

SextOn(Historical setting)



https://indico.fnal.gov/event/43326/contributions/185983/attachments/128391/155302/LossStudy4.pdf

SextOn(GoodDA)



https://indico.fnal.gov/event/43326/contributions/185983/attachments/128391/155302/LossStudy4.pdf

Chromatic Invariant

Normalization

$$p_{x,\delta} = \frac{p_x}{1+\delta}$$

$$\begin{aligned} H &= -\sqrt{(1+\delta)^2 - p_x^2 - p_y^2} - a_s \\ &\simeq \frac{p_x^2 + p_y^2}{2(1+\delta)} + V(\boldsymbol{x}, s) \end{aligned}$$

$$H = \frac{p_{x,\delta}^2 + p_{y,\delta}^2}{2} + \frac{K(s)}{(1+\delta)} \frac{x^2+y^2}{2} + \frac{V(x,y,s)}{(1+\delta)}$$

Chromatic Invariants

In such coordinate system, the *chromatic twiss functions* satisfy

$$\frac{\alpha^2}{\beta}-\left(\alpha'+2\frac{\alpha^2}{\beta}\right)+\frac{K\beta}{(1+\delta)}=\frac{1}{\beta}$$

Since $K=0$ at the NL insert section,

$$\beta(s,\delta)=\beta(s,0)$$

provided that

$$\beta(0,\delta)=\beta(L,\delta)=\beta(0,0)$$
$$\alpha(0,\delta)=-\alpha(L,\delta)=\alpha(0,0)$$

and $\Psi(\delta)=n\pi$ where $s=0$, L are at the NL entrance and and exit respectively, and $\Psi(\delta)$ is the phase-phase-advance over the arc section (from NL exit to entrance)

Therefore, NL potential and invariants can be defined in a usual way, i.e.,

$$\frac{\beta}{(1+\delta)}V\left(c\sqrt{\beta}\boldsymbol{x}_n,\psi\right)=-\frac{\tau}{1+\delta}U(\boldsymbol{x}_n)$$

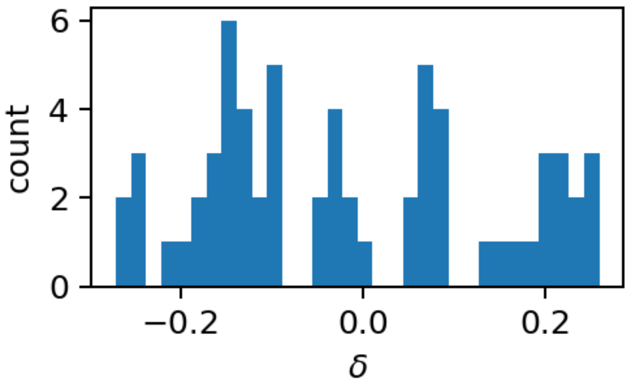
$$H=\frac{\boldsymbol{p}_n^2}{2}+\frac{\boldsymbol{x}_n^2}{2}-\frac{\tau}{1+\delta}U(\boldsymbol{x}_n)$$
$$I=(x_n p_{yn}-y_n p_{xn})^2+p_{xn}^2+\boldsymbol{x}_n^2-\frac{\tau}{1+\delta}W(\boldsymbol{x}_n)$$

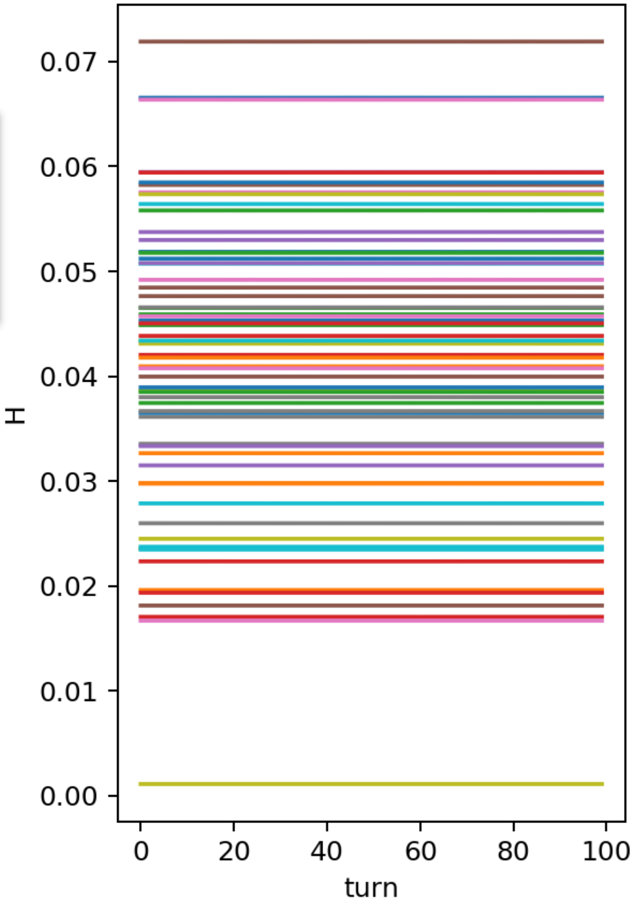
Test 😀

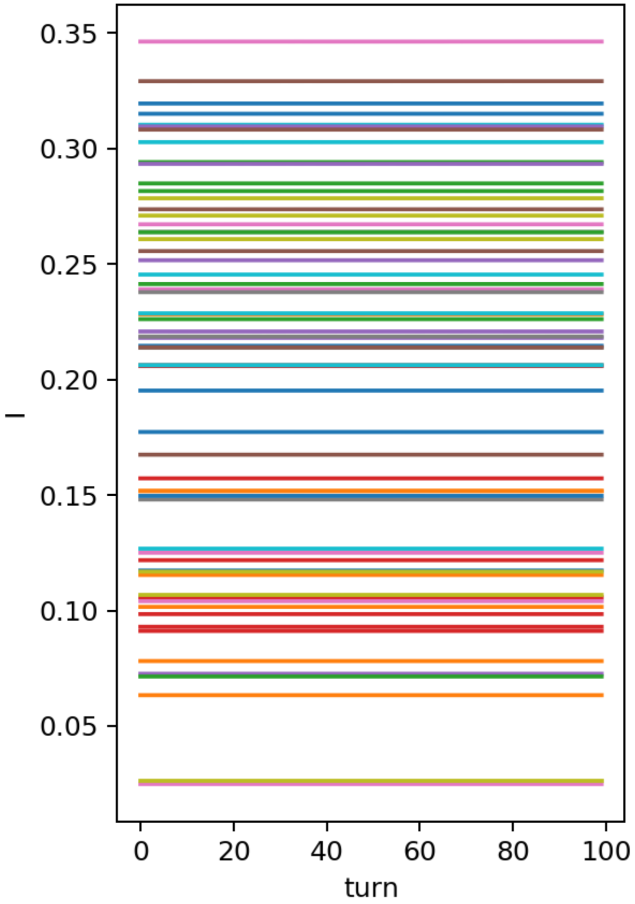
Test using MaryLie with 64 particles, using ideal arc map such that

$$\beta(0,\delta) = \beta(L,\delta) = \beta(0,0)$$
$$\alpha(0,\delta) = -\alpha(L,\delta) = \alpha(0,0)$$

$$\Psi(\delta) = n\pi$$







Problem ☹

However, matching the chromatic twiss functions while keeping chromatic tune-advance to be multiple of half-interger and nonlinear resonance strengths to be nearly zero is practically impossible.

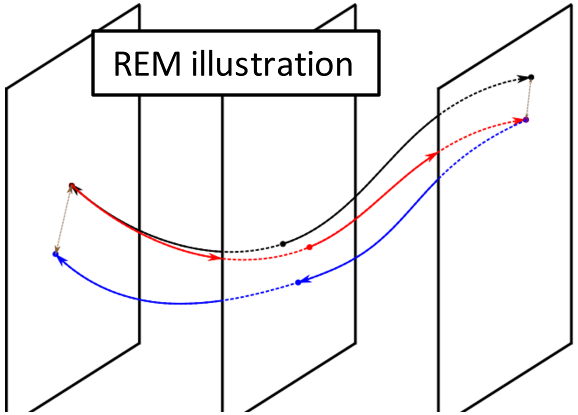
(e.g.) We tried to optimize 12 sextupoles to make the arc map as close to ideal as possible using MaryLie but optimized map did not improve conservation of the invariants

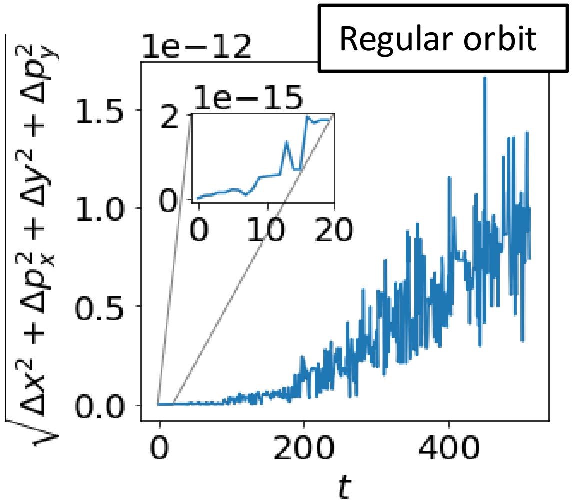
Instead, we took another route: directly optimizing dynamic aperture using a fast chaos indicator

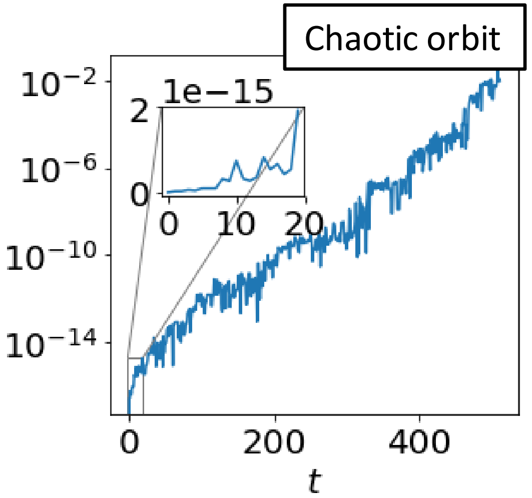
Dynamic Aperture Optimization

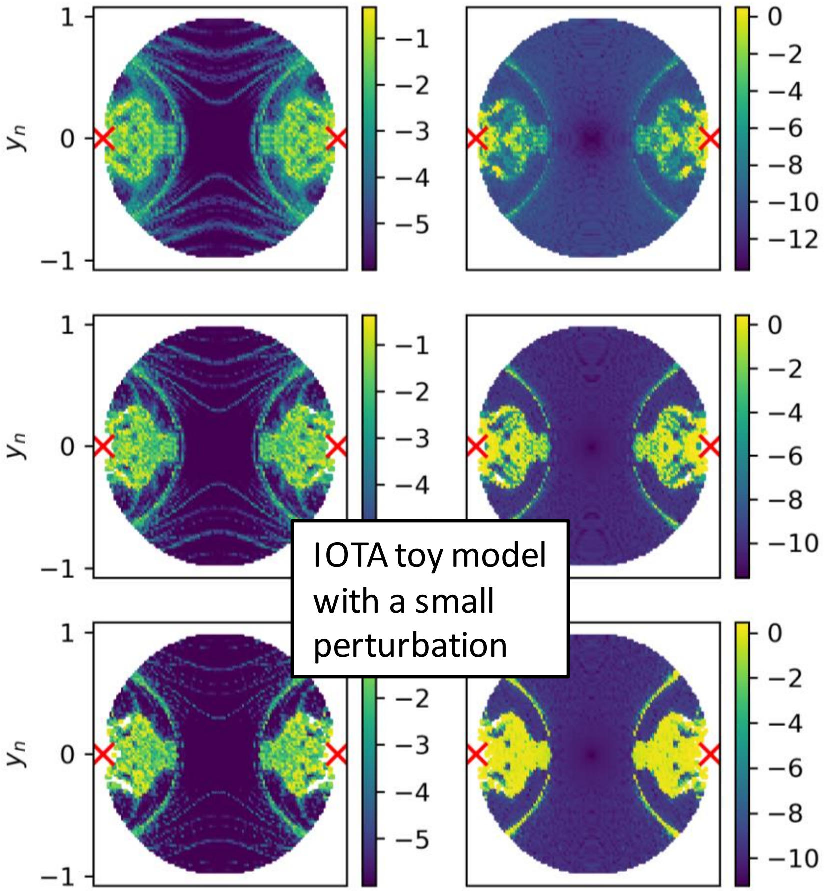
Reversal Erorr Method (REM)

REM is a fast converging chaos indicator which can be used to estimate dynamic aperture efficiently in terms of computational burden.









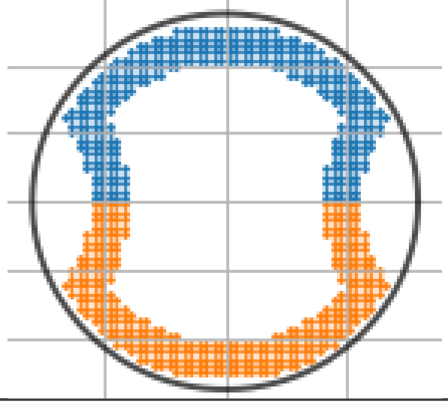
128 turn

256 turn

512 turn

Kilean Hwang, Chad Mitchell, Robert Ryne, "*Rapidly converging chaos indicator for studying dynamic aperture in a storage ring with space charge*", PRAB **23**, 084601(2020)

Particles for REM

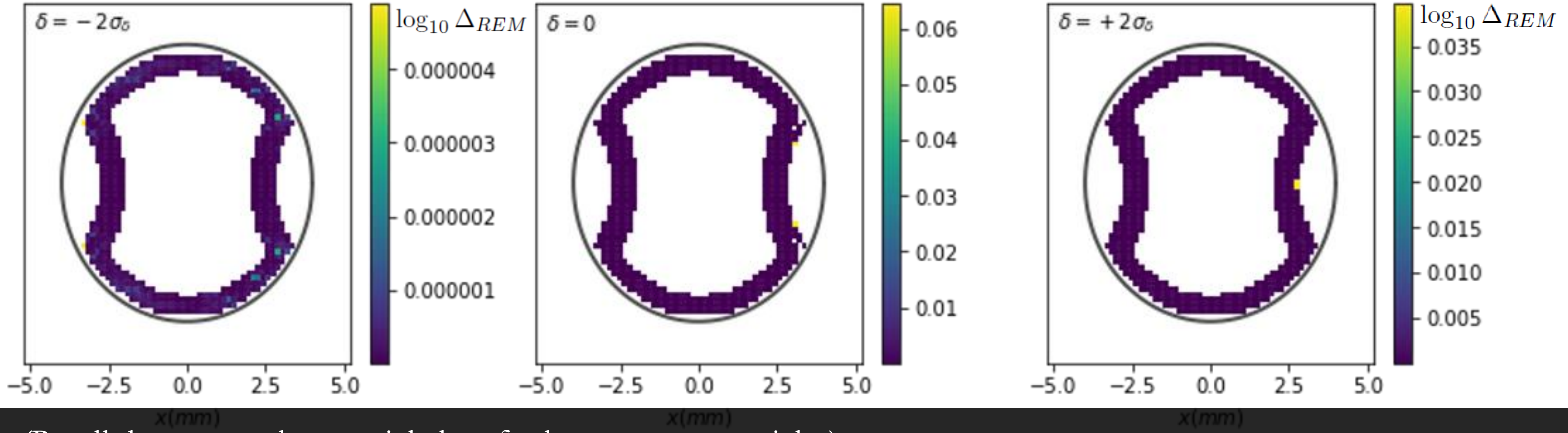


1. prepare particles at the NL mid-point, near the physical aperture and far from sigular points, on y>0, at $\delta = 0, \pm 2\sigma_\delta$

2. Track particles 100 turns forward and then backward. Get the REM error for each particles

$$\Delta_{REM} \equiv \sqrt{\Delta x^2 + \Delta p_x^2 + \Delta y^2 + \Delta p_y^2}.$$

For example, when all the sextupoles have zero strength:



(Recall that we saw less particle loss for lower energy particles)

Define Objective for DA optimization

Define objective function (let's call it by "loss" following ML terminology)

$$\text{loss} = \log_{10}\left(0.7\,\langle\Delta_{REM}\rangle_{\delta=0} + 0.3\,\langle\Delta_{REM}\rangle_{\delta=\pm\sigma_\delta}\right)$$

where the braket is the ensemble average over particles satisfying the condition specified by the subscripts.
In order to penalize the lost particles, we assign to all lost particles the maximum REM error (where the maximum is taken ov er all surviving particles)

Is the loss well defined? (1/3)

In order to check if lower loss gives us less particle loss,
we perform local minimization (over 12 sextupoles) starting from zero strengh sextupoles ($\boldsymbol{x}=\mathbf{0}$) where $\boldsymbol{x}$ is the vector of sextupole strengths. (it took about 10 days)

The initial loss is $\text{loss}_{\boldsymbol{x}=\mathbf{0}} = -3.12$

After local minization, we reached $\text{loss}_{\text{local min}} = -3.63$, and

$\boldsymbol{x}_{\text{local\_min}} = 0.5\sigma$ (6.7e-8, 1.5e-9, -9.9e-8, 1.4e-6, 7.3e-8, -9.9e-9, 3.2e-7, -1.1e-6, 3.8e-7, -1.8e-7, 2.7e-7, 4.6e-7)

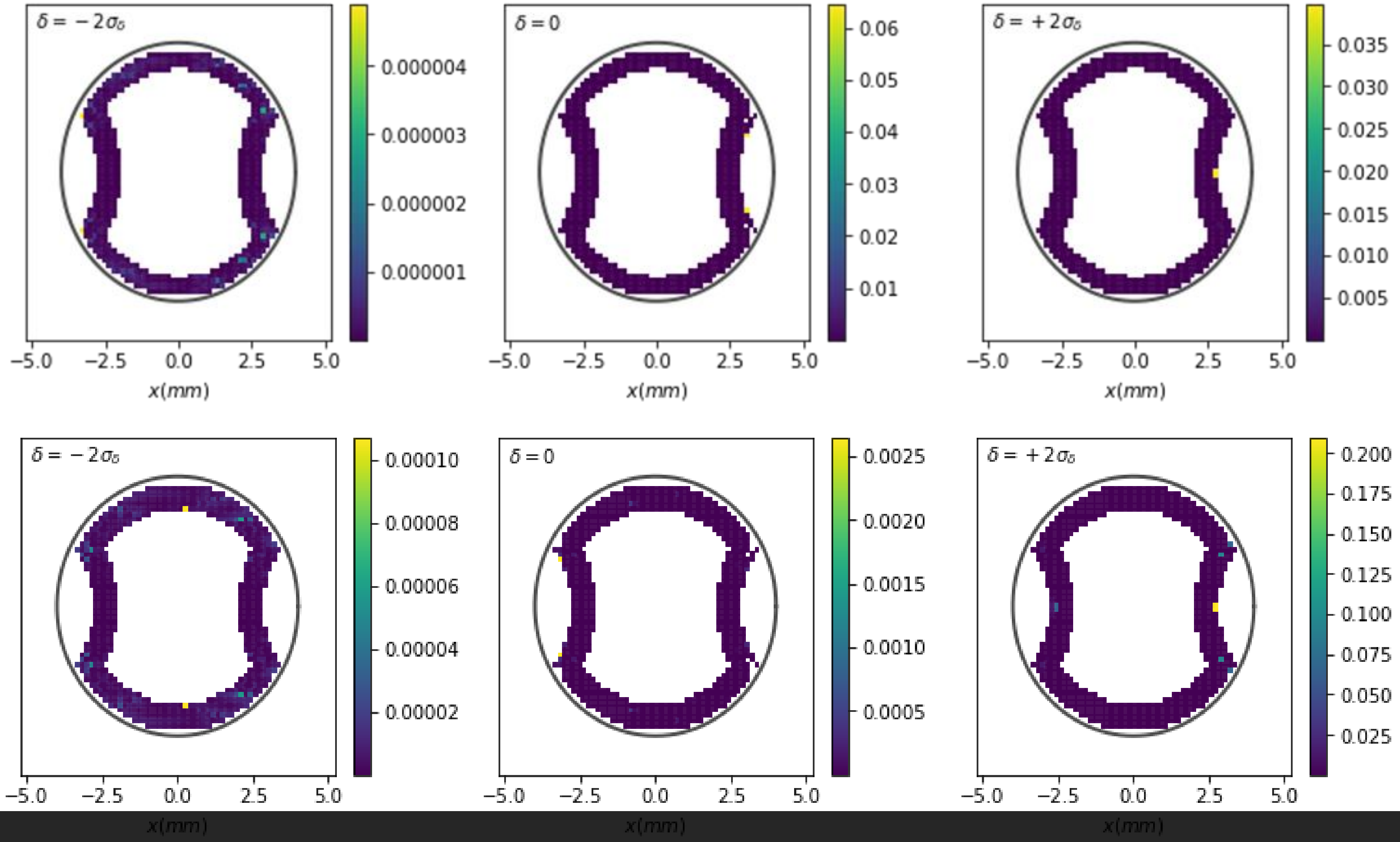
where σ is defined such that the largest sextupole strength of IOTA v.8.4. is at 2.5σ

As we already have loss study data at $\boldsymbol{x}=\mathbf{0}$, we perform another loss study at $\boldsymbol{x}_{\text{local\_min}}$

Is the loss well defined? (2/3)

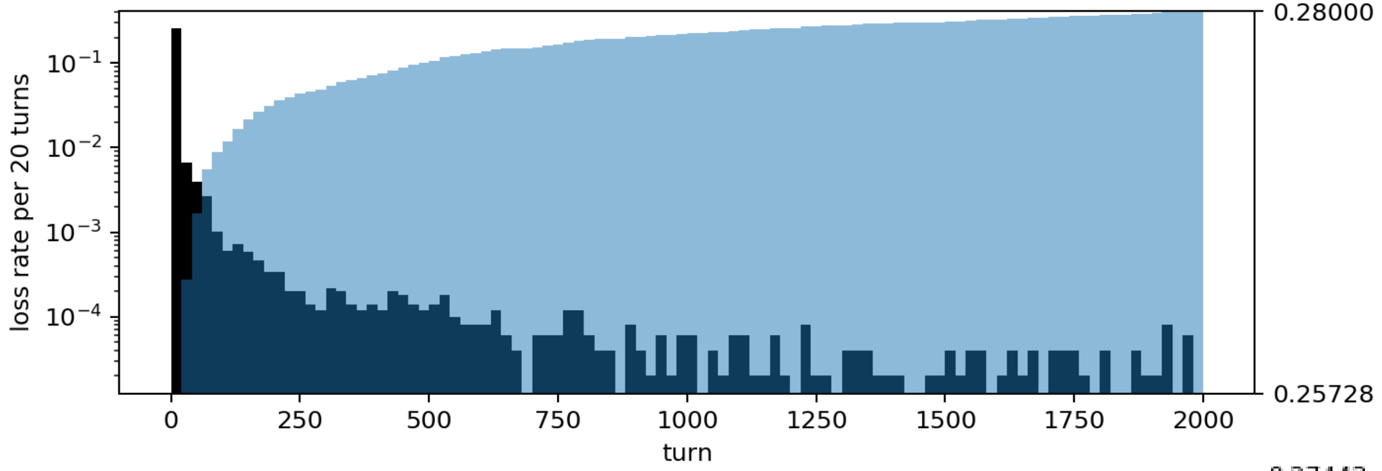
at $\boldsymbol{x}$=$\boldsymbol{0}$

at $\boldsymbol{x}_{\mathrm{local\_min}}$

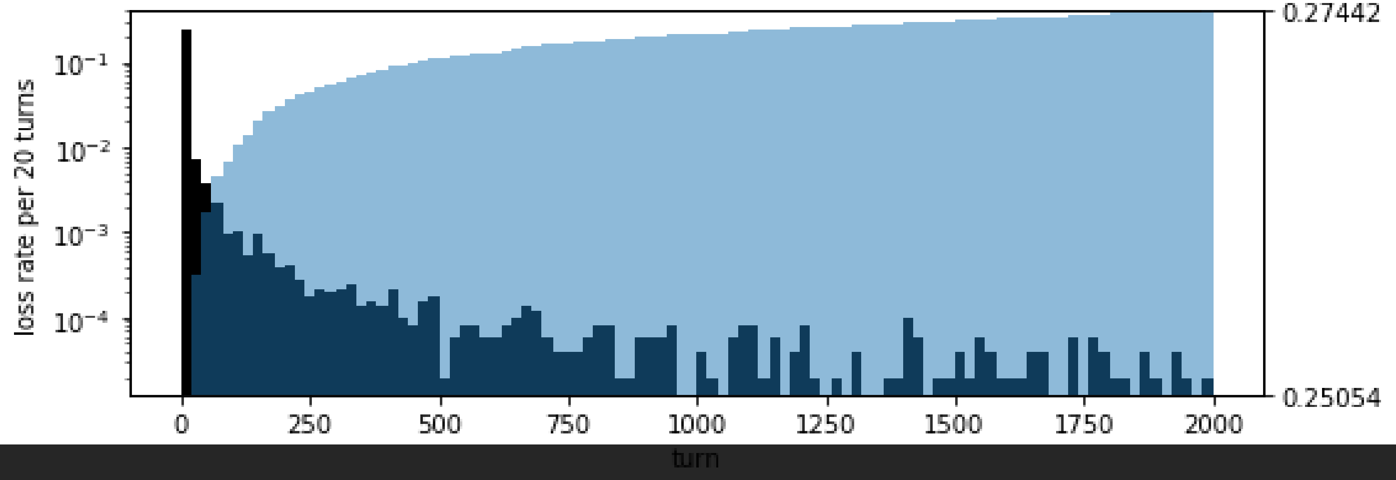


Is the loss well defined? (3/3)

at $\boldsymbol{x}$=$\boldsymbol{0}$



at $\boldsymbol{x}_{\mathrm{local\_min}}$



Global Optimization Strategy

- Computing time for evaluating a loss given a single input x is about 15 minutes.
  (Using a 4 core PC. We are trying to save our allocated computing time in NERSC for later use)

- Direct use of a population based global optimization like genetic-algorithm, particle swarm, and differential evolution is practically impossible.
  (e.g. local minimization took 10 days)

- Instead, we try iterative surrogate model based optimization[*]
  1. Randomly generate population of inputs $\boldsymbol{x}$, and evaluate loss of each to create initial data
  2. Train surrogate model (neural network model) on data
  3. Use global optimizer (differential evolution) on the surrogate model to generate optimized population
  4. Verify the optimized popuation by directly evaluating them
  5. Repeat 2-4

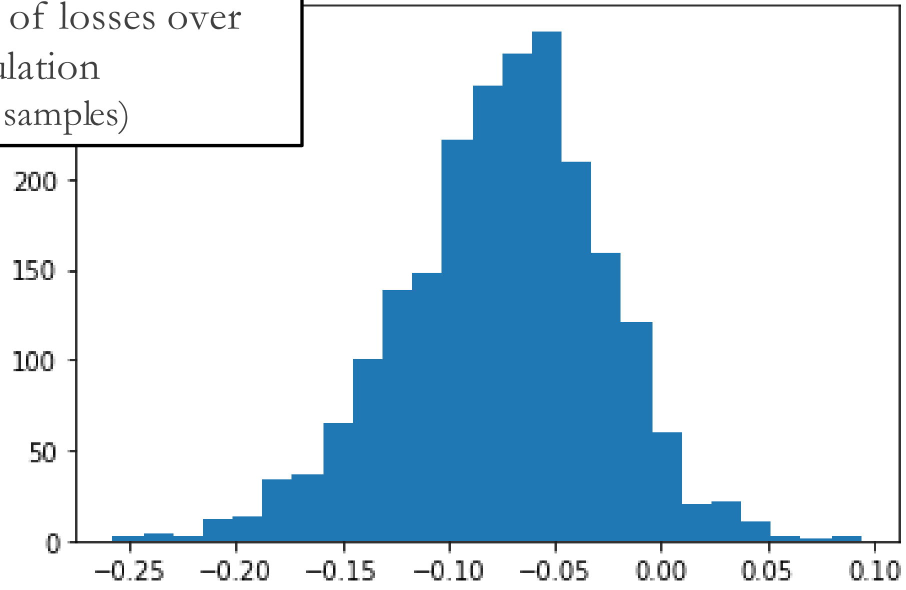
[*] https://link.aps.org/doi/10.1103/PhysRevAccelBeams.23.044601

Initial population

Randomly generated initial population $\boldsymbol{x}$ and evaluated loss $y(\boldsymbol{x})$ for each $\boldsymbol{x}$

- used gaussian random number generate with zero mean and a standard deviation σ , such that the largest sextupole strength of IOTA v.8.4. is at 2.5 σ

Histogram of losses over initial population (about 2200 samples)



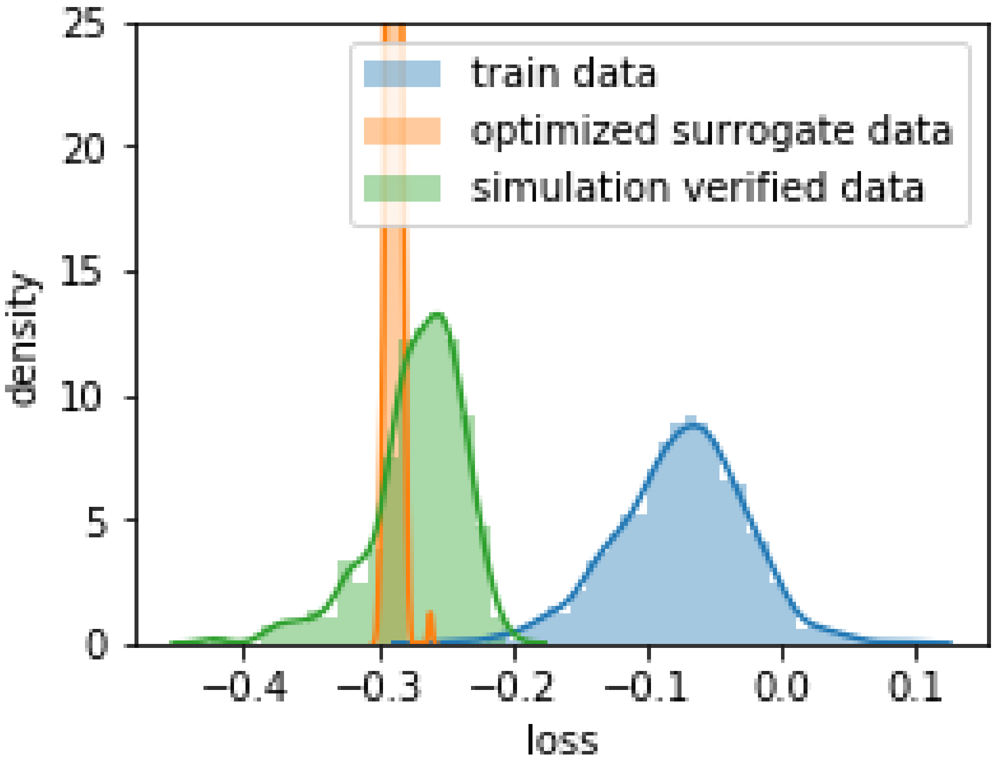
Recall that the loss at all zero sext strength was

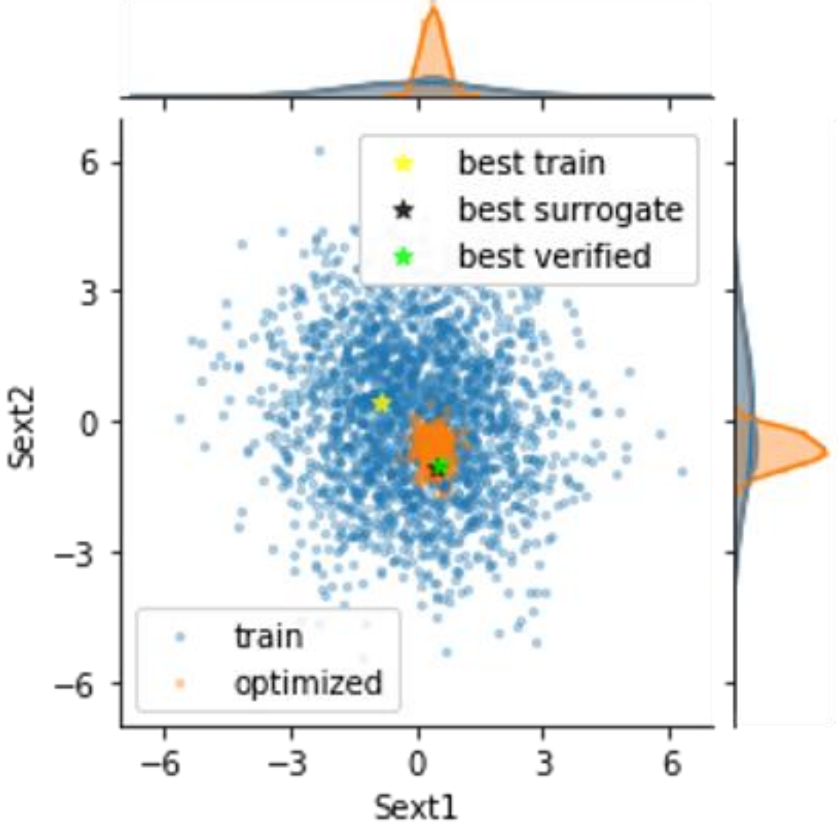
$$\text{loss}_{\boldsymbol{x}=\boldsymbol{0}} = -3.12$$

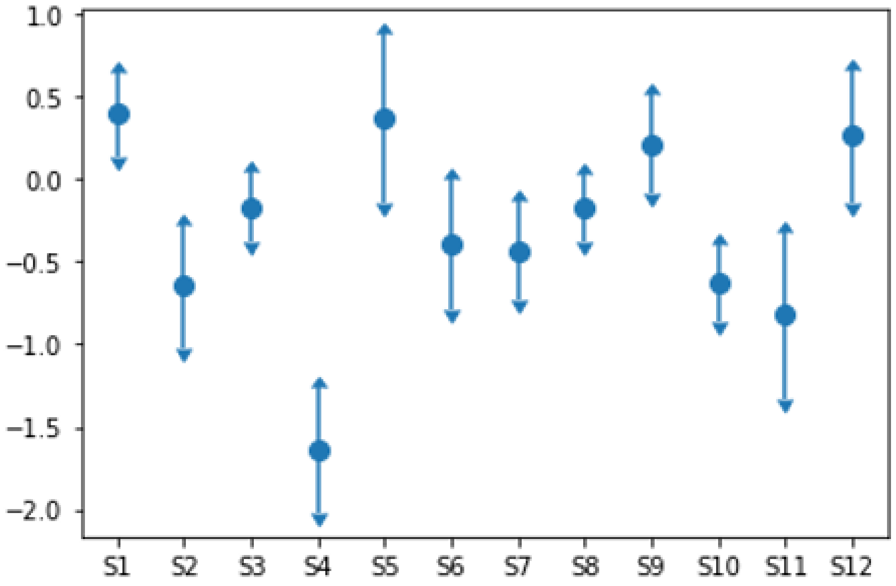
which is significantly smaller than the initial populations.

- is $\boldsymbol{x}$=**0** close to a deep and narrow local minimum ?
- is initial population too few?
- can the iterative surrogate model based optimization find smaller loss than $\boldsymbol{x}$=**0** case?

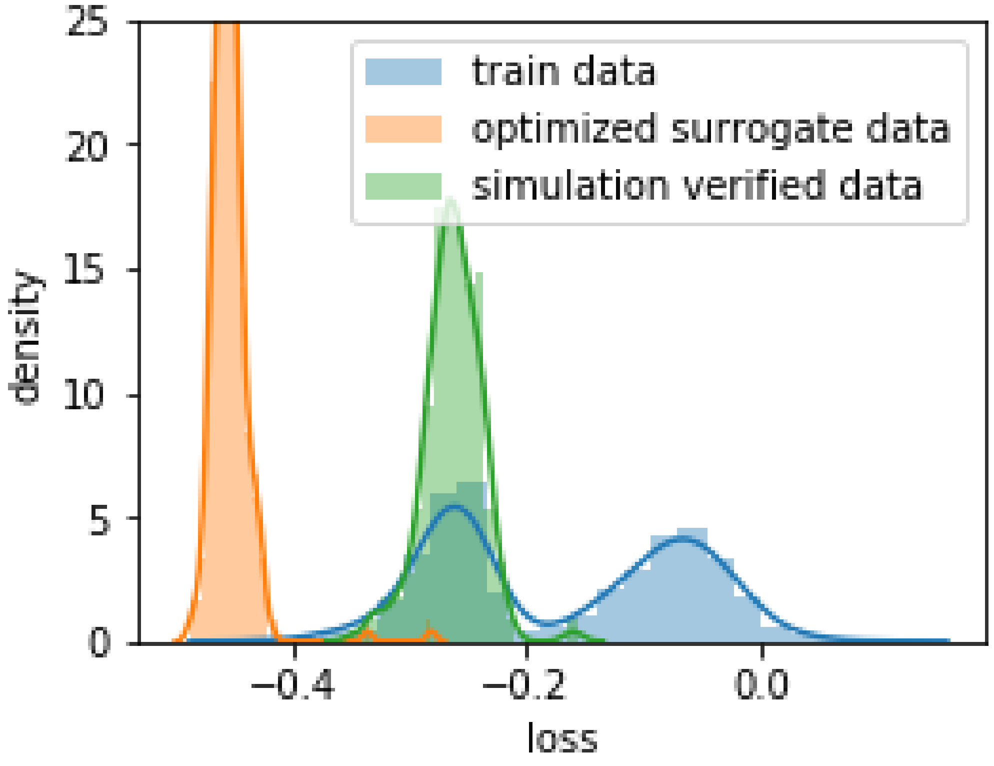
1st iteration 😃

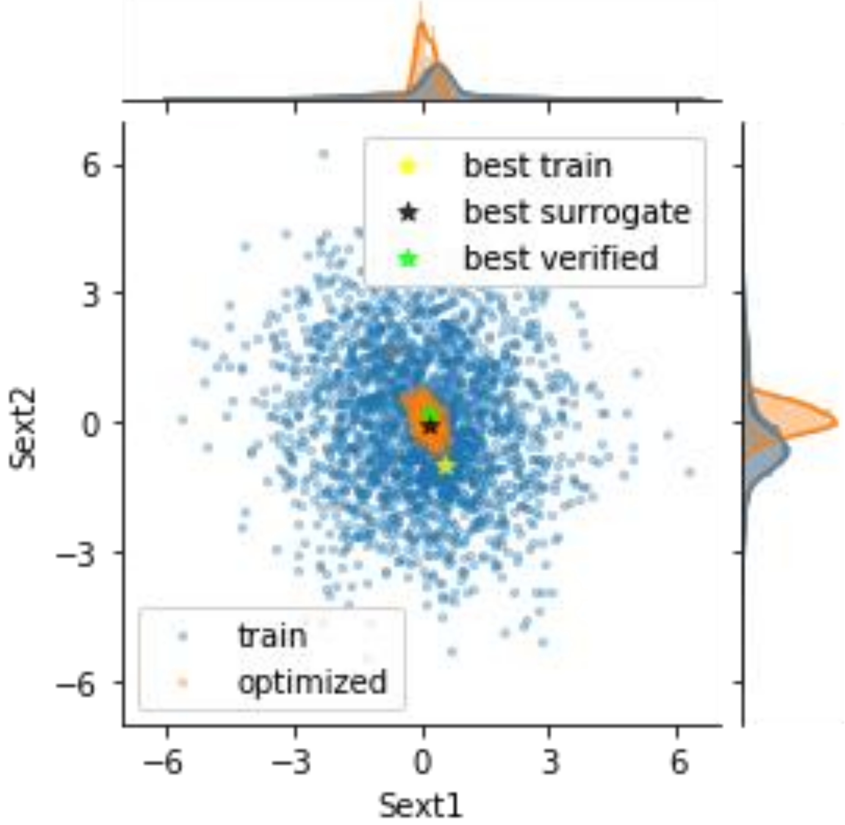


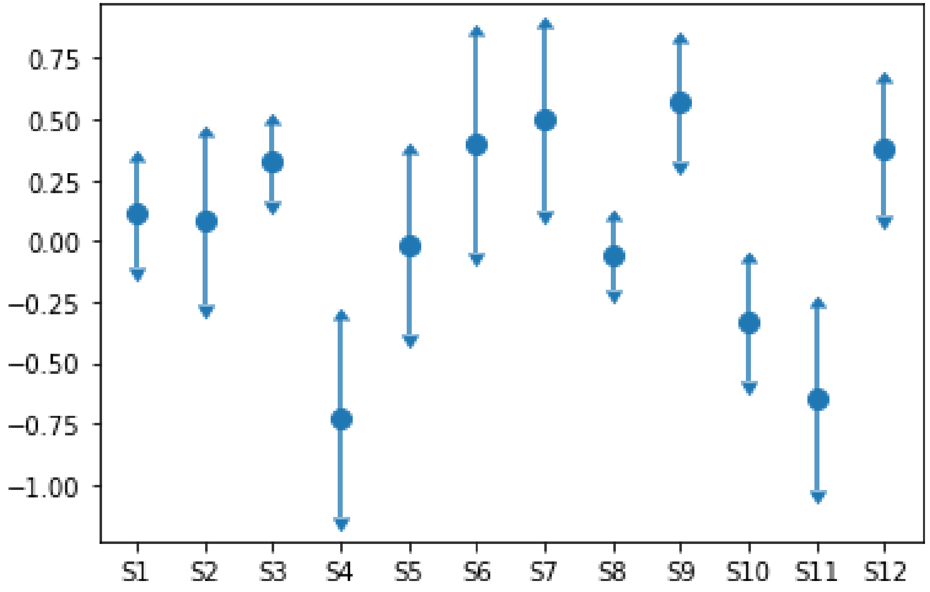




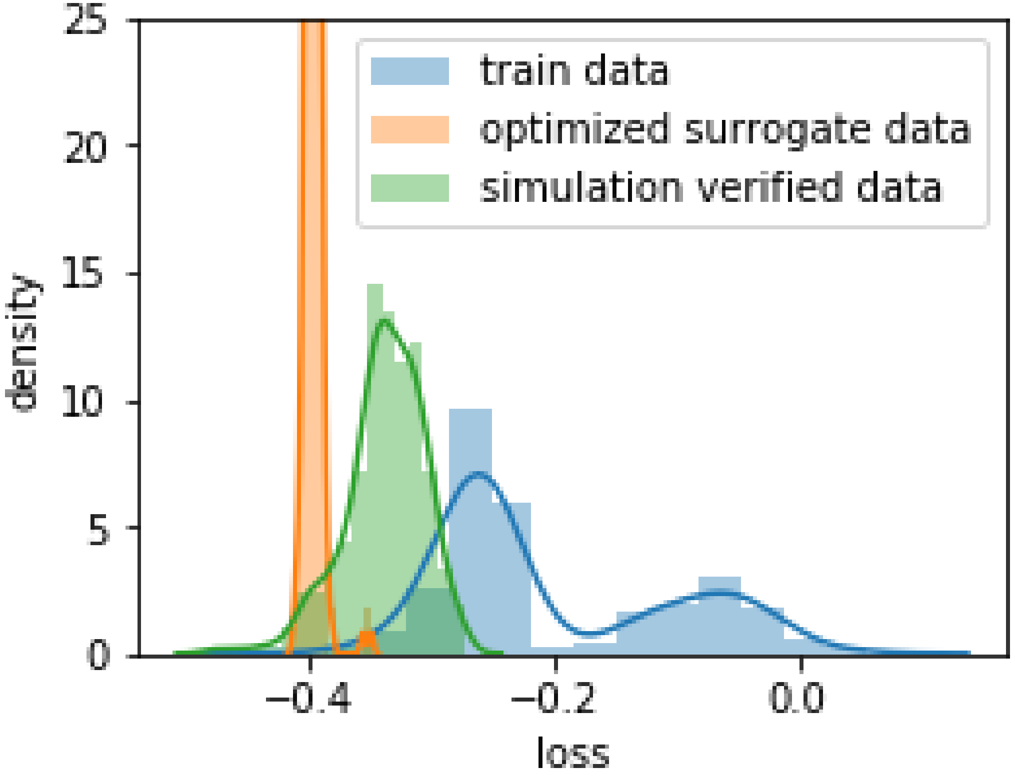
2nd iteration ☹

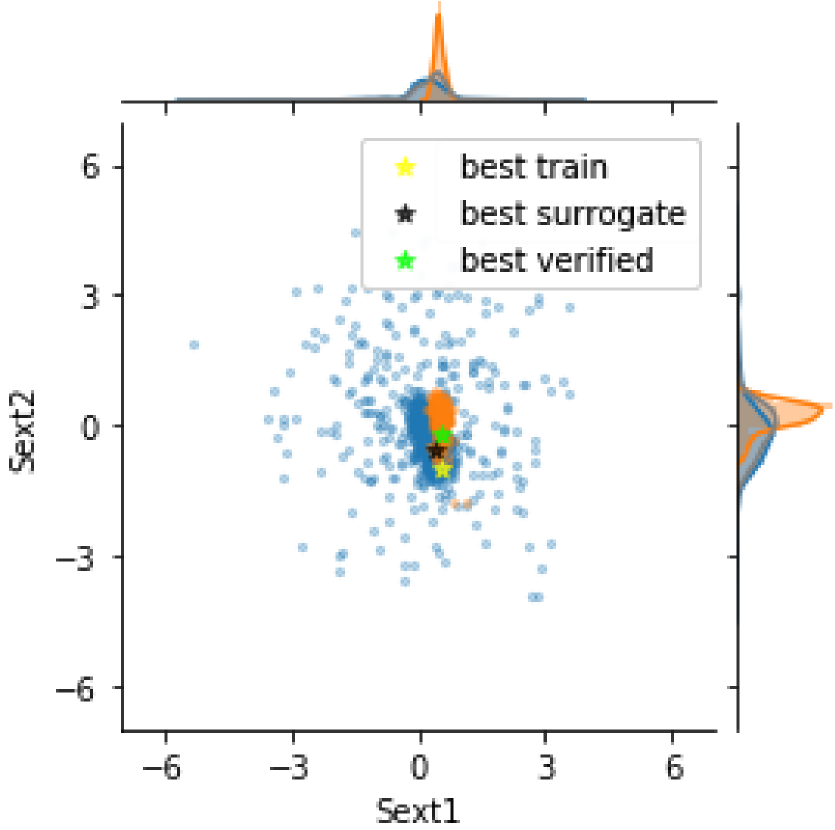


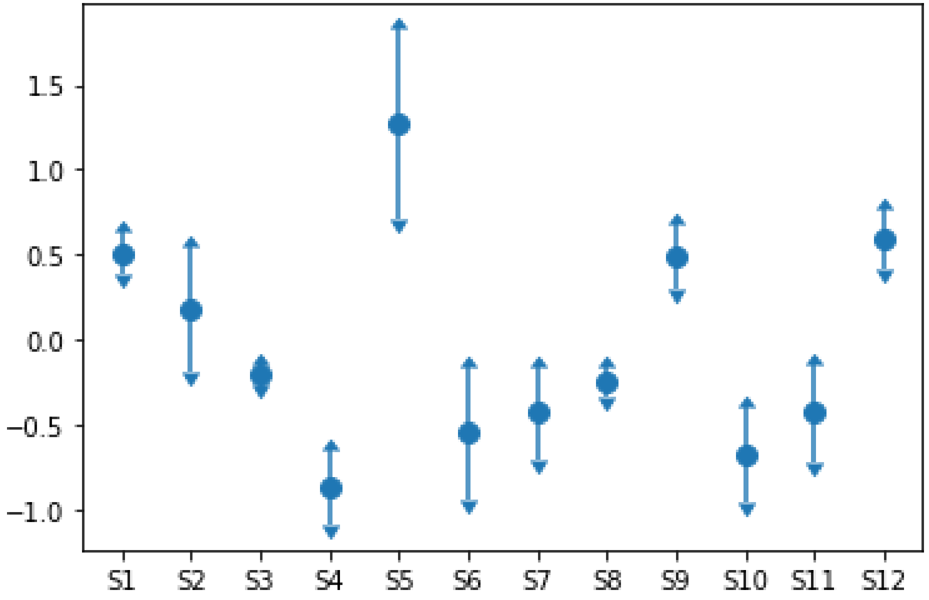




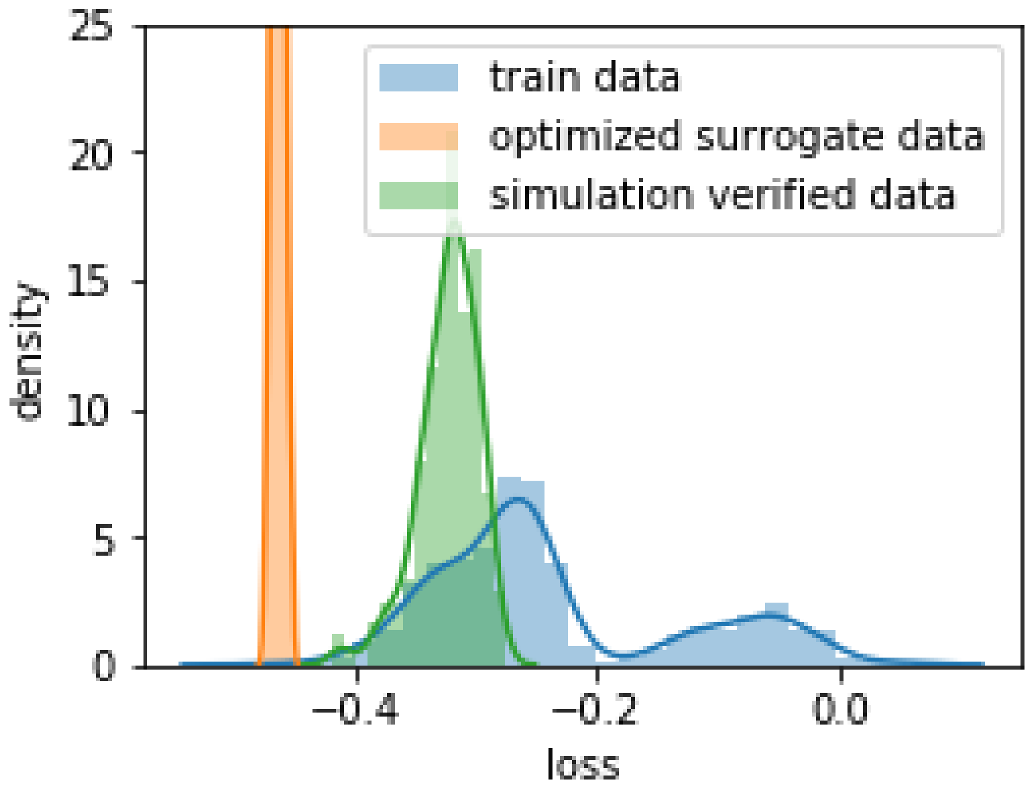
3rd iteration ☺

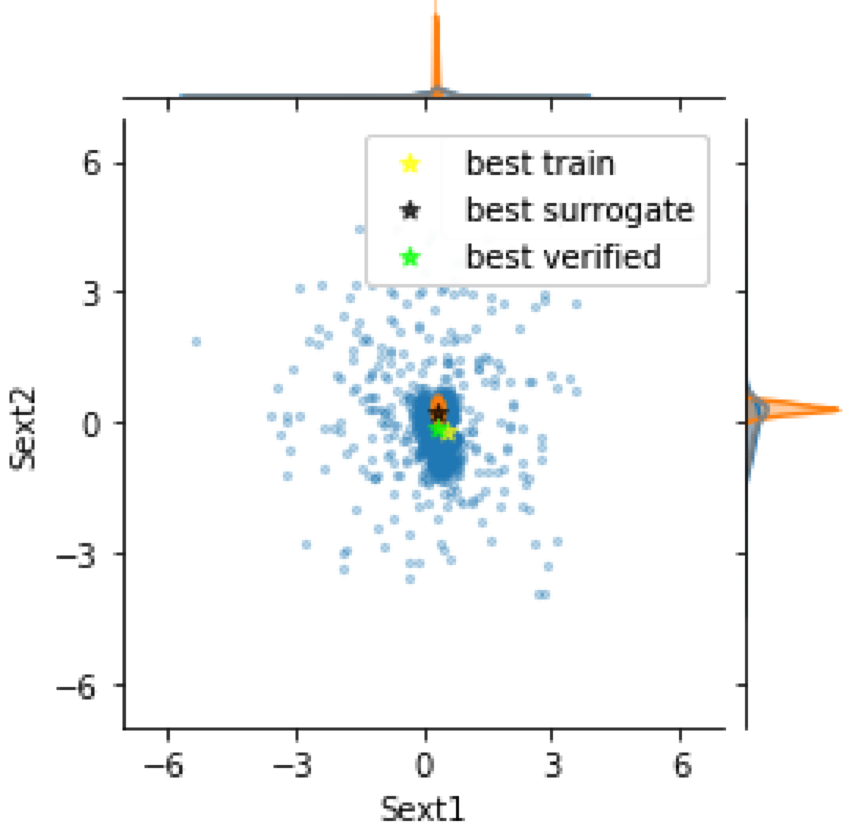


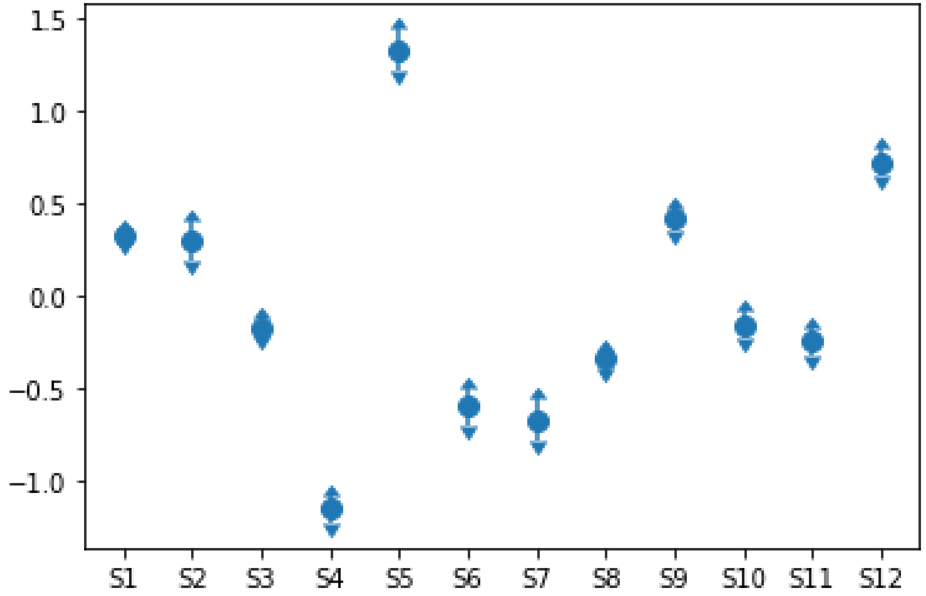




4th iteration ☹







What is going on?

- There is some discrepancy between surrogate model and simulation verification.
  - Not enough data?

    Initially 2200 samples and 256 samples added in each iteration. But the input dimension is hugh (12 sextupoles)
  - Lack of representability of the surrogate model?

    NN surrogate model has two layer of 1024 nodes each.

- $\boldsymbol{x}=\mathbf{0}$ is close to a very deep and narrow minimum?
  - We tested with $\boldsymbol{x}=(0.005\sigma, 0, 0, 0, \ldots)$. The evaluated loss at this test point was -1.5 (recall that the loss at $\boldsymbol{x}=\mathbf{0}$ was –3.12, and the minimum loss at 4th iteration of the optimization was about -0.5)

- NN surrogate model based optimizer converge very fast (many order of magnitude faster).
  - But can it really find global minimum?

Maybe

NN surrogate miss the global minimum ☹

Consider a 1D potential:

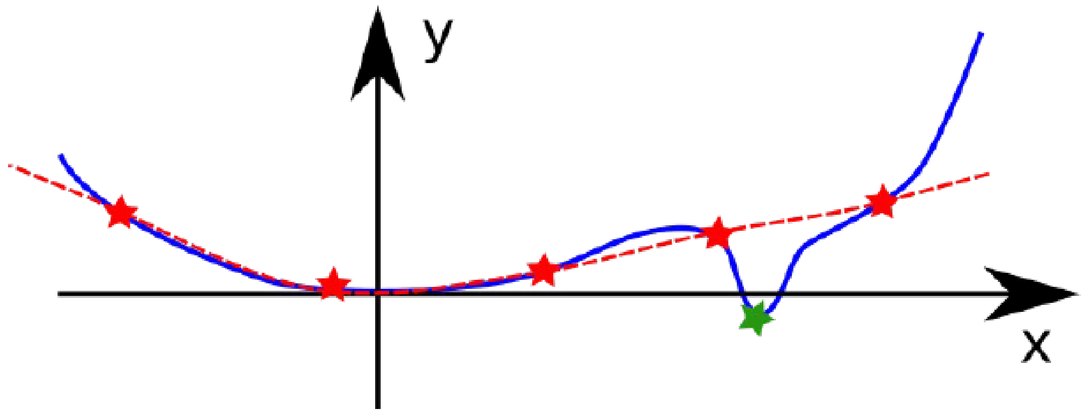


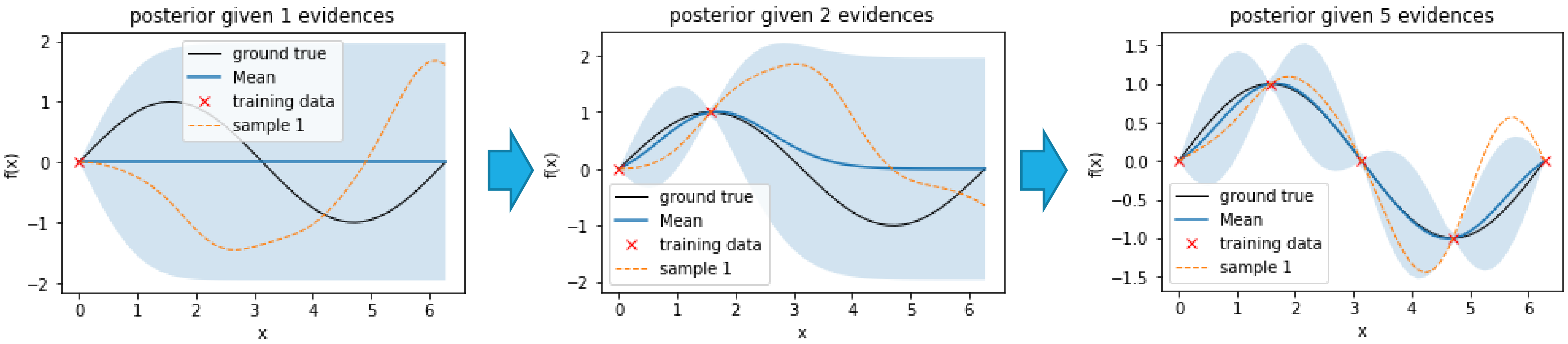
Fig. objective function (y) over input parameter (x). Blue line is the ground true objective function, green star is the ground true global optimum, red stars are (randomly) chosen hypothetical initial population. Red dotted line represent hypothetical surrogate model.

- ❑ This problem can be improved by penalizing the model uncertainty at a test point $\boldsymbol{x}^*$.
  The model uncertainty at $\boldsymbol{x}^*$ is likely large if it is far from training input data $\boldsymbol{x}$

Solution? 😀

Bayesian surrogate model (e.g. Gaussian Process)

Gaussian process surrogate model can give us not only the estimated loss $y(\boldsymbol{x})$ but also uncertainty of the loss $\pm\Delta y(\boldsymbol{x})$. For example,



Then, the global optimizer can use modified loss (called acquisition function) which penalize model uncertainty, e.g. $f(\boldsymbol{x}) = y(\boldsymbol{x}) - \Delta y(\boldsymbol{x})$

Still problematic ☹

Choice of Prior. Curse of dimensionality.

- However, we still need large sample data as the input dimension is large: dim($\boldsymbol{x}$)=12.
- If prior is close to real loss $y(\boldsymbol{x})$, the required sample number can be greatly reduced.
  - However, creating prior over 12 dimensional input is still computationally heavy.

    A prior can be obtained by much coarser simulation. e.g. use smaller number of particles, larger integration step size
- Maybe a dimensionality reduction method can be a solution

  We are still working on it

Summary/Conclusion/FeedBack?

- Chromatic Integrability is practically hard to achieve
- DA optimization over 12 sextupoles was carried out but could not find better solution than the all zero sextupole strength case ($\boldsymbol{x}=\mathbf{0}$).
  - Maybe NN surrogate model based optimizer is missing global minimum
  - But the *curse of dimensionalty* seems to be the main problem
- Even after we improve the optimization method, the optmizer still may not find a better solution than $\boldsymbol{x}=\mathbf{0}$.
  - The average REM error of $\boldsymbol{x}=\mathbf{0}$, is about 100 times smaller than the optmized result and 2200 sample.
- Are all the 12 sextupoles adjustable?
- Is there any viewgraph how the historical sextupole settings improved chromatic invariants or dynamic aperture?
- Any suggestions?