Phys. Rev. Accel. Beams 24, 014601 (2021)

Multiobjective optimization of the dynamic aperture for SLS 2.0 using surrogate models based on artificial neural networks

Autors: Kranjčević, Marija; Riemann, Bernard; Adelmann, Andreas; Streun, Andreas

Ref [1 … 42] can be found in the above mentioned paper

Outline:
Multiobjective optimization
ANN
DA optimization
Results

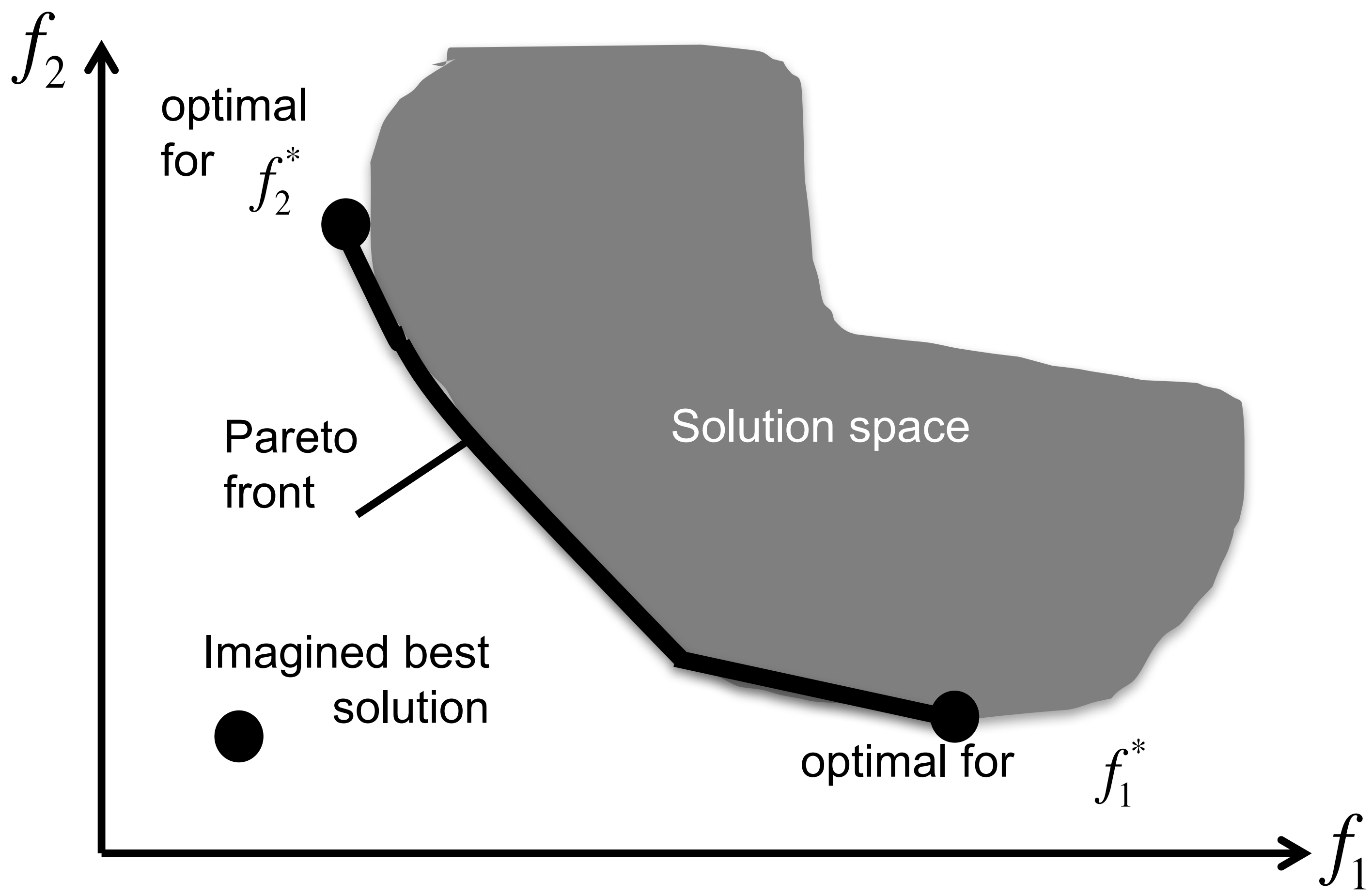
+ some personal remarks

Multiobjective optimization

The optimization problem is multi-objective

$$\min(f_1(x), f_2(x), \ldots)$$

(several criterions to optimize)



and addition constraint has to be full filled

- This has to be quantified mathematical

- The criterions can be contradicting:
Improving one criterion means worsening others

- Find set of optimal solutions instead of single solution (Pareto front)

Multiobjective optimization

- Genetic algorithms algorithms [3,9-14]
- Differential evolution [7,8]
- Particle swarm algorithms [6]

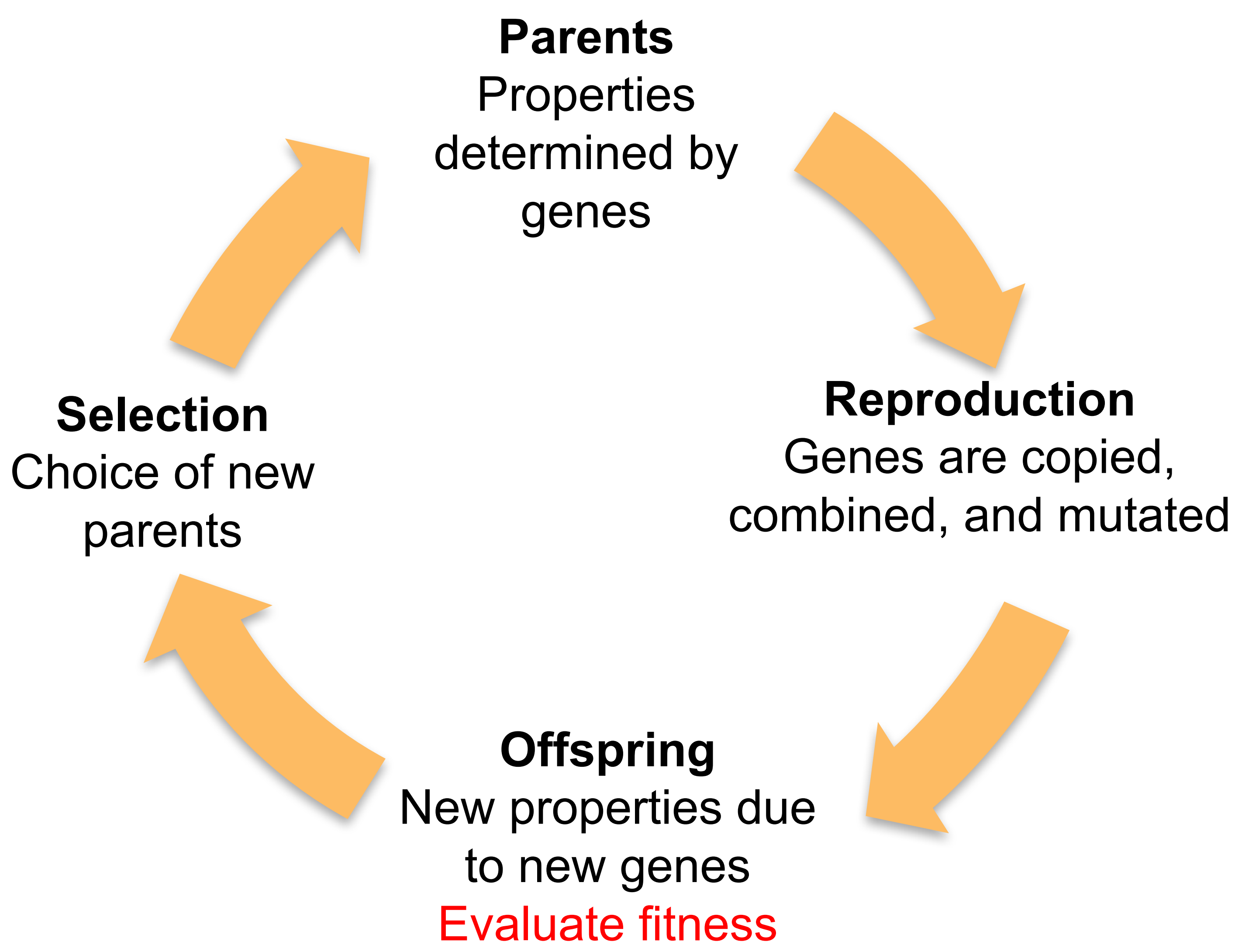
  - Standard for dynamic aperture optimization
  - Also be used for injection optimization [a]
  - MO optimization, in par. MOGA, is time and computer resources consuming

Implementation

[a] S. Appel, O. Boine-Frankenheim and F. Petrov: *Injection optimization in a heavy-ion synchrotron using genetic algorithms*, Nucl. Instr. and Meth. A 852 73–79 (2017)

Nature-inspired optimization

Genetic algorithms



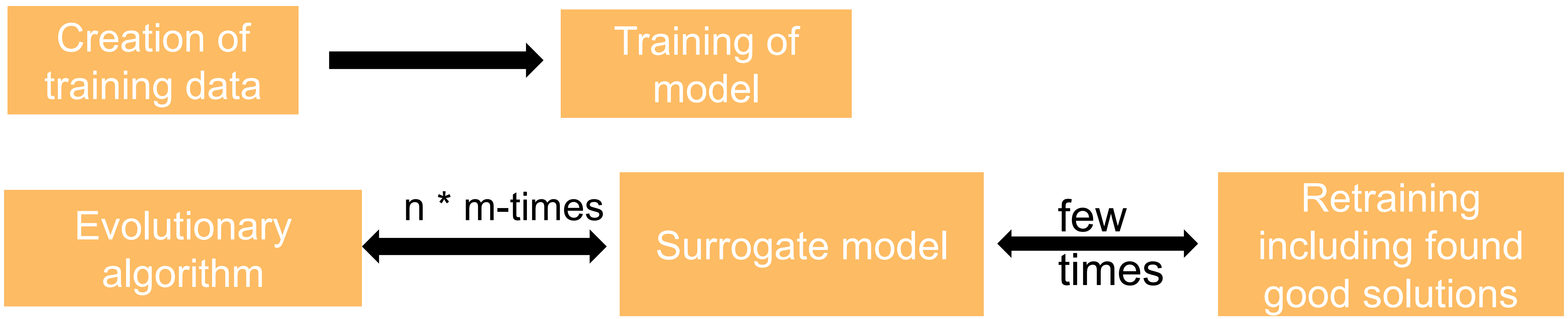
- Nature-inspired algorithms are smart parameter scans:
  - Genetic algorithms
  - Particle swarm algorithms
- Genetic algorithms allow multi-objective optimization
- Equally valid solution form a so-called Pareto front (PA front) [x]
- Search for solutions using techniques such as mutation, selection and crossover
- The fitness measures how good an individual is adapted

[x] A. Konak, Reliab. Eng. Syst. Saf.}, 91 (9), pp. 992--1007, 2006.

Optimizations with surrogate models

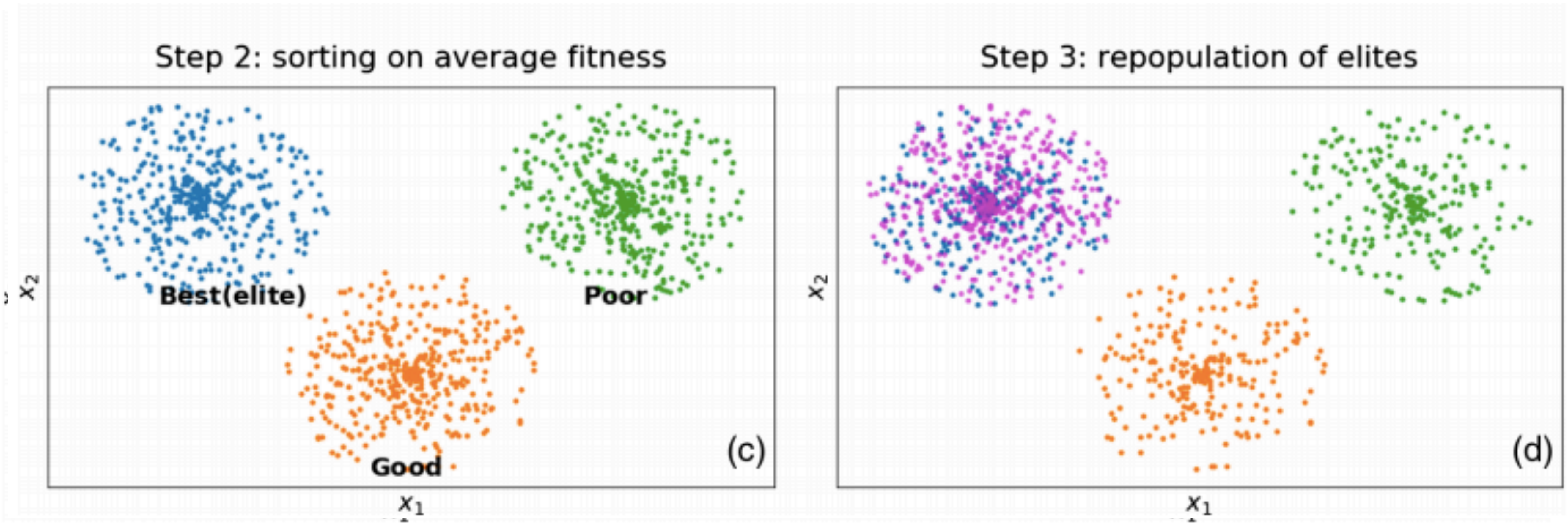
- To speed up: ANN surrogate models are used in the optimization
- For complex problems, retraining during optimization is necessary
- Benefit has been show early in [4,17, 18, d]

Implementation



GA Optimization and ML/AI

- [4] Y. Li, W. Cheng, L. H. Yu, and R. Rainer, Genetic algorithm enhanced by machine learning in dynamic aperture optimization, Phys. Rev. Accel. Beams 21, 054601 (2018)
  - DA optimization for National Synchrotron Light Source II (NSLS-II) Storage Ring, Brookhaven National Laboratory
  - Population is classified into different clusters and the clusters with top average fitness are given “elite” status.



GA Optimization and ML/AI

- [18] A. Edelen, N. Neveu, M. Frey, Y. Huber, C. Mayes, and A. Adelmann, Machine learning for orders of magnitude speedup in multiobjective optimization of particle accelerator systems, Phys. Rev. Accel. Beams 23, 044601 (2020).
  - Argonne Wakefield Accelerator (AWA) Facility
  - Speedup of 144
- [d] Jinyu Wan, Paul Chu, and Yi Jiao, Neural network-based multiobjective optimization algorithm for nonlinear beam dynamics, Phys. Rev. Accel. Beams **23**, 081601 (2020)
  - DA optimization for HEPS, ultralow-emittance storage ring light source being built in Beijing, China
  - The data produced with the standard MOGA are used to train a neural network

GA Optimization and ML/AI

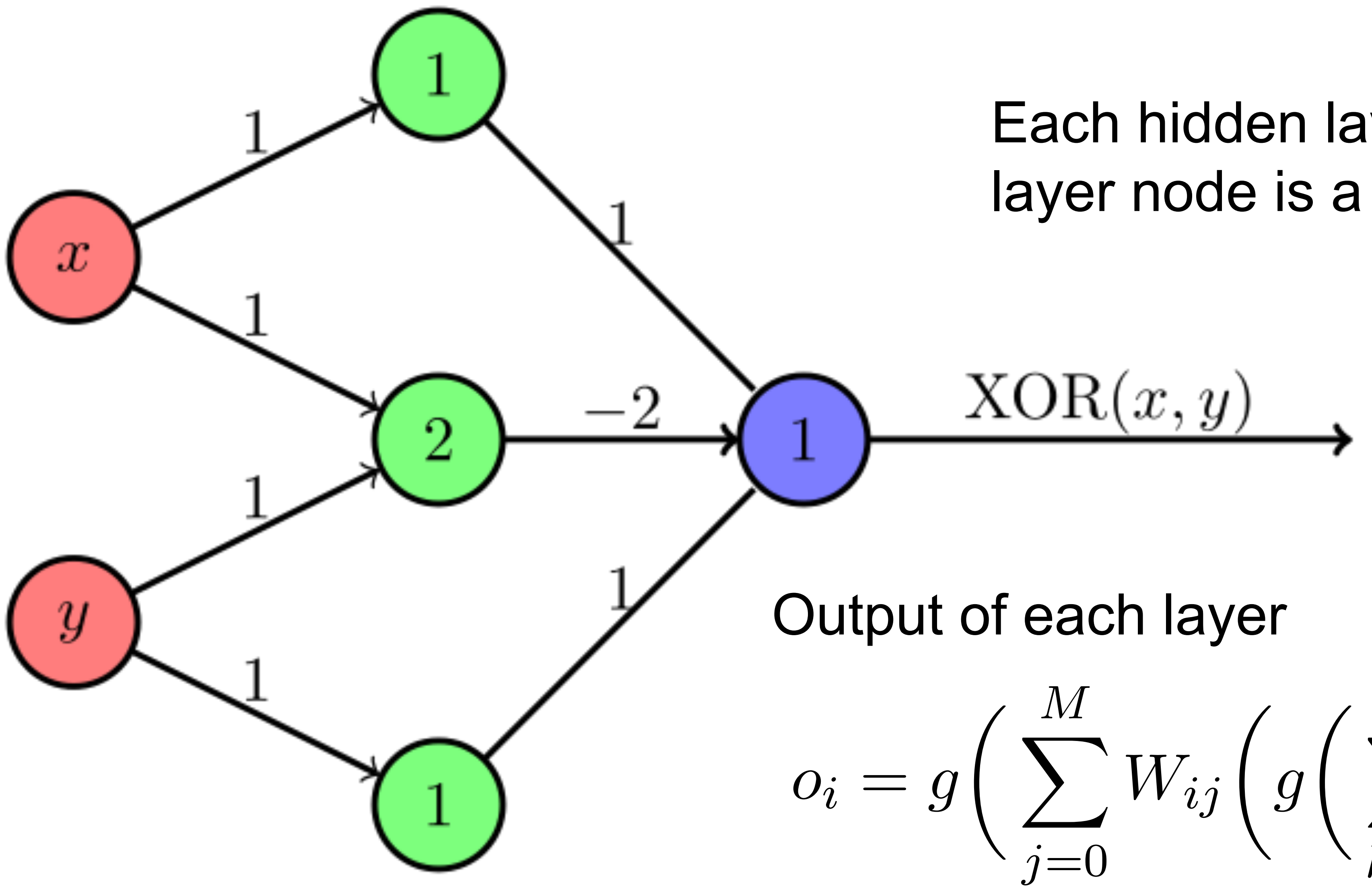
- [17] M. Song, X. Huang, L. Spentzouris, and Z. Zhang, Storage ring nonlinear dynamics optimization with multiobjective multigeneration Gaussian process optimizer, Nucl. Instrum. Methods Phys. Res., Sect. A 976, 164273 (2020) or arXiv:1907.00250.
  - A Gaussian process regression-based model was constructed and used to predict the objective values of a large pool of candidate individuals
  - DA optimization for SPEAR3 upgrade lattice

- Also possible to use: Bayesian Multiobjective Optimization
  - Novel, first paper publish last year [b]
  - Probabilistic model is construed and then exploits during optimization

[b] https://gpflowopt.readthedocs.io/en/latest/notebooks/multiobjective.html

Feed-forward Artificial Neural Network

MartinThoma, Wikipedia



Each hidden layer and output layer node is a perceptron.

Output of each layer

$$o_i = g\left(\sum_{j=0}^{M} W_{ij}\left(g\left(\sum_{k=0}^{N} x_k W_{jk} + b_j\right)\right) + b_i\right)$$

Adding “hidden” layer(s) allow non-linear target functions to be represented

The optimization problem

- Upgrade of the Swiss Light Source to increase the brilliance (smaller emittance)
  - with seven-bend achromats
- Stronger focusing requirements need higher sextupole and higher-order multi-pole fields for chromatic compensation
- More challenging to find large dynamic aperture
- Either be done
  - indirectly: by computing and minimizing dominant resonance driving terms
  - or directly: by computing and maximizing DA and energy acceptance
- The objective functions are defined to maximize the transverse DAs at three different energies and to prevent the tune resonances from being crossed, thus maximizing the energy acceptance and beam lifetime.

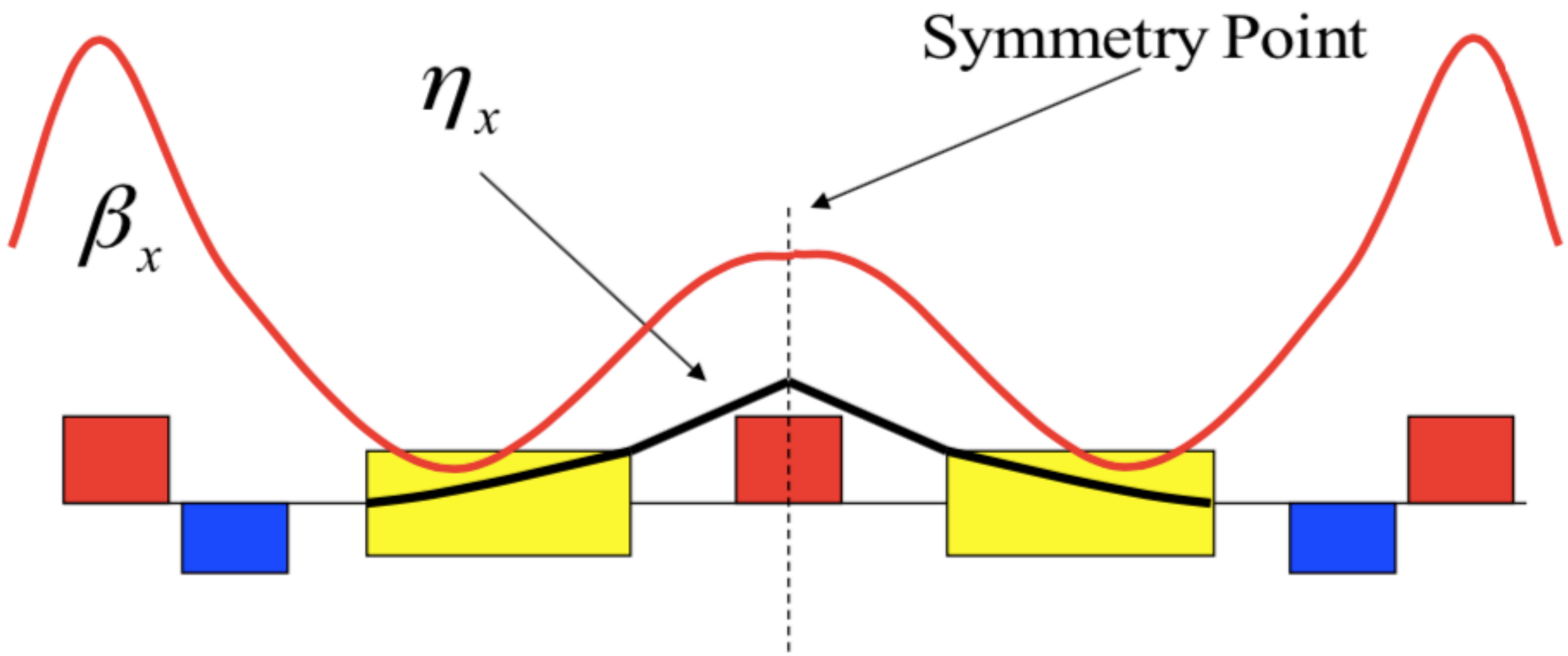
	SLS upgrade [20]
Circumference (m)	287.25
Emittance (pm)	137
Periodicity	3
Topology	12 × 7BA
Q_x	37.383
Q_y	10.280
Natural chromaticity χ_x	−64.9
Natural chromaticity χ_y	−34.5
Peak dispersion (cm)	4.9
# chromatic sextupole families	4
# harmonic sextupole families	9
# octupole families	10

Bend achromat lattice for small emittance

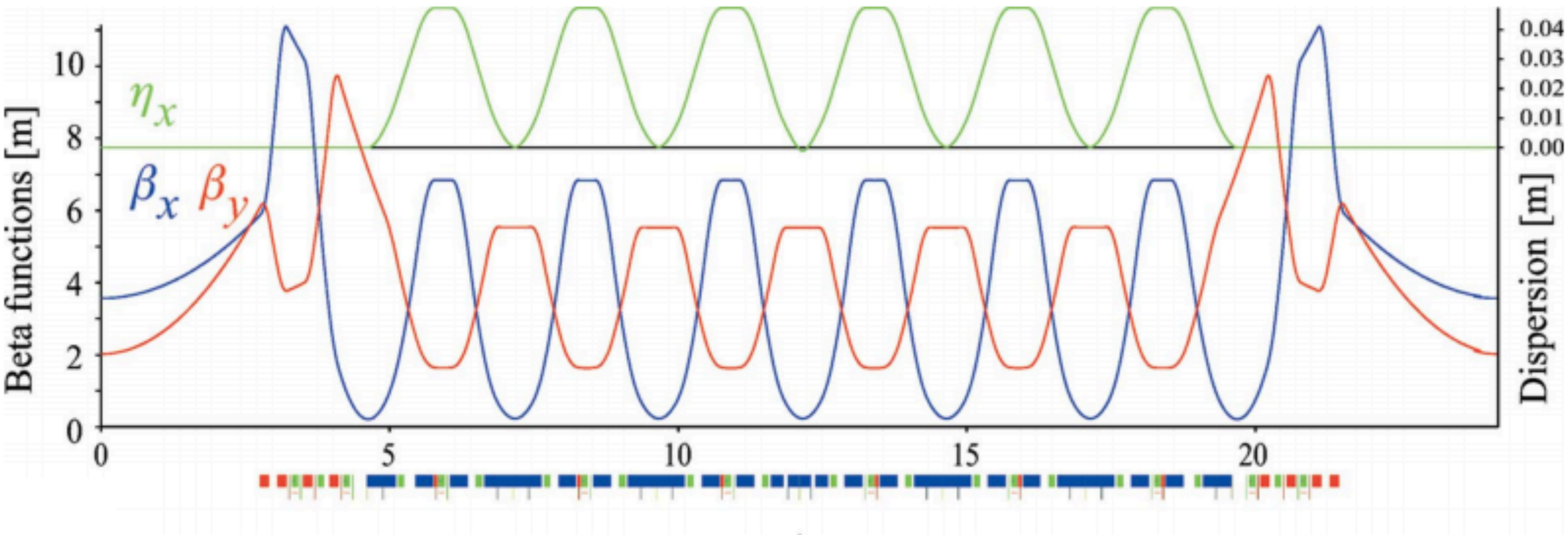
Double Bend Achromat for small emittance lattice

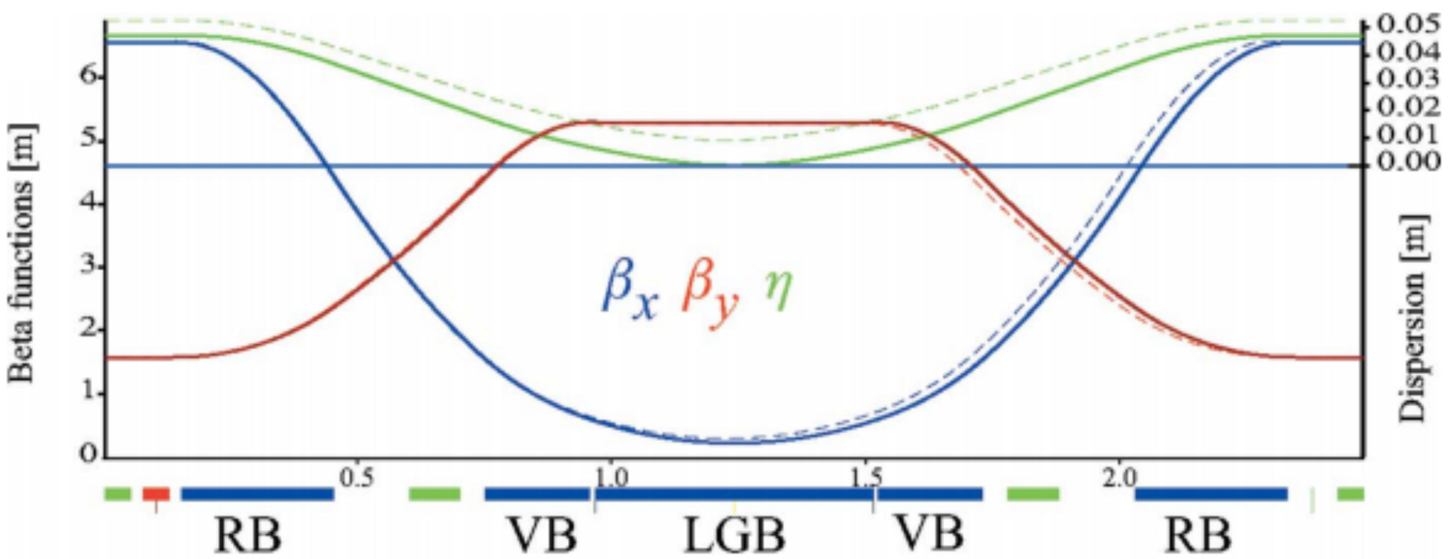
The quadrupoles are used to shape the β- and η-functions inside the dipoles, so as to meet the minimum emittance conditions.

DBA lattice can lead to very strong quadrupoles and large negative chromaticity



The Swiss Light Source Upgrade lattice





RB: reverse bends

VB: vertically focusing combined-function bending magnets

LGB: longitudinal-gradient bends

DA in damping rings

- The dynamic aperture is given as the limit of stable motion of particle amplitudes.
- Generally determined at zero dispersion.
- The tracking time is generally considered to be some fraction of the damping time.
- DA is determined by starting a particle at a large amplitude and reducing the amplitude until the stability condition is reached.
- Attention with is method DA can be wrongly overestimated:
  - Regions of stability separated by unstable regions
- DA is also tracked for off-energy to ensure that the transverse apertures are large enough for dispersive particles.
- Maximizing the momentum aperture is also important for damping rings
  - Can be included by maximizing the off-energy DA and additional constrains

Mathematical definition

- DA objective in Floquet space

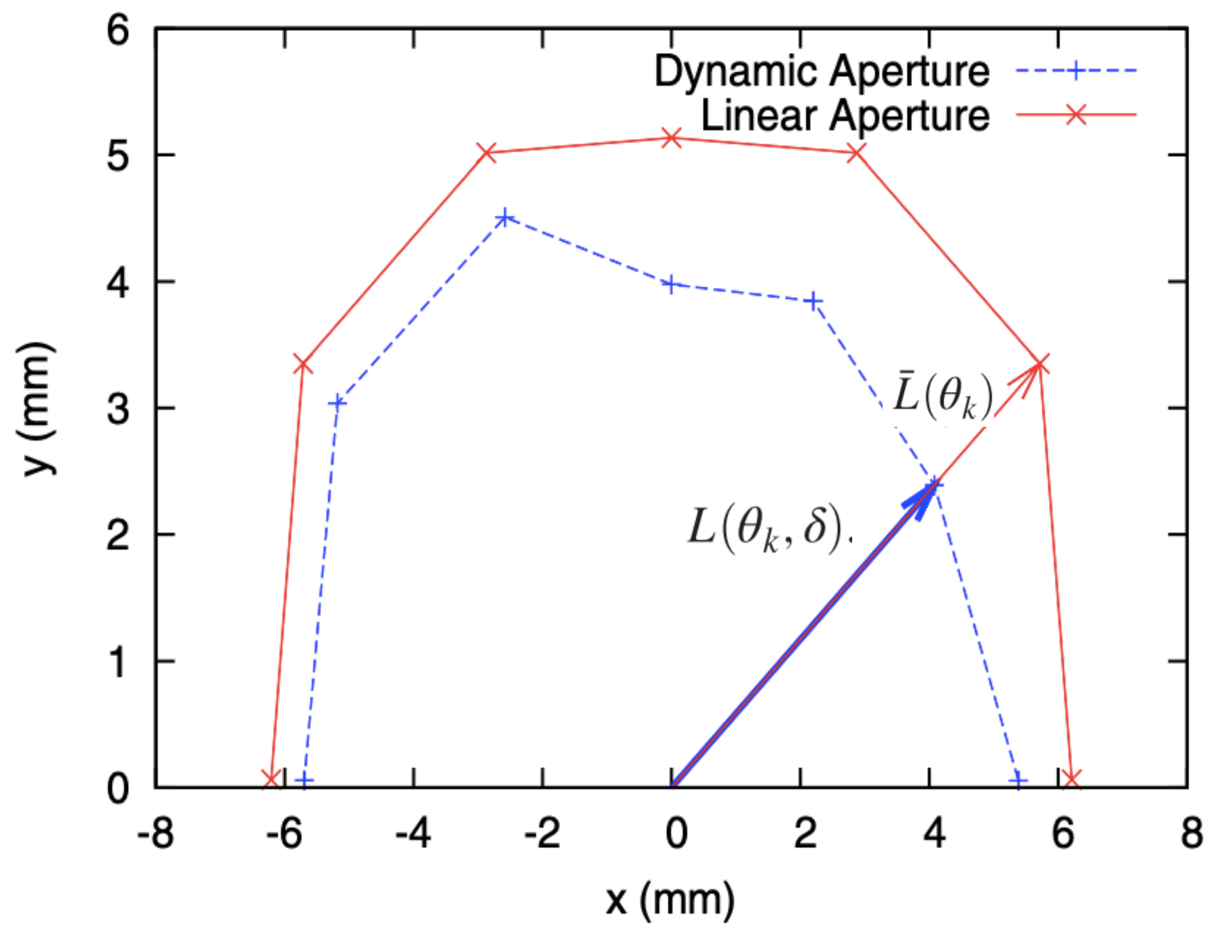
$$\mathrm{DA}_\delta = \frac{1}{2K}\left(f_{0,\delta}^2 + f_{K,\delta}^2 + 2\sum_{k=1}^{K-1} f_{k,\delta}^2\right)$$

- The linear aperture is the smallest aperture found by projecting the chamber from each lattice point to the injection point using a linearized particle momentum map.

- The maximization of the DA corresponds to the minimization of the DA objectives as DA [0,1]

- Optimized nonlinear machine should behave as linear machine

- Length computed using the biased binary search



$$f_{k,\delta} = \frac{\max\{0, \bar{L}(\theta_k) - L(\theta_k,\delta)\}}{\bar{L}(\theta_k)}$$

$\delta =$ -0.03,0,+0.03

Mathematical definition

- Constrain sextupole strength
  - More than two sextupole families available for correction chromaticity and adjust DA

Other families — Unaltered lattice

$$\vec{\xi} = \mathbf{M}\vec{\kappa} + \mathbf{T}\vec{t} + \vec{\xi}_{\mathrm{ua}},$$

Tuning sextupoles families

  - Limit by magnet design and possible chromaticity between [0,1]

To sum up, a design point in the search space is

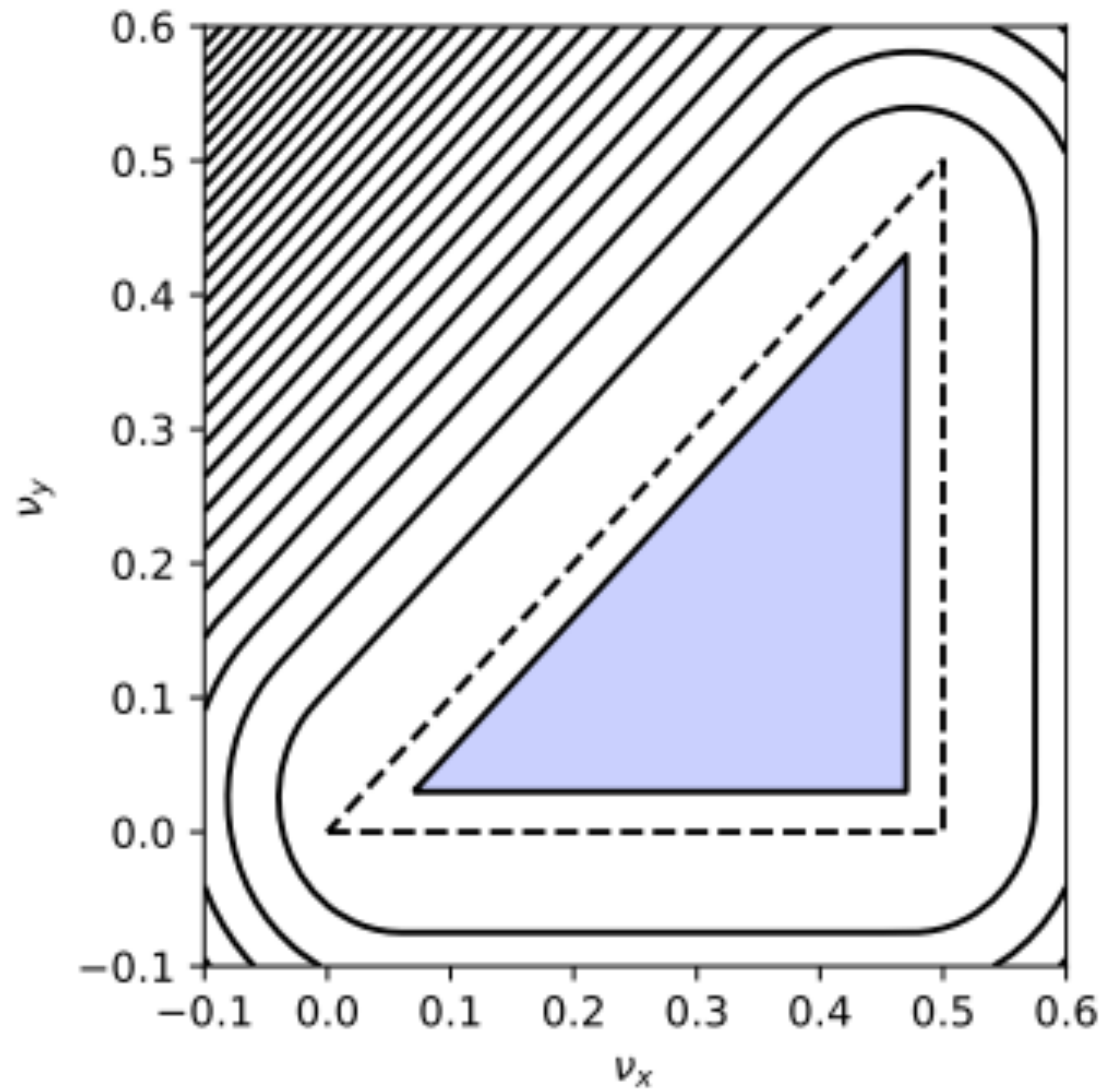
$$\vec{d} = (\xi_x, \xi_y, \kappa_1, \ldots, \kappa_5),$$

where

$$\xi_{x,y} \in [0, \xi_{\max}], \qquad \kappa_i \in [-\kappa_{\max}, \kappa_{\max}].$$

Mathematical definition

- Constrain tune variation
  - ➢ no low-order resonances are crossed
  - Tune footprint distance ctfp is defined, zero in blue area
  - Amplitude-dependent tune footprint distance adts is computed for several points, also zero in blue area



Simulation code is TRACY-3:

Self-Consistent charged particle beam dynamics model based on a Symplectic Integrator.
https://github.com/jbengtsson/tracy-3.5

MOGA optimization

Intel XeonGold 6152 nodes of the PSI Merlin cluster

Used: Three nodes with 132 processes for 829 gen with 300 inv for 48 h

Generation	100	200	300	400	500	829
nof pts better	1	10	17	18	26	31

Objective	$DA_{-\delta}$,	DA_0	DA_δ,	unstable		gen
Design solution	0.032	0.004	0.011	0	0	
point-1	**0.021**	0.003	0.010	0	0	763
point-2	0.031	**0.001**	**0.002**	0	0	769
point-3	0.025	0.001	0.005	0	0	807

ANN surrogate model for fast optimization

- 3×10^4 feasible points out of 7.5×10^4 random points for training (70%), validation (20%) and test set (10%).
- This took 9 h 6 min on five nodes (220 cores)
- Feed-forward ANN with N_{layers} hidden layers and ReLU activation function and mean squared error function and following hyperparmarters:

$$N_{\mathrm{layers}} = 5,\ N_{\mathrm{neurons}} = 64 \text{ and } N_{\mathrm{batch}} = 128$$

- MOGA with ANN model:
  - Speed up of 3.2
  - but the solution quality is not as good, no better solution found
  - ➢ Retraining ANN model during optimization

ANN surrogate model for fast optimization

- The authors showed that is sufficient:
- To retrain two times during optimization at M_{gen} = 50 and M_{gen} = 500
- And a lower data size can be used for the retrain
- More: only 10% are reevaluated with TRACY

	OPT-PILOT	SM (3×10^4)	SM + retrain (2×10^4)	SM + retrain (10^4)	SM + retrain (5000)
nof pts better	31	0	148	368	87
Run-time (reeval all)	48 h	14 h 48 min	12 h 15 min	8 h 31 min	6 h 33 min
Core hours (reeval all)	6336	2847	2325	1593	1210
Speedup (reeval all)	1.0	3.2	3.9	5.6	7.3
Run-time (reeval 10%)		11 h 21 min	8 h 52 min	5 h 5 min	3 h 10 min
Core hours (reeval 10%)		2089	1578	838	465
Speedup (reeval 10%)		4.2	5.4	9.4	15.1

Conclusions of the paper

- Multi-objective genetic algorithm is standard for dynamic aperture energy acceptance optimization for light source.
    - In the paper for the Swiss Light Source upgrade.
- An artificial neural network surrogate models are used for speed up the computation.
- The surrogate model must be retrained during the optimization.
- The faster method makes it possible to include octupole strengths, which could further improve the solution quality.
- Or more accurate and more expensive model can be used: includes nonlinear synchrotron oscillation.