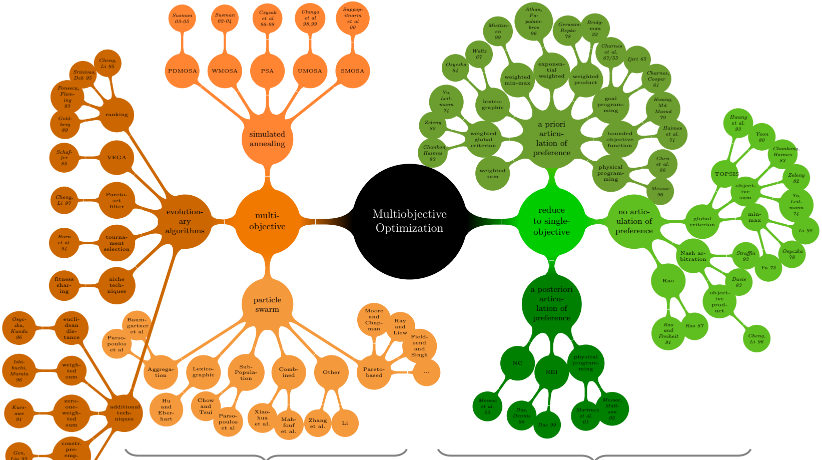
Genetic Multiobjective Optimisation Techniques

A. Adelmann (PSI)

November 13, 2018



- **multiple** Pareto solutions can be found in **one** run
- a posteriori articulation of preference
- "easier": diversity in decision and objective space (non-linear mapping)

- only **one** Pareto solution can be found in **one** run
- preference-based (specify preference for trade-off solution)
- not all can be found in non-convex MOOPS
- all algorithms require a prior knowledge (weights, ε, targets)

PAUL SCHERRER INSTITUT PSI

History

[O.L. De Weck]

Rational people attempt to make the best decision within a specified set of possible alternatives.

- Multiobjective thinking originated in economics: the best referred to decisions taken by buyers and sellers (micro-economics) or governments (macro-economics), which simultaneously optimise or balance several criteria.
- Taxation: an optimal, average level of tax collected (% per $ of economic activity) maximizes the revenue available for the common good, while maintaining a sufficient incentive for individuals to earn income from their own work.

Francis Y. Edgeworth (1845-1926), King's College & Oxford

PAUL SCHERRER INSTITUT PSI

History cont.

Pareto on the other hand was a contemporary of Edgeworth, born in Paris in 1848, graduated from the University of Turin in 1870 (Civil Engineering) with a thesis: *The Fundamental Principles of Equilibrium in Solid Bodies*

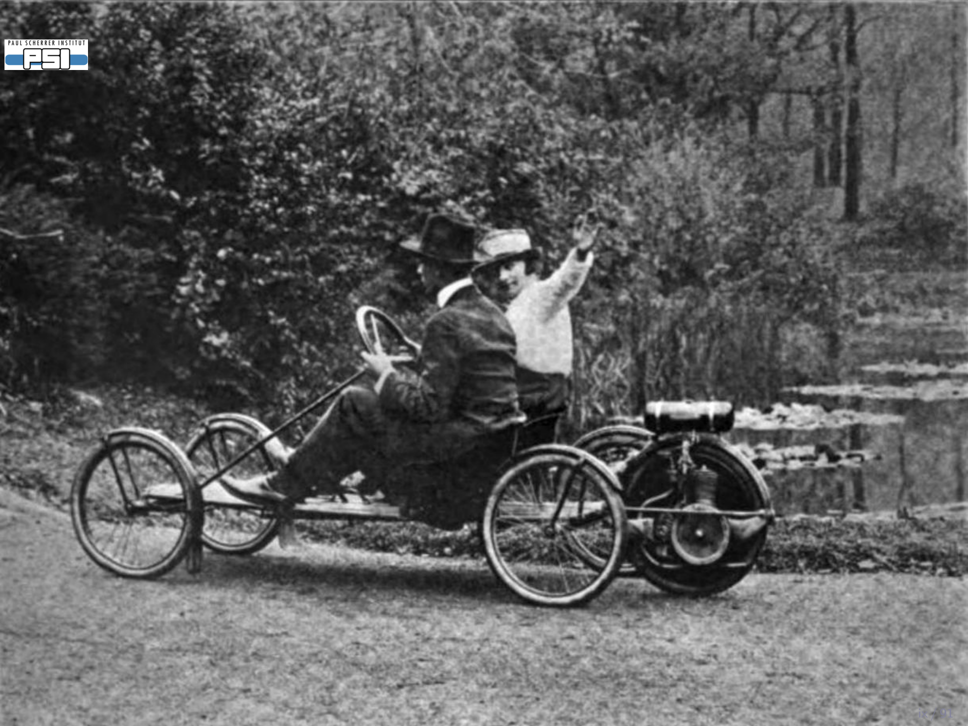
- Pareto took up the study of philosophy and politics and was one of the first to analyse economic problems with mathematical tools
- In 1893, Pareto became the Chair of Political Economy at the University of Lausanne, where he created his two most famous theories:
    1. Circulation of the Elites
    2. The Pareto Optimum

Vilfredo Pareto (1848-1923)

History cont.

- The translation of Pareto's work into English in 1971 spurred the development of multiobjective methods in Applied Mathematics and Engineering.
- The growth of this field manifested itself particularly strongly in the United States with pioneering contributions by (Stadler 1979), (Steuer 1985) among many others.
- Theoretical aspects of multiobjective optimisation can be found in Japan (Sawaragi, Nakayama and Tanino, 1985).
- Over the last three decades the applications of multiobjective optimisation have grown steadily in many areas of Engineering and Design including the Particle Accelerator Community
- A particularly remarkable resource in this area is the website `http://delta.cs.cinvestav.mx/~ccoello/EMOO/` created and maintained by C.A. Coello.

PAUL SCHERRER INSTITUT
PSI

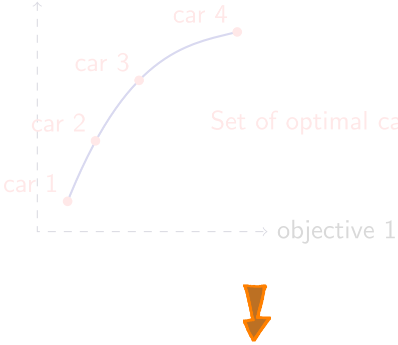


PAUL SCHERRER INSTITUT PSI

Buying a Car

Conflicting criteria → Trade-offs

Conflicting criteria → Trade-offs



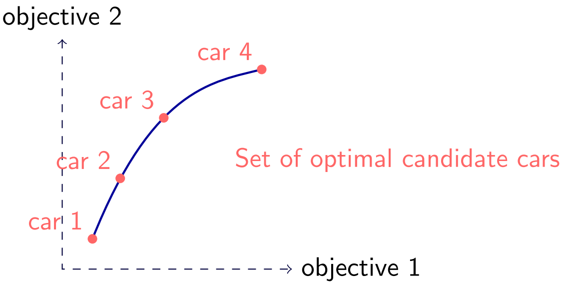
Subjective decision using
higher level information

PAUL SCHERRER INSTITUT PSI

Buying a Car

Conflicting criteria → Trade-offs

Conflicting criteria → Trade-offs



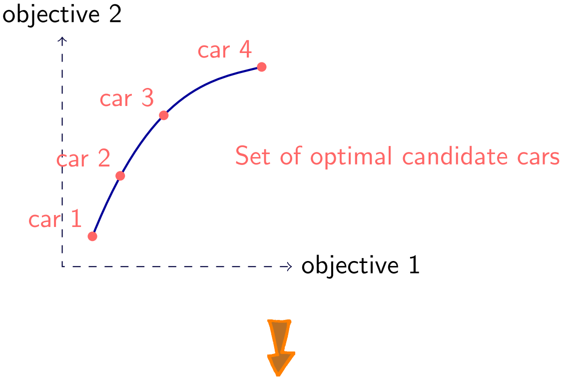
Subjective decision using

higher level information

Buying a Car

Conflicting criteria → Trade-offs

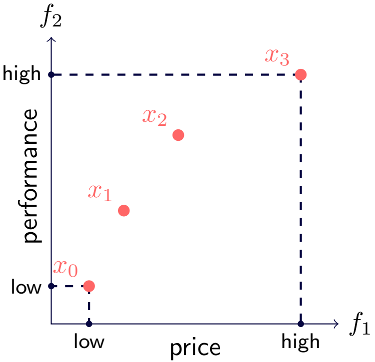
Conflicting criteria → Trade-offs



Subjective decision using

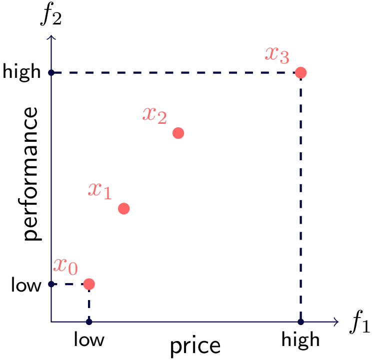
higher level information

Optimality?



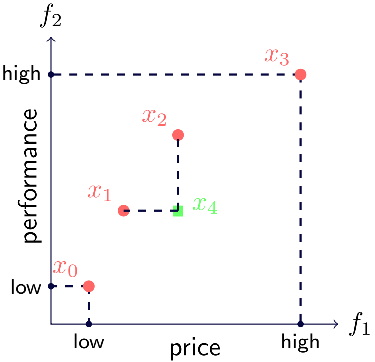
- conflicting objectives:
  minimize price
  maximize performance
- red points are "equally optimal":
  cannot improve one point without hurting at least one other solution
  → Pareto optimality
- x_4 is dominated by x_1 and x_2
- Pareto front

Optimality?



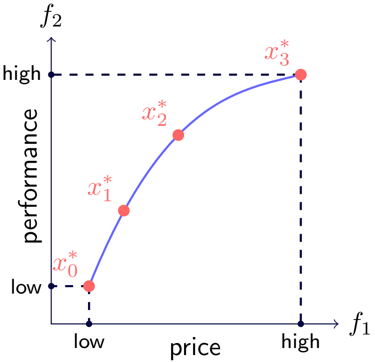
- conflicting objectives:
  minimize price
  maximize performance
- red points are "equally optimal": cannot improve one point without hurting at least one other solution
  $\rightarrow$ Pareto optimality
- x_4 is dominated by x_1 and x_2
- Pareto front

Optimality?



- conflicting objectives:
  minimize price
  maximize performance
- red points are "equally optimal": cannot improve one point without hurting at least one other solution → Pareto optimality
- x_4 is dominated by x_1 and x_2
- Pareto front

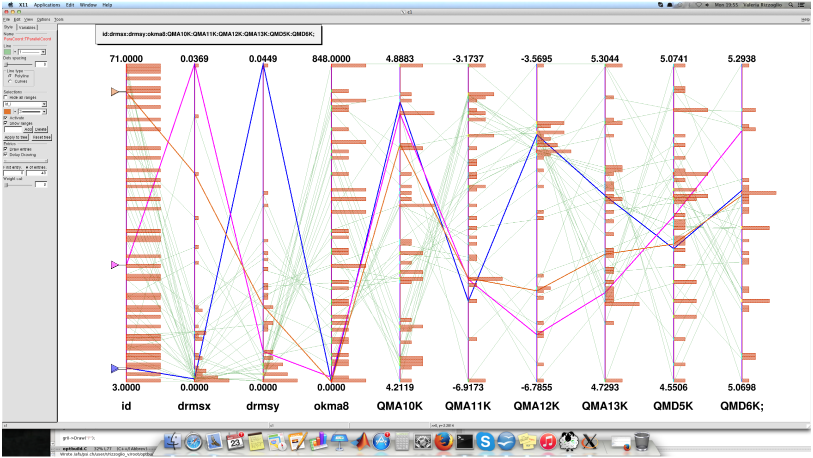
Optimality?



- conflicting objectives:
  minimize price
  maximize performance
- red points are "equally optimal":
  cannot improve one point without hurting at least one other solution
  $\rightarrow$ Pareto optimality
- x_4 is dominated by x_1 and x_2
- Pareto front

High Dimensional Data

Root (CERN)



PAUL SCHERRER INSTITUT PSI

Formulation of the Multiobjective Optimisation Problem

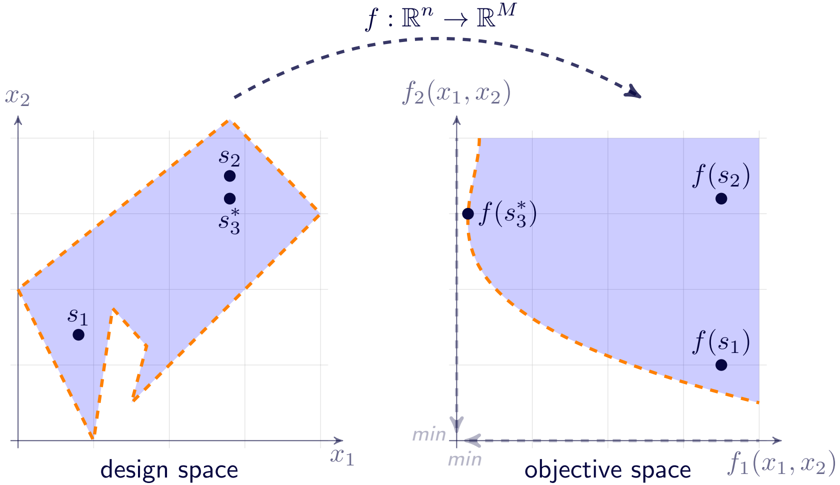
Denoting the feasible domain by $\mathbf{S} \in \mathbb{R}^n$, the problem is to minimise – **simultaneously** – all elements of the objective vector,

Objectives → Design variables

$$
\begin{array}{lll}
\text{min} & f_m(\mathbf{x}) \in \mathbb{R} \text{ and } \mathbf{x} \in \mathbf{S}, & m = 1 \ldots M \\
\text{s.t.} & g_j(\mathbf{x}) \geq 0, & j = 0 \ldots J \\
 & h_k(\mathbf{x}) = 0, & k = 0 \ldots K \\
 & x_i^L \leq \mathbf{x} = x_i \leq x_i^U. & i = 0 \ldots n
\end{array}
$$

Constraints

Formulation of the Multiobjective Optimisation Problem



The (non-linear) mapping $f : \mathbb{R}^n \to \mathbb{R}^M$ from design to objective space.

Formulation of the Multiobjective Optimisation Problem

Multiobjective optimisation methods can be broadly decomposed into two categories

1. Scalarisation approaches: the multiobjective problem is solved by translating it back to a single (or a series of) objective, scalar problems. This requires the formation of an overarching objective function which contains contributions from the sub-objectives in vector J.
2. Pareto approaches

Scalarization I

Weighted Sum Approach

Scalarization methods are based on the assumptions that

1. designer or decision-maker preferences are known before design solutions are found and that
2. the M objectives can be meaningfully combined to express a utility, U, dimensionless scalar quantity expressing the goodness of a particular design.

$$
\begin{array}{lll}
\text{min} & U\{f_m(\mathbf{x})\} \in \mathbb{R} \text{ and } \mathbf{x} \in S, & m = 1 \dots M \\
\text{where} & U = \sum_{q=1}^{M} w_q f_q(\mathbf{x}), \text{with } w_q > 0 \text{ and } \sum_{q=1}^{M} w_q = 1 & \\
\text{s.t.} & g_j(\mathbf{x}) \geq 0, & j = 0 \dots J \\
 & h_k(\mathbf{x}) = 0, & k = 0 \dots K \\
 & x_i^L \leq \mathbf{x} = x_i \leq x_i^U. & i = 0 \dots n
\end{array}
$$

Scalarization II

Weighted Sum Approach

- Formulated in this way the aggregate objective U always forms a strictly convex combination of objectives
- One of the issues in this method is the appropriate choice of λ
- In the case of two equally scaled objectives we get

$$U = \lambda J_1 + (1 - \lambda) J_2. \tag{1}$$

Finding optima for U as λ is changed gradually, in equal intervals, from $0 \ldots 1$ reveals a set of optimal solutions as the weight is gradually shifted from one objective to another.

Formulation of the Pareto Optimal Condition

A point $\mathbf{x}_1$ is dominating $\mathbf{x}_2$

1. the solution $\mathbf{x}_1$ is no worse than $\mathbf{x}_2$ in all objectives
2. the solution $\mathbf{x}_1$ is strytictly better than $\mathbf{x}_2$ in at least one objective.

$$\mathbf{x}_1 \preceq \mathbf{x}_2 \text{ iff } \begin{cases} f_m(\mathbf{x}_1) \geq f_m(\mathbf{x}_2), \ \forall m \in 1 \dots M \\ f_j(\mathbf{x}_1) > f_j(\mathbf{x}_2), \ \exists j \in 1 \dots M \end{cases}$$

The properties of the dominance relation include transitivity

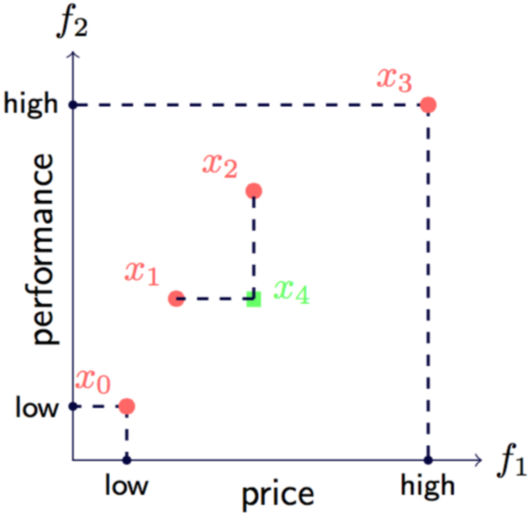
$$x_1 \preceq x_2 \wedge x_2 \preceq x_3 \Rightarrow x_1 \preceq x_3,$$

and asymmetry, which is necessary for an unambiguous order relation

$$x_1 \preceq x_2 \Rightarrow x_2 \npreceq x_1.$$

Using the concept of dominance, the sought-after set of Pareto optimal solution points can be approximated iteratively as the set of non-dominated solutions.

Formulation of the Pareto Optimal Condition



Remarks on Pareto Optimality I

- Deciding if a point truly belongs to the set of Pareto optimal solutions is NP-hard[1] however many efficient heuristics exists.
- A comprehensive or full-factorial evaluation of the design space is often impossible due to the n-dimensionality of the design vector, $\mathbf{x}$, and the required computational effort for obtaining f, g and h.

> Solutions obtained are mere approximations of the Pareto Front.

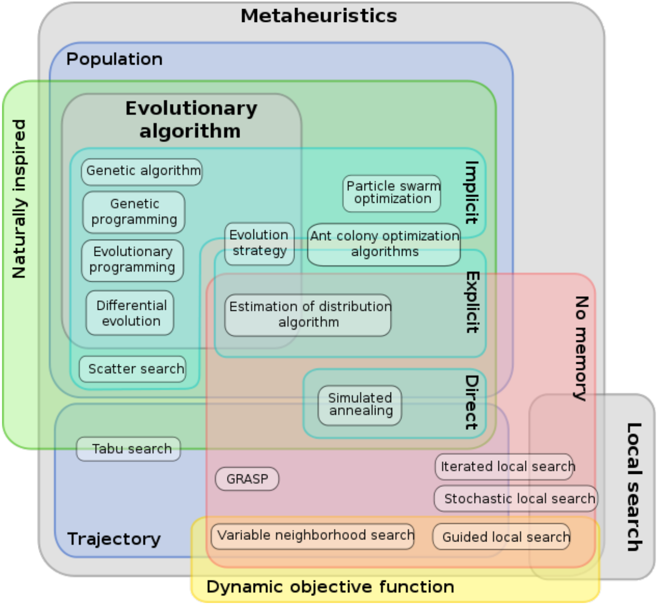
Among the Pareto approaches two in particular have gained increased acceptance and use in recent years:

1. Multiobjective Genetic Algorithms
2. Multiobjective Swarm Optimisation Algorithms

[1] A problem is NP-hard if an algorithm for solving it can be translated into one for solving any NP-problem (nondeterministic polynomial time) problem. NP-hard therefore means "at least as hard as any NP-problem," although it might, in fact, be harder

Genetic Algoritms an Overview I

- A genetic algorithm (GA) is a metaheuristic inspired by the process of natural selection that belongs to the larger class of evolutionary algorithms (EA)
- Directed search algorithms based on the mechanics of biological evolution Developed by John Holland, University of Michigan (1970's)
    - Holland, J.H., "Adaptation in Natural and Artificial Systems", MIT Press, 1975.
- To understand the adaptive processes of natural systems
- To design artificial systems software that retains the robustness of natural systems
- Provide efficient, effective techniques for optimization and machine learning applications
- Widely-used today in business, science and engineering


Metaheuristics
Population
Evolutionary algorithm
Naturally inspired
Genetic algorithm
Genetic programming
Evolutionary programming
Differential evolution
Scatter search
Evolution strategy
Particle swarm optimization
Ant colony optimization algorithms
Implicit
Explicit
Estimation of distribution algorithm
Direct
Simulated annealing
No memory
Local search
Tabu search
GRASP
Iterated local search
Stochastic local search
Variable neighborhood search
Guided local search
Trajectory
Dynamic objective function

PAUL SCHERRER INSTITUT PSI

Genetic Algoritms an Overview cont. I

- GA can be viewed as a general-purpose search method, an optimization method, or a learning mechanism, based loosely on Darwinian principles of biological evolution reproduction and the **survival of the fittest**
- GA maintains a set of candidate solutions called population and repeatedly modifies them
- At each step, the GA selects individuals from the current population to be parents and uses them to produce the children for the next generation
- In general, the fittest individuals of any population tend to reproduce and survive to the next generation with the goal to improve successive generations
- However, inferior individuals can, by chance, survive and also reproduce

Genetic Algoritms an Overview cont. II

- GA is well suited to and has been extensively applied to solve complex design optimization problems because
  - it can handle both discrete and continuous variables
  - on-linear objective and constrain functions
  - no gradient information needed

Evolutionary Algorithms I

- Evolutionary algorithms (EA) are loosely based on nature's evolutionary principles to guide a population of individuals towards an improved solution by honoring the "survival of the fittest" practice.
- This "simulated" evolutionary process preserves entropy (or diversity in biological terms) by applying genetic operators, such as mutation and crossover, to remix the fittest individuals in a population.

A generic evolutionary algorithms consists of the following components:

- *Genes*: traits defining an individual (design variables)
- *Fitness*: a mapping from genes to a set of numeric values (evaluating each objective function) describing the fitness of an individual,
- *Selector*: selecting the k fittest individuals of a population based on some sort of ordering,
- *Variator*: recombination (mutations and crossover) operators for offspring generation.

Evolutionary Algorithms I

Non-dominated sorting

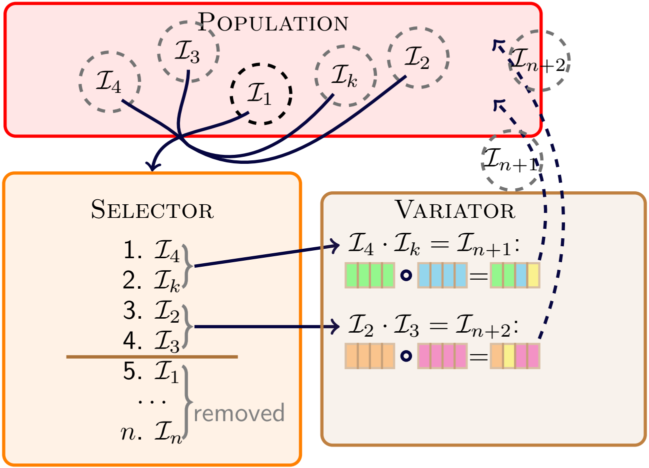
Algorithm: $\forall$ generations

1. initially random population of individuals I_i with a unique set of genes and corresponding fitness
2. In a next step the population is processed by the SELECTOR determining the k fittest individuals.
3. While the k fittest individuals are passed to the VARIATOR, the remaining $n-k$ individuals are eliminated from the population.
4. The VARIATOR mates the k fittest individuals to generate new offspring and applies the recombination operators.
    - Check convergence
5. After evaluating the fitness of all the freshly born individuals a *generation* cycle has completed

Complexity upper bound: $\mathcal{O}(GMN\log N)$ with M number of genes, N the population size and G the number of generations.

Evolutionary Algorithms

A Platform and Programming Language Independent Interface for Search Algorithms [2]



[2] NSGA-II: `http://www.tik.ee.ethz.ch/pisa/`

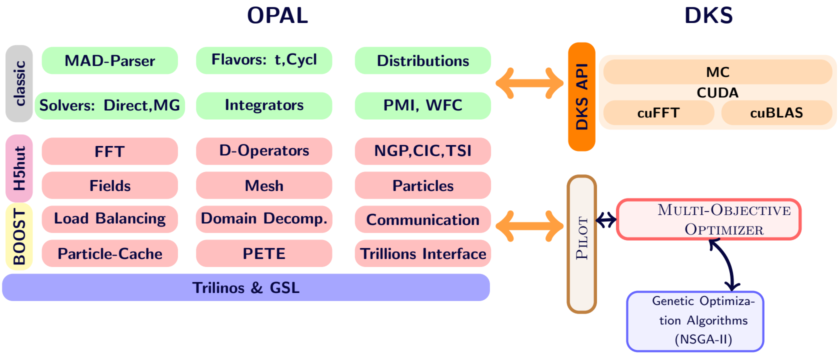
OPAL in a Nutshell I

OPAL is an open-source tool for charged-particle optics in large accelerator structures and beam lines including 3D space charge, particle matter interaction, partial GPU support and multi-objective optimisation.

- OPAL is built from the ground up as a parallel application exemplifying the fact that HPC (High Performance Computing) is the third leg of science, complementing theory and the experiment
- OPAL runs on your laptop as well as on the largest HPC clusters
- OPAL uses the MAD language with extensions
- OPAL is written in C++, uses design patterns, easy to extend
- Webpage: `https://gitlab.psi.ch/OPAL/src/wikis/home`
- the OPAL Discussion Forum: `https://lists.web.psi.ch/mailman/listinfo/opal`
- $\mathcal{O}(40)$ users

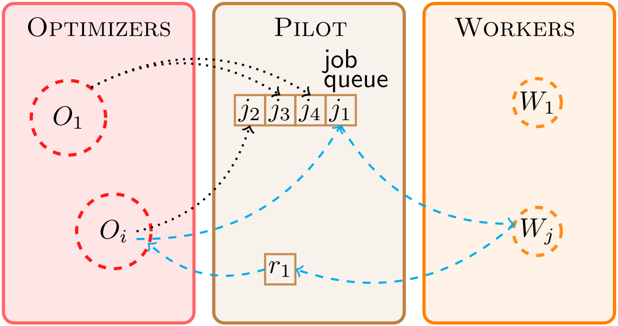
Software Architecture

MPI based + HW accelerators + Optimiser



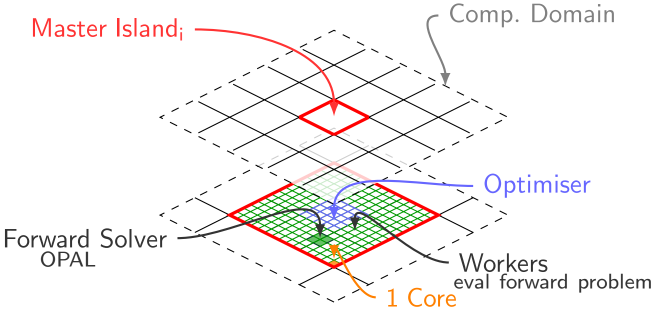
[Y. Ineichen et al., CS-R&D (2012), Y. Ineichen et al., arXiv:1302.2889]

Master/Worker Model



- Asynchronous finite state machine (MPI)
- Multi-Scale optimisation

Island-based Master Model



- using techniques from social network theory
- can solve very challenging problems using largest HPC resources
- PRACE [3] award 2012

[3] Partnership for Advanced Computing in Europe

Solution Exchange

- Introduces additional synchronization points
- Large sets of solutions have to be sent across the network
- This severly limits scalability

➡ Avoiding global synchronization: One-sided communication

- Using put/collect operations (MPI "shared variables")
- Solution set revision information to prevent unnecessary collects

➡ Local solution exchange on "special" graphs

- Implementation of "communication graph" exposes a set of neighbors
- "Route" messages between masters on imposed neighboring network

Solution Exchange

- Introduces additional synchronization points
- Large sets of solutions have to be sent across the network
- This severly limits scalability

Avoiding global synchronization: One-sided communication

- Using put/collect operations (MPI "shared variables")
- Solution set revision information to prevent unnecessary collects

Local solution exchange on "special" graphs

- Implementation of "communication graph" exposes a set of neighbors
- "Route" messages between masters on imposed neighboring network

Solution Exchange

- Introduces additional synchronization points
- Large sets of solutions have to be sent across the network
- This severly limits scalability

Avoiding global synchronization: One-sided communication

- Using put/collect operations (MPI "shared variables")
- Solution set revision information to prevent unnecessary collects

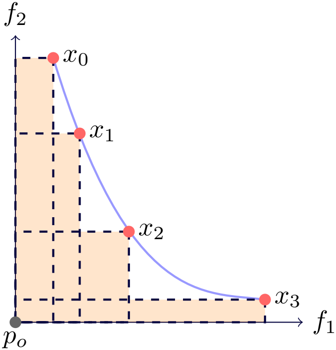
Local solution exchange on "special" graphs

- Implementation of "communication graph" exposes a set of neighbors
- "Route" messages between masters on imposed neighboring network

The FON Problem I

$$
\begin{aligned}
\text{min} \quad & \left[1-\exp\left(-1\left(\left(x_1-\frac{1}{\sqrt{3}}\right)^2+\left(x_2-\frac{1}{\sqrt{3}}\right)^2+\left(x_3-\frac{1}{\sqrt{3}}\right)^2\right)\right),\right. \qquad (2)\\
& \left. 1-\exp\left(-1\left(\left(x_1+\frac{1}{\sqrt{3}}\right)^2+\left(x_2+\frac{1}{\sqrt{3}}\right)^2+\left(x_3+\frac{1}{\sqrt{3}}\right)^2\right)\right)\right]^T\\
\text{s.t.} \quad & -1 \le x_i \le 1, \quad i = 1,2,3.
\end{aligned}
$$

The FON Problem II

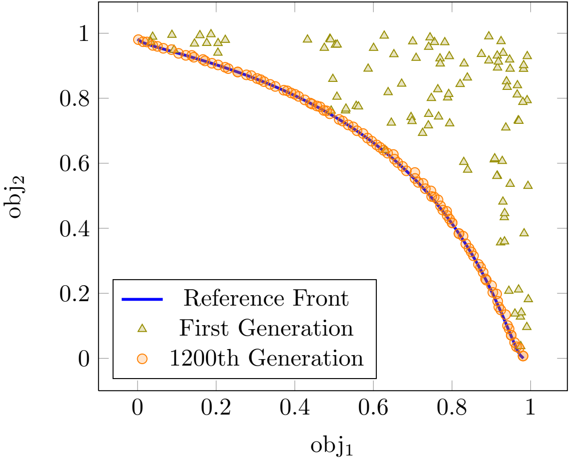


The hypervolume for a two-objective optimization problem corresponds to the shaded area formed by the dashed rectangles spanned by all points on the Pareto front and an arbitrary selected origin p_o.

The FON Problem III

- To that end, we use a metric for comparing the quality of a Paret front.
- Given a point in the Pareto set, we compute the m dimensional volume (for m objectives) of the dominated space, relative a chosen origin.

The FON Problem IV



The FON Problem V

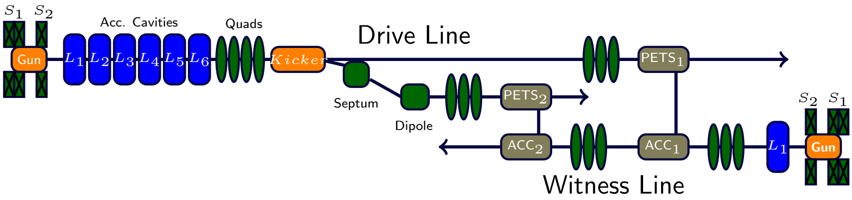
Variator benchmark after 1100 function evaluations using binary crossover and independent gene mutations (each gene mutates with probability $p = \frac{1}{2}$) on a population of 100 individuals.

Table: Convergence of benchmark problem with errors relative to hypervolume of sampled reference solution.

tot. function evaluations	hyper volume	relative error
100	0.859753	3.076×10^{-1}
200	0.784943	1.938×10^{-1}
500	0.685183	4.210×10^{-2}
900	0.661898	6.689×10^{-3}
1100	0.657615	1.749×10^{-4}

Full Staging

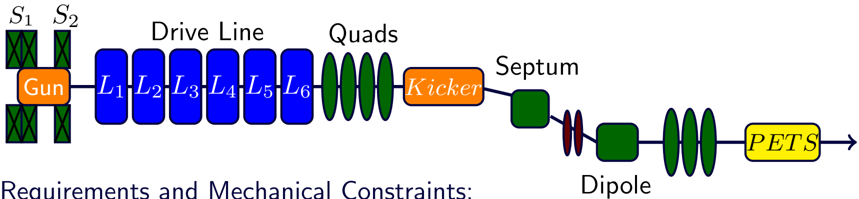
curtesy of Dr. Neveu



- Maintain modular design
- Maximize power in each stage
- Plug and play various structures

PETS: Power Extraction and Transfer Structures

TBA Beam Line Under Design



Requirements and Mechanical Constraints:

- 100% transmission, i.e. reasonable beam size at structure
- Reasonable bunch length at structure (maximize power)
- 1m between kicker and septum
  - for separation $\geq$ 50mm in septum.
- 1.8m between septum and dipole
  - for separation $\geq$ 0.5m of beam lines.
- 15cm between quads for easy installation.
- 0.3m between quads and PETS for yag screen.

GA applied to TBA Beam Line

Variable	Range	Unit
Buck Focusing Solenoid Strength	$300 \leq S_1 \leq 500$	amps
Matching Solenoid Strength	$180 \leq S_2 \leq 280$	amps
Quadrupole Strength	$-8.0 \leq K_i$[4] ≤ 8.0	T/m

Simulation Inputs:

- 6 design variables
- Laser radius is 9 cm
- Laser FWHM 10 ps
- All cavities at -20°

Objectives:

- Transverse beam size, $\sigma_{x,y}$
- Transverse momentum, $\sigma_{px,py}$
- Bunch length, σ_z
- Energy spread, dE

[4] $K_i = [K_1, \ldots, K_4]$

Sketch of an OPAL Inputfile (only optimiser cmd's) I

TBA Beamline

```
dv0: DVAR, VARIABLE=S1 ,   LOWERBOUND=300..
dv1: DVAR, VARIABLE=S2 ,   LOWERBOUND=180 ..
dv2: DVAR, VARIABLE=K1 ,   LOWERBOUND=-8 ..
dv3: DVAR, VARIABLE=K2,   LOWERBOUND=-8 ..
dv4: DVAR, VARIABLE=K3,   LOWERBOUND=-8 ..
dv5: DVAR, VARIABLE=K4,   LOWERBOUND=-8 ..

rmsx: OBJECTIVE,EXPR=statVariableAt(rms_x ,3.1);
rmsy: OBJECTIVE,EXPR=statVariableAt(rms_y,3.1);
rmspx: OBJECTIVE,EXPR=statVariableAt(rms_px,3.1);
rmspy: OBJECTIVE,EXPR=statVariableAt(rms_py,3.1);

rmss: OBJECTIVE,EXPR=statVariableAt(rms_s,3.1);
de: OBJECTIVE,EXPR=fabs(statVariableAt(dE,3.1));
```

Sketch of an OPAL Inputfile (only optimiser cmd's) II

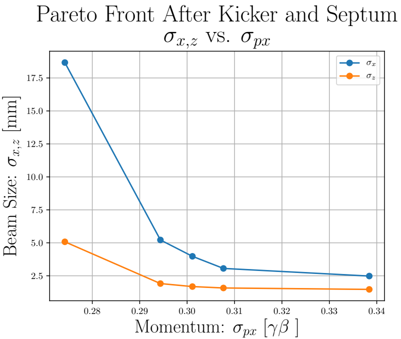
TBA Beamline

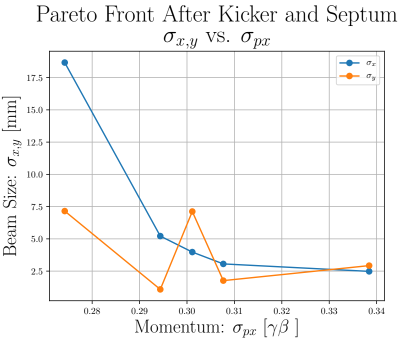
```
OPTIMIZE, INPUT="tmpl/ga-model.tmpl",
  OUTPUT="ga-model", OUTDIR="results",
  OBJECTIVES = {rmsx,rmsy,rmspx,rmspy,rmss,de},
  DVARS = {dv0,dv1,dv2,dv3,dv4,dv5,dv6},
  INITIALPOPULATION=656,
  MAXGENERATIONS=100,
  NUM_MASTERS=1,
  NUM_COWORKERS=8,
   ...
  NUM_IND_GEN=328,
  GENE_MUTATION_PROBABILITY=0.8,
  MUTATION_PROBABILITY=0.8,
  RECOMBINATION_PROBABILITY=0.2;
```

https://gitlab.psi.ch/OPAL/Manual-2.0/wikis/optimiser

PAUL SCHERRER INSTITUT PSI

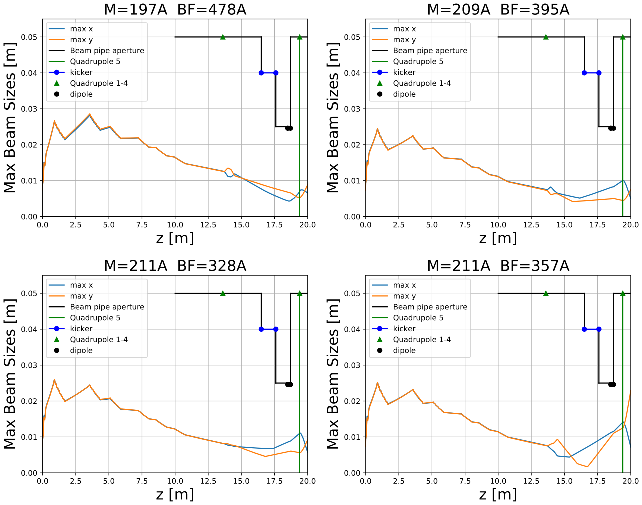
TBA Pareto Fronts





- Looking at entrance of 5th quad
- Location between septum and dipole
- Optimizing here will reduce beam size growth in dipole

Beam Size Results in Optimized Solutions

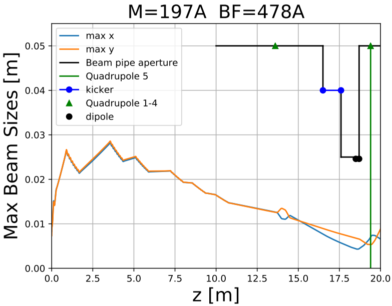


Best Solution

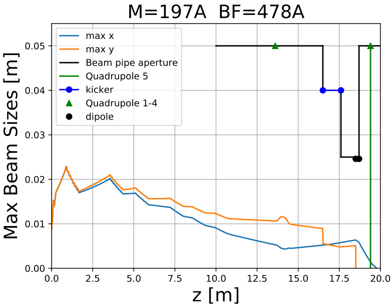
- Symmetric beam not necessary, if transmission is good.
- PETS aperture = 17.6 mm
- Need to adjust matching and quads.
- Energy ≈ 65 MeV

Quad	Value	Unit
Q1	-0.8	amps
Q2	0.9	amps
Q3	0.8	amps
Q4	-1.0	amps

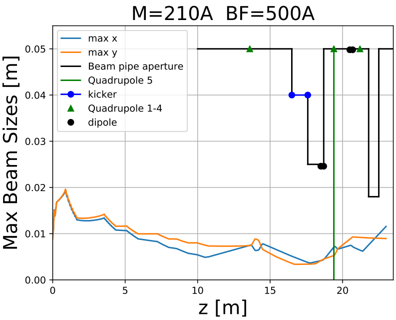
2D Field Maps only



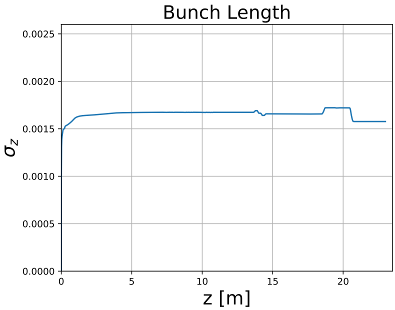
3D Maps and CSR included



Adjusted 3D/CSR Solution



Quad	Value	Unit
Q1	-1.5	amps
Q2	1.6	amps
Q3	1.5	amps
Q4	-1.7	amps
Q5	-2.0	amps
Q6	1.25	amps



- Strengthened all quads by 0.7 A
- No quad strengths are near limits!

Disturbed Isochronicity in Cyclotrons

M. Frey

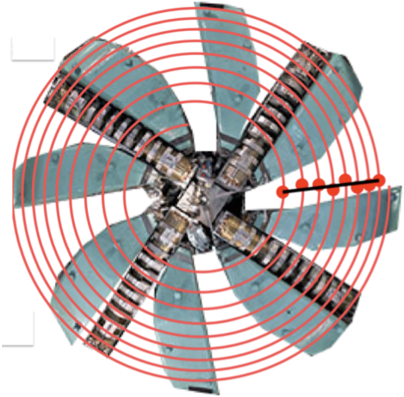
http://www.bt.pa.msu.edu/CPO-10/talks/23Tue/AM1/S1G/23Tue_AM1_1015_S1G_Frey.pdf

- **Discrepancies** in
  - magnetic field **construction** inaccuracies
  - **injection** parameters $(E_{kin}, r, p_r,$
  - element **positioning** (RF cavities)
  - etc.
- **In reality:**

  Additional B-field with **trimcoils**

  $\Longrightarrow$ phase shift

  $\Longrightarrow$ turn radius shift



PAUL SCHERRER INSTITUT PSI

New Trimcoil Model in OPAL

- Radially rational TC profile description

$$TC(r) = B_{\max} \frac{\sum_{i=0}^{n} a_i r^i}{\sum_{j=0}^{m} b_j r^j} \qquad n, m \in \mathbb{N}_0 \wedge TC(r) \in [r_{\min}, r_{\max}]$$

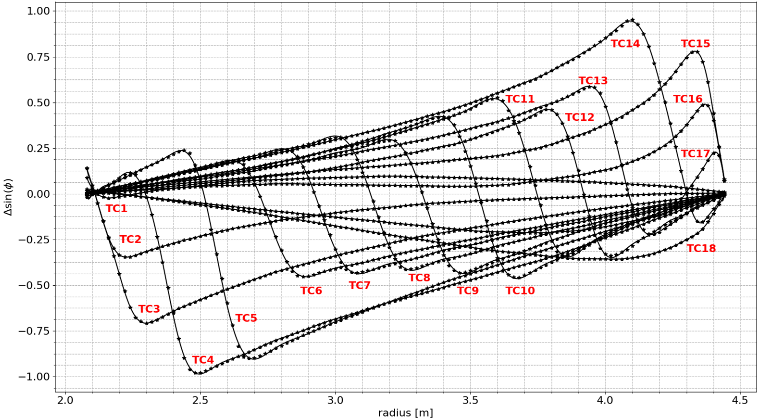
```
tc1:   TRIMCOIL, TYPE = "PSI-PHASE",
                 RMIN = ..., // inner radius [mm]
                 RMAX = ..., // outer radius [mm]
                 BMAX = ..., // B-field peak value [T]
                 COEFNUM   = {a0, a1, a2, a3},
                 COEFDENOM = {b0, b1, b2, b3, b4, b5};

Ring: CYCLOTRON, TRIMCOILTHRESHOLD = ...,
                 // lower limit of TC contribution [T]
                 TRIMCOIL = {tc1, tc2, tc3, ...}
                 ...
                 ;
```

PSI-Ring Trimcoil Model

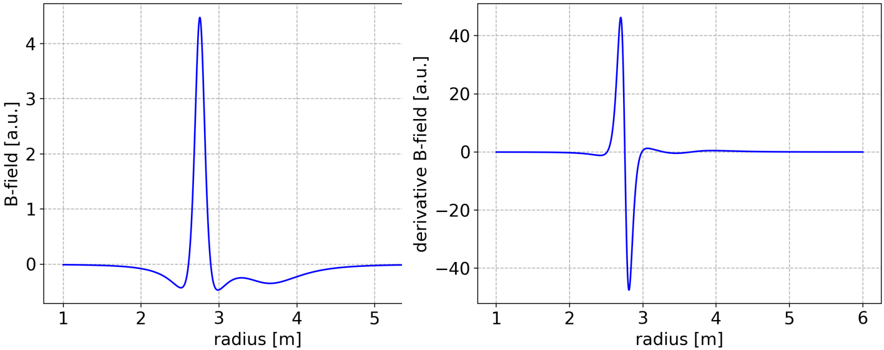
- **Starting point:** Measurement of phase shift effect[5]

$$\Delta B \sim -\frac{d\Delta\sin(\phi)}{dr}$$



[5] S. Adam and W. Joho, PSI Technical Report No. TM-11-13, 1974.

PSI-Ring Trimcoil Model - Example TC6



PAUL SCHERRER INSTITUT PSI

Multi-Objective Optimisation (MOO) in OPAL

- **Built-in MOO**[6]:

$$\begin{array}{lcr} \min & \mathbf{f}(\mathbf{x}), & \dim(\mathbf{f}) = M \in \mathbb{N}^{>0} \\ \text{s.t.} & \mathbf{g}(\mathbf{x}) \geq 0, & \dim(\mathbf{g}) = J \in \mathbb{N}^{0} \\ & -\infty \leq x_i^L \leq \mathbf{x} = x_i \leq x_i^U \leq \infty, & \mathbf{x} \in \mathcal{X} \subset \mathbb{R}^n, \quad n \in \mathbb{N}^{>0} \end{array}$$

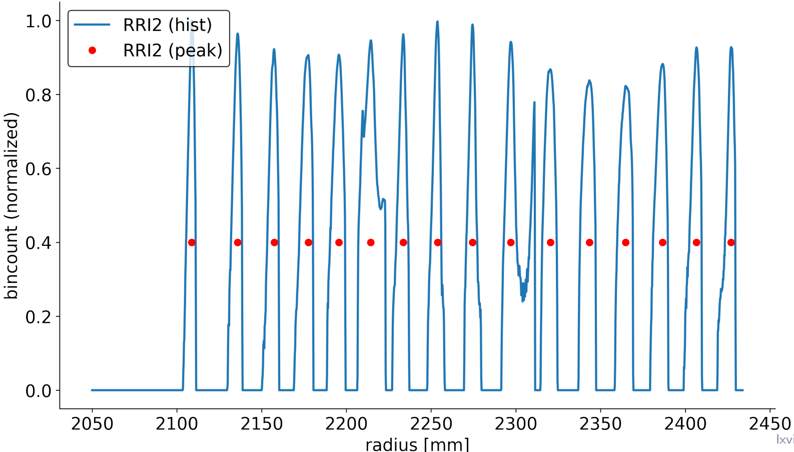
- **Design variables $\mathbf{x}$:** E_{kin}, p_r, φ, TC1 - TC16 max. B-field, etc.
- **Objectives:** Measure between simulation and real data

Note: f is our PSI-Ring model + evaluation of objectives!

[6] Toward massively parallel multi-objective optimisation with application to particle accelerators. PhD Thesis. Y. Ineichen. 2013

Radial Profile Measurement

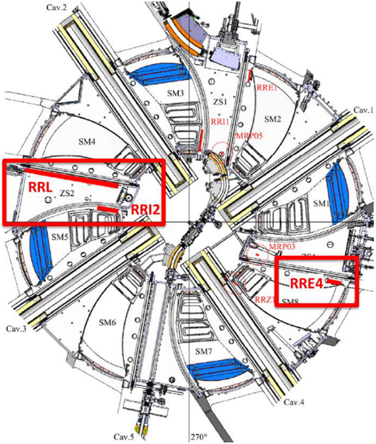
- **Measurements:** Peak intensity of radial profile of probes to distinguish turns



Trimcoil Optimisation in OPAL

- **Simulations:**
  - **Single particle** ⇒ probe hit = turn
  - **Multi particles** ⇒ peak finder routine
- **Good setting:** Radial peak of measurement and simulation at probes are close!
- **RRI2:** turns 1 - 16
- **RRL:** turns 9 - 182
- **RRE4:** turns 177(8) - 188(9)

188(9) turns ⇒ Infeasible number of objectives!



OPAL simulations of the PSI ring cyclotron and a design for a higher order mode flat top cavity. N. J. Pogue, A. Adelmann. Proceedings of IPAC2017. THPAB077. 2017.

PAUL SCHERRER INSTITUT PSI

Problem Reduction

- **Turn - Aggregation:**
  - L_2-norm

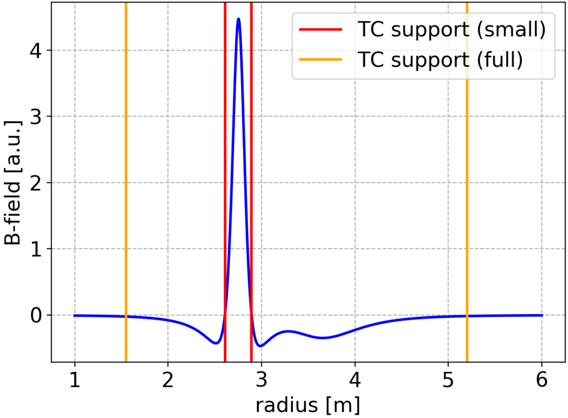
$$\text{err} = \frac{1}{N_{turns}} \sqrt{\sum_{i=1}^{N_{turns}} (p_{i,meas} - p_{i,sim})^2}$$

  - L_∞-norm

$$\text{err} = \max_{i=1,\ldots N_{turns}} |p_{i,meas} - p_{i,sim}|$$

Problem Reduction

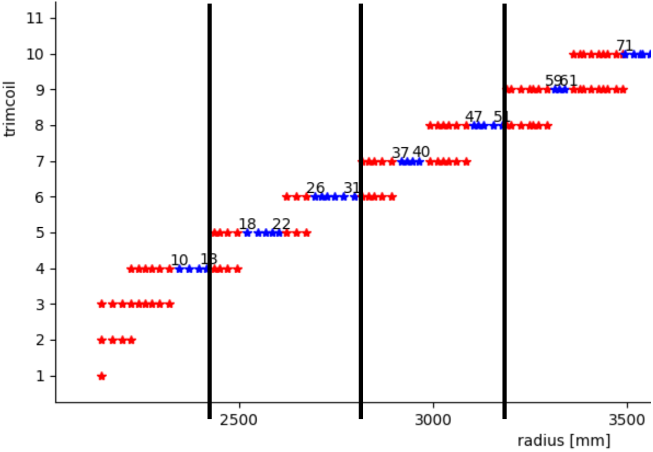
- **TC support reduction:**
  Feasible assumption for neighbouring TCs ⇒ Cancellation of B-field tails



PAUL SCHERRER INSTITUT PSI

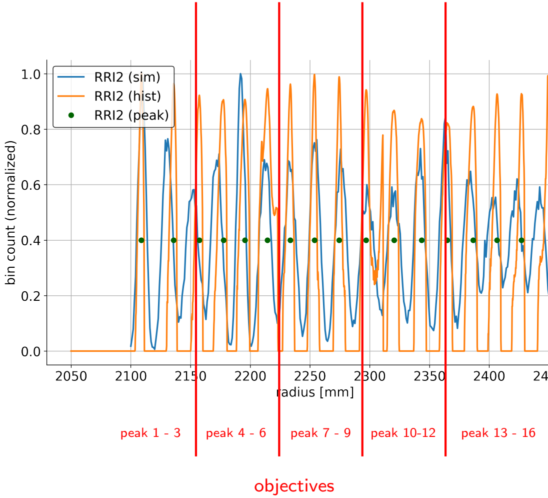
Problem Reduction

- **Optimise on sub-problems:**



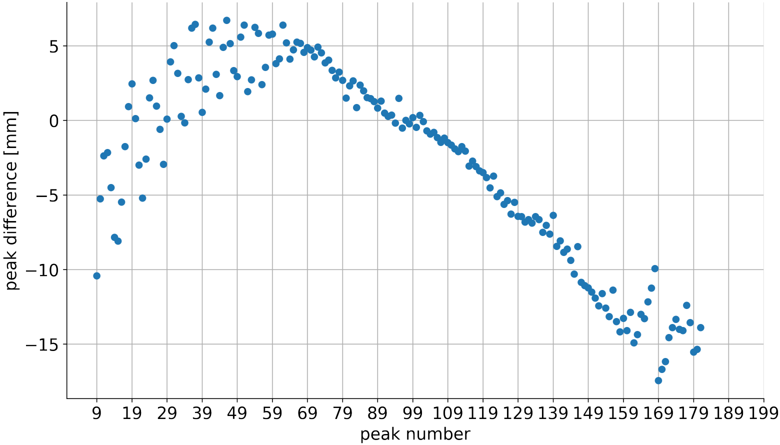
Trimcoil Optimisation in OPAL - Trial 1

- **Goal:**
  Find initial injection values
- **Design variables:**
  - beam energy E_{kin}
  - injection angle
  - injection momentum
  - injection radius
  - TC1 - TC4
- **MOO:** (504 cores)
  #generations 500 +
  #individuals 502
- 5000 particles per individual



Issue of Divergence - Trial 1

- Optimising a few TCs after the others lead to divergence!



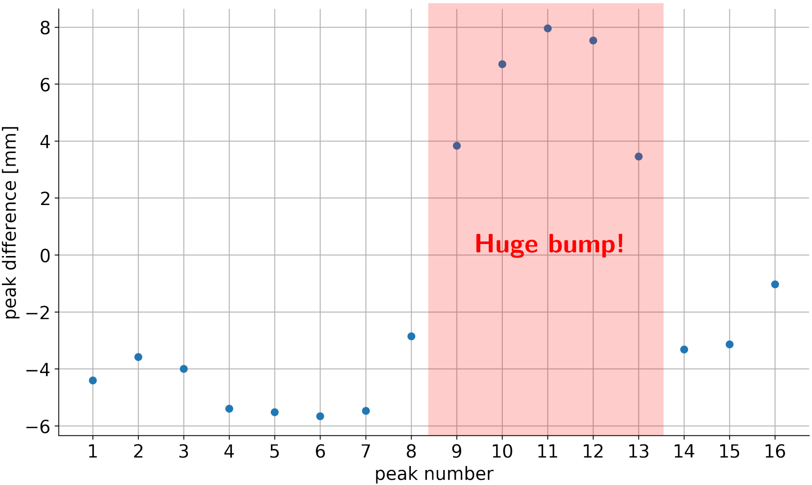
Model Simplification + Design Variable Extension

- **Single particle tracking** instead of bunch (5000 particles) tracking

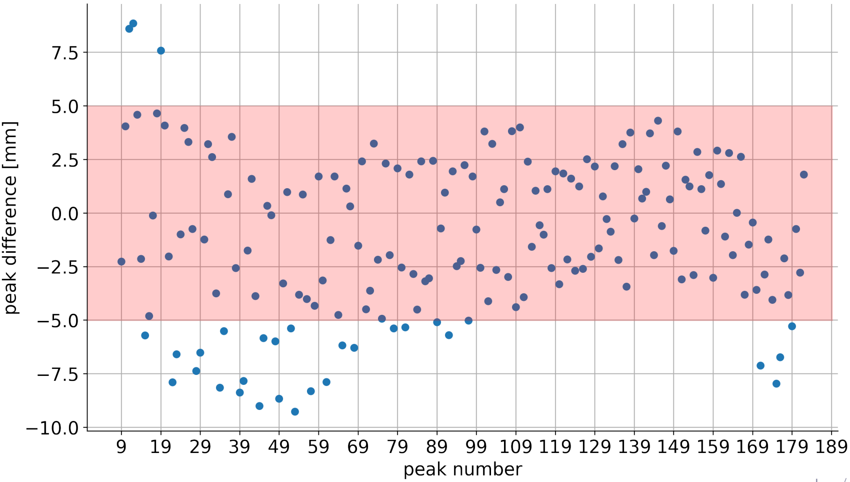
  $\Longrightarrow$ full PSI-Ring simulation in 1 - 2 s
- **Design variables:**
  - injection angle, radius, momentum and energy
  - main cavity voltages
  - phase of Flat-Top cavity
  - voltage of Flat-Top cavity
  - radial position of main cavities
  - radial position of Flat-Top cavity
- **Turn number constraint** to guarantee feasible solutions

PAUL SCHERRER INSTITUT PSI

Injection Probe RRI2



Long Probe RRL1 - No divergence anymore!

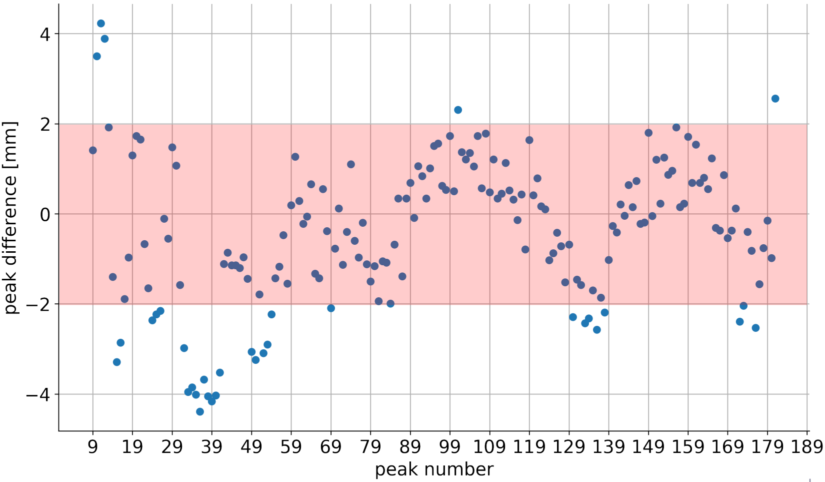


Scanning instead of MOO

- **Issues:**
  - Optimiser suffered with individual selection
  - No further improvements!
  - Changing all parameters at same time might be disadvantageous
- **Idea:** Do simple parameter scanning!
  - Starting from **best MOO result**
  - Iteratively find worst turn and vary parameters to obtain better individual

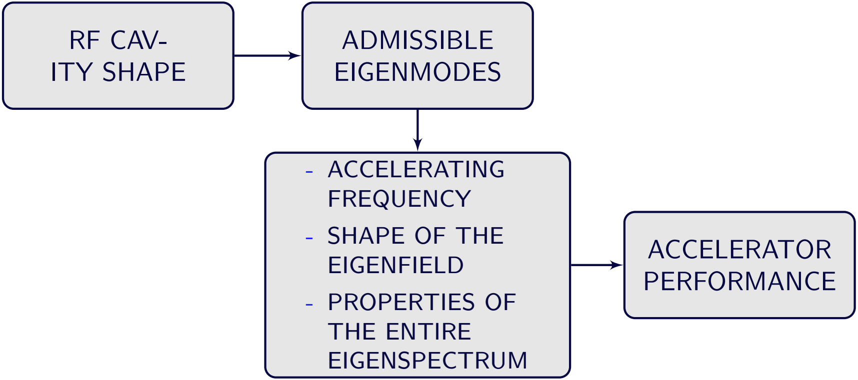
    (check L_∞- and L_2-norm)

Long Probe RRL1



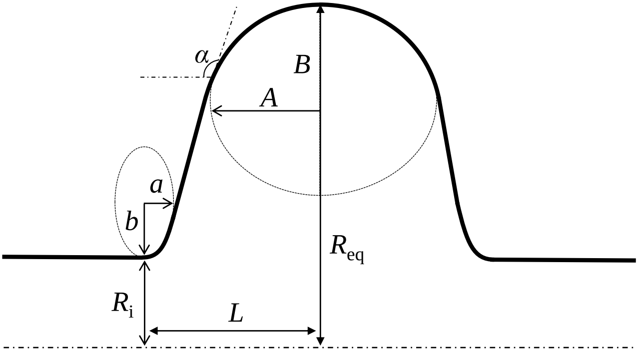
Shape optimization of RF cavities, FCC related

M. Kranjčević, P. Arbenz, ETH Zurich A. Adelmann, Paul Scherrer Institut (PSI), S. Gorgi Zadeh, U. van Rienen, University of Rostock



http://www.bt.pa.msu.edu/CPO-10/talks/23Tue/PM1/S1G/23Tue_PM1_1515_S1G_Kranjcevic.pdf

Single-cell elliptical cavity parameterization



- axisymmetric, variables $R_i, L, A, B, a, b, (R_{eq}) \to \alpha$

Constrained multi-objective optimization problem

- monopole and dipole modes major sources of beam instability

$$\min_{R_i,L,A,B,a,b} \quad (\overbrace{f_0 - f_1}^{F_1}, \overbrace{|f_1 - f_2|}^{F_2}, \overbrace{\frac{R}{Q}_{\perp 1} + \frac{R}{Q}_{\perp 2}}^{F_3}, \overbrace{-G_0 \cdot \frac{R}{Q}_0}^{F_4}),$$

$$\text{subject to} \quad f_0 = 400.79 \text{ MHz}, \quad \alpha \geq 90^\circ$$

- f_0 ... frequency of the fundamental mode
- f_1, f_2 ... frequency of the first and second dipole mode, resp.
- $\frac{R}{Q}_\perp$ ... transverse shunt impedance for the dipole modes[7]
- G_0 ... geometry factor [8]

---

[7] B. P. Xiao et al., IPAC Richmond, VA, USA , 2015. https://doi.org/10.18429/JACoW-IPAC2015-WEPWI059

[8] J. Sekutowicz et al., PAC, Portland, OR, USA, 2003. https://doi.org/10.1109/PAC.2003.1289717

Forward solver

- Maxwell's equations
  - frequency domain
  - axisymmetric domain in 3D[9,10]
  - vacuum; no external fields, sources or charges; PEC
- FEM $\rightarrow$ a GEVP for each azimuthal mode number $m \in \mathbb{N}_0$
- smallest eigenpair for (using half of the cross section)
  - $m = 0$, PEC $\rightarrow$ properties of the fundamental mode (TM_{010})
  - $m = 1$, PEC $\rightarrow$ properties of the dipole mode TM_{110}
  - $m = 1$, PMC $\rightarrow$ properties of the dipole mode TE_{111}

---

[9] P. Arbenz, et al., Appl. Numer. Math. 58 (4): 381-394, 2008. https://doi.org/10.1016/j.apnum.2007.01.019

[10] O. Chinellato, ETH Zurich (Diss. ETH No. 16243), 2005. https://doi.org/10.3929/ethz-a-005067691

Evolutionary algorithm (EA)

- evaluate a random population of individuals I_i, $i = 1, \ldots, N$
- for a predetermined number of generations do
  - variator: for pairs of individuals I_i, I_{i+1}, perform:
    $$crossover(I_i, I_{i+1}), \;\; mutation(I_i), \;\; mutation(I_{i+1})$$
  - evaluate new individuals
  - selector: choose N fittest individuals for the next generation

– massively parallel implementation[11] same as in OPAL

– combined with the axisymmetric Maxwell eigensolver[12]

[11] Y. Ineichen et al., Comput. Sci. Res. Dev. 28 (2) (2013) 185-192. https://doi.org/10.1007/s00450-012-0216-2

[12] M. Kranjčević et al., arXiv:1810.02990,2018

Constraint handling

- $f_0 = 400.79$ MHz
  - given $\boldsymbol{d} = (R_i, L, A, B, a, b)$, find R_{eq} s.t. $f_0 = 400.79$ MHz
  - if $|f_0 - 400.79 \text{ MHz}| \geq 1$ MHz, fine mesh eigensolve avoided (on average, 4 fine eigensolves for each $\boldsymbol{d}$)

- $\alpha \geq 90^\circ$ ... otherwise, the individual is discarded

Results

Euler cluster[13] (Euler I and II) of ETH Zurich FORWARD SOLVE:

- coarse eigensolves . . . 10'000 triangles, 2 s
- fine eigensolves . . . 300'000 triangles, 90 s
  (24 s meshing, 64 s eigenpairs, 2 s objective function values)
- 4 fine eigensolves to find R_{eq} and the properties of TM_{010}
- 2 fine eigensolves to find the properties of TM_{110} and TE_{111}
  (no remeshing)

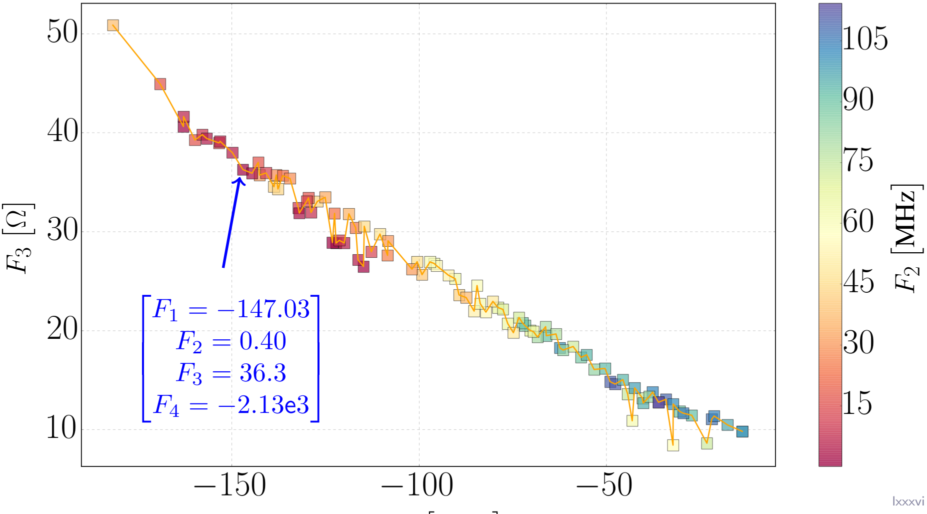
OPTIMIZATION:

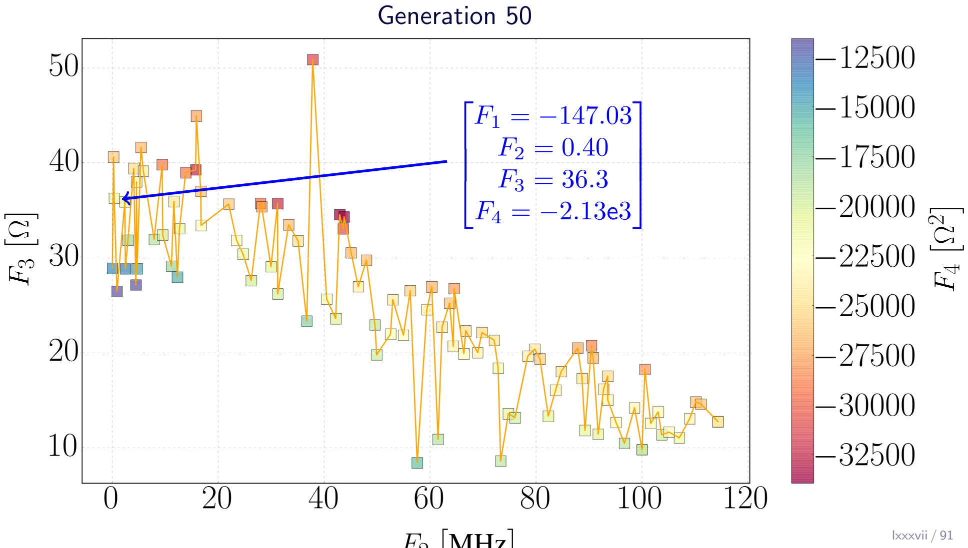
- 13 h for 50 generations with $N = 100$ on 96 processes
  (30% of the individuals discarded)
- initial design variable bounds:

Variable	R_i	L	A	B	a	b
Lower bound [mm]	145	120	40	40	10	10
Upper bound [mm]	160	190	140	140	70	70

[13] https://scicomp.ethz.ch/wiki/Euler

Generation 50





Generation 50
$F_1 = -147.03$
$F_2 = 0.40$
$F_3 = 36.3$
$F_4 = -2.13e3$
$F_3\ [\Omega]$
F_2 [MHz]
$F_4\ [\Omega^2]$
50
40
30
20
10
0
20
40
60
80
100
120
−12500
−15000
−17500
−20000
−22500
−25000
−27500
−30000
−32500

Fundamental mode of the chosen RF cavity

Hands-on: Python Implementation

- DEAP a novel evolutionary computation framework for rapid prototyping and testing of ideas.
- It seeks to make algorithms explicit and data structures transparent.
- It works in perfect harmony with parallelisation mechanism such as multiprocessing.
- `https://github.com/deap/deap`

References

[Y. Ineichen et al., arXiv:1302.2889] A Parallel General Purpose Multi-Objective Optimization Framework, with Application to Beam Dynamics Y. Ineichen, A. Adelmann, A. Kolano, C. Bekas, A. Curioni, P. Arbenz, arXiv:1302.2889, 2013

[Y. Ineichen, ETH Ph.D Thesis (2013)] Y. Ineichen ETH-Diss 21114, 2013

[Y. Ineichen et al., CS-R&D (2012)] Y. Ineichen, A. Adelmann et al., Computer Science - Research and Development, pp. 1-8. Springer, Heidelberg, 2012.

[O.L. De Weck] Multiobjective Optimisation: History and Promisses, CJK-OSM3 `http://strategic.mit.edu/docs/3_46_CJK-OSM3-Keynote.pdf`, 2004