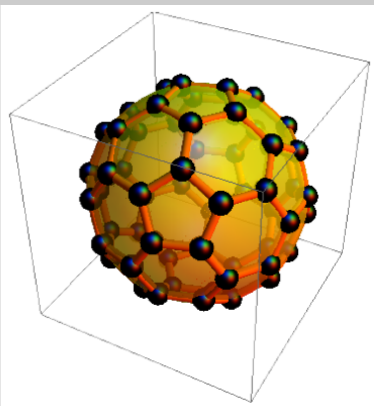
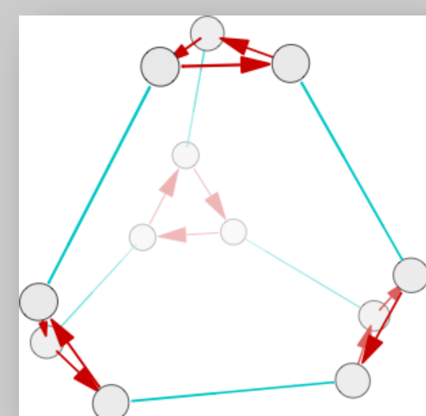
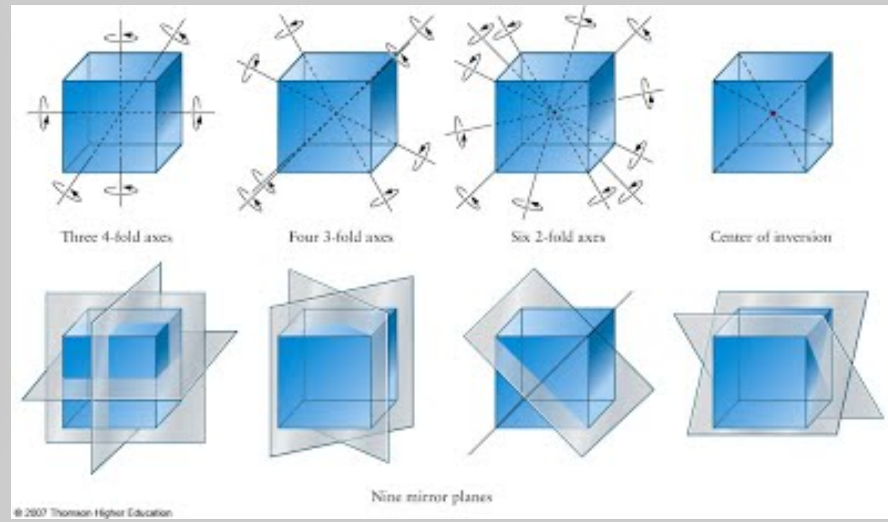
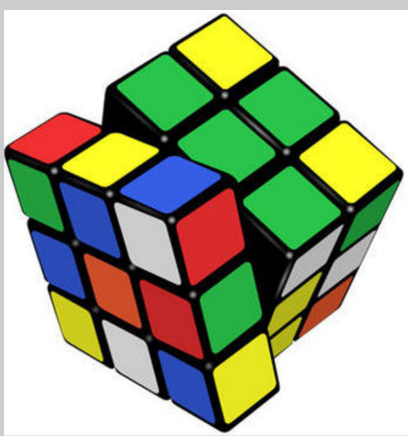
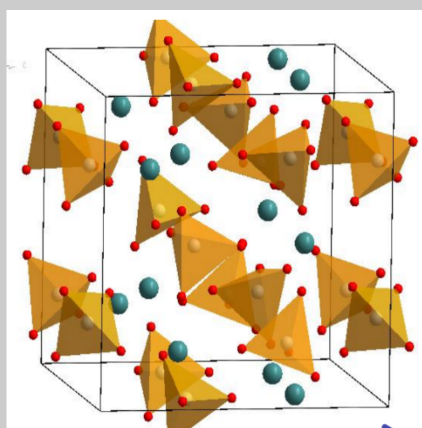
How good is your algorithm ? A group-theoretic framework for assessing and comparing nonlinear optimisation techniques











Stuart Anderson

Space and Atmospheric Physics Group
University of Adelaide

Outline

Principal message : the need for synthetic problems to assess NLO techniques

Optimisation environments : off-line vs real-time, single vs multi-objective

Context for the present study : OTH radar

Basis for the proposed approach : extrema attained under symmetry

Group theoretic formulation and generalisation

Alternative synthesis environments

Results

Conclusions

Complex problems with known exact solutions I

In most operational applications of nonlinear optimisation techniques to large-scale problems, the true solution can never be established with certainty. This makes it difficult if not impossible to determine whether the adopted algorithm is converging to the global extremum, nor can we gauge the true rate of convergence

Moreover, if we have a choice of algorithms, or an algorithm with tunable parameters, it is not always clear which setting is best for the class of problems of interest

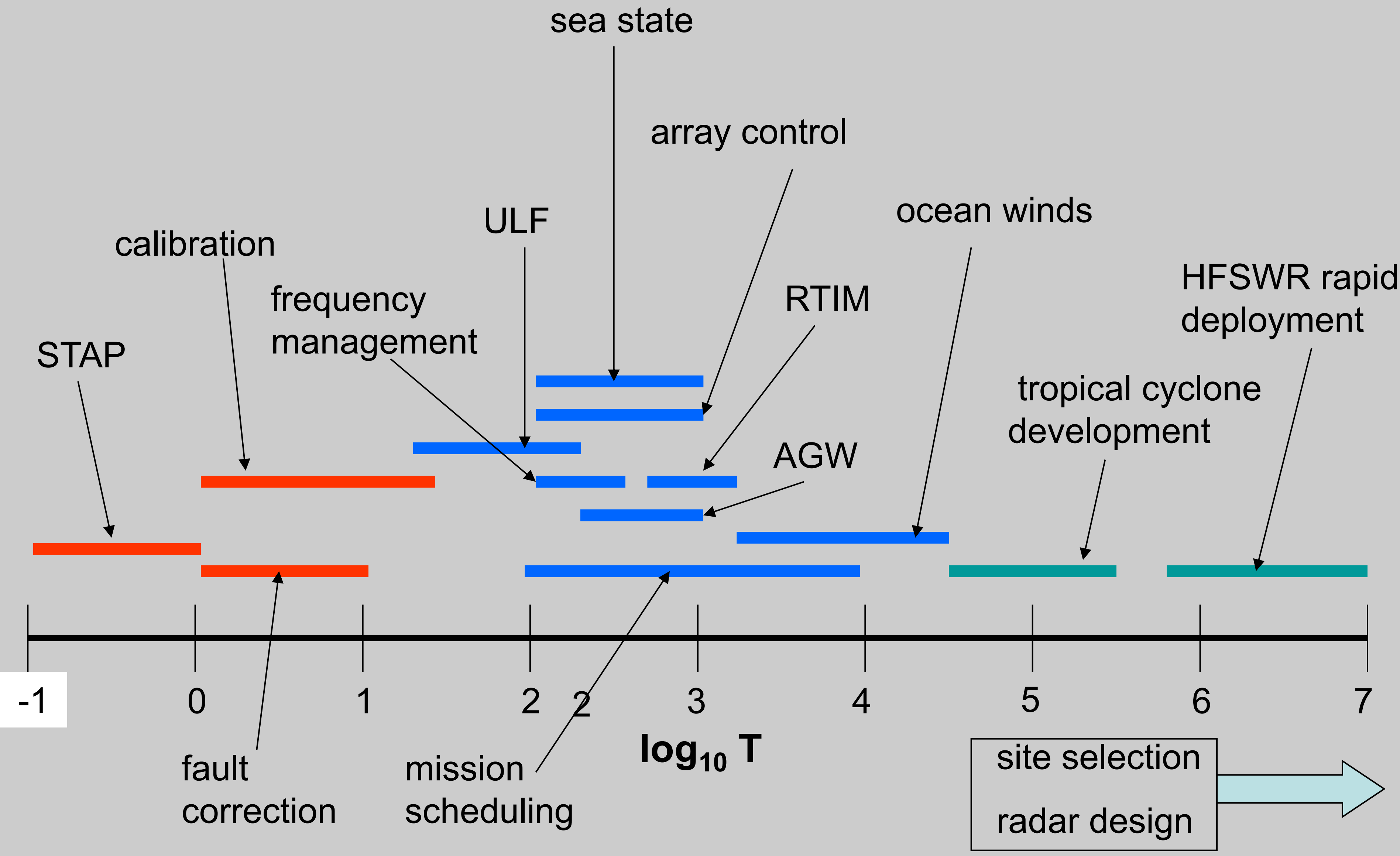
One way to investigate these issues is to apply our techniques to classes of problems which possess complex landscapes, yet for which we know the exact solutions and for which we can control the landscape statistics and the metrical properties around the extrema

Our approach to designing such classes is to exploit the laws of symmetry which characterise minimum energy states in physical systems, and their counterparts in group theory

Motivation behind the work presented here

- Users of geophysical remote sensing systems seek more and more detailed information, so sensor channels grow in number and resolution and the associated inverse problems become intractable using classical techniques
- Moreover, sensors increasingly are required to adapt their parameters to the prevailing environment in real-time to optimise performance
- The associated objective function landscapes are almost invariably very complex and the search for global extrema computationally demanding
- Genetic algorithms have proven effective for dealing with nonlinear optimisation problems with complex landscapes, especially design tasks where time is not a critical concern,
- Their stochastic nature has largely precluded their use in those real-time remote sensing applications where the sensor may have stringent timing constraints on data acquisition and processing
- Accordingly we are motivated to seek ways of accelerating convergence so that we can employ GAs in our remote sensing applications

Characteristic time scales of radar inverse problems and parameter optimisation



Priority tasks for optimisation techniques in off-line and real-time cases

OFF-LINE DESIGN	find the global extremum, ie, the single best design
	map the Pareto front for multi-objective problems
	quantify the penalty of selecting near-optimum solutions which might have other advantages not considered in the optimisation
REAL-TIME APPLICATIONS	deliver convergence to an acceptably accurate approximation of parameters of interest within the rigid data acquisition, processing and distribution schedule
	support real-time adaptation of system degrees of freedom where this will improve the fidelity of instrumental observations and the subsequent inversions to retrieve the desired parameters
	identify occurrences of measurement data where the inversion is likely to be degraded
	if independent data is available for fusion, optimize the quality of the fused product, even at the expense of individual system quality metrics

Priorities for real-time remote sensing applications

(i) For the geophysical parameters of interest, deliver convergence to an acceptably accurate inversion within the rigid data acquisition, processing and distribution schedule

(ii) Support real-time adaptation of sensor degrees of freedom where this will improve the fidelity of instrumental observations and the subsequent inversions to retrieve the geophysical parameters

(iii) Identify occurrences of measurement data where the inversion is likely to be degraded

(iv) If other sensor data is available for fusion, optimise the quality of the fused product even at the expense of individual sensor quality metrics

Priorities for off-line and slow-time applications

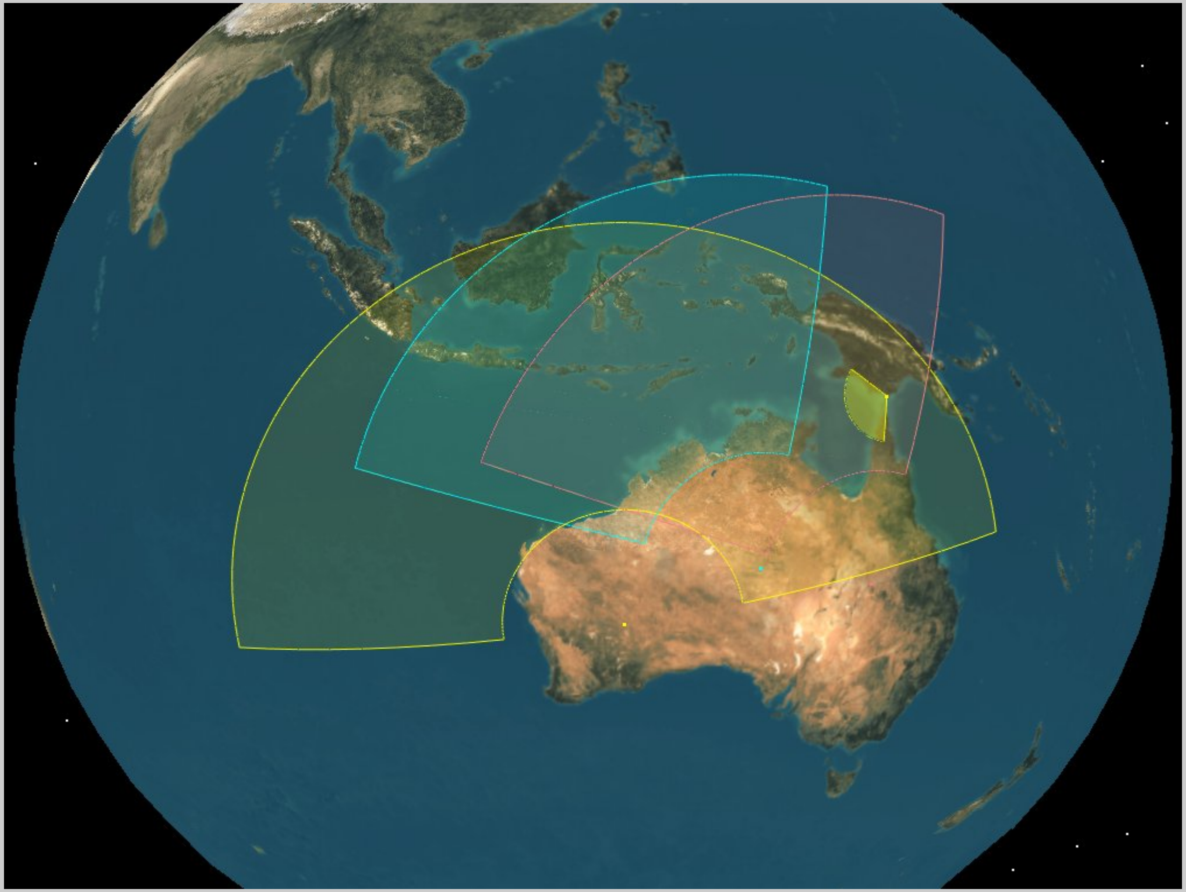
(i) Find the global extremum, ie, the best design

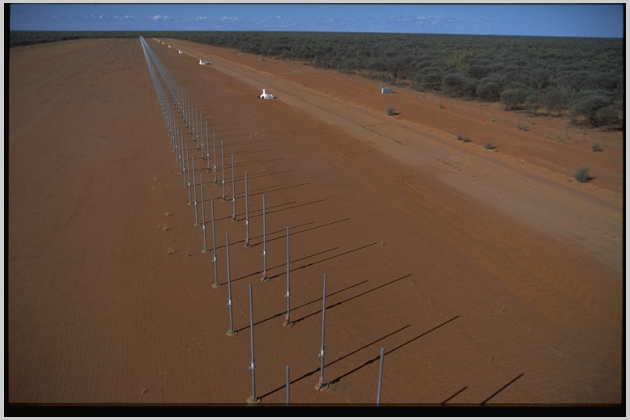
(ii) Map the Pareto front for multi-objective tasking

(iii) Quantify the penalty of selecting near-optimum solutions which might have other advantages not considered in the optimisation

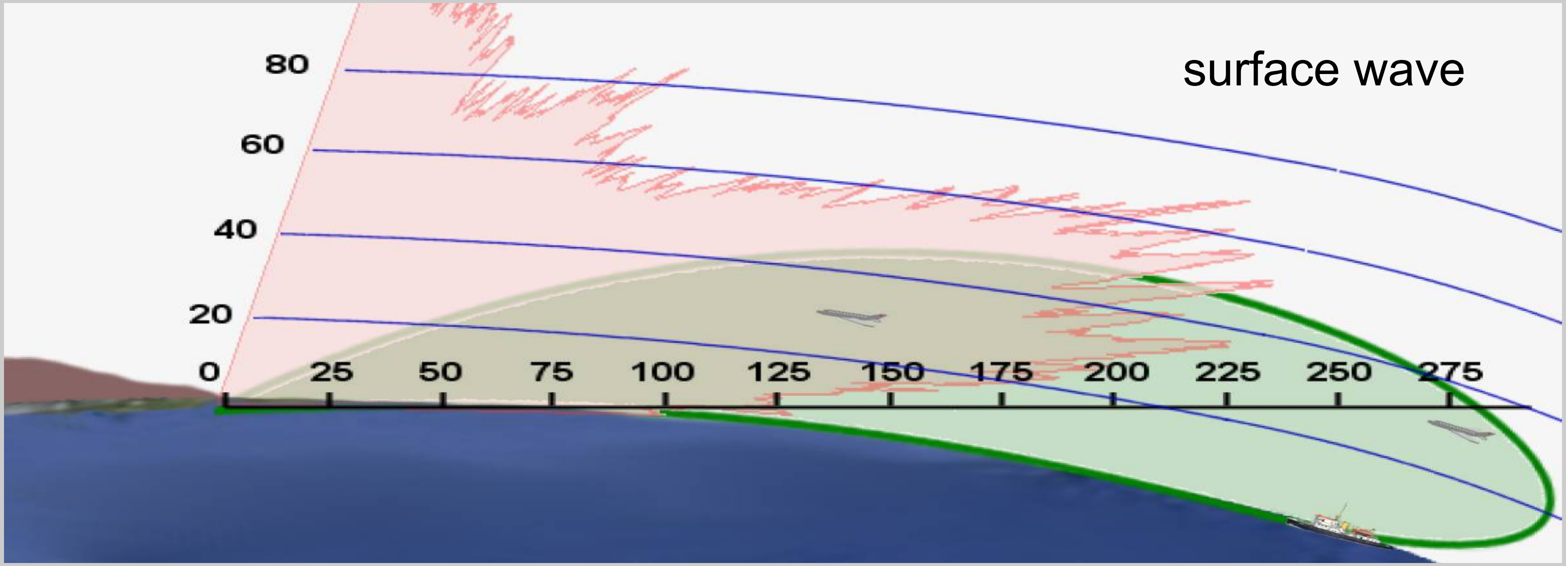
JORN : antenna arrays and nominal coverage

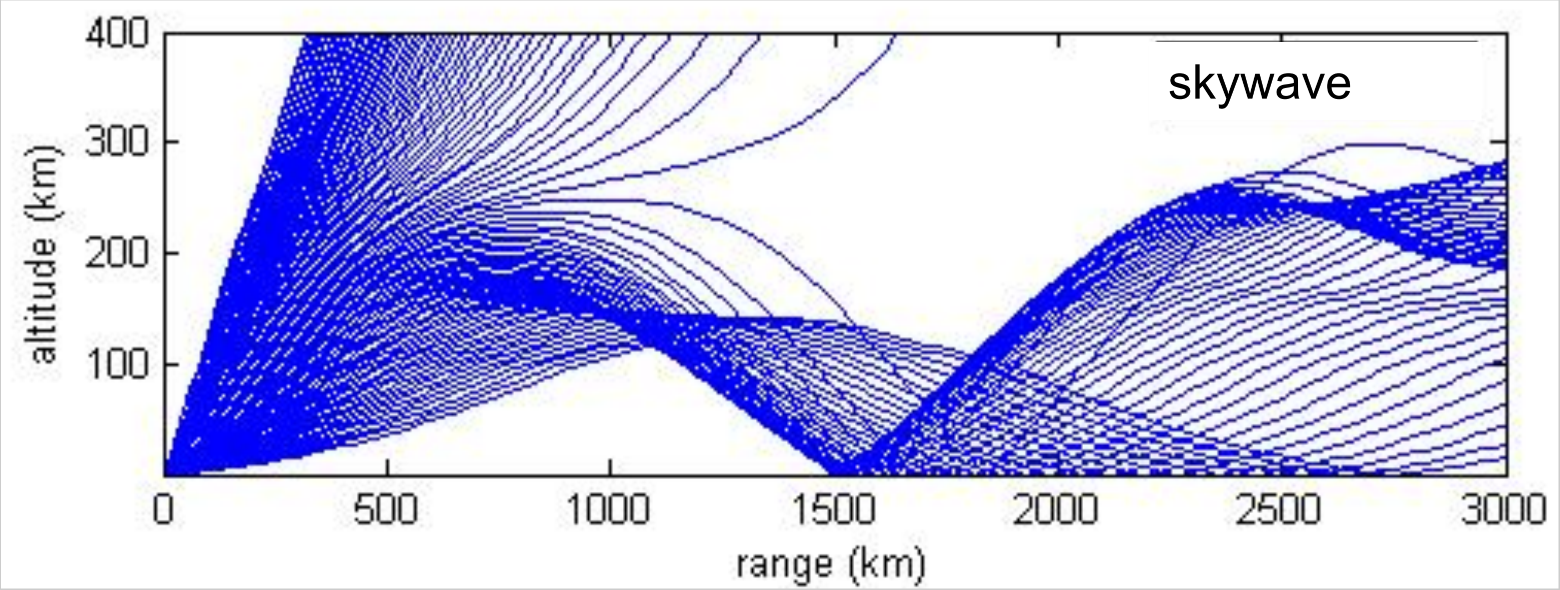
Missions : air and surface surveillance
remote sensing





Propagation modes for over-the-horizon radar





The radar process model

$$S = S_t + S_e + S_{te}$$

$$s = \widetilde{P} \sum_{n_B=1}^{N} \widetilde{R} \left[\prod_{j=1}^{n_B} \boxed{\widetilde{M}_{S(j)}^{S(j+1)}} \widetilde{S}(j)\right] \boxed{\widetilde{M}_{T}^{S(1)}} \widetilde{T} w$$

$$+ \widetilde{P} \sum_{l=1}^{N_J} \sum_{m_B=1}^{M} \widetilde{R} \left[\prod_{k=1}^{m_B} \widetilde{M}_{S(k)}^{S(k+1)} \widetilde{S}(k)\right] \widetilde{M}_{N}^{S(1)} n_l + m$$

w represents the selected waveform

$\widetilde{T}$ represents the transmitting complex, including transmitters and antennas

$\widetilde{M}_{T}^{S(1)}$ represents propagation from transmitter to the first ground scattering region

$\widetilde{S}(j)$ represents all scattering processes in the j-th region

$\widetilde{M}_{S(j)}^{S(j+1)}$ represents propagation from j-th scattering region to the (j+1)-th region

$S(n_B)$ $S(m_B)$ represent the receiver location

n_l represent external noise sources, interferers or jammers

$\widetilde{M}_{N}^{S(1)}$ represents propagation from a noise source to its first ground scattering region

m represents internal noise

$\widetilde{R}$ represents the receiving complex, including antennas and receivers

$\widetilde{P}$ represents the signal processing

s represents the signal decomposition after processing

Inverse problems in OTH radar : Any operator may be the target unknown in the equation

Diagnose perturbed propagation so it can be corrected

Measured Doppler spectra of sea echoes

Design waveform to optimise measurement of interest

$$s = P \; R \; M_R \; S \; M_T \; T \; w$$

Tune signal processing to optimise performance

Calibrate transmitting antenna array

Calibrate receiving antenna array and assign receivers

Estimate sea state from the Doppler spectrum of the scattered signals

Inverse problem # 1 : Estimation of directional wave spectra from HF sea clutter (simplified form)

Ideal distortion-free propagation

Doppler spectra of sea clutter

Ideal distortion-free propagation

$$s = P \; R \; M_R \; \boldsymbol{S} \; M_T \; T \; w$$

Narrow-band signal

3-D spatio-temporal processing

Narrow beam receiving antenna array

Narrow beam transmitting antenna array

Scattering from random sea surface

characterised by directional wave spectrum

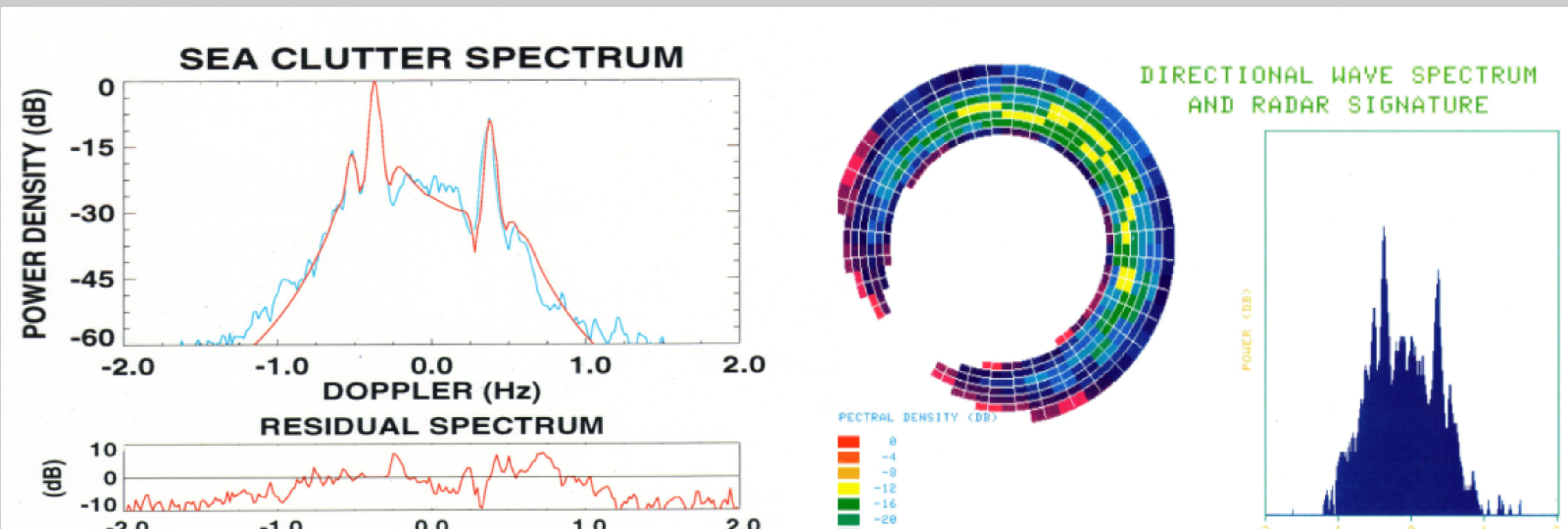
Integral equation relating the scattering operator to the ocean directional wave spectrum

$$S\left(k_{inc},k_{scat}\right)=\int d\kappa_1 F_1\left(k_{inc},k_{scat},\kappa_1\right)S\left(\kappa_1\right)\delta\left(k_{inc}+k_{scat}\pm\kappa_1\right)$$

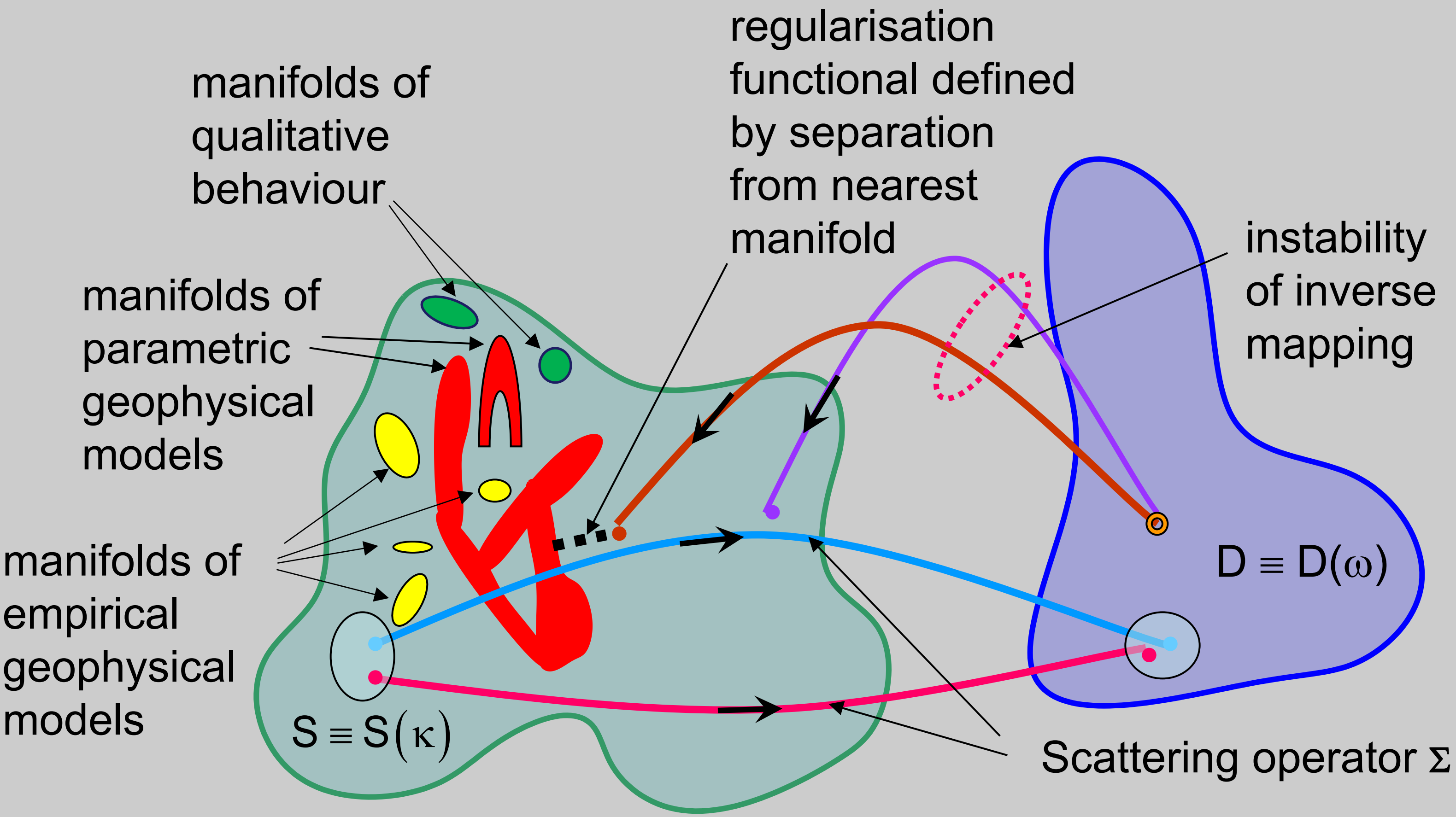
$$+\iint d\kappa_1 d\kappa_2 F_2\left(k_{inc},k_{scat},\kappa_1,\kappa_2\right)S\left(\kappa_1\right)S\left(\kappa_2\right)\times$$

$$\delta\left(k_{inc}+k_{scat}\pm\kappa_1\pm\kappa_2\right)$$

$$+\ldots\ldots$$



Radar echo inversion via regularisation employing both empirical and parametric models



$$S_{opt}(\kappa) = \min_{S(\kappa)\in S}\Big[\ \min_{\Sigma(S(\kappa))\in D}\left\|\Sigma(S(\kappa)) - D_{meas}(\omega)\right\|_D$$

$$+\lambda \min_j\left(\min_{S_\beta \in GM_j}\left\|S(\kappa) - S_\beta\right\|_S\right)\ \Big]$$

Considerations in probing signal design

- The signal should be designed to have high sensitivity to the specific physical phenomenon of interest
- It must also take into account
  - the effects of propagation to and from the target area
  - interaction of the signal with other features of the environment
  - the signal-dependence of the Tx and Rx subsystems
  - the presence of noise and interference
- The remote sensing system may need to adapt its signal in near- real-time to maintain optimal performance
- The inputs to the adaptation process include solutions to both direct and inverse problems, which must be solved within the characteristic timescale of significant environmental variations.
- The signal should be tolerant of non-ideal sensor properties

2-D Tx array

Adapt 56 antenna elements to achieve optimum illumination of radar footprint

Adaptation DoF

(i) phase
(ii) gain
(iii) code

2-D Rx array

Adapt 960 antenna elements to achieve optimum reception from radar footprint

Optimum states :

(i) uniform
(ii) focussed

Real-time optimisation of receiver assignment

Problem #1 : fraction α of failed elements in ULA of cardinality n → ${}^{n}C_{[\alpha n]}$

Problem #2 : assignment of n receivers to 2n-element L-array → ${}^{2n}C_{n}$

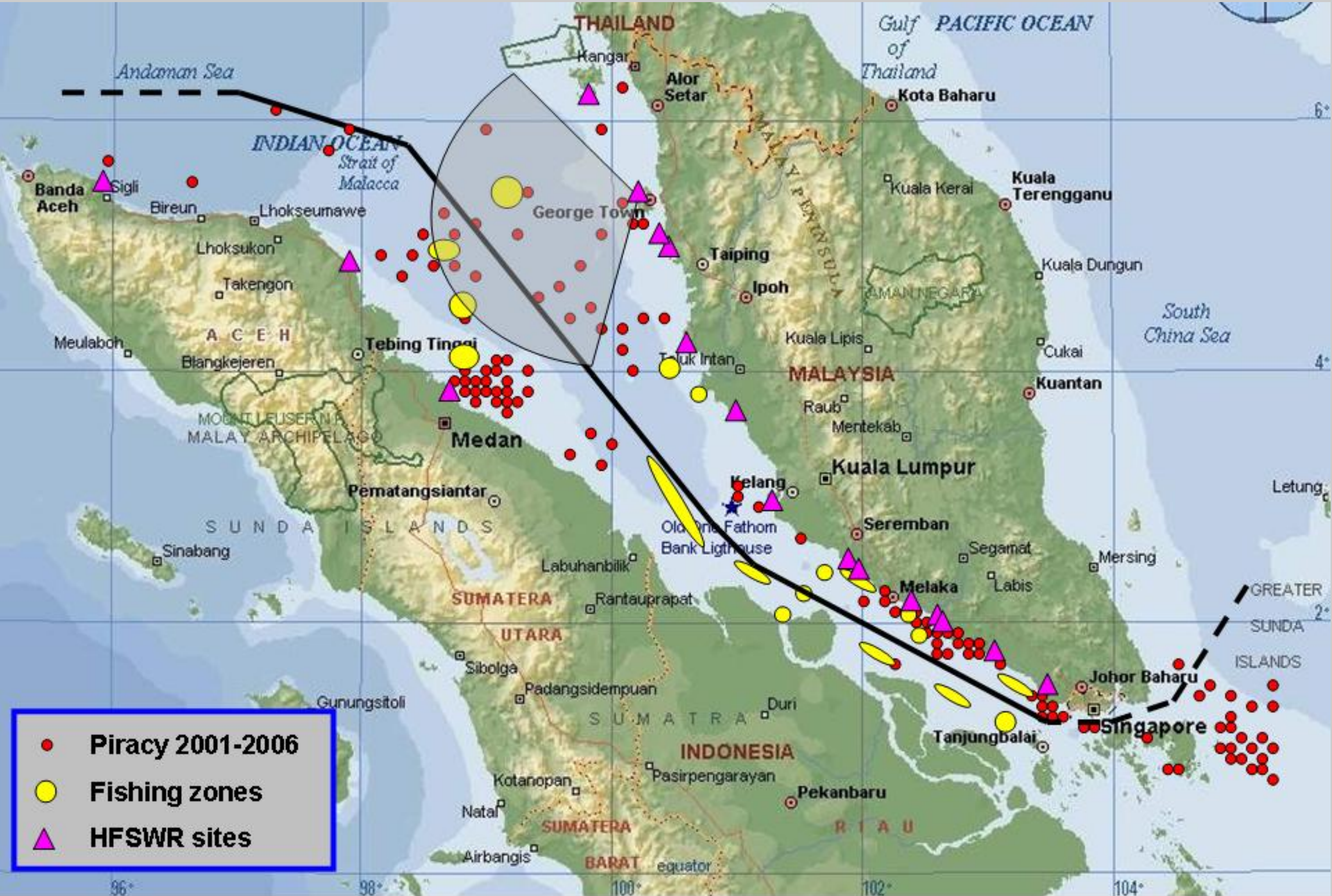
$n = 480$

$\alpha = 0.01$

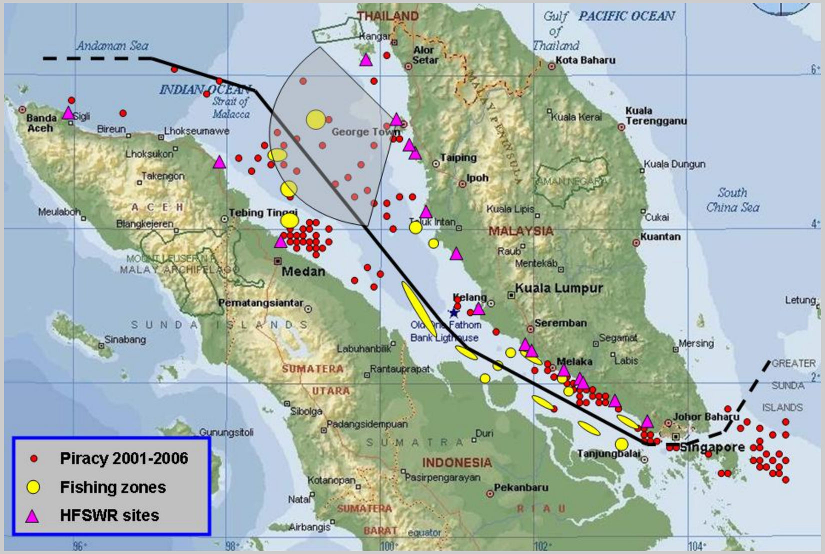
#1 : 2.1×10^{11}

#2 : 2.5×10^{286}

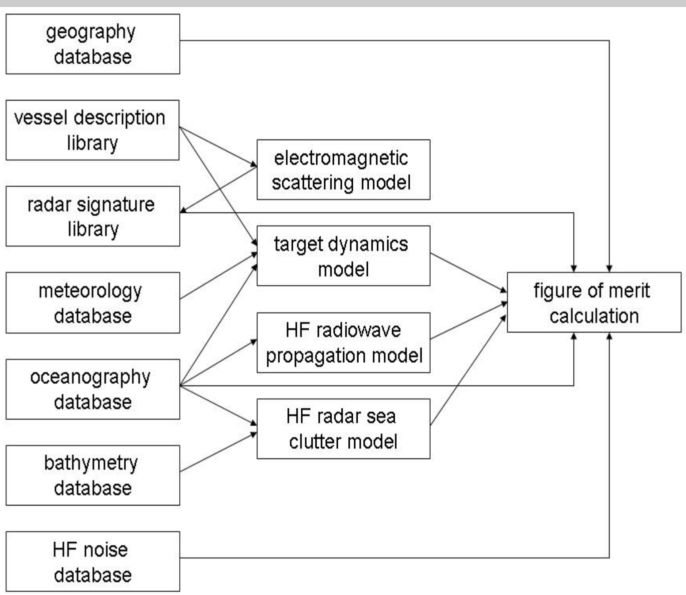
Context for Strait of Malacca HFSWR optimisation study

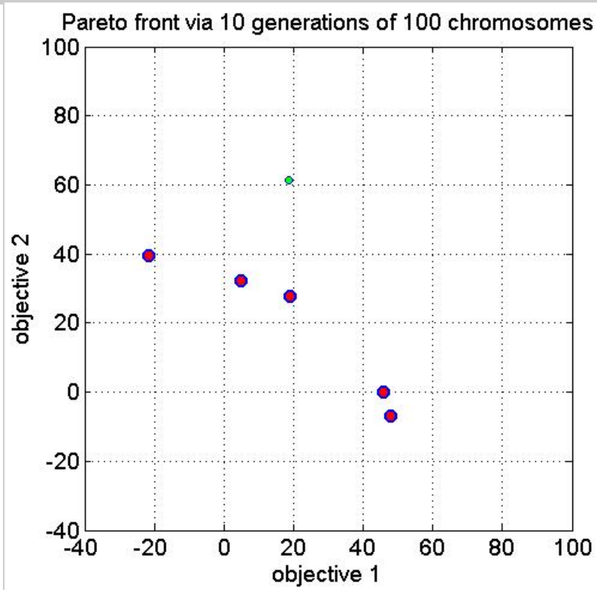


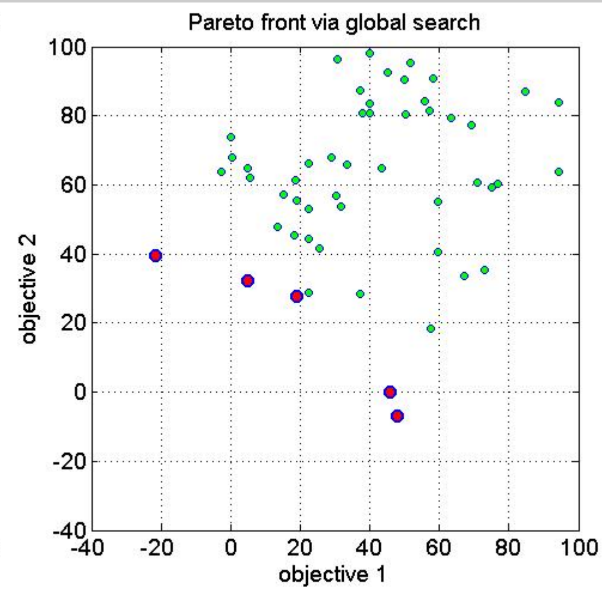
Multi-objective optimisation for the HFSWR site selection



- nominate radar missions to be addressed
- establish geographical priority map
- define performance metrics for each radar mission
- propose the number of radars to be deployed
- select candidate sites
- compute Pareto-optimal solutions







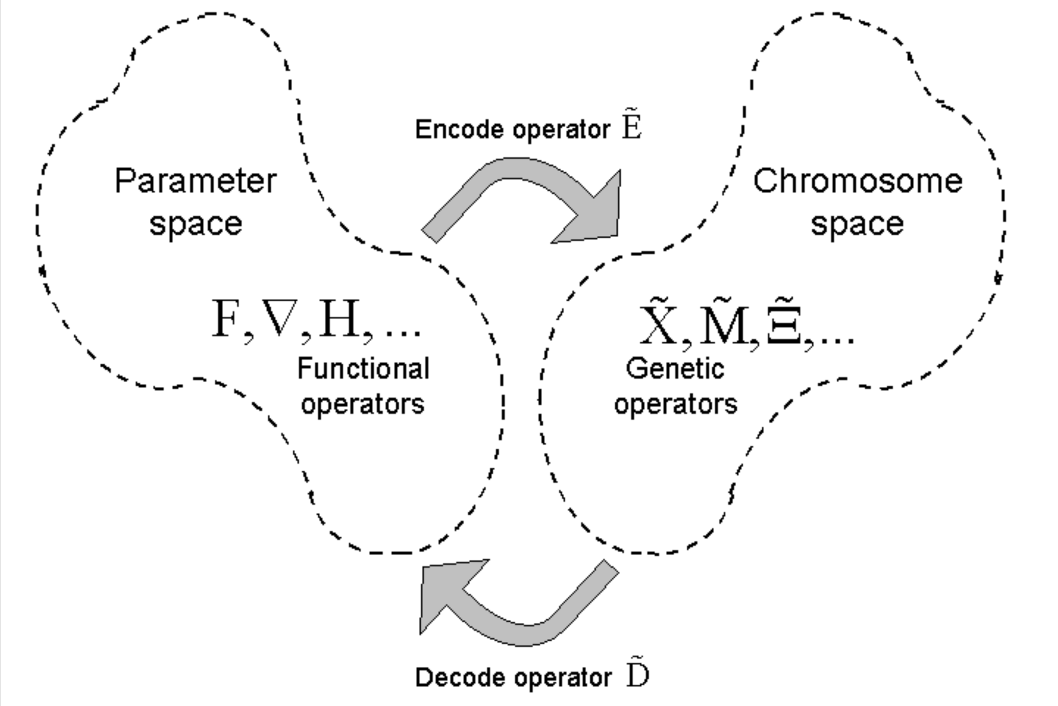
HFSWR radar site options for the rim of the South China Sea

- In the conventional GA methodology, parameter space and chromosome space are treated very differently.

- The fitness function is naturally defined and evaluated on parameter space, where gradient search is effected; the genetic operators perform their evolutionary steps in chromosome space

- Most previous implementations of hybrid methods which combine gradient and GA schema have retained this distinction, to a large extent

- Our approach attempts to exploit a duality between the two domains of representation, whereby the action of a specific operator ("eugenics") involves both spaces.



Hybrid scheme for accelerated convergence

Recall that each chromosome is the image in S_C under a mapping, which we represent by the encode operator, E , of a state represented by a point in S_P

$$E : S_P \rightarrow S_C$$

We also introduce the decode mapping D

$$D : S_C \rightarrow S_P^D$$

which maps each chromosome onto the discretised parameter space S_P^D

Our technique embodies a primitive local search or quasi-gradient mechanism within a new composite genetic operator Ξ which we can write in terms of its ultimate effect as

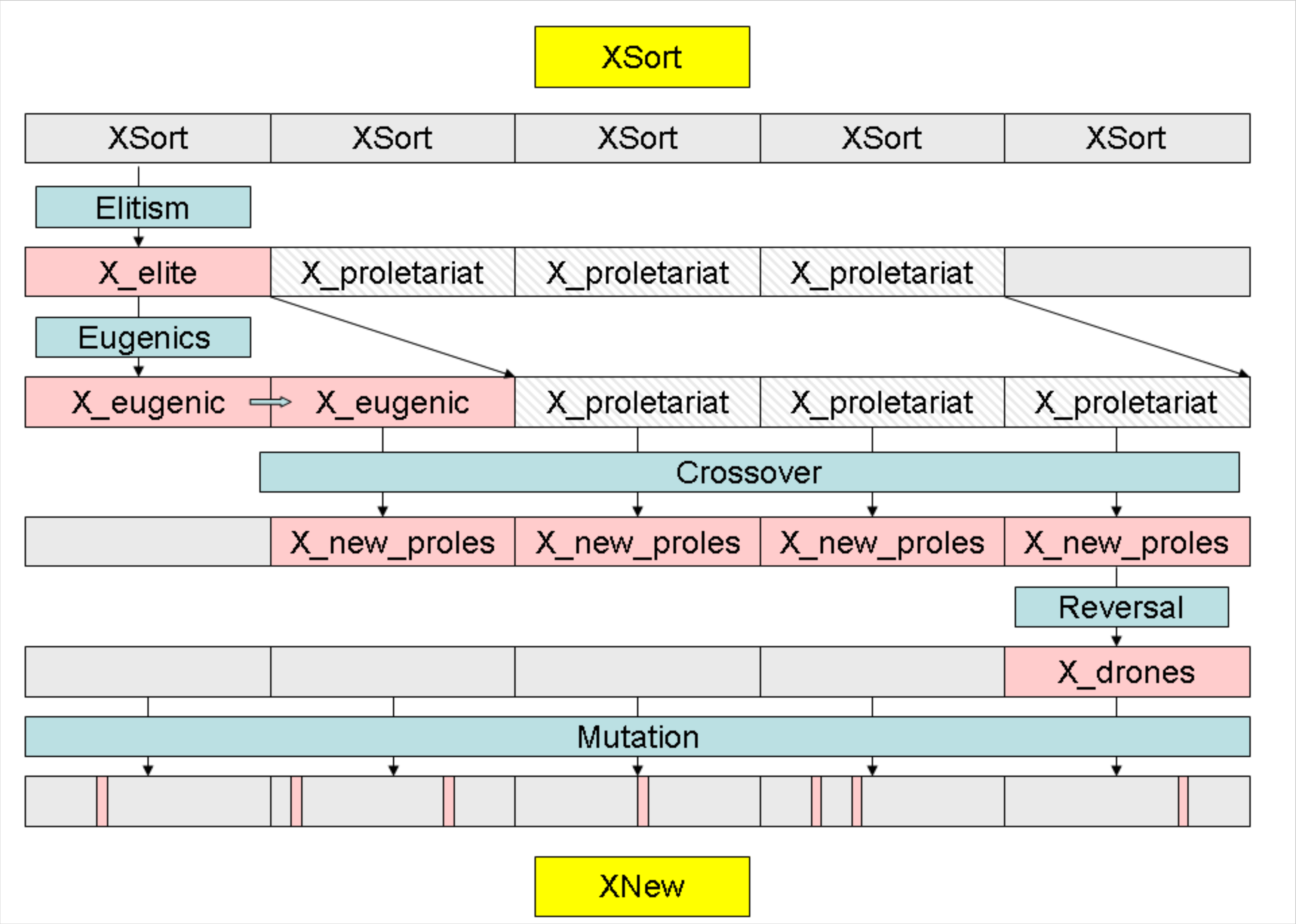
$$\Xi : S_C \rightarrow S_C$$

but which has a more complicated chain of domains, acting on both parameter space and chromosome space, and in a nonlocal sense defined by neighbourhoods in S_C and S_P^D

$$\Xi \quad E \Omega D N_C : S_C \rightarrow \left\{ S_C \right\} \rightarrow \left\{ S_P^D \right\} \rightarrow S_C$$

where N_C defines a neighbourhood of the target chromosome and Ω is the neighbourhood extremum operator in S_P^D

Chromosome selection and transformation including the eugenics operator



Efficiency of the eugenics operator

- modest computational load
- predictable computational load
- fast identification of neighbourhood in S_C
- no need to propose step length or compute Hessian
- avoids need to encode updated solution
- synchronous parallel processing

But how to test it and see if our expectations are met ?

Answer : synthetic 'designer' problems

A connection between symmetry, conservation laws and extrema in physical systems

PROCEEDINGS OF THE ROYAL SOCIETY A

Proc. R. Soc. A (2011) **467**, 3206–3221
doi:10.1098/rspa.2011.0158
Published online 22 June 2011

Extensions of Noether's Second Theorem: from continuous to discrete systems

By Peter E. Hydon[1,*] and Elizabeth L. Mansfield[2]

[1] *Department of Mathematics, University of Surrey, Guildford GU2 7XH, UK*
[2] *SMSAS, University of Kent, Canterbury, Kent CT2 7NF, UK*

A simple local proof of Noether's Second Theorem is given. This proof immediately leads to a generalization of the theorem, yielding conservation laws and/or explicit relationships between the Euler–Lagrange equations of any variational problem whose symmetries depend on a set of free or partly constrained functions. Our approach extends further to deal with finite-difference systems. The results are easy to apply; several well-known continuous and discrete systems are used as illustrations.

Keywords: conservation laws; gauge symmetries; difference equations

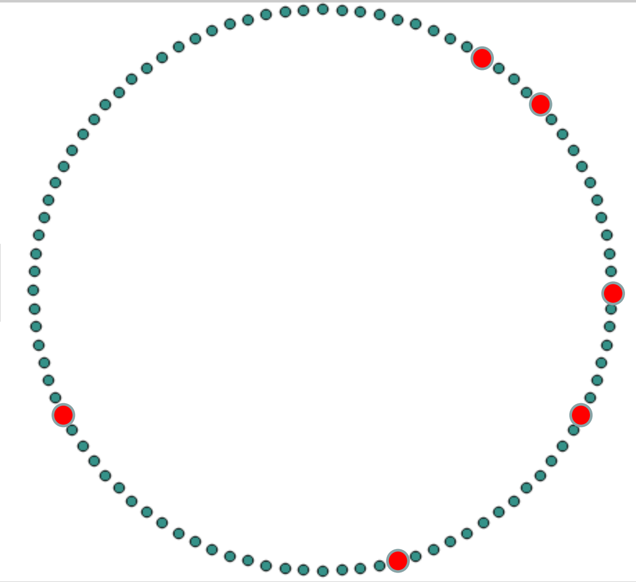
Complex problems with known exact solutions II

In one approach, we consider the spatial configurations of a number of interacting particles confined to a the surface of an n-sphere. For instance, for n = 2 we may take the case of identically-charged particles - electrons, say - confined to a circular hoop.

It can be shown that the state of minimum energy corresponds to the most symmetric state, namely, the electrons equally spaced around the circle. In three dimensions the electrons will reside on the surface of a sphere, and here the problem is a little more interesting as the states with maximum symmetry correspond to the vertices of the Platonic solids.

Typically we employ a generalized Riesz potential, with the objective function taking the form

$$V = \sum_{i \neq j} \sum_{j} \left| r_i - r_j \right|^{-s}$$



Optimal s-energy codes on S^2

Known optimal s-energy codes on $\mathbb{S}^2$

- $s = \log$, Smale's problem, logarithmic points (known for $N = 2 - 6,\ 12$);
- $s = 1$, Thomson Problem (known for $N = 2 - 6,\ 12$)
- $s = -1$, Fejes-Toth Problem (known for $N = 2 - 6,\ 12$)
- $s \to \infty$, Tammes Problem (known for $N = 1 - 12,\ 13, 14,\ 24$)

Limiting case - Best packing

For fixed N, any limit as $s \to \infty$ of optimal s-energy codes is an optimal (maximal) code.

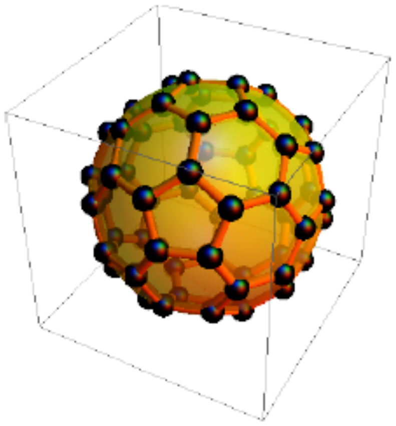
Universally optimal codes

The codes with cardinality $N = 2, 3, 4, 6, 12$ are special (*sharp codes*) and minimize large class of potential energies. First "non-sharp" is $N = 5$ and very little is rigorously proven.

Generalised Thomson problem : Electrostatics in higher dimensions

Thomson Problem (1904) -
("plum pudding" model of an atom)

Find the (most) stable (ground state) energy configuration (**code**) of N classical electrons (Coulomb law) constrained to move on the sphere $\mathbb{S}^2$.



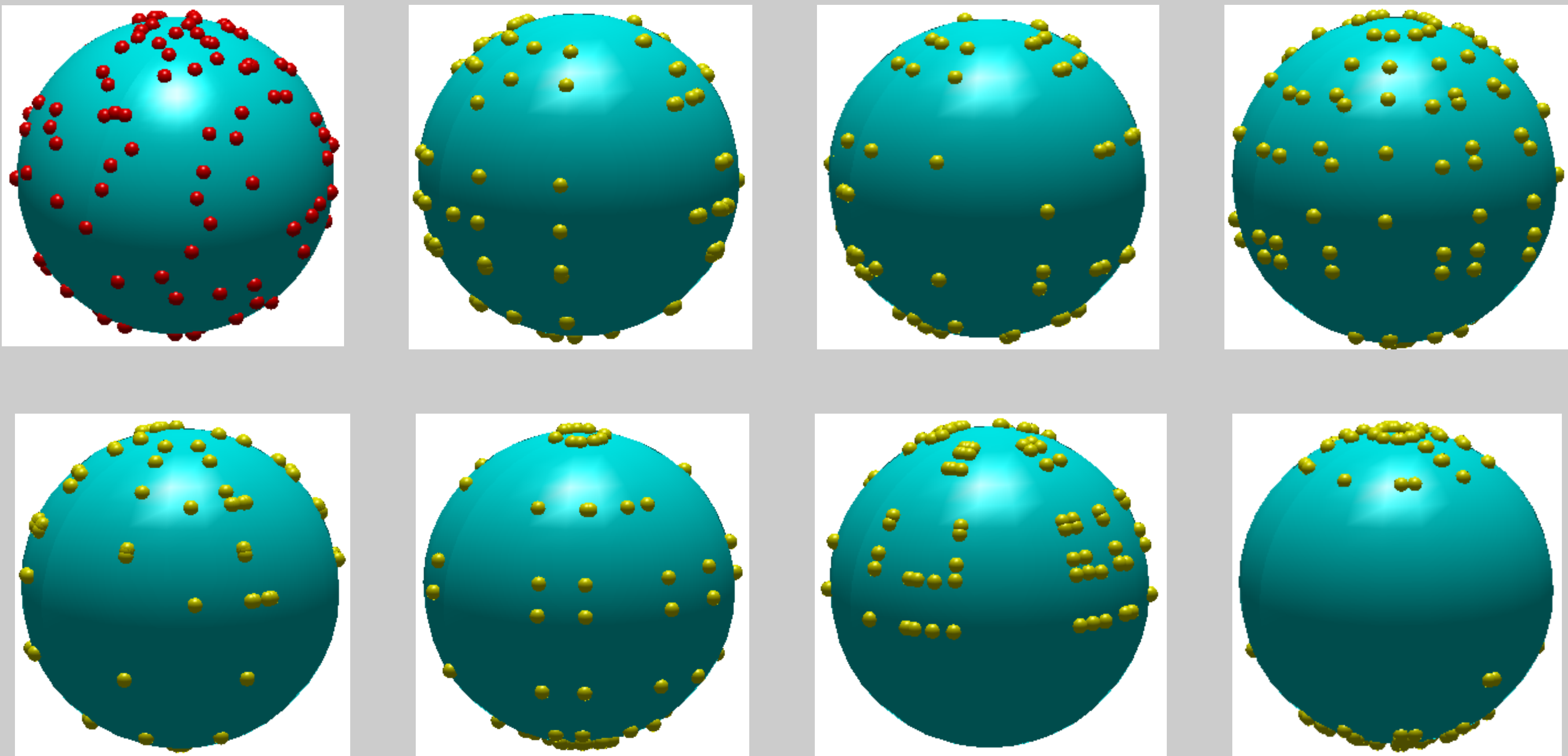
Generalized Thomson Problem ($1/r^s$ **potentials and** $\log(1/r)$)

A code $C := \{\mathbf{x}_1, \ldots, \mathbf{x}_N\} \subset \mathbb{S}^{n-1}$ that minimizes **Riesz *s*-energy**

$$E_s(C) := \sum_{j \neq k} \frac{1}{|\mathbf{x}_j - \mathbf{x}_k|^s}, \quad s > 0, \qquad E_{\log}(\omega_N) := \sum_{j \neq k} \log \frac{1}{|\mathbf{x}_j - \mathbf{x}_k|}$$

is called an **optimal *s*-energy code**.

Intermediate states ranked by Riesz energy (eg electrons on a sphere, interacting via Coulomb potential)



Groups, subgroups and Lagrange's Theorem I

Definition:

An operation on a set G is a function $* : G \times G \to G$.

Definition:

A group is a set G which is equipped with an operation $*$ and a special element $e \in G$, called the identity, such that

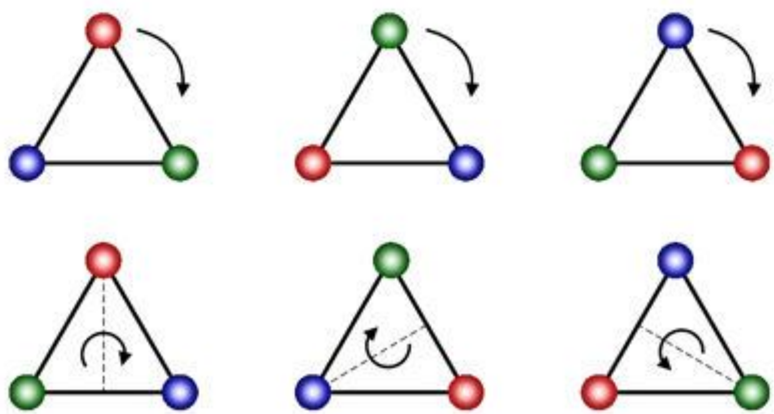
(i) the associative law holds: for every $x, y, z \in G$ we have $x * (y * z) = (x * y) * z$;
(ii) $e * x = x = x * e$ for all $x \in G$;
(iii) for every $x \in G$, there is $x' \in G$ (so-called, inverse) with $x * x' = e = x' * x$.

Definition:

A subset H of a group G is a subgroup if

(i) $e \in H$;
(ii) if $x, y \in H$, then $x * y \in H$;
(iii) if $x \in H$, then $x^{-1} \in H$.

D_3



$a^3 = e$

$b^2 = e$

Definition:

If G is a group and $a \in G$, write

$$\langle a \rangle = \{a^n : n \in \mathbb{Z}\} = \{\text{all powers of } a\};$$

$\langle a \rangle$ is called the cyclic subgroup of G generated by a.

Groups, subgroups and Lagrange's Theorem II

Definition:

A group G is called cyclic if $G = \langle a \rangle$ for some $a \in G$. In this case a is called a generator of G.

Definition:

Let G be a group and let $a \in G$. If $a^k = 1$ for some $k \geq 1$, then the smallest such exponent $k \geq 1$ is called the order of a; if no such power exists, then one says that a has infinite order.

Definition:

If G is a finite group, then the number of elements in G, denoted by $|G|$, is called the order of G.

Theorem:

Let G be a finite group and let $a \in G$. Then

$$\text{order of } a = |\langle a \rangle|.$$

Fermat's Little Theorem:

Let p be a prime. Then $n^p \equiv n \mod p$ for any integer $n \geq 1$.

Groups, subgroups and Lagrange's Theorem III

Lagrange's Theorem: If H is a subgroup of G, then $|G| = n|H|$ for some positive integer n. This is called the *index* of H in G. Furthermore, there exist $g_1, \ldots, g_n$ such that $G = Hr_1 \cup \ldots \cup Hr_n$ and similarly with the left-hand cosets relative to H.

Proof: Take any $r_1 \in G$. Note $|Hr_1| = |H|$. If $Hr_1 \neq G$ then take any $r_2 \in G \setminus Hr_1$. By the lemma, Hr_1, Hr_2 are disjoint so we have $|Hr_1 \cup Hr_2| = 2|H|$. By continuing in this fashion, after n steps for some positive integer n, we will eventually have accounted for all of the elements of G. We will have $|G| = n|H|$ and $G = Hr_1 \cup \ldots \cup Hr_n$.

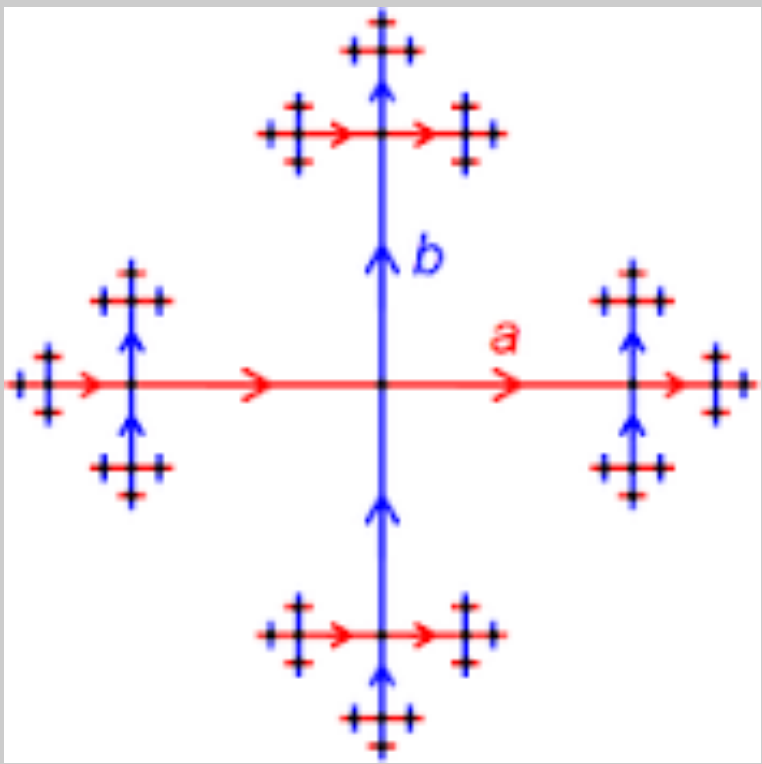
Corollary: Let G be a group and $g \in G$. Then the order of g divides $|G|$.

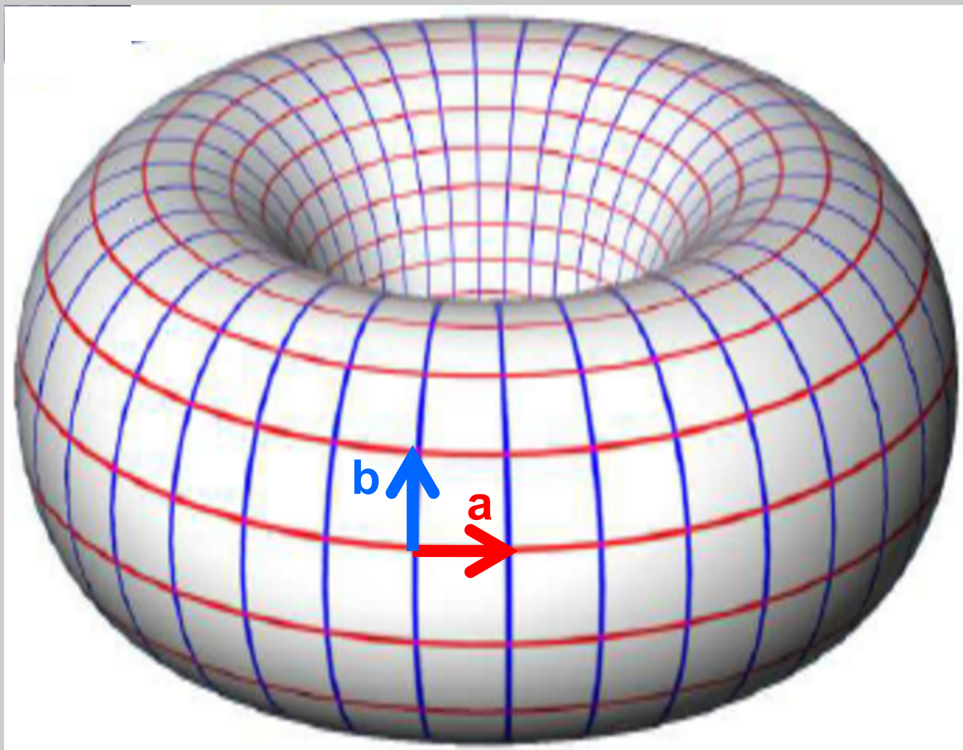
Corollary: Let G be a group of prime order. Then G has no subgroups and hence is cyclic

Presentations of two-generator groups

Cayley graph of X = { a,b | }, the free group of rank 2.

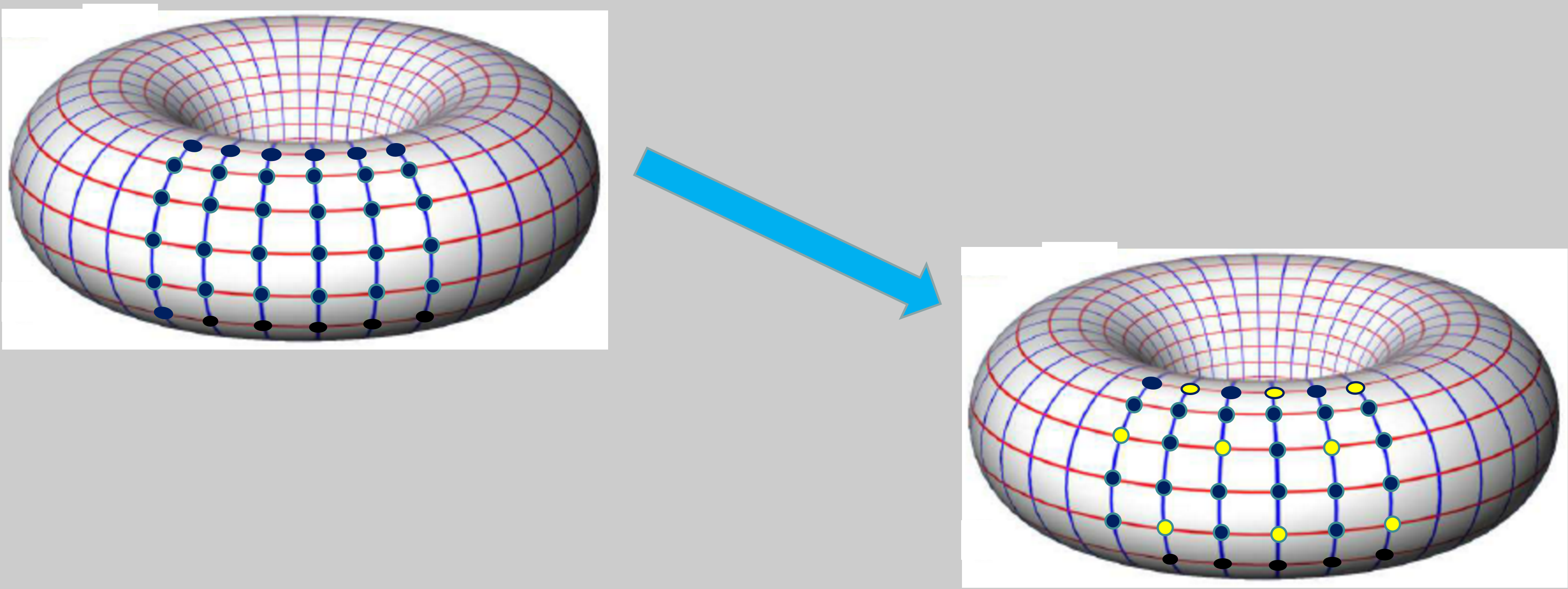
A realisation of X = { a,b | a^M, b^N }, the cyclic group





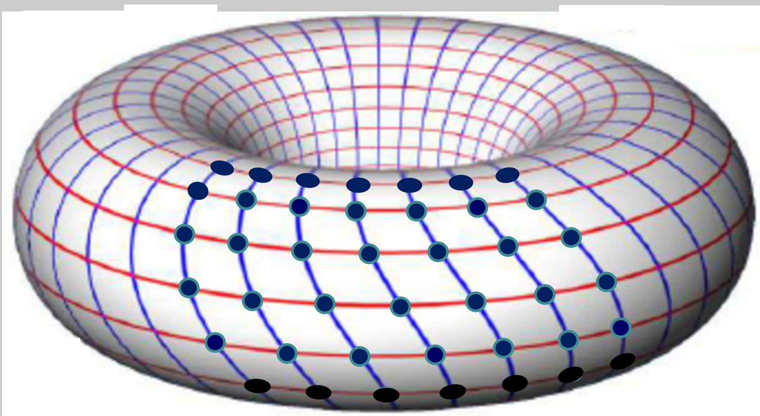
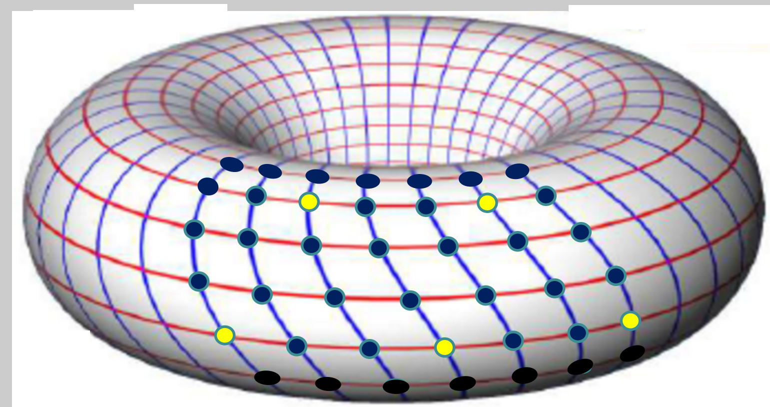
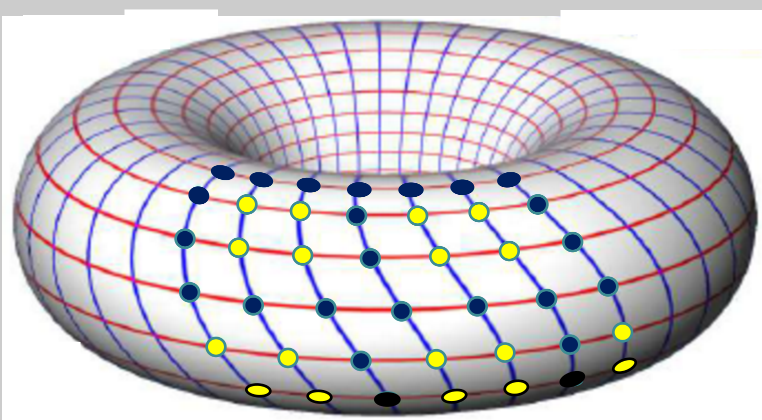
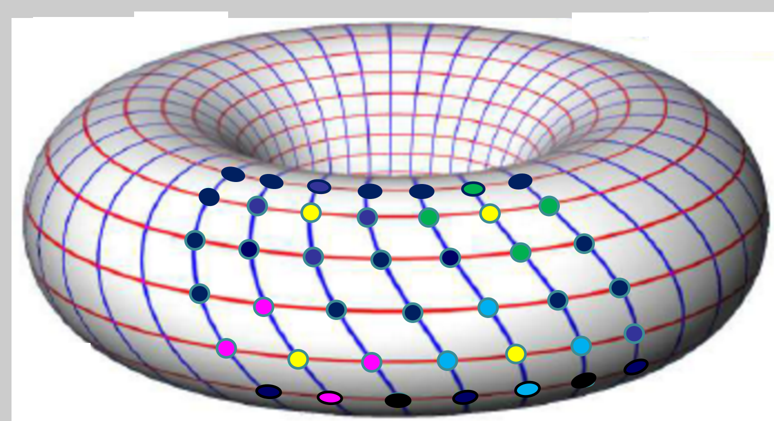
Realisations of the 2-generator cyclic group and a representative subgroup

$$X = \{a, b \mid a^M, b^N\} \qquad \{a, b\} = \left\{ exp\left[i2\pi \frac{1}{M}\right]\hat{\varphi}, exp\left[i2\pi \frac{1}{N}\right]\hat{\theta} \right\}$$

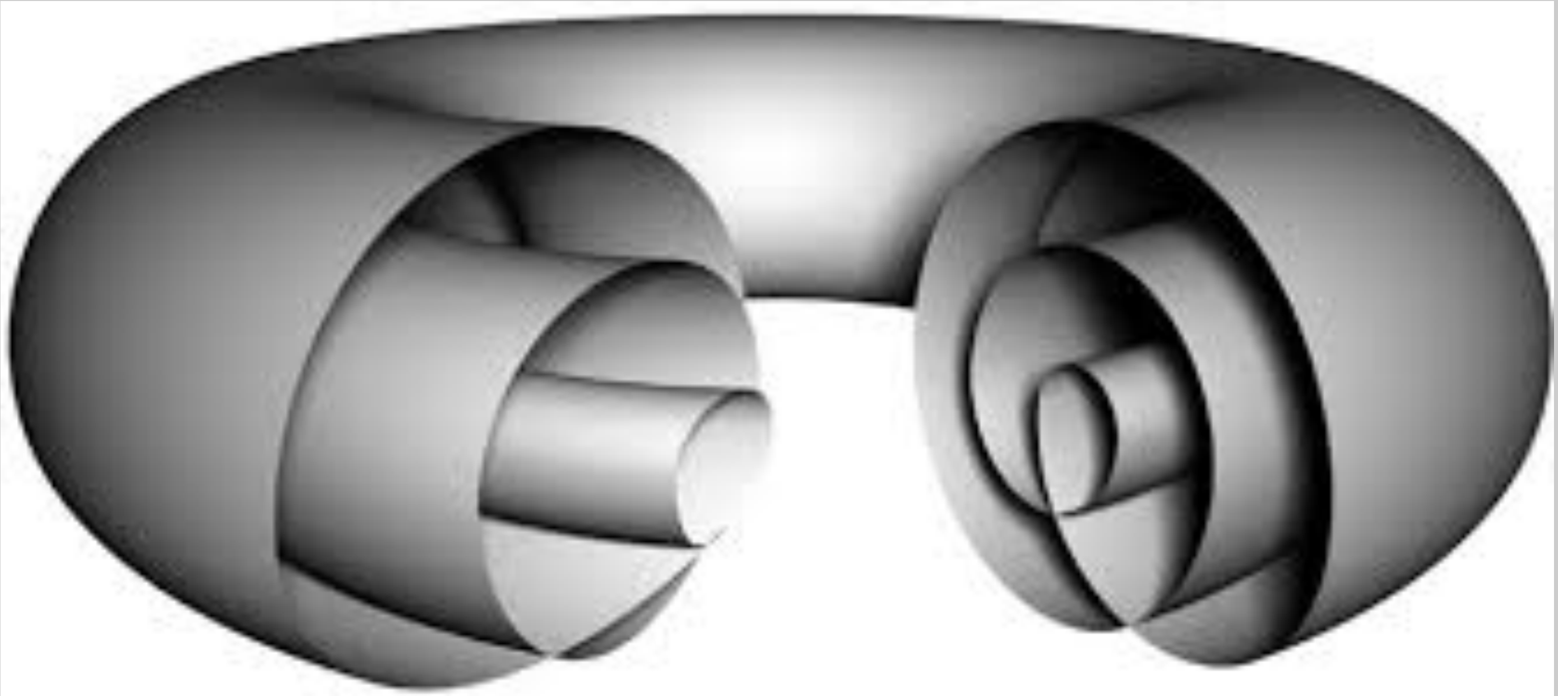


$$X = \left\{ exp\left[i2\pi \frac{j}{M}\right]\hat{\varphi}, exp\left[i2\pi \frac{k}{N}\right]\hat{\theta} \,\middle|\, j = 1, M; k = 1, N \right\}$$

Variations of the torus realisation of the 2-generator cyclic subgroup and nested subgroups

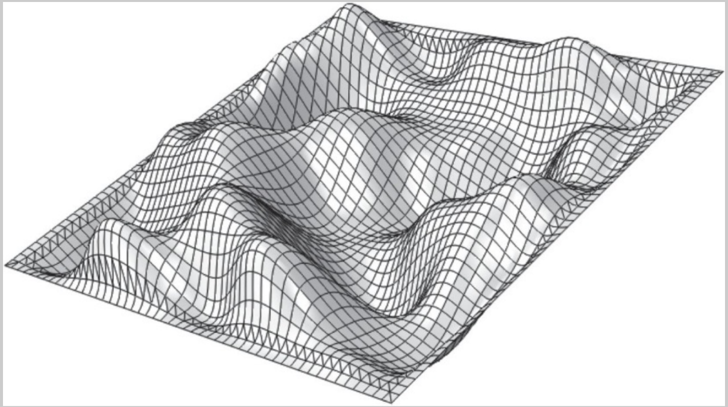
Nested toroid realization of a three-generator cyclic group

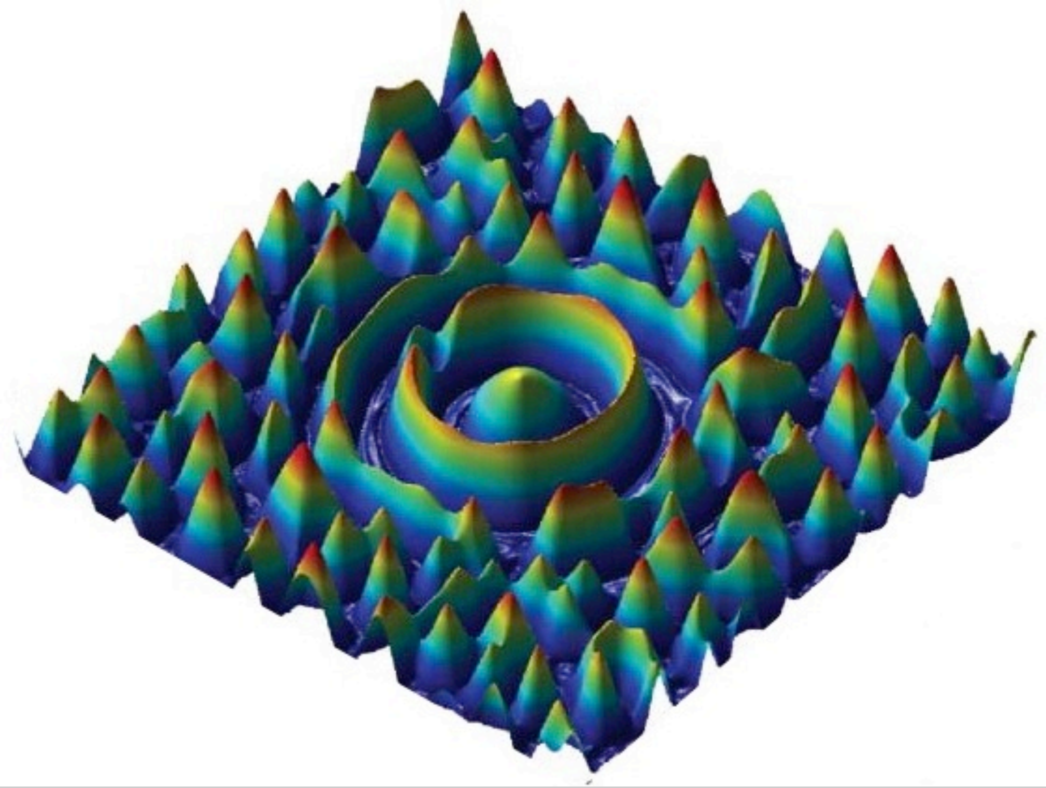


(continuous case illustrated)

Lie groups

Fitness function landscapes : How can we design these by choice of subgroup ?

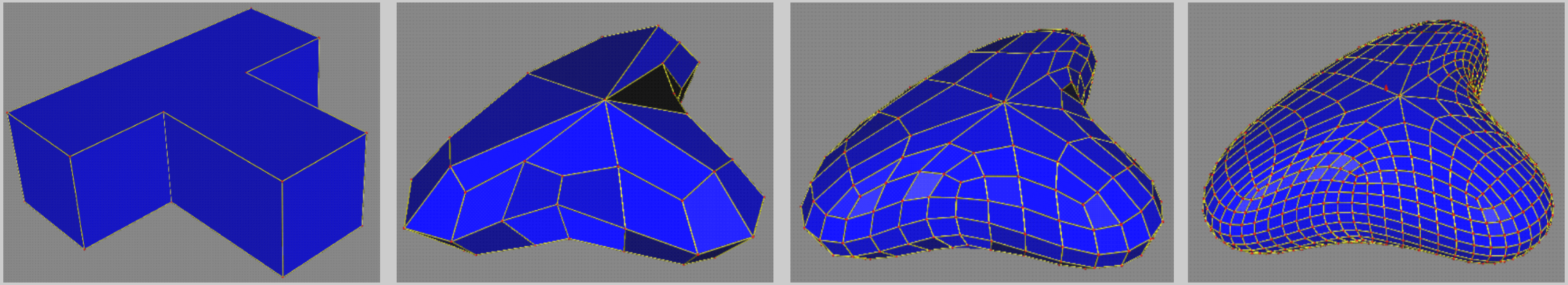




Present method (brute force) :

Construct a 3-dimensional slice display in Matlab and sweep it through the higher-dimensional space along a pseudo-random trajectory

Catmull-Clark surface subdivision : Preserves symmetries

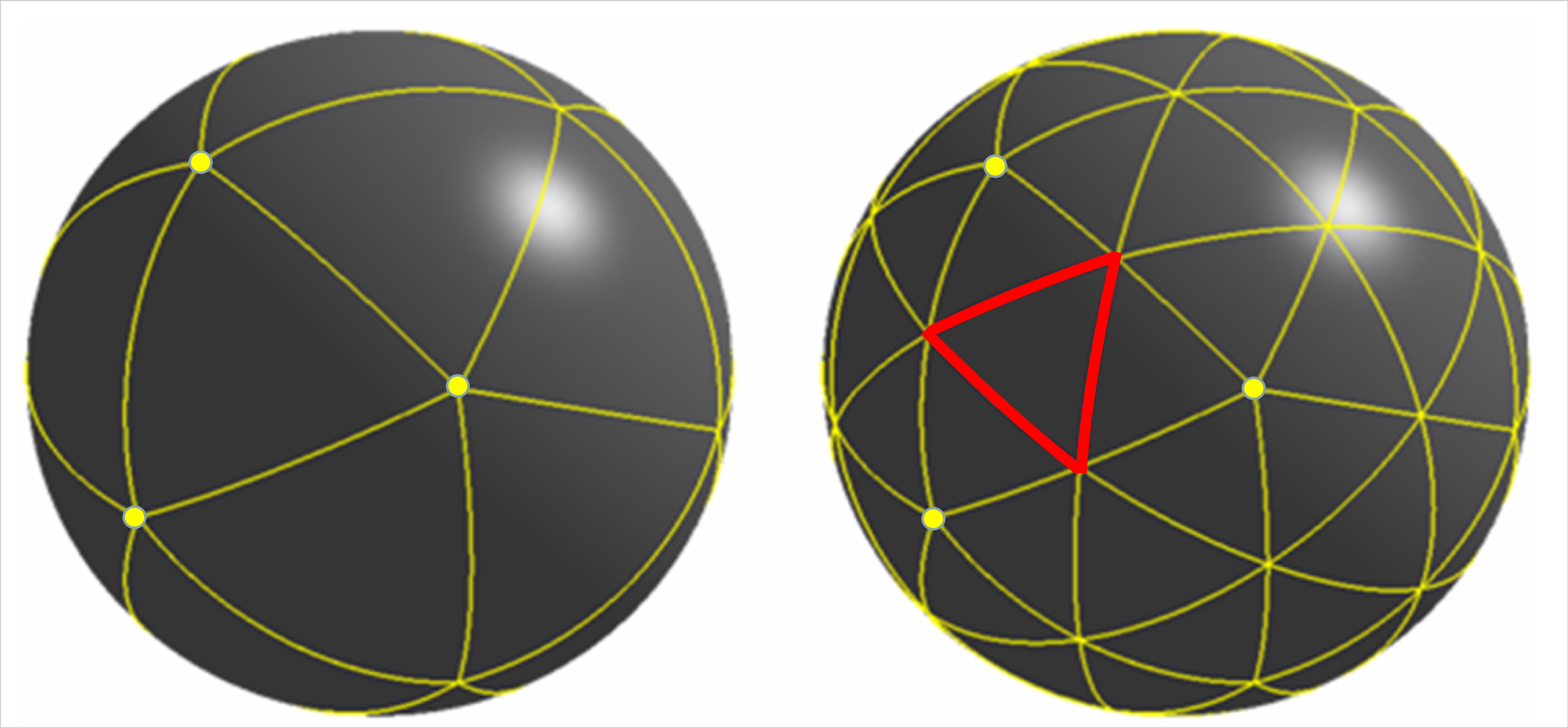


- For each face of the mesh, generate the new face points – which are the average of all the original points defining the face (We note that faces may have 3, 4, 5, or many points now defining them).
- Generate the new edge points – which are calculated as the average of the midpoints of the original edge with the two new face points of the faces adjacent to the edge.
- Calculate the new vertex points – which are calculated as the average of Q, 2R and (n−3)S n , where Q is the average of the new face points of all faces adjacent to the original face point, R is the average of the midpoints of all original edges incident on the original vertex point, and S is the original vertex point.

The mesh is reconnected by the following method.

- Each new face point is connected to the new edge points of the edges defining the original face.
- Each new vertex point is connected to the new edge points of all original edges incident on the original vertex point.

A counter-intuitive problem with refinement : non-uniform Riesz *s*-potentials



Circumscribed spheres about a regular dodecahedron (left) and a refinement of the dodecahedron when midpoints of the arcs connecting adjacent vertices are added (right); both have a high degree of symmetry but only the Platonic dodecahedron has identical potentials at every vertex for all homogeneous scalar interactions

A number theory approach : Fermat's Little Theorem

Fermat's Little Theorem:

Let p be a prime. Then $n^p \equiv n \mod p$ for any integer $n \geq 1$.

By Lagrange's Theorem we get

$$|\langle [a] \rangle| \text{ divides } |\mathbb{Z}_p^\times|,$$

which gives

$$k \mid p-1,$$

since $|\langle [a] \rangle| = k$ and $|\mathbb{Z}_p^\times| = p-1$. So

$$p-1 = kd$$

for some integer d. On the other hand, since k is the order of $[a]$, it follows that for any $n \in [a]$ we have

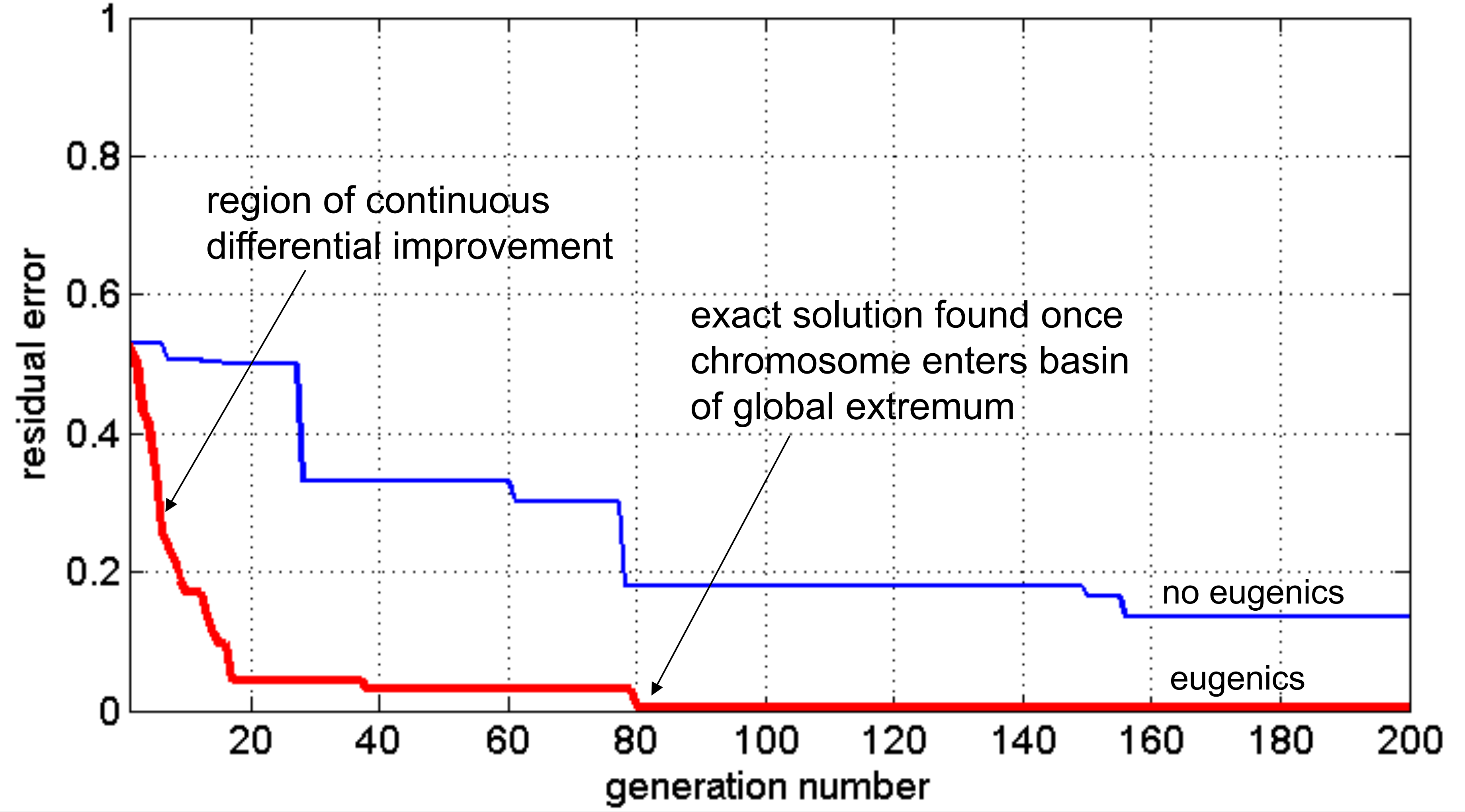
$$n^k \equiv 1 \mod p,$$

hence

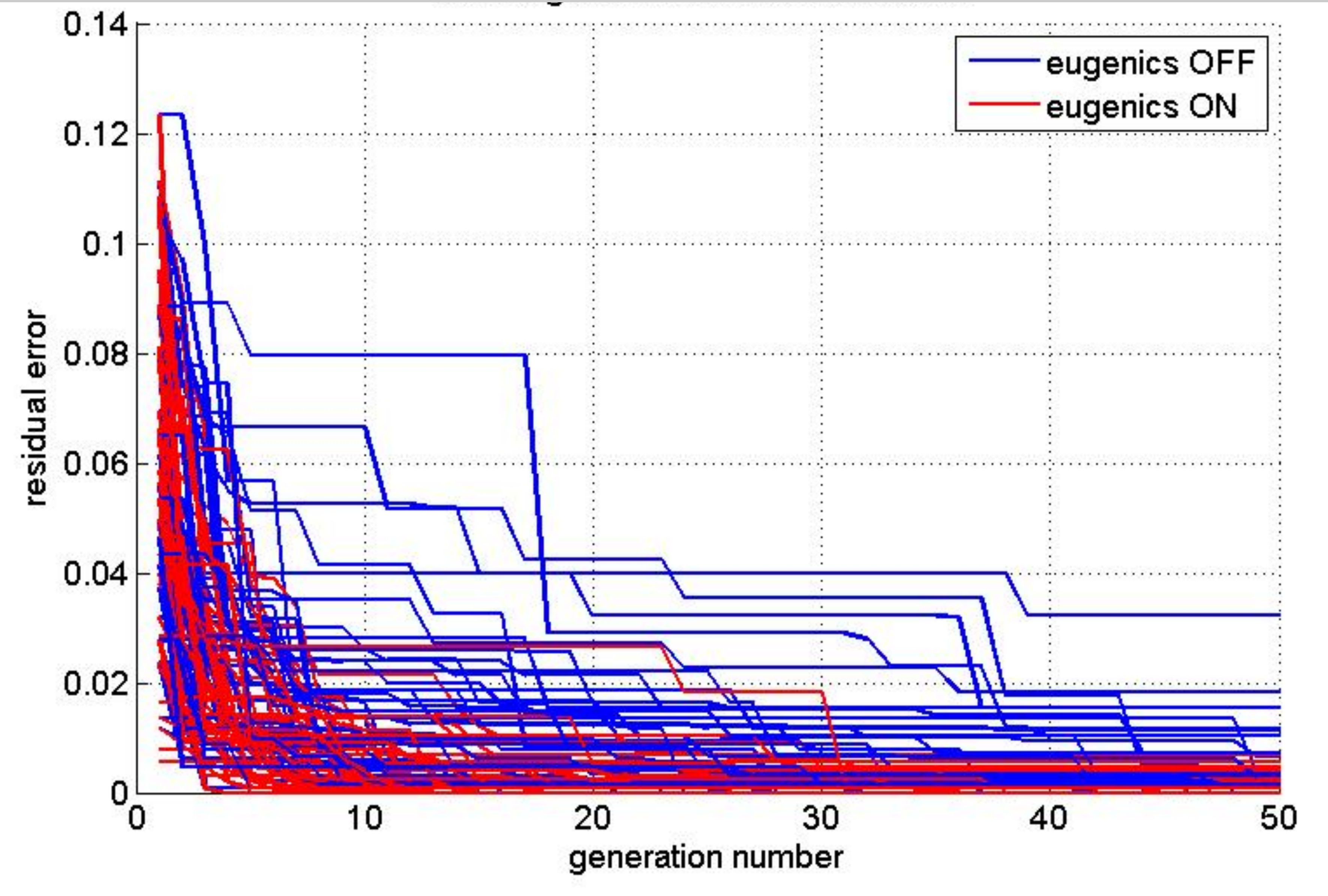
$$n^{kd} \equiv 1^d \equiv 1 \mod p,$$

and the result follows, since $kd = p-1$. ■

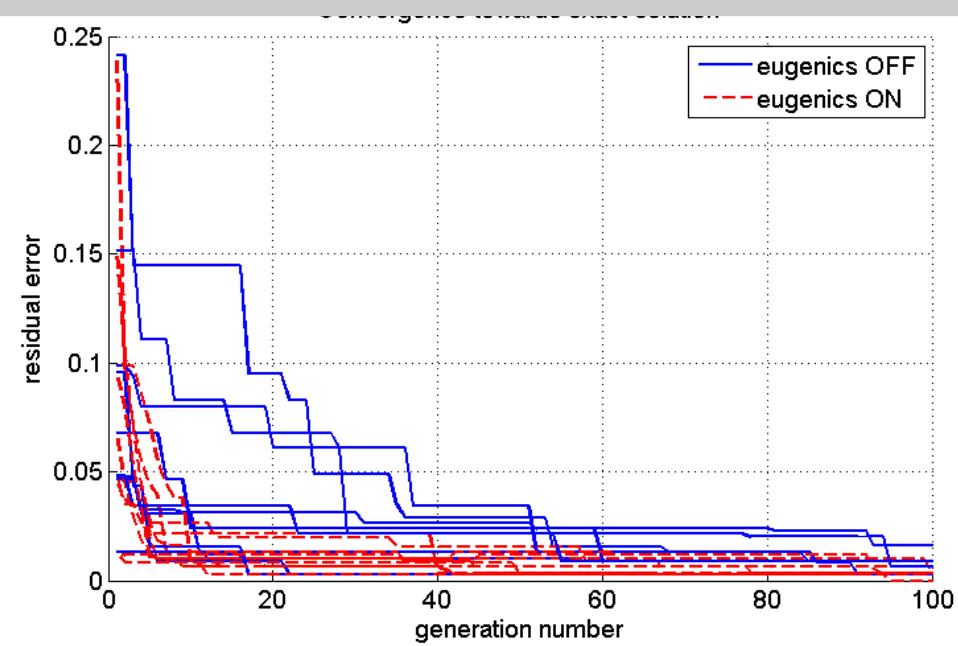
An example illustrating the behaviour we wish to achieve in real-time applications

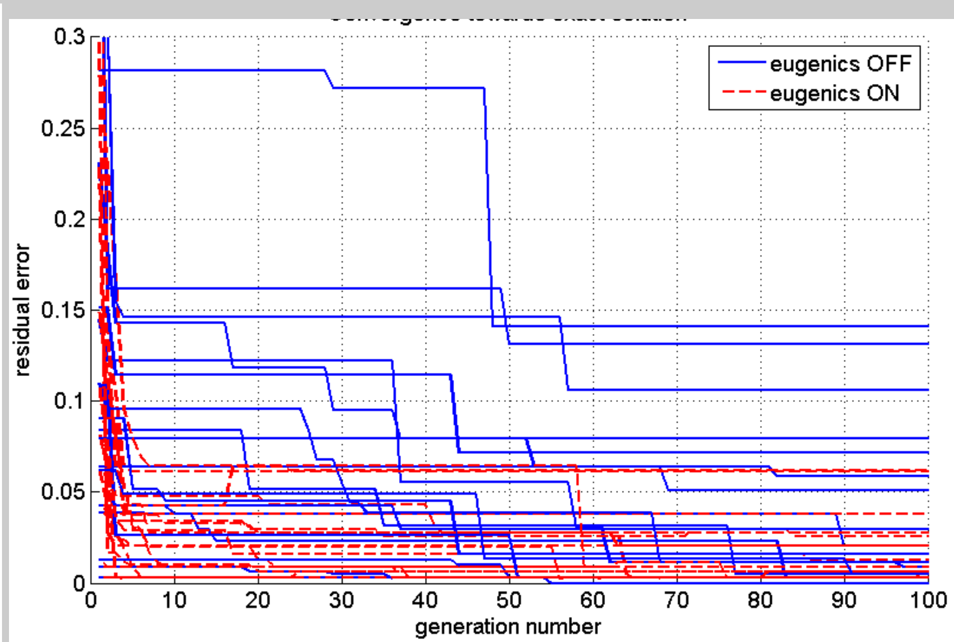


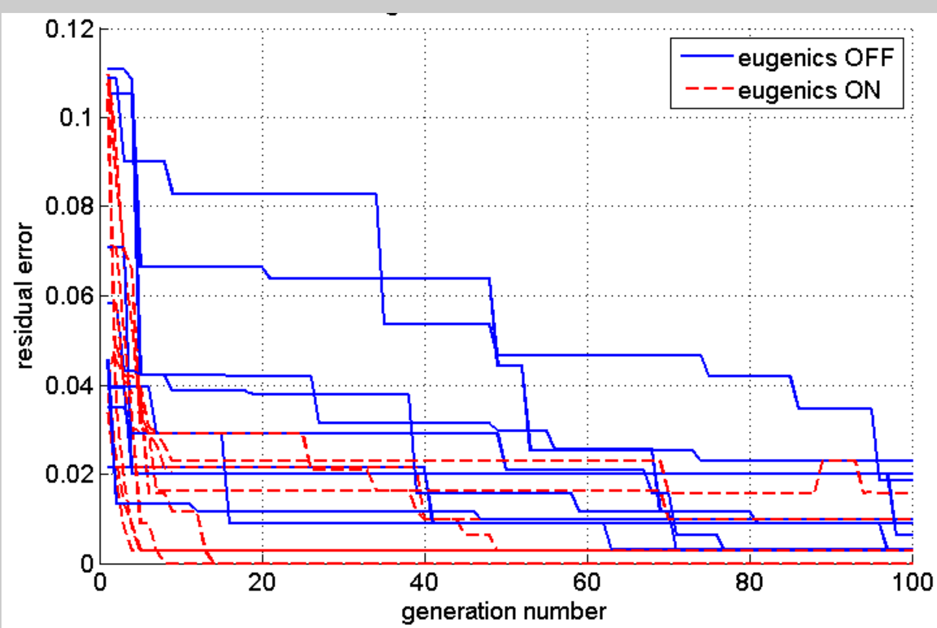
Overlay showing the improved convergence achieved using eugenics

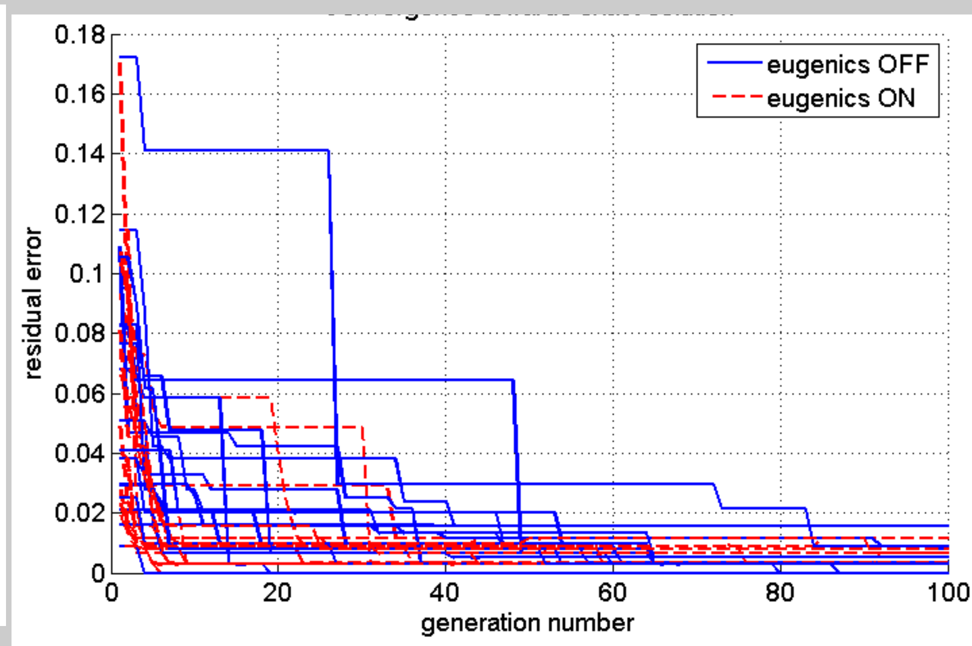


Examples illustrating the merits and limitations of eugenics

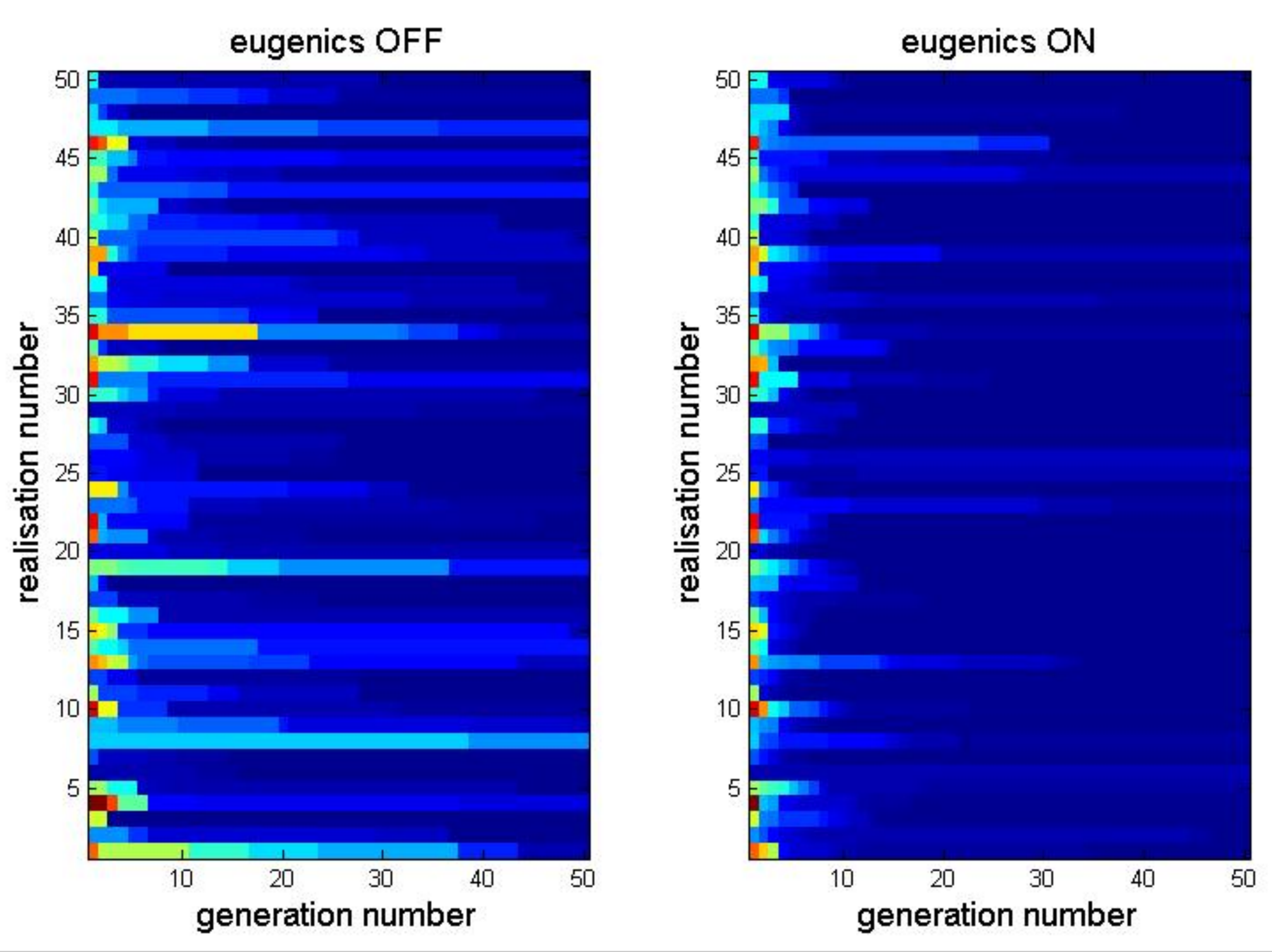




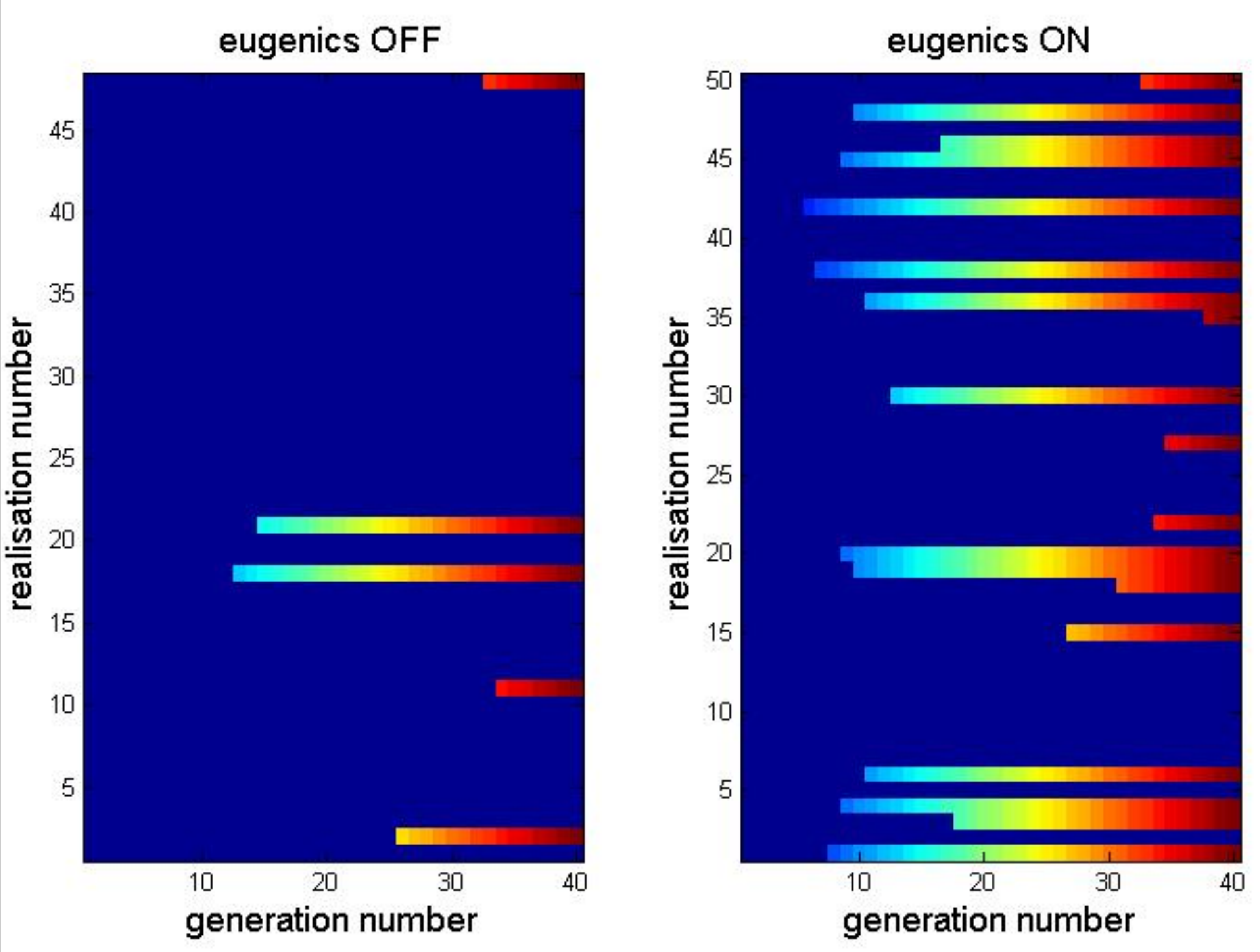




Residual error wrt global extremum with and without eugenics



Time to locate global extremum with and without eugenics



Conclusions

- Nonlinear optimisation (NLO) algorithms play a central role in the design and operation of complex systems such as OTH radar.

- The choice of optimisation algorithm for a given application needs to be based on a good understanding of its performance, including probability of attaining the global extremum, convergence rate, robustness, ability to handle multi-objective problems and so on.

- The use of synthetic problems whose global extrema are precisely known while possessing complex fitness landscapes is a useful way of measuring some aspects of performance.

- We have described and demonstrated this methodology, initially in the form of a simple physical model, then outlining how a group-theoretical framework is appropriate for describing and constructing such synthetic problems.