Optimization and Design of Complex Systems

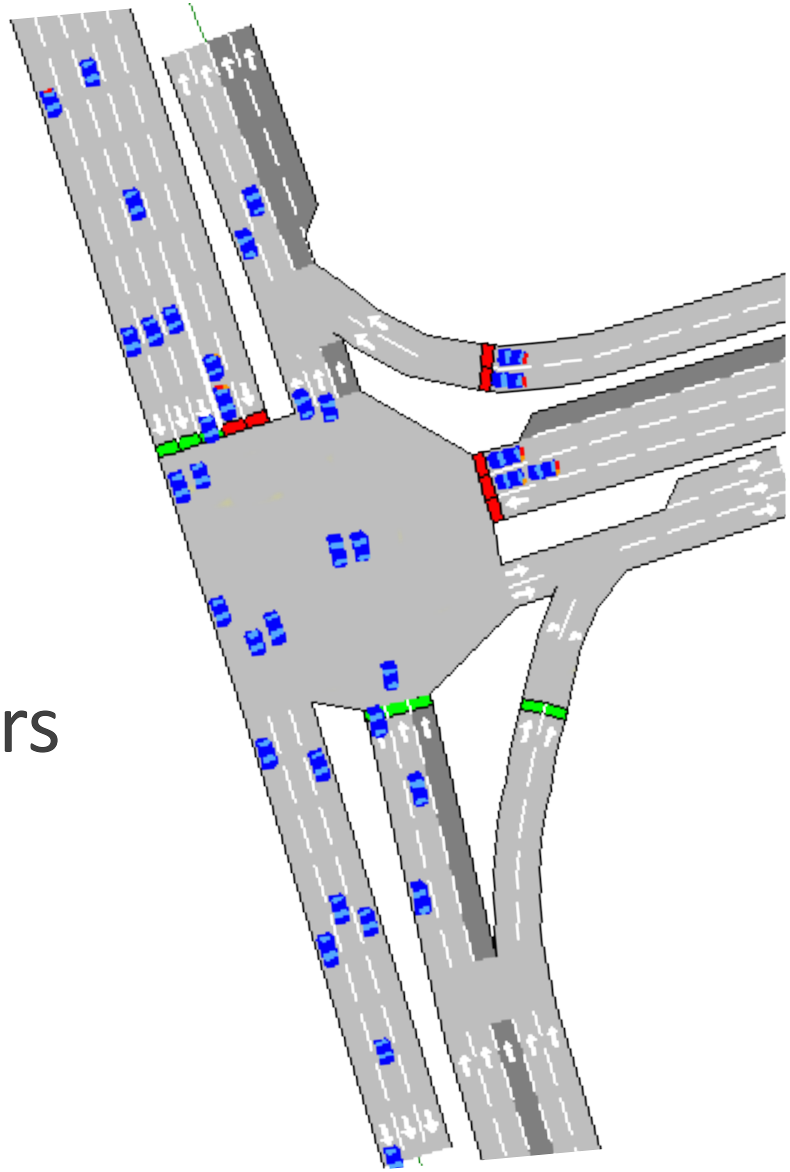
Dealing with uncertainty in simulation-based optimisation

Juergen Branke

We live in a complex world

- Large number of interacting elements
- Emergence
- Can not be understood by analysis of components
- Simulation can capture emergent phenomena

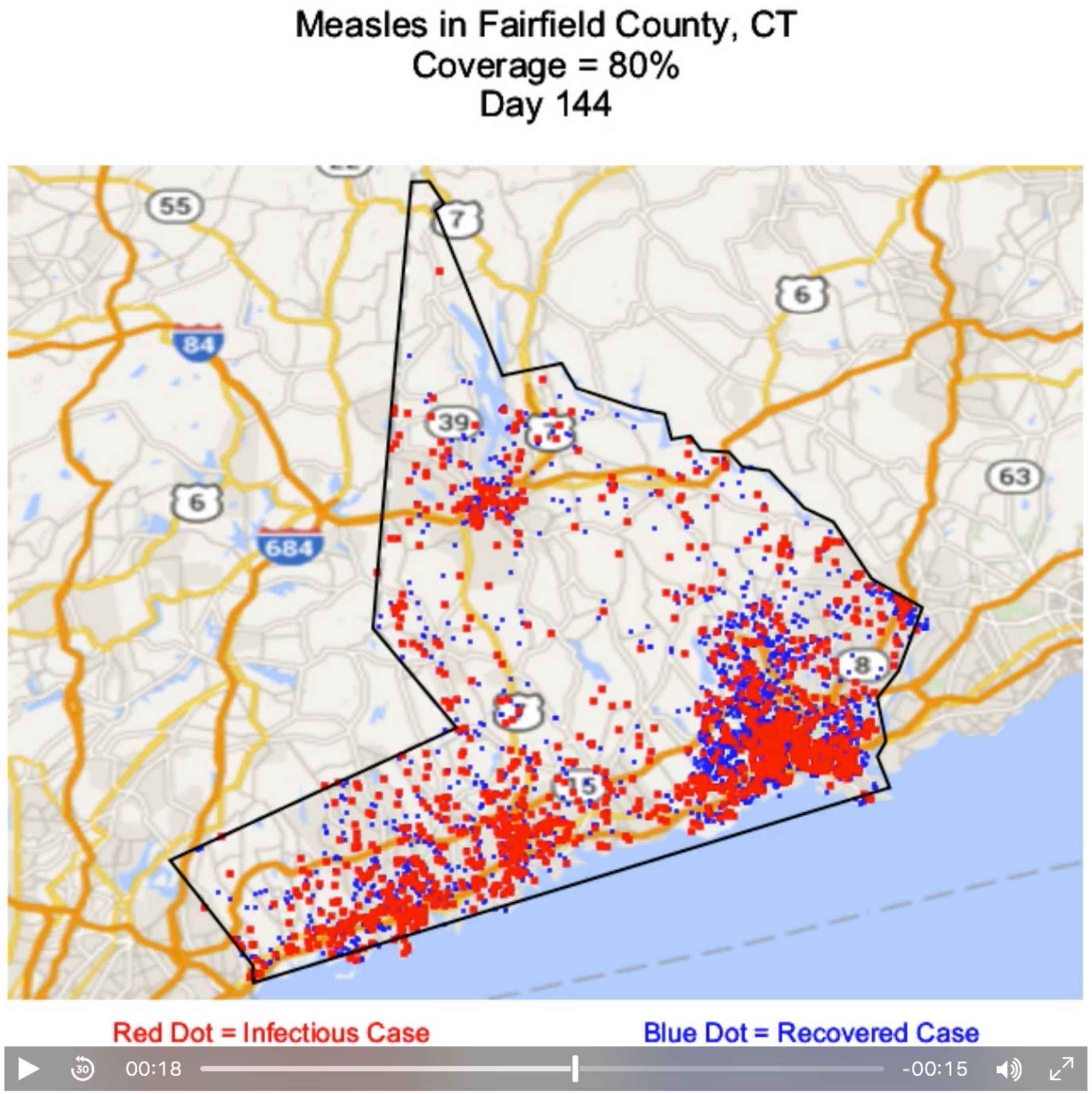
Example: Traffic



- Street networks
- Reactive traffic light controllers

Example: Healthcare

Understanding how diseases spread



Measles in Fairfield County, CT
Coverage = 80%
Day 144

Red Dot = Infectious Case
Blue Dot = Recovered Case

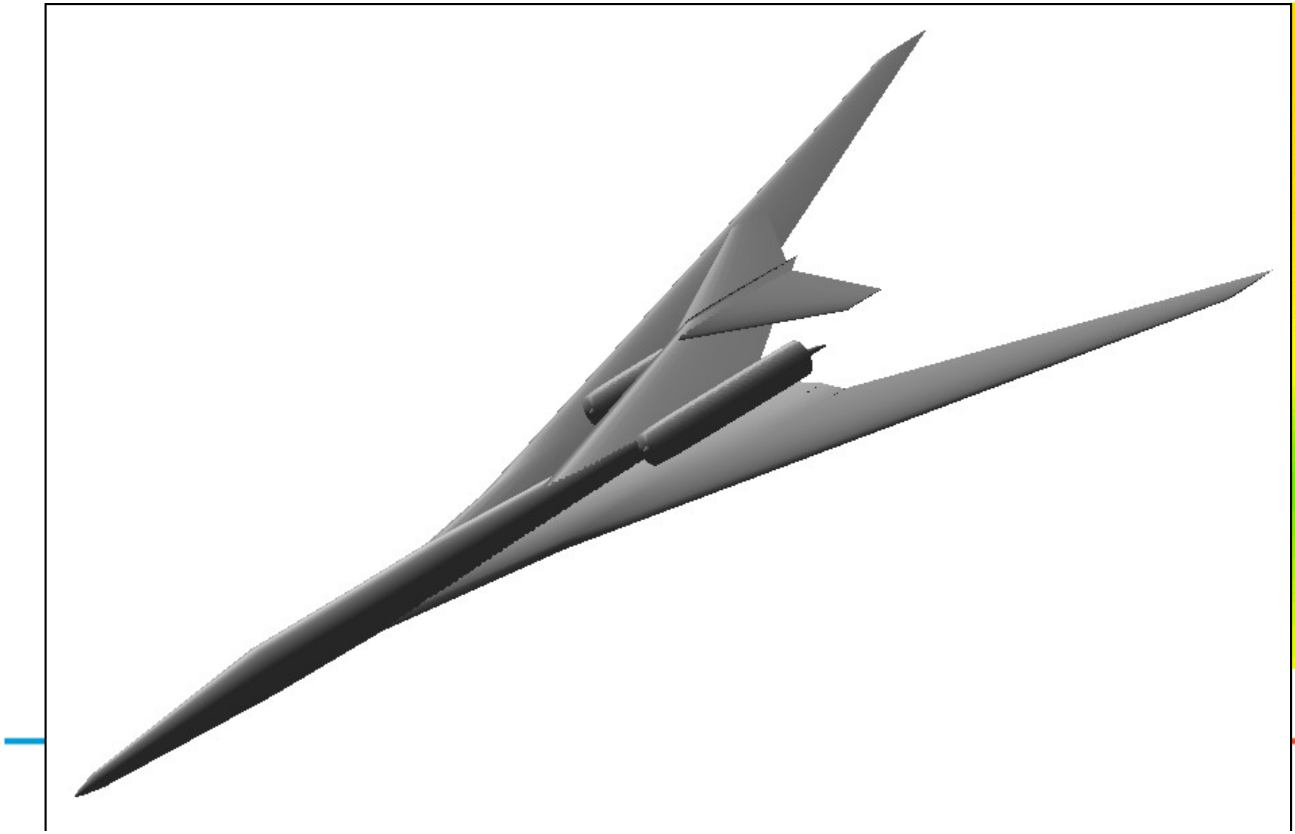
Example: Manufacturing

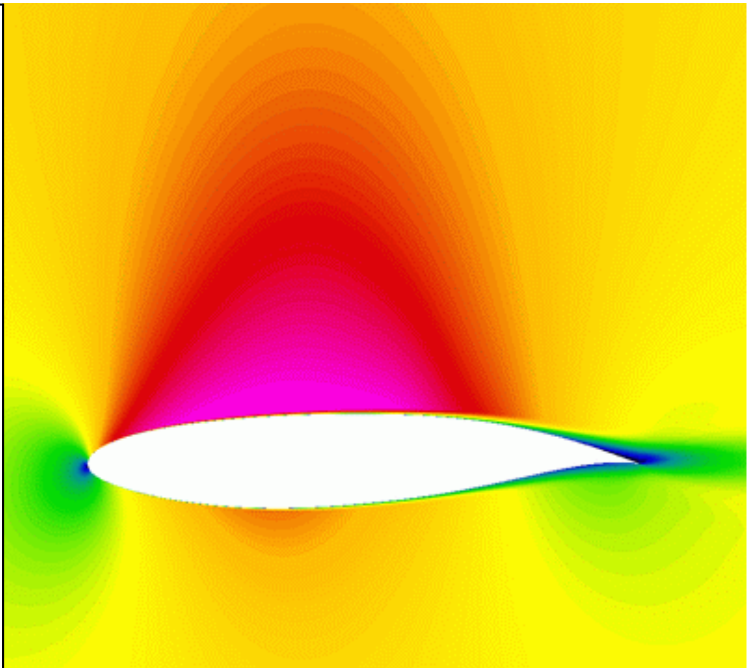
Simulate machine breakdowns, stochastic processing times, complex scheduling rules,

etc.

Example: Engineering

Simulation can replace physical testing





The next step:
Simulation optimisation

- Automatically search vast spaces of parameter settings to find “optimal” settings

Parameters → **Simulation** → Objective value

- Model calibration
- Automated design and optimisation of complex systems

Simulation optimisation examples

- Traffic: Optimise traffic light controller
- Healthcare: Identify optimal vaccination policies
- Manufacturing: Find optimal dispatching rules
- Engineering: Find optimal wing design

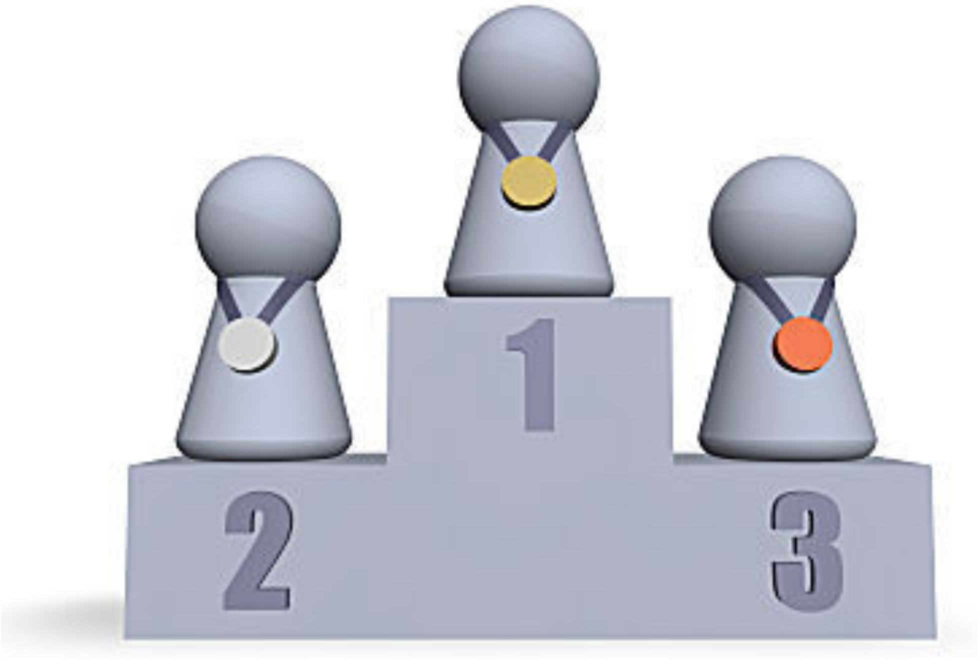
Challenges

- Simulations are mostly black boxes
- Simulations are computationally expensive
- There are often multiple criteria
- Simulations are often stochastic

Outline

- Ranking and Selection
- Black box optimisation
- Optimisation under Noise
- Related topics

Selecting the Best System



Ranking and selection problem

- Select, out of *k* systems, the one with best mean performance
- Let X_{ij} be output of *j*th replication of *i*th system

$$\{X_{ij} : j = 1, 2, \ldots\} \overset{i.i.d.}{\sim} \mathrm{Normal}(w_i, \sigma_i^2,) \quad i = 1, \ldots, k$$

- Sample statistics: $\bar{x}_i$ and $\hat{\sigma}_i^2$ based on n_i observations seen so far
- Order statistics: $\bar{x}_{(1)} \leq \bar{x}_{(2)} \leq \ldots \leq \bar{x}_{(k)}$
- Correct selection if selected system (*k*) is the true best system [*k*]

Standard: Equal allocation

- Sample each system *n* times
- Reduces standard error by $\frac{1}{\sqrt{n}}$

Comparison of m>2 alternatives

- Allocate samples sequentially
- Maximise the value of information

Myopic approach to maximize probability of correct selection

[Chick, Branke, Schmidt: J. of Computing, 2010]

- Assume we can take only one more sample
- If the sample doesn't change selected solution -> information had no value
- Expected value of information is probability of a change in the index of the individual with the best mean

Expected value of information (PCS)

Change of best system if

- system (i) ≠ (k) is evaluated and becomes new best system
- system (k) is evaluated and becomes worse than second best

$$\mathrm{EVI}_{(i)} = \begin{cases} \Phi_{n_{(i)}-1}\left(\dfrac{\bar{x}_{(i)} - \bar{x}_{(k)}}{\sqrt{\frac{\hat{\sigma}^2_{(i)}}{n_{(i)}(n_{(i)}+1)}}}\right) & \text{if } (i) \neq (k) \\ \Phi_{n_{(k)}-1}\left(\dfrac{\bar{x}_{(k-1)} - \bar{x}_{(k)}}{\sqrt{\frac{\hat{\sigma}^2_{(k)}}{n_{(k)}(n_{(k)}+1)}}}\right) & \text{if } (i) = (k) \end{cases}$$

Algorithm

```
Sample each alternative $n_0$ times
Determine sample statistics $\bar{x}_i$ and $\sigma_i^2$ and order
        statistics $\bar{x}_{(1)} \leq \ldots \leq \bar{x}_{(k)}$
WHILE stopping criterion not reached DO
        Take additional sample of system $i$ with
                maximal EVI
        Update sample and order statistics
Pick solution with maximal $\bar{x}_i$
```

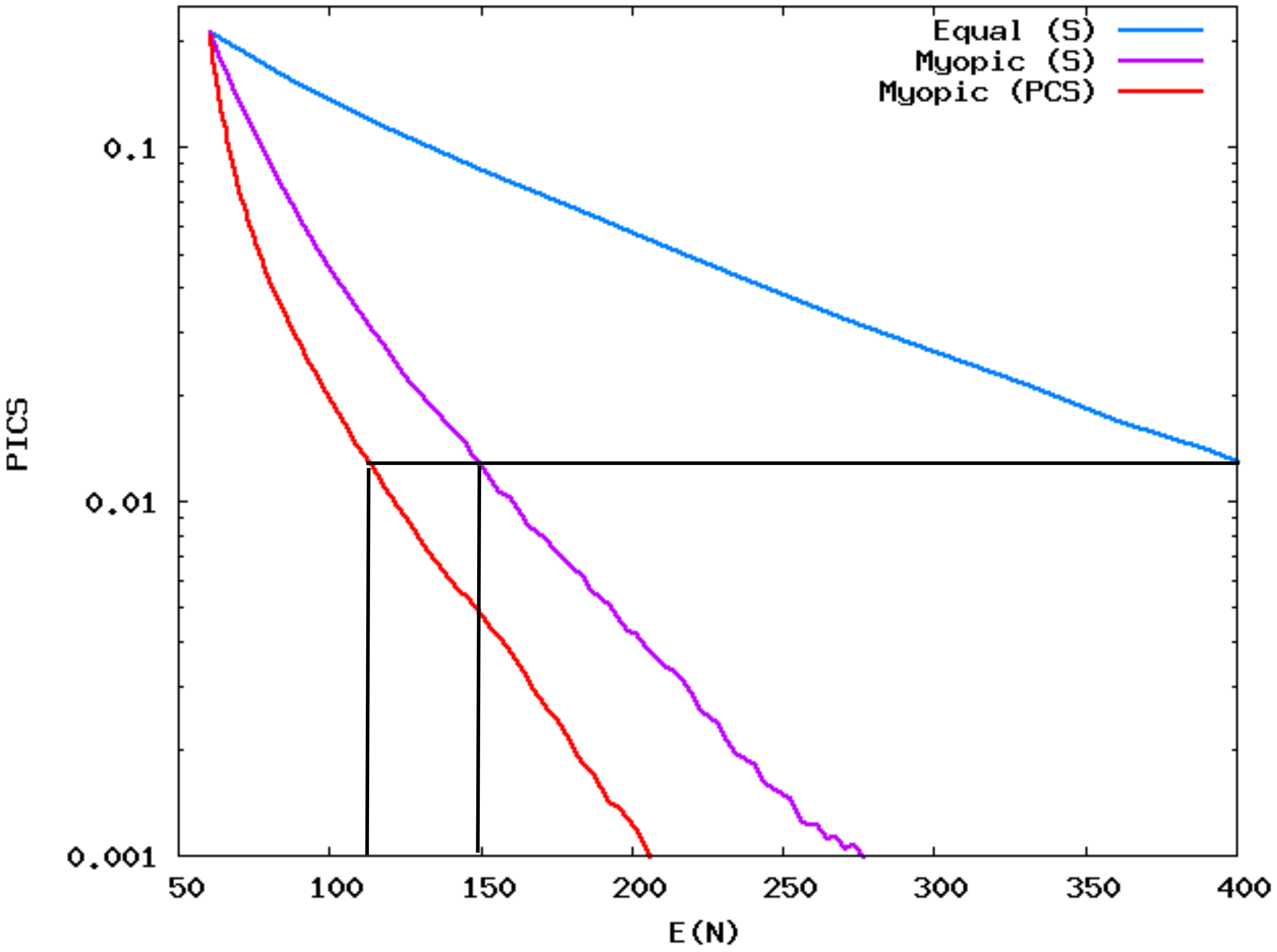
Stopping rule [Branke, Chick, Schmidt, Mngmt Sci, 2007]

- So far: Fixed budget
- Now: Estimate Probability of Correct Selection (PCS)

$$\begin{aligned} \mathrm{PCS}_{\mathrm{Bayes}} &= \Pr(W_{(k)} \geq \max_{j \neq (k)} W_{(j)}) \mid \Xi\} \\ &\geq \prod_{j:(j)\neq(k)} \Pr(W_{(k)} > W_{(j)}) \mid \Xi\} \\ &\approx \prod_{j:(j)\neq(k)} \Phi_{\nu_{(j)(k)}}(d^*_{jk}) \end{aligned}$$

with $d^*_{jk} = (\bar{x}_{(k)} - \bar{x}_{(j)}) \left(\frac{\hat{\sigma}^2_{(k)}}{n_{(k)}} + \frac{\hat{\sigma}^2_{(j)}}{n_{(j)}} \right)^{-1/2}$

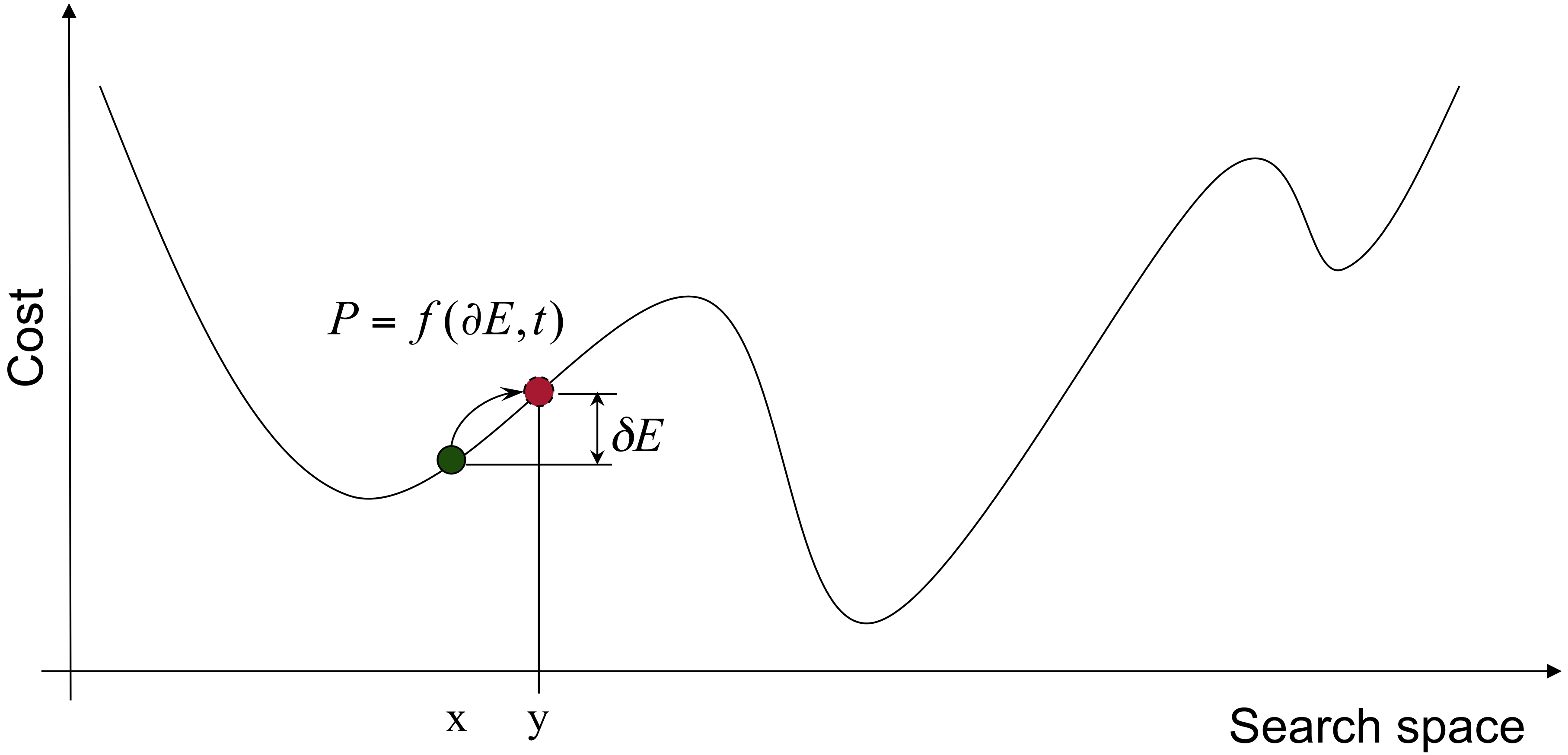
Empirical evaluation (find best out of 10 systems)



Black box optimisation

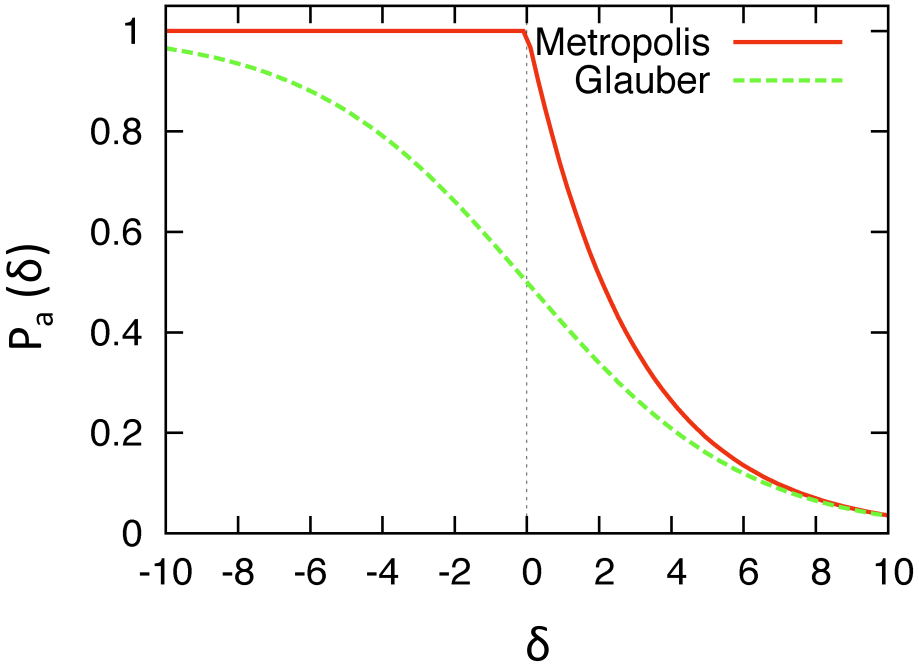
Simulated annealing

Stochastic local search inspired by physical annealing



Simulated Annealing

- Acceptance of solution is probabilistic and depends on quality difference δ and temperature *T*



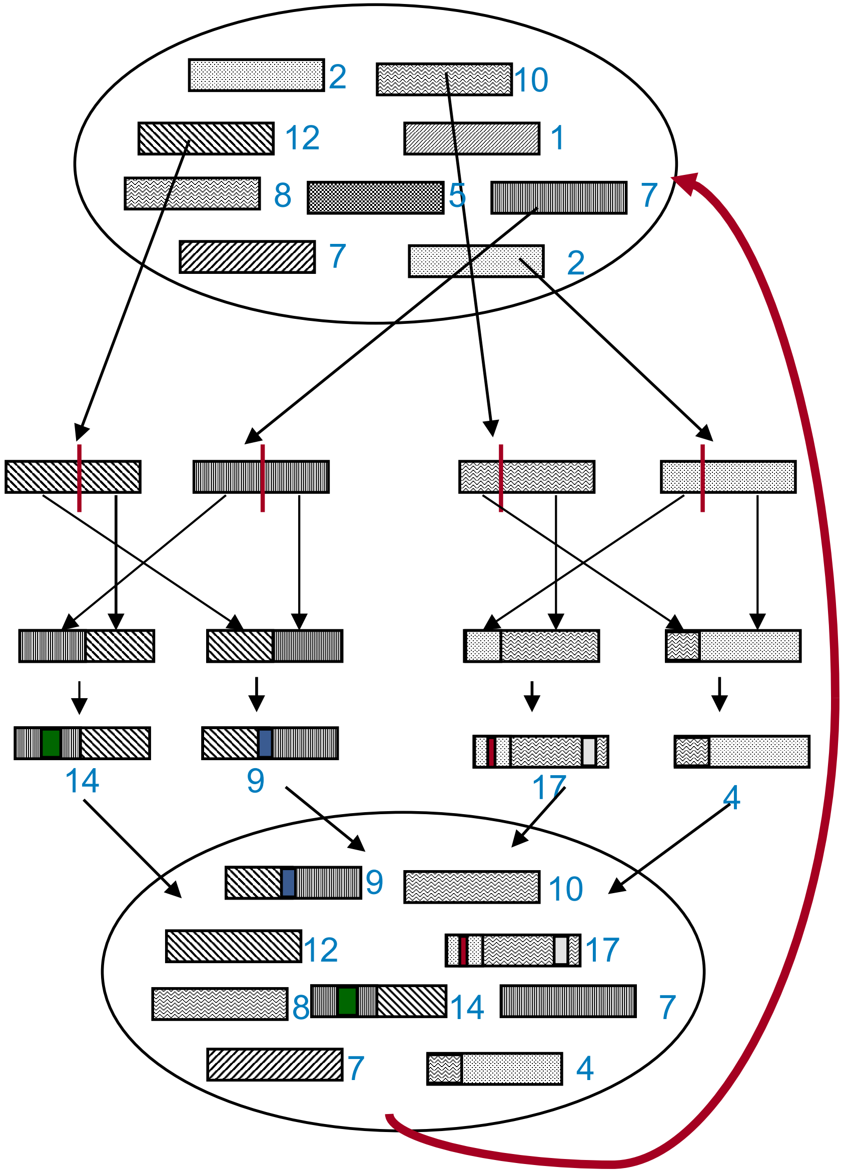
$$\frac{P_a(\delta)}{P_a(-\delta)} = e^{-\delta/T}$$

$$P_a^{Metropolis}(\delta) = \begin{cases} 1 & : \quad \delta \leq 0 \\ e^{-\delta/T} & : \quad \delta > 0 \end{cases}$$

Evolutionary algorithm

```
INITIALIZE population
        (set of solutions)
  EVALUATE Individuals
        according to goal ("fitness")

REPEAT
    SELECT parents
    RECOMBINE parents (CROSSOVER)
    MUTATE offspring
    EVALUATE offspring
    FORM next population
UNTIL   termination-condition
```

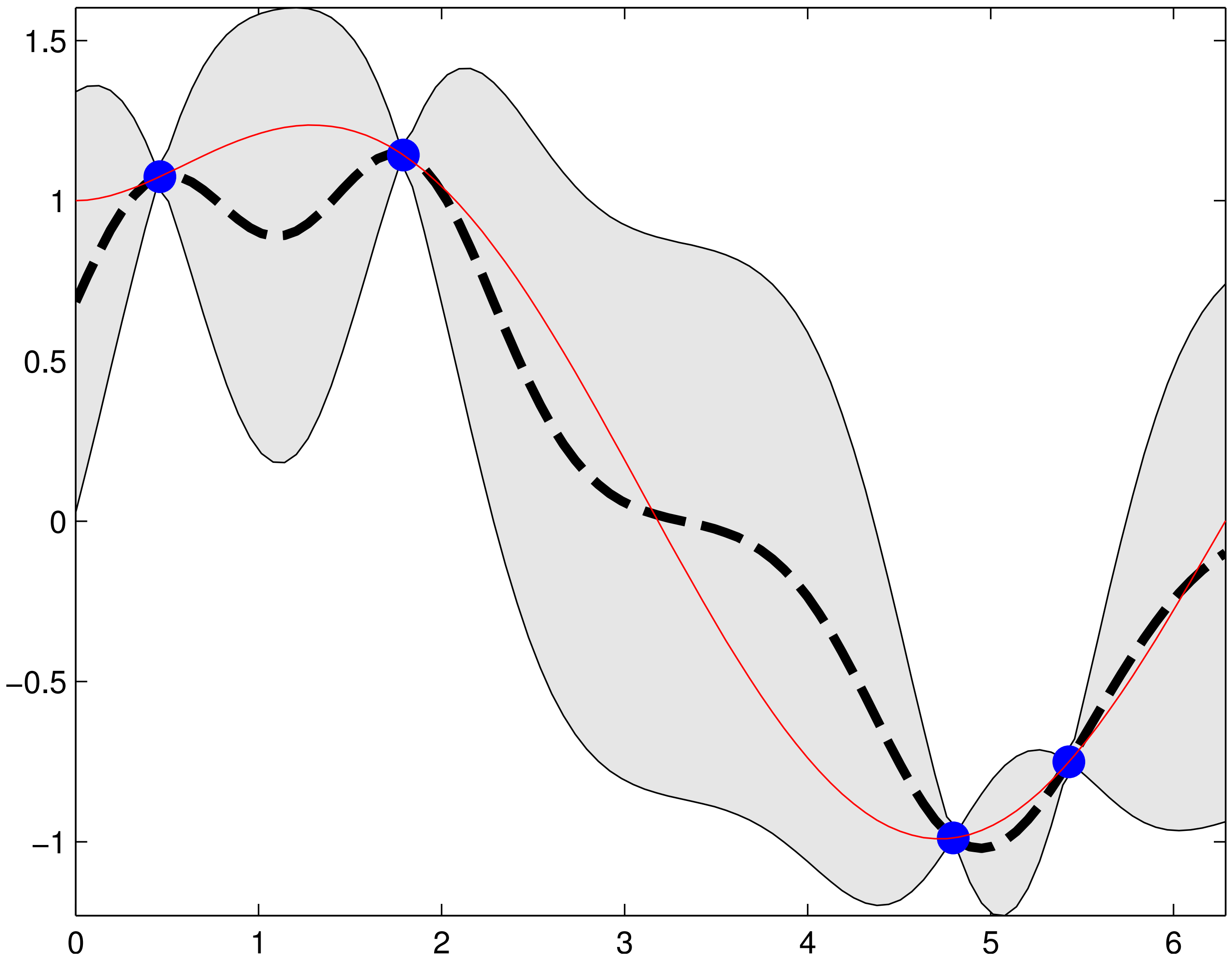


Efficient Global Optimisation (EGO)

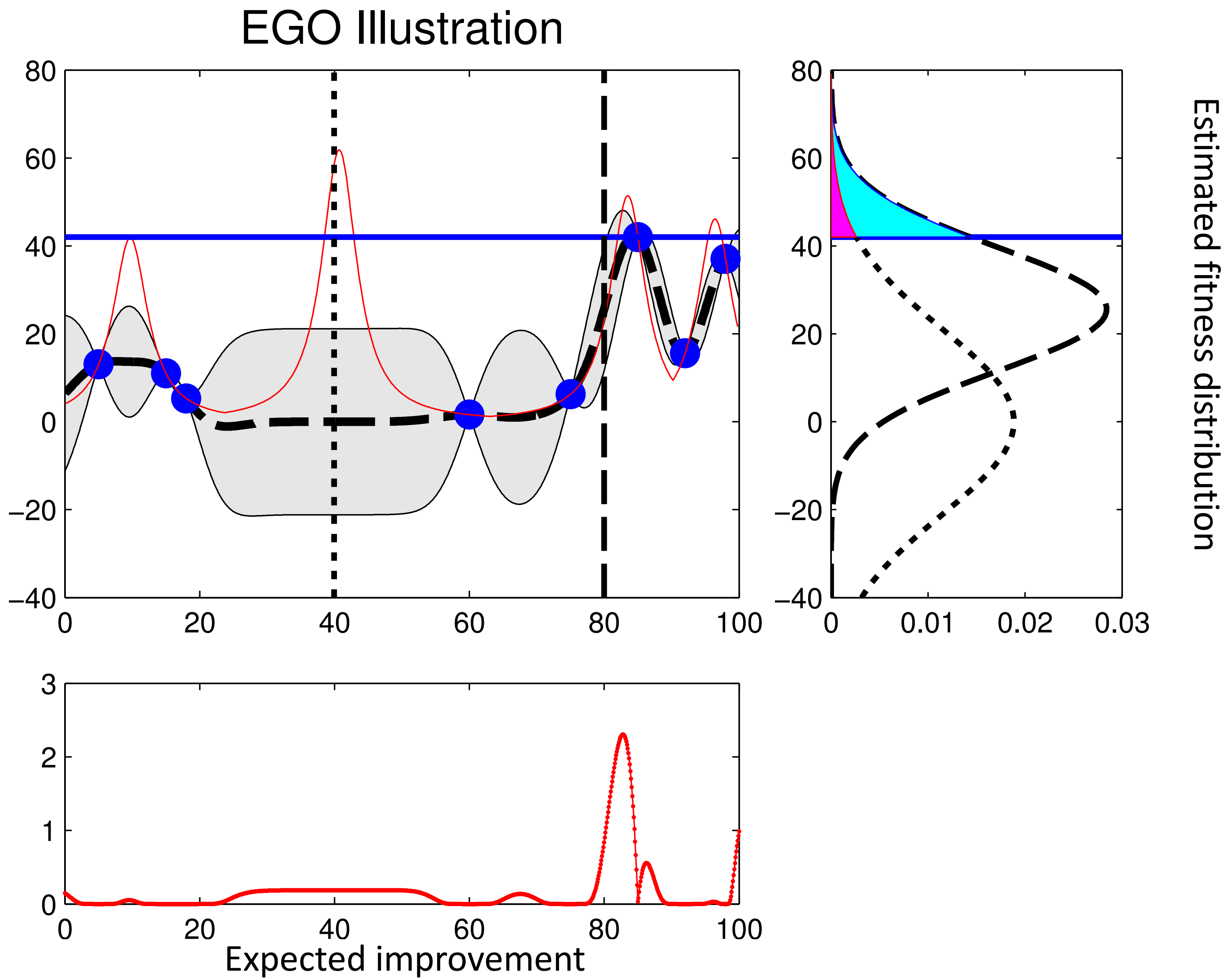
[Jones, Schonlau, Welch 1998]

- Fit a Gaussian Process (GP) to data
- Response model provides information about
    - expected value
    - uncertainty
- Use response model to determine next data point (replaces genetic operators)
- Expected improvement makes explicit trade-off between exploration and exploitation

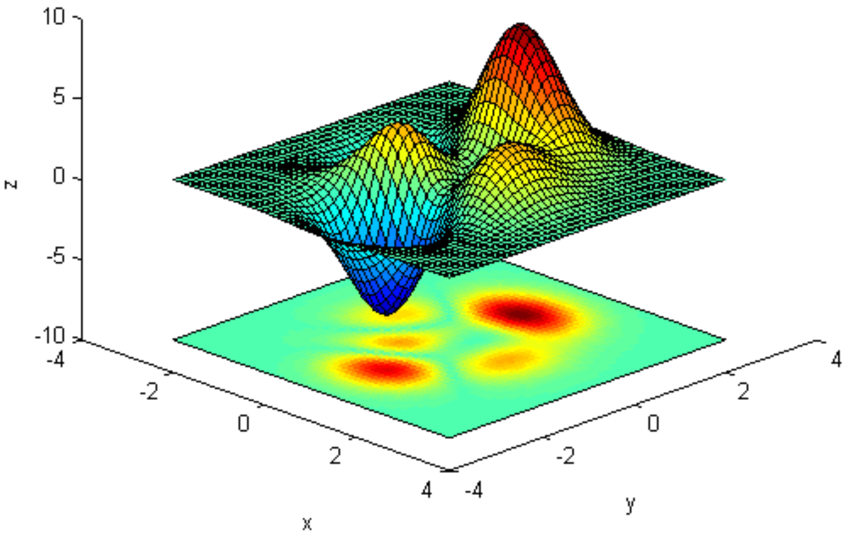
Example: GP in 1 dimension

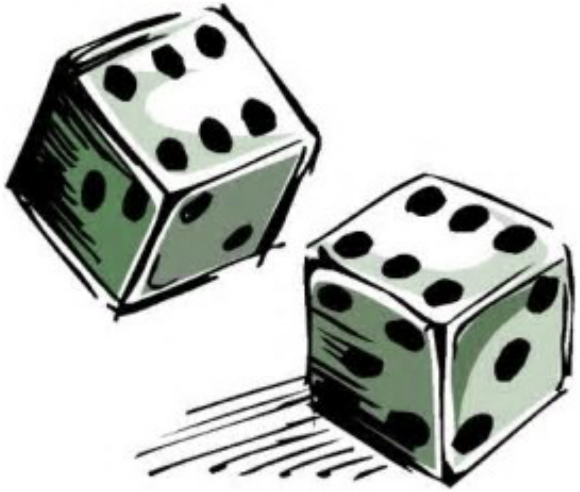


Max expected improvement principle

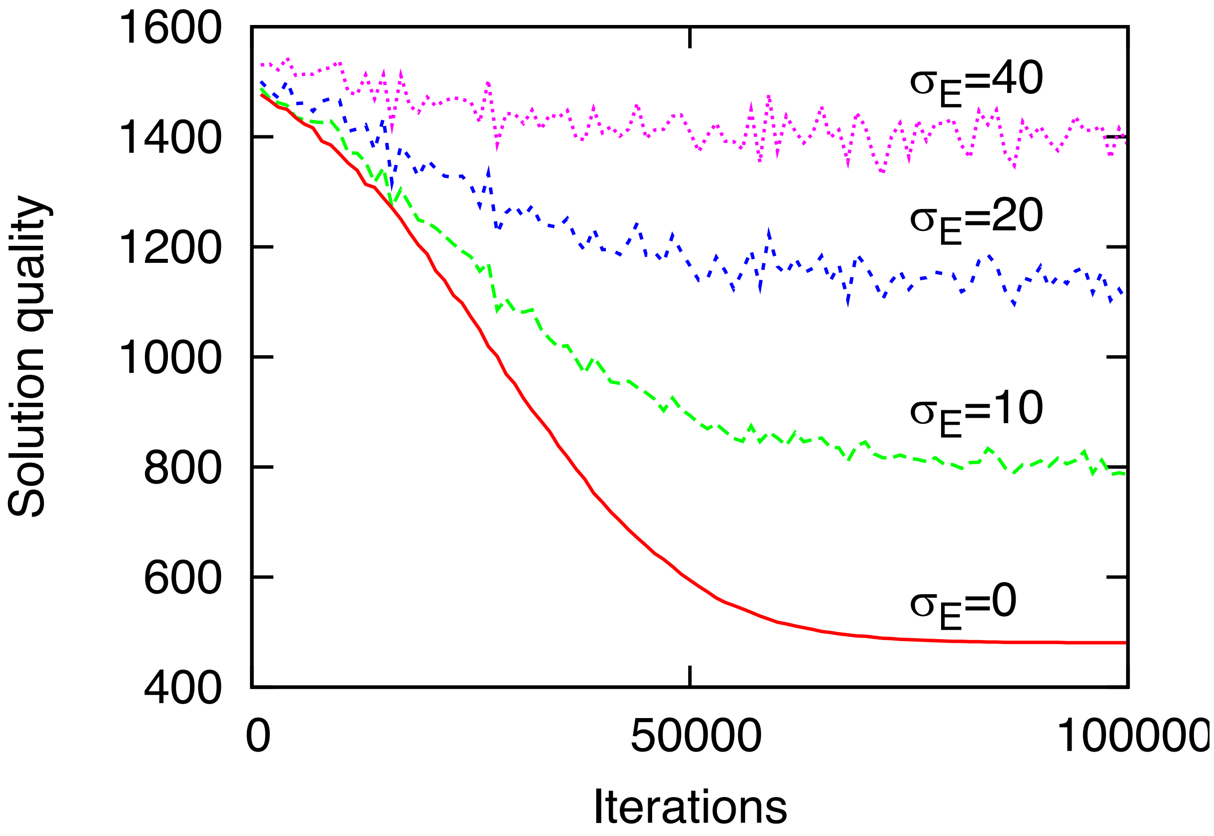


Optimisation under noise





Noise is detrimental for selection



Populations are robust to noise

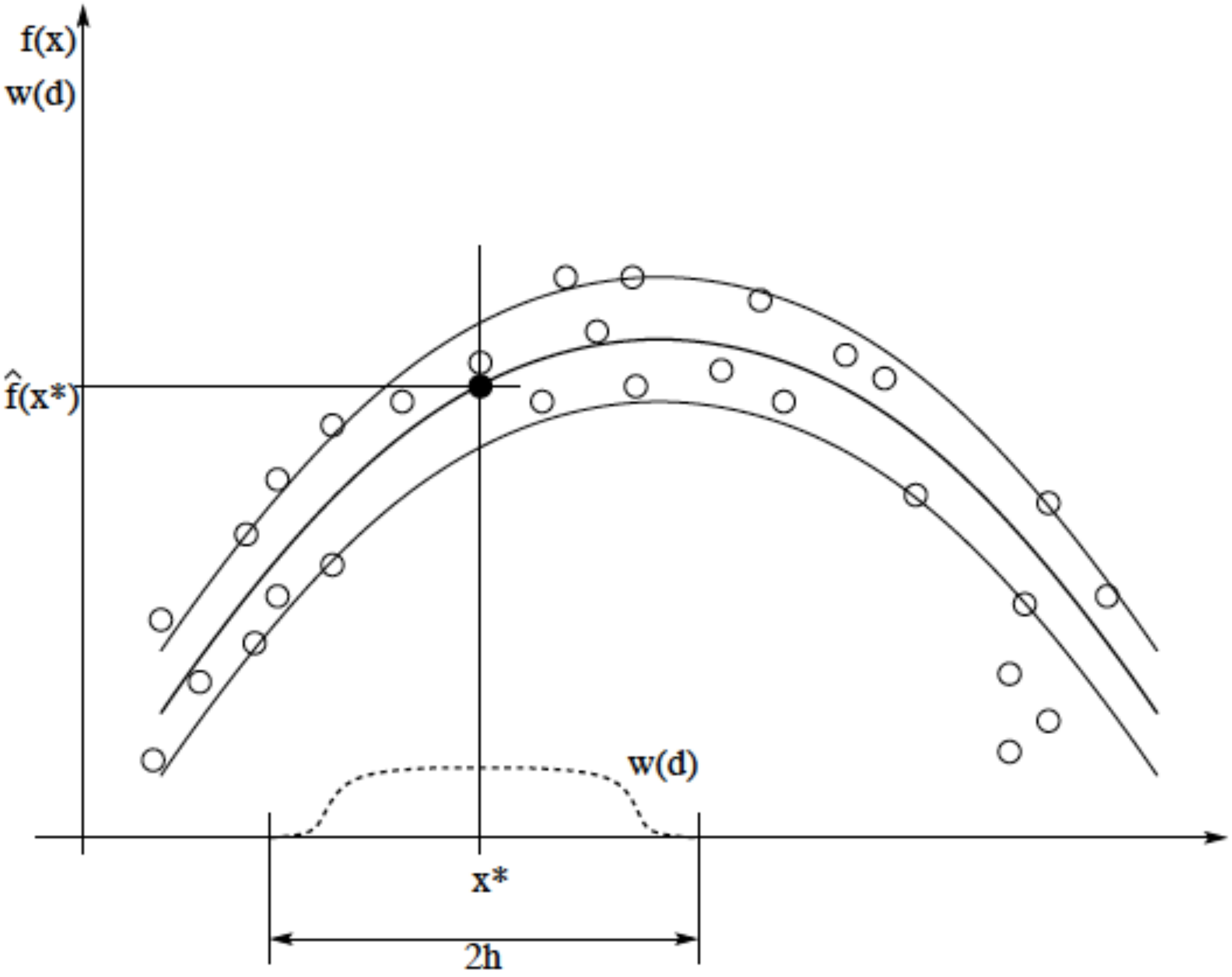
- Implicit averaging over the neighbourhood
- With infinite populations, fitness proportional selection is not affected by noise
- Theory for optimal population sizes in simplified cases
- Black-box Optimization Benchmark competitions show advantages of EAs in noisy environments

CRN and Evolutionary Algorithms

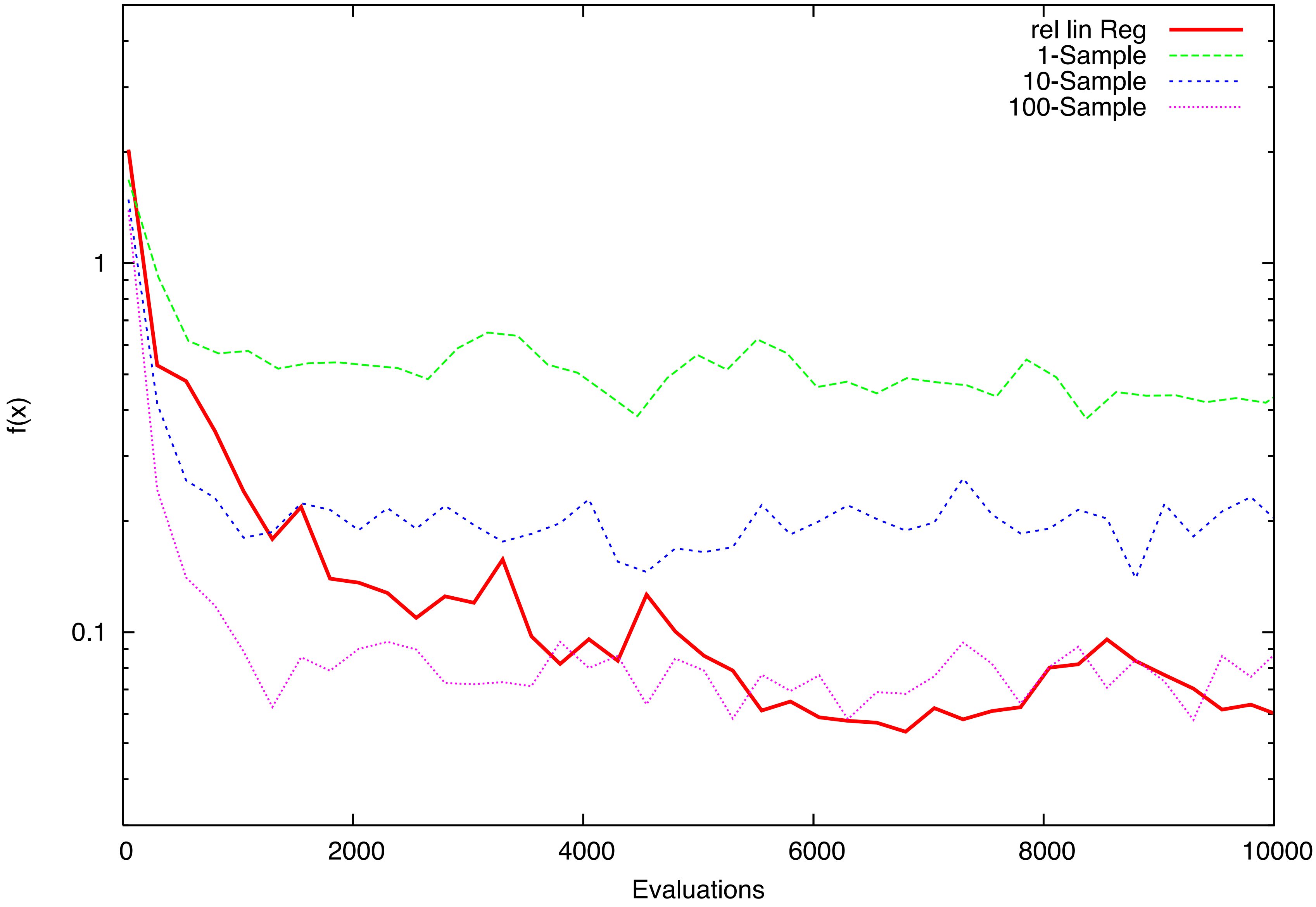
- Use CRN for all individuals to be compared within a generation
  - may drastically improve probability of correct ranking
  - risk of optimizing for one random seed
- Change random number seeds from generation to generation
  - Only individuals that work on a wide range of scenarios will survive for a long time

Use metamodels – average over space

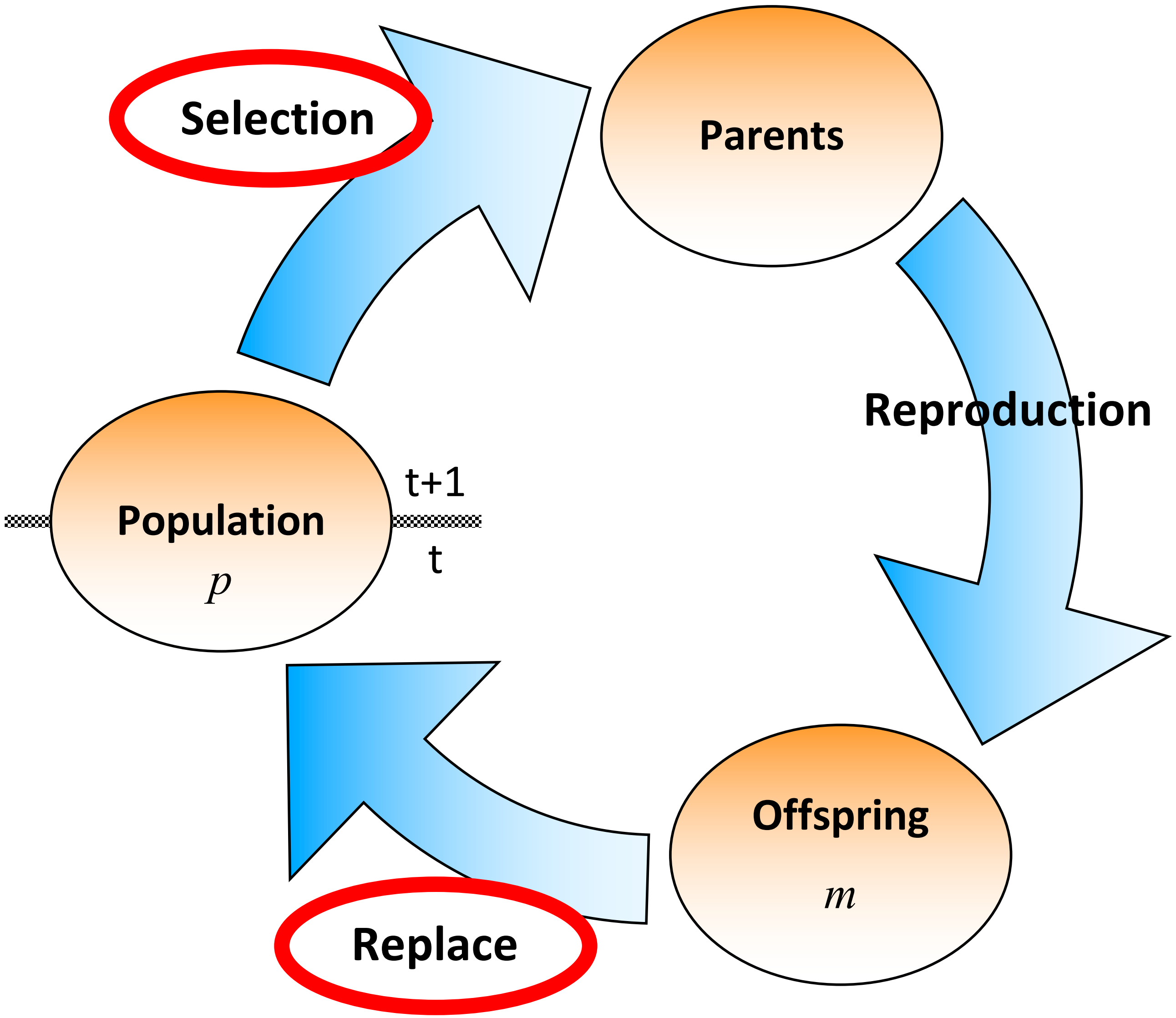
[Branke & Schmidt 2001]



Benefit


rel lin Reg
1-Sample
10-Sample
100-Sample
f(x)
1
0.1
0
2000
4000
6000
8000
10000
Evaluations

Integrating Ranking&Selection



The relevant comparisons

- Steady-State-EA with 2-Tournament

  Population size: 9, offspring: 1

  - Replacement: Worst individual
  - Stopping criterion: Best individual
  - Selection: Best out of {3, 7} and {2, 5}

Observed ranking

Observed ranking

≥	1	2	3	4	5	6	7	8	9	10
1										
2	**x**									
3	**x**									
4	**x**									
5	**x**	**x**								
6	**x**									
7	**x**		**x**							
8	**x**									
9	**x**									
10	**x**	**x**	**x**	**x**	**x**	**x**	**x**	**x**	**x**	

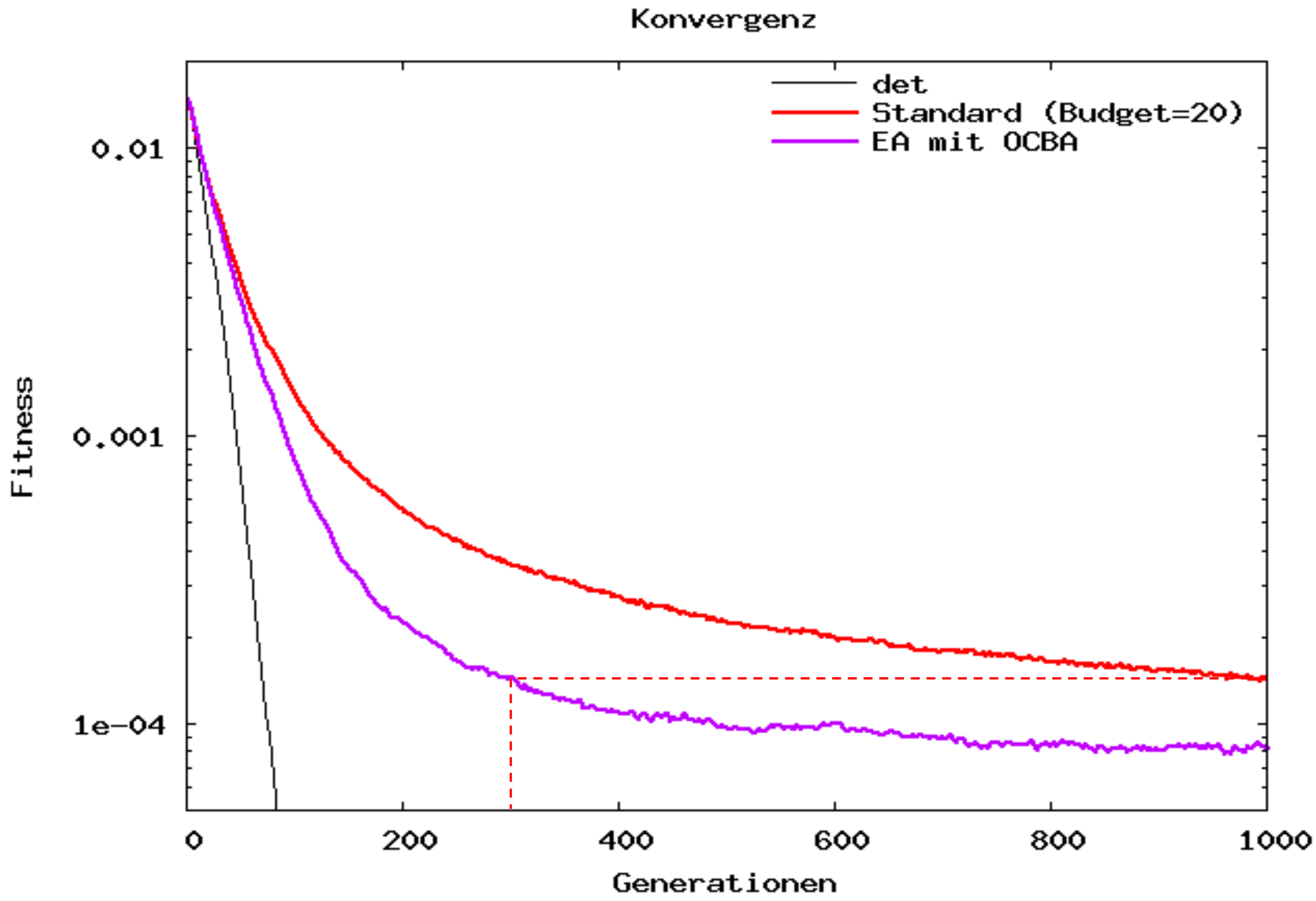
$$\mathrm{PGS}^{EA}_{\mathrm{Slep},\delta^*} = \prod_{(i,j)\in C} \Phi_{\nu_{ij}}((d_{ij}+\delta^*)/s_{ij})$$

Integrating OCBA and EA

Procedure OCBA^{EA}

1. Evaluate each new individual n_0 times. Estimate the ranks
2. Determine set of relevant comparisons C
3. WHILE evidence is not sufficient
    a) allocate new sample to individual according to modified OCBA rule
    b) if ranks have changed, update C

Benefits over the run



Optimal Stochastic Annealing (OSA)

[Ball, Branke, Meisel 2017]

- Tends to deterministically select the better solution
- Uses sequential sampling
- Acceptance criterion modified to maintain detailed balance $\frac{P_a(\delta)}{P_a(-\delta)} = e^{-\delta/T}$
- At every stage, decision to accept, reject or continue
- Acceptance criterion has optimal efficiency (acceptance probability per sample)

OSA acceptance rule

- Based on sum of samples taken so far

$$c_n = \sum_{i=1}^{n} \delta_i$$

- Acceptance probability at current stage:

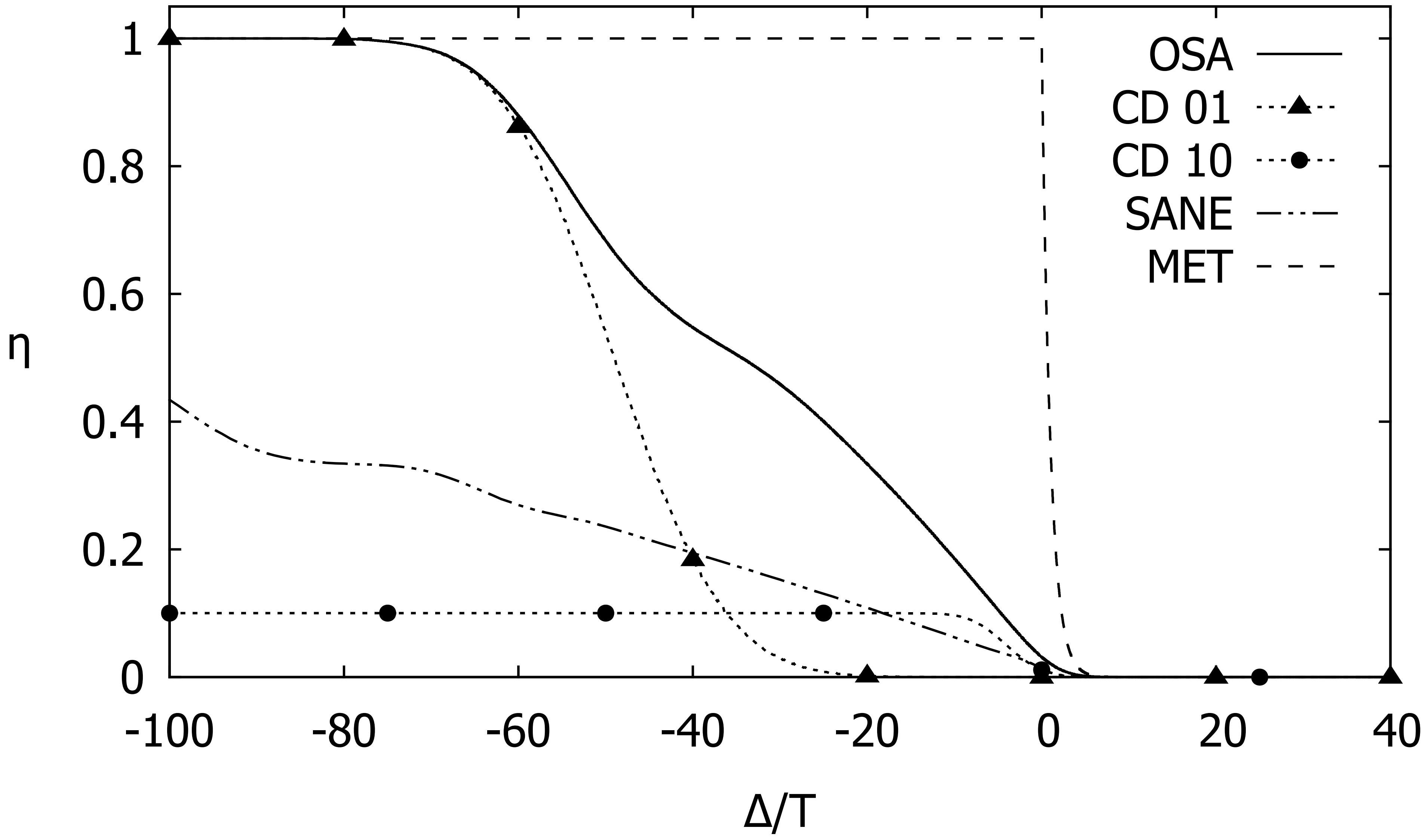
$$A(c_n, c_{n-1}) = \begin{cases} 1 & c_n < -\beta\sigma^2/2 \\ e^{-2(c_n+\beta\sigma^2/2)(c_{n-1}+\beta\sigma^2/2)} & \text{otherwise} \end{cases}$$

- If not accepted, reject if c_n>0
- Continue otherwise

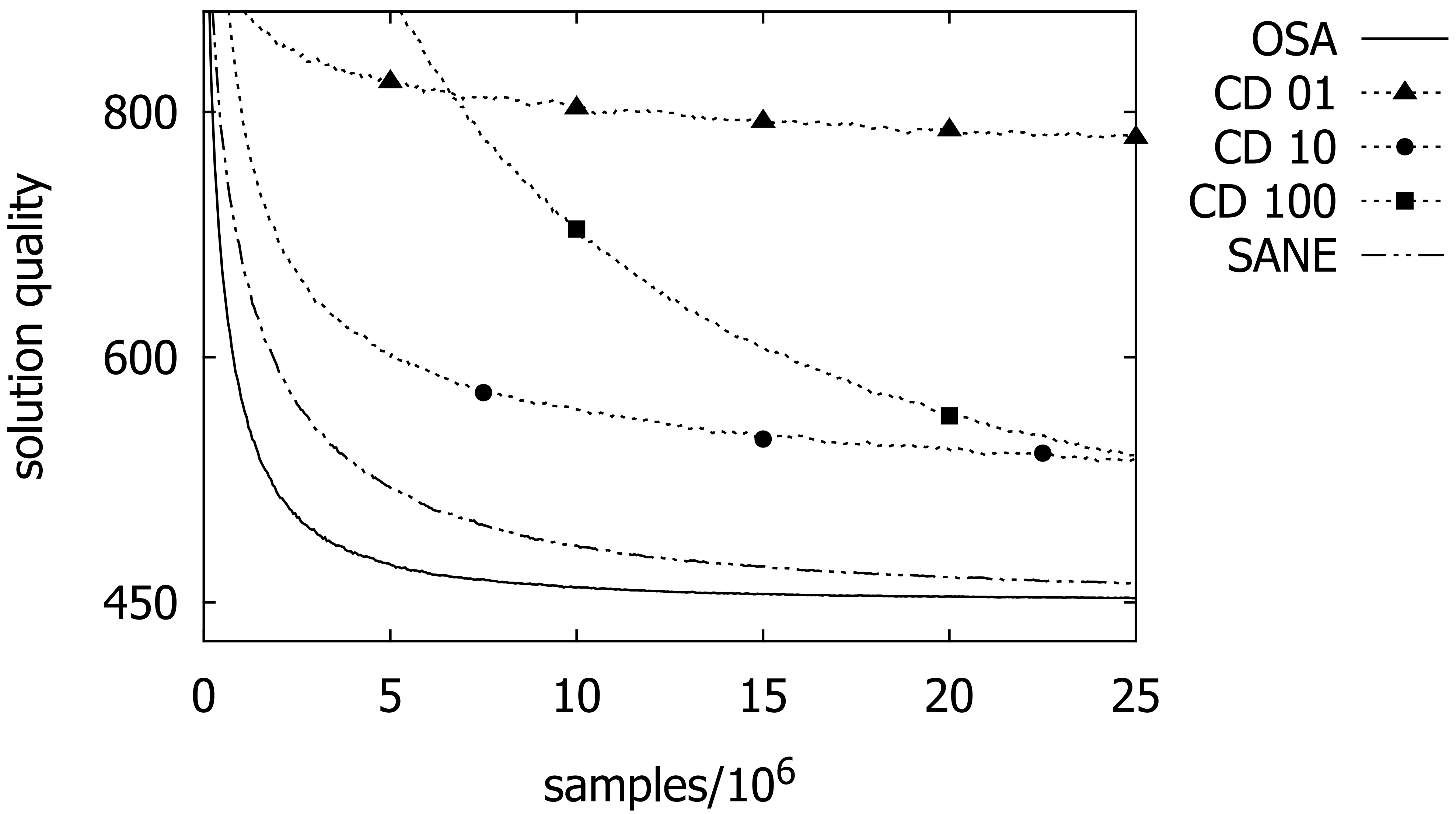
Benchmark algorithms

- SANE [Branke et al. 2007]
- CD1 [Ceperley&Dewing 1999]
  - Adjusted acceptance criterion, obeys detailed balance
- CD10 [Ceperley&Dewing 1999]
  - As CD1, but with 10 samples per move decision

Efficiency high noise (σ/T=10)



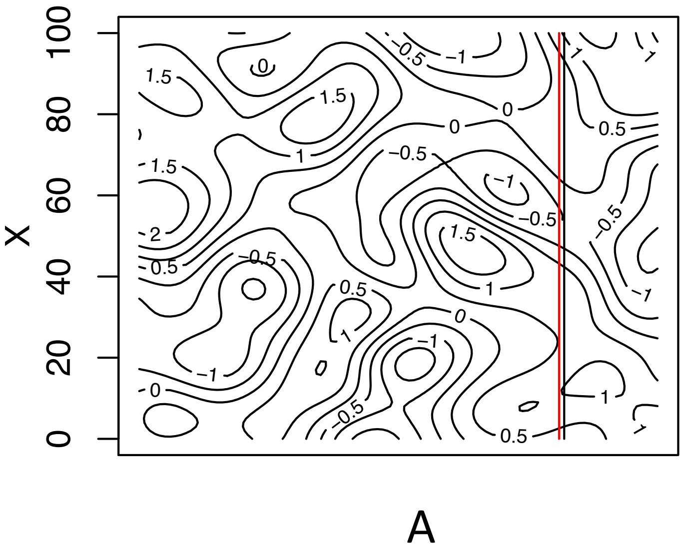
Optimization performance (TSP, σ^2=3200)

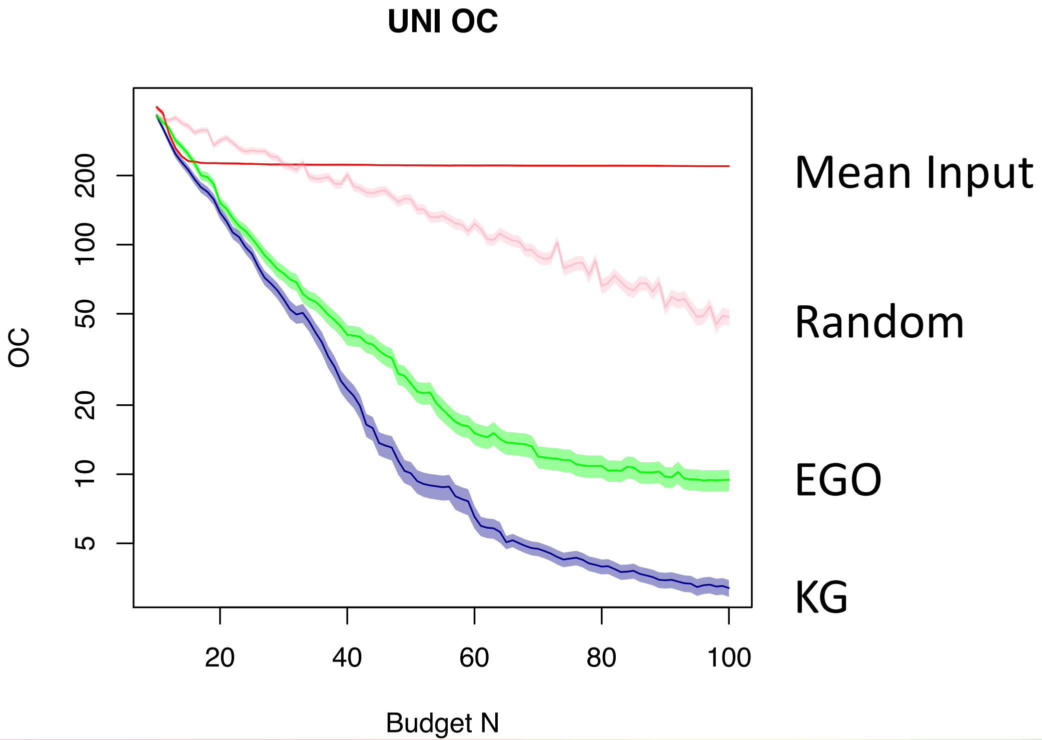


Related topics

Input uncertainty

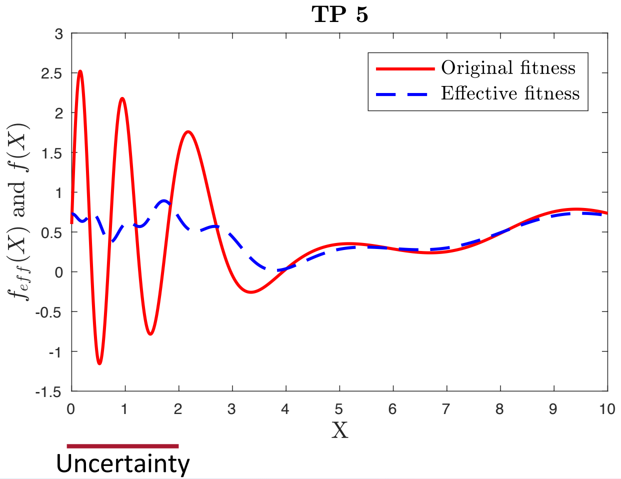
- A simulation model often has parameters estimated by experts of learned from data
- Given a probability distribution of these parameters, we want to find the solution with the best expected performance

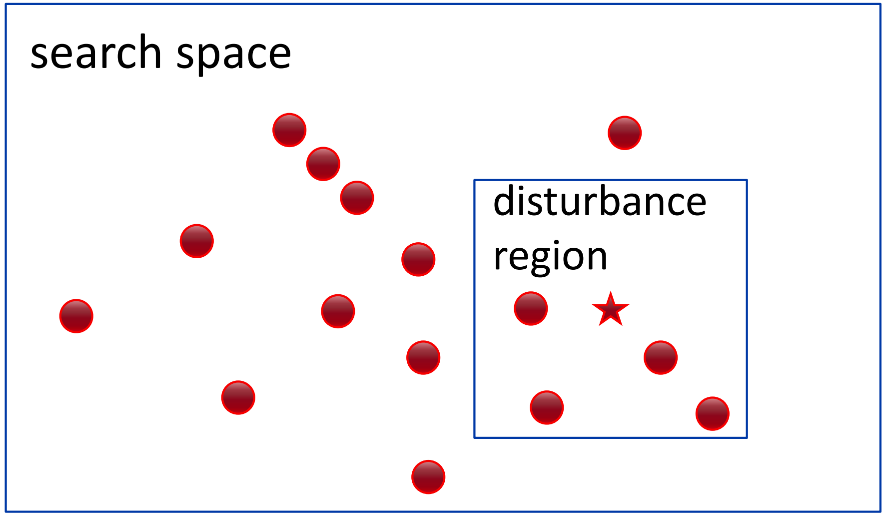




Searching for robust solutions

- Given a probability distribution of manufacturing tolerances, find the solution with the best expected performance
- Re-use previous evaluations
- Where to take new sample to minimise estimation error?

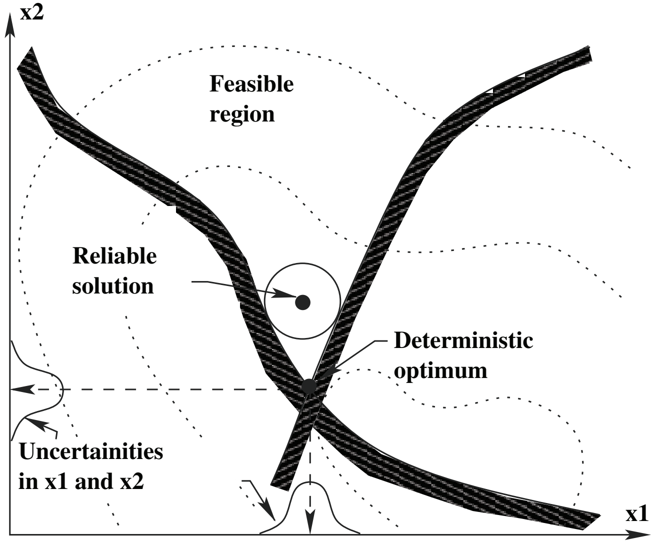


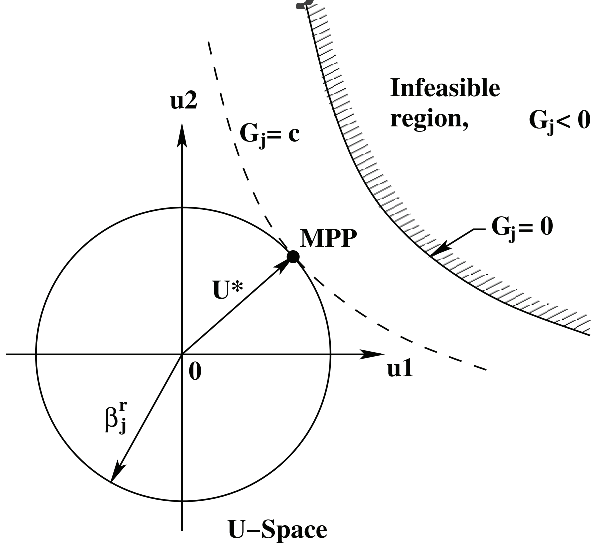

search space

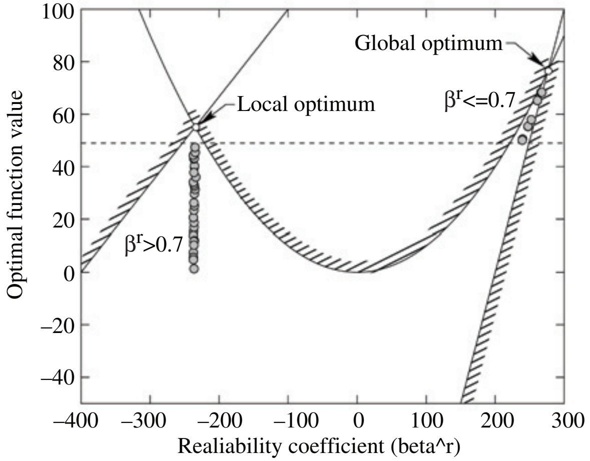
disturbance region

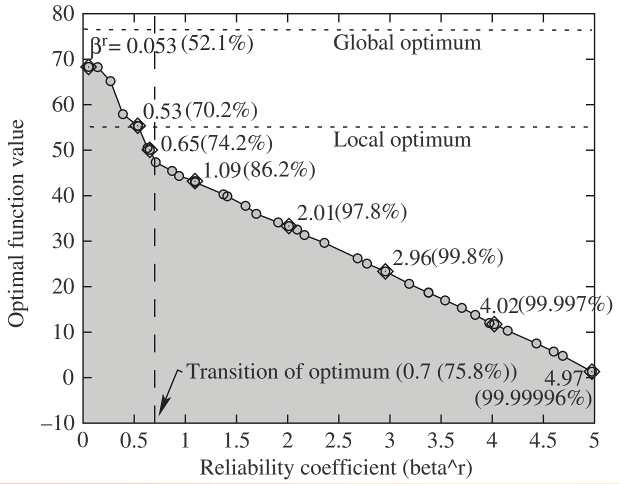
Reliability

- How likely is it that a solution is *feasible?*









Conclusion

- Simulation-based optimisation is powerful tool for design of complex systems
- Evolutionary algorithms, simulated annealing and Bayesian optimisation
- Uncertainty is major challenge
- Reduce uncertainty where most helpful
- Exploit neighbourhood information

Discussion