On the Importance of Information Speed in Structured Populations

Mike Preuss and Christian Lasarczyk*

University of Dortmund,
D-44221 Dortmund, Germany
{mike.preuss,christian.lasarczyk}@cs.uni-dortmund.de

Abstract. A radius-based separation of selection and recombination spheres in diffusion model EAs is introduced, enabling a new taxonomy, oriented towards information flow analysis. It also contains parallel hillclimbers, panmictic EA and an unexplored area. Experiments are performed systematically on five complex binary and real coded problems in search of the best performing variants w.r.t. available optimization time. Additionally, information flow through recombination and selection is emulated by means of a simple model, that produces qualitative similar results.

1 Introduction

It is a generally agreed on opinion that for successfully performing evolutionary search on multimodal and epistatic optimization problems, population diversity maintenance plays an important role. To prevent an evolutionary algorithm (EA) from concentrating on a small search space area, many different operators and structures based on local interaction have been introduced. In contrast to panmictic population models, they deliberately slow down the information flow.

Aranging individuals in topologies and and selecting offspring from local neighborhoods obviously exerts a certain influence on the information flow. This selection type severely limits the distance covered by the offspring as an information carrier. A second influence arises in connection with recombination: Local selection can lead to subpopulations in terms of sets of individuals with low diversity, resulting in high parent-offspring-fitness correlation.

As opposed to this, subpopulation boundaries often reveal incompatibilities with respect to recombination where individuals from different areas usually only bear offspring of much lower fitness than their parents. Fig. 1(a) shows this phenomenon for an exemplary ring topology EA run. Population separation in search space may lead to undesired effects: Subpopulations can get stuck far away from well known optima whereas others get extinct despite being located in promising areas. This happens e.g. if one out of two adjacent subpopulations is able to populate the common border position instead of bad individuals resulting from cross subpopulation recombination.

---

* Authors appear in random order.

Fitness
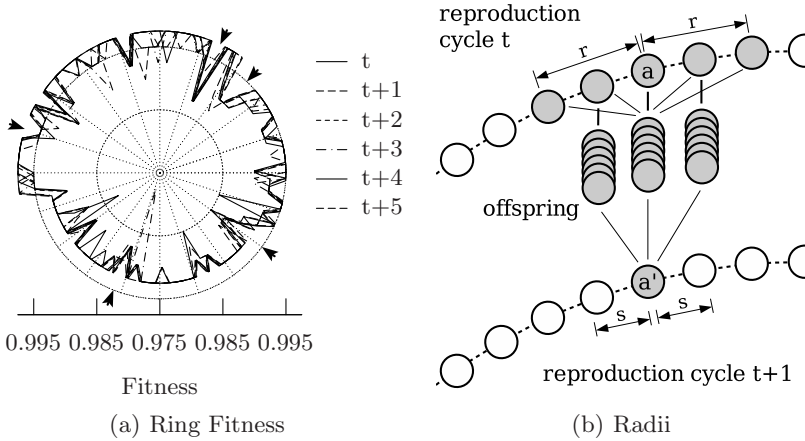
(a) Ring Fitness

(b) Radii

Fig. 1. Left hand: This plot uses radial coordinates to show the individuals fitness against their ring position, presenting six successive generations. Several connected areas exist, seperated by ring postions always showing fitness worse than their neighbors. These permanent fitness slumps indicate, that genomes of adajcent individuals could not be recombined successfully. Right hand: Offspring is generated by a central individual and mating partners from recombination radius r; selection accepts the best individual from the offspring pools within selection radius s.

In this paper, we consider the information flow induced into structured populations by the two ubiquitous non-unary evolutionary operators, recombination (sexual reproduction) and selection. Our aim is to separately observe their influence systematically under different local interaction radii in order to get insight into the mechanisms leading to relatively good or bad optimization performance on multimodal problems.

2   Generalized Local Selection in Structured Populations

Once originating from parallelization efforts, evolutionary algorithms employing structured populations are nowadays regarded as an approach to implicitly preserve diversity and thus prevent premature convergence to a non-optimal solution in case of a complex optimization task [2]. If explicit mechanisms (such as crowding [3], fitness sharing [4], species conservation [5], guided local search [6] or topology based selection in genetic programming [7]) are not considered, mainly two classes of algorithms remain, representing different control and communication methodologies: island model and diffusion model EAs [8]. As the former introduce additional operators like migration, we concentrate on the latter and investigate their relations to multi-hillclimber strategies on the one hand and algorithms using structureless, panmictic populations on the other hand.

Topology issues like size and shape of local neighborhoods in diffusion model EAs have been subject to extensive research (e.g. [9], [10]). Definition of overlapping demes containing all parents available for mating is strictly oriented

towards the needs of the recombination operator. Parent replacement is directly linked with the production of new offspring and usually happens either at the center or a random position of the actual deme.

However, in absence of expensive communication on a parallel hardware, nothing prevents us from copying an accepted individual to another location that belongs to different deme. In analogy to the specification of a recombination range, we may also define a selection range. For each parent individual, it would contain the locations a replacing offspring can originate from during selection. Going to the extremes, the selection range could include the whole population or just one individual. The former would correspond to global selection, the latter is local selection where only direct descendants are allowed to replace a parent. This scheme enables systematic empiric testing of different recombination and selection ranges.

A concrete instantiation of the previously suggested selection range depends on definition of an underlying topology. As we target optimization of very difficult multimodal and epistatic problems, it may be beneficial to slow down information distribution to a minimum and thus keep diversity high for a long period of time. We therefore restrict ourselves to mapping the parent population onto a ring topology. This provides us with maximum characteristic path length and indistinguishable local conditions (2 direct neighbors) for each individual.

We can now define a distance between individuals in algorithmic space, independent from their distance in search space. Distance between two individuals is ascertained by counting edges between them, distance of individuals to themselves is zero. For any arbitrarily chosen center individual, this enables constructing a neighborhood from all individuals found within a certain radius, a method also used in [10]. The whole population is included at radius $\mu/2$ if μ denotes the population size.

Consequently, we introduce two radii for determining recombination and selection ranges. As we want to model speed of information flow rather than analyze the impact of specific concrete operators, we idealize recombination as method that takes two parent individuals and generates one offspring. Likewise, idealized selection is a process that takes a number of offspring individuals and replaces a parent with the best one. Recombination radius r defines the maximum distance of individuals available for mating. As an individual is contained in its own recombination radius for any $r \geq 0$, we always permit selfing, which results in cloning the parent for individuals of constant length. Produced offspring is stored in sets assigned to their first parents until selection.

The distance between this set and the parent individual an offspring may possibly replace is called selection radius s. During selection we process each member of the parent population in random order and detect the best not yet assigned offspring individual from the union of all sets contained in s (see also fig. 1(b)). The chosen offspring always replaces the considered parent individual, resulting in a comma selection. In contrast to other local selection schemes, this allows for individuals being replaced by offspring from outside their recombination radius if $s > r$.

In the following, we refer to EA variants determined by concrete recombination and selection radii as *strategies*. Exploring their taxonomy given by r and s (see fig. 2(a)), we identify several well-known strategies: $(r = 0, s = 0)$, in short $(0,0)$, corresponds to μ independent $(1, \lambda/\mu)$-hillclimbers. The two arrows in fig. 2(a) represent diffusion models: $(r, 0)$ means replacement of the central individual of a deme, whether for $r = s$ strategies, replacement is randomly done within the recombination radius. Note that the $(\mu/2, \mu/2)$ strategy is very similar to a panmictic population structure with global selection, the only difference lies in the enforced minimum offspring number of λ/μ per parent. To our knowledge, EAs with $s > r$ are yet unexplored.

3   EA-Systems and Experiments

The previously described radius approach matches well with environmental selection driven by a birth surplus. Therefore, in all our experiments $\mu = 100$ parents bear $\lambda = 500$ offspring. To enable local selection, strict mating is enforced, so that each parent generates λ/μ offspring with randomly chosen mates from its recombination range. The following paragraphs describe the remaining settings of the utilized evolutionary algorithms and tested problems. Note that we use problem generators ([11]) instead of hand chosen problem instances wherever applicable, to obtain increased predictive power of the experimental results.

3.1   Experiments on Real-Valued Problems

For the real-valued problems, we use an EA instance containing mutation and recombination operators from standard evolution strategies [1]: self-adaptation of one mutation strength for gaussian mutations and dominant (discrete) or intermediary recombination with two parents.

Significance tests yielded that a maximum mutation strength of 1.0, a learning factor $\tau = 0.05$ and dominant recombination for the object parameters are robust settings for all three problems. All three continuous problems are formulated for minimization.

Rastrigin Function. The highly multimodal but strictly symmetrical Rastrigin function has originally been considered as difficult testbed for EA, especially when using small population sizes. However, more recent studies (e.g. [10]) demonstrated that it is efficiently solvable given that the population size exceeds a certain threshold. We use the function in a generalized form mentioned in [1] and generate problem instances by drawing random numbers for the waviness $B \in [2, 7]$, keeping the number of dimensions constant at 50. Search for good strategy parameters yielded dominant mutation strength recombination, low initial mutation strength (0.5) and low minimum mutation strength (10^{-30}). Each run is allowed 10^6 function evaluations and the object parameters are initialized at $5.0 + N(0, 1)$.

Continuous n-Peak Problem. This function generator serves as flexible tool in producing multimodal problem instances of configurable complexity. Its basic concept is to buildup fitness landscapes by combining several peaks of different height and structure as suggested in [12]. We use 100 problem instances with 8 randomly distributed optima in a $d = 10$ dimensional search space. For each peak we also define a dependency matrix to ensure non-separability of the resulting problem, a method inspired by [13]. Equations below summarize fitness computation for an object parameter vector $\mathbf{x}$. Herein, $f(\mathbf{x})$ stands for the fitness function, $g(\mathbf{x}, \mathbf{p})$ is the evaluation result of $\mathbf{x}$ for peak $\mathbf{p}$, $dist(\mathbf{x}, \mathbf{p})$ is the modified euclidean distance, and $dep(\mathbf{x}, \mathbf{p})$ calculates a dependencies term. The parameters s_p, h_p, and r_p denote peak shape, height and radius, respectively. All peaks are distributed in the space bounded by $x_i \in [0, 20]$, with $s_p \in [2, 3]$ and $r_p \in [10, 20]$. Peak heights of local optima are randomly chosen in between 0.7 and 0.9, such that fitness values below 0.1 originate only from search points on the peak containing the global optimum, with $h_p = 1.0$.

$$f(\mathbf{x}) = \min_{\forall p}(g(\mathbf{x}, \mathbf{p})\,), \quad g(\mathbf{x}, \mathbf{p}) = h_p \left(\sqrt[s_p]{\max \left(0, \frac{\text{dist}(\mathbf{x}, \mathbf{p})}{r_p}\right)} - 1 \right) + 1$$

$$\text{dist}(\mathbf{x}, \mathbf{p}) = \sqrt{\sum(x_i - p_i)^2 + \text{dep}(\mathbf{x}, \mathbf{p})}$$

$$\text{dep}(\mathbf{x}, \mathbf{p}) = \sum_{j=1}^{d} \sum_{k=j+1}^{d} (x_j - p_j)(x_k - p_k)D_{pjk}, \quad D_{pjk} := \frac{\text{random}(-0.5, 0.5)}{d - 1 - j}$$

On this problem class, dominant mutation strength recombination, low initial mutation strength (0.5) and low minimum mutation strength value (10^{-30}) prevail against all other tested strategy parameter combinations, as for the Rastrigin function. Also, an observation period of 10^6 function evaluations seems sufficient. Object parameters of initial individuals are randomly distributed, $x_i \in [0, 20]$.

Keane's Bump Function. Keane introduced this function [14] as test problem with several local optima occurring very near to constraint boundaries. For this problem we use the same problem instance with 20 dimensions for all runs. Strategy parameters chosen for this function are: low initial 0.5 mutation strength, high minimum mutation strength 10^{-4}, and dominant mutation strength recombination. A quadratic penalty function ensures strong attraction towards valid function regions. Runs start with object parameters randomly distributed in the valid interval $0 < x_i < 10$ and end after $2 * 10^6$ function evaluations.

3.2   Experiments on Boolean Coded Problems

Normally genetic algorithms (GA) use mating selection, so this is the most important point in which we differ from classic GA. Most other utilized settings are comparable to a standard configuration. We use 1-point-crossover to recombine two individuals and bit-flipping for mutation. Each bit is mutated with equal probability. This probability is equal to the reciprocal value of the individuals length.

NK Fitness Landscapes. Using NK models, every gene locus contributes to fitness. More precisely, its contribution does not only depend on its binary value, but also on the value of K other (adjacent) loci. Each gene of a genome can thus contribute to 2^{K+1} different values. Theses values are drawn randomly with uniform distribution out of $[0, 1]$ and managed in N different tables. The mean value of the $N = 48$ loci is the individuals fitness, which is to be maximized.

L-SAT, a Boolean Satisfiablity Problem. Boolean satisfiablity problems are frequently used as representatives of the class of NP-complete problems. The task is to find an assignment of boolean values to V variables, so that a boolean expression becomes true.

Mitchell et al.[15] presented the random L-SAT problem generator, which creates expressions in conjunctive normal form using three parameters. These expressions consist of C clauses of length L. The L variables are drawn randomly out of a set of V variables, and each of them is negated with a probability of 0.5. When optimized with GAs, fitness of a candidate solution is typically measured by the proportion of satisfied clauses. Here we use $V = 100$ variables and each clause is composed of $C = 3$ (possibly negated) variables.

4   Results

Our experimental design focuses on the influence of different combinations of selection and recombination radius. We thus consider 36 different *strategies* as defined in sec.2, resulting from $r \times s$ with $r, s \in [0, 1, 2, 5, 16, 50]$. We chose this logarithmic distribution of measuring points, because test runs showed promising results in the range of low radii. For each tackled problem class and strategy combination we carry out 100 runs. On the same problem the set of instances is equal for all strategies.

First we want to investigate how effectiveness of strategies depends on the number of objective function evaluations. We call a strategy dominant if its populations show best average best ever fitness. Fig. 2(b) plots the sequence of parameters of dominant strategies at each time step.

The strategy-time-trajectories show consistent gradients for most problems. On short runs, strategies using high radii usually perform best. However, at a later point in time, strategies with lower selection radii start to dominate. These allow for mating with distant individuals, but only offspring at central and neighboring positions can get selected. Figure 3 plots the average best ever fitness progression of selected strategies. Ring-strategies ($r = 1, s = 0$) with minimal deme size show best results from around 10^5 evaluations on most problems. On the long run, multi-hillclimber strategies with ($r = 0, s = 1$) and without ($r = 0, s = 0$) individual "migration by selection" also perform well, the latter starting faster, but stagnating earlier. The former eventually take the lead at nearly 10^6 evaluations for the n-Peak and the two L-Sat problems. Both strategies are not allowed to use sexual reproduction (apart from selfing) and thus have no means to mix genetic information of two individuals. Nevertheless, recombination seems
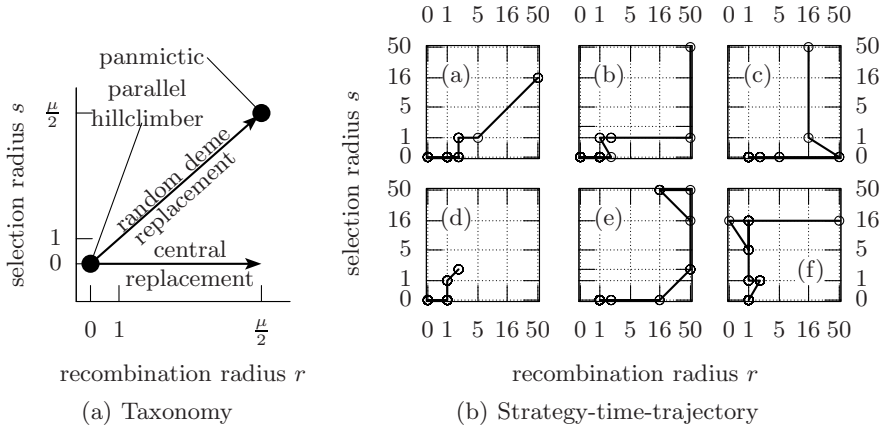
(a) Taxonomy   (b) Strategy-time-trajectory

Fig. 2. Left hand: Taxonomy of population models using selection and recombination radius, with the positions of well-known EA variants. Right hand: Strategy-time-trajectory of considered problems. For each measuring point in time, we plot the strategy displaying best average best ever fitness. At first, high radii dominate, but later on strategies with low radii take the lead. (a) L-Sat ($C = 400$), (b) L-Sat ($C = 1200$), (c) NK, (d) n-Peak, (e) Rastrigin, (f) Keane's Bump.

to be crucial for solving Keane's bump and Rastrigin problems and at least important for the "NK" problem where ring-strategies stay dominant.

Figure 2(b) first suggests that strategies with high selection and low recombination radius are medium-term advantageous on Keane's Bump function. However, fig. 3 shows that this part of the strategy-time-trajectory corresponds to the early stages of the optimization, where most strategies achieve similar results. This is true until the ($r = 1, s = 1$)-strategy starts to dominate. In our opinion, the short dominance of strategies with low recombination and high selection radius is not statistically significant. Choosing radii that comply with $s > r$ does not seem to make sense in general. For these strategies recombination and selection radii are effectively equal, because an individual chosen for mating within a small neighborhood may just have been transferred into it from a far away position during the previous reproduction cycle.

5 Radius Dependent Information Speed Model

The previously identified qualitatively similar relations between strategies raise the question, whether these results are reproducible by means of a simple simulation model. Such a model shall only take information transfer via recombination and selection into account, ignoring mutation. We thus suggest to consider each position on the ring topology as host of an information carrier, representing an amount of information with a single real value $x \in [0, 1]$. Initially we deposit information quantity 1 at one position and set all other positions to value 0.
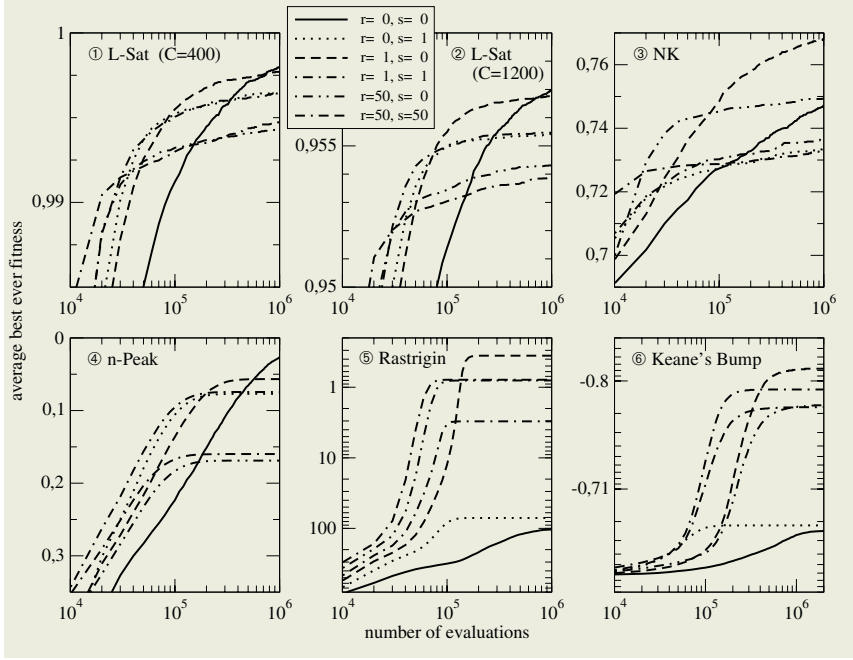
Fig. 3. Time dependent average best ever fitness on considered problems for a small subset of analyzed strategies, scaled differently to stress important fitness ranges. Higher curves show better fitness.

We then measure how fast this information is transfered to the opposite ring position. We assume neutral fitness, i.e. we only model the genetic drift effect.

Each simulation time step consists of randomly determining a position and replacing its information carrier. Therefore, we randomly select an other information carrier a within its selection radius s and chose a random mating partner b out of the recombination radius r of a. Assigned information quantity $I(o)$ of offspring o depends on information quantities of its parents and their diversity. It is drawn randomly from a normal distribution $N((I(a)+I(b))/2, |I(a)-I(b)|/2)$, with values outside $[0, 1]$ adjusted to the limits.

Fig. 4 plots the average amount of information transfered to the opposite position for different (r,s) configurations, averaged from 1000 replicates. Settings using recombination radius $r = 0$ are only capable of transferring information as a whole by selection and thus always yield discrete values zero or one at the opposite position. Setting $(r = 0, s = 0)$ prohibits information transfer completely and always leads to zero values at the opposite position.

The ring-strategy $(r = 1, s = 0)$ shows desirable behavior in this simple model, too. Note that it yields high information quantities despite an expectedly deferred start.

Further comparison with fitness plots of fig. 3 reveals, that high radii lead to a fast increase of information quantity in the beginning but also favor early
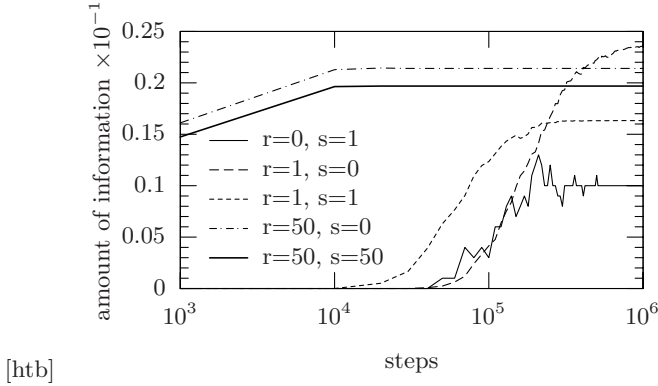
Fig. 4. Amount of information at the opposite ring position using a simple model of information transfer by means of recombination and selection.

stagnation. Low radii settings need much more time to start information transfer, but show progress for longer periods of time. Again, recombination seems to be important, enabling transfer of parts of information in contrast to the "all or nothing"-principle of pure selection.

6   Conclusion and Outlook

An improved success rate by increased diversity on the one hand and a loss of computing time on non-promising regions of search space on the other hand are the two conflicting effects of local recombination and selection. These effects are adjustable by the degree of locality, controlled by selection and recombination radii, which have to be chosen with respect to the available optimization time.

Using the radius based approach we identified three important strategies for optimizing complex problems with an EA: Panmictic population structures provide early results whereas for long runs, we suggest use of a local-selection EA with minimal recombination radius $(1, 0)$. Alternatively, a multi-hillclimber approach may be worthwile although it seemingly fails on some problems like NK landscapes, Keane's Bump or the Rastrigin function. This can be due to bad parameterization (e.g. mutation strength) or the inability to aggregate matching building blocks.

In the presented approach, we explicitly modelled the speed of information flow in EAs. Beside rings we considered other topologies as small-world networks [16] that are well known as topologies of natural systems. Thus, we shortened average path lengths by adding shortcuts instead of increasing radii. As those shortcuts connect very diverse subpopulations, we again met the problem of unsuccessful recombination.

In our opinion, investigation and control of information flow is a promising approach to improve and better understand the dynamics of EAs. Nevertheless, research on this area shall be guided by examination of recombineability.

References

1. Beyer, H.G., Schwefel, H.P.: Evolution strategies: A comprehensive introduction. Journal Natural Computing **1** (2002) 3–52
2. Eiben, A.E., Smith, J.E.: Introduction to Evolutionary Computing. Springer (2003)
3. De Jong, K.A.: An analysis of the behavior of a class of genetic adaptive systems. PhD thesis, University of Michigan (1975)
4. Goldberg, D.E., Richardson, J.: Genetic algorithms with sharing for multimodal function optimization. In Grefenstette, J., ed.: Genetic Algorithms and their Applications (ICGA'87), Lawrence Erlbaum Associates (1987) 41–49
5. Li, J.P., Balazs, M.E., Parks, G.T., Clarkson, P.J.: A species conserving genetic algorithm for multimodal function optimization. Evolutionary Computation **10** (2002) 207–234
6. Voudouris, C.: Guided local search – an illustrative example in function optimisation. BT Technology Journal **16** (1998) 46–50
7. Lasarczyk, C.W.G., Dittrich, P., Banzhaf, W.: Dynamic subset selection based on a fitness case topology. Evolutionary Computation **12** (2004) in print.
8. Alba, E., Tomassini, M.: Parallelism and evolutionary algorithms. IEEE Transactions on Evolutionary Computation **6** (2002) 443–462
9. Sarma, J., De Jong, K.: An analysis of the effects of neighborhood size and shape on local selection algorithms. In Voigt, H., Ebeling, W., Rechenberg, I., eds.: Parallel Problem Solving from Nature – PPSN IV, Berlin, Springer (1996) 236–244
10. Sprave, J.: Linear neighborhood evolution strategy. In Sebald, A.V., Fogel, L.J., eds.: Proc. EP'94, Singapore, World Scientific (1994) 42–51
11. De Jong, K.A., Potter, M.A., Spears, W.M.: Using problem generators to explore the effects of epistasis. In Bäck, T., ed.: ICGA, San Francisco, CA, Morgan Kaufmann (1997) 338–345
12. De Jong, K.A., Spears, W.M.: An analysis of the interacting roles of population size and crossover in genetic algorithms. In Schwefel, H.P., Männer, R., eds.: Proceedings PPSN I, Springer-Verlag (1990) 38–47
13. Salomon, R.: Re-evaluating genetic algorithm performance under coordinate rotation of benchmark functions: A survey of some theoretical and practical aspects of genetic algorithms. BioSystems **39** (1996) 263–278
14. Keane, A.J.: Experiences with optimizers in structural design. In Parmee, I.C., ed.: Proc. Adaptive Computing in Engineering Design and Control 94, Plymouth, UK. (1994) 14–27
15. Mitchell, D., Selman, B., Levesque, H.: Hard and easy distributions of SAT problems. In: Proceedings of AAAI–92, San Jose, California, AAAI Press, Menlo Park, California, USA (1992) 459–465
16. Watts, D.J., Strogatz, S.H.: Collective dynamics of 'small–world' networks. Nature **393** (1998) 440–442