The Ising Model on the Ring: Mutation Versus Recombination

Simon Fischer and Ingo Wegener[*]

FB Informatik, LS2, Univ. Dortmund, 44221 Dortmund,
Germany
{simon.fischer,ingo.wegener}@uni-dortmund.de

Abstract. The investigation of genetic and evolutionary algorithms on Ising model problems gives much insight how these algorithms work as adaptation schemes. The Ising model on the ring has been considered as a typical example with a clear building block structure suited well for two-point crossover. It has been claimed that GAs based on recombination and appropriate diversity-preserving methods outperform by far EAs based only on mutation. Here, a rigorous analysis of the expected optimization time proves that mutation-based EAs are surprisingly effective. The $(1 + \lambda)$ EA with an appropriate λ-value is almost as efficient as usual GAs. Moreover, it is proved that specialized GAs do even better and this holds for two-point crossover as well as for one-point crossover.

Keywords: Ising model, mutation vs. recombination, expected optimization time, fitness sharing

1 Introduction

Nowadays, genetic algorithms (GAs) and evolutionary algorithms (EAs) are mainly applied as optimization algorithms. Holland [3] has designed GAs as adaptation systems. The building block hypothesis (see Goldberg [2]) claims that GAs work by combining different building blocks in different individuals by crossover (or recombination). There is a long debate on the role of mutations in this context.

Naudts and Naudts [7] have presented the Ising model as an interesting subject for the investigation of GAs and EAs. Ising [4] has described the model now called Ising model to study the theory of ferromagnetism. In its most general form, the model consists of an undirected graph $G = (V, E)$ and a weight function $w\colon E \to \mathbb{R}$. Each vertex $i \in V$ has a positive or negative spin $s_i \in \{-1, +1\}$. The contribution of the edge $e = \{i, j\}$ equals $f_s(e) := s_i \cdot s_j \cdot w(e)$. The fitness $f(s)$ of the state s equals the sum of all $f_s(e)$, $e \in E$, and has to be maximized.

[*] Supported in part by the Deutsche Forschungsgemeinschaft (DFG) as part of the Collaborative Research Center “Computational Intelligence” (SFB 531) and by the German Israeli Foundation (GIF) in the project “Robustness Aspects of Algorithms.”

K. Deb et al. (Eds.): GECCO 2004, LNCS 3102, pp. 1113–1124, 2004.

The Ising problem in its general form is NP-hard. Nevertheless, there are quite efficient algorithms for this problem (Pelikan and Goldberg [8]). For the investigation of the adaptation capabilities of simple GAs and EAs, one is interested in the case where $w(e) = 1$ for all $e \in E$. By an affine transformation, we consider the state space $\{0, 1\}^n$ instead of $\{-1, +1\}^n$. The fitness $f(s)$ equals the number of monochromatic edges, i. e. edges connecting vertices of equal spin or color. The states 0^n and 1^n are the only optimal states for connected graphs. Connected monochromatic subgraphs are schemata of high fitness and, therefore, building blocks. However, the fitness function has the property of spin-flip symmetry, i. e., $f(s) = f(\overline{s})$ for all states s and their bitwise complement $\overline{s}$. Therefore, 0-colored building blocks compete with 1-colored building blocks.

The Ising model on the ring is of particular interest. The ring is a graph on $V = \{1, \ldots, n\}$ with edges $\{i, i + 1\}$, $1 \leq i \leq n - 1$, between neighbored vertices and the turn-around edge $\{n, 1\}$. Building blocks are also blocks in the string (if the positions 1 and n are considered as neighbored) and two-point crossover can cut out a building block. Extensive experiments on GAs for this problem have been reported by van Hoyweghen [9], van Hoyweghen, Goldberg, and Naudts [12], and van Hoyweghen, Naudts, and Goldberg [11]. These papers contain also discussions how the algorithms work and some theoretical results but no run time analysis. In recent years, the rigorous run time analysis of EAs has led to interesting results. Here, this approach is applied to the Ising model on the ring.

Sections 2, 3, and 4 analyze mutation-based algorithms. Experiments have led to the conjecture that these algorithms are quite inefficient for the Ising model. The authors of the papers mentioned above do not explicitly state such a conjecture but they and many others have argued in discussions that mutation-based EAs will need exponential optimization time. In Section 2, we analyze randomized local search (RLS) flipping one bit per step and applying a plus-strategy for selection. This simple algorithm finds the optimum in an expected number of $O(n^3)$ steps and the constants in the O-term are surprisingly small. Based on this analysis, a similar bound is obtained in Section 3 for the (1+1) EA.

In Section 4, we analyze parallel variants of the algorithms, parallel RLS (PRLS) and the (1+λ) EA, respectively. They produce λ offspring per generation and select a best individual. For $\lambda = n/\log n$, the expected optimization time consists of $O(n^2 \log n)$ generations and $O(n^3)$ fitness evaluations. This analysis follows the line of research started by Jansen and De Jong [5] and Jansen, De Jong, and Wegener [6]. In Section 5, we compare our results with the experiments on GAs.

It would be even more interesting to obtain also bounds on the expected optimization time of GAs. We are not able to do this for the GAs used in experiments which apply an island model to preserve diversity. We analyze in Section 6 the GA introduced by Culberson [1] and known as GIGA (Gene Invariant GA) and in Section 7 an idealized GA with fitness sharing. Both algorithms are tailored to cope with the given problem and perform better than RLS and the (1+1) EA. We finish with some conclusions.

2 Randomized Local Search

Randomized Local Search (RLS) chooses the first search point $x \in \{0,1\}^n$ uniformly at random. Afterwards, it chooses a position $i \in \{1,\dots,n\}$ uniformly at random, computes x' by flipping bit i of x, and replaces x by x' iff $f(x') \geq f(x)$. We are interested in the expected number of f-evaluations until $x \in \{0^n, 1^n\}$.

Instead of maximizing f, we investigate the equivalent problem of minimizing the number i of monochromatic blocks on the ring. This number is even for non-optimal points and has to be decreased from at most n to 1. For $2 \leq i \leq n$ and i even, let $t_i(n)$ be the expected time until i is decreased if we start with a worst search point with i blocks. We estimate the expected run time by the sum of all $t_i(n)$ and the term 1 for the initialization step.

By the pigeon-hole principle, there is one block whose length is bounded above by $N := \lfloor n/i \rfloor$. We investigate a shortest block B of the first search point x. If i is not decreased, the length of B can change at most by 1 per step. We distinguish relevant steps (either decreasing i or changing the length of B) from the other steps called non-relevant. First, we only investigate the relevant steps. It is possible that some block $B' \neq B$ gets shorter than B and vanishes earlier. Pessimistically, we ignore this. Only if B grows to length $N+1$ we switch our interest to another block whose length is at most N. Pessimistically, we assume that this length equals N. Then we obtain the following Markoff chain on $\{0,1,\dots,N\}$ where the state j describes the length of the considered block. If $j \in \{2,\dots,N-1\}$, by symmetry, the transition probability $p(j,j-1) = p(j,j+1) = 1/2$. By the discussion above, state "$N+1$" is replaced by N and $p(N,N-1) = p(N,N) = 1/2$. State 1 is untypical, since there are two bits whose flip increases the block length but only one decreasing it. Hence, $p(1,0) = 1/3$ and $p(1,2) = 2/3$. Let $T_N(j)$ be the expected time until reaching state 0 when starting in state j.

Lemma 1. *If $j \geq 1$, $T_N(j) = 4N - 1 + N(N-1) - (N-j+1)(N-j) = 2Nj + 2N + j - j^2 - 1$.*

Proof. We fix N and omit the index N. We prove by induction that

$$T(j) = 2 \cdot (N - j + 1) + T(j-1),$$

if $j \geq 2$. By the law of total probability,

$$T(N) = 1 + (1/2) \cdot T(N) + (1/2) \cdot T(N-1)$$

implying that $T(N) = 2 + T(N-1)$. If $j < N$, by induction hypothesis,

$$\begin{aligned} T(j) &= 1 + (1/2) \cdot T(j+1) + (1/2) \cdot T(j-1) \\ &= 1 + (N-j) + (1/2) \cdot T(j) + (1/2) \cdot T(j-1). \end{aligned}$$

Solving for $T(j)$, this proves the claim. Finally,

$$\begin{aligned} T(1) &= 1 + (1/3) \cdot T(0) + (2/3) \cdot T(2) \\ &= 1 + (2/3) \cdot (2 \cdot (N-1) + T(1)) \end{aligned}$$

implying that $T(1) = 4N - 1$. This proves the lemma for $j = 1$ and, if $j \geq 2$,

$$T(j) = 2 \cdot (N - j + 1) + 2 \cdot (N - j + 2) + \cdots + 2 \cdot (N - 1) + 4N - 1 \qquad (*)$$

which implies the lemma. □

Equation $(*)$ implies that T_N is monotone increasing and concave, i. e.,

$$T_N(j+1) - T_N(j) \leq T_N(j) - T_N(j-1).$$

In order to estimate the expected number of relevant steps, it is sufficient to sum up all $T_{\lfloor n/i \rfloor}(\lfloor n/i \rfloor), i \in I := \{j \mid 2 \leq j \leq n, j \text{ even}\}$. Since $T_N(N) = N^2 + 3N - 1$, we obtain

$$\begin{aligned}\sum_{i \in I} T_{\lfloor n/i \rfloor}(\lfloor n/i \rfloor) &\leq n^2 \sum_{i \in I}(1/i^2) + 3n \sum_{i \in I}(1/i) - \lfloor n/2 \rfloor \\ &\leq 0.411 \cdot n^2 + 1.5 \cdot n \ln n + n.\end{aligned}$$

We are interested in the probability that a step is relevant. There are 4 positions such that the length of B changes if one of the corresponding bits flips and the length of B is at least 2. If B has length 1, there are only 3 such positions. The expected waiting time until one of k bits flips is exactly n/k. In order to get good bounds, we estimate the expected number of relevant steps where the block length equals 1. Since the probability of reaching state 0 and finishing a phase equals 1/3, the expected number of steps in state 1 equals 3 independent of i. Hence, $(3/2)n$ of the relevant steps have to be multiplied by $n/3$ and the other ones by $n/4$ to obtain an upper bound on the expected run time.

Theorem 1. *The expected number of steps until RLS finds an optimum for the Ising model on the ring is bounded above by*

$$T_{\text{RLS}}(n) = 0.103 \cdot n^3 + 0.375 \cdot n^2 \cdot (\ln n + 1).$$

This bound is pessimistic in the following aspects:

- the first search point can have less than the maximal number of blocks,
- the first search point with i blocks can contain a block which is shorter than $\lfloor n/i \rfloor$,
- larger blocks can get shorter than the considered block.

In any case, the bound of Theorem 1 is surprisingly small when considering the discussions about this problem. Experiments have shown that, in the case $i = 2$, the shorter block has an average block length of $0.28n$ when reaching this phase. It is easy to obtain the following result.

Theorem 2. *Starting with two blocks of length εn and $(1-\varepsilon)n$, $0 < \varepsilon \leq 1/2$ a constant, the expected number of steps until RLS finds an optimum for the Ising model on the ring is $\Theta(n^3)$.*

3 The (1+1) EA

The (1+1) EA can be considered as the simplest evolutionary algorithm. It works like RLS with the exception of the search operator. The mutant x' is obtained from x by flipping each bit of x independently of the others with probability $1/n$. Steps without flipping bits do not count since they do not lead to a fitness evaluation. Let $e = 2.718\ldots$ be the Eulerian constant.

Theorem 3. *The expected number of steps until the* (1+1) *EA finds an optimum for the Ising model on the ring is bounded above by* $T_{(1+1)}(n) = (e-1) \cdot (1 + o(1)) \cdot T_{\mathrm{RLS}}(n) \le 0.177 \cdot n^3 + o(n^3)$.

Proof. We use the same ideas as in the proof of Theorem 1. In particular, we concentrate our analysis on the length of one block and we consider first only relevant steps, i. e., steps changing the length of the chosen block. We investigate another block if the chosen block has a length larger than $\lfloor n/i \rfloor$. The main idea is that we do not estimate the number of steps directly but we compare the (1+1) EA with RLS. For this purpose, we investigate some stochastic processes "between" RLS and the (1+1) EA.

We start with RLS* which applies the search operator of the (1+1) EA but only mutants x' where exactly one bit has flipped are considered for selection. Then the expected run time increases by the expected waiting time for a step flipping exactly one bit. This waiting time equals $e-1$ for a Poisson distribution with $\lambda = 1$. Here we obtain a factor of $(e-1) \cdot (1+o(1))$ since the number of flipping bits is asymptotically Poisson distributed. This indeed is the essential factor why the (1+1) EA is slower than RLS. If the number of blocks is not too large, the probability that a step flipping more than one bit is relevant is much less than the corresponding probability for steps flipping one bit. The reason is that the other flipping bits typically increase the number of blocks.

Nevertheless, there are relevant steps flipping more than one bit and there are relevant steps changing the length of the considered block by more than 1. For each search point x let $p_k^+(x)$ be the probability that the next step is accepted and produces a search point where the length of the considered block B has been increased by k and let $p_k^-(x)$ be the corresponding probability for decreasing the length of B. We know from Section 2 that $p_k^+(x)$ may be larger than $p_k^-(x)$. To simplify the analysis, we investigate two further stochastic processes called (1+1) EA$_{\mathrm{sym}}$ and RLS$_{\mathrm{sym}}$. They are based on the algorithms (1+1) EA and RLS, respectively, but, if $p_k^+(x) > p_k^-(x)$, the probability of increasing the length of B is decreased to $p_k^-(x)$. As before, we consider another block if the length of B is larger than $\lfloor n/i \rfloor$. First, RLS$^*_{\mathrm{sym}}$ is obviously faster than RLS*. We show that the expected run time of the (1+1) EA$_{\mathrm{sym}}$ is bounded by the upper bound proven for RLS* and, therefore, also for RLS$^*_{\mathrm{sym}}$ and, later, we compare the (1+1) EA and the (1+1) EA$_{\mathrm{sym}}$.

Let A_t be the algorithm working t steps like the (1+1) EA$_{\mathrm{sym}}$ and afterwards like RLS$^*_{\mathrm{sym}}$. We prove by induction on t that the expected run time of A_t is not larger than the upper bound obtained for RLS*. This is true for $t = 0$, since $A_0 = \mathrm{RLS}^*_{\mathrm{sym}}$. For the induction step, we compare A_t and A_{t+1}. They are identical for the first t steps and we consider the (random) search point x after t

steps. The probability of a relevant step is for the (1+1) $\mathrm{EA}_{\mathrm{sym}}$ not smaller than for $\mathrm{RLS}^*_{\mathrm{sym}}$. We compare the algorithms conditioned on some events and prove the claim for each of the cases. If the next step is neither relevant for A_t nor for A_{t+1}, the claim is obvious since the upper bound for RLS* only depends on the length of the considered block. The perhaps larger probability of a relevant step of A_{t+1} is only in favor of A_{t+1}. Finally, we have to compare the effect of relevant steps. Instead of having steps changing the length of B by $+1$ and -1 (with the same probability), we now may change the length of B by $+k$ and $-k$ (with the same probability). Afterwards, we apply $\mathrm{RLS}^*_{\mathrm{sym}}$ in both cases. The upper bound for RLS* (and also $\mathrm{RLS}^*_{\mathrm{sym}}$), namely the function T_N from Section 2 is increasing and concave. Therefore, a $\pm k$-step instead of a ± 1-step reduces the expected run time, i. e., $(T(j+k)+T(j-k))/2 < (T(j+1)+T(j-1))/2$, if $k \geq 2$. For $t \to \infty$, we obtain the claim.

Finally, we have to compare the (1+1) EA and the (1+1) $\mathrm{EA}_{\mathrm{sym}}$. We investigate a phase of length $n^{7/2}$. By Markoff's inequality, the probability that the (1+1) $\mathrm{EA}_{\mathrm{sym}}$ needs more than $n^{7/2}$ steps is $O(n^{-1/2}) = o(1)$. In this case, we repeat the arguments for the next phase leading to an additional $1+o(1)$ factor. In the following, we investigate a phase of length $n^{7/2}$. Events which altogether have a probability of $o(1)$ can be ignored since then the phase can be considered as unsuccessful also leading to a $1+o(1)$ factor.

Let k be the length of the considered shortest block B, w. l. o. g. a block of ones. If $k \geq 4$, the string contains $0^4111^{k-4}110^4$. We consider the substrings 0^411 and 110^4. The probability that a phase contains a step with at least four flipping bits at these positions is $o(1)$ and this event can be ignored. Steps with at most three flipping bits at these positions do not eliminate one of the blocks. The situation is symmetric with respect to lengthenings and shortenings of B.

We are left with the situation $k \leq 3$. Recalling the analysis of RLS in Section 2, it is easy to obtain the result that the $(1+1)^*\mathrm{EA}_{\mathrm{sym}}$ has an expected number of $O(n)$ steps where $k \leq 3$. By Markoff's inequality, we can ignore runs where this number is larger than $n^{3/2}$. The probability that a phase contains a step with at least two flipping bits in the substring $0^k1^k0^k$ is $o(1)$.

Finally, decreasing the length of B from k to 0 does not imply that we decrease the number of blocks. A new block may be created somewhere else. The probability of no bit flipping elsewhere is at least e^{-1}. Hence, with a probability of $1-e^{-1}$ we are still in the same situation as before, i. e., we have the same values of $k \in \{1,2\}$. This happens on average $e/(e-1)$ times, each time increasing the expected run time by $O(n^2)$. Hence, we have proved the theorem. □

It is worth noticing that we were not able to prove such a small bound by analyzing the (1+1) EA directly. It was helpful to analyze the simpler algorithm RLS and to compare RLS and the (1+1) EA.

Finally, we prove a lower bound similarly to the lower bound of Theorem 2.

Theorem 4. *Starting with two blocks of length εn and $(1-\varepsilon)n$, $0 < \varepsilon \leq 1/2$ a constant, the expected number of steps until the $(1+1)$ EA finds an optimum for the Ising model on the ring is $\Theta(n^3)$.*

We omit the proof here. The essential argument is that the function T_N from Lemma 1 is concave but the curvature is not strong. In particular, $T_N(j-k)+$

$T_N(j+k) = T_N(j-1) + T_N(j+1) - c_k$ where c_k only depends on k. Since steps changing the block length by k have a probability of $\Theta(n^{-k})$, we do not save too much by steps changing the block length by at least 2.

4 Parallel RLS and the (1+λ) EA

A GA works with a population of $s(n)$ individuals and, in most cases, run time is defined as the number of generations. The number of fitness evaluations is larger by a factor of $s(n)$. For RLS and the (1+1) EA, the number of generations equals the number of fitness evaluations. In order to have a fair comparison with GAs, we consider population-based RLS and (1+1) EA. Parallel RLS (PRLS) or (1+λ) RLS creates λ children from the parent x using the search operator of RLS. The children are created independently. Selection chooses x if all children are worse and chooses one of the fittest children uniformly at random otherwise.

For $s(n) = n$ we get an expected number of $O(n^2 \log n)$ generations and $O(n^3 \log n)$ fitness evaluations. While reducing $s(n)$ to $n/\log n$ does not affect the number of generations, it reduces the number of fitness evaluations to $O(n^3)$.

Theorem 5. *The expected number of generations until* $(1+\lfloor n/\log n \rfloor)$ *RLS finds the optimum for the Ising model on the ring is bounded above by* $O(n^2 \log n)$ *and the expected number of fitness evaluations by* $O(n^3)$.

Proof. It is sufficient to investigate the number of generations since each generation consists of $\lfloor n/\log n \rfloor$ fitness evaluations. Again, let B be the considered block. The probability that no child shortens or lengthens B equals $(1 - c/n)^{\lfloor n/\log n \rfloor} = 1 - \Theta(1/\log n)$. The expected waiting time for a generation with a child changing B equals $\Theta(\log n)$. If x contains i blocks, the expected number of children with the same number of blocks as x is $\Theta(i/\log n)$ and the probability that this number is bounded by $O(i/\log n)$ is at least 1/2 (Markoff's inequality).

If $i \geq \log n$, the probability of choosing a child where B is changed, if such a child is created, is $\Omega(\log n/i)$. The waiting time for such a step is $O(i/\log n)$. Hence, each step has a probability of $\Omega(1/i)$ of being relevant. By Lemma 1, the expected number of relevant steps to decrease i is $O(n^2/i^2)$ and this takes $O(n^2/i)$ generations on average. For all i, $\log n \leq i \leq n$ and i even, we obtain a bound of $O(n^2 \log n)$.

If $i < \log n$ and one child changes B, the probability that all other children have more blocks equals $(1 - \Theta(i/n))^{\lfloor n/\log n \rfloor - 1}$ which is bounded below by a positive constant. Then the generation is relevant. Hence, the expected waiting time for a relevant generation equals $\Theta(\log n)$ and the expected number of generations is bounded by $O((n^2 \log n)/i^2)$. Considering all $i < \log n$ and even, this gives an additional term of $O(n^2 \log n)$. □

The (1+λ) EA applies the search operator of the (1+1) EA and produces independently λ children from the parent which is the only individual of the current population. We have to be careful with the selection operator. It is likely that many children are a replica of the parent. In order to guarantee exploration of the search space, we select the parent x only if all children $y \neq x$ have a worse fitness than x. Otherwise, we randomly select an individual among the fittest children $y \neq x$.

Combining the methods from Section 3 and Theorem 5 we obtain the following result.

Theorem 6. *The expected number of generations until the* $(1 + \lfloor n/\log n \rfloor)$ *EA finds the optimum for the Ising model on the ring is bounded above by* $O(n^2 \log n)$ *and the expected number of fitness evaluations by* $O(n^3)$.

5 A Comparison with GA Experiments

We have no doubt that crossover can play an essential role for the Ising model on the ring. A theoretical fundament for this argument will be presented in Sections 6 and 7. Here, we want to argue that mutation-based EAs are better than expected in many papers. Van Hoyweghen [9] claims that "the presence of spin-flip symmetry in the one-dimensional Ising model prevents an unspecialized GA to find an optimum in a reasonable amount of time." Van Hoyweghen, Goldberg, and Naudts [10] indicate in this context that "the Ising model shows that for a certain class of optimization problems niching becomes a necessity for a GA to solve these problems." Our results have shown that unspecialized EAs solve this problem in reasonable time. The upper bounds on the expected run times of RLS ($0.103n^3 + 0.375n^2(\ln n + 1)$ and even $117,957$ for $n = 100$) and of the (1+1) EA (by a factor of 1.72 slower than RLS) show this even for populations of size 1. The time bounds are much better, namely $O(n^2 \log n)$, if $n/\log n$ children are generated in parallel. Hence, the optimization is finished in a reasonable amount of time without any niching. Van Hoyweghen [9] has considered the case of GAs for $n = 100$ and a population size of 100. The best parameters for tournament selection and two-point crossover lead to an average number of $35,857$ generations. This can be decreased to $10,881$ using SAWing (Stepwise Adaptation of Weights). With an Island model and a distributed GA there is a good chance that $10,000$ generations suffice. In all these cases a population of size $s(n) \geq 100$ is used. In general, it is claimed that a population size of $10.9n^{0.57}$ suffices. These algorithms need less generations than the mutation-based algorithms examined in this paper but they do not beat RLS with respect to the expected number of fitness evaluations (at least for $n = 100$).

6 The Expected Run Time of GIGA

Although mutation-based algorithms are surprisingly efficient for the Ising model on a ring, it is believed that GAs can be faster. It is difficult to analyze the effect of crossover if one is interested in the expected optimization time. We are not able to analyze distributed GAs. Therefore, we analyze GAs which are specialized to work on the Ising model on the ring. We cannot expect to obtain the same good time bounds for unspecialized GAs.

In this section, we analyze a simple variant of GIGA (Gene Invariant Genetic Algorithm) introduced by Culberson [1]. The population has size 2 and consists of a search point $x \in \{0,1\}^n$ and its bitwise complement $\overline{x}$. In the initialization step, x is chosen uniformly at random. Later, a new pair of search points $(y, \overline{y})$ is produced from $(x, \overline{x})$ by crossover. Since $f(x) = f(\overline{x})$, the new pair $(y, \overline{y})$ replaces $(x, \overline{x})$ if $f(y) \geq f(x)$. Since we want to cut out a block in x and to replace it by its bitwise complement, two-point crossover seems to be the appropriate recombination operator. Let us consider the effect of crossover at the positions j and k, $0 \leq j < k < n$. A position p is called border of x, if $x_p \neq x_{p+1}$ or $x_n \neq x_1$ if $p = 0$. Let i be the number of blocks of x.

Case 1: The positions j and k are not borders. Then y has $i+2$ blocks and $(y, \overline{y})$ is not accepted.
Case 2: Exactly one of the positions j and k is a border. Then y also has i blocks and $(y, \overline{y})$ is accepted but the fitness is not changed.
Case 3: The positions j and k are borders. If $i > 2$, y has $i-2$ blocks. If $i = 2$, y has one block. In any case, $(y, \overline{y})$ is accepted and the fitness is improved.

As long as x is not optimal, $i \geq 2$ and there are $\binom{i}{2}$ among $\binom{n}{2}$ pairs of positions which lead to an improved fitness. Hence, the expected optimization time can be bounded above by (remember that $I = \{i \mid 2 \leq i \leq n, i \text{ even}\}$)

$$\sum_{i \in I} \binom{n}{2} \Big/ \binom{i}{2} \leq 0.70 \cdot n \cdot (n-1).$$

With probability $1 - o(1)$, the initial value of i is at least $n/3$. We obtain the following results.

Theorem 7. *The expected number of steps until GIGA with two-point crossover finds an optimum for the Ising model on the ring is bounded above by* $0.70 \cdot n \cdot (n-1)$ *and bounded below by* $0.69 \cdot n^2 - o(n^2)$.

We can generalize GIGA to (1+λ) GIGA where λ offspring pairs are produced independently and a best one is chosen if it is not worse than the parent. We analyze the (1+n) GIGA. The probability of producing a better offspring is bounded below by a positive constant, if $i > n^{1/2}$, and by $\Omega(i^2/n)$, otherwise. Hence, the expected number of generations equals $\Theta(n)$.

Theorem 8. *The expected number of generations until the* $(1+n)$ *GIGA with two-point crossover finds an optimum for the Ising model on the ring equals* $\Theta(n)$*, the expected number of fitness evaluations equals* $\Theta(n^2)$.

Surprisingly, one-point crossover is almost as good as two-point crossover. The probability that two consecutive steps with one-point crossover decrease i is $\Theta(i^2/n^2)$ as for one step of two-point crossover. This leads to the following result.

Theorem 9. *The expected number of fitness evaluations until GIGA or the* $(1+n)$ *GIGA with one-point crossover finds an optimum for the Ising model on the ring equals* $\Theta(n^2)$*. For the* $(1+n)$ *GIGA, the number of generations equals* $\Theta(n)$.

7 The Expected Run Time of a GA with Fitness Sharing

The variant of GIGA analyzed in Section 6 is highly specialized. Diversity in the population of size 2 is guaranteed by choosing always individuals with the maximal Hamming distance. Here, we consider a GA with the unusually small population size 2 where diversity is supported by fitness sharing. Populations are multisets. In fitness sharing, the closeness of x and y is measured by

$$S(x, y) := \max\{1 - d(x, y)/\sigma, 0\}$$

where d is an appropriate distance measure and σ is a critical value deciding when x and y are so far from each other that they do not share fitness. In our case, d is the Hamming distance and $\sigma := n$ since we like to produce individuals with large Hamming distance. Then, for population P, $S(x, P)$ is the sum of all $S(x, y)$, $y \in P$. The shared fitness of x in the population P is defined by

$$f(x, P) := f(x)/S(x, P)$$

if f is the real fitness. Finally, $f(P)$ is defined as the sum of all $f(x, P)$, $x \in P$.

The following GA applies two-point crossover to produce two children and mutations flipping each bit independently with probability $1/n$.

Algorithm 1 *(Steady-state GA with population size 2 and fitness sharing)*

```
1.)  The initial population $P$ consists of two individuals chosen independently
     and uniformly at random.
2.)  Selection for reproduction. Both individuals $x$ and $y$ are chosen.
3.)  Offspring creation. One of the Steps 3a and 3b is chosen uniformly at random.
3a.) $x' := \text{mutate}(x),\ y' := \text{mutate}(y),\ P' := P \cup \{x', y'\}$.
3b.) $(\tilde{x}, \tilde{y}) :=$ two-point-crossover$(x, y),\ x' := \text{mutate}(\tilde{x}),$
     $y' := \text{mutate}(\tilde{y}), P' := P \cup \{x', y'\}$.
4.)  Selection of the next generation. Choose a population $P \subseteq P'$ of size 2 with
     the maximal $f(P)$-value.
```

Since we work with populations of very small size, it is not too time-consuming to choose in Step 4 a population with the largest f-value. This reflects the real idea behind fitness sharing. The shared fitness of the population should be large.

Let the population P consist of the individuals x and y with a Hamming distance of $d = d(P)$. Let $i(z)$ be the number of borders within the individual z and let $i = i(P) := i(x) + i(y)$. Then $f(z) = n - i(z)$ and

$$f(x, P) = \frac{n - i(x)}{1 - H(x, x)/n + 1 - H(x, y)/n} = \frac{n - i(x)}{2 - d/n}$$

and

$$f(P) = \frac{2n - i}{2 - d/n}.$$

Hence, we can increase P by decreasing i and/or by increasing d. As long as we do not decrease i, we hope to increase d. If $d = n$, we have two complementary individuals and two-point crossover is a good operator to decrease i. Since $0 \leq f(P) \leq 2n$ and f cannot decrease because of the plus-strategy for selection, we try to analyze the expected time until f has been increased at least by a constant additive term c. For this purpose, we classify the possible populations P:

- type OPT contains all P where $i \leq 1$, i. e., at least one individual is optimal,
- type $A(i)$, $i \geq 2$, contains all P where $i = i(P)$ and $d = n$,
- type B contains all P where $2 \leq i \leq n$ and $d < n$, and
- type C contains all P where $i > n$ and $d < n$.

Theorem 10. *The expected number of fitness evaluations until the steady-state GA with population size 2 and fitness sharing finds an optimum for the Ising model on the ring is bounded above by $O(n^2)$.*

Proof. All populations of type $A(i)$ have the same fitness $2n - i$. After having increased the fitness, we will never accept a population of type $A(i)$. Moreover, if $P = \{x, y\}$ is of type $A(i)$, then $y = \overline{x}$. The expected waiting time until two-point crossover creates a population P' of type $A(i-4)$ is bounded by $O(n^2/i^2)$, see Section 6. Then $f(P') - f(P) = 4$. By standard arguments, the expected time with populations of type A is bounded by $O(n^2)$.

For populations of type B or C, we prove that the probability of increasing the fitness by at least $1/4$ is bounded below by $\Omega(1/n)$. We have to wait for at most $8n$ of such steps which proves the theorem.

Let $P = \{x, y\}$ be of type B. Since $d < n$, $x \neq \overline{y}$. Let j be the rightmost position where $x_j = y_j$. Then $x_{j+1} \neq y_{j+1}$ (where $n+1$ is identified with 1 since we are on a ring). W. l. o. g. $x_j = x_{j+1}$ and $y_j \neq y_{j+1}$. With a probability of $\Omega(1/n)$, we choose Step 3a and only bit j is flipped when producing y'. Then $f(y') \geq f(y)$ and $H(x, y') = H(x, y) + 1$. The population $P' = \{x, y'\}$ is a possible successor population and

$$\begin{aligned} f(P') - f(P) &\geq \frac{2n-i}{2-(d+1)/n} - \frac{2n-i}{2-d/n} \\ &= \frac{(2n-i)\cdot(2-d/n) - (2n-i)\cdot(2-d/n-1/n)}{(2-(d+1)/n)\cdot(2-d/n)} \geq \frac{1}{4}, \end{aligned}$$

since the numerator equals $2 - i/n \geq 1$ and the denominator is at most 4.

Type-C populations can be handled in a similar way. Since $i \geq n$, one individual has a block of length 1 which can be eliminated by a 1-bit flip. □

Finally, we can consider a GA with population size 2 and fitness sharing which produces $P'_1, \ldots, P'_n$ by performing Step 3 n times independently in parallel. Then it selects a population $P \subseteq P'_i$ for some i which has the largest $f(P)$-value.

Theorem 11. *The expected number of generations until the GA with population size 2, fitness sharing, and n pairs of offspring per generation finds an optimum for the Ising model on the ring is bounded above by $O(n)$.*

Conclusions

The Ising model is a good model to analyze the adaptation capabilities of EAs and GAs. In particular, the Ising model on the ring leads to surprising results. Mutation-based algorithms and even randomized local search are much more efficient than expected in the GA community. This is especially true if we consider the number of generations in the case of producing more than one offspring. Nevertheless, recombination can decrease the expected optimization time. This has been proved rigorously for two specialized GAs which work with very small populations. It is an open problem to analyze generic GAs with niching for the Ising model on the ring.

References

1. J. Culberson. Genetic invariance: A new paradigm for genetic algorithm design. Technical Report 92–02, University of Alberta, 1992.
2. D. E. Goldberg. *Genetic Algorithms in Search, Optimization, and Machine Learning*. Addision-Wesley, Reading, MA, 1989.
3. J. H. Holland. *Adaptation in Natural and Artificial Systems*. University of Michigan, MI, 1975.
4. E. Ising. Beiträge zur Theorie des Ferromagnetismus. *Z. Physik*, (31):235–258, 1925.
5. T. Jansen and K. De Jong. An analysis of the role of offspring population size in EAs. In *Proc. of the Genetic and Evolutionary Computation Conference (GECCO 2002)*, pages 238–246, San Mateo, CA., 2002. Morgan Kaufmann.
6. T. Jansen, K. De Jong, and I. Wegener. On the choice of the offspring population size in evolutionary algorithms. *Submitted to Evolutionary Computation*, 2003.
7. B. Naudts and J. Naudts. The effect of spin-flip symmetry on the performance of the simple GA. In *Proc. of 5th Conf. on Parallel Problem Solving from Nature (PPSN - V)*, number 1498 in LNCS, pages 67–76, 1998.
8. M. Pelikan and D. E. Goldberg. Hierarchical BOA solves Ising spin glasses and MAXSAT. In *Proc. of the Genetic and Evolutionary Computation Conference (GECCO 2003)*, number 2724 in LNCS, pages 1271–1282, 2003.
9. C. Van Hoyweghen. *Symmetry in the Representation of an Optimization Problem.* PhD thesis, Univ. Antwerpen, 2002.
10. C. Van Hoyweghen, D. E. Goldberg, and B. Naudts. Building block superiority, multimodality and synchronization problems. Technical report, Illinois Genetic Algorithms Laboratory, 2001. IlliGAL Rep. No. 2001020.
11. C. Van Hoyweghen, D. E. Goldberg, and B. Naudts. From twomax to the Ising model: Easy and hard symmetrical problems. In *Proc. of the Genetic and Evolutionary Computation Conference (GECCO 2002)*, pages 626–633, San Mateo, CA, 2002. Morgan Kaufmann.
12. C. Van Hoyweghen, B. Naudts, and D. E. Goldberg. Spin-flip symmetry and synchronization. *Evolutionary Computation*, (10):317–344, 2002.