Presented at EC'94, The IEEE Conf. on Evolutionary Computation, part of the IEEE World Conf. on Computational Intelligence, June 27 - June 29, 1994.

Fitness Landscapes and Difficulty in Genetic Programming

Kenneth E. Kinnear, Jr.
Adaptive Computing Technology
62 Picnic Rd.
Boxboro, MA 01719
kim.kinnear@adapt.com

***Abstract*—The structure of the fitness landscape on which genetic programming operates is examined. The landscapes of a range of problems of known difficulty are analyzed in an attempt to determine which landscape measures correlate with the difficulty of the problem. The autocorrelation of the fitness values of random walks, a measure which has been shown to be related to perceived difficulty using other techniques, is only a weak indicator of the difficulty as perceived by genetic programming. All of these problems show unusually low autocorrelation. Comparison of the range of landscape basin depths at the end of adaptive walks on the landscapes shows good correlation with problem difficulty, over the entire range of problems examined.**

I. INTRODUCTION

Genetic Programming, like all algorithms which depend on some form of evolutionary adaptation, operates within the context of a *fitness landscape.* The concept of a fitness landscape, introduced by Sewell Wright [1932], refers to the mapping from the genomes of a population of individuals to their fitness, and a visualization of that mapping. Recently there has been renewed interest in the examination of the landscape characteristics of various problems. In this work, the landscape characteristics of several specific problems presented to a genetic programming system are analyzed.

Analysis of the structure of the fitness landscapes is of both theoretical and practical interest, in that it might allow us to relate the difficulty of the problem to the structure of the landscape. The analysis of the landscape is frequently computationally more tractable than simply trying to solve the problem with genetic programming.

A. Genetic Programming

Programming paradigms that exploit evolutionary techniques are becoming more common. The techniques used in the work described here derive from Genetic Algorithms (GA) developed by John Holland [1975,1992]. Genetic Algorithms typically work on fixed length, linear representations of genetic material, and operate on this material with fitness proportionate selection, reproduction, crossover, and mutation [De Jong 1987, Goldberg 1989]. John Koza has extended the work done by Holland and many others in the GA field by using analogies of GA genetic operators directly on tree structured programs (typically LISP functions) and calls this approach Genetic Programming (GP) [Koza 1992]. GP differs from traditional GA in that it uses hierarchical genetic material, and this genetic material is also not of fixed size. While GP doesn't mimic nature as closely as does GA, it does offer the opportunity to directly evolve programs of unusual complexity, without having to define the structure or the size of the program or genetic material in advance.

B. Introduction to Fitness Landscapes

What is a fitness landscape? First, imagine the space of all possible programs that can be generated by a particular genetic programming system applied to a particular problem. With each program is associated a *fitness*, a real number which reflects how well this program solves the problem. Then, imagine that this space of all possible programs is mapped into the x-y plane. Finally, imagine that the fitness of each one of these programs is plotted on the z-axis. This creates a surface where the peaks are the locations of programs with poor fitness, and the basins show the locations of the programs with good fitness. Discovering the best solution to the problem then becomes equivalent to searching over this landscape for the deepest basin.

The initial random population of a genetic programming run locates individuals at widely separated points on this landscape. Each point corresponds to a single individual program in the population. Each succeeding generation creates a number of new individuals from the existing ones, and locates new points on this fitness landscape — points that are lower down in deeper basins, if the process is one that is effectively progressing toward solution.

The particular mapping from the space of programs into the x-y plane as part of the visualization process of a landscape is important. If each point in the x-y plane represents a separate individual program, then which programs are close to which other programs? Programs that are close together on the x-y plane are those that are most likely to be created from one another through the operation of the particular genetic operators and representation employed. It follows that the genetic operators have a large effect on determining the fitness landscape for a particular problem and the *neighbors* of a particular individual on the x-y plane of the landscape are those which can be generated by one operation of the genetic operators [Altenberg 1994, Weinberger 1990].

This highlights a key aspect of evolutionary adaptation, the *correlation* of an offspring's fitness to that of the parents'. There must be some correlation of fitness between offspring and parents for any sort of evolutionary adaptation to proceed

— that is to say matings of parents with good fitness should produce, generally, offspring of good fitness. If there is no variation between the fitness of the offspring and parents, then there is no opportunity for improvement as the adaptive process proceeds. If, on the other hand, there is no correlation at all, then selecting good parents to create the offspring will have no effect, and the process becomes one of random search where selecting parents is just a way to randomize the search, and the search is not directed in any way by the parents' fitness. This correlation is directly represented in a fitness landscape by the *ruggedness* of the landscape.

The ruggedness of the landscape has a direct bearing on the difficulty of the problem. Evolutionary techniques such as genetic programming will have increasing difficulty in locating the deepest basins on landscapes that display greater ruggedness. It is perhaps easiest to visualize in the opposite case, that of a landscape with a single large basin, the bottom of which is the best solution. In this case, it is relatively easy for the adaptive processes to proceed directly to the bottom of the basin.

In the opposite case, where the landscape is quite rugged, perhaps even locally or globally discontinuous, the individuals may get trapped on local minima without every finding the (or a) global minimum, or the population may simply wander aimlessly across the landscape, in which case the adaptive process has degenerated into random search.

C. Recent Work in Landscape Analysis

The introduction by Stuart Kauffman of landscapes with *tunable ruggedness* [Kauffman 1989, 1993] marks the beginning of renewed interest in landscape analysis. These landscapes are termed the *NK* landscapes. Briefly, the individuals which populate these landscapes consist of fixed length bit strings of length *N*, and the fitness of each individual is based on the values of the bits contained within it in a particular way. The contribution of each bit to the fitness of the individual is based not only on its own value, but on the value of *K* other bits (where *K* is constant for a given population). Kauffman has determined that if K=1 that the landscape is smooth with a single large basin, and that as *K* increases towards *N* that the ruggedness increases, until when K=N, the landscape is completely rugged and any process that uses evolutionary adaptation will be equivalent to random search. These results are independent of the particular assignment of fitness values to bits in the individuals.

Edward Weinberger [1990, 1991] has performed groundbreaking work in landscape analysis, using the *NK* family of landscapes to investigate how the autocorrelation of the fitness of individuals along the steps of a random walk relates to the ruggedness of the landscapes. He has done extensive analyses of the distribution and depth of local optima as well as a wide variety of other landscape features.

Peter Stadler and Wolfgang Schnabl [1992] have investigated the correlation structure proposed by Weinberger for the Travelling Salesman Problem. Peter Stadler and Robert Happel [1992] have investigated the correlation structure of the graph bi-partitioning problem, and Stadler [1992] has investigate the low autocorrelation binary string problem and the matching problem.

Examination of the autocorrelation of random walks for a very different landscape, that of RNA folding, has been reported by Walter Fontana, Peter Schuster and Peter Stadler along with several others [Fontana, et. al. 1993a, 1993b], [Schuster, et. al. 1993].

Bernard Manderick, Mark de Weger, and Piet Spiessens [1991] have applied landscape analysis techniques to the fitness landscapes of genetic algorithms for the *NK* landscapes and the Travelling Salesman Problem. They have examined the autocorrelation function of the landscapes, determined the correlation length, and explored the fitness correlation of various genetic operators.

Marc Lipsitch has generated a range of landscapes by iterations of an elementary cellular automaton. A genetic algorithm was then used to simulate adaptation on these landscapes with interesting results [Lipsitch 1991].

John Koza [1994] has analyzed the distribution of fitness values in a large number of randomly generated individuals for several problems with and without the use of Automatically Generated Functions (ADF), an idea proposed by Nick Keenan.

II. EXPERIMENTAL APPROACH

The work reported here investigates the connection between the empirically determined features of the landscape and the difficulty of the problem as perceived by genetic programming.

A. Problems to Analyze

The problems analyzed in this work fall into two classes. The first class is concerned with evolving general sorting algorithms capable of sorting any sequence of numbers. Four different examples of this problem are used, covering a wide range of difficulty. This problem was first introduced in [Kinnear 1993a] and the relative difficulty of the various forms of the problem is examined in [Kinnear 1993b].

The second class of problems used are the boolean even parity problems. In these problems, the output is `t` if an even number of the inputs are `t`. Variations of 3, 4, and 6 inputs create progressively more difficult problems. Versions of the 3, 4, and 6 even-parity problems using Automatically Defined Functions (ADF) were also used for comparison and are substantially easier. For more details on this class of problems see [Koza 1992, 1993, 1994], and [Kinnear 1994].

The genetic operators used for each problem were crossover (.70 probability), mutation, create, and hoist (.10 probability each). See [Kinnear 1993a, or 1994] for details.

B. Comparing Difficulty

An intuitive approach to the comparison of difficulty is to compare the *cumulative probability of success* (*CPS*) for each problem. The CPS is simply the probability that a solution to the problem is found by running the problem to a particular generation. The CPS curves for all of the problems reported here (except the 6 even-parity, which is so difficult that no solutions were generated in 20 runs with a population size of 1000) are shown in Figure 1.

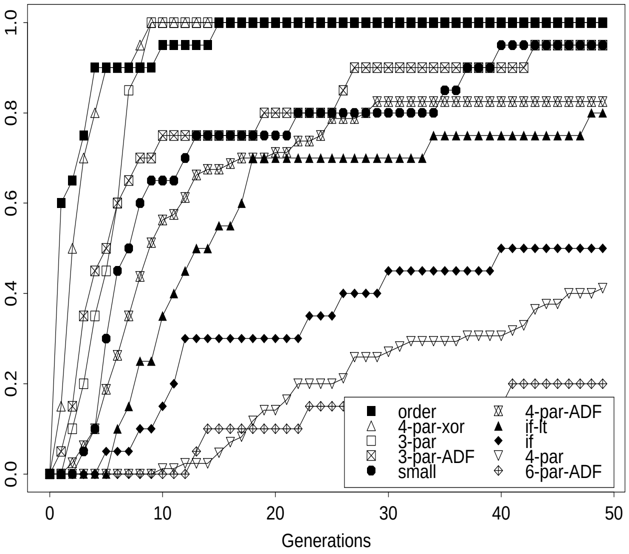

Figure 1: Relative Difficulty

While this provides a visual way to separate the easy from the hard, it doesn't yield a quantifiable result. The quantifiable relative difficulty used is the *number of individuals that must be processed to yield a solution with 99% probability*, after [Koza 1992, 1993]. The *difficulty index* is simply the number of individuals required divided by 1000, shown in the table below.

Problem	Difficulty Index (K individuals)	Class
order	10	sorting
4-par-xor	12	even-parity
3-par	18	even-parity
3-par-ADF	36	even-parity-ADF
small	50	sorting
4-par-ADF	65	even-parity-ADF
if-lt	76	sorting
if	169	sorting
4-par	448	even-parity
6-par-ADF	660	even-parity-ADF
6-par	very large	even-parity

III. LANDSCAPE AUTOCORRELATION

A. Definition

The most common landscape analysis technique is to examine the autocorrelation of a series of fitnesses of individuals derived from a random walk [Weinberger 1990]. In this approach, adapted from Weinberger for a situation with multiple parents, a time series of fitness values is produced by selecting an initial parent with uniform random selection and recording its fitness. Then a genetic operator is selected with whatever selection technique is usually employed, in this case a probabilistic one heavily weighted towards crossover. When required, a second parent is selected using uniform random selection. Finally, the genetic operation selected previously is performed. The resulting individual's fitness is evaluated, and recorded as the next step in the series. This individual becomes the new parent and the process is repeated.

The results analyzed in this work used random walk lengths of 1000 individuals, and represent an average over a minimum of 10 (and usually 20) independent walks.

The autocorrelation of the random walks relates the fitness of two individuals which are s steps apart $R(t,s)$:

$$R(t, s) = \frac{E(f_t \times f_{t+s}) - E(f_t) \times E(f_{t+s})}{var(f)}$$

where E is the expected or mean value, f_t is the fitness of the individual at the t-th step, f_{t+s} is the fitness of the individual s steps farther along the walk, and $var(f)$ is the variance of the entire time series. This is calculated for all 1000 - s possible pairs in each walk.

For this technique to be effective, the fitness landscapes must be assumed to be *isotropic,* i.e. the statistics of a random walk on the landscape do not depend on the particular random walk used. The random walks must also be assumed *stationary*, i.e. the joint distribution of a contiguous subset of the random variables that make up the fitnesses recorded in a random walk does not change based on where the subset begins.

B. Results

The autocorrelation of random walks for many of the problems listed in Section III are shown in Figure 2, below. The more rapidly the autocorrelation line falls to zero, the less correlated the landscape, at least as measured by this technique, and presumably the harder the problem is to solve for genetic programming.

There are several things to note in these plots. Most obviously, there is essentially no overall correlation between the difficulty index listed above and the difficulty determined from the curves in Figure 2. Even within the two problem classes, sorting and boolean even-parity, there is limited correlation between difficulty as measured by the difficulty index and difficulty inferred from the autocorrelation plots. Only within the 3,4, and 6 even-parity problem without ADF's is

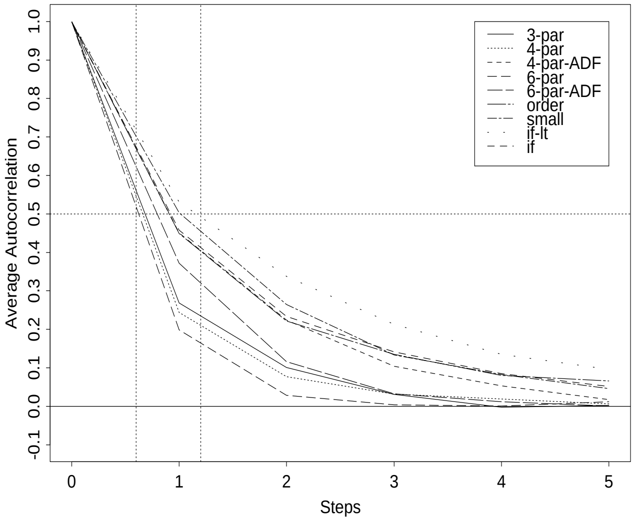

Figure 2: Autocorrelation of Random Walks
Mean of between 10 and 20 walks of 1000 points each.

there a matchup (the hardest three plots, closest to the origin in Figure 2). Within the class of sorting problems, three out of four are indistinguishable, with if-lt looking by far the easiest, when it in fact is not. When comparing this graph with others in the literature, note that the x-axis has been expanded considerably to allow examination of the detail in this graph.

If the results presented in Figure 2 are examined with respect to what Manderick, de Weger, and Spiessens [1991] call the correlation length, i.e. the number of steps for which the autocorrelation falls to 0.5 (the horizontal dotted line) then we can see that all of the problems analyzed above fall into the range of 0.6 to 1.2 steps for their correlation length. This statistic implies that these landscapes are not well correlated at all, and that these should be very difficult problems indeed. Some comparative correlation lengths from the literature are presented here:

Problem	Correlation Length	Reference
Matching Problem (N=100)	17.5	[Stadler 1992]
Low Autocorrelation Binary String N=100.	12.5 (N/8)	[Stadler 1992]
Graph Bi-Partitioning Problem 50 vertices, connectivity 0.2	4	[Stadler and Happel 1992]
NK Landscape with *K* = N/2 N = 96, *K*=48.	1	[Manderick, de Weger, Spiessens 1991]
All problems described here	0.6 to 1.2	

C. *Discussion*

It appears that the landscape autocorrelation is not a particularly good way to determine the relative difficulty between or within the classes of problems that have been examined here. Certainly all of these problems appear to operate in highly uncorrelated landscapes, and thus appear very difficult, based on the rapidity that the autocorrelation falls off toward zero and the resultant short correlation length. It may well be that these *are* in fact very hard problems, especially with the small population sizes used here, and that the landscape autocorrelation technique is unable to discriminate between subtle differences on poorly correlated landscapes. The differences in difficulty of the problems examined in this work may also simply not show up as differences in landscape autocorrelation at all. Lipsitch [1991] documents a case where three similar problems all had similar landscape correlation measures, yet his genetic algorithm found one problem dramatically more difficult.

Table 2
Results for Adaptive Walks, averaged over 2000 walks, maximum of 30 neighbors considered

Problem	Difficulty Index (K individuals)	Original Fitness			Final Fitness			Adaptive Walk Length		
		min.	mean	max.	min.	mean	max.	min.	mean	max
order	10	2240	5069	113900	1640	3817	6455	0.000	5.186	29.000
4-par-xor	12	400	800	1600	0	581	800	0.000	1.595	7.000
3-par	18	200.0	800.3	1200.0	100.0	282.9	400.0	0.000	1.131	4.000
3-par-ADF	36	200.0	400.4	600.0	0.0	323.3	400.0	0.000	0.739	4.000
small	50	2140	5302	69040	1780	4147	6355	0.000	3.734	14.000
4-par-ADF	65	500	801	1100	0.0	738.9	800.0	0.000	0.572	4.000
if-lt	76	1850	5300	68910	1790	3888	6455	0.000	4.077	15.000
if	169	1625	4860	67380	1625	3901	5635	0.000	3.592	14.000
4-par	448	600.0	801.5	1000.0	400.0	694.4	800.0	0.000	1.032	5.000
6-par-ADF	660	3000	3199	3400	2000	3178	3200	0.000	0.183	5.000
6-par	very large	3000	3200	3300	2800	3160	3200	0.000	0.395	3.000

IV. Adaptive Walks

A. Definition

The other landscape analysis technique employed in this work is an examination of *adaptive walks* [Weinberger 1991, Fontana et. al. 1993a]. In an adaptive walk, an initial parent is selected using uniform random selection, and its fitness is evaluated. From this parent multiple offspring are created using, when required for crossover, different second parents. The first offspring which is more fit then the initial parent is selected as a new parent, and the process continues until no offspring can be found with a fitness which is better than its parent. Deciding how many offspring to examine before declaring that the process has reached a local optimum requires a generalization of the mutation based approach discussed in [Weinberger 1991, Fontana et. al. 1993a]. In those cases, there were a defined number of neighbors that could be examined, and the process terminated when all of them evaluated to a fitness equal to or worse than that of the parent. For the purposes of this work, *neighbor* has been defined as any individual that can be reached through the action of one genetic operation. Clearly, all of these neighbors cannot be examined in order to decide that the process has reached a local optimum, so rather arbitrarily the number 30 was chosen as a limit.

B. Results

The fundamental results from the adaptive walks are presented in their entirety in Table 2. Several factors from these results were examined in an attempt to find correlations to the difficulty index developed in Section II.

The mean walk length indicates the number of steps required to reach a local optimum. Figure 3 plots this against the difficulty index. In Figures 3, 4, and 5 the rightmost points on the x-axis are not represented to scale although they are on the y-axis.

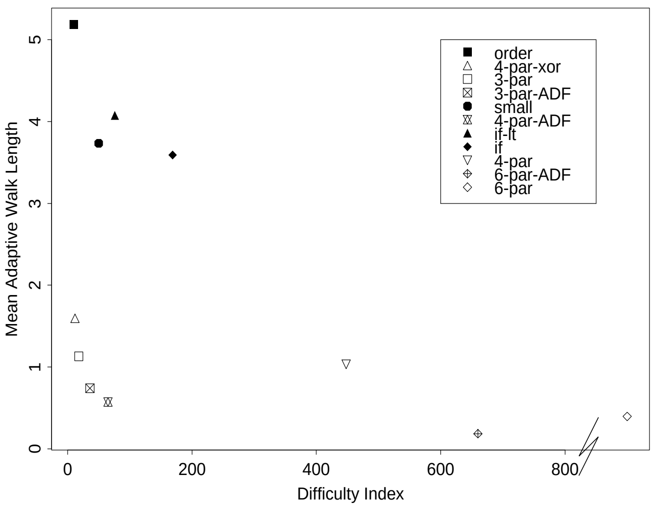

Figure 3: Mean Adaptive Walk Length

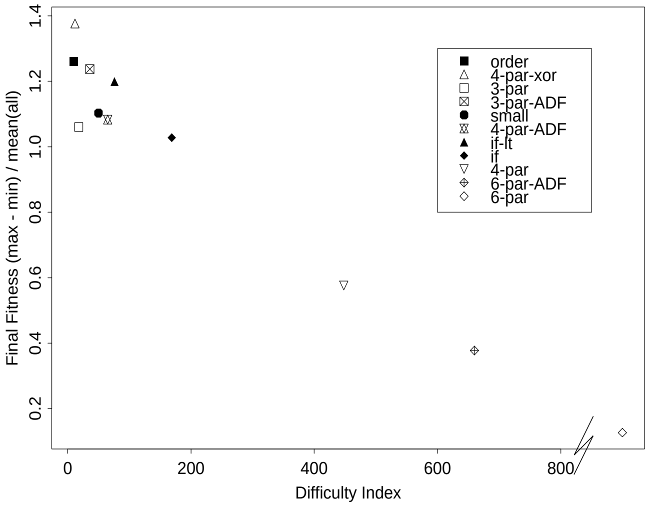

Figure 4: Extreme Basin Depths

The results fall into couple of groups. The solid symbols are all sorting runs, and there is a rough correlation among them as to difficulty. The hollow symbols are all non-ADF even-parity runs, and there is little correlation among them — the 3 parity problem looks about like the 4 parity problem. For problems like the sorting problem, we might use this measure as a rough gauge of difficulty, but it doesn't generalize to the even-parity problems.

Looking further for correlations, if the difference between the maximum and minimum final fitnesses is divided by the mean of the final fitness, some correlation does appear, shown in Figure 4. What does this mean? This measure represents the extremes of the basin depths reached by the adaptive walks, normalized for different fitness measures. An interesting property, but suspect due to high dependence on single occurrences. A different, more stable approach would be to average several of the maximum values and subtract the average of several minimum values from them, then normalize by divid-

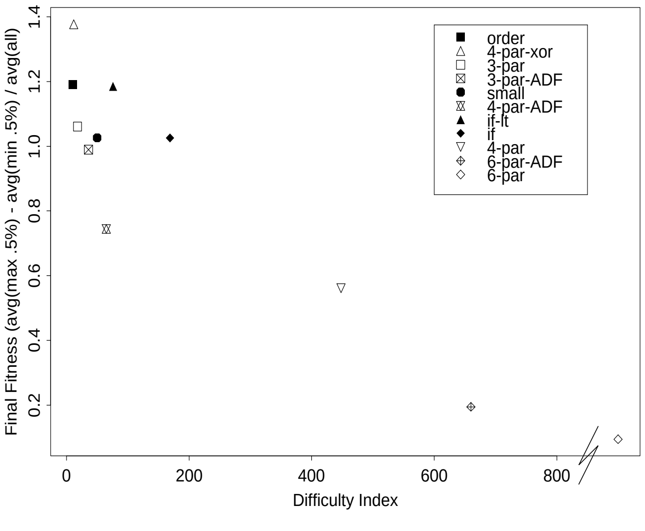

Figure 5: Average of .5% Extreme Basin Depths

ing by the mean of the entire set. This was done, using a set size of 0.5% for the maximum and minimum, and the results are shown in Figure 5. This shows considerably less correlation, with the ADF points moving the most. There are two interpretations of this result. First, this entire technique is just a statistical anomaly. Alternatively, the ADF runs could be sensitive to specific, low probability occurrences. The second view is supported by results from Chapter 17 of Genetic Programming II [Koza 1994], which show that ADF's regularly have exceptionally few occurrences of very highly fit and unfit individuals in the initial random generation.

Similar plots of the original fitness do not show any similar correlation.

C. Discussion

Where there was some correlation in the difficulty and the landscape autocorrelation for the non-ADF even-parity runs, but not for the sorting runs, we now see the reverse. Walk length seems to correlate well with difficulty for sorting, but not for even-parity runs.

Straining a bit to find some factor that correlates to difficulty in the adaptive walk scenario, examination of the basin depth range at the end of an adaptive walk appears to give a rough correlation with difficulty. Whether this linkage holds up under additional testing remains to be seen.

V. CONCLUSIONS

The landscapes of several genetic programming problems have been shown to be different from each other, as well as less correlated than might be expected from the difficulty experienced by genetic programming in solving them. When compared to other problems from the literature, these appear unusually difficult. This may be an artifact of the operation of genetic programming, or in fact these may be very difficult problems.

The range of basin depths at the end of a large number of adaptive walks is seen as a potential indicator of the difficulty of a problem.

ACKNOWLEDGMENTS

Thanks to Lee Altenberg, who first made it clear to me how a fitness landscape is not independent of the genetic operators used. Walter Fontana generously provided a copy of a key paper. Special thanks to Karen Kinnear for many illuminating technical discussions as well as editorial assistance.

REFERENCES

Altenberg, L. (1994) “The Evolution of Evolvability in Genetic Programming,” in *Advances in Genetic Programming,* K. E. Kinnear, Jr., Ed. Cambridge, MA: MIT Press.

De Jong, K. A. (1987) “On Using Genetic Algorithms to Search Program Spaces,” in *Proceedings of the 2nd International Conference on Genetic Algorithms,* J. Grefenstette, Ed. Hillsdale, NJ: Lawrence Erlbaum Associates.

Fontana, W. (1993) Stadler, P. F., Bornberg-Bauer, E. G., Griesmacher, T., Hofacker, I. L., Tacker, M., Tarazona, P., Weinberger, E. D., Schuster, P., “RNA folding and combinatory landscapes”, Physical Rev. E **47,** 2083.

Fontana, W. (1993) Konings, D. A. M, Stadler, P. F., Schuster, P., “Statistics of RNA Secondary Structures”, Biopolymers **33,** 1389.

D. E. Goldberg, (1989) *Genetic Algorithms in Search, Optimization, and Machine Learning*. Reading, MA: Addison-Wesley Publishing Company.

Holland, J. H. (1975, 1992) *Adaptation in Natural and Artificial Systems*. 1975, Ann Arbor, MI: University of Michigan Press and 1992, Cambridge, MA: MIT Press.

Kauffman, S. A. (1989) “Adaptation on Rugged Fitness Landscapes”, in *Lectures in the Sciences of Complexity: Volume I in the Santa Fe Institute Studies in the Sciences of Complexity*. Redwood City, CA: Addison-Wesley Publishing Company.

Kauffman, S. A. (1993) *The Origins of Order: Self-Organization and Selection in Evolution.* New York, NY: Oxford University Press.

Kinnear, K. E. Jr. (1993a) “Evolving a Sort: Lessons in Genetic Programming,” in *Proceedings of the 1993 International Conference on Neural Networks*, New York, NY: IEEE Press.

Kinnear, K. E. Jr. (1993b) “Generality and Difficulty in Genetic Programming: Evolving a Sort,” in *Proceedings of the Fifth International Conference on Genetic Algorithms*, S. Forrest, Ed. San Mateo, CA: Morgan Kaufmann.

Kinnear, K. E. Jr. (1994) “Alternatives in Automatic Function Definition: A Comparison of Performance”, in *Advances in Genetic Programming*, K. E. Kinnear, Jr. Ed. Cambridge, MA: MIT Press.

Koza, J. R. (1992) *Genetic Programming: On the Programming of Computers by Means of Natural Selection*. Cambridge, MA: MIT Press.

Koza, J. R. (1993) “Hierarchical Automatic Function Definition in Genetic Programming,” in *Foundations of Genetic Algorithms, 2*, L. D. Whitley, Ed. San Mateo, CA: Morgan Kaufmann.

Koza, J. R. (1994) *Genetic Programming II.* Cambridge, MA: MIT Press.

Lipsitch, M. (1991) “Adaptation on Rugged Landscapes Generated by Iterated Local Interactions of Neighboring Genes”, in *Proceedings of the Fourth International Conference on Genetic Algorithms,* R. K. Belew and L. B. Booker, Eds., San Mateo, CA: Morgan Kauffman.

Manderick, B., de Weger, M. and Spiessens, P. (1991) “The Genetic Algorithm and the Structure of the Fitness Landscape”, in *Proceedings of the Fourth International Conference on Genetic Algorithms,* R. K. Belew and L. B. Booker, Eds., San Mateo, CA: Morgan Kauffman.

Stadler, P. F. and Schnabl, W. (1992) “The landscape of the travelling salesman problem”, Physics Letters A **161**, 337-344.

Stadler, P. F. (1992) “Correlation in Landscapes of Combinatorial Optimization Problems”, Europhys. Lett. **20** (6), 479-482.

Stadler, P. F. and Happel, R. (1992) “Correlation structure of the landscape of the graph-bipartitioning problem”, J. Phys. A: Math. Gen. **25,** 3103-3110.

Weinberger, E. D. (1991) “Local properties of Kauffman's *N-k* model: A tunably rugged energy landscape”, Physical Rev. A **44**, 6399.

Weinberger, E. D. (1990) “Correlated and Uncorrelated Fitness Landscapes and How to Tell the Difference”, Biol. Cybern. **63**, 325-336.

Wright, S. (1932) “The Roles of Mutation, Inbreeding, Crossbreeding and Selection in Evolution”, in *Proceedings of the Sixth International Congress on Genetics* **1**, 356-366.