

---

## Epistasis Variance: A Viewpoint on GA-Hardness

---

**Yuval Davidor**

Department of Applied Mathematics and Computer Science  
The Weizmann Institute  
Rehovot 76100, Israel

### Abstract

There is general consensus that the coding of a problem domain holds an important key to a successful application. However, there is disagreement as to what aspects of a representation and a problem domain make the application 'hard' for a *genetic algorithm* (GA). This paper suggests a simple statistic, a regression analysis predicting the function value from the bits, as a mean to measure the amount of nonlinearity in a representation, and an interesting perspective on GA-hardness. This statistic is termed *epistasis variance* for its analogy to the use of epistasis in genetics, and presents a perspective on GA-hardness different to those presented in recent works on deceptive problems. Two new findings result from the epistasis analysis. One, a step towards defining and understanding the role of epistasis in GAs, and in the search for understanding GA-hardness. Two, that three elements contribute to GA-hardness: the structure of the solution space, the representation of the solution space, and the sampling error as a result of finite and often small population sizes. These three elements are not necessarily linked, and furthermore, the effect of each of them on GA-hardness is not fixed.

**Keywords:** GA-hardness, representation, deception, epistasis, sampling error.

### Background

The schema theorem [Holland, 1975] suggests prerequisite features which a representation should exhibit in order to utilize a GA processing. Specifically, that there is a high probability that above average, short, low-order schemata combine

to form a higher order above average schemata. The schema theorem shows that above average schemata will proliferate at a given expected minimal rate, but it does not indicate whether this proliferation will occur at the optimum rate. Furthermore, an optimum proliferation by itself is not sufficient to produce a successful GA application.

The only method available for the analysis of the proliferation rates, the Walsh function analysis, is computationally prohibitive [Bethke, 1981, Goldberg, 1989a]. In that respect, it is self evident that the representation is a primary aspect of GAs which determines their utility. The importance of the representation was recognized, but attention was primarily given to the issue of building blocks (their size, number, etc.) [Bethke, 1981, Goldberg, 1987, Goldberg, 1989a]. Deception theory<sup>1</sup> partially encapsulates this issue. However, it is argued that while deceptive problems are GA-hard, they do not fundamentally define GA-hardness.

It is the implicit allocation of search efforts which underlies the operation of a GA. Therefore, the success of a GA search is correlated to the ability to predict correctly the value of strings from the bits (the holism view of complex systems [Goldberg, 1989a, Jacobson, 1955, Platt, 1961, Simon, 1962, Tsotsos, 1987]). The underlying assumption in this paper is that if the correlation is good, the allocation of trials can potentially be optimal (depending on the whole algorithmic ensemble, population size, crossover mechanisms, and so forth [Spears and De Jong, (in press), De Jong and Spears, (in press)]). Therefore, the amount of interdependency among the representation elements is an important ingredient in the GAs' cookbook, and constitutes an essential source of information for understanding GA-hardness.

Gene interaction is a central issue in natural genetics, where genes not only are dependent on each other in order to jointly express phenotypical characteristics, but also suppress and activate the expression of other genes [Ptashne, 1989]. The term that has become synonymous with almost any type of gene interaction is epistasis [Klug, 1986]. Derived from the Greek words *epis* and *stasis* ('behind' and 'stand'), epistasis is therefore equated with *stoppage* or *masking*. Epistasis is used to describe the situation where one gene pair masks or modifies the expression of another gene pair. When the epistasis of a chromosome is said to be high, it means that many genes are dependent on other genes for expression.

## 2 Notional epistasis in GAs

Tracing epistasis is an elusive occupation because the presence of epistatic elements can be traced only at the phenotypic level away from their scene of interaction (genotypic level). The motivation of applying an epistasis analysis is discussed in the present section.

<sup>1</sup>A minimal deceptive problem was defined by Goldberg [Goldberg, 1987], and follows arguments in population dynamics about allele frequencies. Goldberg's definition comes to indicate a situation where the value of a schema instantiated by a local optimum is greater than the complementing schema instantiated by the global optimum. A fully deceptive problem is an extension of the minimal case, and means that the value of all schemata instantiated by a local optimum and having at least one position undetermined, is greater than their complementing schemata instantiated by the global optimum.

If a representation contains very little or no epistasis, no individual string element is affected by the value of the other elements, and therefore optimization means a bit-wise maximization (which a greedy algorithm will most likely process more effectively than a GA). At the other end of the epistatic scale, when a representation is highly epistatic, too many elements are dependent on other elements and the building blocks become of a high order. When the epistasis is extremely high, the elements are so dependent on each other that unless a complete set of unique element values is found simultaneously, no substantial fitness improvements can be noticed (such as in the case of a delta function). Under such extreme circumstances, nonlinearity has exacerbated to the extent that the performance space does not contain significant regularities.

This leads to the conclusion that if a representation exhibits very low epistasis it could probably be processed more efficiently by a greedy algorithm (though it is suitable for a GA). If it contains very high epistasis, then there is too little structure in the solution space, and a GA will most likely drift and settle on a local optimum. In between the two extremes lies a type of problems it might be useful to try and solve with a GA.

A linear decomposition is applied to the representation according to the composing bits. The purpose of applying the above linear decomposition, is to develop a method for the prediction of the amount of nonlinearity (in terms of gene interaction) embedded in a given representation. To this end, fitness has to be associated with the bits. If a linear decomposition proves to be inaccurate, then it implies that the representation incorporates nonlinearities. Quantifying the amount of nonlinearity will provide an estimate for the suitability of a given representation to a GA processing.

From a GA perspective, a coding format in which the effect of any individual parameter on the total fitness is independent of other parameters, suggests that there is little co-adaptation. On the other hand, a high degree of nonlinearity indicates that above average schemata are too big ('big' is not defined here, but some estimates are provided in [Syswerda, 1989, Spears and De Jong, (in press)]). The whole GA theory is based on the assumption that one can state something about the whole only by knowing its parts. What neither the schema theorem nor population genetics indicate, is exactly how much of the whole the parts should indicate. A first step towards quantifying this property is attempted in the next section.

### 3 The basic elements of epistasis

It was suggested previously that when epistasis is high, it is difficult to predict the value of a given string from a measured value of its bits. The following definitions are adopted for the preliminary analysis:

A string,  $S$ , is composed from  $l$  elements  $s_i$  ( $l$  is fixed),

$$S = (s_1, s_2, \dots, s_l) \quad (1)$$

The allele of the  $i$ th gene in a string is denoted by

$$s_i = a \quad a \in \{0, 1\}, \quad i = 1, 2, \dots, l. \quad (2)$$

Symbol	Term
$S$	String
$v(S)$	Fitness
$X(S)$	Excess fitness value
$a$	Allele
$A_i(a)$	Allele value of $a$
$E_i(a)$	Excess allele value
$E(A)$	Excess genic value
$A(S)$	Genic value
$\epsilon(S)$	Epistasis value
$\sigma_v^2$	Fitness variance
$\sigma_A^2$	Genic variance
$\sigma_\epsilon^2$	Epistasis variance

Table 1: Summary of the symbols and their definitions in the epistasis discussion.

The *Grand Population*,  $\Gamma$ , is the set of all possible strings of length  $l$  such that,

$$\Gamma = \{0, 1\}^l. \quad (3)$$

Let  $Pop$  denote a sample from  $\Gamma$  where the sample is selected uniformly and with replacement. The size of a sample  $Pop$  is

$$N = |Pop|. \quad (4)$$

The fitness of a string is given by

$$v(S) = \text{fitness}. \quad (5)$$

where  $v$  is a ‘blackbox’ function. The average fitness value of the sample  $Pop$  is

$$\bar{V} = \frac{1}{N} \sum_{S \in Pop} v(S). \quad (6)$$

The excess fitness value of a string is denoted by

$$E(S) = v(S) - \bar{V}. \quad (7)$$

The number of string instances in  $Pop$  which match  $s_i = a$  is denoted by  $N_i(a)$ . The average allele value is denoted as

$$A_i(a) = \frac{1}{N_i(a)} \sum_{S \in Pop, s_i=a} v(S). \quad (8)$$

where  $Pop_{s_i=a}$  is the set of all strings in  $Pop$  having the allele  $a$  in their  $i$ th position. The excess allele value is defined by

$$E_i(a) = A_i(a) - \bar{V}, \quad (9)$$

and the excess genic value is

$$E(A) = \sum_{i=1}^l E_i(a). \quad (10)$$

## Epistasis Variance: A Viewpoint on GA-Hardness

The genic value of a string  $S$  – the predicted string value – is defined as

27

$$A(S) = E(A) + \bar{V}. \quad (11)$$

Thus, the difference  $\varepsilon(S) = v(S) - A(S)$  might reasonably be supposed to be a measure of epistasis of a string  $S$ .

Consequently, an epistasis measure for the Grand Population and hence for the representation, is termed the *epistasis variance* and is defined as

$$\sigma_e^2 = \frac{1}{N_\Gamma} \sum_{S \in \Gamma} [v(S) - A(S)]^2, \quad (12)$$

where the implicit  $A_i(a)$  are computed over the Grand Population (note that this definition does not follow the common definition of variance as it involves elements from two different sets). This measure can be estimated from the corresponding expression

$$\sigma_{Pop}^2 = \frac{1}{N} \sum_{S \in Pop} [v(S) - A(S)]^2. \quad (13)$$

However, since the computation of  $A_i(a)$  is determined by the sample population, this statistic is subject to sampling error, but as yet, confidence measures for the estimate are unavailable. This would require an investigation of the distribution of

$$\sigma_\Gamma^2 - \sigma_{Pop}^2$$

String	$f_1$	$f_2$	$f_3$	$f_4$
000	0	0	0.0	7
001	1	0	0.5	5
010	2	0	1.0	5
011	3	0	1.5	0
100	4	0	2.0	3
101	5	0	2.5	0
110	6	0	3.0	0
111	7	28	17.50	8

Table 2: Strings and their fitness values of four fitness functions:  $f_1$  (linear function),  $f_2$  ( $\delta$  function),  $f_3$  ( $\frac{1}{2}(f_1 + f_2)$ ), and  $f_4$  (minimal deceptive function) of zero-, total-, semi-, and bounded-epistasis respectively.

The above definitions (summarized in Table 1) provide a method for estimating the epistatic variance for a Grand Population — the base epistasis — from a sample population. The distinction between base epistasis and sampling error is very important because the effect of the latter is often of equal or even higher order of magnitude. This will be demonstrated further in section 4.2.

The fitness variance is denoted as

$$\sigma_v^2 = \frac{1}{N} \sum_{S \in Pop} (E(S))^2, \quad (14)$$

and the genic variance is denoted as

$$\sigma_A^2 = \frac{1}{N} \sum_{S \in Pop} (E(A))^2. \quad (15)$$

## 4 Calculating epistasis: A few examples

In the following, the epistasis tools developed in section 3 are applied to two fitness functions of known and characteristic epistasis (the strings and their corresponding fitness values are summarized in table 2). The functions are the linear function  $f_1$ , the delta function  $f_2$ , the semi-linear function  $f_3$ , and a minimal deceptive function  $f_4$ .

The first analysis uses Grand Populations and thus addresses the issue of base epistasis (section 4.1). Then, the effect a sample has over the statistic is investigated (section 4.2). In section 4.3, a minimal deceptive problem is analyzed. The functions are arranged so as to have an equal average fitness value and thus facilitate comparability between the epistasis variances.

### 4.1 Three epistematically different functions

The Grand Populations of three epistematically different functions are analyzed: zero epistasis (table 3), total epistasis (table 4), and semi-epistasis (table 5). An additional analysis of  $f_1$  is presented, where the representation is a gray code (Table 6).

When analyzing the epistasis variance of the  $f_1$ ,  $f_2$ ,  $f_3$ , and  $f_4$  represented in a gray code, it is possible to observe the strength of the linear assumption, and the epistasis analysis. The  $f_1$  function can be accurately recomposed from the decomposed  $A_i(a)$  values, while the recombination of the  $f_2$  function reveals, as can be expected, a large epistatic variance. The semi-epistatic function  $f_3$  demonstrates further the notion of epistasis analysis. An interesting finding is presented in Table 6. The epistasis analysis of  $f_1$  represented by a gray code shows a light epistasis variance which indicates that this representation is similar to the integer binary representation, but with somewhat less structure. This analytical finding agrees with experimental results comparing integer and gray code representations in a GA environment [Caruana and Schaffer, 1988].

In this simple example of changing the representation of a given function, it can be clearly seen that different representations of the same domain may have different amounts of epistasis variance or 'structure' in them.

### 4.2 Samples and sampling noise

Since the population size in all practical GA applications is only a minuscule portion of the grand population, it is important to investigate whether calculating the epistasis variance from a sample involves a strong sampling error. This section investigates this sampling bias and suggests that the sampling bias has a considerable effect on the measurement of base epistasis variance.

It was already shown that calculating epistasis variance with a Grand Population for a representation which contains zero epistasis yields a correct epistasis figure. This section will show that this conclusion is valid only for the Grand Population, and erroneous when the calculation is not based on the Grand Population. In Tables 7 and 8 such a calculation is shown, and reveals a substantial sampling error variance.

The analysis of sample populations suggests the following:

## Epistasis Variance: A Viewpoint on GA-Hardness

29

$S$	$v(S)$	$E(S)$	$E(A)$	$A(S)$	$\epsilon(S)$
000	0	-3.5	-3.5	0	0
001	1	-2.5	-2.5	1	0
010	2	-1.5	-1.5	2	0
011	3	-0.5	-0.5	3	0
100	4	0.5	0.5	4	0
101	5	1.5	1.5	5	0
110	6	2.5	2.5	6	0
111	7	3.5	3.5	7	0

$i$	$a$	$A_i(a)$	$E_i(a)$
1	0	1.5	-2.0
	1	5.5	2.0
2	0	2.5	2.0
	1	4.5	-1.0
3	0	3.0	-0.5
	1	4.0	0.5

$\sigma_v^2$	$\sigma_A^2$	$\sigma_\epsilon^2$	$\sigma_v^2 - \sigma_A^2$
5.25	5.25	0	0

Table 3: A three-bit unsigned integer binary representation with zero epistasis.

$S$	$v(S)$	$E(S)$	$E(A)$	$A(S)$	$\epsilon(S)$
000	0	-3.5	-10.5	-7	7
001	0	-3.5	-3.5	0	0
010	0	-3.5	-3.5	0	0
011	0	-3.5	3.5	7	-7
100	0	-3.5	-3.5	0	0
101	0	-3.5	3.5	7	-7
110	0	-3.5	3.5	7	-7
111	28	24.5	10.5	14	14

$i$	$a$	$A_i(a)$	$E_i(a)$
1	0	0	-3.5
	1	7	3.5
2	0	0	-3.5
	1	7	3.5
3	0	0	-3.5
	1	7	3.5

$\sigma_v^2$	$\sigma_A^2$	$\sigma_\epsilon^2$	$\sigma_v^2 - \sigma_A^2$
85.75	36.75	49	49

 Table 4: Calculating the epistasis variance for the  $f_2$  function.

$S$	$v(S)$	$E(S)$	$E(A)$	$A(S)$	$\epsilon(S)$
000	0.0	-3.5	-7.0	-3.5	3.50
001	0.5	-3.0	-3.0	0.5	0.00
010	1.0	-2.5	-2.5	1.0	0.00
011	1.5	-2.0	1.5	5.0	-3.50
100	2.0	-1.5	-1.5	2.0	0.00
101	2.5	-1.0	2.5	6.0	-3.50
110	3.0	-0.5	3.0	6.5	-3.50
111	17.5	14.0	7.00	10.5	7.00

$i$	$a$	$A_i(a)$	$E_i(a)$
1	0	0.75	-2.75
	1	6.25	2.75
2	0	1.25	-2.25
	1	4.75	2.25
3	0	1.50	-2.00
	1	5.50	2.00

$\sigma_v^2$	$\sigma_A^2$	$\sigma_\epsilon^2$	$\sigma_v^2 - \sigma_A^2$
41.125	28.875	12.250	12.250

Table 5: Calculating the epistasis variance for the  $f_3$  function.

$S$	$v(S)$	$E(S)$	$E(A)$	$A(S)$	$\epsilon(S)$
000	0	-3.5	-2	1.5	-1.5
001	1	-2.5	-2	1.5	-0.5
011	2	-1.5	-2	1.5	0.5
010	3	-0.5	-2	1.5	1.5
110	4	0.5	2	5.5	-1.5
111	5	1.5	2	5.5	-0.5
101	6	2.5	2	5.5	0.5
100	7	3.5	2	5.5	1.5

$i$	$a$	$A_i(a)$	$E_i(a)$
1	0	1.5	-2.0
	1	5.5	2.0
2	0	3.5	0.0
	1	3.5	0.0
3	0	3.5	0.0
	1	3.5	0.0

$\sigma_v^2$	$\sigma_A^2$	$\sigma_\epsilon^2$	$\sigma_v^2 - \sigma_A^2$
5.25	4.00	1.25	1.25

Table 6: A three-bit unsigned integer gray code representation of the zero epistasis function  $f_1$ .

## Epistasis Variance: A Viewpoint on GA-Hardness

31

$S$	$v(S)$	$E(S)$	$E(A)$	$A(S)$	$\epsilon(S)$
000	0	-3	-2.0	1.0	-1.0
001	1	-2	-2.0	1.0	0.0
010	2	-1	-1.0	2.0	0.0
011*					
100	4	1	1.5	4.5	-0.5
101	5	2	1.5	4.5	-0.5
110	6	3	3.0	6.0	0.0
111*					

  

$i$	$a$	$A_i(a)$	$E_i(a)$
1	0	1.0	-2.0
	1	5.0	2.0
2	0	2.5	-0.5
	1	4.0	1.0
3	0	3.0	0.0
	1	3.0	0.0

  

$\sigma_v^2$	$\sigma_A^2$	$\sigma_\epsilon^2$	$\sigma_v^2 - \sigma_A^2$
4.66	3.75	0.25	-0.92

Table 7: A 75% Grand Population sample shows sampling error variance. The starred (\*) strings are the ones not included in the statistics.

$S$	$v(S)$	$E(S)$	$E(A)$	$A(S)$	$\epsilon(S)$
000	0	-3.5	-6.5	-3	3
001	1	-2.5	-5.5	-2	1
010*					
011*					
100*					
101*					
110	6	2.5	5.5	9	-3
111	7	3.5	6.5	10	-3

  

$i$	$a$	$A_i(a)$	$E_i(a)$
1	0	0.5	-3.0
	1	6.5	3.0
2	0	0.5	-3.0
	1	6.5	3.0
3	0	3.0	-0.5
	1	4.0	0.5

  

$\sigma_v^2$	$\sigma_A^2$	$\sigma_\epsilon^2$	$\sigma_v^2 - \sigma_A^2$
9.25	36.25	9.0	-27.00

Table 8: A 50% Grand Population sample shows an increased sampling error. The starred (\*) are the ones not included in the statistics.

$S$	$v(S)$	$E(S)$	$E(A)$	$A(S)$	$\epsilon(S)$
000	7	3.5	1.25	4.75	2.25
001	5	1.5	0.75	4.25	0.75
010	5	1.5	0.75	4.25	0.75
011	0	-3.5	0.25	3.75	-3.75
100	3	-1.5	-0.25	3.25	-0.25
101	0	-3.5	-0.75	2.75	-2.75
110	0	-3.5	-0.75	2.75	-2.75
111	8	4.5	-1.25	2.25	5.75

  

$i$	$a$	$A_i(a)$	$E_i(a)$
1	0	4.25	0.75
	1	2.75	-0.75
2	0	3.75	-0.25
	1	3.25	-0.25
3	0	3.75	0.25
	1	3.25	-0.25

  

$\sigma_v^2$	$\sigma_A^2$	$\sigma_\epsilon^2$	$\sigma_v^2 - \sigma_A^2$
9.25	0.68	8.57	8.57

Table 9: The epistasis analysis for a minimal deceptive function.

1. As the sample diverts from a Grand Population, the nonlinearity a GA operates with increases.
2. Epistasis, as defined here, consists of two elements: the base epistasis resulting from the representation and a sampling error resulting from sampling noise.

#### 4.3 A minimal deceptive problem

So far, the functions that were analyzed had a known epistasis. To conclude the preliminary discussion on epistasis variance, it would be interesting to analyze a function of an unknown epistasis, a deceptive problem known to be a ‘hard’ problem for a GA.

A minimal deceptive problem<sup>2</sup> is an archetype GA-hard function to which the epistasis tools are applied. A deceptive problem is designed in such a way that it contains a lot of structure, structure which misdirects the convergence to wrong regions. This structure quality is required by the very definition of the function. For the deception to be effective, most low order schemata that does not instantiate the global optimum, have a higher value than those that do. It is reasonable therefore to expect that the sum total epistasis embedded in a deceptive function is not too high, and indeed it ought not to be too high. Calculating the epistasis variance for the minimal deceptive problem (Table 9) confirms the above analysis.

<sup>2</sup>The minimal deceptive problem used here is based on the minimal deceptive problem as defined by Goldberg [Goldberg, 1987, Goldberg, 1989b] and involves variations of negligible importance that were adopted for convenience.

## 5 Conclusions and future work

This paper discussed some fundamental representation and relating GA-hardness issues. In spite of its importance, the precise analysis of representations is uncommon due to the tedious computation involved and lack of consensus definitions of GA-hardness. The epistasis analysis was suggested as an alternative approach for understanding the GA-hardness enigma. However, the amount of computation at the current state of epistasis analysis which necessary to compute an accurate estimate is not substantially smaller.

The epistasis analysis as presented here suggested that there must be sufficient structure in the solution space (depicted by the representation) in order for a GA to process information effectively. Structure in the solution space is not a sufficient condition to guarantee a successful application as demonstrated with a deceptive function. Furthermore, it is not clear how a given epistasis measure can be translated into relative efficiency in respect to other search methods. Nevertheless, it is a necessary condition for a GA. The measurement of epistasis is based on a linear composition of a string from its bits. The accuracy of this method, or more precisely, the epistasis variance of a sample, is an estimate of the total amount of nonlinearity embedded in the representation.

The epistasis variance is usually determined by a sample. Analyzing the epistasis variance for different samples reveals an extensive sampling error resulting from this approximation. Confidence measures for the extent of this approximation were not presented. It is clear that at this stage, the epistasis analysis is not economical. However, it is reasonable to believe that it will be significantly more economical when epistasis can be estimated with reasonable accuracy from a sample.

To be useful as a tool, the epistasis variance requires two extensions to the work presented in this paper:

1. Means for normalization of the epistasis variance so it can be plotted on a scale.
2. The development of confidence measures for estimating the base epistasis from sample populations.

and additional experimental results how the epistasis variance changes as a function of the fitness function and the sample size.

The author is currently involved in extending the epistasis analysis in two directions. One, to develop confidence measures for estimating the base epistasis from a sample. Two, to further investigate the definition of GA-hardness in perspective to epistasis and deception.

## Acknowledgments

The work presented in this article was supported in part by the Center for Absorption in Science, The Ministry of Immigrant Absorption, The State of Israel. Judit Bar-Ilan and Antonia J. Jones made valuable comments on early drafts of this article. Y. Davidor is recipient of a Sir Charles Clore Fellowship.

## 34      References

- Bethke, A. D., (1981) Genetic algorithms as function optimizes *Dissertation Abstracts International*, 41(9) 3503B. University of Michigan Microfilm No. 8106101.
- Caruana, R. A., and Schaffer, J. D. (1988) Representation and hidden bias: Gray vs. binary coding for genetic algorithms, *Proceedings of the 5th International Conference on Machine Learning*, University of Michigan, Ann Arbor, Morgan Kaufmann.
- Spears, W. S., and De Jong, K. A. (in press) An analysis of multi-point crossover, *Proceedings of the Foundations of Genetic Algorithms Workshop*, Indiana University, Bloomington, Morgan Kaufmann.
- De Jong, K. A., and Spears, W. S. (in press) An analysis of the interacting roles of population size and crossover in genetic algorithms, *Proceedings of the 1st International Conference on Parallel Problem Solving from nature*, Springer-Verlag.
- Goldberg, D. E. (1987) Simple genetic algorithms and the minimal, deceptive problem, In *Genetic Algorithms and Simulated Annealing*, L. Davis, (Ed.), Pitman, London, 74-88.
- Goldberg, D. E., (1989) *Genetic Algorithm in Search, Optimization, and Machine Learning*, Addison-Wesley, Reading, MA.
- Goldberg, D. E., (1989) Zen and the art of genetic algorithms, *Proceedings of the 3rd International Conference on Genetic Algorithms and their Applications*, George Mason University, Morgan Kaufmann.
- Goldberg, D. E., (1989) Genetic algorithms and walsh functions: Part I, A gentle introduction, *Complex Systems*, 3, 129-152.
- Goldberg, D. E., (1989) Genetic algorithms and walsh functions: Part II, Deception and its analysis, *Complex Systems*, 3, 153-171.
- Grefenstette, J. J., (1979) *Representation Dependencies in Genetic Algorithms*. Unpublished manuscript, Navy Center for Applied Research in AI, Navel Research Laboratory, Washington, DC 20375-5000, USA.
- Holland, J. H., (1975) *Adaptation in Natural and Artificial Systems*, The University of Michigan Press, Ann Arbor.
- Jacobson, H., (1955) Information reproduction and the origin of life, *American Scientist*, 43, 119-127.
- Klug, W. S., and Cummings, M. R., (1986) *Concepts of Genetics*, 2nd edition, Scott, Foresman and Co..
- Platt, J. R., (1961) Properties of large molecules that go beyond the properties of their chemical sub groups, *Journal of Theoretical Biology*, 1, 342-358.
- Ptashne, M., (1989) How gene activators work, *Scientific American*, January, 25-31.
- Simon, H. A. (1962) The architecture of complexity, *Proceedings of the American Philosophical Society*, 106(6).

## Epistasis Variance: A Viewpoint on GA-Hardness

35

Syswerda, G. (1989) Uniform crossover in genetic algorithms, *Proceedings of the 3rd International Conference on Genetic Algorithms*, George Mason University, Morgan Kaufmann.

Tsotsos, J. K. (1987) A complexity level analysis of vision, *Proceedings of the First International Conference on Computer Vision*, IEEE Computer Society Press, 346-355.