Information Characteristics and the Structure of Landscapes

Vesselin K. Vassilev
School of Computing
Napier University
Edinburgh, EH14 1DJ
Scotland, UK
v.vassilev@dcs.napier.ac.uk

Terence C. Fogarty
School of Computing
Napier University
Edinburgh, EH14 1DJ
Scotland, UK
t.fogarty@dcs.napier.ac.uk

Julian F. Miller
School of Computer Science
The University of Birmingham
Birmingham, B15 2TT
England, UK
j.miller@cs.bham.ac.uk

Abstract

Various techniques for statistical analysis of the structure of fitness landscapes have been proposed. An important feature of these techniques is that they study the ruggedness of landscapes by measuring their correlation characteristics. This paper proposes a new *information analysis* of fitness landscapes. The underlying idea is to consider a fitness landscape as an ensemble of objects that are related to the fitness of neighboring points. Three information characteristics of the ensemble are defined and studied. They are termed: *information content*, *partial information content*, and *information stability*. The information characteristics of a range of landscapes with known correlation features are analyzed in an attempt to reveal the advantages of the information analysis. We show that the proposed analysis is an appropriate tool for investigating the structure of fitness landscapes.

Keywords

Fitness landscapes, information content, partial information content, information stability, information collapse.

1 Introduction

The notion of a *fitness landscape*, introduced by Wright (1932), has become an important concept in evolutionary computation. Three components constitute a fitness landscape. The first is a set of genotypes. The second is a mapping that associates a numerical value to each genotype, known as a fitness function. The final component is an operator that defines a neighborhood relationship within the set of genotypes (Jones, 1995). Recently, the landscapes of a range of problems of known difficulty have been analyzed in an attempt to determine the relation between the landscape structure and the performance of the evolutionary algorithms (EAs) (Goldberg, 1989; Grefenstette and Baker, 1989; Lipsitch, 1991; Manderick et al., 1991; Mitchell et al., 1991; Horn and Goldberg, 1995). In this paper, a new method for analysis of the structure of fitness landscapes is explored. The aim is to define a measure that could give us more information about the landscape structure

than is available using known statistical analysis techniques (Weinberger, 1990; Hordijk, 1996). Such an information measure has been proposed (Vassilev, 1997b) and has been used to evaluate a study of a landscape smoothing technique (Vassilev, 1997a), which was itself inspired by the idea that a landscape can be considered as a superposition of smoother sublandscapes (Stadler, 1995; Slavov and Nikolaev, 1999). Here the original idea of information analysis (Vassilev, 1997a) has been extended. The analysis has been applied to a number of landscapes to explore their structure and, thus, to reveal the advantages of the approach.

A number of techniques for a correlation analysis of the structure of landscapes have been developed. Weinberger (1990) investigated how the autocorrelation function of fitness values of points in a random walk related to the ruggedness of a landscape. The autocorrelation function of random walks for various landscapes has also been explored by Manderick et al. (1991). Together with the correlation length, they examined the fitness correlation coefficient of the corresponding evolutionary operators. Lipsitch (1991) applied a modification of the correlation analysis to landscapes generated by iterations of elementary cellular automata. The intention was to explore the relationship between the nature of the local interactions of neighboring genes, the characteristics of the generated landscapes, and the adaptive capabilities of the populations on these landscapes. Another landscape analysis method, based on correlation, has been proposed by Hordijk (1996). He used the Box and Jenkins (1970) approach on a statistical time series in order to extend the correlation analysis proposed by Weinberger (1990).

The work, reported here, studies a different analysis of the structure of fitness landscapes called *information analysis*. It is inspired by the concept that in algorithmic information theory (Chaitin, 1987), the information content of an individual system is a measure of how difficult it is to describe that system. Others take the information content to define a measure of the ruggedness of a system (Theiler, 1990; Barnsley, 1993).

The information analysis is based on the assumption that each fitness landscape can be considered as an ensemble of various objects, which are characterized by their size, form, and distribution. To analyze these characteristics we define three information features of landscapes. These are called *information content*, *partial information content*, and *information stability*. The first two are based on measures of the amount of information contained in the ensemble of objects. Each object consists of a point in the fitness landscape and its nearest neighbors. The information stability is the highest possible difference in the fitness values of two neighboring points. It can be inferred from a process of filtering out the information content. These features allow us to obtain information about the structure of the landscapes that cannot be given by the other analysis methods.

An important assumption made here is that the fitness landscapes are *statistically isotropic*. A landscape is statistically isotropic when the sequence of fitness values, obtained by a random walk on the landscape, forms a stationary random process for the assumed joint distribution of fitness values.

Section 2 introduces the concept of landscapes. It also gives a brief description of the Nk model of tunably rugged landscapes and includes a discussion of some recently used statistical landscape analysis techniques. Section 3 presents the information analysis. In Section 4, we apply the analysis to Nk landscapes and study their structure. The generality and the advantages of the information analysis are discussed in Section 5. The section also gives an example of how the analysis can lead to a better understanding of the nature of evolutionary processes in the case of digital circuit design (Miller et al., 1997). Conclusions are given in Section 6.

2   Fitness Landscapes

The fitness landscapes are search spaces defined over elements called *phenotypes* that are represented by their *genotype*. A *fitness value* is assigned to each genotype, and it reflects on the ability of the phenotype to survive and reproduce. The fitness value of a genotype is evaluated by a *fitness function* f that measures how *good* the encoded phenotype is. We denote the set of all genotypes by $\mathcal{R}$ and define their fitness values to be real numbers on the interval $\mathcal{I}$. Thus, a fitness function is defined as

$$f: \mathcal{R} \rightarrow \mathcal{I} \tag{1}$$

In an evolutionary search, the relations between points on the landscape are defined by the nature of the evolutionary operator ϕ that is used to move over the landscape. The operator ϕ acts on a k-tuple of members from $\mathcal{R}$ and produces an l-tuple of genotypes. A pair of genotypes will be related when the genotypes can be reached from one another by applying the evolutionary operator to the corresponding landscape points exactly once. In its most general form, a landscape $\mathcal{L}$ is a directed graph (V_ϕ, E_ϕ) whose vertices, $V_\phi = V_\phi^k \cup V_\phi^l$, are genotypes labeled with fitness values and whose edges, $E_\phi \subseteq V_\phi^k \times V_\phi^l$, are connections between the genotypes. The connections are defined by the evolutionary operator (Jones, 1995). The sets V_ϕ^k and V_ϕ^l are defined as

$$V_\phi^\theta = \{(r, f^\theta(r)) \mid r \in \mathcal{R}^\theta\} \tag{2}$$

where θ is either k or l, and $f^\theta: \mathcal{R}^\theta \rightarrow \mathcal{I}$ assign a fitness value to each vertex. The described model has two important advantages. First, it allows us to define terms such as local and global optimum, neighborhood, basin of attraction, etc. (Appendix A). Second, it allows us to associate unique fitness landscapes with specific evolutionary operators in agreement with the concept "one operator, one landscape" (Jones, 1995). Primarily, this paper considers simple landscapes associated with a bitflip operator. In this case, each landscape point is a genotype, and two genotypes are neighbors if they differ in a single gene. However, in Section 5 we consider a much more involved model in which both mutation and crossover landscapes are investigated.

The structure of fitness landscapes influences the ability of an evolutionary algorithm to perform an efficient search (Kauffman, 1989; Manderick et al., 1991; Mitchell et al., 1991). There are several characteristics associated with the landscape optima that define the structure of fitness landscapes. These are their number, type, magnitude, and the sizes of their basins of attraction. These characteristics are well known in the Evolutionary Computation community. They have been studied in investigations of the landscape deceptiveness (Goldberg, 1987, 1989), modality (Altenberg, 1995; Horn and Goldberg, 1995), and ruggedness (Kauffman, 1989; Weinberger, 1990) in order to understand the nature of evolutionary search under different conditions.

There are various techniques for studying the structure of landscapes, and a number of models with tunable ruggedness have been proposed. These are discussed in the following two subsections.

2.1   Tunably Rugged Landscapes

A model of landscapes with tunable ruggedness, known as Nk landscapes, has been proposed by Kauffman (1989) in order to explore the relationship between the local interactions

among the genes in genotypes and the population flow on rugged landscapes.[1] The Nk model represents a family of landscapes with tunable ruggedness defined by two parameters: the length of the genotype, N, and the number of genes that epistatically influence each gene, k ($0 \leq k < N$). The fitness of each genotype, $g \in \mathcal{R}$, is the mean of fitness contributions of the genes and is given by

$$f(g) = \frac{1}{N} \sum_{i=1}^{N} f_i(g_i; g_{i_1}, ..., g_{i_k}) \tag{3}$$

where the fitness contribution of the ith gene is formed by the ith allele and k others that could be either the nearest neighbors (adjacent neighborhood model) or randomly chosen genes (random neighborhood model). Therefore, for each locus, a table of 2^{k+1} elements is generated that determines the fitness contribution of the locus.

It has been demonstrated that when the parameter k is increased towards N the fitness landscape ruggedness also increases (Kauffman, 1989; Weinberger, 1990). Thus, when $k = 0$, each fitness contribution depends only on the corresponding gene. There is a single optimum, and it is reachable from all other genotypes via hillclimbing. The fitness difference between neighboring genotypes is small, and, therefore, the landscape is smooth. For $k = N - 1$ the epistasis in the genotypes is high. The expected number of local optima is $\frac{2^N}{N+1}$, and the landscape is maximally rugged. The relationship between the parameter k and the landscape ruggedness is given by the autocorrelation function $R(d)$. The exact forms of $R(d)$ for both the random and the adjacent neighborhood models are given by

$$R(d) = (1 - \frac{d}{N})(1 - \frac{k}{N-1})^d \tag{4}$$

and

$$R(d) = 1 - \frac{k+1}{N}d + \frac{1}{\binom{N}{d}} \sum_{j=1}^{\min(k, N+1-d)} (k - j + 1)\binom{N - j - 1}{d - 2} \tag{5}$$

respectively, where d is the Hamming distance (Weinberger and Stadler, 1993; Stadler, 1995; Altenberg, 1997).

2.2 Landscape Analysis Techniques

The structure of a fitness landscape can be investigated by measuring the degree of correlation between points on the landscape (Weinberger, 1990; Lipsitch, 1991; Manderick et al., 1991; Hordijk, 1996). The degree of correlation between landscape points depends on the difference between their fitness values. The smooth landscapes are highly correlated because the landscape points have similar fitness values. If the fitness difference is high, the landscape is rugged, and the correlation is low. Therefore, if we measure the correlation between points, we can assess the landscape ruggedness.

---

[1] Another model of tunably rugged landscapes has been proposed by Lipsitch (1991). His model is also based on the local interactions within the genotypes. Briefly, the landscapes are constructed by estimating the phenotypes created through l iterations of elementary cellular automata (Gutowitz, 1991) initialized with genotypes. The ruggedness can be tuned by a special parameter λ that determines the amount of information in the automata transition rule (Langton, 1990; Li et al., 1990). Using this parameter, Li et al. (1990) have divided the elementary cellular automata rules into six classes. These classifications are based on the differences in the behavior of elementary cellular automata. Lipsitch (1991) has shown that fitness landscapes can be classified according to the classification of cellular automata rules that generate them. Unfortunately, this model has not been investigated enough and, therefore, we find it inappropriate to use in this paper.

A technique for studying correlation in a fitness landscape has been proposed by Weinberger (1990). He generates a sequence of fitness values, $\{f_t\}_{t=1}^n$, by a random walk on the landscape using a bitflip operator and examines the *autocorrelation function*

$$\rho(s) = \frac{E[f_t f_{t+s}] - E[f_t]E[f_{t+s}]}{V[f_t]} \tag{6}$$

where $E[f_t]$ and $V[f_t]$ are the expectation and the variance, respectively, of the time series. The autocorrelation function indicates the correlation between points that are separated by a distance s. Another measure suggested by Weinberger (1990) is *correlation length*. This is the distance beyond which the majority of points become uncorrelated. Because of the exponential decay of the correlation, Weinberger (1990) defines the correlation length as follows:

$$\tau = -\frac{1}{\ln(\rho(1))} \tag{7}$$

where $\rho(1)$ is the autocorrelation (Equation 6) of neighboring points. Thus, the autocorrelation function calculated on a time series obtained by a random walk on a Nk landscape should empirically correspond to the ruggedness of the landscape. The plots, depicted in Figure 1, show the autocorrelation functions of Nk landscapes with random neighborhoods for $k = 0, 2, 5, 10, 19$ and $N = 20$, and their correlation lengths. These were calculated on random walks of $10,000$ steps (the random walks are performed by a bitflip operator). The figure reveals that the correlation decreases as k is increased, which agrees with our expectations.
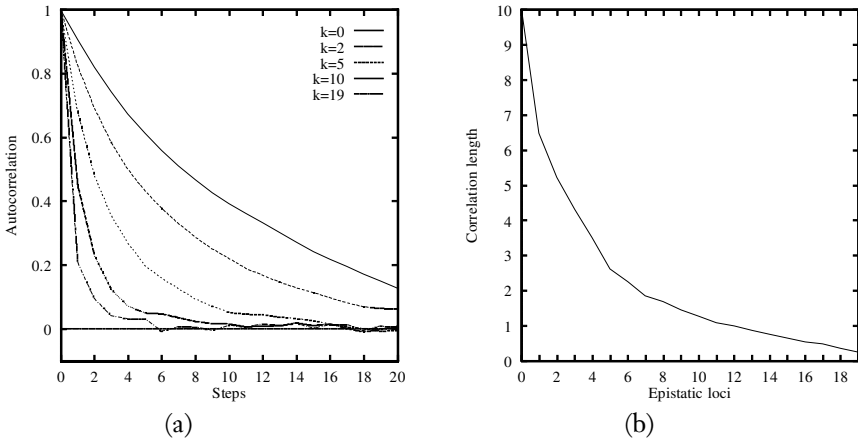

Figure 1: Correlation characteristics of Nk landscapes with random neighborhoods for $N = 20$ and different values of k: (a) autocorrelations and (b) correlation lengths.

Various modifications of the correlation analysis have been devised. Manderick et al. (1991) suggested, together with the autocorrelation function and a modified form of the correlation length (the distance for which the autocorrelation function is $1/2$), that one could calculate an estimate of a *correlation coefficient* associated with the evolutionary operator. The correlation coefficient is defined as follows:

$$\rho_{OP}(F_i, F_{i+s}) = \frac{Cov[F_i, F_{i+s}]}{\sqrt{V[F_i]V[F_{i+s}]}} \tag{8}$$

where F_i is a set of fitness values of genotypes within the ith population. $Cov[F_i, F_{i+s}]$ is the covariance between F_i and F_{i+s}, and $V[F_i]$ is the variance (in the original definition s is 1). Thus, the correlation between two populations separated by s applications of the evolutionary operator can be assessed. Note that the correlation coefficient from Equation 8, which was originally defined as an estimate of the usefulness of the corresponding evolutionary operator, is indirectly related to the structure of the landscape. Now we define a simple correlation analysis based on these ideas. The aim is to show that the aforementioned analysis technique leads to similar conclusions to those found by Weinberger (1990). We propose to measure the relationship of the distance between the populations and their correlation coefficient, where the distance is the number of applications of the bitflip operator to a randomly chosen population. Thus, we define a population correlation function as

$$\rho'(s) = \rho_{OP}(F_0, F_s) \qquad (9)$$

We use a similar definition of the correlation length to that of Manderick et al. (1991). The correlation length now is defined to be the distance s_l for which $\rho'(s_l) = \frac{1}{2}$. As before, we calculate the population correlation functions and correlation lengths of Nk landscapes with random neighborhoods, $k = 0, 2, 5, 10, 19$ and $N = 20$. The correlations are computed on random walks of a population with size $1,000$ (the random walks are performed by a bitflip operator applied to each element of the population). The plots, depicted in Figure 2, reveal a similar correlation structure to that shown in Figure 1. Thus, we conclude that the two analyses ultimately do not fundamentally differ.
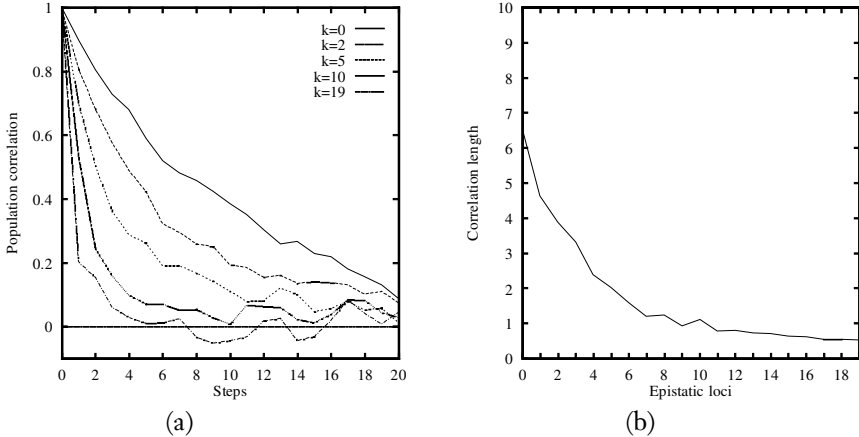

Figure 2: Correlation characteristics of Nk landscapes with random neighborhoods for $N = 20$ and different values of k: (a) population correlations and (b) correlation lengths (Equation 9).

Another interesting approach is the statistical analysis proposed by Hordijk (1996) who used the Box and Jenkins (1970) approach in order to extend the correlation analysis of Weinberger (1990). The idea of utilizing the time series analysis to explore the structure of landscapes is not unknown. It has been suggested by Weinberger (1990), who asserts that the *autoregressive* (AR) model of order 1 captures the statistics of walks on the Nk landscapes, and, therefore, the AR(1) model is applicable to the Nk landscapes. Hordijk's analysis is different. He proposes time series analysis to be the *autoregressive moving-average* (ARMA)

model, an extension of the AR model. The analysis gives us the autocorrelations together with a stochastic model that describe the time series more accurately.

3   Information Analysis

We have seen that the correlation analyses have been used to study the structure of fitness landscapes. Unfortunately, they do not go far enough and give us only a vague notion of the structure of the landscapes. The landscape is a structure involving a number of important features. To analyze them, we consider that a fitness landscape can be seen as an ensemble of basic objects each of which is represented by a point and the possible outcomes that may be produced by the corresponding evolutionary operator at that point. There are several kinds of objects, and they can be classified into three principal classes: flat areas (each point together with its neighbors have the same fitness), isolated points (each point has higher or lower fitness than all its neighbors), and a third class of all points that are neither flat nor isolated points. We propose three information measures that characterize the structure of a fitness landscape defined over a time series obtained by a walk on the landscape. The information measures are:

1. *Information content* that is an entropic measure of the time series.

2. *Partial information content* that relates to the modality encountered on the landscape path.

3. *Information stability* that is the highest fitness difference between neighboring points reached in the time series.

In general, the information content of a system can be regarded as the amount of information that is required to reconstruct the system exactly. There are two aspects to this: the number of states and the distribution of elements among the states in a system. Consider one extreme where the elements of a system are in the same state. To reconstruct the system we merely have to define the state of a single element and replicate that for all remaining elements. In the other extreme, where the elements of the system are in completely different states, the amount of information needed to describe the system could be enormous. Classical information theory captured one aspect of this where the information content was defined as a function of the distribution of the elements over the states of a system. This is also known as the Shannon entropy (Shannon, 1948). The entropy has been also used as a basic concept in quantifying the fractal dimension of a system as a measure of its ruggedness (Theiler, 1990; Barnsley, 1993). Another aspect of information content is captured by algorithmic information theory (Chaitin, 1987). The information content of an object has been defined as the smallest number of bits required by a program to print out the object. The measures proposed above in the *information analysis* of a landscape are inspired by these two concepts, and they come to characterize the distribution (information content) and the number (partial information content) of the optima encountered in a walk on the landscape.

The distribution of local optima on a landscape is itself composed of two aspects. First is the variety of "shapes" on the landscape. This relates to the local neighborhood of a landscape point and is captured by the suggested information content. Second is the magnitude of the landscape optima. The upper bound of this aspect is captured by the information stability.

Of themselves these three information characteristics, though interesting, do not markedly specify the structure of a landscape any more exactly than the correlation analysis. Thus, we pose the following question: is it possible that by altering the natural scale by which we observe a landscape (to ignore certain non-essential features), we can learn more about the landscape structure? We try to answer this question in the next section.

3.1 Information Characteristics of Landscapes

Consider a sequence of fitness values $\{f_t\}_{t=0}^n$, which are real numbers taken from the interval $\mathcal{I}$ and obtained by a *walk* on a landscape $\mathcal{L}$. The sequence is a time series that represents a path in $\mathcal{L}$ and contains information about the structure of the landscape. The aim is to extract this information by representing the time series as an ensemble of objects. The ensemble can be defined as a string $S(\varepsilon) = s_1 s_2 s_3...s_n$ of symbols $s_i \in \{\bar{1}, 0, 1\}$ given by

$$s_i = \Psi_{f_t}(i, \varepsilon) \tag{10}$$

where

$$\Psi_{f_t}(i, \varepsilon) = \begin{cases} \bar{1}, & \text{if } f_i - f_{i-1} < -\varepsilon \\ 0, & \text{if } |f_i - f_{i-1}| \le \varepsilon \\ 1, & \text{if } f_i - f_{i-1} > \varepsilon \end{cases} \tag{11}$$

for any fixed ε. The parameter ε is a real number taken from the interval $[0, l_{\mathcal{I}}]$, where $l_{\mathcal{I}}$ is the length of the interval $\mathcal{I}$. Note that the parameter ε determines the accuracy of calculation of the string $S(\varepsilon)$. If $\varepsilon = 0$, the function Ψ_{f_t} will be very sensitive to the differences between the fitness values, and $S(\varepsilon)$ will be determined as precisely as possible. When the parameter ε is $l_{\mathcal{I}}$, $S(\varepsilon)$ will be a string of 0s.

The string $S(\varepsilon)$ contains information about the structure of the landscape. Note that the function Ψ_{f_t} associates each edge of the path with an element from the set $\{\bar{1}, 0, 1\}$. Each *object* of the path is represented by a string, $s_i s_{i+1}$, which is a sub-block of length two of the string $S(\varepsilon)$. One can think of $S(\varepsilon)$ as a sequence of elements (a sample) of the *incidence* matrix of the landscape underlying graph. The incidence matrix of a landscape is related to the landscape's graph *Laplacian* matrix whose eigenvectors are the orthogonal basis of eigenfunctions of the Fourier transform of the landscape (Stadler, 1995).

3.1.1 Information Content

We define an entropic measure of the ensemble of the sub-blocks of length two of string $S(\varepsilon)$. This is

$$H(\varepsilon) = - \sum_{p \ne q} P_{[pq]} \log_6 P_{[pq]} \tag{12}$$

and is referred to as the *information content*. The measure is an estimate of the variety of "shapes" in the ensemble, and, thus, it characterizes the ruggedness of the landscape path that is represented by $S(\varepsilon)$ with respect to the flat areas in the path. The probabilities $P_{[pq]}$ are frequencies of the possible blocks pq of elements from set $\{\bar{1}, 0, 1\}$. They are defined as

$$P_{[pq]} = \frac{n_{[pq]}}{n} \tag{13}$$

where $n_{[pq]}$ is the number of occurrences of pq in $S(\varepsilon)$. Note that the sum in Equation 12 is taken over sub-blocks of $S(\varepsilon)$ that are composed of two different symbols. Hence, the

inequality

$$\sum_{p \neq q} P_{[pq]} \leq 1 \tag{14}$$

is fulfilled. The aim is to define the entropic measure with respect to the relation between the number of the objects pq for which $p \neq q$ and the number of all possible sub-blocks of length two of string $S(\varepsilon)$.

The logarithm from Equation 12 is taken with base six since this is the number of blocks composed by two different symbols taken from $\{\bar{1}, 0, 1\}$. In this way, the information content is scaled in the interval $[0, 1]$. In our calculations below, the string $S(\varepsilon)$ is considered with periodic boundary conditions since the investigated landscapes are assumed to be statistically isotropic.

3.1.2 Partial Information Content

An important feature of the landscape path $\{f_t\}_{t=0}^n$ that is related to its ruggedness is the modality of the path. This can be assessed by measuring the amount of information of the ensemble that is represented by the string $S(\varepsilon)$ in a new way. Note that the modality of a landscape path cannot be characterized by its information content $H(\varepsilon)$ since this is an estimate of the diversity of objects associated with the landscape optima. To explore the modality of a path sampled by a walk on a landscape, we assume that this is a characteristic related only to the number of optima in the path. Thus, the objects are defined to be the optima, irrespective of the fact that they might be isolated optima, plateaus, etc. (Appendix A).

Consider the string $S(\varepsilon)$ that is associated with the time series $\{f_t\}_{t=0}^n$. We construct a new string $S'(\varepsilon)$ of elements from $S(\varepsilon)$ in the following way: $S'(\varepsilon)$ is empty if $S(\varepsilon)$ is a sequence of 0s; otherwise it is defined as $S'(\varepsilon) = s_{i_1} s_{i_2} ... s_{i_\mu}$ where $s_{i_j} \neq 0$, and $s_{i_j} \neq s_{i_{j-1}}$ for $j > 1$. Thus, by ignoring a certain non-essential part of $S(\varepsilon)$, we obtain string $S'(\varepsilon)$ with length μ, and the quantity μ indicates the modality of the landscape path. Note that the string $S'(\varepsilon)$ has the form "$\bar{1}1\bar{1}...$", and this is the shortest string that represents the slopes of the corresponding landscape path. For instance, if the landscape path is maximally multimodal, $S(\varepsilon)$ could not be modified, and its length will remain unchanged. The length of $S'(\varepsilon)$ scaled in the interval $[0, 1]$ is called *partial information content* and it is given by

$$M(\varepsilon) = \frac{\mu}{n} \tag{15}$$

where n is the length of $S(\varepsilon)$. Define the function $\Phi_S(i, j, k)$ to count the slopes of the optima that are represented by string $S(\varepsilon) = s_1 s_2 s_3 ... s_n$ as

$$\Phi_S(i, j, k) = \begin{cases} k, & \text{if } i > n \\ \Phi_S(i+1, i, k+1), & \text{if } j = 0 \text{ and } s_i \neq 0 \\ \Phi_S(i+1, i, k+1), & \text{if } j > 0, s_i \neq 0 \text{ and } s_i \neq s_j \\ \Phi_S(i+1, j, k), & \text{otherwise} \end{cases} \tag{16}$$

The evaluation of μ can now be written as $\Phi_S(1, 0, 0)$. The partial information content $M(\varepsilon)$ is 0 when the landscape path is flat and there are no slopes in the path. When the landscape path is maximally multimodal, $M(\varepsilon)$ is 1. For a given partial information content $M(\varepsilon)$, the number of optima of the corresponding landscape path can be calculated as $\lfloor \frac{nM(\varepsilon)}{2} \rfloor$.

An important assumption made here is that the elements of $S'(\varepsilon)$ are instances of slopes with different characteristics, such as length and angle. In fact, the diversity of slopes can be studied by moving the parameter ε. This is considered in the following section.

3.1.3 Information Stability

The information content and the partial information content characterize the time series $\{f_t\}_{t=0}^n$ with a certain accuracy. The accuracy of the estimations depends on the parameter ε that in turn defines two functions of the information measures versus accuracy. One can think of the parameter ε as a magnifying glass through which the landscape can be observed. For small values of ε, the function Ψ_{f_t} from Equation 11 will be very sensitive to the difference between the fitness values, i.e., the glass will make each element of the landscape visible. If ε is zero, then the accuracy of the estimations of $H(\varepsilon)$ and $M(\varepsilon)$ is high. In contrast, for $\varepsilon = l_{\mathcal{I}}$, the information content and the partial information content of $S(\varepsilon)$ are 0, i.e., for such ε, the landscape path will be determined as completely flat. The role of the parameter ε is illustrated in Figure 3 where the landscape profile for different values of ε is shown.
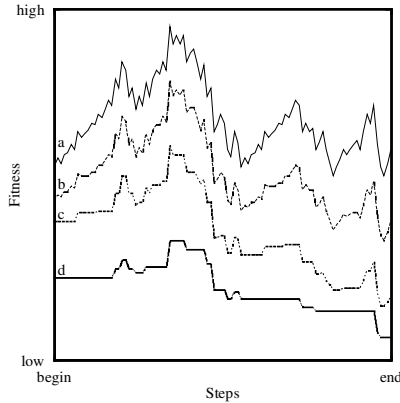

Figure 3: The profile of a landscape path for different values of parameter ε: (a) 0, (b) 0.02, (c) 0.05, and (d) 0.1.

The smallest value of ε for which the landscape becomes flat is called the *information stability*. This is the value $\varepsilon^\star$ for which $S(\varepsilon^\star)$ is a string of 0s.

3.2 Information Characteristics of a Simple Landscape Path

This section gives a simple illustration of an estimation of the information characteristics for a sequence of fitness values. Consider a sequence $F = \{f_t\}_{t=0}^5 = \{0, 0.01, 0.05, 0.2, 0.21, 0.9\}$ of values from the interval $[0, 1]$. To calculate the information characteristics $H(\varepsilon)$ and $M(\varepsilon)$ for a given value of ε, we need to construct the string $S(\varepsilon) = s_1 s_2 s_3 s_4 s_5$ for this ε $(0 \leq \varepsilon \leq 1)$. For instance, let ε be 0.05. According to the formulae from Equations 10 and 11, the elements s_1, s_2, and s_4 are the symbol 0; s_3 and s_5 are the symbol 1 since $f_1 - f_0$, $f_2 - f_1$, and $f_4 - f_3$ are less than 0.05, and $f_3 - f_2$ and $f_5 - f_4$ are greater than 0.05, respectively. Hence, the string $S(\varepsilon = 0.05)$ is 00101. We now calculate the probabilities $P_{[pq]}$ for each p and q, where $p \neq q$. In this particular

Table 1: Information characteristics of the landscape path F (Section 3.2).

ε	$S(\varepsilon)$	$P_{[01]}$	$P_{[10]}$	$H(\varepsilon)$	$M(\varepsilon)$
0	11111	0	0	0	1/5
0.01	01101	2/5	2/5	0.4091	1/5
0.05	00101	2/5	2/5	0.4091	1/5
0.16	00001	1/5	1/5	0.3593	1/5
0.68	00001	1/5	1/5	0.3593	1/5
0.69	00000	0	0	0	0
1	00000	0	0	0	0

example, we are only interested in the frequencies of sub-blocks 01 and 10 since symbol $\bar{1}$ is not represented in $S(\varepsilon = 0.05)$. The number of occurrences of block 01 in S is equal to 2, which is also the number of occurrences of 10 ($n_{[01]} = n_{[10]} = 2$). The evaluation of $n_{[01]}$ and $n_{[10]}$ is illustrated by

$$\mathbf{00\widehat{10}10} \Longrightarrow n_{[01]} = 2 \text{ and } \mathbf{00\widehat{10}10} \Longrightarrow n_{[10]} = 2 \qquad (17)$$

Since the string is taken with periodic boundary conditions, we consider that the first number is also last (in Equation 17 it is the number 0 that is given in the end of the strings with non-bold type). Consequently, the probabilities $P_{[01]}$ and $P_{[10]}$ are $\frac{2}{5}$ each since the number of all sub-blocks of length two of S is equal to 5. Thus, the information content is $H(\varepsilon = 0.05) = -\frac{2}{5} \log_6 \frac{2}{5} - \frac{2}{5} \log_6 \frac{2}{5}$, which is approximately 0.4091.

To estimate the partial information content we apply the formula from Equation 16 to the string $S(\varepsilon = 0.05)$. The partial information content $M(\varepsilon = 0.05)$ is $\frac{\mu}{5}$, where $\mu = \Phi_S(1, 0, 0)$ is 1.

The information characteristics of the landscape path for different values of ε are given in Table 1. The information stability is comparatively high (approximately 0.69), and the information content $H(0)$ is zero. Consequently, the landscape path is smooth, however, not flat since $H(0) < H(0.01)$. The partial information content $M(0)$ is $1/5$. This implies a slope in the landscape path.

3.3 Properties of the Information Characteristics

For a better understanding of the information characteristics, some interesting properties are given below.

1. For each sequence of fitness values, the information stability $\varepsilon^\star$ can always be calculated. This follows from the definition of string $S(\varepsilon)$. Since $S(l_\mathcal{I})$ is a string of 0s, we conclude that $\varepsilon^\star \leq l_\mathcal{I}$.

2. Let $\varepsilon^\star$ be the information stability of a sequence of fitness values. The information content and the partial information content are equal to zero for each ε greater than $\varepsilon^\star$, i.e., $\varepsilon^\star \leq \varepsilon \leq l_\mathcal{I}$. Therefore, it is unnecessary to calculate the information characteristics for $\varepsilon > \varepsilon^\star$.

3. In the case of a *regular walk* on a landscape, the information content and the partial information content of the time series become positive constants for each $\varepsilon \in [0, \varepsilon^\star)$. Note that the term *regularity* has nothing to do with the notion of regular graphs. We say that a time series $\{f_t\}_{t=0}^n$ is generated by a *regular* walk on a landscape when the time series obeys

$$f_{t+1} = f_t \pm \kappa c \tag{18}$$

where c is a constant and κ is a variable that can be -1, 0, or 1. A landscape path is generated by an *irregular* walk when the condition in Equation 18 is not fulfilled. In practice, it is rare to encounter purely regular walks on a landscape. The walks usually have a low degree of regularity. If we consider Equation 18 in the form $f_{t+1} = f_t \pm \kappa \sum_i c_i$ where c_i are different constants, then the degree of regularity of a landscape is given by the number of different κc_i, that is, the number of all possible differences of fitness values.

4. There can exist a sequence of fitness values, and parameters ε_1 and ε_2, for which $H(\varepsilon_1) < H(\varepsilon_2)$, where $0 \leq \varepsilon_1 < \varepsilon_2 \leq l_\mathcal{I}$. A typical example for which $\varepsilon_1 = 0$ and $\varepsilon_2 = 0.01$ is the path with information characteristics shown in Table 3.2. The landscape paths associated with this class are characterized by relatively small flat landscape areas. For such paths, $H(\varepsilon)$ is an increasing function for low values of ε since each path as an ensemble consists mainly of two types of objects. We say that the flat landscape areas *prevail* over the ruggedness in a time series if $H(\varepsilon)$ is a decreasing function.

5. The partial information content decreases toward 0 as ε increases, and the steepness of the function $M(\varepsilon)$ indicates the diversity of the optima when they are classified by their magnitude.

3.4 Information Characteristics and Landscape Modality

The modality of fitness landscapes is an important feature that strongly influences the ability of an evolutionary algorithm to search. Unfortunately, however, the relationship between modality and search is still vague. For instance, contrary to our expectations, it has been shown that unimodal landscapes could be hard for climbing (Horn and Goldberg, 1995), and highly multimodal landscapes could be transformed to easy ones for evolutionary search (Altenberg, 1995). On the other hand, there are not, at present, appropriate statistical analysis methods that could be used to perform an accurate study of this landscape characteristic. The impediments that might be faced in a study of the modality of a landscape are twofold: first, to choose an appropriate *walk* on the landscape and, thus, to generate a time series of fitness values that can give us information about this landscape feature; second, to choose an appropriate method for analysis of the time series. In this section we are interested in how the information characteristics, introduced above, are related to the modality of a landscape path without paying attention to the algorithm that generates the time series.

We consider that the modality is a landscape feature that is related to three basic characteristics of the local optima: the number of local optima, the number of groups of local optima, and the degree of isolation of the local optima. The relation between the number of local optima and the landscape modality is straightforward. A landscape is unimodal, bimodal, or multimodal when the number of optima is one, two, or more, respectively. However, landscapes that have equal numbers of optima do not present identical difficulty for an evolutionary search algorithm. In order to distinguish such landscapes we consider

that the landscape modality is also related to the size of the basins of attraction of the optima, that is, the degree of isolation of the optima (Appendix A) and the flat fitness landscape areas that are defined by the number of groups of optima. We say that peaks compose a group of optima if they are surrounded by a flat landscape area. We assume that to define an analysis of the landscape modality we must be able to investigate these local optima characteristics.

Now, using the described information characteristics, we intend to explore the modality of a simple example of a landscape path that is composed of groups of isolated optima. The aim is to attain understanding of how the information characteristics are related to the modality of the path.

Consider a time series obtained by a walk on a landscape. Let us assume that for a certain value of the parameter ε, the landscape path $\mathcal{P}$ that originates from the corresponding ensemble of objects is constituted from groups of isolated optima. The path is simply the string $S(\varepsilon_{\mathcal{P}})$, and it has the form

$$001\bar{1}1\bar{1}1\bar{1}0...01\bar{1}...00 \qquad (19)$$

Each group of optima starts and finishes with the blocks 01 and $\bar{1}0$, respectively. We require symbol 0 in order to specify the existence of a group of optima. The asymmetry "$01...\bar{1}0$" in our simple example appears to assure us that peaks exist in the landscape path.

The calculation of the information content and the partial information content of $\mathcal{P}$ is straightforward. Let m, k, and n be the number of peaks, the number of groups, and the length of $S(\varepsilon_{\mathcal{P}})$, respectively, where $k \leq m \ll n$. Then the probabilities $P_{[1\bar{1}]}$ and $P_{[\bar{1}1]}$ are equal to $\frac{m}{n}$ and $\frac{m-k}{n}$, respectively, and the probabilities $P_{[01]}$ and $P_{[\bar{1}0]}$ are equal to $\frac{k}{n}$. Considering the formulae from Equations 12 and 15, we obtain that the information content is

$$H_{\mathcal{P}} = 2\frac{k}{n}\log_6\frac{n}{k} + \frac{m-k}{n}\log_6\frac{n}{m-k} + \frac{m}{n}\log_6\frac{n}{m} \qquad (20)$$

and the partial information content is

$$M_{\mathcal{P}} = \frac{2m}{n} \qquad (21)$$

The relationship between the information content and the parameters m and k is shown in Figure 4. The depicted plots represent the information content versus the number of groups of peaks for different values of m. The figure reveals that for a certain number of groups of peaks the information content increases with the number of peaks. The information content also indicates how the peaks are grouped in the landscape path. Unfortunately, however, this information characteristic cannot accurately quantify the numbers of peaks and groups in the path. Figure 4 demonstrates that paths with different modality can have equal information content. Even if we estimate the partial information content that measures the number of peaks, different landscape paths with equal numbers of peaks may have the same information content.

Another measure of the landscape path, represented as an ensemble of objects, that may be useful in a study of the structure of a landscape is the estimate of the variety of *flat* and *smooth* sections of the landscape. For us, these are the blocks 00, 11, and $\bar{1}\bar{1}$. Thus, we propose the entropic measure

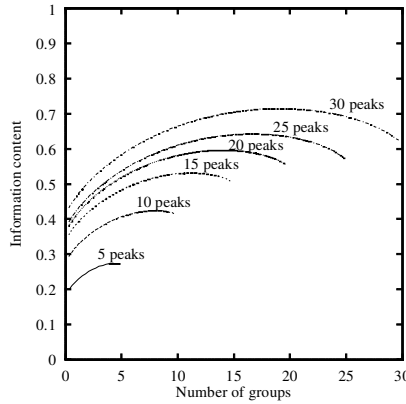$$h(\varepsilon) = -\sum_{p\in\{\bar{1},0,1\}} P_{[pp]}\log_3 P_{[pp]} \qquad (22)$$

Figure 4: The relationship between the information content $H_{\mathcal{P}}$ and the number of peaks and groups for $n = 90$.

which we call *density-basin information.* The probabilities $P_{[pp]}$ are frequencies of sub-blocks pp of the string $S(\varepsilon)$ (Equation 13). The motivation to introduce this coefficient is twofold. First is estimating the *density* of peaks. For instance, the density-basin information of $\mathcal{P}$ is

$$h_{\mathcal{P}} = \frac{n - 2m - k}{n} \log_3 \frac{n}{n - 2m - k} \tag{23}$$

and it decreases when the number of groups increases (for a fixed value of m), i.e., the density of peaks becomes lower. Second is specifying whether the optima in the time series are isolated. Note that the information characteristics $H_{\mathcal{P}}$ and $M_{\mathcal{P}}$ will remain unchanged if we assume that the optima of $\mathcal{P}$ are not isolated (the optima have the form "$\dots111\bar{1}\bar{1}\bar{1}\dots$"). Therefore, by increasing the basin of attraction of the peaks, we change the density-basin information of a landscape path.

The information content and the density-basin information can be used to determine whether a landscape path is maximally multimodal. Thus, we propose the following assertion.

THEOREM 1: *If $H(\varepsilon) = \log_6 2$ and $h(\varepsilon) = 0$ for $\varepsilon = 0$, then the explored landscape path is either maximally multimodal, or an increasing or decreasing step function.*

Since $h(0) = 0$ and $H(0) > 0$, it follows that each block $s_i s_{i+1}$ of the string $S(0) = s_1 s_2 s_3 \dots s_n$ is nonhomogeneous, i.e., there are no flat and smooth sections in the landscape path. Furthermore, when $h(0)$ is zero, the minimal information content of the path presented by $S(0)$ is $\log_6 2$. This is due to the fact that the information content will be minimal when $s_i = s_{i+2}$ and $s_i \neq s_{i+1}$. Consequently, $S(0)$ is the string $pqpq\dots pq$, where $p \neq q$, and $P_{[pq]} = P_{[qp]} = \frac{1}{2}$. It is evident that there are three possibilities for the explored path. The path is

- maximally multimodal when p is $\bar{1}$ and q is 1 (p is 1 and q is $\bar{1}$).

- an increasing step function when p is 0 and q is 1 (p is 1 and q is 0).

- a decreasing step function when p is 0 and q is $\bar{1}$ (p is $\bar{1}$ and q is 0).

The increasing and decreasing step functions are easy for evolutionary search since the optimum is reachable by simple climbing and descending techniques, respectively. Alternatively, the maximally multimodal landscapes are difficult for search. However, there is a class of maximally multimodal landscapes that can be more easily searched. These are the landscapes that allow *regular adaptive* walks (Equation 18). Thus, the aforementioned assertion allows us to state the following: a landscape path is either maximally multimodal, an increasing or decreasing step function and is easy for evolutionary search if $H(\varepsilon) = \log_6 2$ and $h(\varepsilon) = 0$ for each $\varepsilon \in [0, \varepsilon^\star)$, where $\varepsilon^\star$ is the information stability.

3.5 Information Analysis

The analysis starts by performing a *walk* on the landscape using the evolutionary operator by which the neighborhood relationship of the landscape points is defined. For each step, the fitness value of the current point is recorded. Thus, for a certain number of steps a time series, $\{f_t\}_{t=0}^n$ will be obtained. Then the information functions $H(\varepsilon)$ and $M(\varepsilon)$ are calculated. In addition, we obtain the information stability $\varepsilon^\star$. We also suggest to calculate the density-basin information $h(0)$. Note that the accuracy of this procedure depends on the density of values of ε taken from the interval $[0, l_{\mathcal{I}}]$. A higher accuracy can be attained when the values of ε are given by $\varepsilon_i = i\delta$, where $0 \le i \le \frac{\varepsilon^*}{\delta}$ and $\delta \le \min\{|f_t - f_{t-1}|$ such that $0 < t \le n\}$.

4 Information Analysis of the Nk Landscapes

In this section, we apply the information analysis to the Nk model (Section 2.1) in order to investigate the structure of these landscapes. The aim is twofold: to illustrate the use of the information analysis and to find out what the analysis can tell us about the structure of the Nk landscapes.

Since the information analysis is based on the fitness difference of neighboring points, we calculate the difference of a pair of genotypes at Hamming distance 1. Thus, for a pair of genotypes g and $\tilde{g}$ we obtain

$$f(g) - f(\tilde{g}) = \frac{1}{N} \sum_{i=0}^{N-1} f_i(g) - \frac{1}{N} \sum_{i=0}^{N-1} f_i(\tilde{g}) = \frac{1}{N} \sum_{i \in C} (f_i(g) - f_i(\tilde{g})) \tag{24}$$

where C is a subset of loci that influence the fitness values when a genotype g differs from $\tilde{g}$ in one gene. If we denote $\alpha_i = f_i(g) - f_i(\tilde{g})$ then

$$f(g) - f(\tilde{g}) = \frac{1}{N}(\alpha_{i_0} + \dots + \alpha_{i_k}) \tag{25}$$

Note that $\alpha_{i_0}, \dots, \alpha_{i_k}$ is a sequence of independent random variables from the interval $(-1, 1)$. Using Equation 25 we can suggest the following:

1. The information content is strongly dependent on the distribution of the sequence $\alpha_{i_0}, \dots, \alpha_{i_k}$, which is the distribution of the loci fitness values. Usually, it is a uniform or U-shaped distribution (Kauffman, 1989) with mean 0. Hence, we can hypothesize that the probabilities $Pr\{f(g) - f(\tilde{g}) < -\varepsilon\}$ and $Pr\{f(g) - f(\tilde{g}) > \varepsilon\}$ are approximately equal for each value of ε, and the probability $Pr\{f(g) - f(\tilde{g}) = 0\}$ is negligible for the information content and the partial information content. Consequently, we expect the Nk landscapes to be multi-peaked without flat landscape areas.

2. The information stability is less than $\frac{k+1}{N}$ since

$$|f(g) - f(\tilde{g})| = \frac{1}{N}|\sum_{j=0}^{k} \alpha_{i_j}| \leq \frac{1}{N} \sum_{j=0}^{k} |\alpha_{i_j}| < \frac{k+1}{N} \qquad (26)$$

However, we expect the information stability to increase non-linearly as k increases because of the distribution of the sequence $\alpha_{i_0}, ..., \alpha_{i_k}$.

3. The information stability decreases toward 0 as N increases. The reason is twofold. First, increasing the parameter N causes the difference between the fitness values of each pair of neighboring points to decrease at least with rate $\frac{1}{N}$. Second, as the parameters k and N increase, with k being proportional to N, the sum $\sum_{j=0}^{k} \alpha_{i_j}$ becomes smaller since the distribution of the set $\alpha_{i_0}, ..., \alpha_{i_k}$ for large N has mean 0. This is known as *complexity catastrophe* (Kauffman, 1989). Consequently, when k and N both increase, the difference between the fitness values of each pair of neighboring points still decreases faster than $\frac{1}{N}$. However, it is not clear how the information content will change when the parameter N is increased, and what will be the contribution of the landscape objects to the information characteristics.

In order to reveal the "information" structure of the Nk landscapes, we perform two series of experiments in which we explore how the information characteristics are related to the parameters k and N. The next subsections give an experimental setup, followed by results of the analysis for k and N.

4.1 Experimental Setup

We apply the information analysis to Nk landscapes using the random neighborhood model. In our experiments, each landscape point is a genotype (a string of 0s and 1s), and two genotypes are neighbors if they differ from one another in one gene (note that we investigate the Nk model instead of the mutation and crossover landscapes generated by applying the corresponding evolutionary operators to this search space). Since the underlying graph is the Hamming graph, to perform a walk on the landscape we use a bitflip operator starting from a randomly chosen genotype.

As we mentioned above, we perform two series of experiments in which we study how the information characteristics are changed when we vary the parameters k and N. In the first series of experiments, N is set to 20. We measure the information characteristics $H(0)$, $M(0)$, $h(0)$, and $\varepsilon^{\star}$ for each value of k from 0 to $N-1$. We also calculate the information functions $H(\varepsilon)$ and $M(\varepsilon)$ for $k = 0, 2, 5, 10, 19$. The results for each k are calculated on a random walk of $10,000$ steps (isotropy has been assumed). In order to investigate how the modality of Nk landscapes is related to the parameter k, we measure the information content and the partial information content of time series obtained by a modification of the adaptive walks used by Kauffman (1989) and Weinberger (1990). The algorithm is implemented as follows: start from a randomly chosen point on the landscape, generate all neighbors of the current point, determine the fitness values of the neighbors, select a fitter one at random, continue to move until reaching a maximum, then save the optimum and move downhill in the same manner by choosing a worse neighbor at random. The walk terminates after a given number of steps unless a previously visited optimum is reached. The results for each $k = 0, 2, 5, 10, 19$ are averaged over 300 adaptive walks of no more than $1,000$ steps (standard deviations are also given).

The second series of experiments concern the question of how the information characteristics relate to the Nk landscapes when we vary the length of the genotype, N. We measure the information characteristics $H(0)$, $M(0)$, $h(0)$, and $\varepsilon^\star$ for several values of N from 0 to 900 and $k = 2$. To investigate the *complexity catastrophe* phenomenon we vary the parameter N and calculate the information content and stability of several landscapes under two conditions: first, a fixed value of k ($k = 2$) and, second, different values of k ($k = N - 1$). Again, the results for each pair of k and N are calculated on a random walk of 10,000 steps.

In our calculations, the parameter ε takes values from 0 to 1 given by $i\delta$, where $0 \leq i$, and δ is set to 0.0001.

4.2 The Parameter k and the Information Structure of Nk Landscapes
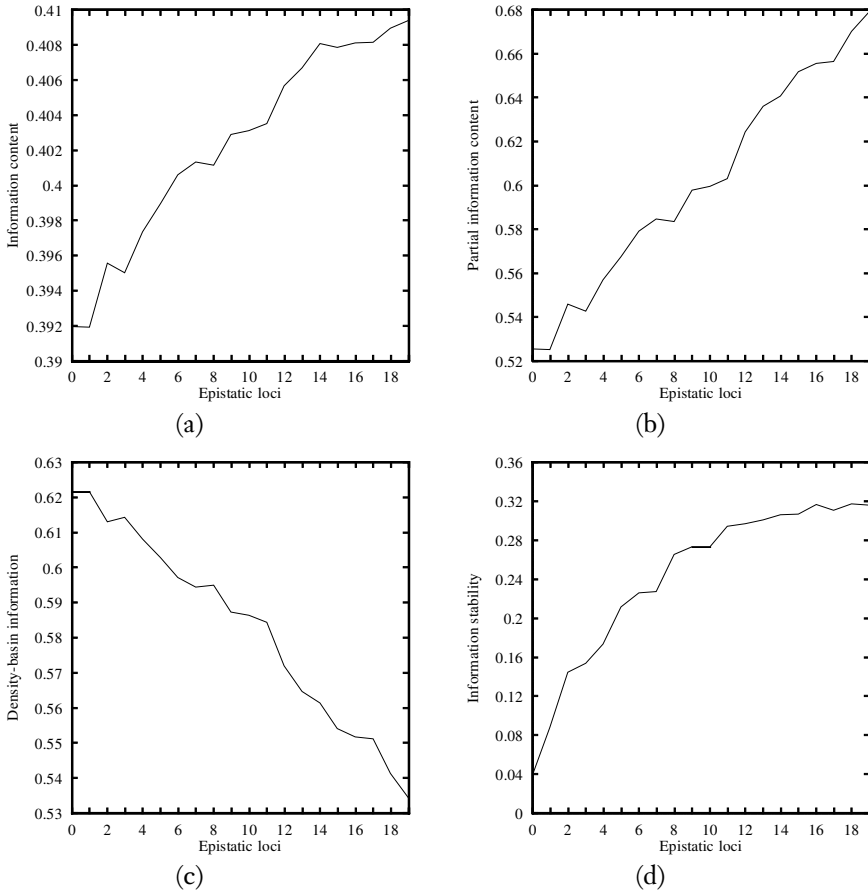

Figure 5: The information characteristics of Nk landscapes with random neighborhoods for $N = 20$ versus epistatic loci: (a) information content $H(0)$, (b) partial information content $M(0)$, (c) density-basin information $h(0)$, and (d) information stability.

The information characteristics of the Nk landscapes with random neighborhoods for different k are depicted in Figure 5. We can see that the information content $H(0)$,

the partial information content $M(0)$, and the information stability $\varepsilon^\star$ increase, and the density-basin information $h(0)$ decreases as the parameter k increases. Consequently, the landscapes become more rugged for higher values of k. Note, however, that the information content $H(0)$ only increases over quite a small interval. This indicates that, by moving k from 0 to $N-1$, we do not change drastically the structural type of the landscapes. Note that $H(0)$ is close to $\log_6 2$, and $M(0)$ is significantly higher than 0 for each value of k. Thus, given the assertion proposed in the previous section, we surmise that the Nk landscapes are composed of two kinds of local optima, and the number of *plateau forms* is negligible for our statistics. The conclusion is also supported by Table 2. The table represents the

Table 2: The probabilities of the landscape information objects for $\varepsilon = 0$ and different values of k, calculated on random walks of $10,000$ steps.

k	$P_{\bar{1}\bar{1}}$	$P_{\bar{1}0}$	$P_{\bar{1}1}$	$P_{0\bar{1}}$	P_{00}	P_{01}	$P_{1\bar{1}}$	P_{10}	P_{11}
0	0.2344	0	0.2655	0	0	0	0.2655	0	0.2346
2	0.2273	0	0.2728	0	0	0	0.2728	0	0.2271
5	0.2236	0	0.2836	0	0	0	0.2836	0	0.2092
10	0.1997	0	0.2997	0	0	0	0.2997	0	0.2009
19	0.162	0	0.3396	0	0	0	0.3396	0	0.1588

probabilities of landscape objects used in the calculations of the information characteristics $H(0)$ and $h(0)$. The results show that $Pr\{f(g) - f(\tilde{g}) < -\varepsilon\} \approx Pr\{f(g) - f(\tilde{g}) > \varepsilon\}$ and $Pr\{f(g) - f(\tilde{g}) = 0\} \approx 0$, which is to be expected. Hence, if we look at the plots of $H(0)$, $M(0)$, and $h(0)$ we can surmise that, by increasing k, we increase the number of local optima (the plots of $H(0)$ and $M(0)$), and the landscape optima become more and more isolated (the plot of $h(0)$). At the same time, the difference between the fitness values of the neighboring points become greater since the information stability increases (note that these conclusions assume that there are no plateau forms in the Nk landscapes). To support this, we calculate the information characteristics $H(0)$ and $M(0)$ of the time series obtained by adaptive walks on the Nk landscapes for $k = 0, 2, 5, 10, 19$ and $N = 20$. The results are given in Table 3. Note that, as k reaches $N-1$, the modality of the landscapes increases, however, when $k = N - 1$, the landscape is still not maximally multimodal.

Table 3: The information characteristics $H(0)$ and $M(0)$ of Nk landscapes with random neighborhoods for $N = 20$ and different values of k. The information characteristics are averaged over 300 adaptive walks of no more than $1,000$ steps.

k	$H(0)$	$M(0)$
0	0.08708 ± 0.00001	0.05979 ± 0.00001
2	0.11612 ± 0.00224	0.06809 ± 0.00147
5	0.14999 ± 0.00349	0.09084 ± 0.00289
10	0.2 ± 0.00363	0.1366 ± 0.00382
19	0.28911 ± 0.00421	0.26166 ± 0.00691

Additional information about the structure of the Nk landscapes can be derived from

the plots depicted in Figure 6. The figure presents the information functions $H(\varepsilon)$ and $M(\varepsilon)$ for different values of k. It is shown that the random walks on the landscapes are irregular (Section 3.3), and the degree of irregularity increases with the number of interactions within the genotypes. The degree of irregularity of a landscape (Section 3.3) is an estimate
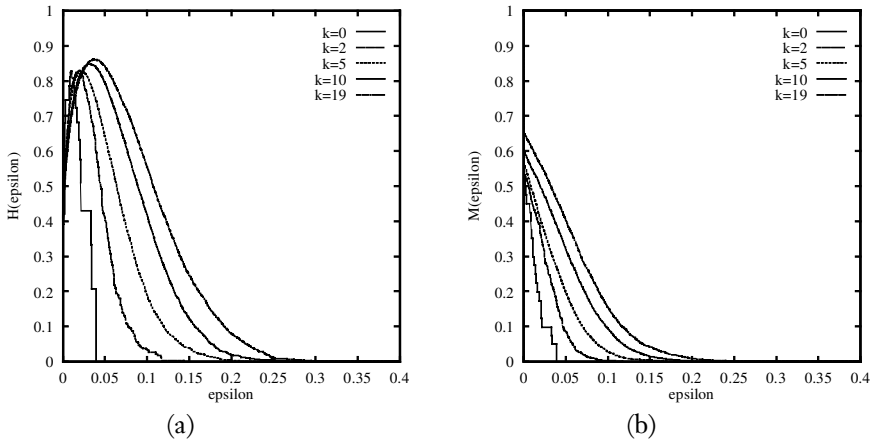

Figure 6: The information functions (a) $H(\varepsilon)$ and (b) $M(\varepsilon)$ of Nk landscapes with random neighborhoods for $N = 20$ and different values of k.

of the complexity of the walks among the points on the landscape. If a landscape is regular (Equation 18), the structure determined by the walks on the landscape is "homogeneous", and the information functions $H(\varepsilon)$ and $M(\varepsilon)$ will be constant functions when ε is less than the information stability. Figure 6 demonstrates that the complexity of the walks on the Nk landscapes increases as the interactions between the genes within the genotypes increase.

Another conclusion that can be derived from Figure 6 is that the diversity of local optima classified by their altitude increases as the parameter k increases. Note that the functions $H(\varepsilon)$ and $M(\varepsilon)$ are steeper for lower values of k. The steepest information functions characterize the smoothest landscape, while the less steep functions correspond to the more rugged landscapes. However, for $k > 0$, the steepness of $H(\varepsilon)$ and $M(\varepsilon)$ decreases when the functions become close to 0 for large ε. Consequently, there is a small number of neighboring points with a significantly different fitness values. This could be one of the reasons why the Nk landscapes are difficult for evolutionary search for $k \gg 0$.

4.3 The Complexity Catastrophe Causes an Information Collapse

We study how the information structure of the Nk landscapes changes when we vary the parameter N. Kauffman (1989) also investigated this and demonstrated that the fitness values of the attainable optima decrease toward 0.5 when k and N increase. He called this phenomenon *complexity catastrophe*. Here, we study this in terms of the information analysis and, in particular, when N is varied for a fixed value of k and when k and N increase with k being proportional to N.

The information characteristics of Nk landscapes for $k = 2$ and several values of N are depicted in Figure 7. The figure shows the information content, the partial information content, the density-basin information for $\varepsilon = 0$, and the information stability plotted as

functions of the parameter N. Each plot is evaluated on random walks of $10{,}000$ steps. We found it unnecessary to plot the functions $H(\varepsilon)$ and $M(\varepsilon)$ since their behavior differed little from that shown in Section 4.2.
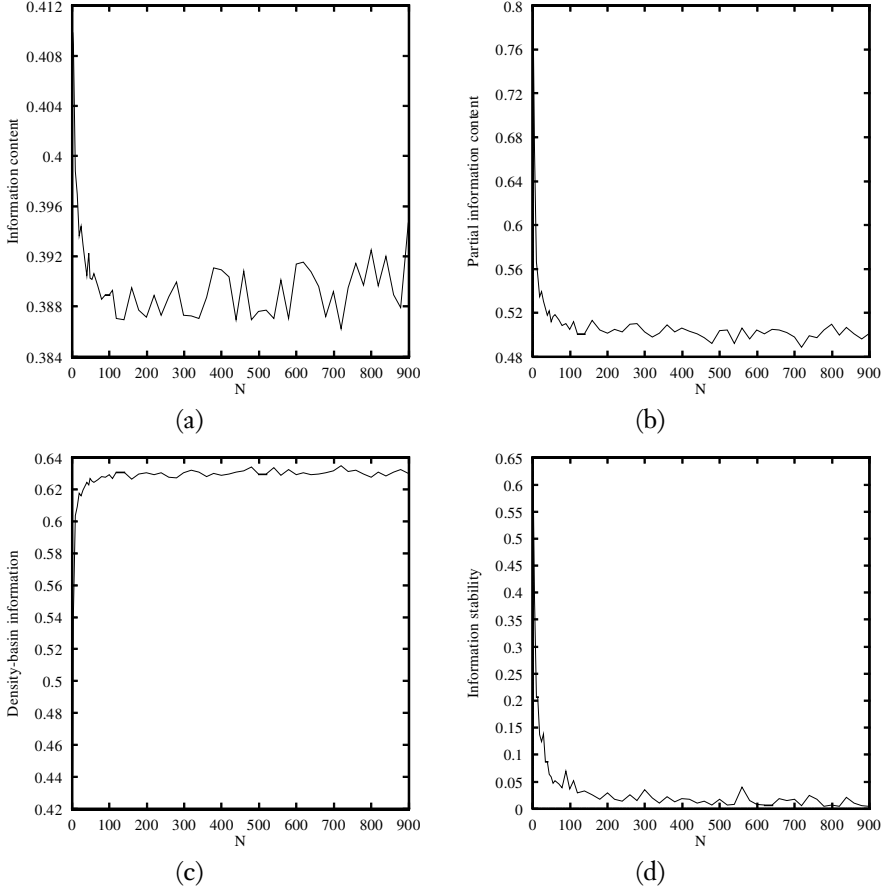

Figure 7: The information characteristics of Nk landscapes with random neighborhoods for $k = 2$ and different values of N: (a) $H(0)$, (b) $M(0)$, (c) $h(0)$, and (d) information stability.

The plots in Figure 7 suggest that the Nk landscapes become smoother as N increases. It can be seen that the information content and the partial information content decrease, while the density-basin information quickly increases to constant levels for small values of N. It can be seen that the levels attained by the information content, the partial information content, and the density-basin information are approximately $\log_6 2$, $\frac{1}{2}$, and $\log_3 2$, respectively. It follows that, for high values of N, it is equally likely that, during the random walk, the fitness will increase or decrease. The information stability decays in a similar manner to that of the other information characteristics. The plot in Figure 7 implies that as N increases, the information stability decreases toward 0 rapidly at first and more gradually for N approximately greater than 150. These results imply that the sizes of the basins of attraction of the optima increase with N.

The decays observed for the information characteristics when k is equal to $N - 1$ are different. We compare the information content and the information stability of several Nk landscapes with random neighborhoods. The results are shown in Figure 8. The left hand side of the figure shows the plots of (a) the information content and (b) the information stability obtained for $k = 2$ and $k = N - 1$. The right hand side of Figure 8 shows how the difference (discrepancy) between the plots of the information characteristics ((a) information content and (b) information stability) is related to the parameter N.
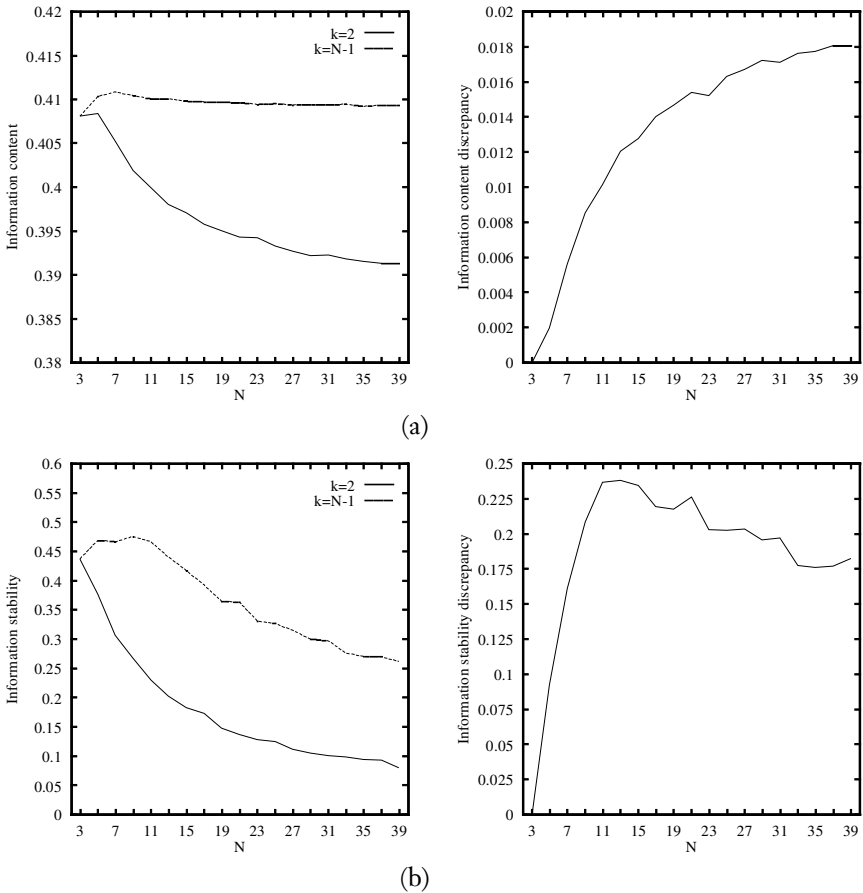

Figure 8: The information decay of Nk landscapes with random neighborhoods: (a) information content and (b) information stability. The left figures show information characteristics, and the right figures show discrepancy between the information characteristics for $k = N - 1$ and $k = 2$.

The plots in Figure 8(a) suggest that as the number of genes N increases, the information content decreases, and the degree to which the information decays depends on the constancy of the ratio of k to N. When the parameter k is fixed, the information content decreases rapidly. When the parameter k increases such that it is proportional to N, the information content decreases linearly. Consequently, by increasing the parameter N, we do not drastically change the structure of these landscapes.

In contrast to the behavior of the information content with increasing N, the infor-

mation stability decreases in a different manner. Once again, the degree of the information decay depends on k, however, the relation between k and the information stability is different from that exhibited by the information content. The plots in Figure 8(b) show that when k is proportional to N, the information stability decreases toward 0 faster than the plot corresponding to a fixed value of k. We suppose that for a very large value of N, a critical level k_c of the interactions among the genes in the genotypes exists for which the information stability increases as k increases towards k_c and decreases as k increases from k_c to $N-1$.

5   Discussion

The previous sections have shown how the described fitness landscape analysis is a useful tool for studying the structure of landscapes. The technique is simple and has the following advantages:

1. It allows a measure of the diversity of the local optima. Calculating the information content of a landscape, we obtain information about the landscape profile (the possible *shapes* of landscape optima). Applying the analysis to the Nk model, we showed that the Nk landscapes are multi-peaked without plateau forms. The nature of the landscape structure is only slightly changed when we vary the parameters of the model.

2. The analysis can assist in a study of the landscape modality. We introduced the partial information content that is an estimate of the modality when applied to a time series obtained by an adaptive walk. In addition, we suggested the density-basin information that characterizes the diversity of flat and smooth sections of the explored time series. The similarity between the plots of the information characteristics, depicted in the previous section, again implied that the nature of the Nk landscape structure is relatively invariant to changes in the parameters of the model.

3. Filtering the information content and the partial information content, we constructed information functions $H(\varepsilon)$ and $M(\varepsilon)$ that revealed how the landscape structure changes with the scale of observation. This allowed a measurement of the degree of regularity of walks on a landscape, and, thus, we showed that by increasing the parameter k, we increase the irregularity of walks on Nk landscapes.

4. By measuring the regularity of walks on maximally multimodal landscapes, we can determine whether or not they are easy for evolutionary search (the assertion in Section 3.4). We suspect that by studying the degree of regularity of a maximally multimodal landscape, the landscape might be transformed so that it becomes easier for evolutionary search.

5. Calculating the information functions, we obtain the information stability of the landscape that is an estimate of the largest fitness difference encountered in neighboring points. The above experiments showed that for higher values of k, the information stability of Nk landscapes grows non-linearly. This is caused by the dependence on the distribution of the fitness values. Furthermore, by studying the complexity catastrophe phenomenon, we demonstrated that for large values of N, a critical level of interactions between the genes exists beyond which the information stability decreases towards 0 and, hence, tends to an information collapse.

Thus far we demonstrated how the information analysis can be applied to landscapes, particularly Nk landscapes, in order to characterize their structure. In our investigation, the sequences of fitness values were obtained by walks which were performed by a bitflip operator. We gave an example of how to apply and to analyze the results of such an investigation. With an understanding of the analysis, other types of landscapes can be studied. Essentially, one only needs to change the algorithm for generating the time series. Note that the evolutionary operators define the neighborhood in the landscapes, and, for instance, to perform a walk on a crossover landscape, we should apply the corresponding crossover operator in order to move from one landscape point to another.

The approach presented is general and directly applicable to different variants of evolutionary algorithms such as messy genetic algorithms (Goldberg et al., 1989) and genetic programming (Koza, 1992). For instance, the information analysis was employed by Nikolaev and Slavov (1998) in studying how the novel *context-preserving* mutation operator for genetic program trees is better than other types. The generality and the advantages of the analysis are also demonstrated in the following study of the onepoint mutation and uniform crossover landscapes associated with a two-bit multiplier evolved by an evolutionary algorithm. This is an example of digital circuit evolution (Miller et al., 1997; Miller and Thomson, 1998) and has been recently studied in the field of evolvable hardware (Sipper et al., 1997).

Digital circuit evolution refers to the design of electronic circuitry in which configurations of logic gates for some pre-specified computational task can emerge in a population of gate arrays using artificial evolution. The correlation characteristics of the fitness landscapes associated with the digital circuit evolution on an idealized model of a field-programmable gate array were studied by Vassilev et al. (1999). The difficulty in studying the structure of these landscapes stems from the genotype representation that allows us to evolve the functionality and connectivity of the gate array. The genotypes are constructed over three configuration spaces which are defined over two completely different alphabets. The spaces represent all the possible configurations of functionality and the internal and output connectivity of the encoded gate array. This gives rise to highly non-isotropic landscapes which makes their investigation much more convoluted.

In Vassilev et al. (1999), a model for studying the structure of circuit evolution landscapes was proposed (Appendix B). The model is based on the idea that a landscape might be decomposed to subspaces that are suitable for a statistical investigation (Hordijk, 1997). Here we employ the model in order to study the information characteristics of the subspaces of onepoint mutation and uniform crossover landscapes defined on a two-bit multiplier that is evolved by an evolutionary algorithm (Appendix B). Since there are three configuration spaces, we estimate the information characteristics of six time series obtained by random walks[2] of length $100,000$ on the corresponding onepoint mutation and uniform crossover subspaces.

---

[2] A random walk on a mutation landscape is implemented as follows: start from a genotype (in our case, it is the evolved two-bit multiplier), generate all neighbors of the current point by mutation and evaluate their fitness values, choose randomly one neighbor and record its fitness, generate all neighbors of the new point, which becomes "current", and so on (Weinberger, 1990). A random walk on a crossover landscape is implemented by the algorithm given in Wagner and Stadler (1998) (see also Stadler and Wagner (1997)). In short, it can be described as follows: start with a pair of genotypes, generate a set of offsprings by applying the crossover operator and evaluate their fitness values, from those randomly choose one (record its fitness) and mate it with a randomly chosen genotype, etc. until the termination conditions are satisfied. Here, we perform the walks with respect to the investigated subspace (Appendix B).
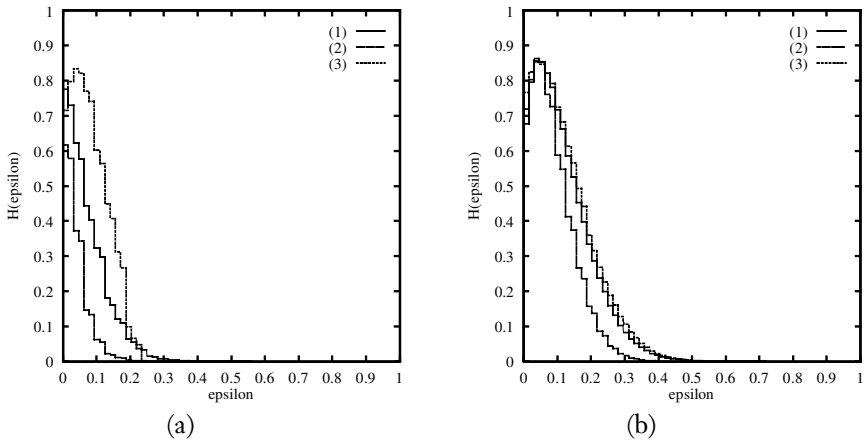
Figure 9: The information functions $H(\varepsilon)$ of (1) functionality, (2) internal connectivity, and (3) output connectivity subspaces of the two-bit multiplier landscape obtained by (a) onepoint mutation and (b) uniform crossover operators.

Figure 9(a) represents the information functions $H(\varepsilon)$ of the (1) functionality, (2) internal connectivity, and (3) output connectivity subspaces of the onepoint mutation landscape. The subspaces are characterized with flat landscape areas since the information content $H(0)$ of each subspace is significantly higher than $\log_6 2$. The plots of $H(\varepsilon)$ also imply subspaces with significantly different profiles. For instance, the information functions of the functionality and internal connectivity subspaces decrease as ε increases, while $H(\varepsilon)$ of the output connectivity subspace increases as ε increases from 0 to approximately 0.047. Consequently, the plateau forms in the functionality and internal connectivity subspaces prevail in the corresponding ensembles of information objects. This is not so for the output connectivity subspace.

Figure 9(b) represents the information functions $H(\varepsilon)$ of the (1) functionality, (2) internal connectivity, and (3) output connectivity subspaces of the uniform crossover landscape. Again, the information content $H(0)$ of each subspace is significantly higher than $\log_6 2$, which implies the existence of plateaus on the landscape. The depicted information functions $H(\varepsilon)$ reveal that these subspaces have similar profiles. The subspaces are characterized with multi-peak configurations which prevail over the flat landscape areas. Consequently, the uniform crossover landscape has higher modality than the onepoint mutation one, which implies a landscape with higher ruggedness. It agrees with Vassilev et al. (1999) who suggested that the uniform crossover operator is not the right operator for this optimization problem.

The plots in Figure 9 suggest a certain similarity between the mutation and crossover subspaces. Since the plots of $H(\varepsilon)$ represent step functions, we surmise that the landscapes consist of a small number of altitude levels which, together with the flat landscape areas, imply vast and sharply differentiated landscape plateaus. It suggests that an evolutionary strategy with simple elitism and a small population might be much more effective to search in these landscapes which agrees with the findings in Miller (1999).

6 Conclusions

Often when we investigate the structure of a fitness landscape, we measure the correlation characteristics of the landscape, which is the easiest reliable way to assess the landscape ruggedness. However, the correlation measures provide generalized and often insufficient information about the landscape structure. Many features of the structure of a landscape influence the ability of a population to *search*, and it is difficult from studying only the ruggedness of the landscape to determine the likelihood of the evolutionary search to succeed.

In this paper, a concept for studying the structure of fitness landscapes was proposed. We chose a standard analysis of rugged systems and applied it to fitness landscapes, ignoring certain non-essential information about the structure of the landscapes. We introduced a technique for analyzing the profile of a landscape that is based on three basic information characteristics: information content, partial information content, and information stability. The first two information characteristics were defined as estimates of the amount of information of the landscape, while the information stability was defined as a result of filtering out the estimated information content. To measure the amount of information in a landscape, represented as an ensemble of objects, does not necessarily imply measuring the ruggedness of the landscape. At this point, the information analysis differs significantly from many others statistical approaches. For instance, consider an estimate of the correlation characteristics of a landscape. The lower the correlations, the more rugged the landscape. The information analysis is different. The analysis gives us a notion of what is the interplay between the smooth, rugged, and flat landscape areas.

The information analysis was applied to the Nk model of landscapes. The aim was to investigate what the analysis says about the structure of landscapes with well known characteristics. The results of the study indicated that the technique was reliable and capable of providing useful information about the landscape structure. The usefulness of the analysis has been also demonstrated by the study of the two-bit multiplier landscapes, which revealed how the analysis could be helpful for a better understanding of the evolutionary search on these landscapes. To summarize, we demonstrated that the proposed statistical analysis technique is an appropriate tool for exploring the structure of fitness landscapes, and, further, it gives information that is not easily attainable using the other statistical analyses.

Acknowledgments

A Fitness Landscapes – Basic Definitions

The appendix provides some definitions related to the theory of fitness landscapes, most of them given by Jones (1995).

Consider a landscape $\mathcal{L}$ defined over a graph (V_ϕ, E_ϕ) where V_ϕ and E_ϕ are the sets of vertices and edges, respectively, and ϕ is the evolutionary operator that specifies the

neighborhood in the graph. Also, let f be a fitness function that assigns each vertex from V_ϕ with a real number from a certain interval.

DEFINITION A.1 (Neighborhood and Neighbor): *Given an evolutionary operator ϕ, the* **neighborhood** *of $v \in V_\phi$ is the set*

$$N_\phi(v) = \{w \in V_\phi \mid \phi(v) = w\} \qquad (27)$$

If $w \in N_\phi(v)$ we say that w is a **neighbor** *of v.*

DEFINITION A.2 (Neighborhood and Neighbor 2): *Let A be a subset of V_ϕ. For an evolutionary operator ϕ, the* **neighborhood** *of A is*

$$N_\phi(A) = \{w \in V_\phi \backslash A \mid \phi(v) = w \text{ and } v \in A\} \qquad (28)$$

The vertex w is a **neighbor** *of A if $w \in N_\phi(A)$.*

DEFINITION A.3 (Peak): *A vertex $v \in V_\phi$ is a* **peak** *iff, for each $w \in N_\phi(v)$,*

$$f(v) > f(w) \qquad (29)$$

DEFINITION A.4 (Global Maximum (Optimum)): *A* **global maximum** *of a landscape is a vertex $v \in V_\phi$ such that*

$$f(v) \geq f(w) \qquad (30)$$

for all $w \in V_\phi$.

DEFINITION A.5 (Local Maximum (Optimum)): *A* **local maximum** *of a landscape is a vertex $v \in V_\phi$ such that*

$$f(v) \geq f(w) \qquad (31)$$

for all $w \in N_\phi(v)$.

DEFINITION A.6 (Plateau): *A* **plateau** *is a set $P \subset V_\phi$ of two or more vertices so that for each $v_0, v_n \in P$ a subset $v_1, ..., v_{n-1}$ exists where $f(v_i) = f(v_{i+1})$ and $v_{i+1} \in N_\phi(v_i)$ for every i, $0 \leq i < n$.*

DEFINITION A.7 (Basin of Attraction): *A* **basin of attraction** *of a vertex v_n is the set of vertices*

$$
\begin{aligned}
B_\phi(v_n) \quad = \quad & \{v_0 \in V_\phi \mid \exists v_1, ..., v_{n-1} \in V_\phi \text{ with } v_{i+1} \in N_\phi(v_i) \text{ and } f(v_{i+1}) > f(v_i) \\
& (\text{or } f(v_{i+1}) < f(v_i) \text{ if minimising}) \text{ for each } i, 0 \leq i < n\} \qquad (32)
\end{aligned}
$$

We consider that the number of vertices in the basin of attraction of a vertex, $v \in V_\phi$, defines the *size* of the basin. Thus, *isolated optima* are the peaks with smallest basins of attraction. We also say that the size of a basin of attraction specifies the *degree of isolation* of a particular optimum. The higher the degree of isolation, the smaller the basin of attraction.

B The Two-Bit Multiplier Landscapes

In this appendix, we present the model of the digital circuit evolution landscapes and the two-bit multiplier whose fitness landscapes are studied in the paper (further details can be found in Vassilev et al. (1999)).

B.1 Evolving Digital Circuits – Landscapes

We consider a genotype representation of an idealized field-programmable gate array based on that proposed in Miller and Thomson (1998). The structure of the genotype is dependent on the parameters of the gate array. These are the number of inputs, the number of rows and columns, the number of allowed logic functions, the radius range of internal connectivity, which is also known as *levels-back* parameter, and, of course, the functionality of the evolved circuit or the fitness function that also determines the number of outputs.

The genotype is a composition of three different parts which are responsible for: first, the gates functionality, second, the array internal connectivity, and third, the array outputs. For convenience, we adopt the term *chromosome* to mean a genotype part. Thus, each genotype becomes a composition of three chromosomes with different length that are defined over two completely different alphabets. The "gate functionality" chromosomes are strings over alphabet α with a length equal to the number of gates. The "internal connectivity" and "array outputs" chromosomes are defined over alphabet β, and they are strings with a length equal to the number of gates and the number of array outputs, respectively.

The alphabet α is a set of integers that represents the allowed logic functions. Therefore, the alphabet size l_α is the number of logic functions used in the circuit design. The alphabet β is related to the size of the neighborhood of the cells and array outputs that is dependent upon the levels-back parameter. Again, the alphabet is a set of integers, however, they are reference numbers of the elements of a neighborhood. Hence, the size of β is $l_\beta = \begin{cases} nL, & \text{if } L \leq m \\ nm + n_I, & \text{otherwise} \end{cases}$, where n is the number of rows, m is the number of columns, n_I is the number of inputs of the gate array, and L is the levels-back.

Since each genotype consists of three chromosomes, we assume that the original landscape for a given evolutionary operator is a product of three configuration spaces defined over alphabets α and β. Consider the hypergraphs $\mathcal{G}_{l_\alpha}^{nm}$, $\mathcal{G}_{l_\beta}^{3nm}$, and $\mathcal{G}_{l_\beta}^{no}$ that represent the configuration spaces of the chromosomes responsible for functionality, connectivity, and output connections, respectively. Let ϕ denote an evolutionary operator. The digital circuit evolution landscapes are

$$\{\mathcal{G}_{f_i}, f_i, \phi\}_{i=0}^{l_\beta^{3nm+n_O}-1}, \quad \{\mathcal{G}_{g_i}, g_i, \phi\}_{i=0}^{l_\alpha^{nm} l_\beta^{no}-1}, \quad \{\mathcal{G}_{h_i}, h_i, \phi\}_{i=0}^{l_\alpha^{nm} l_\beta^{3nm}-1} \tag{33}$$

The graphs $\mathcal{G}_{f_i}$, $\mathcal{G}_{g_i}$, and $\mathcal{G}_{h_i}$ are obtained by assigning each vertex from $\mathcal{G}_{l_\alpha}^{nm}$, $\mathcal{G}_{l_\beta}^{3nm}$, and $\mathcal{G}_{l_\beta}^{no}$, respectively, with a fitness value. The fitness values are provided by fitness functions $\{f_i\}_{\forall i}$, $\{g_i\}_{\forall i}$, and $\{h_i\}_{\forall i}$, which are defined as follows

$$(1) \quad \forall i(\mathbf{c}_i \in \mathcal{G}_{l_\beta}^{3nm} \times \mathcal{G}_{l_\beta}^{no}), \forall \mathbf{x} \in \mathcal{G}_{l_\alpha}^{nm} : f_i(\mathbf{x}) = F(\mathbf{x} \cdot \mathbf{c}_i)$$

$$(2) \quad \forall i(\mathbf{c}_i \in \mathcal{G}_{l_\alpha}^{nm} \times \mathcal{G}_{l_\beta}^{no}), \forall \mathbf{x} \in \mathcal{G}_{l_\beta}^{3nm} : g_i(\mathbf{x}) = F(\mathbf{x} \cdot \mathbf{c}_i)$$

$$(3) \quad \forall i(\mathbf{c}_i \in \mathcal{G}_{l_\alpha}^{nm} \times \mathcal{G}_{l_\beta}^{3nm}), \forall \mathbf{x} \in \mathcal{G}_{l_\beta}^{no} : h_i(\mathbf{x}) = F(\mathbf{x} \cdot \mathbf{c}_i) \tag{34}$$

The function F is defined over the genotype space (the operator "$\cdot$" is considered to merge the strings in a special way so that the genotype structure is obtained), and it evaluates the percentage of correctness of the represented circuit. Thus, for each family of landscapes we have a group of fitness functions, and each fitness function estimates only a part of the genotype. Hence, its index is determined by the constant string $\mathbf{c}_i$ which is the remainder of the genotype.

B.2   The Two-Bit Multiplier

The two-bit multiplier is evolved on a 3×4 array of logic cells. The levels-back parameter is 4, and the number of allowed two-input logic functions is 10 (Vassilev et al., 1999). The genotype is shown in Table 4, and the corresponding phenotype is depicted in Figure 10. The most significant bits are inputs 1 and 3, and output 1.

Table 4: The genotype of the two-bit multiplier evolved on a 3×4 gate array. The column "Cell" gives the label of the cell whose genetic information is listed on the right side at column "Allele(Gene)". The column "Allele(Gene)" represents the genotype where the label of the gene is given in parentheses.

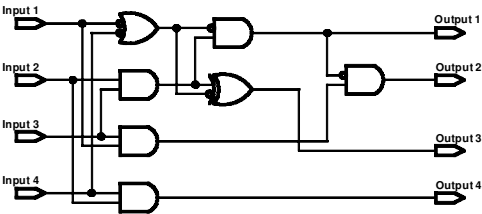| Cell	Allele(Gene)			
4	$3_{(0)}$	$0_{(1)}$	$3_{(2)}$	$15_{(3)}$
5	$0_{(4)}$	$3_{(5)}$	$1_{(6)}$	$6_{(7)}$
6	$2_{(8)}$	$0_{(9)}$	$2_{(10)}$	$6_{(11)}$
7	$2_{(12)}$	$4_{(13)}$	$1_{(14)}$	$9_{(15)}$
8	$1_{(16)}$	$2_{(17)}$	$2_{(18)}$	$6_{(19)}$
9	$5_{(20)}$	$6_{(21)}$	$5_{(22)}$	$15_{(23)}$
10	$4_{(24)}$	$1_{(25)}$	$7_{(26)}$	$14_{(27)}$
11	$8_{(28)}$	$4_{(29)}$	$2_{(30)}$	$7_{(31)}$
12	$3_{(32)}$	$1_{(33)}$	$7_{(34)}$	$6_{(35)}$
13	$4_{(36)}$	$8_{(37)}$	$8_{(38)}$	$11_{(39)}$
14	$6_{(40)}$	$11_{(41)}$	$12_{(42)}$	$7_{(43)}$
15	$0_{(44)}$	$12_{(45)}$	$9_{(46)}$	$15_{(47)}$
Outputs	$11_{(48)}$	$14_{(49)}$	$13_{(50)}$	$12_{(51)}$

Figure 10: The schematic of the two-bit multiplier, obtained by artificial evolution.

References

Altenberg, L. (1995). The schema theorem and price's theorem. In Whitley, L. D. and Vose, M. D., editors, *Foundations of Genetic Algorithms*, Volume 3, pages 23–49. Morgan Kaufmann, San Francisco, California.

Altenberg, L. (1997). Nk fitness landscapes. In Bäck, T., Fogel, D. B. and Michalewicz, Z., editors, *Handbook of Evolutionary Computation*, Chapter B2.7:2. Oxford University Press, New York, New York.

Barnsley, M. (1993). *Fractals Everywhere*. Academic Press Professional, San Diego, California.

Box, G. and Jenkins, G. (1970). *Time Series Analysis, Forecasting and Control*. Holden Day, San Francisco, California.

Chaitin, G. J. (1987). *Information, Randomness & Incompleteness*. World Scientific, Singapore.

Goldberg, D. (1987). Simple genetic algorithms and the minimal deceptive problem. In Davis, L., editor, *Genetic Algorithms and Simulated Annealing*, pages 74–88. Pitman, London, England.

Goldberg, D. (1989). Genetic algorithms and walsh functions: Part ii, deception and its analysis. *Complex Systems*, 3:153–171.

Goldberg, D., Korb, B. and Deb, K. (1989). Messy genetic algorithms: Motivation, analysis and first results. *Complex Systems*, 3:493–530.

Grefenstette, J. J. and Baker, L. E. (1989). How genetic algorithms work: A critical look at implicit parallelism. In Schaffer, J. D., editor, *Proceedings of the Third International Conference on Genetic Algorithms*, pages 20–27, Morgan Kaufmann, San Mateo, California.

Gutowitz, H. (1991). *Cellular Automata: Theory and Experiment*. MIT Press, Cambridge, Massachusetts.

Hordijk, W. (1996). A measure of landscapes. *Evolutionary Computation*, 4(4):335–360.

Hordijk, W. (1997). Correlation analysis of the synchronising-ca landscape. *Physica D*, 107:255–264.

Horn, J. and Goldberg, D. (1995). Genetic algorithm difficulty and the modality of fitness landscapes. In Whitley, L. D. and Vose, M. D., editors, *Foundations of Genetic Algorithms*, Volume 3, pages 243–269. Morgan Kaufmann, San Francisco, California.

Jones, T. (1995). *Evolutionary Algorithms, Fitness Landscapes and Search*. Ph.D. thesis, University of New Mexico, Albuquerque, New Mexico.

Kauffman, S. (1989). Adaptation on rugged fitness landscapes. In Stein, D., editor, *Lectures in the Sciences of Complexity*, SFI Studies in the Sciences of Complexity, pages 527–618. Addison-Wesley, Reading, Massachusetts.

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

Langton, C. G. (1990). Computation at the edge of chaos: Phase transitions and emergent computation. *Physica D*, 42:12–37.

Li, W., Packard, N. H. and Langton, C. G. (1990). Transition phenomena in cellular automata rule space. *Physica D*, 45:77–94.

Lipsitch, M. (1991). Adaptation on rugged landscapes generated by iterated local interactions of neighboring genes. In Belew, R. K. and Booker, L. B., editors, *Proceedings of the Fourth International Conference on Genetic Algorithms*, pages 128–135, Morgan Kaufmann, San Mateo, California.

Manderick, B., de Weger, M. and Spiessens, P. (1991). The genetic algorithm and the structure of the fitness landscape. In Belew, R. K. and Booker, L. B., editors, *Proceedings of the Fourth International Conference on Genetic Algorithms*, pages 143–150, Morgan Kaufmann, San Mateo, California.

Miller, J. F. (1999). An empirical study of the efficiency of learning boolean functions using a cartesian genetic programming approach. In Banzhaf, W., Daida, J., Eiben, A. E., Garzon, M. H., Honavar, V., Jakiela, M. and Smith, R. E., editors, *Proceedings of the First Genetic and Evolutionary Computation Conference*, Volume 2, pages 1135–1142, Morgan Kaufmann, San Francisco, California.

Miller, J. F. and Thomson, P. (1998). Aspects of digital evolution: Evolvability and architecture. In Eiben, A. E., Bäck, T., Schoenauer, M. and Schwefel, H.-P., editors, *Parallel Problem Solving from Nature V*, Volume 1498 of *Lecture Notes in Computer Science*, pages 927–936, Springer, Berlin, Germany.

Miller, J. F., Thomson, P. and Fogarty, T. (1997). Designing electronic circuits using evolutionary algorithms arithmetic circuits: A case study. In Quagliarella, D., Periaux, J., Poloni, C. and Winter, G., editors, *Genetic Algorithms and Evolution Strategies in Engineering and Computer Science*, pages 105–131. Wiley, Chichester, England.

Mitchell, M., Forrest, S. and Holland, J. (1991). The royal road for genetic algorithms: Fitness landscapes and GA performance. In Varela, J. and Bourgine, P., editors, *Proceedings of the First European Conference on Artificial Life*, pages 245–254, MIT Press, Cambridge, Massachusetts.

Nikolaev, N. and Slavov, V. (1998). Concepts of inductive genetic programming. In Banzhaf, W., Poli, R., Schoenauer, M. and Fogarty, T., editors, *Proceedings of the First European Workshop on Genetic Programming*, Volume 1391 of *Lecture Notes in Computer Science*, pages 49–59, Springer-Verlag, Berlin, Germany.

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, 27(3):379–423.

Sipper, M., Sanchez, E., Mange, D., Tomassini, M., Pérez-Uribe, A. and Stauffer, A. (1997). A phylogenetic, ontogenetic, and epigenetic view of bio-inspired hardware systems. *IEEE Transactions on Evolutionary Computation*, 1(1):83–97.

Slavov, V. and Nikolaev, N. (1999). Genetic algorithms, fitness sublandscapes and subpopulations. In Reeves, C. and Banzhaf, W., editors, *Foundations of Genetic Algorithms*, Volume 5, pages 199–218. Morgan Kaufmann, San Francisco, California.

Stadler, P. F. (1995). Towards theory of landscapes. In Lopéz-Peña, R., Capovilla, R., García-Pelayo, R., Waelbroeck, H. and Zertuche, F., editors, *Complex Systems and Binary Networks*, pages 77–163. Springer-Verlag, Berlin, Germany.

Stadler, P. F. and Wagner, G. P. (1997). Algebraic theory of recombination spaces. *Evolutionary Computation*, 5(3):241–275.

Theiler, J. (1990). Estimating fractal dimension. *Journal of Optical Society of America*, 7(6):1055–1073.

Vassilev, V. K. (1997a). Information analysis of fitness landscapes. In Husbands, P. and Harvey, I., editors, *Proceedings of the Fourth European Conference on Artificial Life*, pages 116–124, MIT Press, Cambridge, Massachusetts.

Vassilev, V. K. (1997b). An information measure of landscapes. In Bäck, T., editor, *Proceedings of the Seventh International Conference on Genetic Algorithms*, pages 49–56, Morgan Kaufmann, San Francisco, California.

Vassilev, V. K., Miller, J. F. and Fogarty, T. C. (1999). Digital circuit evolution and fitness landscapes. In *Proceedings of the Congress on Evolutionary Computation*, Volume 2, pages 1299–1306, IEEE Press, Piscataway, New Jersey.

Wagner, G. P. and Stadler, P. F. (1998). Complex adaptations and the structure of recombination spaces. In Nehaniv, C. and Ito, M., editors, *Algebraic Engineering*. World Scientific, Singapore. Santa Fe Institute Report 97-03-029.

Weinberger, E. D. (1990). Correlated and uncorrelated fitness landscapes and how to tell the difference. *Biological Cybernetics*, 63:325–336.

Weinberger, E. D. and Stadler, P. F. (1993). Why some landscapes are fractal. *Journal of Theoretical Biology*, 163:255–275.

Wright, S. (1932). The roles of mutation, inbreeding, crossbreeding and selection in evolution. In Jones, D. F., editor, *Proceedings of the Sixth International Conference on Genetics*, Volume 1, pages 356–366.