Bitwise Regularity and GA-Hardness.

Benoît Leblanc, Evelyne Lutton.
INRIA - Rocquencourt, B.P. 105, F-78153 LE CHESNAY Cedex, France
Tel: +33 (0)1 39 63 55 23 - Fax: +33 (0)1 39 63 59 95
e-mail: Benoit.Leblanc@inria.fr, Evelyne.Lutton@inria.fr
`http://www-rocq.inria.fr/fractales/`

Abstract— **We present in this paper a theoretical analysis that relates an irregularity measure of a fitness function to the so-called GA-deception. This approach is a continuation of a work [18] that has presented a deception analysis of Hölder functions. The analysis developed here is a generalization of this work in two ways: we first use a "bitwise regularity" instead of a Hölder exponent as a basis for our deception analysis, second, we perform a similar deception analysis of a GA with uniform crossover. We finally propose to use the bitwise regularity coefficients in order to analyze the influence of a chromosome encoding on the GA efficiency, and we present experiments with Gray encoding.**

Keywords— **Genetic Algorithms, optimization, bitwise regularity, deception analysis, fractals, Hölder functions.**

I. Introduction.

Theoretical investigations on GA and EA in general concern mainly convergence analysis (and convergence speed analysis on a locally convex optimum for EA), influence of the parameters, and GA-hardness analysis. For GA, our main concern here, these analyses are based on different approaches :

- Proofs of convergence based on Markov chain modeling [6], [3], [1], [20].
- Deceptive functions analysis, based on Schema analysis and Holland's original theory [14], [8], [9], [11], which characterizes the efficiency of a GA, and allows to shed light on "GA-hard" functions.
- Some rather new approaches are based on an explicit modelization of a GA as a dynamical system [16], [22].

Deception has been intuitively related to the biological notion of epistasis [5], which can be understood as a sort of "non-linearity" degree. It can also be related, to the so called "fitness landscape" analyses (see for example [19]). In any ways, it basically depends on :

- the parameter setting of the GA,
- the shape of the function to be optimized,
- the chromosome encoding , i.e. the "way" of scanning the search space.

In a previous work [18] it has been proven that some tools, that have been developed in the framework of fractal theory, can be used in order to refine a deception analysis of Genetic Algorithms. This work has mainly related an irregularity measure (Hölder exponent) of the function to be optimized to its deceptiveness. We first recall in section II these results, that allow to model the influence of some of the GA parameters. The main hypothesis of this previous analysis is that the fitness function can be considered as the sampling of an underlying continuous Hölder function. In section III we will then present a generalization of this work that considers another regularity measure of the fitness function, the bitwise regularity, and which does not support the Hölder hypothesis anymore.

The GA modeled in this framework is the so-called canonical GA, i.e. with proportional selection (roulette wheel selection), one point crossover and mutation, at fixed rates p_c and p_m all along the GA run. In section IV we then present a similar theoretical analysis for a canonical GA with uniform crossover, which is an operator that is largely used in real world applications.

Besides the intuitive fact that it relates the irregularity of the fitness function to its "difficulty", one important application of this theoretical analysis is that it provides a mean to "measure" (of course to a certain extent, due to the intrinsic limitations of deception theory) the influence of the chromosome encoding. We present in section V some experimentations with the Gray encoding that prove the interest of such an approach.

II. Background and previous work.

In this section we quickly remind the definitions of schemata, deception and Hölder exponents.

A. Schemata.

Schemata has been widely studied in the field of GA, and are the basis of the deception analysis. A schema corresponds to a subset of the space $\Omega^l = \{0,1\}^l$ (the space of binary strings of length l for a GA using binary encoding), or more precisely a hyper-plan of Ω^l. An additional symbol '*', representing a wildcard ('0' or '1') is used to represent a schema. For example, if $l = 4$, the strings $i_1 = 0101$ and $i_2 = 1101$ are the two elements of the schema $H = *101$. The order of a schema, $O(H)$, is defined as the number of fixed positions in H, and the defining length, $\delta(H)$, as the distance between the first and the last fixed positions of H. A fundamental theorem about schemata is the following :

Theorem 1 (Schema theorem, Holland ([14]))

For a given schema H let :

- m(H,t) be the expected number of representatives of the schema H in the population $P(t)$ (t indexes the number of the generation) : $m(H,t) = |H \cap P(t)|$.

- $\tilde{f}(H,t)$ be the mean fitness value of the representatives of H in the population $P(t)$:
$$\tilde{f}(H,t) = \frac{1}{|H \cap P(t)|} \sum_{i \in H \cap P(t)} f(i)$$
- $\bar{f}(t)$ be the mean fitness value of the individuals of $P(t)$:
$$\bar{f}(t) = \frac{1}{|P(t)|} \sum_{i \in P(t)} f(i)$$
- p_c and p_m be respectively the (one point) crossover and mutation probabilities.

Then :

$$m(H,t+1) \geq m(H,t) * \frac{\tilde{f}(H,t)}{\bar{f}(t)} * \left[1 - p_c \frac{\delta(H)}{l-1} - O(H) p_m\right]$$

B. Deception analysis.

A famous consequence of the schema theorem is that the schemata having a short defining length, a small order and a mean fitness better than the population mean fitness will be more and more represented in the successive generations (such schemata are called *building blocks*, [10]). This remark leads to the conclusion that if the global optimum of the fitness function f is the intersection of such good building blocks, a GA will easily find it. On the contrary, if the intersection of these buildings blocks is a secondary optimum, the population will preferably converge onto it, missing the global one. In this situation the GA will be considered to have failed[1] and f will be called *deceptive.*

More formally, Goldberg ([8], [9]) defined *static deception*: the *selection* results in an expected greater mean fitness for the set of individuals selected for reproduction, than for the preceding population. But this mean value will be changed by the application of genetic operators. It follows that the GA is attracted toward the optima of a function f', defined for each point of Ω^l as its expected fitness value after the application of crossover and mutation. The function f will be called deceptive for a GA with a given parameter setting, if the global optima of f' and f differ. This function may be calculated with the help of the Walsh basis :

Definition 1 (Walsh polynomials) They form an orthogonal basis of the set of functions defined on Ω^l :

$$\psi_j(x) = \prod_{t=0}^{l-1} (-1)^{x_t j_t} = (-1)^{\sum_{t=0}^{l-1} x_t j_t} \qquad (1)$$

Where x_t and j_t denote the the values of the t^{th} bit of the binary decomposition of x and j.

The projection of a function f on this basis is :

$$f(x) = \sum_{j=0}^{2^l-1} w_j \psi_j(x) \quad \text{with} \quad w_j = \frac{1}{2^l} \sum_{x=0}^{2^l-1} f(x) \psi_j(x)$$

The coefficients w_j are called *Walsh coefficients* and are strongly related to schemata. Roughly, a given w_j is related to schemata having fixed bits at the position where j has '1' in its binary decomposition. Consequently, the *adjusted Walsh coefficients* ("adjusted" according to genetic operators) may be calculated :

$$w'_j = w_j (1 - p_c \frac{\delta(j)}{l-1} - 2 p_m O(j)) \qquad (2)$$

Where $O(j)$ denotes the number of '1' in the binary decomposition and $\delta(j)$ the distance between the first and the last '1'.

$$f'(x) = \sum_{j=0}^{2^l-1} w'_j \psi_j(x) \qquad (3)$$

Defining the two following sets (near optimal sets of f and f') for a given ϵ :

$$N_\epsilon = \{x \in [0..2^l] \ / \ |f(x) - f^*| \leq \epsilon\}$$

and

$$N'_\epsilon = \{x \in [0..2^l] \ / \ |f'(x) - f'^*| \leq \epsilon' = \frac{f'^* - w_0}{f^* - w_0} \epsilon\}$$

the definition of static deception follows([9]) :

Definition 2: A function-coding combination is **statically deceptive** at the level ϵ when $N_\epsilon - N'_\epsilon \neq 0$.

C. Deception analysis on Hölder functions.

The work presented in [18] aims to characterize the deception of a given function f, considered as the binary encoding of the sampling of a Hölder function on the interval $[0,1]$:

Definition 3 (Hölder function of exponent h [7])
Let (X, d_X) and (Y, d_Y) be two metric spaces. A function $F : X \to Y$ is called *Hölder function of exponent* $h > 0$, if for each $x, y \in X$ such that $d_X(x,y) < 1$, we have :

$$d_Y(F(x), F(y)) \leq k . d_X(x,y)^h \qquad (4)$$

for some $k > 0$.

Although a Hölder function is always continuous, it needs not to be differentiable, and if it is Hölder with exponent h, it is Hölder with exponent h' for all $h' \in]0,h]$. Intuitively, we may characterize a Hölder function of low exponent h as more "irregular" than a Hölder function of higher h.

It is possible to consequently establish a relation between h and $|f - f'|$. To reach this point the following basis is used :

Definition 4 (Haar polynomials) They form an orthogonal basis of the set of functions defined on Ω^l :

$$H_{2^q+m}(x) = \begin{cases} 1 & \text{for } (2m)2^{l-q-1} \leq x < (2m+1)2^{l-q-1} \\ -1 & \text{for } (2m+1)2^{l-q-1} \leq x < (2m+2)2^{l-q-1} \\ 0 & \text{otherwise in } \{0, \ldots, 2^l - 1\} \end{cases} \qquad (5)$$

With $q \in [0..(l-1)]$, $m \in [0..(2^q - 1)]$ and $j = 2^q + m \in [0..(2^l - 1)]$. The projection of f on this basis is :

$$f(x) = \sum_{j=0}^{2^l-1} h_j H_j(x) \quad \text{and} \quad h_j = \frac{1}{2^{l-q}} \sum_{x=0}^{2^l-1} f(x) H_j(x) \qquad (6)$$

---

[1] Only in an optimization perspective. Recast in a more general context, the success of a GA may not only be related to its ability to find a global optimum at each trial, but rather to rapidly find good solutions.

As the Haar coefficients may be bounded :

$$\forall j = 2^q + m, \quad |h_j| \leq \frac{k}{2} 2^{-h(q+1)} \tag{7}$$

the authors of [18] have proved the following theorem :

Theorem 2: Let f be the sampling on l bits of a Hölder function of exponent h and constant k, defined on $[0,1]$, and let f' be defined as in (3). Then :

$$\forall x \in \{0, \ldots, 2^l - 1\} \;\; |f(x) - f'(x)| \leq k * B(p_m, p_c, l, h)$$

with

$$\begin{aligned} B(p_m, p_c, l, h) = \quad & \frac{p_c}{l-1} 2^{-h} \left[\frac{2^{-l(h+1)} - 1}{2^{-(h+1)}} \right. \\ & \left. + \frac{(1 - 2^{l-h})(2^{-hl} - 1) - l2^{-hl}(1 - 2^{-h})}{(2^{-h} - 1)^2} \right] \\ & + p_m \frac{2^{-h}}{(2^{-h} - 1)^2} \left[1 + 2^{-hl}(l2^{-h} - l - 1) \right] \end{aligned}$$

We can see that if B grows, f has more and more chances to be deceptive. Without going into details, B is a function of 4 real parameters having the following behavior :

- B decreases with h.
- B increases with p_m, p_c.
- B increases with l, when l is small, reaches a maximum for a given value l_{max}, and decreases for $l > l_{max}$.

III. A BITWISE REGULARITY CHARACTERIZATION.

The previous analysis is based on an irregularity characterization with respect to an underlying distance that is the Euclidian distance on $[0,1]$. This approach is straightforward for fitness functions defined on $\mathbb{R}$, and in the general case it is always possible to consider the fitness function as the sampling of an underlying one-dimensional Hölder function. It is however less evident in this latter case that the Hölder exponent reflects in a simple way the irregularity of the fitness function (it may appear for example more irregular than it is in a multidimensional space). This is the reason why we present in this paper a similar irregularity analysis but with respect to the Hamming distance on the set of binary strings. Another justification is also that the use of Hamming distance is more closely related to the action of genetic operators.

A. Bitwise regularity coefficients.

Consequently to preceding considerations, we introduce the following coefficients, that are derived from Hölder grained exponents with respect to a distance proportional to the Hamming distance (see [17] for a more detailed justification) :

Definition 5 (Bitwise regularity coefficients) Let f be a function defined on Ω^l :

$$\forall q \in \{0, \ldots, l-1\}, \; C_q = \sup_{x \in \Omega^l} \{|f(x) - f(x'_{l-q-1})|\}$$

with x'_{l-q-1} and x differing only with respect to one bit at the position $(l - q - 1)$.[2]

In other terms, the C_q coefficient represents the maximum fitness variation due to a bit flip at the position $(l - q - 1)$. Therefore, we can show that :

$$\forall j = 2^q + m, \quad |h_j| \leq \frac{C_q}{2}$$

In the same way as in [18], with the help of the Haar basis the following theorem has been established (see [17] for a demonstration) :

Theorem 3: Let f be a function defined on Ω^l with bitwise regularity coefficients $(C_q)_{q \in \{0, \ldots, l-1\}}$, and let f' be defined as in (3). Then $\forall x \in \Omega^l$:

$$\begin{aligned} |f(x) - f'(x)| \leq \quad & \frac{p_c}{l-1} * \sum_{q=0}^{l-1} C_q * \left(\frac{1 + 2^q(q-1)}{2^q} \right) \\ & + p_m * \sum_{q=0}^{l-1} C_q * (q+1) \end{aligned}$$

Furthermore, this result still holds when the order of the C_q is reversed, so the final bound is the one minimizing the preceding expression.

We also have to note that the bits do not have the same role in this bound expression. In fact their relative weight is strictly increasing with respect to the index q. Sorting (either in increasing or decreasing order) would then minimize this bound suggesting that the simple change of coding consisting in a permutation on the bits would make the function easier. This feature can be explained by the fact that the one point crossover disrupts more easily a combination of a few genes spread at each extremities of the chromosome than if these genes were grouped at one extremity. Reordering the bits in order to sort the bitwise regularity coefficients is then equivalent to group the most "sensible" genes at one extremity of the chromosome. Some experiments presented in [17] partially support this hypothesis, but also reveal that other phenomenons (as epistasis for example [5]) have to be taken into account in order to predict the sensibility of GA to such encoding changes.

B. Bitwise regularity coefficients compared to Hölder exponent.

If we suppose that the fitness function f is the sampling on l bits of a Hölder function of exponent h and constant k, defined on $[0,1]$, the bound of *theorem 3* is lower than the bound of *theorem 2*.

One can easily show, (see [17]), that :

$$C_q \leq \quad k * 2^{-(q+1)h} \tag{8}$$

as we have :

$$|h_j| \leq \frac{C_q}{2} \;\text{ and }\; |h_j| \leq \frac{k}{2} * 2^{-(q+1)h}$$

and as the bound on $|f - f'|$ is a linear function of the bounds on the $|h_j|$, it follows immediately that the bound of *theorem 3* is the lowest. Moreover, the estimation of the bitwise regularity coefficients is computationnaly cheaper than the estimation of the Hölder exponent and its associated constant k.

---

[2] The less significant bit being at position 0.

IV. DECEPTION ANALYSIS OF A GA WITH UNIFORM CROSSOVER.

As we have seen, the bound on $|f - f'|$ derived from the bitwise regularity coefficients C_q depends on their relative order, due to the use of the one point crossover. The aim of this section is to establish analogous results for the uniform crossover ([21], for which the positional bias no longer exists).

First, we establish the expression the Walsh adjusted coefficients of a GA with uniform crossover. The only change is to replace the schema disruption probability p_d for this version of crossover :

$$p_d \leq \left(1 - \left(\frac{1}{2}\right)^{O(H)-1}\right) \tag{9}$$

This upper bound is obtained by observing that once the first fixed bit of the schema is allocated to one of the offsprings, it will always survive if all other fixed bits are allocated to the same offspring. As for the one point crossover p_d is conservatively set to $\left(\frac{\delta(h)}{l-1}\right)$, if we set p_d to the upper bound (9), then the new adjusted Walsh coefficients are :

$$w'_j = w_j \left[1 - p_c \left(1 - \left(\frac{1}{2}\right)^{O(j)-1}\right) - 2p_m O(j)\right]$$

Notice that $O(j)$ no longer depends on the defining length of the schema. Furthermore as the order of a schema is invariant with respect to a permutation on the bits, the following theorem has been proven (see [17] for a demonstration) :

Theorem 4: Let f be a function defined on Ω^l with bitwise regularity coefficients $(C_q)_{q \in \{0,\ldots,l-1\}}$, and let f' be defined as in (3). Then for all permutation σ defined on the set $\{0, \ldots, l-1\}$, $\forall x \in \Omega^l$:

$$|f(x) - f'(x)| \leq p_c * \sum_{q=0}^{l-1} C_{\sigma^{-1}(q)} + p_m * \sum_{q=0}^{l-1} C_{\sigma^{-1}(q)} * (q+1) \tag{10}$$

We immediately see that this upper bound is minimal when the $C_{\sigma^{-1}(q)}$ are ordered in decreasing order.

Practically, if it is possible to get the C_q values (or good estimations), it is hard to draw conclusions from the value of the bound (10). But if we consider the effect of an encoding change on it, it is interesting to see if its variation is experimentally correlated to the performances of the GA. Intuitively, the hypothesis is formulated as follows : if an encoding change (such as a Gray code) induces a decrease of the bound (10), the GA should perform better with this new encoding, and conversely. We present experiments with the Gray code in the next section.

V. EXPERIMENTS.

All the experiments presented here are based on a simple GA with uniform crossover. Several functions were tested with different encodings. The specified parameters are :

- l : number of bits per chromosome.
- N : population size.
- *Gen* : number of generations for a run.
- *Runs* : number of runs.
- p_c : (uniform) crossover probability.
- p_m : mutation probability.

l	N	*Gen*	*Runs*	P_c	p_m
30	128	100	100	1.0	0.005

Fct. & Code	*Bound*	*Avg.*	*Max*	*Ratio*
*F*1 & *Code*1	**131**	**75.59**	**78.64**	**0.04**
*F*1 & *Code*2	210.3	74.01	78.62	0
*F*1 & *Code*3	235.2	73.42	78.47	0
*F*1 & *Code*4	313.6	70.58	78.11	0

TABLE I
PARAMETERS AND PERFORMANCES FOR F1.

For each function-encoding combination and for each generation, the average of the population mean fitness value, the average of the best individual fitness value, and the ratio of populations containing a global optima are recorded. These performances at the last generation are dispalyed in tables I to VII.

The tested encoding change is the Gray code, that is :

$$K : \Omega^l \to \Omega^l \quad , \quad K(x) = g \;\text{ with}$$

$$g_i = \begin{cases} x_{(l-1)} & \text{if } i = (l-1) \\ (x_{(i+1)} \; XOR \; x_i) & \text{if } (l-1) < i \leq 0 \end{cases}$$

A. Function f_1.

This function is one of the *De Jong Five-Functions Test Bed* [15], turned into a maximization problem : $F1(X) = (max(f_1) - f_1(X))$, with :

$$f_1(X) = \sum_{i=1}^{3} \left(X^{(i)}\right)^2 \quad \text{with} \quad -5.12 \leq X^{(i)} \leq 5.12$$

This function is 3-dimensional and each component is defined on 10 bits. Four different mappings from Ω^{10} to $[-5.12, 5.12]$ have been experimented. Let x be any of the $X^{(i)}$, coded on 10 bits, $b_9 b_8 \ldots b_1 b_0$:

- *Code*1 : a classical signed integer binary encoding, mapped to $[-5.12, 5.12]$: $x = \frac{1}{100}(-1)^{b_9} * \sum_{j=0}^{8} 2^j * b_j$
- *Code*2 : an unsigned binary integer encoding, mapped to $[-5.12, 5.12]$: $x = \frac{1}{100}\left(\sum_{j=0}^{9} 2^j * b_j - 512\right)$
- *Code*3 : same as *Code*1 but with Gray encoding for $b_8 b_7 \ldots b_1 b_0$.
- *Code*4 : same as *Code*2 but with Gray encoding for $b_9 b_8 \ldots b_1 b_0$.

In table I, we see that the bound is increasing with each new encoding, and that the performances of the GA decrease as predicted, though it is measured only with *Avg.* and *Max*. The runs that found the global optimum were very rare, since a lot of solutions have a fitness value very close to the optimum (due the absence of scaling, the GA is unable to distinguish them).

l	N	Gen	$Runs$	P_c	p_m
24	512	50	100	1.0	0.005

$Fct.$ & $Code$	$Bound$	$Avg.$	Max	$Ratio$
$F2$ & $Code1$	**13350**	**3880**	3910	0.17
$F2$ & $Code2$	15610	3829	3910	0.1
$F2$ & $Code3$	16570	3832	3910	0.19
$F2$ & $Code4$	19070	3741	3910	0.28

TABLE II

Parameters and performances for F2.

l	N	Gen	$Runs$	P_c	p_m
16	200	40	100	1.0	0.005

$Fct.$ & $Code$	$Bound$	$Avg.$	Max	$Ratio$
$M2$ & $Code1$	**4.86**	**0.9542**	**1**	**1**
$M2$ & $Code2$	5.76	0.9056	0.952	0.42

TABLE III

Parameters and performances for M2.

B. Function f_2.

This function is also one of the *De Jong Five-Functions Test Bed*, turned into a maximization problem : $F2(X) = (max(f_2) - f_2(X))$, with :

$$f_2(X) = 100 * ((X^{(1)})^2 - X^{(2)})^2 + (1 - X^{(1)})^2$$

with $X^{(i)} \in [-2.048, 2.048]$.

It is a function of a 2-dimensional parameter whose components are coded on 12 bits. The same 4 encodings as for $F1$ are tested.

Once again, a lot of points have fitness values very close to the optimum, so the ratio of populations containing it is more or less random (see table II). In fact, it would require far more than 4 digits to distinguish the Max performances.

The Avg performances follow, in order, the predictions of the bound, except for the comparison between $Code2$ and $Code3$, for which the bound increase is the lowest and the performances are roughly identical.

C. Function M2.

This is a sampling of a 1-dimensional function from [12] :

$$\forall x \in [0,1], \quad M2(x) = e^{-2(ln2)(\frac{x-0.1}{0.8})^2} sin^6(5\pi x)$$

Two encodings are tested :

- $Code1$: a classical unsigned integer encoding mapped to [0,1]
- $Code2$: Gray version of $Code1$.

Here the Gray encoding induces an increase of the bound and a decrease of the performances as predicted (see table III).

l	N	Gen	$Runs$	P_c	p_m
18	512	300	100	1.0	0.005

$Fct.$ & $Code$	$Bound$	$Avg.$	Max	$Ratio$
$M7$ & $Code1$	18.9	1.605	1.972	0
$M7$ & $Code2$	**13.8**	**2.164**	**2.458**	**0.16**

TABLE IV

Parameters and performances for M7.

l	N	Gen	$Runs$	P_c	p_m
24	512	100	100	1.0	0.005

$Fct.$ & $Code$	$Bound$	$Avg.$	Max	$Ratio$
$M7$ & $Code1$	**43.8**	**21.35**	24	1
$M7$ & $Code2$	57.1	20.64	24	1

TABLE V

Parameters and performances for EPI6.

D. Function M7.

This function extracted from [4], is massively multimodal and deceptive. It is composed of sub-functions defined on 6 bits, which reach their maximum value for two mirror strings. Here we used 3 of them ($l = 18$).

Two encodings are tested :

- $Code1$: the classical encoding.
- $Code2$: Gray version of $Code1$.

Here the Gray encoding induces a decrease of the bound and an increase of the performances, as predicted (see table IV).

E. Function EPI6.

This function is the concatenation of 6 epistatic subfunctions EPI defined on 4 bits :

$$EPI(b_3 b_2 b_1 b_0) = \begin{cases} \sum_{i=0}^{2}(1 - b_i) & \text{if } b_3 = 0 \\ 1 + \sum_{i=0}^{2} b_i & \text{if } b_3 = 1 \end{cases}$$

Two encodings are tested :

- $Code1$: the classical encoding.
- $Code2$: Gray version of $Code1$.

Here the Gray encoding induces an increase of the bound and a decrease of the performances, as predicted (see V).

F. Function W1.8.

This fitness function is a sampling of a Weierstrass function ([7]) of dimension 1.8 (i.e. the Hölder exponent is 0.2). The encodings are the same as for $M2$.

Here the increase of the bound is relatively small and only the $Ratio$ performance decreases as predicted (see table VI).

l	N	Gen	$Runs$	P_c	p_m
16	512	100	100	1.0	0.005

$Fct.$ & $Code$	$Bound$	$Avg.$	Max	$Ratio$
$W1.8$ & $Code1$	**21.6**	2.29	2.699	**0.95**
$W1.8$ & $Code2$	22.65	**2.347**	**2.701**	0.74

TABLE VI

Parameters and performances for W1.8.

l	N	Gen	$Runs$	P_c	p_m
12	64	50	500	1.0	0.005

$Fct.$ & $Code$	$Bound$	$Avg.$	Max	$Ratio$
$FBM1.5$ & $Code1$	4.658	**1.009**	1.087	0.344
$FBM1.5$ & $Code2$	**4.609**	1.006	**1.088**	**0.36**

TABLE VII

Parameters and performances for FBM1.5.

G. Function FBM1.5.

This fitness function is a sampling of Fractional Brownian Motion ([7]) of dimension 1.5 (Hölder exponent 0.5). The encodings are the same as for $M2$.

The decrease of the bound is very small, compared to previous tests, and except for the *Avg.*, the performances slightly increase as predicted (see table VII).

VI. Conclusions.

The previous tests show that the bound calculated from the bitwise regularity coefficients is a quite reliable tool to compare encodings as long as its variations are significant enough : when the bound variations are high, the GA behaves according to the predictions, when they are low (as for $Code2$ and $Code3$ of functions $F2$, for functions W1.8 and FBM1.5) the GA behavior is less predictable.

These limitations can be explained in many ways. The one that seems to us the most appropriate is of the same nature as the *Static Building Blocks Hypothesis*, pointed out in [13]. If we consider cautiously the calculation of f', which is the basis of the *static deceptivity* analysis, we note that it is assumed that each allele is equally represented at each position. This point of view, even if not largely exposed here (see [17] for details), should be considered with care in order to continue the work presented here and suggests that a dynamical modelization of the GA behavior would be more appropriate. The nonuniform Walsh-schema transform [2] could be the basis of such an improvement.

References

[1] A. Agapie. Genetic algorithms : Minimal conditions for convergence. *Artificial Evolution, Nimes, France, October 1997,*. Springer Verlag, 1997.

[2] C.L. Bridges and D.E. Goldberg. The nonuniform walsh-schema transform. In G.J.E. Rawlins, editor, *Foundations of Genetic Algorithms.*, pages 13–22. Morgan Kaufmannn publishers, San Mateo, 1991.

[3] R. Cerf. Asymptotic convergence of genetic algorithms, *Artificial Evolution, European Conference, AE 95, Brest, France, September 1995, Selected papers*, volume Lecture Notes in Computer Science 1063, pp 37–54. Springer Verlag, 1995.

[4] K. Deb, D.E. Goldberg and J. Horn. Massive multimodality, deception and genetic algoritms. In B. Manderick R. Männer, editor, *PPSN-II*, pp 37–46. Amsterdam : North-Holland, 1992.

[5] Y. Davidor. Epistasis variance : A viewpoint on ga-hardness. In G.J.E. Rawlins, editor, *Foundations of Genetic Algorithms.*, pp 23–35. Morgan Kaufmannn publishers, San Mateo, 1991.

[6] T.E. Davis and J.C. Principe. A Simulated Annealing Like Convergence Theory for the Simple Genetic Algorithm . In *Proceedings of the Fourth ICGA*, pp 174–182, 1991. 13-16 July.

[7] K.J. Falconer. *Fractal Geometry : Mathematical Foundation and Applications.* John Wiley & Sons, 1990.

[8] D.E. Goldberg. Genetic algorithms and walsh functions: Part i, a gentle introduction. In *Complex Systems*, volume 3, pp 123–152. 1989.

[9] D.E. Goldberg. Genetic algorithms and walsh functions: Part ii, deception and its analysis. In *Complex Systems*, volume 3, pages 153–171. 1989.

[10] D.E. Goldberg. *Genetic Algorithms in Search, Optimization, and Machine Learning.* Addison-Wesley, 1989.

[11] D.E. Goldberg. Construction of high–order deceptive functions using low–order walsh coefficients. IlliGAL Report 90002, University of Illinois at Urbana–Champaign, 1990.

[12] D.E. Goldberg and J. Richardson. Genetic algorithms with sharing for multimodal function optimization. In J.J. Grefenstette, editor, *Genetic Algorithms and their Applications*, pages 41–49. Lawrence Erlbaum Associates, 1987.

[13] J.J. Grefenstette. Deception considered harmful. In L.D. Whitley, editor, *Foundations of Genetic Algorithms 2.*, pp 75–91. Morgan Kaufmannn publishers, San Mateo, 1992.

[14] J. Holland. *Adaptation in Natural and Artificial Systems.* Ann Arbour, University of Michigan Press, 1975.

[15] K.A. De Jong. *An analysis of the behavior of a class of genetic adaptative systems.* PhD thesis, University of Michigan, 1975.

[16] J. Juliany and M. D. Vose. The genetic algorithm fractal. *Evolutionary Computation*, 2(2):165–180, 1994.

[17] B. Leblanc and E. Lutton. Bitwise regularity coefficients as a tool for deception analysis. Technical report, INRIA, France, to appear.

[18] E. Lutton and J. Lévy-Véhel. Some remarks on the optimization of hölder functions with genetic algorithms. Technical Report 2627, INRIA, France, 1995.

[19] S. Rochet, G.Venturini, M. Slimane, and E.M. El Kharoubi. A critical and empirical study of epistasis measures for predicting GA performances : a summary. *Artificial Evolution, Nimes, France, October 1997, Selected papers*, Springer Verlag, 1997.

[20] G. Rudolph. Asymptotical convergence rates of simple evolutionary algorithms under factorizing mutation distributions. *Artificial Evolution, Nimes, France, October 1997,*. Springer Verlag, 1997.

[21] G. Syswerda. Uniform crossover in genetic algorithms. In J.D. Schaffer, editor, *Proceedings of the Third ICGA.* Morgan Kaufmannn publishers, 1989.

[22] M.D. Vose. Formalizing genetic algorithms. In *Genetic Algorithms, Neural Networks and Simulated Annealing Applied to Problems in Signal and Image processing.* May 1990. University of Glasgow.