Fluctuating Crosstalk as a Source of Deterministic Noise and its Effects on GA Scalability

Kumara Sastry
Paul Winward
David E. Goldberg
Cláudio Lima

IlliGAL Report No. 2005025
November, 2005

Illinois Genetic Algorithms Laboratory
University of Illinois at Urbana-Champaign
117 Transportation Building
104 S. Mathews Avenue Urbana, IL 61801
Office: (217) 333-2346
Fax: (217) 244-5705

Fluctuating Crosstalk as a Source of Deterministic Noise and its Effects on GA Scalability

Kumara Sastry[1], Paul Winward[1], David E. Goldberg[1], Cláudio Lima[2]

[1]Illinois Genetic Algorithms Laboratory,
Department of General Engineering,
University of Illinois at Urbana-Champaign

[2]DEEI-FCT,
University of Algarve,
Campus de Gambelas,
8000-117 Faro, Portugal

{ksastry,winward,deg}@uiuc.edu, clima@ualg.pt

Abstract

This paper explores how fluctuating crosstalk in a deterministic fitness function introduces noise into genetic algorithms. We model fluctuating crosstalk or nonlinear interactions among building blocks via higher-order Walsh coefficients. The fluctuating crosstalk behaves like exogenous noise and can be handled by increasing the population size and run duration. This behavior holds until the strength of the crosstalk far exceeds the underlying fitness variance by a certain factor empirically observed. Our results also have implications for the relative performance of building-block-wise mutation over crossover.

1 Introduction

Many diverse operators and variants of the genetic algorithms (GAs) have been studied over the years as GA practitioners and theorists strive to develop what has been termed elsewhere as *competent GAs* — GAs that solve a boundedly hard class of problems quickly, reliably, and accurately (Goldberg, 2002). One area of study that has received considerable attention is the relative performance of mutation to crossover - two key operations proposed from GA beginnings.

Recently, Sastry and Goldberg (2004) presented an unbiased comparison between the computational costs associated with crossover in a selectorecombinative GA to that of mutation. In that study, the mutation algorithm exploits its linkage knowledge to greedily change one building block (BB) at a time. In deterministic problems, the mutation algorithm outperforms the GA. However, in the presence of constant exogenous Gaussian noise, the situation flips as the selectorecombinative GA comes out on top.

One question that arose from that study was how these operators would fare in the presence of crosstalk, or what is sometimes referred to as epistasis. Goldberg (2002) conjectures that one type of crosstalk, fluctuating crosstalk, can induce similar effects to explicit Gaussian noise, although still deterministic. He explains how a GA can converge to the global optimum if the same relaxations are applied as if external noise was present: supply a larger population of individuals and allow more time for population convergence. Furthermore, the needed population and convergence time follow facetwise models derived for fitness functions with additive Gaussian noise when the crosstalk signal falls below a certain critical point. The purpose of this paper is to construct a bounding test function that demonstrates this effect when the fluctuation noise is varied from the nonexistent to the very high, and understand this in the light of recent theoretical decomposition of problem difficulty as pointed out elsewhere (Goldberg, 2002).

This paper is organized as follows. We first present a brief background to crosstalk in the context of a three-way decomposition to problem difficulty. In section 3, we describe how to represent fluctuating crosstalk through Walsh transformations and explain how many deterministic functions possess some form of fluctuating crosstalk. The next section compares our results to known models of population size, convergence time, and function evaluations. A threshold is observed for when fluctuating crosstalk deviates from these models based on the expected population size when crosstalk acts as a signal rather than noise. The paper concludes with a connection of our results to mutation and crossover performance and briefly addresses future work.

2 Crosstalk: A Facet of Problem Difficulty

Although GAs process populations of individual bit strings, it is more useful for analytical purposes to think of GAs as processing minimal, sequentially superior BBs of a solution (Goldberg, 2002). BB-wise problem difficulty can stem from three sources: intra-BB difficulty (deception), inter-BB difficulty (scaling), and extra-BB difficulty (noise). A more complete discussion of these sources can be found elsewhere (Goldberg, 2002). As discussed there, we briefly review an indirect source of difficulty termed crosstalk, or what might be called *inter-BB epistasis*.

Epistasis in GA literature refers to the nonlinear interactions among genes or sets of genes, and is widely known to be a contributing factor of GA-hardness. Historically, there have been three primary approaches to tackling epistasis (Kumar, 2002). The first approach relies on a priori measurements of problem difficulty which can either be direct or indirect measurements of epistasis. Davidor (Davidor, 1991) first suggested that the amount of epistasis determines problem hardness for a GA. However, Naudts (Naudts & Kallel, 1998) observes that it is the distribution and structure of the epistasis that contributes to problem difficulty, and not only the amount of it. He shows these a priori measures are sensitive to non-linear scalings and require further theoretical development for robust and accurate measures. He introduces the concept of a site-wise optimization measure but careful probing reveals that it fails to correctly identify difficulty in several cases. A related area of these measurements is epistasis approximation through random sampling (Davidor, 1991) but Naudts (Naudts & Kallel, 1998) finds random sampling a poor solution for approximating values of such measures. Heckendorn (Heckendorn & Whitley, 1999) derives local bitwise epistasis measures that scale well to overall difficulty for certain polynomial functions. He proposes techniques of argument centering and parity truncation to reduce epistasis for this class of functions. The second common approach is to use variants of the Factorized Distribution Algorithm (FDA) — a probabilistic model building GA (PMBGA) with multivariate interactions. Although such techniques directly confront epistasis, the FDA only guarantees optimal solutions for particular gene interactions and requires prior information generally not available. The third approach calls for a change in problem representation but direct approaches on the problem require a hefty amount of ingenuity and manipulation. Indirect representational approaches like the popular Gray Code "shifting" technique show practical promise but the clarifying of the relationship between the need for shifting and epistasis remains to be done.

In this paper, we validate a fourth approach first proposed by Goldberg (Goldberg, 2002). In discussing problem difficulty, Goldberg identified the three forms of crosstalk as a mapping to the three primary sources of problem difficulty. Hence, a simple but effective way to optimize deterministic functions in the presence of crosstalk is to apply known techniques of handling deception, scaling, and exogenous noise. It becomes readily apparent of the mapping to deception and scaling, and these cases are well addressed in the literature. In the remainder of this section we explain those mappings, and especially that between fluctuating crosstalk and exogenous noise, with some examples.

Consider the additively-separable objective fitness function

$$f(x) = f_1(x_1x_2x_3x_4) + f_2(x_5x_6x_7) + f_3(x_8x_9).$$

The optimal solution to $f(x)$ will be the concatenation of the target bits of the three optimal BBs. In this example, each BB is independent of one another since the constituent bits of any one BB are not found in any other BB. Building block independence makes the solution search much easier since the marginal fitness boost from a particular discovered BB remains constant throughout the run. Contrast this condition to one in which we introduce *reinforcing crosstalk* by adding a fourth term that gives a positive value only when the second and third BBs have reached their target values:

$$f(x) = f_1(x_1x_2x_3x_4) + f_2(x_5x_6x_7) + f_3(x_8x_9) + f_4(x_5x_6x_7x_8x_9).$$

From this equation we can see how crosstalk or epistasis refers to the nonlinear interaction among BBs. Initially, when the proportion of individuals with correct target bits in positions 5 to 9 are small, the GA processes BBs in the same manner as if crosstalk didn't exist. As soon as those bits are found, an added fitness is assigned to the individual, thereby causing those individuals with the correct second and third BBs to grow exponentially over time. The relative contribution to fitness for other individuals is devalued compared to the bonus supplied by reinforcing crosstalk. Hence, reinforcing crosstalk is essentially a problem of scaling, and a competent GA that can handle scaling can handle reinforcing crosstalk.

Consider what happens under *punishing crosstalk* when a negative marginal fitness is assigned when the second and third BB target bits are found. This leads to a scaling of BB values, as the other BBs will converge more quickly than those being punished. The preceding statement holds as long as the punishing signal is within some range relative to the variance of the nonpunished BBs. If the punishing signal is excessive, however, what may have been initially thought to be a good solution is no longer optimal. This is a case of deception, and requires further processing to refine the BBs. Hence, a competent GA that can handle the core difficulty of deception can handle excessive punishing crosstalk.

With this background, we are prepared to consider the heart of what this paper considers — *fluctuating crosstalk*. Suppose that instead of always punishing or rewarding the individual once the correct target bits are found, we add or subtract some positive weight w to the fitness function based on the parity over the target bits; $+w$ for even parity and $-w$ for odd parity. A natural question to consider is how drastic shifts in fitness because of a single bit change can be overcome during the course of BB decision making. Fortunately, two things act in our favor. First, the initial populations are assumed sufficiently randomized such that the target bits over BBs participating in the crosstalk are random. Hence, the net effect on fitness due to fluctuating crosstalk is zero since there are equal numbers of even-parity individuals as there are odd-parity individuals over those target bits. The second factor in our favor is the fact that ultimately, the GA will converge even with a possible selection stall. Thus, towards the end of the run fluctuation crosstalk behaves as either reinforcing crosstalk or punishing crosstalk and a GA that can handle those can effectively handle fluctuating crosstalk.

3 Principled Modeling of Fluctuating Crosstalk through Walsh Coefficients

Having explained fluctuating crosstalk in the context of problem difficulty, we must now address how we model it in a principled way to account for any conceivable fluctuating crosstalk subfunction. This will be accomplished by representing the fitness function in another basis, the *partial-parity* or *Walsh basis*. The Walsh basis is significant to GAs because it allows for bounding measures of deceptiveness and for faster schema-average fitness calculation. And in our case, it allows us to represent our fluctuating crosstalk example f_4 as a partial, signed sum of the Walsh coefficients. We begin with some notation and definitions.

For our discussion, we take the intuitive approach introduced by Goldberg (1989). Let $\mathbf{x} = x_l x_{l-1} \ldots x_2 x_1$ be the l-bit string representing the coding of the decision variables for an arbitrary individual. We now introduce the auxiliary string positions y_i that are mapped to from the bitwise string positions x_i for $i = 1, \ldots, l$ by:

$$y_i = \begin{cases} 1, & \text{if } x_i = 0 \\ -1, & \text{if } x_i = 1. \end{cases}$$

This definition allows the following multiplication to act as an exclusive or operator (XOR). The j^{th} Walsh function $\psi_j(y)$ where $0 \leq j \leq 2^l - 1$ is calculated as:

$$\psi_j(y) = \prod_{i=1}^{l} y_i^{j_i}, \; y_i \in \{-1, 1\}$$

where j_i is the i^{th} bit of the binary representation of j and y is understood to be the transformed x using the auxiliary mapping.

For example, for $j = 19 = 10011_2$, ψ_{19}(1 -1 1 -1 1)$= y_1 y_2 y_5 = (1)(-1)(1) = -1$. Said another way, $\psi_j(x)$ is a partial-parity function that returns a -1 or a $+1$ as the number of ones in the argument that match the corresponding bits in j is odd or even. Hence, $\psi_{31}(01101) = \psi_{31}(00001) = -1$ since 31 in base 2 is 11111 and the parity of every bit in the argument is considered.

In the canonical basis, fitness values are obtained by referencing a table of bitstrings and their objective fitness values. Bethke (1981) showed that any fitness function[1] $f(x)$ over a finite domain can be rewritten as a partial signed sum of the Walsh coefficients, given by

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

For an arbitrary fitness function, we can imagine that some portion of the partial signed sum of the Walsh coefficients has direction - meaning that this smaller sum is acyclic in the limit and represents what the GA seeks to solve. The remaining portion fluctuates with possibly irregular periods and shapes - but not contributing to the overall direction of the function[2]. This latter portion will likely involve higher-order Walsh coefficients since a higher order indicates that more bits interact with another (since more bits of the index are set). Consider the inclusion of the

[1] f maps the l-bit strings into the reals: $f : \{0,1\}^l \rightarrow R$. We call any non-negative figure of merit a fitness function and although this notation may differ from the biological use of the term, it is standard GA practice.

[2] We constrain ourselves to an intuitive discussion of the foregoing material and avoid the rigid formalities of continuity, limits, and functional analysis for the sake of understanding the identification of the fluctuating crosstalk portion in a fitness function.

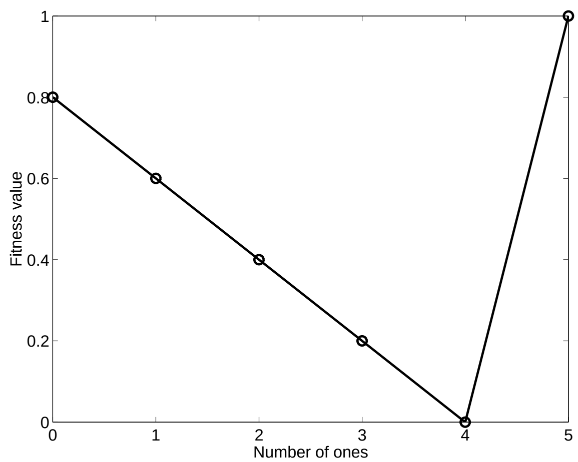

Figure 1: A 5-bit trap function

Walsh coefficient w_{2^l-1} for example. This means taking the parity of the entire string. It provides no direction but merely acts as a source of deterministic noise. Other coefficients could be included as part of this fluctuating crosstalk but for this paper we assume only a full parity.

4 Effect of Crosstalk on GA Scalability

Exogenous noise[3] is a problem for GAs because it interferes with the decision making process as a GA tries to identify the dividing lines between building blocks and infer the superiority of one partition member over another. In this section, we compare exogenous noise effects to those wrought by deterministic noise and empirically validate model adherence. We present facetwise models of population, convergence time, and function evaluations when exogenous noise is present and examine at what point fluctuating crosstalk diverges from these models. However, we first need to introduce the test problem that was common to the three experiments.

4.1 Test Problem

To test the effects of deterministic noise on population size, convergence time, and function evaluations, our test problem consists of an on-average deceptive, concatenated trap function. The use of a trap function assumes a knowledge of linkage information but this well suits our situation since we want to focus on the effects of fluctuating crosstalk and assume that other factors are favorably constant. Such a function allows for efficient exchanging of BBs since each BB is independent of each other. The concatenated trap also allows testing of bounded difficulty since both the order of the trap and number of traps play a role in determining problem difficulty. Each BB consists of a 5-bit trap function with a max fitness of 1 at 11111 and other fitness values given by figure 1.

The Walsh coefficient w_{2^l-1} is then added to or subtracted from this temporary fitness based on the full parity of the string to obtain the final fitness. Modeling epistasis in this manner translates to the highest bitwise interaction possible where every bit interacts with every other bit. Note that the inclusion of an even number of BBs (m, ranging from 4 to 50) preserved optimal solutions under

[3]In this paper, we treat exogenous noise as coming from outside the problem as a single packet, and is viewed as zero-mean, Gaussian noise (Goldberg, 2002) as a close approximation to the true distribution of the noise.

full parity although individual BBs were penalized. We also chose uniform-BB crossover to avoid disrupting discovered BBs. The GA utilized binary tournament selection ($s = 2$) and no mutation.

For each of the following plots, data was obtained by performing 30 *bisection runs* of 50 independent GA trials. In a single bisection run, the population size is adjusted after each set of 50 trials until the minimum population size is obtained that yields an average of $m - 1$ correctly discovered BBs. The interested reader may refer to (Sastry, 2001) for population size adjusting details. We then report the average of the averages over all bisection runs.

4.2 Population Size

One simple but effective way to overcome the negative effects of exogenous noise is is to supply a larger population. Increasing the population size sufficiently ensures that in the initial generations the target BBs are present and discovered by the averaging out of the noise. It also assists in decision making over time so the GA might reliably choose the best partition member. Practical population sizing bounds on selectorecombinative GAs using generation-wise modeling were given by Goldberg, Deb, and Clark (1992). Those early bounds provided a guarantee of good solution quality in a selectorecombinative GA with a sufficiently large population and have shown to well approximate solution quality in the more recent Bayesian Optimization Algorithm (BOA) (Pelikan, Goldberg, & Cantú-Paz, 2000). They were known to be somewhat conservative for the typical selectorecombinative GA, however, and tighter bounds provided by Harik, Cantú-Paz, Goldberg, and Miller (1999) are derived from the gambler's ruin problem which considers the accumulated correctness of deciding between partition members. This model, which also accounts for exogenous noise, is given by:

$$n = -\frac{\sqrt{\pi}}{2d} 2^k \log(\alpha) \sqrt{\sigma_f^2 + \sigma_N^2} \tag{1}$$

where d is the *signal* or difference in fitness between the best and second best individuals, k is the BB size, α is the error tolerance, σ_f^2 is the fitness function variance, and σ_N^2 is the noise variance.

Without noise, this reduces to

$$n_0 = -\frac{\sqrt{\pi}}{2d} 2^k \log(\alpha) \sigma_f \tag{2}$$

and with a little arithmetic we see that equation 1 can be cast in terms of n_0 by

$$n = n_0 \sqrt{1 + \frac{\sigma_N^2}{\sigma_f^2}}. \tag{3}$$

In our case, $\sigma_N^2 = w_{2^l-1}^2$ and $\sigma_f^2 = m\sigma_{BB}^2$ where σ_{BB}^2 is the 5-bit trap variance. This leads to

$$n = n_0 \sqrt{1 + \frac{w_{2^l-1}^2}{m\sigma_{BB}^2}}. \tag{4}$$

Dividing both sides by n_0 reveals an important population size ratio:

$$n_r = \frac{n}{n_0} = \sqrt{1 + \frac{w_{2^l-1}^2}{m\sigma_{BB}^2}}. \tag{5}$$

The above model was specifically derived with exogenous noise in mind but what effect on required population size does fluctuating crosstalk have? When the crosstalk signal is low, we

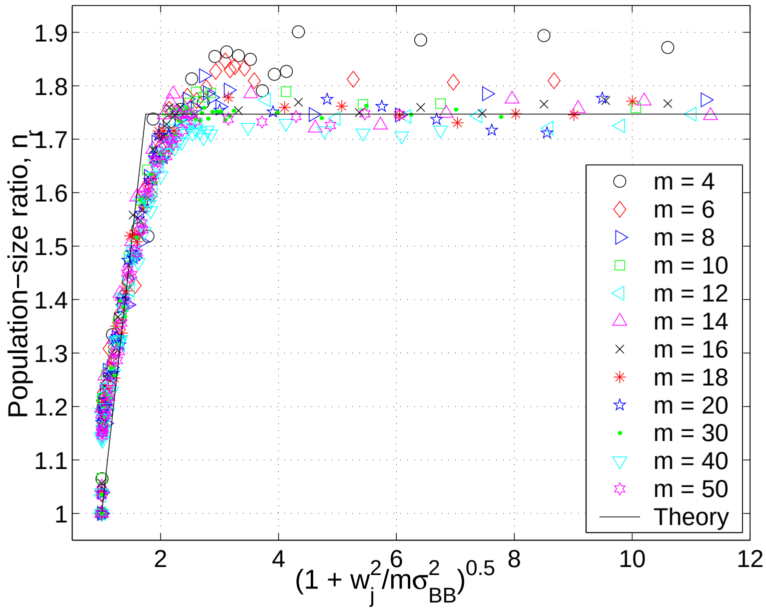

Figure 2: Population size requirements for optimal convergence when fluctuating crosstalk is present ($w_j = w_{2^l-1}$) and problem size m varies from 4 to 50 BBs. Initially, the effects of the crosstalk follow the model (rising bolded line) of exogenous noise closely but then level off when the problem becomes parity-dominated. The flat bolded line is the average required population for all problem sizes tested.

would expect the GA to first solve BBs for the fitness function, since solving these will yield higher marginal contributions than those BBs of the crosstalk. The crosstalk merely interferes with the decision making, acting as a source of deterministic noise. As the crosstalk signal increases, the problem shifts to a parity-dominated one whereby individuals with even parity are first discovered and then varied to find the secondary benefits of solving the trap function. In this sense the directional function is actually perturbing the crosstalk. Once on the parity-dominated side, the required population will level off since the population is large enough for the GA to find the optimal or near-optimal solution and supplying a larger crosstalk signal only creates an effective reduction in the marginal value of solving BBs from the new perturbation source.

Figure 2 illustrates these principles clearly. Before the parity-dominated point but for a given crosstalk signal, required population size grows as problem size m grows. More building blocks induce higher collateral noise since more BBs incur larger numbers of schemata to consider, and the resulting changes to a single BB due to crossover are obfuscated by the simultaneous variations among other crossover-induced BB changes. What is not as intuitive though is why the population size tends to level off where it does.

This can be explained by what we term the *parity-induced filtering effect*. We begin by considering the behavior of the GA as it nears the situation described by the middle of the curve. Here, the partial parity function plays a major role in BB processing just as the previous deterministic function still does. For full l-bit parity and early on in the GA processing, half of the population will lose its market share since half will be penalized for its odd number of ones. For the other half, the GA is processing the deterministic function. Hence, it requires roughly twice the population needed when no noise is present ($n_r \approx 2$). This effect is compounded by the subsequent generation of individuals. Of course, this only applies for the initial generations where schema diversity is high. It should also be remembered that the directional function still plays a nontrivial role at this

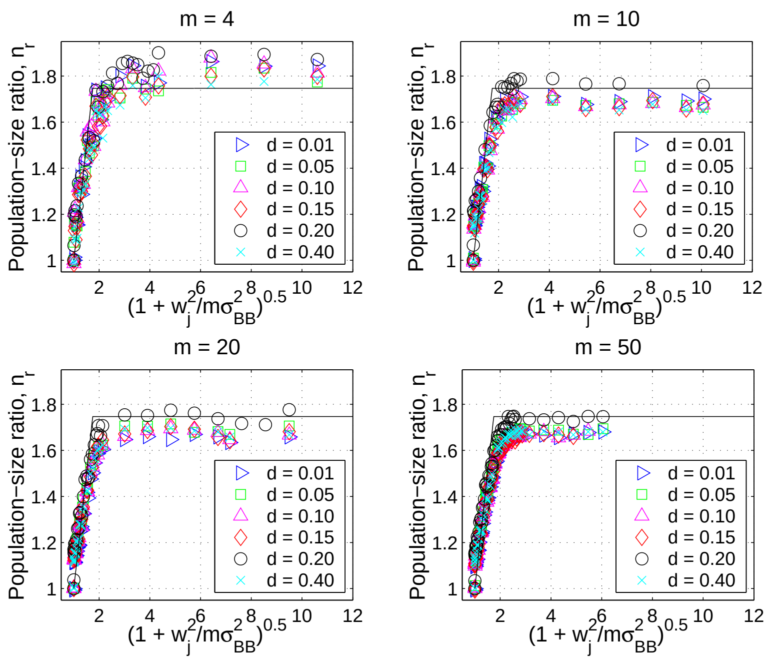

Figure 3: Population size requirements for optimal convergence when fluctuating crosstalk is present ($w_j = w_{2^l-1}$) and problem size m varies. Varying the signal (the difference from the first and second best fitness) makes little impact on the parity-induced filtering effect.

critical point and we would not expect a true doubling of the population.

If we assume an upperbound of a doubling in required population size, as empirically observed, we can estimate the critical point for the corresponding crosstalk signal:

$$n_r = 2 = \sqrt{1 + \frac{w_{2^l-1}^2}{m\sigma_{BB}^2}} \tag{6}$$

and hence

$$w_{2^l-1}^2 = 3m\sigma_{BB}^2. \tag{7}$$

Now, recall equation 5:

$$n_r = \frac{n}{n_0} = \sqrt{1 + \frac{w_{2^l-1}^2}{m\sigma_{BB}^2}}.$$

This represents the ratio of the required population size for some fitness function with noise to the required population size for the same function without noise. Note that d has entirely dropped out of the equation. Hence, this population enlargement factor due to noise does not depend on signal. This is empirically validated in figure 3.

4.3 Convergence Time

Our scalability analysis would not be complete without considering the effects of deterministic noise on convergence time. With a little reasoning we see that the filtering effect elongates convergence time similarly to growing the required population size. This may be thought of as a bump in selection pressure since with s-tournament selection the GA now needs $2s$ individuals to have the same *quality of choices* in selection. Individuals of the wrong parity are immediately discarded, and can only be chosen if all s individuals are placed into the selection pool.

Various convergence time models have been employed over the years. We forgo the development of convergence time models here but the interested reader may refer to Sastry (2001) and Goldberg (2002) for such cronologies. For our discussion, we note that convergence time models based on quantitative genetics (Bulmer, 1985; Falconer, 1989) are proving especially useful (Mühlenbein & Schlierkamp-Voosen, 1993; Mühlenbein & Schlierkamp-Voosen, 1994; Thierens & Goldberg, 1994a; Thierens & Goldberg, 1994b; Bäck, 1995; Miller & Goldberg, 1995; Miller & Goldberg, 1996; Voigt, Mühlenbein, & Schlierkamp-Voosen, 1996; Goldberg, 2002). A facetwise model using selection pressure in the presence of exogenous noise that we employ here is that of Miller and Goldberg (1995, 1996):

$$t_c = \frac{\pi\sqrt{l}}{2I}\sqrt{1+\frac{\sigma_N^2}{\sigma_f^2}} \tag{8}$$

where I is the *selection intensity* (Bulmer, 1985; Bäck, 1995; Blickle & Thiele, 1995). As we saw with population size, the ratio of the convergence time under noise to that without noise does not depend on the signal. We start by considering the convergence time needed when noise is absent:

$$t_{c0} = \frac{\pi\sqrt{l}}{2I}. \tag{9}$$

By casting equation 8 in terms of t_{c0}, and then dividing both sides by t_{c0}, we obtain the convergence time ratio:

$$t_{c,r} = \frac{t_c}{t_{c0}} = \sqrt{1+\frac{w_{2^l-1}^2}{m\sigma_{BB}^2}}. \tag{10}$$

Note again that the convergence time doesn't depend on the trap signal d. We observe from figure 4 that the predicted plots follow the model well for varying signals and building blocks.

4.4 Function Evaluations

The number of function evaluations needed to reliably find the optimal solution is a product of convergence time and population size. This measure of time is the true test of performance and based on our results of population size and convergence time, we expect results to follow the model initially until the critical point given by equation 7. Afterwards, the parity-dominated side behaves as explained previously. Figure 5 confirms this but also gives way to some immediate conclusions.

5 Future work

This study modeled fluctuating crosstalk using the highest order Walsh coefficient possible representing the case when every bit interacts with every other bit. This is a fair assumption since the smaller ordered coefficents in a Walsh transformation contribute to the directional function but the point at which this breaks down is unclear. Preliminary results using w_{2^l-2} have yielded deterministic noise behavior presented here while a smaller order coefficient such as w_2 has not.

We also seek to determine a more precise transition point between crosstalk-as-noise and crosstalk-as-signal. This requires a closer look into the cumulative effect of the subsequent generation of points derived from odd-parity individuals.

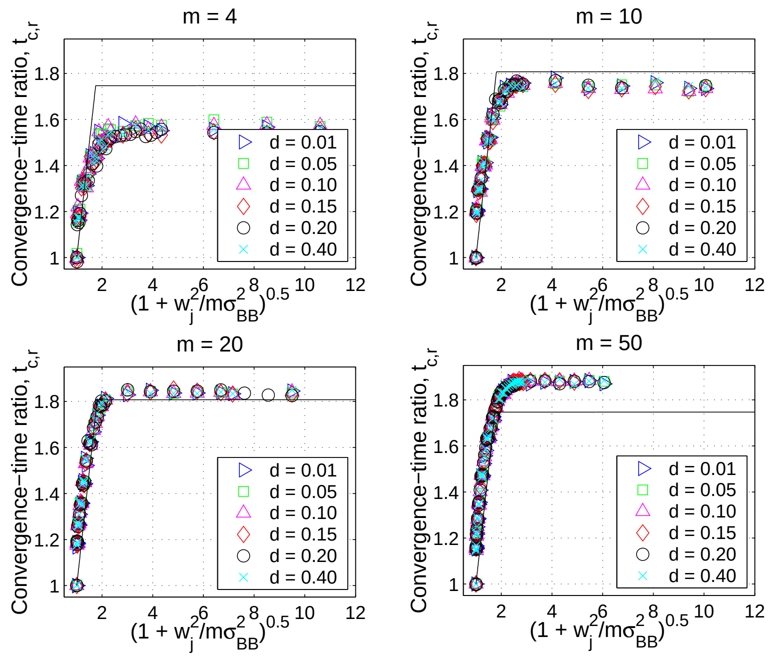

Figure 4: Convergence time requirements for optimal convergence when fluctuating crosstalk is present ($w_j = w_{2^l-1}$) and problem size m varies. Varying the signal (the difference from the first and second best fitness) makes little impact on the parity-induced filtering effect.

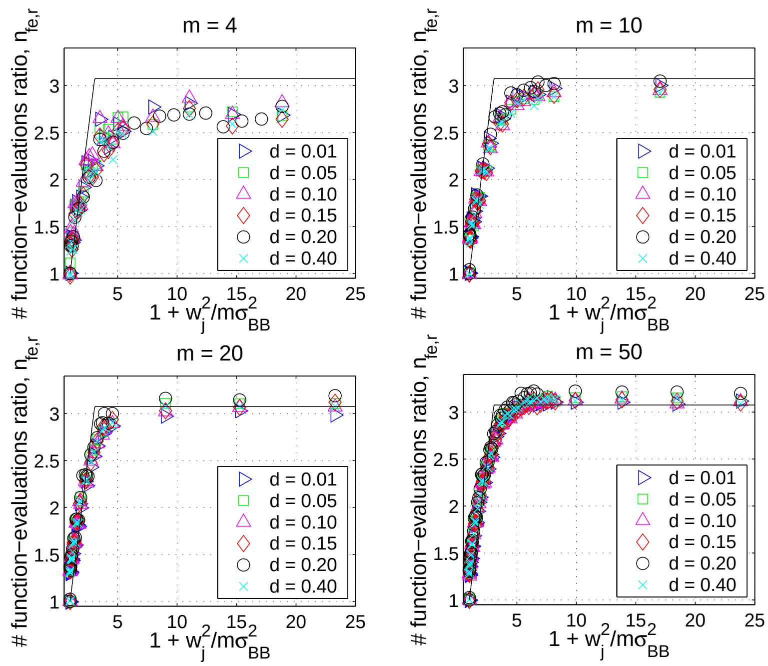

Figure 5: Functional evaluation requirements for optimal convergence when fluctuating crosstalk is present ($w_j = w_{2^l-1}$) and problem size m varies. Varying the signal (the difference from the first and second best fitness) makes little impact on the parity-induced filtering effect.

Another useful direction is to consider various forms of epistasis such as if only a portion of the entire chromosome is used to determine the parity of the individual. Such epistasis may be uniformly distributed as parity bits within each BB, confined to entire BBs, or be a mixture of both. Of course, bits may also be involved in multiple parity evaluations. We seek to consider these effects on GA scalability and under what conditions the effects of fluctuating crosstalk may be modeled as exogenous noise.

6 Summary and Conclusions

We have illustrated the introduction of noise in a deterministic fitness function via fluctuating crosstalk. We modeled fluctuating crosstalk with higher-order Walsh coefficients and showed that fluctuating crosstalk behaves like additive exogenous noise until the crosstalk variance far exceeds the underlying fitness variance by a certain threshold we empirically observe. While the crosstalk behaves similarly to external noise, its effects can be handled in a similar manner by increasing the population size and run duration.

Returning to where we started, the question that motivated this study was a prior performance comparison (Sastry & Goldberg, 2004) between mutation and crossover with and without exogenous noise. We have shown that fluctuating crosstalk acts as a source of deterministic noise and affects GA performance similarly to exogenous noise. For a precise discussion of when mutation is preferred over the selectorecombinative or vice versa, the reader should consult the aforementioned work.

7 Acknowledgments

The authors would like to acknowledge Fernando Lobo for his helpful discussions. This work was sponsored by the Air Force Office of Scientific Research, Air Force Materiel Command, USAF, under grant F49620-03-1-0129, the National Science Foundation under ITR grant DMR-03-25939 at Materials Computation Center, and ITR grant DMR-01-21695 at CPSD, and the Dept. of Energy under grant DEFG02-91ER45439 at Fredrick Seitz Materials Research Lab. The U.S. Government is authorized to reproduce and distribute reprints for government purposes notwithstanding any copyright notation thereon.

References

Bäck, T. (1995). Generalized convergence models for tournament—and (μ, λ)—selection. *Proceedings of the Sixth International Conference on Genetic Algorithms*, 2–8.

Bethke, A. D. (1981). *Genetic algorithms as function optimizers*. Doctoral dissertation, The University of Michigan.

Blickle, T., & Thiele, L. (1995). A mathematical analysis of tournament selection. *Proceedings of the Sixth International Conference on Genetic Algorithms*, 9–16.

Bulmer, M. G. (1985). *The mathematical theory of quantitative genetics*. Oxford: Oxford University Press.

Davidor, Y. (1991). Epistasis variance: A viewpoint on GA-hardness. *foga91*, 23–35.

Falconer, D. S. (1989). *Introduction to quantitative genetics* (Third ed.). New York, NY, USA; London, UK; Sydney, Australia: John Wiley and Sons.

Goldberg, D. E. (1989). Genetic algorithms and Walsh functions: Part I, a gentle introduction. *Complex Systems*, *3*(2), 129–152. (Also TCGA Report 88006).

Goldberg, D. E. (2002). *Design of innovation: Lessons from and for competent genetic algorithms*. Boston, MA: Kluwer Acadamic Publishers.

Goldberg, D. E., Deb, K., & Clark, J. H. (1992). Genetic algorithms, noise, and the sizing of populations. *Complex Systems*, *6*, 333–362. (Also IlliGAL Report No. 91010).

Harik, G., Cantú-Paz, E., Goldberg, D. E., & Miller, B. L. (1999). The gambler's ruin problem, genetic algorithms, and the sizing of populations. *Evolutionary Computation*, *7*(3), 231–253. (Also IlliGAL Report No. 96004).

Heckendorn, R. B., & Whitley, D. (1999). Predicting epistasis from mathematical models. *Evolutionary Computation*, *7*(1), 69–101.

Kumar, V. (2002). Tackling epistasis: A survey of measures and techniques. *Assignment from an advanced GEC course taught at UIUC by D. E. Goldberg*.

Miller, B. L., & Goldberg, D. E. (1995). Genetic algorithms, tournament selection, and the effects of noise. *Complex Systems*, *9*(3), 193–212. (Also IlliGAL Report No. 95006).

Miller, B. L., & Goldberg, D. E. (1996). Genetic algorithms, selection schemes, and the varying effects of noise. *Evolutionary Computation*, *4*(2), 113–131. (Also IlliGAL Report No. 95009).

Mühlenbein, H., & Schlierkamp-Voosen, D. (1993). Predictive models for the breeder genetic algorithm: I. continous parameter optimization. *Evolutionary Computation*, *1*(1), 25–49.

Mühlenbein, H., & Schlierkamp-Voosen, D. (1994). The science of breeding and its application to the breeder genetic algorithm (BGA). *Evolutionary Computation*, *1*(4), 335–360.

Naudts, B., & Kallel, L. (1998). *Some facts about so-called ga-hardness measures* (Tech. Rep. No. 379). France: Ecole Polytechnique, CMAP.

Pelikan, M., Goldberg, D. E., & Cantú-Paz, E. (2000). Linkage learning, estimation distribution, and Bayesian networks. *Evolutionary Computation*, *8*(3), 314–341. (Also IlliGAL Report No. 98013).

Sastry, K. (2001). *Evaluation-relaxation schemes for genetic and evolutionary algorithms*. Master's thesis, University of Illinois at Urbana-Champaign, General Engineering Department, Urbana, IL. (Also IlliGAL Report No. 2002004).

Sastry, K., & Goldberg, D. E. (2004). Let's get ready to rumble: Crossover versus mutation head to head. *Proceedings of the 2004 Genetic and Evolutionary Computation Conference*, *2*, 126–137. Also IlliGAL Report No. 2004005.

Thierens, D., & Goldberg, D. E. (1994a). Convergence models of genetic algorithm selection schemes. *Parallel Problem Solving from Nature*, *3*, 116–121.

Thierens, D., & Goldberg, D. E. (1994b). Elitist recombination: An integrated selection recombination GA. *Proceedings of the First IEEE Conference on Evolutionary Computation*, 508–512.

Voigt, H.-M., Mühlenbein, H., & Schlierkamp-Voosen, D. (1996). The response to selection equation for skew fitness distributions. *Proceedings of the International Conference on Evolutionary Computation*, 820–825.