EC Theory: A Unified Viewpoint

Christopher R. Stephens and Adolfo Zamora

Instituto de Ciencias Nucleares, UNAM, Circuito Exterior
A. Postal 70-543, México D.F. 04510
{stephens,zamora}@nuclecu.unam.mx

Abstract. In this paper we show how recent theoretical developments have led to a more unified framework for understanding different branches of Evolutionary Computation (EC), as well as distinct theoretical formulations, such as the Schema theorem and the Vose model.In particular, we show how transformations of the configuration space of the model – such as coarse-grainings/projections, embeddings and coordinate transformations – link these different, disparate elements and alter the standard taxonomic classification of EC and different EC theories. As an illustration we emphasize one particular coordinate transformation between the standard string basis and the "Building Block" basis.

1 Introduction

The development of Evolutionary Computation (EC) up to now has been principally empirical and phenomenological, theory only playing a relatively limited role. In its more "scientific" guise it is related to an older, more mature field – population genetics. Mathematically speaking it is the study of certain classes of heuristic algorithms based on populations of objects (for a rigorous exposition in the context of Genetic Algorithms (GAs) see [1]). From the "engineering" point of view it is an area where the analogical use of "natural selection" appears as a moulding force in the creation of more "competent" problem-solvers [2].

In distinction to other more mature areas of science there is not really a clear consensus on what should be the task of EC theory. Is it to provide recipes for practitioners, to provide exact computational models, to allow a deeper understanding of a complex system, all of these, none of these, or what?

Different approaches to EC theory have been proposed in the past, often with proponents who have been disparaging of the others' point of view. These include schema theories [3], the Vose model [4], the statistical mechanics approach [5] and more. Is there a model that is superior to all others? Often, models are judged by their clarity, simplicity, and ability to explain and predict. Is there a framework that does this best?

A theoretical model is also often judged by how well it unifies a range of phenomena. As there are many different flavours of Evolutionary Algorithm (EA) – GAs, Genetic Programming (GP), Evolution Strategies (ES) etc. – one may ask if there is a theoretical framework that encompasses them all? If not, then which is the framework with the broadest applicability?

E. Cantú-Paz et al. (Eds.): GECCO 2003, LNCS 2724, pp. 1394–1405, 2003.

In many sciences a large part of theory is associated with taxonomy – classification with respect to natural relationships. In EC, various high-level taxonomic labels are at our disposal, such as GP, GAs, ES etc. Whether these labels are optimal, or even useful other than in an historic sense, however, is a debatable point. Taxonomy allows us to understand commonality between different things. Subsequently we must understand why such commonality exists. For instance, the periodic table was initially an empirical and phenomenological construct until the atomic theory gave it a firm "microscopic" foundation. What is the "periodic table" for EC? Does such a construct exist? If nothing of this nature existed it would be deeply worrying as it would mean that a theoretical treatment of each and every EA and/or problem would be different. It is clear however that there is commonality. The question is more – can it be suitably formalized?

At the other extreme one could claim a type of "hyperuniversality", such as was present in the original version of the Building Block Hypothesis [6], which claimed that all GAs behaved in the same way in finding an optimum – via fit, short schemata. We now know that this, in its strict form, is wrong, being rather an engineering rule-of-thumb with only limited validity, and that such a degree of hyperuniversality does not exist. Nevertheless, a prime job of EC theory should be to tell us what EAs and problems, or classes of EAs and problems, are likely to lead to similar outcomes or behaviour. It does not need to be elaborated on that a deeper understanding of this would be of great use to practitioners.

In this article we wish to give a more unified presentation of EC theory showing how recent developments [7,8,9,10,11,12] have pointed to a new understanding of the relationships between the different branches of EC and the different theoretical formulations.

2 Some Fundamentals of EC Theory

We first describe, briefly and non-rigorously, some fundamentals of EC theory that could be applied to many different types of EA. We try to maintain as much generality as possible, in particular to show how a unified theoretical framework, encompassing most standard EAs, can be developed. Formally, an EA is a stochastic algorithm that takes as input a population of "objects" (strings, trees etc.) and a fitness function, at a given time, and gives as output the population at a later time. The objects live on a configuration space X, of dimensionality N_X, with elements $i \in X$. We denote a population by $\mathbf{P} = (n_1, n_2, ..., n_{N_X})$, where n_i represents the proportion of objects of type i in the population. Each object is assigned a quality or fitness, via a fitness function $f_X : X \to R^+$.

A dynamics is imposed via an evolution operator which in the infinite population limit leads to

$$\mathbf{P}(t+1) = \mathcal{S}\left(\mathbf{P}(t), \mathsf{f}\right) \circ \mathcal{T}\left(\mathbf{P}(t), \mathsf{p}\right) , \tag{1}$$

where $\circ$ denotes the Schur product of vectors. The transmission term $\mathcal{T}(\mathbf{P}(t), \mathsf{p})$ describes the probability of transforming one object into another one by mutation, crossover, or other genetic operators, the explicit transmission mechanism

being encoded by the parameters p. The term $\mathcal{S}(\mathbf{P}, \mathsf{f})$ describes the selection forces acting on P with the parameters f determining the fitness function.

For selection the number of parameters necessary depends on the type of fitness function and the amount of degeneracy of f_X. For instance, for one-max, only N fitness values are needed due to the degeneracy of the genotype-phenotype map. Mutation usually only depends on one parameter – the mutation probability. Two-parent recombination depends on the set of recombination distributions, $\{\lambda_{ijk}(m)\}$, that characterize the transferral of genetic material from parents to offspring, where $\lambda_{ijk}(m)$ is the probability to form an offspring object, i, given two parent objects, j and k, and a crossover "mode", m, i.e. a rule for redistributing genetic material between parent and offspring objects.

We mentioned previously that taxonomy is important without specifying what should be classified. One may think that it is EAs themselves. An EA alone however, is in some sense a "black box" which takes a "problem" (usually a fitness landscape and an initial population) as input and then gives an output (the population at a later time). A given EA, though, may have very different characteristics with respect to a given measure on one problem versus another, hence, one is led to study EA-problem pairs rather than EAs in isolation. We will call an EA/problem pair a "model", characterizing a particular model by a set of parameters p and f, and taking the models to live on a space, $\mathcal{E}$. In principle one could put a metric on $\mathcal{E}$ and talk about how "close" one model is to another. A less rigorous, but more pragmatic, approach is to think of two models as being "close" if they lead to "similar" behaviour. Of course, to do this one must define "similarity measures". A simple example might be the time taken for a particular correlation between two loci in a population to decrease by a given factor. Continuity on $\mathcal{E}$ would lead one to believe that models with similar parameter values should behave similarly.

Population flows take place on X. All the main branches of EC – GP, GAs, ES etc. – fall into this general framework. The chief differences lie more in what objects are being represented in X and what specific operators generate the dynamics. For instance, in GAs the i represent fixed length strings. In GP they are program trees and in machine code GP [13] or Grammatical evolution [14] they are variable length strings. "Coarse grained" representations, such as phenotypes or schemata also offer very useful representations.

3 Generic Genetic Dynamics

The space of models, $\mathcal{E}$, is very large if one thinks of all possible genetic operators. Selection, mutation and crossover, however, form a very generic set and we will now restrict attention to them. Formally at least, the following also applies to GP as well as GAs:

$$P_i(t+1) = \sum_j \mathcal{P}_{ij} P_j^c(t) \tag{2}$$

where $P_i^c(t)$ is the probability to find objects of type i after selection and crossover. The matrix elements of the mutation matrix, $\mathcal{P}$, give the probabil-

ity to mutate object j to object i. In the simple case of fixed length GAs for instance, $\mathcal{P}_{ij} = p^{d^H(i,j)}(1-p)^{N-d^H(i,j)}$, where d^H_{ij} is the Hamming distance between the two strings and N is the strings' length. For mutation Hamming distance is clearly a very natural metric. Explicitly $P^c_i(t)$ is given by

$$P^c_i(t) = (1-p_c)P'_i(t) + \sum_m \sum_j \sum_k \lambda_{ijk}(m) P'_j(t) P'_k(t) \tag{3}$$

where $P'_i(t)$ is the probability to select i. $P'_i = (f_i/\bar{f}(t))P_i$ for proportional selection, where $\bar{f}$ is the average population fitness. $\lambda_{ijk}(m)$ is an interaction term between objects, i.e. objects j and k are selected and crossed over ("interact") to potentially form an object i, depending not only on j and k but also on the particular recombination mode. In the case of homologous crossover these modes are just crossover masks with $\sum_m$ being the sum over all possible masks while for non-homologous crossover they are more general. Equations (2) and (3), as an exact representation of the dynamics, in the case of fixed-length GAs, where a crossover mode is simply a mask, are equivalent to the Vose model or, indeed, to earlier formulations in population biology. These equations however are also valid for objects other than fixed-length strings.

$\lambda_{ijk}(m) = 0$ unless the mode m creates object i from j and k. This is unlikely and hence the vast majority of interactions are zero. e.g. in GAs with binary alleles, for a given i and m, $\lambda_{ijk}(m)$ is a 2^N-dimensional square matrix. However, only of the order of 2^N matrix elements are non-zero. Thus, the microscopic representation is very inefficient. This also hold for more complicated types of object.

4 Coarse Graining and Coordinate Transformations

Rather than considering one type of representation as being more "general" or fundamental than another it is useful to think of transforming between different representations. We will consider three basic types of transformation: coarse grainings, "coordinate" transformations, and embeddings, concentrating mainly on the first two. Such transformations give flexibility in terms of what particular representation we may find most suitable for a problem and also give a more unified framework within which we may view different elements of EC, such as GP and GAs, in a more coherent light. More importantly they can facilitate an understanding of the dynamical equations associated with the true effective degrees of freedom of the model. These effective degrees of freedom will more often than not be aggregations of the underlying "microscopic" degrees of freedom and may be made more manifest via a coordinate transformation, embedding or coarse-graining/projection. Additionally, it may be the case that effective degrees of freedom most naturally emerge in an approximation to the dynamics rather than the exact dynamics.

4.1 Coarse Graining

The generic dynamics discussed in the previous section is described by an exponentially large number of coupled, non-linear difference equations representing the microscopic degrees of freedom, i.e. the completely specified objects themselves. In the absence of recombination, the equations are essentially linear and the problem reduces down to finding the eigenvalues and eigenvectors of the selection-mutation matrix. However, save in very simple problems, even this simpler problem is formidable. Crossover adds yet another layer of complexity. Naturally, in such problems one wishes to find the correct effective degrees of freedom so as to be able to affect an effective reduction in the dimensionality of the problem. Such a reduction can be affected by an appropriate coarse graining.

We can formalize these considerations by introducing a general coarse graining operator, $\mathcal{R}(\eta, \eta')$, which coarse grains from the variable $\eta \in X_\eta$ to the variable $\eta' \in X_{\eta'} \subset X_\eta$. Thus, the action of $\mathcal{R}$ is a projection. Given two such coarse grainings we have

$$\mathcal{R}(\eta, \eta')P_\eta(t) = P_{\eta'}(t) \qquad \mathcal{R}(\eta, \eta'')P_\eta(t) = P_{\eta''}(t) \tag{4}$$

However, given that $\mathcal{R}(\eta', \eta'')P_{\eta'}(t) = P_{\eta''}(t)$ we deduce that

$$\mathcal{R}(\eta, \eta'') = \mathcal{R}(\eta, \eta')\mathcal{R}(\eta', \eta'') \tag{5}$$

i.e., the space of coarse grainings has a semi-group structure. This type of structure is known, by an abuse of language, as the Renormalization Group. The naturalness of a particular coarse graining transformation will be to a large extent determined by how the transformed dynamics looks.

Considering (1), then given that $\mathcal{R}(\eta, \eta')P_\eta(t) = P_{\eta'}(t)$ the dynamics under a coarse graining is governed by $\mathcal{R}(\eta, \eta')\mathcal{S}\left(\mathbf{P}_\eta(t), \mathsf{f}\right) \circ \mathcal{T}\left(\mathbf{P}_\eta(t), \mathsf{p}\right)$, where $\mathcal{S}\left(\mathbf{P}_\eta(t), \mathsf{f}\right)$ and $\mathcal{T}\left(\mathbf{P}_\eta(t), \mathsf{p}\right)$ are the dynamical operators associated with the variables η. If this can be written in the form $\mathcal{S}\left(\mathbf{P}_{\eta'}(t), \mathsf{f}'\right) \circ \mathcal{T}\left(\mathbf{P}_{\eta'}(t), \mathsf{p}'\right)$ with suitable "renormalizations", $\mathsf{f} \longrightarrow \mathsf{f}'$ and $\mathsf{p} \longrightarrow \mathsf{p}'$ of the model's parameters, then the dynamics is form covariant or invariant under this coarse graining. Note that we are here considering a more general notion of invariance than the idea of "compatibility" [1] (see [15] for a discussion of the relationship between the two). In the case of selection only, the coarse graining transforms the fitness

$$f_\eta \longrightarrow f_{\eta'} = \mathcal{R}(\eta, \eta')f_\eta = \sum_{\eta \in \eta'} f_\eta P_\eta(t) / \sum_{\eta \in \eta'} P_\eta(t). \tag{6}$$

An important point to note here is that, generically, a coarse graining gives rise to a *time dependent* coarse-grained fitness.

Of course, there are many types of coarse graining procedure all of which lead to a dimensional reduction. Such reductions can sometimes come about in a relatively "trivial" fashion, such as in the case of the genotype-phenotype map, where the dynamics is invariant in the case of selection and in the absence of mixing operators. In fact, it is strictly invariant not just form invariant, as there is no "renormalization" necessary of any parameter or variable and we have $f_{\eta'} =$

$\mathcal{R}(\eta, \eta')f_\eta = f_\eta$, where, here, η' represents the phenotype and η the genotype. A concrete example is that of the "needle-in-a-haystack" landscape where the fitness landscape is degenerate for all genotypes except one, the "needle". For selection only, as there are only two phenotypes, there is a reduction in the size of the configuration space from 2^N to 2, i.e. a reduction in the number of degrees of freedom from N to 1. However, if we include in the effect of mutation we see that there is an induced breaking of the genotype-phenotype symmetry due to the fact that sequences close to the master sequence in Hamming distance have more offspring than the equally fit genotypes that are further away. In this case $\mathcal{R}(\eta, \eta')\mathcal{S}\left(\mathbf{P}_\eta(t), \mathsf{f}\right) \circ \mathcal{T}\left(\mathbf{P}_\eta(t), \mathsf{p}\right) \neq \mathcal{S}\left(\mathbf{P}_{\eta'}(t), \mathsf{f}'\right) \circ \mathcal{T}\left(\mathbf{P}_{\eta'}(t), \mathsf{p}'\right)$ and the natural effective degrees of freedom are Hamming classes rather than phenotypes.

Another important class of coarse grainings is that of "schemata", which we will denote by α with $P_\alpha(t)$ its relative frequency at time t. In this case the action of the coarse graining operator is: $\mathcal{R}(x, \alpha)P_x(t) = P_\alpha(t) = \sum_{x \in \alpha} P_x(t)$. Schemata have a simple geometric interpretation in the binary case, a particular schema being represented by an $(N - N_2)$-dimensional hyperplane in X which passes through the 2^{N-N_2} vertices that represent the loci that have been coarse grained. A schema partition then consists of 2^{N_2} of such 2^{N-N_2}-dimensional hyperplanes. Thus, there is an effective dimensional reduction from a 2^N-dimensional configuration space X to a 2^{N_2}-dimensional one, X_α, and a corresponding reduction in the number of degrees of freedom from N to N_2.

Unlike the simple case of the coarse graining to phenotype, here the coarse-grained fitness is time-dependent, with the "renormalized" fitness now being a highly non-trivial function of the original microscopic fitness, being defined as $f_\alpha = \mathcal{R}(x, \alpha)f_x = \sum_{x \in \alpha} f_x P_x(t) / \sum_{x \in \alpha} P_x(t)$. As there exist 3^N possible schemata a full schema basis is over-complete and non-orthonormal. However, the space of schemata is not the natural one for recombination, as we shall see.

4.2 Embeddings

In the case of embeddings one passes from a lower to a higher dimensional configuration space. An example would be that of passing from a representation where objects are represented by variable-length strings, of up to maximum size N_m with binary alleles, to a fixed-length representation of size N_m, by including a third allele value that specifies that there was no corresponding bit in the variable length case. The original configuration space is of dimension $2(2^{N_m} - 1)$. However, due to the addition of a third allele the dimension of the embedding space is 3^{N_m}. Of course, for these more general transformations development of the operators and the theory necessary to maintain syntactic correctness of the offspring is a open issue. In this case, one might be better off using the theory for variable length structures already developed in GP. Lest the reader think that this type of transformation is somewhat exotic we may mention the case of protein structure where the protein can be specified at the level of a linear sequence of amino acids but which later forms a two-dimensional secondary structure and finally a three-dimensional tertiary structure.

4.3 Coordinate Transformations

Coordinate transformations are a standard tool in the mathematical analysis of physical models where one seeks a set of variables that is well-adapted to the internal structure of the model and hence simplifies the structure of the equations that have to be analyzed. For our purposes we will only need *linear* transformations which can be described in terms of matrices. We restrict our discussion of explicit examples to the case of binary strings. In this case the standard string basis is: $x = (x_1, \ldots, x_N)$ with $x_i = 0, 1$ and the configuration space the N-dimensional Boolean hypercube.

The three alternative bases we will consider: the Walsh basis, the Taylor basis and the Building Block Basis are all related to the standard basis via linear transformations. In (14), (15) and (16) we show the explicit transformation matrices for the case of three loci.

Walsh (Fourier) Basis. Probably the most important alternative basis is the *Walsh basis* ψ, consisting of *Walsh functions*, $\psi_I(x) = 1/\sqrt{|X|}\prod_{j\in I} x_j$, where I is a subset of $\{1, \ldots, N\}$ and $x_j = \pm 1$. The Walsh functions are normed and orthogonal and of *order* $|I|$, the number of loci that are multiplied. The Walsh-transform, $\hat{f}$, of a function f is defined implicitly by $f(i) = \sum_I \hat{f}(I)\psi_I(i)$ Multiplying with $\psi_K(x)$ and summing over all $i \in X$ we obtain

$$\sum_{i\in X} f(i)\psi_K(i) = \sum_I \hat{f}(I) \sum_{i\in X} \psi_I(i)\psi_K(i) = \sum_I \hat{f}(I)\delta_{IK} = \hat{f}(K) \tag{7}$$

In matrix form $\hat{f} = \mathbf{\Psi} f$, where the matrix $\mathbf{\Psi}$ has the Walsh functions ψ_K as its rows. One of the most important properties of the Walsh functions is that they are eigenfunctions of the mutation operator $\mathcal{P}$ that satisfies $\mathcal{P}\psi_I = (1-2|I|/N)\psi_I$ The mutation operator is therefore diagonal in the Walsh basis. Equation (2) reads in these coordinates

$$\hat{P}_I(t+1) = \sum_i (\mathbf{\Psi}\mathcal{P})_{Ii} P_i^c = \left(1 - 2\frac{|I|}{N}\right)\hat{P}_I^c \tag{8}$$

The Walsh basis will be particularly useful if the transformed selection-crossover term $\hat{P}_I^c$ also has a simple form.

Taylor Series. While the standard basis and the Walsh basis ψ are orthonormal, this is not necessarily the case in general. In [16], for instance, the *Taylor series* of a landscape on the Boolean hypercube is introduced in terms of the polynomials $\tau_I(i) = \prod_{i\in I} \tilde{x}_i$, such that $f(i) = \sum_I \tilde{f}(I)\tau_I(i)$. Let us write $i \subset I$ if $\tilde{x}_i = x_i = 1$ for all elements of I. We define the matrix $\mathbf{\Upsilon}$ by $\mathbf{\Upsilon}_{Ii} = 1$ if $i \subset I$ and $\mathbf{\Upsilon}_{Ii} = 0$ otherwise, i.e., $\mathbf{\Upsilon}_{Ii} = \tau_I(i)$. Thus, we can write the Taylor series expansion in the form $f = \mathbf{\Upsilon}\tilde{f}$, i.e., $\tilde{f} = \mathbf{\Upsilon}^{-1} f$. The matrix $\mathbf{\Upsilon}$ is invertible [16] but is neither normalized nor orthogonal. i.e. the basis functions $\tau_I(i)$ do not form an orthonormal basis. In fact, we have $\sum_{i\in X} \tau_I(i)\tau_k(i) = 2^{N-|I\cup K|}$ since $\tau_I(i)\tau_k(i) = 1$ whenever $x_i = 1$ for all elements of $I \cup K$, and 0 otherwise.

The Building Block Basis. Coordinate transformations such as the Walsh transform and the Taylor series are general. Here, however, we wish to consider a particular coordinate transformation that arises as an almost inevitable consequence of the action of recombination. We saw that representing recombination in terms of the fundamental, microscopic objects is very inefficient due to the sparsity of the interaction matrix. This is an indication that individual objects are not the natural effective degrees of freedom for recombination. So what are? To form the string `111` with a mask `100` one can join strings `111`, `110`, `101`, and `100` with either `111` or `011`. In other words, for the first parent the second and third bit values are unimportant and for the second the first bit value is unimportant. Thus, it is natural to coarse grain over those strings that give rise to the desired target for a given mask.

If one picks, arbitrarily, a vertex in X, associated with a string i, one may perform a linear coordinate transformation $\Lambda : X \to \tilde{X}$ to a basis consisting of all schemata that contain i. For instance, for two bits $X = \{\texttt{11}, \texttt{10}, \texttt{01}, \texttt{00}\}$, while $\tilde{X} = \{\texttt{11}, \texttt{1}*, *\texttt{1}, **\}$. The invertible matrix $\mathbf{\Lambda}$ is such that $\mathbf{\Lambda}_{\alpha i} = 1 \iff i \in \alpha$.

$$\mathbf{\Lambda} = \begin{array}{c|cccccccc} & 111 & 110 & 101 & 011 & 100 & 010 & 001 & 000 \\ \hline \texttt{111} & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \texttt{11}* & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ \texttt{1}*\texttt{1} & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ *\texttt{11} & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ \texttt{1}** & 1 & 1 & 1 & 0 & 1 & 0 & 0 & 0 \\ *\texttt{1}* & 1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 \\ **\texttt{1} & 1 & 0 & 1 & 1 & 0 & 0 & 1 & 0 \\ *** & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{array} \tag{9}$$

$$\mathbf{\Upsilon} = \begin{array}{c|cccccccc} & ooo & loo & olo & ool & llo & lol & oll & lll \\ \hline 111 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 110 & 1 & 1 & 1 & 0 & 1 & 0 & 0 & 0 \\ 101 & 1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 \\ 011 & 1 & 0 & 1 & 1 & 0 & 0 & 1 & 0 \\ 100 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 010 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 001 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 000 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array} \tag{10}$$

$$\mathbf{\Psi} = \frac{1}{\sqrt{8}} \begin{array}{c|cccccccc} & 111 & 110 & 101 & 011 & 100 & 010 & 001 & 000 \\ \hline 111 & -1 & +1 & +1 & +1 & -1 & -1 & -1 & +1 \\ 110 & +1 & -1 & -1 & -1 & -1 & -1 & +1 & +1 \\ 101 & +1 & -1 & +1 & -1 & -1 & +1 & -1 & +1 \\ 011 & +1 & -1 & -1 & +1 & +1 & -1 & -1 & +1 \\ 100 & -1 & -1 & -1 & +1 & -1 & +1 & +1 & +1 \\ 010 & -1 & -1 & +1 & -1 & +1 & -1 & +1 & +1 \\ 001 & -1 & +1 & -1 & -1 & +1 & +1 & -1 & +1 \\ 000 & +1 & +1 & +1 & +1 & +1 & +1 & +1 & +1 \end{array} \tag{11}$$

We denote the associated coordinate basis the Building Block basis (BBB)[1] in that one may think of the elements of this basis as the BBs[2] that are joined together by crossover to form i.[3] The BBB is complete but not orthonormal. Note that the vertex i by construction is a fixed point of this transformation. Apart from the vertex i, the points in $\tilde{X}$, being schemata, correspond to higher dimensional objects in X. For instance, $1*$ and $*1$ are one-planes in X while $**$ is the whole space. In the BBB one may transform (3) to find

$$\tilde{P}^c_\alpha(t+1) = (1-p_c)\tilde{P}'_\alpha(t) + \sum_{m=1}^{2^N} \sum_{\beta,\gamma} \tilde{\lambda}_{\alpha\beta\gamma}(m)\tilde{P}'_\beta(t)\tilde{P}'_\gamma(t) \tag{12}$$

where $\tilde{\lambda}_{\alpha\beta\gamma}(m) = \mathbf{\Lambda}_{\alpha i}\lambda_{ijk}\mathbf{\Lambda}^{-1}_{\beta j}\mathbf{\Lambda}^{-1}_{\gamma k}$.

The advantage of this new representation is that the properties and symmetries of crossover are much more transparent. For instance, $\tilde{\lambda}_{\alpha\beta\gamma}(m)$ is such that for a given mask only interactions between BBs that construct the target schema are non-zero, i.e., $\tilde{\lambda}_{\alpha\beta\gamma}(m) = 0$, unless γ corresponds to a schema which is the complement of β with respect to α. Also, $\tilde{\lambda}_{\alpha\beta\gamma}(m) = 0$ unless β is equivalent to m, whereby equivalent means that for any 1 in the mask we have a 1 at the corresponding locus in β and for any 0 we have a $*$. These two important properties mean that the summations over β and γ in (12) disappear to leave only the sum over masks with an "interaction" constant $p_c(m)$ which depends only on the mask. For example, for two bits, if we choose as vertex 11, then 11 may interact only with $**$, while $1*$ may interact only with $*1$.

In X this has the interesting interpretation that for a target schema α of dimensionality $(N-d)$ only geometric objects "dual" in the d-dimensional subspace of X that corresponds to α may interact. In other words, a k-dimensional object recombines only with a $(N-d-k)$-dimensional object. Additionally, a $(N-d)$-dimensional object may only be formed by the interaction of higher dimensional objects. In this sense interaction is via the geometric intersection of higher dimensional objects. For example, the point 11 can be formed by the intersection of the two lines $1*$ and $*1$. Similarly, 1111 can be formed via intersection of the three-plane $1 * **$ with the line $*111$ or via the intersection of the two two-planes $11 * *$ and $* * 11$.

As mentioned, one of the primary advantages of the BBB representation is that the sums over j and k disappear thus obtaining

$$P^c_i(t) = (1-p_c)P'_i(t) + \sum_{m=1}^{2^N} p_c(m)P'_{i_m}(t)P'_{i_{\bar{m}}}(t) \tag{13}$$

[1] This basis is implicit in the work [7,8] but has only been considered in more detail recently [17].

[2] Note that these BBs are not the same as their well known counterparts in GA theory [6] being dynamic, not static, objects. Neither are they necessarily short or fit.

[3] Given the arbitrariness of the choice of vertex there are in fact 2^N equivalent BBBs each transformable to any other by a permutation.

where $P'_{i_m}(t)$ is the probability to select the BB i_m (note that the mask uniquely specifies which element, i_m, of the BBB to choose) and $P'_{i_{\bar{m}}}(t)$ is the probability to select the BB $i_{\bar{m}}$, which is uniquely specified as the complement of i_m in i. Both i_m and $i_{\bar{m}}$ are elements of the BBB associated with i. The above equation clearly shows that recombination is most naturally considered in terms of the BBB. In the standard basis there were of the order of 2^{2N} elements of λ_{ijk} to be taken into account for a fixed i. In the BBB there is only one term. Of course, the coarse grained averages of i_m and $i_{\bar{m}}$ contain 2^N terms, still, the reduction in complication is enormous. Thus, crossover naturally introduces the idea of a coarse graining, the associated effective degrees of freedom being the BBs we have defined. This is an important point as it shows that evolution is acting in the presence of crossover most naturally at the level of populations, the BBs representing populations with a certain degree of "kinship" to the target object.

Inserting (13) in (2) we can try to solve for the dynamics. However, in order to do that we must know the time dependence of x_m and $x_{\bar{m}}$. Although the number of BB basis elements is 2^N we may generalize and consider the evolution of an arbitrary schema, α. To do this we need to sum with $\sum_{x\in\alpha}$ on both sides of the equation (2). This can simply be done to obtain [7,8] again the form (2), where this time the index α runs only over the 2^{N_2} elements of the schema partition and where again $\mathbf{M}_{\alpha\beta} = p^{d^H(\alpha,\beta)}(1-p)^{N-d^H(\alpha,\beta)}$. In this case however $d^H(\alpha,\beta)$ is the Hamming distance between the two schemata. For instance, for strings with three loci the schemata partition associated with the first and third loci is $\{1*1, 1*0, 0*1, 0*0\}$. In this case $d^H(1,2) = 1$ and $d^H(1,4) = 2$. $P^c_\alpha(t) = \sum_{x\in\alpha} P^c_x(t)$ is the probability of finding the schema α after selection and recombination. Note the form invariance of the equation after coarse graining. To complete the transformation to schema dynamics we need the schema analog of (13). This also can be obtained by acting with $\sum_{x\in\alpha}$ on both sides of the equation. One obtains

$$P^c_\alpha(t) = (1 - p_c N_\alpha) P'_\alpha(t) + \sum_{m\in\mathcal{M}_r} p_c(m) P'_{\alpha_m}(t) P'_{\alpha_{\bar{m}}}(t) \tag{14}$$

where α_m represents the part of the schema α inherited from the first parent and $\alpha_{\bar{m}}$ that part inherited from the second. $N(\alpha)$ is the number of crossover masks that affect α, relative to the total number of masks with $p_c(m) \neq 0$, the set of such masks being denoted by $\mathcal{M}_r$. Obviously, these quantities depend on the type of crossover implemented and on properties of the schema such as defining length. Note that the BBB naturally coarse grains here to the BBB appropriate for the schema α as opposed to the string x.

Thus, we see that the evolution equation for schemata is form invariant, there being only a simple multiplicative renormalization of the recombination probability p_c. This form invariance, shown in [7], demonstrates that BB schemata in general are a preferred set of coarse grained variables and, more particularly, the BBB is a preferred basis in the presence of recombination. It has also been shown [1] that schemata, more generally, are the only coarse graining that leads to invariance in the presence of mutation and recombination.

Considering again the structure of (13) and (14) we see that variables associated with a certain degree of coarse graining are related to BB "precursors" at an earlier time, which in their turn ... etc. This hierarchical structure terminates at order-one BBs as these are unaffected by crossover. Thus, for example, the level one BB combinations of 111, i.e., BBs that lead directly upon recombination to 111 are: $11* : **1$, $1*1 : *1*$ and $1** : *11$. The level two BBs are $1**$, $*1*$ and $**1$. Thus, a typical construction process is that BBs $1**$ and $*1*$ recombine at $t = t_1$ to form the BB $11*$ which at some later time t_2 recombines with the BB $**1$ to form the sequence 111.

In the case of recombination note also that the coarse graining operator associated with the BBs satisfies

$$\mathcal{R}(\eta, \eta') = \mathcal{R}(\eta^m, \eta'^m)\mathcal{R}(\eta^{\bar{m}}, \eta'^{\bar{m}}) \tag{15}$$

where $\mathcal{R}(\eta^m, \eta'^m)$ represents the action of the coarse graining on the BB $\mathcal{S}$ while $\mathcal{R}(\eta^{\bar{m}}, \eta'^{\bar{m}})$ represents the action on the BB $\bar{m}$.

5 Conclusions

In this paper, based on a coarse-grained, or schema-based, formulation of genetic dynamics we have seen how various branches of EC, in particular GAs and GP, can be understood in a more unified framework. Additionally, we have explicitly demonstrated how to pass between different formulations of genetic dynamics, such as the Vose model and schema-based models, via coordinate transformations on the configuration space, showing how the most natural basis for crossover – the BBB – can be obtained from the standard string basis. The emphasis here has been on how transformations of the configuration space of the model – such as coarse-grainings/projections, embeddings and coordinate transformations – link these different, disparate elements and alter the standard taxonomic classification of EAs and theories of EAs. We firmly believe that a more profound understanding of EC in general will result from a deeper understanding of these, and other, transformations. Finally, it is important to note that such a unified viewpoint has already led to several notable advances, such as proofs of the well known Geiringer's theorem to the case of variable-length strings in the case of subtree crossover [18] and homologous crossover [19], as well as an extension, in the case of fixed length strings, to the case of a non-flat landscape with weak selection [9].

Acknowledgements. CRS would like to thank Riccardo Poli and Peter Stadler for very fruitful collaborations and acknowledges support from DGAPA project ES100201 and Conacyt project 30422-E. AZ would like to thank the Instituto de Ciencias Nucleares for their hospitality.

References

1. Michael D. Vose. *The simple genetic algorithm: Foundations and theory.* MIT Press, Cambridge, MA, 1999.
2. David E. Goldberg. *The Design of Innovation.* Kluwer Academic Publishers, Boston, 2002.
3. J. Holland. *Adaptation in Natural and Artificial Systems.* University of Michigan Press, Ann Arbor, USA, 1975.
4. A.E. Nix and M.D. Vose. Modeling genetic algorithms with Markov chains. *Annals of Mathematics and Artificial Intelligence*, 5(1):79–88, April 1992.
5. Adam Prügel-Bennett and Jonathan L. Shapiro. An analysis of genetic algorithms using statistical mechanics. *Physical Review Letters*, 72:1305–1309, 1994.
6. David E. Goldberg. *Genetic Algorithms in Search, Optimization and Machine Learning.* Addison Wesley, Reading MA, 1989.
7. C.R. Stephens and H. Waelbroeck. Effective degrees of freedom of genetic algorithms and the block hypothesis. In T. Bäck, editor, *Proceedings of ICGA97*, pages 31–41, San Francisco, CA, 1997. Morgan Kaufmann.
8. C.R. Stephens and H. Waelbroeck. Schemata evolution and building blocks. *Evol. Comp.*, 7:109–124, 1999.
9. C.R. Stephens. Some exact results from a coarse grained formulation of genetic dynamics. In *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO-2001)*, pages 631–638, San Francisco, California, USA, 7-11 July 2001. Morgan Kaufmann.
10. Riccardo Poli. Exact schema theorem and effective fitness for GP with one-point crossover. In *Proceedings of the Genetic and Evolutionary Computation Conference*, pages 469–476, Las Vegas, July 2000. Morgan Kaufmann.
11. Riccardo Poli. Exact schema theory for genetic programming and variable-length genetic algorithms with one-point crossover. *Genetic Programming and Evolvable Machines*, 2(2):123–163, 2001.
12. W.B. Langdon and R. Poli. *Foundations of Genetic Programming.* Springer Verlag, Berlin, New York, 2002.
13. Peter Nordin. *Evolutionary Program Induction of Binary Machine Code and its Applications.* PhD thesis, der Universitat Dortmund am Fachereich Informatik, 1997.
14. Michael O'Neill and Conor Ryan. Grammatical evolution. *IEEE Transaction on Evolutionary Compuation*, 5(4), August 2001.
15. A. Aguilar, J. Rowe, and C.R. Stephens. Coarse graining in genetic algorithms: Some issues and examples. In E. Cantú-Paz, editor, *Proceedings of GECCO 2003*, Chicago, 9-13 July 2003. Springer-Verlag.
16. Edward D. Weinberger. Fourier and Taylor series on fitness landscapes. *Biological Cybernetics*, 65:321–330, 1991.
17. C.R. Stephens. The renormalization group and the dynamics of genetic systems. *Acta Phys. Slov.*, 52:515–524, 2002.
18. R. Poli, J.E. Rowe, C.R. Stephens, and A.H. Wright. Allele diffusion in linear genetic programming and variable-length genetic algorithms with subtree crossover. In *Proceedings of EuroGP 2002, LNCS, Vol. 2278*, pages 211–227. Springer-Verlag, 2002.
19. R. Poli, C.R. Stephens, J.E. Rowe, and A.H. Wright. On the search biases of homologous crossover in linear genetic programming and variable-length genetic algorithms. In *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO-2002)*, pages 211–227. Morgan Kaufmann, 2002.