CHAPTER 4. THE RENORMALIZATION GROUP IN GENETICS

This final chapter intends to give a brief introduction to the use of operators of a genetic system based on bit strings; in particular shows how crossover is distinguished in a 'building block basis' which determines a hierarchical solution to the genetic dynamics, with a reference to the renormalization group.

4.1 Dynamical equations

An illustration of how schemata may benefit the ease of a particular computation within the binary string sets is in the case of the evolution of a string I of fixed length ℓ by means of crossover events with a probability p_c. For such a string, the dynamical equation stating the relation of recurrence for its population P_I is

$$P_I(t+1) = (1-p_c)P_I(t) + p_c \sum_n p_c(m_n)\, \lambda_I^{JK}(m_n) P_J(t) P_K(t) \tag{69}$$

where the first contribution relates to no cross-over at time t in P_I which preserves the string a time later, and the second contribution considers all the possible ways in which P_I may be created from any string parents P_J and P_K, by summing over all the incurring n *crossover masks*.

A crossover mask m for I is another binary string of length ℓ which now dictates the rules for recombination: whenever a 1 is found in any locus of the string, the correspondent attribute in the first (or second) parent will be assigned to the offspring; finding a 0 indicates that the allele in question shall be taken from the second (or first) parent. In general, if a probability $p_c(m)$ is a priori specified to every allowed mask, any homologous crossover operator can be defined[16].

$\lambda_I^{JK}(n)$ recognises whether the transformation correspondent to a mask m_n and string parents P_J and P_K lead to P_I.

For simplicity, in a string of $\ell = 2$, in order to create string "11" the only masks possible in the previous definition are either 01 (first bit from first parent, second bit from second parent) or 10 (vice versa). A probability of ½ for each mask to be implemented is assumed. A diagram to reach 11 from a population of string parents at time t would look like in figure (19).

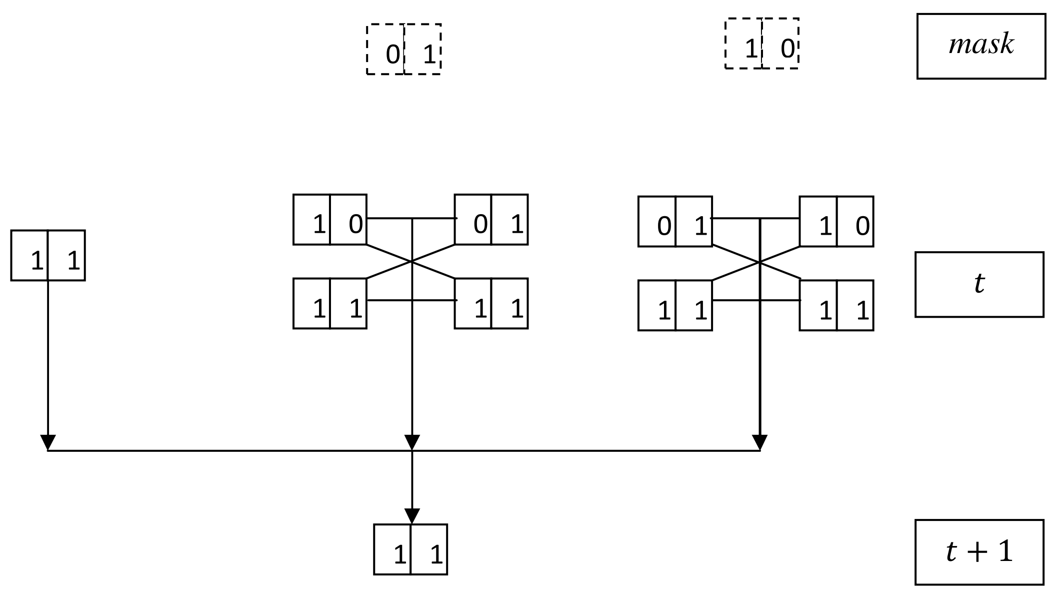

Figure 19. Ways to attain the $\ell = 2$ string 11 by cross-over from parent strings.

Equation (69) includes every combination involved:

$$P_{11}(t+1) = (1-p_c)P_{11}(t) + p_c \cdot \frac{1}{2}\left[\left(P_{10}(t)\cdot P_{01}(t) + P_{10}(t)\cdot P_{11}(t) + P_{11}(t)\cdot P_{01}(t) + P_{11}(t)\cdot P_{11}(t)\right) + \left(P_{01}(t)\cdot P_{10}(t) + P_{01}(t)\cdot P_{11}(t) + P_{11}(t)\cdot P_{10}(t) + P_{11}(t)\cdot P_{11}(t)\right)\right]$$

$$= (1-p_c)P_{11}(t) + p_c \cdot \left\{\frac{1}{2}\left[\left(P_{10}(t)\cdot P_{01}(t) + P_{10}(t)\cdot P_{11}(t) + P_{11}(t)\cdot P_{01}(t)\right) + \left(P_{01}(t)\cdot P_{10}(t) + P_{01}(t)\cdot P_{11}(t) + P_{11}(t)\cdot P_{10}(t)\right)\right] + P_{11}(t)\cdot P_{11}(t)\right\} \tag{70}$$

As this is the simplest case for recombination-only driven evolution, it is clear that for ‘real life’ situations involving a large number of loci/attributes/bits (along with an even larger number of masks), the calculation needed to describe one single string becomes difficult even computationally.

A building block defined on the basis of schemata groups different strings as one single schema, and it can be shown that it is equivalently subject to the dynamical equations of recombination as the former strings were without the loss of any information of relevance to the process. This is a consequence of noting that for a mask 01 it is unnecessary to state what the first parent’s second bit is, and so on. A simplification is shown in figure 20.

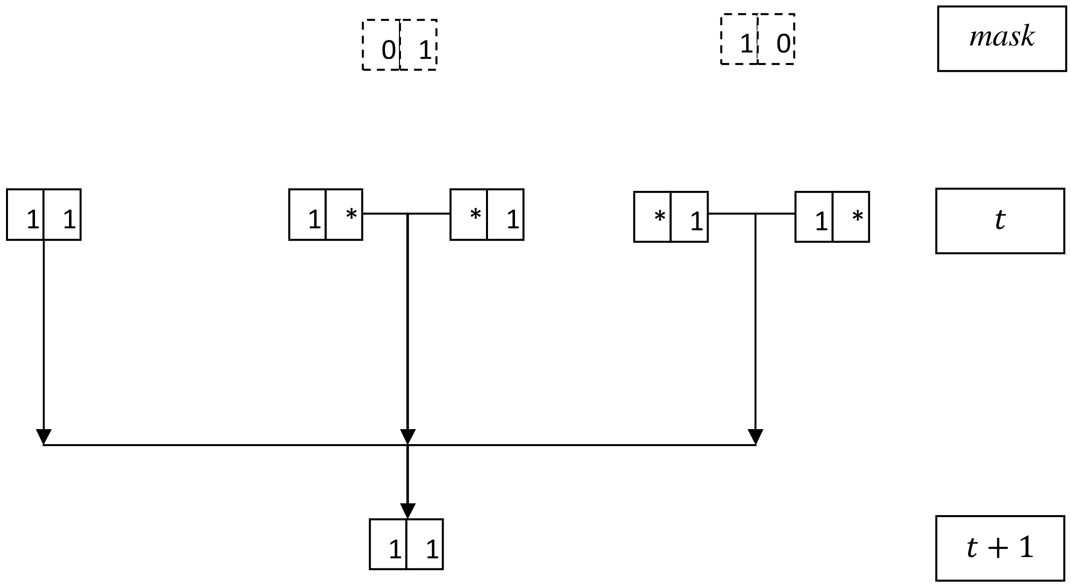

Figure 20. Ways to reach $\ell = 2$ string 11 by cross-over from constituent building blocks.

Now, it is possible to re-write equation (70), including all of the combinations involved in an effective building block representation:

$$P_{11}(t+1) = (1-p_c)P_{11}(t) + p_c \cdot \frac{1}{2}\left[\left(P_{1*}(t) \cdot P_{*1}(t)\right) + \left(P_{*1}(t) \cdot P_{1*}(t)\right)\right] \tag{71}$$

As eight original contributions have now been reduced to three.

4.2 From genotypes to building blocks

The study of genetic dynamics involves the evolution of populations of strings by operators of selection, mutation and recombination. We previously mentioned them when discussing the Holland schema theorem –equation (24)- applied to schemata, where each contribution of the equation could be clearly separated. Now, for strings we rewrite as:

Selection only	Mutation only[6]	Recombination only
$\frac{f_i(t)}{\bar{f}(t)} P_I(t)$	${p_m}^{\delta_{IJ}^H}(1-p_m)^{\ell-\delta_{IJ}^H} P_J(t)$	$(1-p_c)P_I(t) + p_c \sum_n p_c(m_n)\, \lambda_I^{JK}(m_n) P_J(t) P_K(t)$

Table 1. How to assort $P_I(t+1)$from the population of strings I, J and K a time earlier, under the referred regime.

For the case of selection and mutation operators, these have linear equations that can be relatively easy to map into two-dimensional statistical-mechanics problems wherein techniques such as the transfer matrix may be of use[32]. As a consequence of binary input (or more in general) recombination involves at least one nonlinear contribution. However, a simpler representation is made under the observation that much of the information of the parents is 'lost', that is, useless to the production of I in crossover. Schemata, as shown in the previous sub-section, are particularly good when trying to reduce the number of contributions accountable towards the objective string.

More formally, from a population $P(t) = \{c_i(t)\}$ of strings from a configuration space $\mathcal{G}$, where $c_i(t)$ is a set of genotypes (formerly attributes) that may be present in P at a time t (in our previous example 11, 10, 01 and 00); the transformation $\Lambda: \mathcal{G} \longrightarrow \tilde{\mathcal{G}}$ to a set consisting of schemata of $c_i(t)$ with elements of $\Lambda_{ij} = [c_j \in \xi_i]$ [7] has an associate *building*

[6] δ_{IJ}^H is the Hamming distance between strings I and J, the number of positions for which attributes differ. For schemata we considered only their order, the number of cases that with probability p_m could destruct the schema due to mutation.
[7] The Iverson bracket.

block basis, which can be thought of as representing the natural effective degrees of freedom of crossover.

Such can be equivalent to “block spins” of the renormalization group techniques[32] understood in terms of a coarse graining operator $\mathcal{R}(\eta,\eta')P(\eta,t) = P(\eta',t)$ from a variable η to another η'. As a result of

$$\mathcal{R}(\eta',\eta'')P(\eta',t) = P(\eta'',t) \tag{72}$$

we can write

$$\mathcal{R}(\eta,\eta'') = \mathcal{R}(\eta,\eta')\mathcal{R}(\eta',\eta'') \tag{73}$$

Or in general, as result of its iterative structure, we obtain the recurrence relation[32,16].

$$\mathcal{R}^{(n+1)}\left(\eta,\eta^{(n+1)}\right) = \mathcal{R}\left(\eta^{(n)},\eta^{(n+1)}\right)\mathcal{R}^{(n)}\left(\eta,\eta^{(n)}\right) \tag{74}$$

In addition, following that masks are intrinsic to the crossover operator, an application to the renormalization group techniques is the following: the selection of spins to be ‘deleted’ during a single iteration to obtain the renormalization group equations is analogous to the use of a crossover mask which implements a coarse graining in a bit chain.

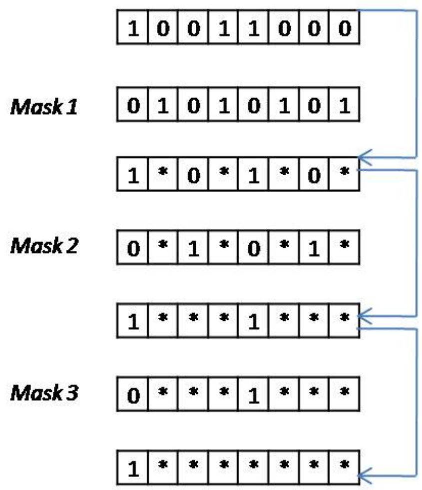

Figure 21. Just as with the renormalization group we chose which elements to be set aside, the selection of crossover masks tells how strings will be counted when deleting even loci in the string.

As seen earlier, the building block basis can serve to simplify the equations as it gives new effective variables for recombination, or degrees of freedom. We will show in the next subsection a brief example on how it serves to decouple or at least help to solve the system in a hierarchy.

4.3 Hierarchy dynamics

Equation (71) provides the description of the dynamics for the case of string 11 under single-point crossover, as a function of its building blocks: P_{1*} and P_{*1}. These are just:

$$P_{1*}(t+1) = (1-p_c)P_{1*}(t) + p_c\big[\big(P_{1*}(t)\cdot P_{**}(t)\big)\big] = P_{1*}(t) \qquad (75.a)$$

analogously,

$$P_{*1}(t+1) = P_{*1}(t) \tag{75.b}$$

Being order-one schemata, their values at a time t are constant as these do not restrain to recombination. Also, $P_{**} = 1$ at every time, hence we give the alternate representation of the system:

$$\begin{pmatrix} P_{11} \\ P_{1*} \\ P_{*1} \\ P_{**} \end{pmatrix}_{t+1} = \begin{pmatrix} 1-p_c & \frac{p_c}{2}P_{*1} & \frac{p_c}{2}P_{1*} & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} P_{11} \\ P_{1*} \\ P_{*1} \\ P_{**} \end{pmatrix}_{t} \tag{76}$$

Note the structure of an upper triangular matrix which suggests the hierarchical path to solve for P_{11} in its building block basis. For the case where $\ell = 3$ we evaluate:

$$P_{111}(t+1) = (1-p_c)P_{111}(t) + p_c \cdot \tfrac{1}{2}\left[\left(P_{11*}(t) \cdot P_{**1}(t)\right) + \left(P_{1**}(t) \cdot P_{*11}(t)\right)\right] \tag{77}$$

With the sources being,

$$P_{11*}(t+1) = (1-p_c)P_{11*}(t) + p_c \cdot \tfrac{1}{2}\left[\left(P_{11*}(t) \cdot P_{***}(t)\right) + \left(P_{1**}(t) \cdot P_{*1*}(t)\right)\right]$$

$$= \left(1 - \tfrac{p_c}{2}\right) P_{11*}(t) + \tfrac{p_c}{2} P_{1**}(t) \cdot P_{*1*}(t) \tag{78.a}$$

$$P_{*11}(t+1) = \left(1 - \tfrac{p_c}{2}\right) P_{*11}(t) + \tfrac{p_c}{2} P_{*1*}(t) \cdot P_{**1}(t) \tag{78.b}$$

Analogously, the equations for order-one building blocks are linear and independent of t. The evolution of three loci and single point crossover is shown by matrix:

$$\begin{pmatrix} P_{111} \\ P_{11*} \\ P_{1*1} \\ P_{*11} \\ P_{1**} \\ P_{*1*} \\ P_{**1} \\ P_{***} \end{pmatrix}_{T+1} = \begin{pmatrix} 1-p_c & \frac{p_c}{4}P_{**1} & 0 & \frac{p_c}{4}P_{1**} & \frac{p_c}{4}P_{*11} & 0 & \frac{p_c}{4}P_{11*} & 0 \\ 0 & 1-\frac{p_c}{2} & 0 & 0 & \frac{p_c}{4}P_{*1*} & \frac{p_c}{4}P_{1**} & 0 & 0 \\ 0 & 0 & 1-\frac{p_c}{2} & 0 & \frac{p_c}{4}P_{**1} & 0 & \frac{p_c}{4}P_{1**} & 0 \\ 0 & 0 & 0 & 1-\frac{p_c}{2} & 0 & \frac{p_c}{4}P_{**1} & \frac{p_c}{4}P_{*1*} & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} P_{111} \\ P_{11*} \\ P_{1*1} \\ P_{*11} \\ P_{1**} \\ P_{*1*} \\ P_{**1} \\ P_{***} \end{pmatrix}_{T}$$

Again the triangular form emphasises that the solution for coarser blocks may be found after the solutions for lower order blocks are known, in a clear analogy with the counters example of chapter two –see equation (35)-. The building block basis here shows that the crossover operator is hierarchical as the evolution of schemata of decreasing order is known first. In general, $2^{\ell}-1$ coupled quadratic equations can be turned into uncoupled linear equations in this basis.