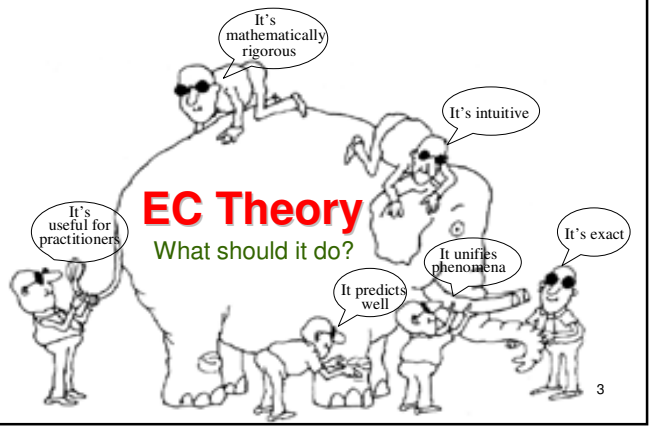
Taxonomy and Coarse Graining in Evolutionary Computation

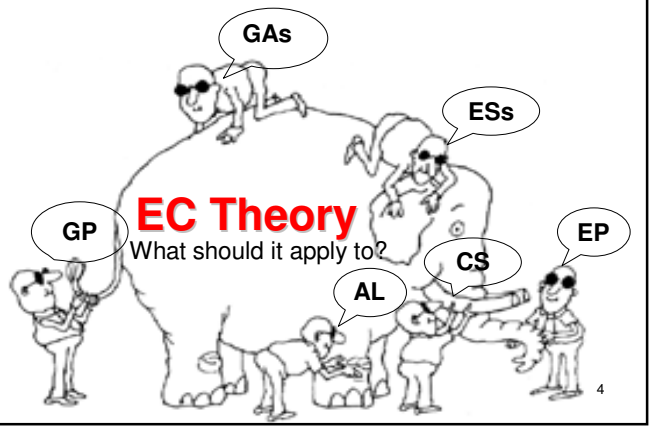
Chris Stephens,
Instituto de Ciencias Nucleares, UNAM
and
Department of Computer Science, University of Essex
Tutorial GECCO2005, 28/06/2004
stephens@nucleares.unam.mx

Taxonomy

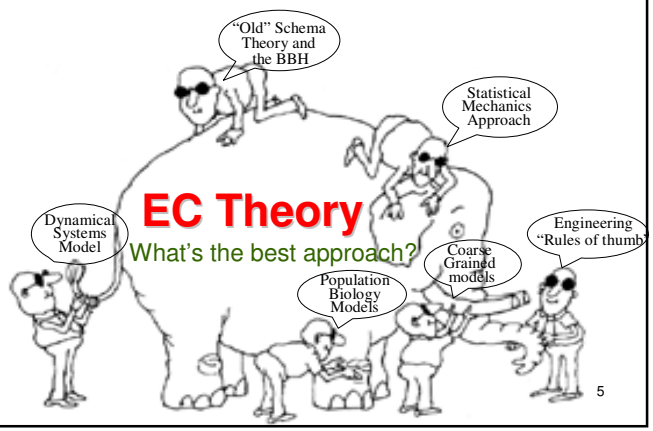
The Problem of Taxonomy…



The Problem of Taxonomy…



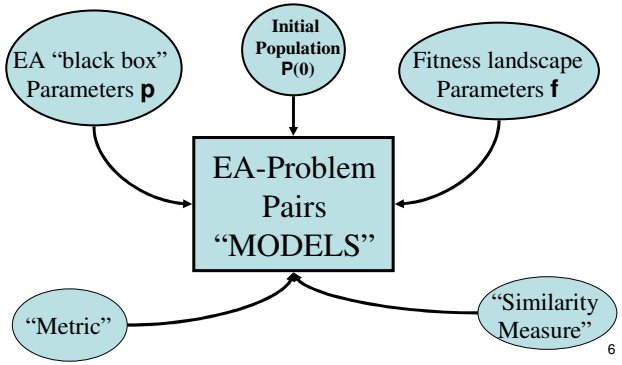
The Problem of Taxonomy…



5

The Problem of Taxonomy…

What's a useful Taxonomy for algorithms, fitness landscapes (i.e. "problems") …?



6

The Space of EAs

How "far apart" are a GA with one-point crossover, pc=0.8, mutation, p=0.08, and a NK fitness landscape, N=27, K=3 and GP for K-SAT, K=4, sub-tree crossover, pc=0.5, mutation, p=0.05?

How "far apart" are a giraffe and a grasshopper?

How "far apart" are hydrogen and uranium?

How "far apart" are a GA with one-point crossover, pc=0.8, mutation, p=0.08, and a NK fitness landscape, N=27, K=3 and a GA with one-point crossover, pc=1, mutation, p=0.1, and a NK fitness landscape, N=35, K=3?

How "far apart" are a giraffe and a horse?

How "far apart" are sodium and potassium?

Taxonomy is easier with "distance" measures

7

Taxonomy

Phenomenology – we want more of that! e.g. "periodic table"

Theory – what can it tell us? E.g. "electronic structure"

History – contingency, that's what we've had

Universality

8

Phenomenology

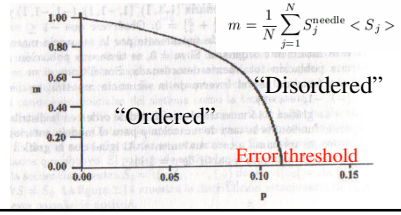
The first prerequisite for a successful Taxonomy

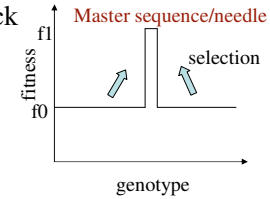
9

Universality – when is the devil in the details?

Eigen model/Needle-in-a-haystack
Characteristic of viruses and real world
BRITTLE problems (it works or it doesn't!)

Qualitative behavior dominated
by existence of error threshold –
doesn't depend on "details" - **UNIVERSAL**



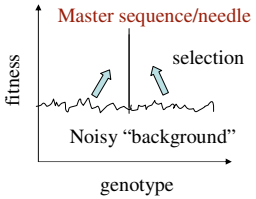


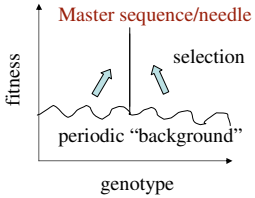
Value of critical mutation
rate does depend on details
(N, f1, f0 …)
– **NON-UNIVERSAL**

10

Universality – when is the devil in the details?

Same **UNIVERSALITY CLASS**
as NIAH

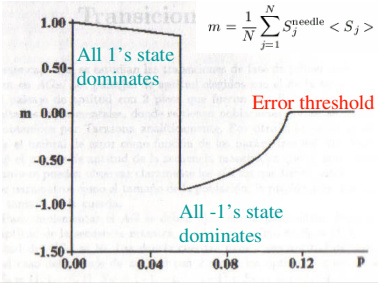


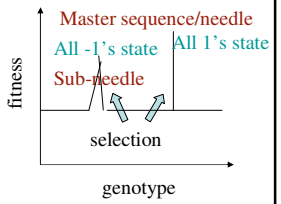


11

Universality – when is the devil in the details?

NOT in the same **UNIVERSALITY CLASS** as NIAH!





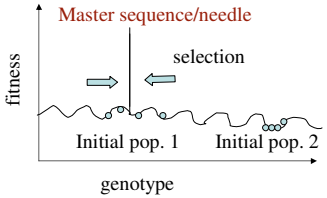
Corresponds to a system where there's
several "it kind of works" states as
well as a "it definitely works" state

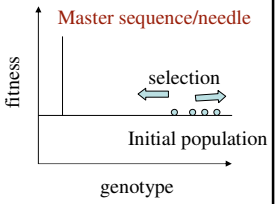
12

Universality – when is the devil in the details?

Same **UNIVERSALITY CLASS** as NIAH? YES

What typically happens?





Universality – when is the devil in the details?

Phase transition for K-SAT

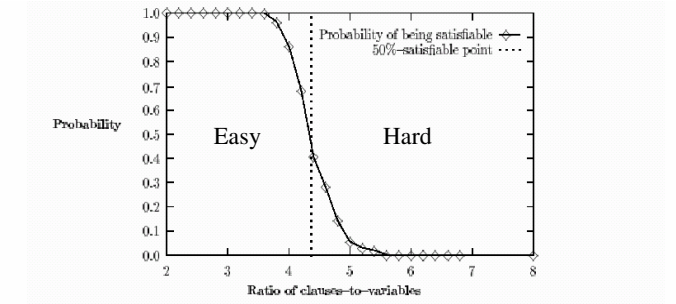


Figure 4: Probability of satisfiability of 50–variable formulas, as a function of the ratio of clauses-to-variables.

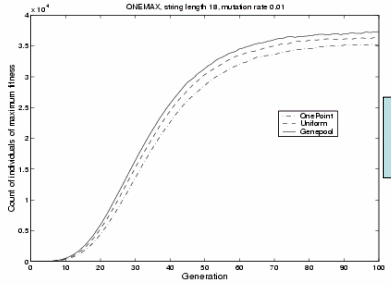
From: Mitchell et al.(1992) Hard and easy distributions of SAT problems

Similarity measures

Need objective criteria by which to judge the degree of affinity between different models. There are many possibilities…e.g.

- Average population fitness vs. time
- Best in population versus time
- Diversity versus time
- "Order parameter" (e.g. % of population that is optimal as function of EA parameters)
- Time to find optimum
- "Hardness"
- "Robustness"
- Fixed points (asymptotic behavior)

Example of a similarity measure



< 10% difference between One-point and genepool

"Model"/"Toy" problems tell us much more than you think!

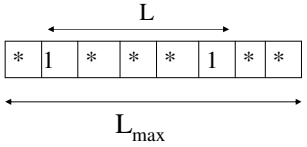
Figure 2: The number of optimal individuals for different types of recombination

Wright, Rowe, Poli and Stephens – GECCO2002

Example of a similarity measure

Stephens, Waelbroeck and Aguirre – FOGA 7

$M(L) = (n_{opt}(L) - n_{opt}(L_{max}))/ n_{opt}(L_{max})$



$n_{opt}(L)$ = no. of optimal 2-schemata/total no. possible per string

Interested in whether short or long building blocks are preferred. M(L) > 0 ⇨ preference for short blocks, M(L) < 0 ⇨ preference for long blocks

Example of a similarity measure

Experiment: popsize = 5,000; L_{max} = 8; 30 runs

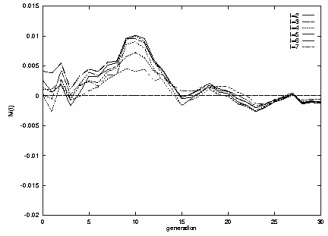


Figure 1: Graph of $M(l)$ versus t in the unitation model with $p_c = 0$.

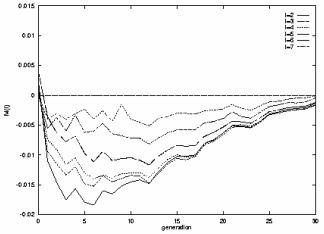


Figure 2: Graph of $M(l)$ versus t in unitation model with $p_c = 1$.

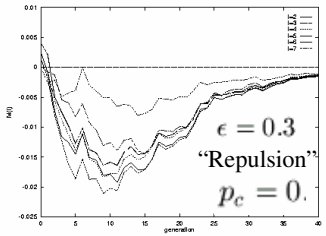
Without crossover – no preference for one size versus another

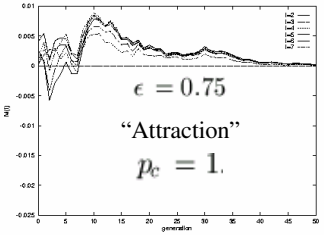
With crossover – large schemata grow, short schemata diminish – opposite of Building Block Hypothesis

Example of a similarity measure

$$f(c_i) = \sum_j 1_j + \frac{\epsilon}{N^{\pm}} \sum_{jk \in C_i} l_{jk}^{\pm 1}$$

Add pair epistasis: + →repulsion; - →attraction





Results for $p_c = 1.$ with no epistasis are similar to those with $p_c = 0.$ and an epistatic repulsion between bits

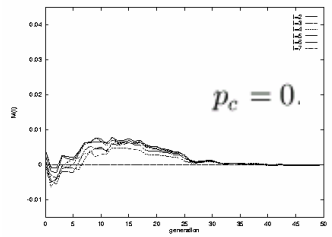
Results for $p_c = 0.$ with no epistasis are similar to those with $p_c = 1.$ and an epistatic attraction between bits

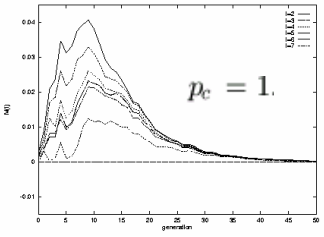
UNIVERSALITY

Example of a similarity measure

"Deceptive" landscape

$f(11) = 3,\ f(01) = f(10) = 1,\ f(00) = 2;$ for every pair





Without crossover - no preference for one size versus another

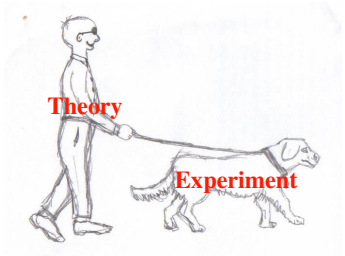
With crossover – short blocks preferred

Theory

Provides a mathematical framework in which
to develop and understand a Taxonomy

21

The Problem of Theory…



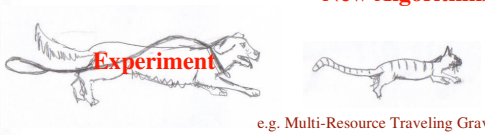
The "ideal"

22

The Problem of Theory…

In EC …

New Applications
New Algorithms



e.g. Multi-Resource Traveling Gravedigger Problem with Variable Coffin Size

"Most algorithms are NEVER used (except by the people who created them)" - Darrell Whitley, GECCO 2003 tutorial

23

The Problem of Theory…

The EC Expectation Gap

What theoreticians think practitioners are and what practitioners think theoreticians should be

What practitioners think theoreticians are and what theoreticians think practitioners should be

24

The Problem of Theory…

"Professors in every branch of the sciences prefer their own theories to truth; the reason is that their theories are private property, but truth is common stock" – Charles Caleb Colton, Lacon (1825).

"It is the nature of an hypothesis, when once a man has conceived it, that is assimilates everything to itself as proper nourishment, and, from the first moment of your begetting it, it generally grows the stronger by everything you see, hear, read, or understand" – Laurence Sterne, Tristram Shandy (1767).

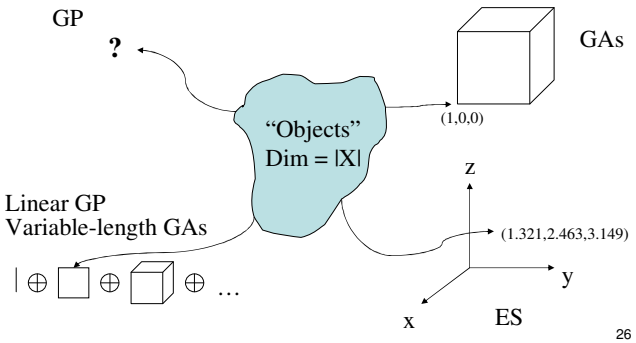
"EC theory is hard!" - Chris Stephens (most weeks).

The EC "Expectation Gap" ⟷ The Population Biology "Lesson"

"How does this help practitioners…?" – most referees

25

EC Theory – the "Bare Necessities"

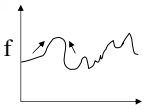


26

EC Theory – the "Bare Necessities"

Objects have fitness: $f_X : X \rightarrow R^+$

Objects have interactions:

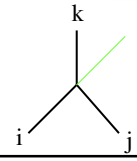


$Object_i \rightarrow Object_i$ P'_i Selection

$Object_i \longleftrightarrow Object_j$ $\mathcal{P}_{ij}$ Mutation

$Object_i + Object_j \longleftrightarrow Object_k + Object_l$ Recombination

m – recombination "mode"



$$\mathbf{P}(t+1) = \mathcal{H}(\mathbf{p}, \mathbf{f}, \mathbf{P}(t))$$

Dynamics

27

EC - the "Microscopic" Theory

$$P_I(t+1) = \sum_J M_I{}^J \left((1-p_c)P'_J + p_c \sum_M \sum_{KL} \frac{1}{2}\left(p(M) + p(\bar{M})\right) \lambda_J{}^{KL}(M) P'_K P'_L \right)$$

$P_I(t+1)$ - Probability to find "object" I

P'_J - Probability to select "object" J

$M_I{}^J$ - Probability to mutate "object" J to "object" I

$p(M)$ - Conditional probability for recombination mask/mode M

p_c - Probability to implement recombination

$\lambda_J{}^{KL}(M)$ - Conditional probability that given "objects" K and L and mode M "object" J is created (= 0, 1).

Sums are over all possible recombination modes and all objects J and K. e.g. for GA and homologous crossover 2^{3N} terms

28

EC – the "Microscopic" Theory

$$P_I(t+1) = \sum_J M_I{}^J \left((1-p_c)P'_J + p_c \sum_M \sum_{KL} \frac{1}{2}\left(p(M) + p(\bar{M})\right) \lambda_J{}^{KL}(M) P'_K P'_L \right)$$

Probability that "object" J is mutated to "object" I

Probability that "object" J is cloned

Probability that parent "objects" K and L are selected and "mixed" to form child "object" J via mode M

This is an extension of the Fisher/Wright model of population genetics and equivalent to the dynamical systems/Vose model in the case of GAs

Here it's written in a form that emphasizes the different explicit processes that can occur to form a genotype I

EC – the "Microscopic" Theory

These are 2^N coupled, non-linear, first order difference equations. The underlying mathematics is conceptually quite simple. Understanding what the equations can and are telling us isn't!

So, what can we do with them?

First thing – simplify by considering only a reduced set of genetic operators

EC – the "Microscopic" Theory Selection Only

$P_I(t+1) = \frac{f_I}{\bar{f}(t)} P_I(t)$ The average fitness is the problem – contains contributions from all strings

So, change variables $P_I(t) = \dfrac{x_I(t)}{\sum_I x_I(t)}$

$x_I(t+1) = f_I x_I(t)$ $\qquad x_I(t) = f_I^t x_I(0)$

Exact solution for selection only (infinite population)

$$P_I(t) = \frac{f_I^t P_I(0)}{\sum_I f_I^t P_I(0)}$$

Need to have full landscape (up to 2^Nvalues) to predict the dynamics exactly

EC – the "Microscopic" Theory

How do we go beyond the selection only case?

For mutation-selection – **conceptually** straightforward – **formally** can find the eigenvalues and eigenvectors of **W = MF**

Construction due to crossover is more difficult! For instance, what are the "regularities" of $\lambda_J{}^{KL}(M)$

Brydges & Goldberg
Whitley

Example: 2-bit GA with $p(M) = 1/4$ for all M, $I = (11)$

$\lambda_{(11)}((00)) =$ $\begin{pmatrix} 0&0&0&0 \\ 0&0&0&0 \\ 0&0&0&0 \\ 1&1&1&1 \end{pmatrix}$ represents $\boxed{0\,0}\boxed{1\,1}$ ⇓ $\boxed{1\,1}$

$\lambda_{(11)}((01)) =$ $\begin{pmatrix} 0&0&0&0 \\ 0&0&0&0 \\ 0&1&0&1 \\ 0&1&0&1 \end{pmatrix}$ represents $\boxed{0\,1}\boxed{1\,0}$ ⇓ $\boxed{1\,1}$

Ugly!

EC – the "Microscopic" Theory String construction

$$\sum_{J,K=00,\ 01,\ 10,\ 11} \lambda_{11}{}^{JK}(10) = (P_{00} \quad P_{01} \quad P_{10} \quad P_{11}) \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} P_{00} \\ P_{01} \\ P_{10} \\ P_{11} \end{pmatrix}$$

$$= P_{10}P_{01} + P_{10}P_{11} + P_{11}P_{01} + P_{11}P_{11}$$

For N bits there are $2^N \times 2^N \times 2^N \times 2^N$ possible terms

Target strings; masks; First parents; Second parents

So, how do we see what recombination is really doing?

EC – the "Microscopic" Theory

Can iterate the equations and represent the solution graphically -

$$P_I(t) = \sum_J \left|\begin{matrix} I & t \\ J & t{=}0 \end{matrix}\right| + \sum_{JKL} \sum_M \sum_{n=0}^{t} \left|\begin{matrix} I & t \\ J & \\ K & L \end{matrix}\right| \quad t = n$$

Term exclusively due to constructive effect of recombination

$\left|\begin{matrix} I & t \\ J & t' \end{matrix}\right| = G_{IJ}(t,t')$ Probability that object J propagates from t to t' and converts to I on the way

$\bigcirc = \frac{1}{2}(p(M) + p(\bar{M}))\lambda_J{}^{KL}(M)\frac{f_K}{\bar{f}(t)}\frac{f_L}{\bar{f}(t)}$ Measures strength of interaction between objects J, K and L

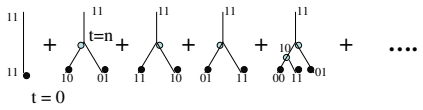
$I \bullet = P_I(t)$

EC – the "Microscopic" Theory

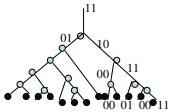
Iterate … by **recursively** substituting for • until get to t = 0

Example – 2-bits, 1-point crossover

Process where 11 is formed by crossover of 10 and 01 at t = n for any n

$P_{11}(t) =$ 

$+ \ldots.$



$+ \ldots.$

Each tree tells us the probability of forming 11 by a given **process**. In principle can see which processes are most important. But … tree depth bounded only by t! COMPLICATED! TOO MANY DIAGRAMS (PROCESSES)

Diagrams are pictorial representations of the different processes that can occur

EC – the "Microscopic" Theory

General (Feynman) rules – including selection and mutation too:

1) Draw all possible tree diagrams that contribute to creation of "object"
2) For each internal line ———— attach a propagator

$$G_{IJ}(t,t') = (1-p_c)^{t-t'}\frac{(\mathbf{FM})_{IJ}^{t-t'}}{\sum_I (\mathbf{FM})_{IJ}^{t-t'} P_J(t')}$$

3) To each vertex ⬭ attach a weight

$$\frac{1}{2}(p(M) + p(\bar{M}))\lambda_J{}^{KL}(M)\frac{f_K}{\bar{f}(t)}\frac{f_L}{\bar{f}(t)}$$

4) To each root • attach a factor $P_I(t')$
5) Carry out integration over time for all vertices

These rules, analogous to what is used in quantum field theory, generate the algebraic expressions that describe the different probabilistic processes that can occur in the dynamics of an EA

EC – the "Microscopic" Theory

So what do we have so far from the "microscopic" theory?

Exact	Yes
Mathematically rigorous	Yes ?
Unifies Phenomena	Yes/No
Intuitive	No
Predicts well	No
Useful for Practitioners	No

37

EC Theory – the "old stuff"

Let's compare with the old Schema theorem and Building Block Hypothesis approach

Exact	No
Mathematically rigorous	Yes ??
Unifies Phenomena	Yes/No
Intuitive	Yes/No
Predicts well	No ?
Useful for Practitioners	Yes/No

38

Coarse Graining

39

Coarse Graining

Why?

What?

How?

40

Coarse Graining

Why?

1. **Emergence of "Effective Degrees of Freedom" (EDOF)/Collectivity/Universality**
2. **Curse of dimensionality/intractable dynamics**

Coarse-grained degrees of freedom are combinations of the underlying "microscopic" degrees of freedom. EDOF are those coarse-grained degrees of freedom that are important for the dynamics

41

Coarse Graining

What?

1. "Direct" dimensional reduction
2. Phenotypes
3. Schemata
4. Hyperschemata
5. Building Blocks
6. Lowest cumulants of fitness distribution
7. "Normal (e.g. Walsh) modes"
8. Others

What is the most natural coarse graining depends on the operators and their corresponding parameters, the fitness landscape and the population.

42

Coarse Graining

How?

1. Phenotype Dynamics
2. Schemata Dynamics
3. Hyperschemata Dynamics
4. Building Block Dynamics
5. Aggregation of Markov chain
6. Truncation of cumulants
7. Walsh analysis
8. Others

Is it exact?

43

Coarse Graining By Coordinate Transformations

Identifying "Effective Degrees of Freedom" by transforming to the most appropriate coordinate system

44

Coarse graining via Coordinate Transformations

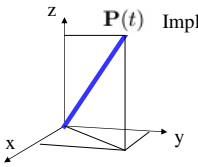
The fundamental equation for "everything" written like this…

$$P_I(t+1) = \sum_J M_I{}^J \left((1-p_c)P'_J + p_c \sum_M \sum_{KL} \frac{1}{2}\left(p(M)+p(\bar{M})\right)\lambda_J{}^{KL}(M) P'_K P'_L \right)$$

is COVARIANT, i.e. has the same content in ANY coordinate system

$\mathbf{P}(t)$ is INVARIANT in any coordinate system, its components

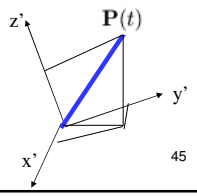
$P_I(t)$ however, do change



Implemented by rotation matrix

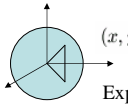
$$P_I = \sum_J \Lambda_I{}^J P_J$$

Coordinate transformation Implemented linearly



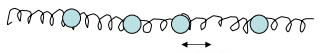
Coarse graining via Coordinate Transformations

- Appropriate choice of coordinate system can make manifest the **Effective Degrees of Freedom** and greatly facilitate calculations



$(x,y,z) \to (R,\theta,\phi)$

Exploit spherical symmetry



$u(k) = \frac{1}{\sqrt{2}} \sum_{x_i} \exp(ikx_i) u(x_i)$

Normal modes - waves

Coordinate system used up to now is the "object" system – e.g. strings, trees etc. ⇒ OK when EDOF are strings, trees etc.

Appropriate in "strong" selection regime

But, for instance, mutation matrix non-diagonal in object/string basis

Coarse graining via Coordinate Transformations

Mutation …

Transform coordinates to **Walsh basis** (for fixed length binary strings) using transformation matrix $\mathbf{\Lambda^w}$ with elements

bitcount

$$\Lambda^w{}_I{}^J = \frac{1}{2^{N/2}}(-1)^{\#(I\otimes J)}$$

then $$M_I{}^J \to \sum_{I'J'} \Lambda^w{}_I{}^{J'} M_{I'}{}^{J'} \Lambda^w{}_{J'}{}^J = (1-2p)^{|I|}\delta_I{}^J$$

and $$P_I(t) \to \hat{P}_I(t) = \sum_{I'} \Lambda^w{}_I{}^{I'} P_{I'}(t)$$

Gives as solution $\hat{P}_I(t) = (1-2p)^{|I|}\hat{P}_I(0)$ Exact solution for mutation only (infinite population)

Higher order Walsh modes decay more quickly. Fixed point is $\hat{P}^*_0 = 1/2^{N/2}$ and corresponds to a random population

Coarse graining via Coordinate Transformations

Mutation …

For example, for two bits

$$\mathbf{\Lambda^w} = \frac{1}{2}\begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{pmatrix}$$

$$\mathbf{M} = \begin{pmatrix} (1-p)^2 & p(1-p) & p(1-p) & p^2 \\ p(1-p) & (1-p)^2 & p^2 & p(1-p) \\ p(1-p) & p^2 & (1-p)^2 & p(1-p) \\ p^2 & p(1-p) & p(1-p) & (1-p)^2 \end{pmatrix} \to \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & (1-2p) & 0 & 0 \\ 0 & 0 & (1-2p) & 0 \\ 0 & 0 & 0 & (1-2p)^2 \end{pmatrix}$$

"Frequencies" of "normal modes"

Coarse graining via Coordinate Transformations

Mutation …

In Walsh basis …

- Mutation matrix is diagonal
- Selection matrix is non-diagonal
- Crossover – O(n) Walsh coefficients made up from crossing O(m) and O(n-m) coefficients
- "Normal modes" not simply interpretable
- Useful for landscape analysis
- **Gives exact solution for mutation only**

Appropriate in "strong" mutation regime

Coarse graining via Coordinate Transformations

Mutation and Selection…

For mutation EDOF are discrete versions of normal modes

But … selection is non-diagonal.
In Walsh basis selection matrix is

$$\begin{pmatrix} \hat{F}_{00\,00} & \hat{F}_{00\,01} & \hat{F}_{00\,10} & \hat{F}_{00\,11} \\ \hat{F}_{01\,00} & \hat{F}_{01\,01} & \hat{F}_{01\,10} & \hat{F}_{01\,11} \\ \hat{F}_{10\,00} & \hat{F}_{10\,01} & \hat{F}_{10\,10} & \hat{F}_{10\,11} \\ \hat{F}_{11\,00} & \hat{F}_{11\,01} & \hat{F}_{11\,10} & \hat{F}_{11\,11} \end{pmatrix}$$

In principle, can find solution by linear coordinate transformation $\tilde{P}_I(t) = \sum_J \tilde{\Lambda}_I{}^J P_J(t)$

which diagonalizes the mutation-selection matrix **W = MF**

Such that $W \to \tilde{W}$ with elements $\lambda_I \delta_I{}^J$

where λ_I are the eigenvalues of **W** and the columns of the coordinate transformation matrix $\tilde{\mathbf{\Lambda}}$ are given by the corresponding eigenvectors

Coarse graining via Coordinate Transformations

This gives a solution $\tilde{P}_I(t) = \dfrac{\lambda_I^t \tilde{P}_I(0)}{\sum_I \lambda_I^t \tilde{P}_I(0)}$ in the transformed coordinates

and a solution $P_I(t) = \dfrac{\sum_{JK} \tilde{\Lambda}_I{}^J \lambda_J^t (\tilde{\Lambda}^{-1})_J{}^K P_K(0)}{\sum_{IK} \lambda_I^t (\tilde{\Lambda}^{-1})_I{}^K P_K(0)}$ in the original string coordinates

Note that it is a sum of exponentials

The largest eigenvalue $\lambda_{\max}$ gives the fixed point

$$P_I^* = \frac{\sum_K \tilde{\Lambda}_I{}^{\max} \tilde{\Lambda}^{-1}{}_{\max}{}^K P_K(0)}{\sum_K \tilde{\Lambda}^{-1}{}_{\max}{}^K P_K(0)}$$

But it's difficult to determine $\tilde{\mathbf{\Lambda}}$

Coarse graining via Coordinate Transformations

- Can formulate the dynamics so that it is independent of the coordinate system used
- Can get explicit, exact solutions for the dynamics of selection only or mutation only systems
- For selection and mutation together can get formal solutions that have a constrained functional form (sum of exponentials)
- All these solutions are more naturally found, and expressed, in particular coordinate systems – strings/objects for selection and Walsh/Fourier modes for mutation

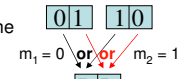
But what about recombination? Does there exist a "preferred" coordinate system for it, where the dynamics looks "simple"?

Coarse graining for Recombination

Crossover…

$$(\lambda^{\delta})_I{}^{JK}(M) = \prod_{r=1}^{N}\left(\bar{m}_r\delta_{i_r}{}^{j_r} + m_r\delta_{i_r}{}^{k_r}\right) \quad \text{where} \quad \delta_{i_r}{}^{j_r} \begin{matrix} = 1 \text{ if } i_r = j_r \\ = 0 \text{ otherwise} \end{matrix}$$

e.g. for 2 bits with one point crossover



i.e. r th bit of offspring i can only come from the r th bit of either parent j or parent k

$$P_{I_1I_2}(t+1) = p(01)\sum_{J_1J_2}\sum_{K_1K_2}\delta_{I_1J_1}\delta_{I_2K_2}P'_{J_1J_2}P'_{K_1K_2} + p(10)\sum_{J_1J_2}\sum_{K_1K_2}\delta_{I_1K_1}\delta_{I_2J_2}P'_{J_1J_2}P'_{K_1K_2} = p(01)P'_{I_1*}P'_{*I_2} + p(10)P'_{*I_2}P'_{I_1*}$$

16 terms in sum for each mask

Only one term per mask!

where $P'_{I_1*} = \frac{f(I_1*)}{\bar{f}(t)}P_{I_1*}(t)$ is the selection probability for the schema I_1*

Coarse graining for Recombination

In the general case

Building blocks of J with respect to the recombination mode m

$$P_I(t+1) = \sum_J M_I{}^J\left((1-p_c)P'_J + p_c\sum_m \frac{1}{2}(p_c(m)+p_c(\bar{m}))P'_{J_m}P'_{J_{\bar{m}}}\right)$$

Coarse grained exact evolution equation

2^N terms

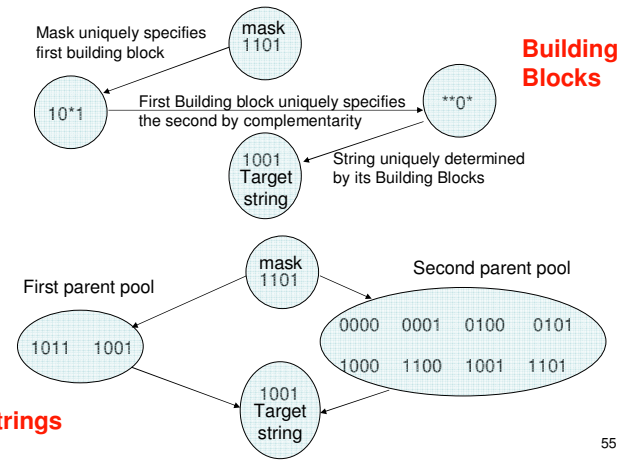
2^{3N} terms

$$P_I(t+1) = \sum_J M_I{}^J\left((1-p_c)P'_J + p_c\sum_M\sum_{KL}\frac{1}{2}(p(M)+p(\bar{M}))\lambda_J{}^{KL}(M)P'_KP'_L\right)$$

Microscopic exact evolution equation

Even if we are considering only strings, schemata naturally appear as the best way to describe string construction via recombination. There are a very particular set of schemata associated with each string – the Building Block schemata. They are the **ONLY** way a string can be built up by recombination

Coarse graining for Recombination



Coarse graining for Recombination

$$P_I(t+1) = \sum_J M_I{}^J\left((1-p_c)P'_J + p_c\sum_m \frac{1}{2}(p_c(m)+p_c(\bar{m}))P'_{J_m}P'_{J_{\bar{m}}}\right)$$

Can further coarse grain, e.g. $I \to I_m = \sum_{I_r\in\bar{m}} I \qquad \sum_{I_2=0,1} I_1I_2I_3 = I_1 * I_3$

$$P_{I_m}(t+1) = \sum_{J_m} M_{I_m}{}^{J_m}\left((1-p_c)P'_{J_m} + p_c\sum_{m'}\frac{1}{2}(p_c(m')+p_c(\bar{m}'))P'_{J_{mm'}}P'_{J_{m\bar{m}'}}\right)$$

This is an Exact Schema Theorem

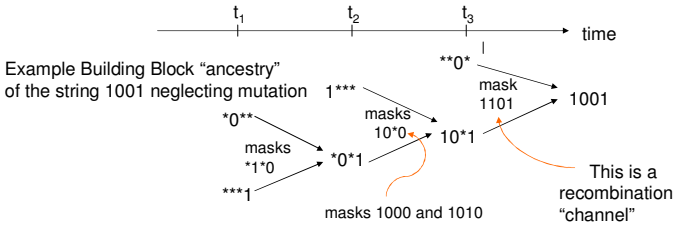
Building Blocks of the Schema J_m

Define effective fitness $P_{I_m}(t+1) = \frac{f^{\text{eff}}_{I_m}(t)}{\bar{f}(t)}P_{I_m}(t)$

The EA "works" by favoring Schemata of high effective fitness.

The role of recombination is to recombine effectively fit Building Block Schemata into effectively fit higher order Building Block Schemata

Coarse graining for Recombination



So when is recombination "useful"? Define

$$\Delta_I(m) = P'_I - P'_{I_m} P'_{I_{\bar{m}}}$$

SWLD Coefficient
(Selection Weighted Linkage Disequilibrium)

Coarse graining for Recombination

Recombination via a particular channel defined by a mask m increases/decreases the proportion/effective fitness of a given string or schemata I when

$$\Delta_I(m) \begin{matrix} < 0 \\ > 0 \end{matrix}$$ respectively

If $\Delta_I(m) < 0$, "channel" is "non-deceptive" ⇨ higher probability to select the Building Blocks of the string/schemata than the string/schemata itself

If $\Delta_I(m) > 0$, "channel" is "deceptive" ⇨ lower probability to select the Building Blocks of the string/schemata than the string/schemata itself

Standard Two-bit deception: f(0*) > f(1*) ⇨ $\Delta_I(m) > 0$

i.e. $P'_{11}(t) - P'_{1*}(t)P'_{*1}(t) > 0$

Coarse graining for Recombination

$\Delta_I(m)$ depends on the population, the fitness landscape, the crossover mask and the target string

For a uniform population, $\Delta_I(m) = 0$ for any landscape or mask
Recombination is neutral

For a random population, for a non-epistatic landscape $\Delta_I(m) < 0$
More recombination is better

For a random population, for a "deceptive" landscape/mask combination where $\Delta_I(m) > 0$, less recombination is better

In general, the formation of a string/schemata will involve a mix of non-deceptive channels and deceptive channels

Coarse graining for Recombination

Recombinative dynamics is most naturally written in terms of Building Block schemata. Our equations are exactly true for fixed length representations and homologous crossover operators.

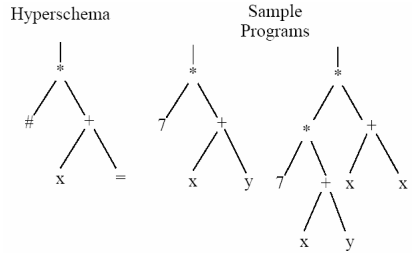
"Structurally" they are also true for other representations and non-homologous operators. However, the analogs of Building Block Schemata and recombination masks are more complicated and subtle. For instance, for Genetic Programming…

Poli's Extension to GP

First, what is the generalization of a Building Block schema?

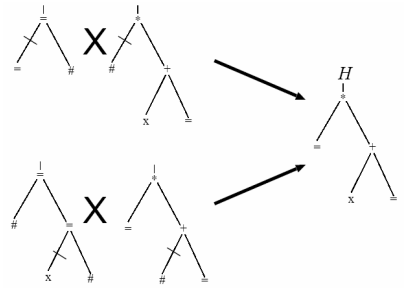
A *GP hyperschema* is a tree with internal nodes from $F \cup \{ = \}$ and leaves from $T \cup \{ =, \# \}$, where F is the function set and T the terminal set

= is a "don't care" symbol which stands for exactly one node, while # stands for any valid subtree.



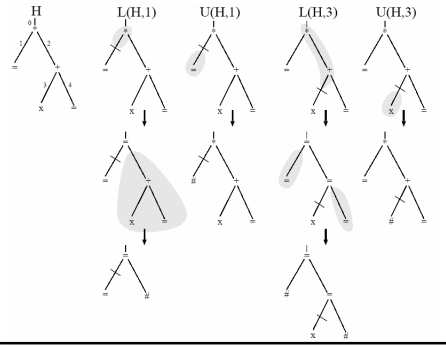
Poli's Extension to GP

For instance, for one point crossover Upper and lower Building Block hyperschemata express the structure the parents must have to produce offspring in a schema H for each crossover point



Poli's Extension to GP

The lower BB *L(H,i)* is obtained by replacing all the nodes between crossover point *i* and the root node with = nodes, and all the subtrees connected to the new = nodes with # nodes
The upper BB *U(H,i)* is obtained by replacing the subtree below crossover point *i* with a # node



Poli's Extension to GP

One can now write the equation for selection and recombination with respect to a given crossover operator in terms of Building Block hyperschemata instead of program trees

For example: For one-point crossover and no mutation, in Poli's notation

$$E[m(H,t+1)/M] = (1-p_{xo})p(H,t) + p_{xo}\sum_{j,k}\frac{1}{NC(G_j,G_k)} \times \sum_{i \in C(G_j,G_k)} p(L(H,i) \cap G_j, t)\, p(U(H,i) \cap G_k, t)$$

Notice the same structural form as for fixed length representation

G1, *G2*, … are all the possible fixed-size-and-shape schemata of order 0

NC(Gj,Gk) is the number of nodes in the tree fragment representing the common region between programs with shape *Gj* and programs with shape *Gk*

C(Gj,Gk) is (the set of indices of the crossover points) in such a common region

Poli's Extension to GP

• The terms *L(H,i)∩ Gj* and *U(H,i) ∩ Gk* in the exact GP schema theorem are fixed-size-and-shape Building Block hyperschemata

• Like a GA, GP with one-point crossover builds higher-order hyperschemata by juxtaposing lower-order ones

• For this to happen the Building Block hyperschemata need not necessarily be all fitter than average, "short" or even low-order.

• Generalizes to non-homologous crossovers such as subtree crossover

Coarse graining for Recombination via Coordinate Transformations

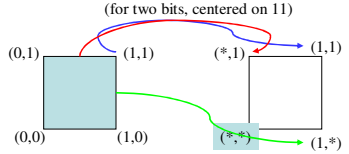
So, do the Building Block schemata form a coordinate basis? If so, what's the corresponding transformation matrix? Consider …

$\Lambda =$

	111	110	101	011	100	010	001	000
111	1	0	0	0	0	0	0	0
11*	1	1	0	0	0	0	0	0
1*1	1	0	1	0	0	0	0	0
*11	1	0	0	1	0	0	0	0
1**	1	1	1	0	1	0	0	0
1	1	1	0	1	0	1	0	0
**1	1	0	1	1	0	0	1	0
***	1	1	1	1	1	1	1	1

Building Block basis (BBB)

(for two bits, centered on 11)



Coordinate transformation maps vertices to hyperplanes

$$\Lambda_N = \Lambda_1^{\otimes N} \qquad \Lambda_1 = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$$

When transforming to the BBB $(\lambda')_I{}^{JK} = \lambda_R{}^{ST}\Lambda_I{}^R(\Lambda^{-1})_S{}^J(\Lambda^{-1})_T{}^K$

$(\lambda')_I{}^{JK}$ = 0 for a given I and m for all J and K except for one element, where "J = m" (the Building Block picked out by the mask) and K is the "complementary" Building Block, i.e. the non-zero element is always on the off-diagonal

Coarse graining for Recombination via Coordinate Transformations

Example: 2-bit GA with $p(M) = 1/4$ for all M, $I = (11)$

$$\lambda^{\mathrm{m}}_{(11)} = \frac{1}{4}\begin{pmatrix} 0&0&0&1 \\ 0&0&1&0 \\ 0&1&0&0 \\ 1&0&0&0 \end{pmatrix}$$

• In Building Block basis interaction matrix is skew diagonal

• Mask simply tells you which skew diagonal elements interact, e.g. mask 101011 points to building block 1*1*11 which interacts with *1*1** to give 111111

As BBB is a basis, can reconstruct any string frequency by inverting the transformation. For example, for the BBB centered on 11

$$P_{10} = P_{1*} - P_{11} \quad P_{01} = P_{*1} - P_{11} \quad P_{00} = P_{11} - P_{1*} - P_{*1} + P_{**}$$

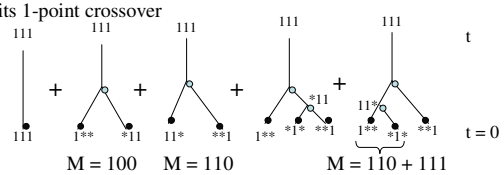
How does the BBB make the dynamics of a recombinative EA easier to analyze?

Coarse graining via Coordinate Transformations

Go back to the recursive solution attempted when examining the dynamics in the string basis. i.e. Iterate … by recursively substituting for • until get to t = 0

Example – 3-bits 1-point crossover

$P_{111}(t) =$



M = 100 M = 110 M = 110 + 111

I | t = $G_{II}(t,t')$ Probability that "Building Block" I propagates from t to t'

I | t' ⬭ = $\frac{1}{2}(p(M) + p(\bar{M}))\lambda_J{}^{KL}(M)\frac{f_K}{\bar{f}(t)}\frac{f_L}{\bar{f}(t)}$ Measures strength of interaction between "Building Blocks" J, K and L

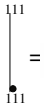
K • = $P_I(t)$

Skew-diagonal – only conjugate "Building Blocks" interact!

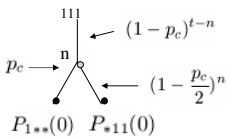
Coarse graining via Coordinate Transformations

Each tree tells us the probability of forming 111 by a given process. In principle can see which processes are most important. Tree depth bounded by N or t – whichever is smaller. MUCH SIMPLER THAN STRING ("OBJECT") BASIS! MUCH FEWER DIAGRAMS (PROCESSES) TO CONSIDER.

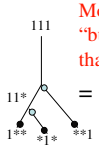
Example: 1-point crossover, no selection or mutation



$= (1-p_c)^t P_{111}(0)$



Moral: No point putting in "building blocks" of higher order than one!



$= 2(1-p_c)\left((1-\frac{p_c}{2})^t - (1-p_c)^t\right)$

$= 2(1-p_c)\left((1-p_c)^t - 2(1-\frac{p_c}{2})^t + 1\right) P_{1**}(0)P_{*1*}(0)P_{**1}(0)$

Dominates in long time limit – Geiringer's theorem

Coarse graining via Coordinate Transformations

So, can determine which processes are most important (probable). Can we evaluate the effect of all the processes together? Yes! E.g., for one-point crossover without selection and mutation, in the continuous time limit

$$P_I(t) = \sum_{n=0}^{N-1} e^{\frac{-np_c}{N-1}}(1-e^{\frac{-p_c}{N-1}})^{N-n}\mathcal{P}(n+1)$$

where $\mathcal{P}(n+1) = \sum_i \Pi_{n_i=1}^{N-n} P_{n_i}(0)$ depends only on the initial conditions

Each $P_{n_i}(0)$ is the probability for the BB i which crossover can combine to give genotype I. The product is over the different numbers of Building Blocks and the sum over the different permutations for a given number. For example, for N=3 and

$I = 111$ $\mathcal{P}(1) = P_{1**}(0)P_{*1*}(0)P_{**1}(0)$, $\mathcal{P}(2) = P_{11*}(0)P_{**1}(0) + P_{1**}(0)P_{*11}(0)$ (two permutations) and $\mathcal{P}(3) = P_{111}(0)$.

Asymptotic behavior is governed by n=0 which gives the Geiringer limit

Coarse graining via Coordinate Transformations

So, in the Building Block basis …

- Building Blocks schemata are the natural EDOF for recombination
- They are dynamical and not necessarily "short" or "fit"
- **They are the ONLY way in which higher order "objects" can be built up by recombination**
- Generically, the "construction" term dominates
- BBB is complete but not orthonormal
- There are |X| equivalent BBB (related by simple permutations)
- Only "dual" objects (i.e. complementary BBs) interact, e.g. line and plane intersect at a point
- **Gives exact solution for recombination only**
- Selection complicated because schema fitnesses are population dependent

Appropriate in "strong" crossover regime

Coarse graining via Coordinate Transformations

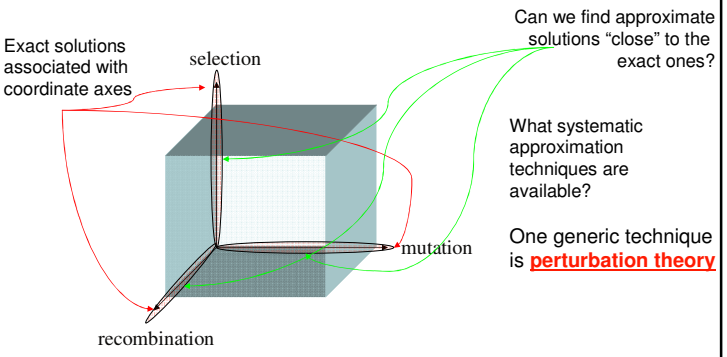
Operator / Basis	Selection	Mutation	Recombination
String	Simple/Natural Selection matrix diagonal Exact solution	Complicated Mutation matrix not diagonal	Very complicated Highly redundant
Walsh	Unnatural Selection matrix not diagonal	Simple/natural Mutation matrix diagonal Exact solution	Complicated
Building Block	Complicated BB fitnesses are time dependent	Simple Mutation matrix triangular	Simple/natural Recombination matrix skew diagonal Exact solution

Coarse graining via Coordinate Transformations

- The most "natural" description of the dynamics of an EA depends on the genetic operators used. Selection looks simplest in terms of the string/object basis. Mutation looks simplest in the Walsh basis and recombination in the BBB.
- An exact solution appears naturally in terms of the microscopic string/object degrees of freedom for selection only dynamics
- Via a coarse graining implemented using a coordinate transformation, exact solutions can also be found for mutation only dynamics (Walsh basis) and for recombination only dynamics (BBB)

If we have an exact solution for selection only, i.e. p = 0, what do we think a solution for p = $1/10^{10}$ should look like? Very different or similar? What about p = $1/10^{6}$? What about p = $1/10^{4}$? What about p = 0.01, 0.1 …?

Beyond the exact solutions



Beyond the exact solutions

Example of a perturbative solution: consider mutation and selection in the unnormalized variables $x_I(t)$ where

$$P_I(t) = \frac{x_I(t)}{\sum_I x_I(t)}$$

then

$$x_I(t+1) = \sum_{JK} M_I{}^J F_J{}^K x_K(t)$$

Where we write $F_I{}^J = \underbrace{(\delta_I{}^J + \epsilon F'_I{}^J)}$ **or** $M_I{}^J = \underbrace{(\delta_I{}^J + \epsilon M'_I{}^J)}$

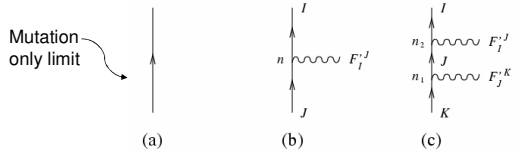
natural for weak selection — natural for weak mutation

and consider ϵ to be small, seeking a solution of the form

$$P_I(t) = \sum_{n=0}^{\infty} \epsilon^n P_I^{(n)}(t)$$

Beyond the exact solutions

For the weak selection case we should use the Walsh basis. Once again, the terms can be represented diagrammatically



Diagramatic representation of the $O(1)$, $O(\epsilon)$ and $O(\epsilon^2)$ perturbative terms

The first two diagrams represent

$$\hat{x}_I(t) = (1-2p)^{|I|t}\hat{x}_I^{(0)}(0) + \epsilon \sum_J \sum_{n=0}^{t-1} (1-2p)^{|I|(t-n)} F'_I{}^J (1-2p)^{|J|n} \hat{x}_J^{(0)}(0)$$

Propagation of Walsh mode I - mutation only

Walsh transform of Fitness – causes Interaction between I and J

Propagation of Walsh mode J - mutation only

Beyond the exact solutions

As $\hat{P}_I(t) = \frac{1}{2^{\frac{N}{2}}} \frac{\hat{x}_I(t)}{\hat{x}_0(t)}$ the perturbative solution can be written as

$$\hat{P}_I(t) = (1-2p)^{|I|t}\hat{P}_I(0) + \epsilon(1-2p)^{|I|t} \sum_J \frac{\left(1-(1-2p)^{(|J|-|I|)t}\right)}{1-(1-2p)^{(|J|-|I|)}} \hat{F}'_I{}^J \hat{x}_J(0)$$

$$-\epsilon(1-2p)^{|I|t} 2^{\frac{N}{2}} \hat{P}_I(0) \sum_J \frac{\left(1-(1-2p)^{|J|t}\right)}{1-(1-2p)^{|J|}} \hat{F}'_0{}^J \hat{x}_J(0);$$

Note that the only term that is NOT exponentially damped is the J = 0 term. The fixed point is given by

$$\hat{P}^*_I = \frac{\epsilon}{2^{\frac{N}{2}}} \frac{\hat{F}'_I{}^0}{((1-2p)^{-|I|}-1)}$$

and determines how far from the random population fixed point the system has moved due to selection effects. Numerically, considered three cases …


```
(2,3) Average fitness and Best string for Eigen model in the string basis.

Number of Bits: 7
Mutation Ratio: 0.1
       Epsilon: 1.0
  Maximum Time: 50

Fitness values: f_1111111=1.5 and f_i=1.0 (for i!=1111111)
Uniform initial conditions P_i(0)=1/(2^N).

(4,5) Average fitness and Best string for the 1+E(counting ones) model
in the string basis.

Number of Bits: 7
Mutation Ratio: 0.1
       Epsilon: 1.0
  Maximum Time: 50

Fitness values: f_i = 1 + E*(sum of 1s in string i).
We choose E=0.0714... so that f_i ranges between 1.0 and 1.5.
Uniform initial conditions P_i(0)=1/(2^N).

(6,7) Average fitness and Best string for Random model in the string basis.

Number of Bits: 7
Mutation Ratio: 0.1
       Epsilon: 1.0
  Maximum Time: 50

Fitness values f_i in [1, 1+R], R=random number in [0,0.5]
Uniform initial conditions P_i(0)=1/(2^N).
```


Beyond the exact solutions

For example, for the Eigen model (needle-in-a-haystack) and with a BIG perturbation – 50% difference in fitness between the needle and the hay

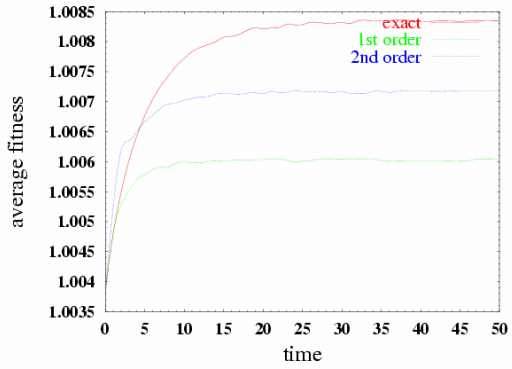


Figure 2. Average fitness – Eigen model – 7 bits – Planarity deviation: 50 %

Beyond the exact solutions

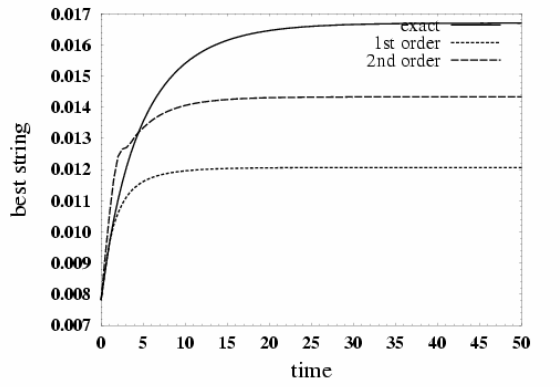


Figure 3. Best string – Eigen model – 7 bits – Planarity deviation: 50 %

Beyond the exact solutions

- So, perturbation theory implemented diagrammatically not only gives deeper qualitative insights into the relation between the different genetic operators but also shows strong potential as a quantitative tool for approximating the dynamics of EAs
- At this level it is giving approximate solutions to the underlying "microscopic" dynamics. This means keeping track of all strings or Walsh modes. This is problematic for larger values of N
- Means must be found to see to what extent a further coarse graining may be carried out so that fewer degrees of freedom have to be tracked. This involves coarse grainings that act as projections

81

Coarse Graining By Projections

Making intractable dynamics more tractable

82

Coarse graining via Projections

Introduce a general coarse-graining operator $\mathcal{R}(\eta, \eta')$

Which coarse grains from the variables $\eta \in X_\eta$ to the variables $\eta' \in X_{\eta'} \subset X_\eta$

Given two such coarse grainings:

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

but $\mathcal{R}(\eta', \eta'')P_{\eta'}(t) = P_{\eta''}(t)$

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

i.e. coarse grainings form a semi-group – **"Renormalization Group"**

83

Coarse graining via Projections

Dynamics coarse grains via

$$\mathcal{R}(\eta, \eta')\mathcal{H}(\mathbf{p}, \mathbf{f}, \mathbf{P}_\eta(t))$$

If this can be written in the form

$$\mathcal{H}(\mathbf{p}', \mathbf{f}', \mathbf{P}_{\eta'}(t))$$

with suitable "renormalizations"

$\mathsf{f} \longrightarrow \mathsf{f}'$ and $\mathsf{p} \longrightarrow \mathsf{p}'$

then the dynamics is form covariant or invariant under the coarse graining. If $\mathsf{f} = \mathsf{f}'$ and $\mathsf{p} = \mathsf{p}'$ dynamics is "compatible"

84

Coarse graining via Projections

Examples: Compatible Coarse grainings

1. Selection and Phenotypes
   - Unitation, e.g. 2^N genotypes → (N+1) phenotypes
   - Eigen model (NIAH), e.g. 2^Ngenotypes → 2 phenotypes
2. Mutation and Crossover and Schemata
   - 2^N genotypes → 2^{N_2} coarse-grained genotypes

Incompatible Coarse grainings

1. Selection, Mutation and Crossover and Schemata
   - 2^N genotypes → 2^{N_2} coarse grained genotypes
   - $f_\alpha = \mathcal{R}(x,\alpha) f_x = \sum_{x\in\alpha} f_x P_x(t) / \sum_{x\in\alpha} P_x(t)$. time-dependent

Coarse graining via Projections

In BBB for 1-point crossover…

$$P_{111}(t+1) = (1-p_c)P_{111}(t) + \frac{p_c}{2}(P_{1**}(t)P_{*11}(t) + P_{11*}(t)P_{**1}(t))$$

"Zap" (projection) $111 \to 11*$

$$P_{11*}(t+1) = (1-p_c)P_{11*}(t) + \frac{p_c}{2}(P_{1**}(t)P_{*1*}(t) + P_{11*}(t)P_{***}(t))$$

$$\Rightarrow \quad P_{11*}(t+1) = (1-\frac{p_c}{2})P_{11*}(t) + \frac{p_c}{2}P_{1**}(t)P_{*1*}(t)$$

Note – coarse grained (projected) 3-bit equation same as "microscopic" 2-bit equation with "renormalization" $p_c \to \frac{p_c}{2}$

$$P_{11}(t+1) = (1-p_c)P_{11}(t) + p_c P_{1*}(t)P_{*1}(t)$$

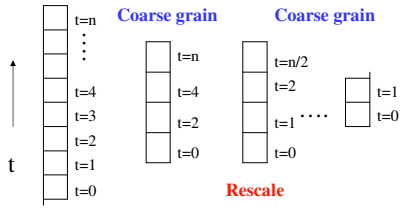
FORM INVARIANCE

Coarse graining via Projections

- Generalizes to the case of variable-length GAs and GP; Building Block Schemata →Building Block Hyperschemata; "form invariance" of equations over different types of EA and form invariant upon coarse graining to schemata/hyperschemata;
- Gives exact form of the Schema Theorem and generalizes it to EAs other than GAs
- Neglecting the "construction" terms leads to standard Holland Schema Theorem as an approximation

Coarse Graining by Projection - "Divide and Conquer"

Example: 1-bit



Can we coarse grain an n generation problem to a one generation problem? Much easier to solve the dynamics over only one generation!

$X_1(t)$ – unnormalized incidence vector

p – mutation rate

$$\begin{pmatrix} X_1(t+2) \\ X_0(t+2) \end{pmatrix} = \underbrace{\begin{pmatrix} (1-p)f_1 & pf_0 \\ pf_1 & (1-p)f_0 \end{pmatrix}^2} \begin{pmatrix} X_1(t) \\ X_0(t) \end{pmatrix}$$

Evolves bit two time steps in landscape f(1), f(0) with mutation p

Coarse Graining by Projection - "Divide and Conquer"

$$\begin{pmatrix} X_1(t'+1) \\ X_0(t'+1) \end{pmatrix} = \begin{pmatrix} (1-p_1')f_1' & p_0'f_0' \\ p_1'f_1' & (1-p_0')f_0' \end{pmatrix} \begin{pmatrix} X_1(t') \\ X_0(t') \end{pmatrix}$$

Evolves bit one time step in "renormalized" landscape f'(1), f'(0) with asymmetric mutation rates p'(1) and p'(0)

Equivalent dynamics (all we did was "change names"!, i.e. "renormalize")

$$f_1' = (1-p_1)f_1^2 + p_1 f_0 f_1$$
$$f_0' = (1-p_0)f_0^2 + p_0 f_0 f_1$$
$$p_1' = p_1\left(\frac{(1-p_1)f_1 + (1-p_0)f_0}{(1-p_1)f_1 + p_1 f_0}\right)$$
$$p_0' = p_0\left(\frac{(1-p_0)f_0 + (1-p_1)f_1}{(1-p_0)f_0 + p_0 f_1}\right)$$

89

Coarse Graining by Projection - "Divide and Conquer"

Evolution of mutation/selection GA over n time steps with fitness landscape f(1), f(0) and mutation rates p(2) and p(1) is same as that of a GA with "renormalized" landscape and mutation rates, f'(1), f'(0), p'(2), p'(1) over n/2 time steps!

UNIVERSALITY

Fixed points of Renormalization Group transformation:
|ln(f(1)/f(0))| = 0, p(1) = p(0) = 0; no selection/mutation – "FERROMAGNETIC"
|ln(f(1)/f(0))| = infinity, p(1) = p(0) = 0; strong selection – "FROZEN"
|ln(f(1)/f(0))| = constant, p(1) + p(0) = 1; neutral evolution – "PARAMAGNETIC"

90

Coarse Graining by Projection - "Divide and Conquer"

- Iterated map takes you to a problem with fewer degrees of freedom – NOT associated with "trivial" symmetries.
- Linearization around the fixed points of the equations give the late time asymptotics
- Can understand "universality" of behaviour
- Can coarse grain in both "space" and "time"
- Coarse graining can almost never be done exactly
- Have to decide what coarse graining is most appropriate for a given model

91

EC Theory – Coarse-grained

So what do we have so far?

Exact	Yes
Mathematically rigorous	Yes ?
Unifies Phenomena	Yes
Intuitive	Yes
Predicts well	No (but getting there!)
Useful for Practitioners	Yes/No

92

The Bottom Line …

- Present taxonomy in EC theory is inadequate
- Taxonomy can be greatly improved by using "distance" measures
- Taxonomy and universality can also be much better understood using an appropriate coarse graining
- Coarse grained genetic dynamics unifies and makes compatible different areas of EC and different previous theoretical formulations
- GAs and GP – different sides of the same coin
- Old Schema theory/BB hypothesis and Vose type models – different sides of the same coin
- Coarse graining, perturbation theory and the Renormalization Group offer a generic methodology for approximating genetic dynamics

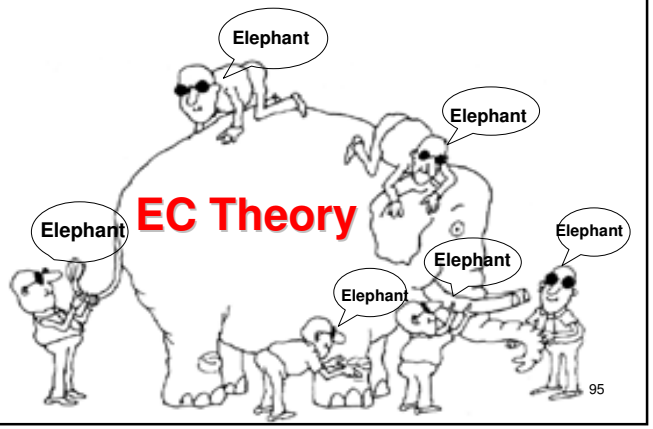
93

The Bottom Line …

- Theory in EC is **NOT** particularly well developed. If it was there wouldn't be such a huge expectation gap between theoreticians and practitioners. No systematic approximation techniques for attacking problems from first principles
- Practitioners have to realize what is and isn't theoretically feasible (theoretical population biologists have spent nearly a century achieving things that "practitioners" would scorn).
- Practitioners could really help by stress testing theory (too much testing of theory in the hands of people who make up the theory and too much testing of "never to be used" algorithms by practitioners)

94

The Bottom Line …



95

Acknowledgements

Collaborators:

Henri Waelbroeck, Riccardo Poli, Alden Wright, Peter Stadler, Chryssomalis Chryssomalakos, Jon Rowe, Bill Langdon, Conor Ryan, Miguel Nicolau

Students and postdocs: Adolfo Zamora, Rosalia Aguirre, Jaime Mora, Odon Palacios, Andres Aguilar, Ian Garcia, Jorge Cervantes, Rosario Cruz

Valuable Critics:

Michael Vose, Dave Goldberg, Anil Menon, Ken De Jong, Nic McPhee, Robert Heckendorn

Funding: CONACyT; DGAPA, UNAM; EPSRC; Leverhulme Trust, DIAS, SFI Ireland

96

References

CRS and H. Waelbroeck, ICGA97, 34-41 (1997)
CRS and H. Waelbroeck, Evol. Comp. 7, 109-124 (1999)
CRS, H. Waelbroeck and R. Aguirre, FOGA5, 117-135 (1999)
CRS, GECCO2001, 631-638 (2001)
CRS, Acta Phys. Slov. 52, 515-524 (2002)
R. Poli, GECCO2000, 469-476 (2000)
R. Poli, Genetic Prog. and Evol. Mach. 2, 123-163 (2001)
W.B. Langdon and R. Poli, Foundations of Genetic Programming, Springer-Verlag (2002)
R. Poli et al, EuroGP2002, 211-227, Springer-Verlag (2002)
P.F. Stadler and CRS, Comments in Theor. Biol. 8, 389 (2003)
R. Poli et al, GECCO2002, 211-227 (2002)
CRS and A. Zamora, GECCO2003
CRS and C. Chryssomalakos, GECCO2004
CRS and R. Poli, in Frontiers of Evolutionary Computation, Kluwer (2004)

References

Vose, M.D. (1999) *The Simple Genetic Algorithm: Foundations and Theory*, (MIT Press, Cambridge MA).

Bürger, R. (2000) "The Mathematical Theory of Selection, Recombination and Mutation", Wiley Series in Mathematical and Computational Biology.

Alden H. Wright. The exact schema theorem. http://www.cs.umt.edu/CS/FAC-/WRIGHT/papers/schema.pdf, January 2000.

Altenberg, L. (1995). The Schema Theorem and Price's Theorem. In Whitley, L. D. and Vose, M. D., editors, *Foundations of Genetic Algorithms 3*, pages 23–49, Estes Park, Colorado, USA. Morgan Kaufmann.

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

Goldberg, D. E. (1989b). Genetic algorithms and Walsh functions: II. Deception and its analysis. *Complex Systems*, 3(2):153–171.

Prügel-Bennett, A. and Shapiro, J. L. (1994). An analysis of genetic algorithms using statistical mechanics. *Physical Review Letters*, 72:1305–1309.

Whitley, D. (1992). An executable model of a simple genetic algorithm. In Whitley, D., editor, *Foundations of Genetic Algorithms Workshop (FOGA-92)*, Vail, Colorado.

J. Rowe and C. Reeves, Genetic Algorithms – Principles and Perspectives A Guide to GA Theory, Kluwer 2002.