CSE/ECE 848 Introduction to Evolutionary Computation

Module 2, Lecture 8, Part 1a
Introduction to Schema Theory

Erik D. Goodman, Executive Director
BEACON Center for the Study of Evolution in Action
Professor, ECE, ME, and CSE

Explaining Why a GA Works – Intro to GA Theory

- This lecture: introduction to schema theory
- Next few lectures: some classical results:
  - Schema theorem – explains how building blocks are assembled
  - Implicit parallelism – explains that each evaluation provides information on many possible candidate solutions
  - k-Armed Bandit problem – how the GA allocates search effort among schemata

What is a GA DOING? -- Schemata and Hyperstuff

- Schema -- adds "*" to alphabet, means "don't care" – any value
- One schema, two schemata (*forgive* occasional misuse in Whitley)
- Definition: *ORDER* of schema H -- o(H): # of non-*'s
- Def.: *Defining Length* of a schema, $\Delta(H)$: distance *between* first and last non-* in a schema; for example: Δ (**1*01*0**) = 5 (= number of positions where 1-pt crossover can disrupt it).

  (NOTE: diff. xover → diff. relationship to defining length)
- Strings or chromosomes or individuals or "solutions" are order L schemata, where L is length of chromosome (in bits or loci). Chromosomes are *INSTANCES* (or members) of lower-order schemata

Cube and Hypercube

Vertices are order ? schemata

Edges are order ? schemata

Planes are order ? schemata

Cubes (a type of hyperplane) are order ? schemata

8 different order-1 schemata (cubes): 0***, 1***, *0**, *1**, **0*, **1*, ***0, ***1

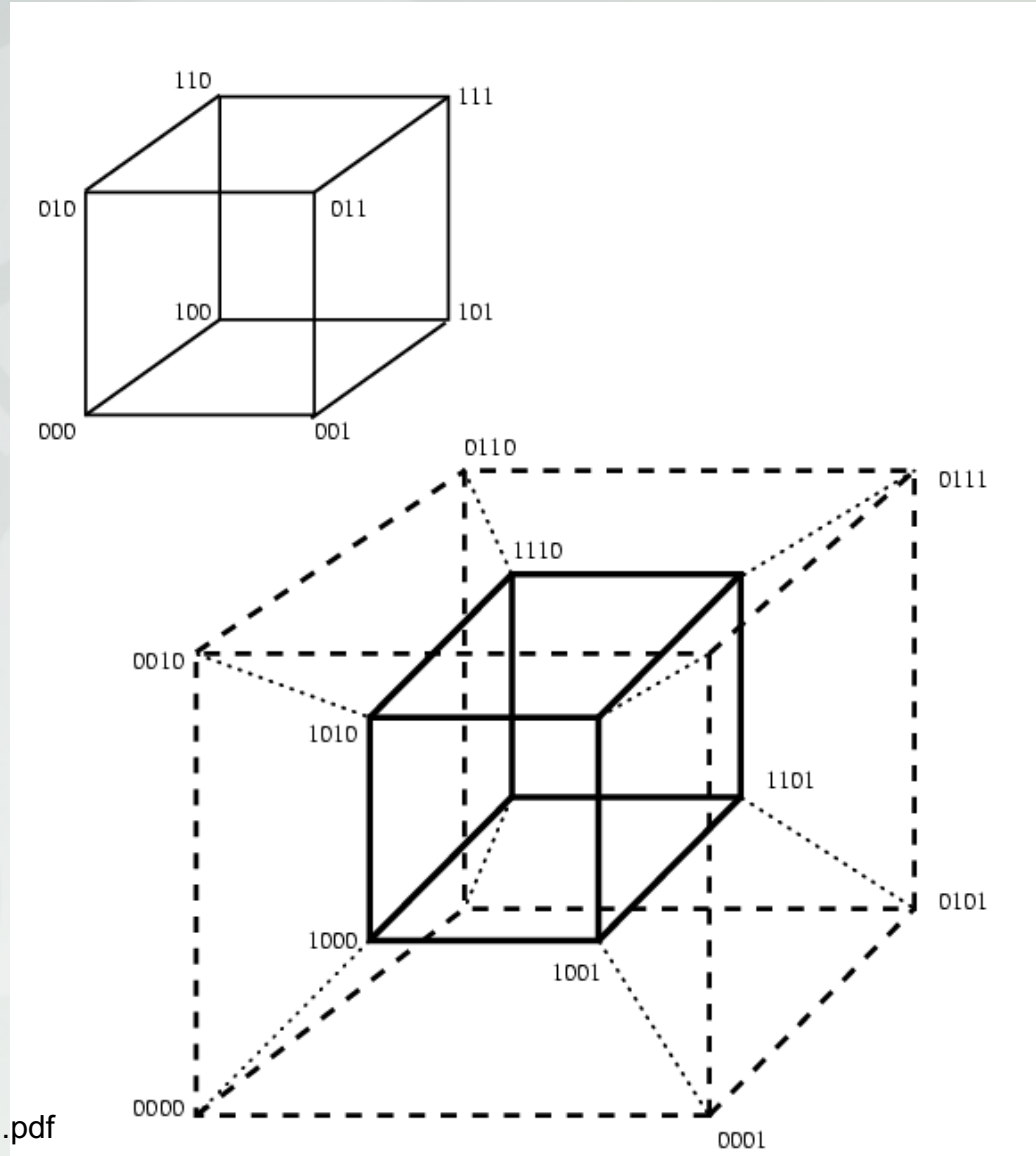

Figure from Whitley, A Genetic Algorithm Tutorial, https://www.cs.colostate.edu/~genitor/MiscPubs/tutorial.pdf

Hypercubes, Hyperplanes, etc.

- A string is an instance of 2^L-1 schemata or a member of that many hyperplane partitions (-1 because ****…*** all *’s, *the whole space*, is *not* counted as a schema, per Holland)
- List them, for L=3: example: string 101

  101, 10*, 1*0, *01, **1, *0*, 1**
- What is the fitness of a schema H?

  We define the fitness of schema H at any time t*: f(H, t)*,

  as the average fitness of the INSTANCES of schema H in the current population (since we can’t calculate its theoretical fitness without evaluating ALL of its instances)

GA Sampling of Hyperplanes

- So, in general, string of length L is an instance of 2^L-1 schemata
- But how many schemata are there in the whole search space?

  (how many choices each locus?)
- Since one string instances 2^L-1 schemata, how much does a population tell us about schemata of various orders?
- *Implicit parallelism:* one string's fitness tells us something about relative fitnesses of more than one schema