Genetic Algorithms

Presented at
Computer Society Meeting
5 June 1997, Ottawa

by

Prof. V. Rao Vemuri
Department of Applied Science
University of California, Davis
Livermore, CA

vemuri1@llnl.gov

OUTLINE

- **What are Genetic Algorithms?**
- **Where are They Useful?**
- **A Simple Example**
- **More Realistic Examples**

WHAT ARE GAs?

- INSPIRED BY BIOLOGICAL EVOLUTION
- SEARCH AND OPTIMIZATION PROCEDURES
- METAPHORICAL EVOLUTIONARY PROCESSES
- GOOD IN RUGGED SEARCH SPACES
- GOOD FOR MULTIMODAL ENVIRONMENTS
- PROBLEM INDEPENDENT

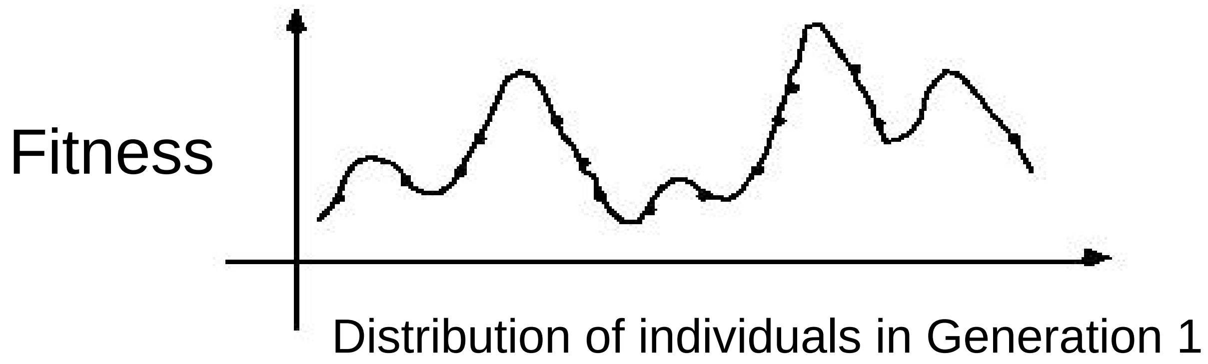
MAIN IDEAS

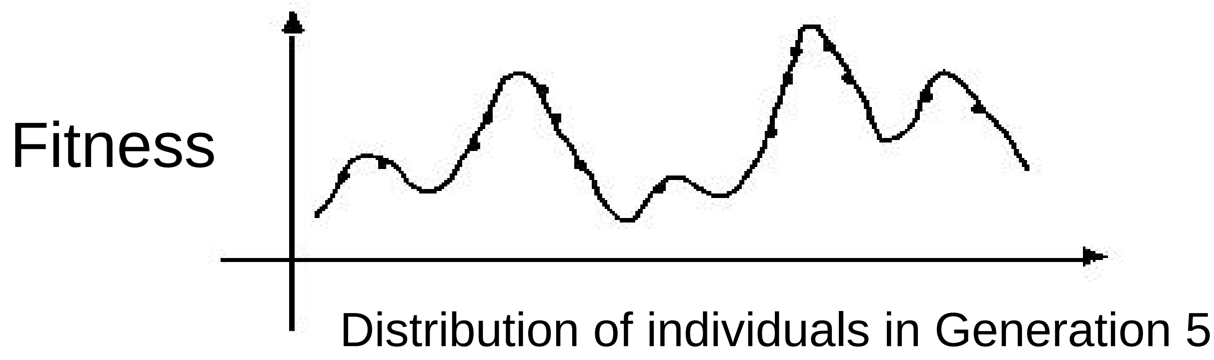
- SOLUTIONS ENCODED AS CHROMOSOMES
- MAINTAINS POPULATION OF SOLUTIONS
- EVALUATE FITNESS OF ALL SOLUTIONS
- SELECT THOSE WITH DESIRABLE TRAITS
- MATE THEM VIA CROSSOVER (EXPLOITATION)
- MUTATE THEM (EXPLORATION)
- REPEAT UNTIL "CONVERGENCE"

CHARACTERISTICS OF GAs

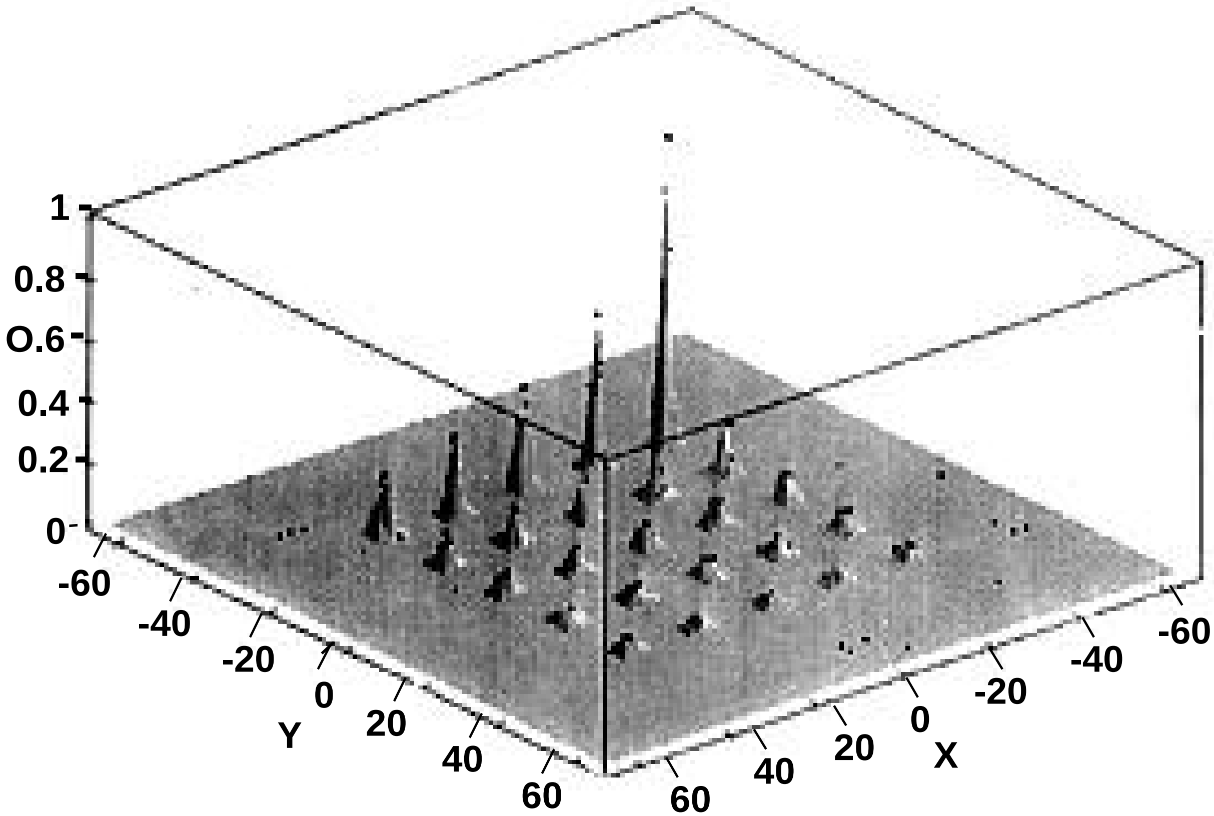
- CAN SOLVE HARD PROBLEMS RELIABLY
- EASY TO INTERFACE WITH EXISTING MODELS
- EASY TO HYBRIDIZE WITH OTHER METHODS
- EXTENDIBLE
- NEEDS VERY LITTLE DOMAIN KNOWLEDGE

SEARCH USING GA

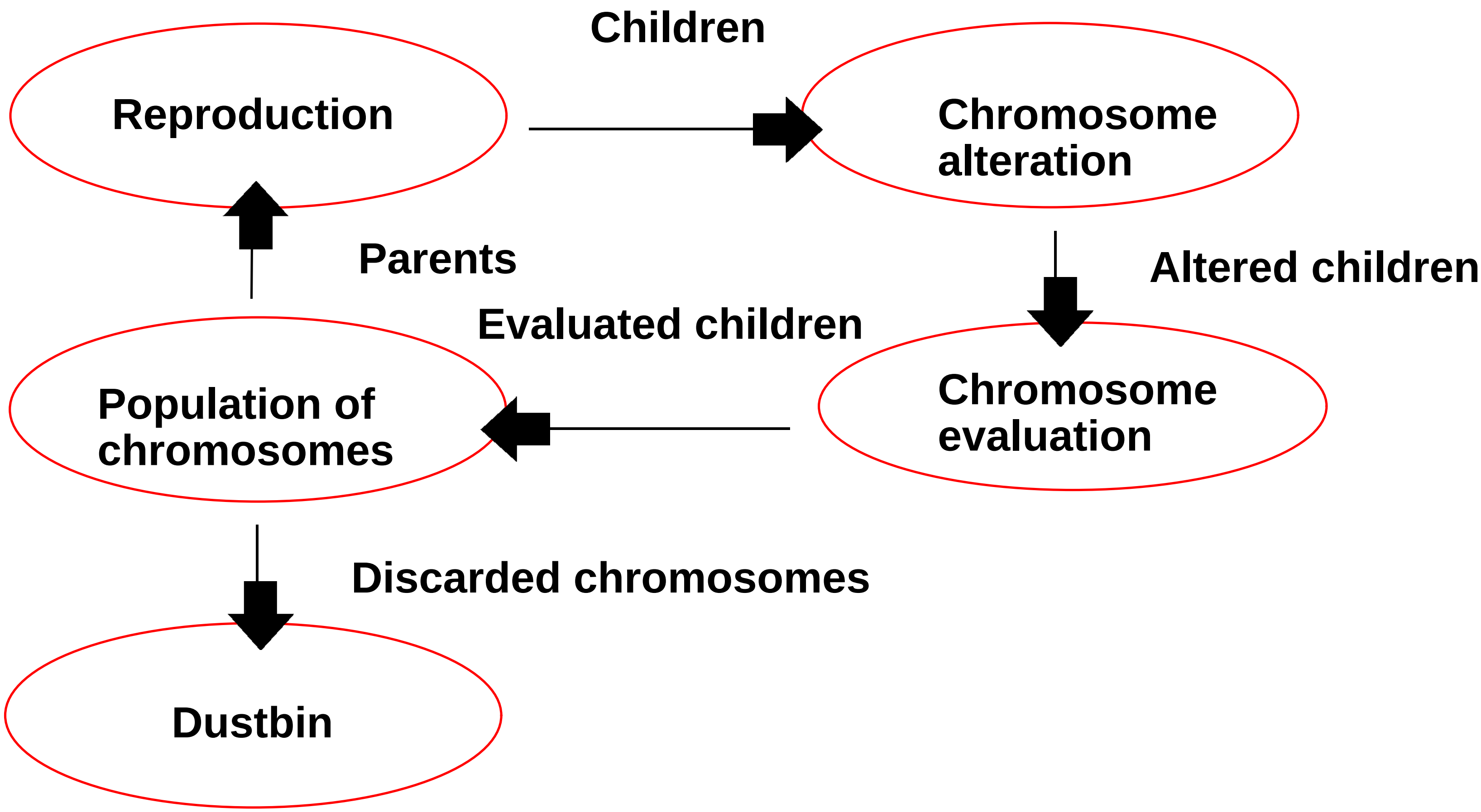




MULTI-MODAL OPTIMIZATION



HOW A GENETIC ALGORITHM WORKS



GENETIC OPERATORS

- **SELECTION:**
  SURVIVAL OF THE FITTEST

- **CROSSOVER:**
  COMBINES INDIVIDUALS TO GENERATE NEW COMBINATIONS OF PARTIAL SOLUTIONS

- **MUTATION:**
  INTRODUCES FEATURES THAT ARE NOT PRESENT IN THE PARENTS

SELECTION OPERATOR

- **CALCULATE FITNESS OF ALL MEMBERS**
- **SELECT THOSE WITH BETTER FITNESS SCORES**
- **USE ONE OF MANY SELECTION SCHEMES**

  - ROULETTE WHEEL SELECTION
  (A.K.A. Fitness Proportionate Reproduction, or FPR)

  - TOURNAMENT SELECTION

A SIMPLE CROSSOVER OPERATION

PARENT 1	**1 0 0 <u>1 1 1</u> 0 0**
PARENT 2	**0 1 0 <u>0 0 1</u> 1 0**
CHILD 1	**1 0 0 0 0 1 0 0**
CHILD 2	**0 1 0 1 1 1 0 0**

Other types of Crossovers exist

MUTATION OPERATION

PARENT	**1 0 0 <u>1</u> 1 1 0 0**
CHILD	**1 0 0 <u>0</u> 1 1 0 0**

SAMPLE PROBLEM

- A FRANCHISER WANTS PROFIT-MAKING STRATEGY
- OPTIONS ARE: AFFORDABLE VS FANCY

  HOT DOGS VS CREPES

  SOFT DRINK VS WINE
- FORECAST ON COSTS AND SALES AVAILABLE

ENCODING THE PROBLEM

• DEFINE A 3-BIT CHROMOSOME

BIT 3	BIT 2	BIT 1

• **ENCODE AS FOLLOWS**

- BIT 3 : 0 AFFORDABLE, 1, FANCY

- BIT 2 : 0 HOT DOGS, 1, CREPES

- BIT 1 : 0 SOFT DRINKS, 1, WINE

FITNESS EVALUATION

- EACH OPTION CAN BE REPRESENTED BY A 3-BIT SEQUENCE
- IN THIS SMALL PROBLEM THERE ARE ONLY 8 OPTIONS
- ASSUME THAT THE PROFIT MADE ON EACH OPTION CAN BE CALCULATED BY A FORMUL, SUCH AS

PROFIT = DECIMAL VALUE OF 3-BIT SEQUENCE

A PEEK AT THE CORRECT SOLUTION

OPTION	PROFIT	OPTION	PROFIT
000	$ 0	100	$ 4
001	$ 1	101	$ 5
010	$ 2	110	$ 6
011	$ 3	111	$ 7

EXHAUSTIVE ENUMERATION IMPRACTICAL

- **CONSIDER TRAVELLING SALESPERSON PROBLEM**
- **ASSUME**

  20-CITY PROBLEM TAKES ABOUT 1 HOUR
- **THEN**

  21-CITY PROBLEM TAKES ABOUT 20 HOURS

  22-CITY PROBLEM TAKES ABOUT 20 DAYS

  •

  •

  25-CITY PROBLEM TAKES ABOUT 6 CENTURIES!!

A POSSIBLE SOLUTION BY GA

- TRY 4 RANDOMLY SELECTED STRATEGIES AT 4 DIFFERENT OUTLETS
- OBSERVE HOW OUTLETS PERFORM FOR A WEEK
- COMPARE THEIR PERFORMANCE
- PICK THOSE THAT ARE DOING WELL: MATING POOL
- GENERATE NEW STRATEGIES FROM MATING POOL USING CROSSOVER AND MUTATION
- GO BACK TO FIRST STEP

LET'S TRY

X(i)	F(i)	M.p	F(i)	X(i)	F(i)
011	3	011	3	111	7
001	1	110	6	010	2
110	6	110	6	110	6
010	2	010	2	010	2
Total F	12		17		17
Worst	1		2		2
Ave	3		4.5		4.25
Best	1		6		7

WHAT DID THE FRANCHISER LEARN?

(1) $3 IS AN ESTIMATE OF THE AVERAGE FITNESS (PROFIT)OF THE SEARCH SPACE

- **Estimate Based on 4 Samples**

(2) THAT

- **110 is 200% better than estimated average**
- **010 is 2/3 as good as estimated average**
- **001 is 1/3 as good as estimated average**

WHAT DO WE DO NEXT?

OPTION 1: BRUTE FORCE METHOD

- **CONTINUE SEARCHING FOR BETTER SOLUTIONS**

DRAWBACK:

- **UNIVERSE CREATED 15 BILLION YEARS AGO**
- **AT BILLION SOLUTIONS/SEC, WE WOULD HAVE SEARCHED ONLY 2**90 SOLUTIONS**
- **THIS MEANS THAT OUR CHROMOSOMES CANNOT BE LONGER THAN 90 BITS!!**

ANOTHER LESSON LEARNED

OPTION 2: GREEDY METHOD

- **BECAUSE 110 IS 200% BETTER THAN THE AVERAGE, GRAB IT. DO NOT WORRY ABOUT POSSIBLE BETTER SOLUTIONS**
- **EXPLORING EVERY NEW POINT COSTS $6 - $3 = $3, on the average**
- **NOT EXPLORING COSTS $7 - $6 = $1, on the average**
- **THIS IS EXPLORATION VS EXPLOITATION!**

SUMMARY: A SIMPLE GA

GENERATE RANDOMLY AN INITIAL POPULATION

EVALUATE FITNESS OF THE POPULATION

for *GENERATION* = 1 TO *MAX_GENERATIONS*

SELECTION: CREATE MATING POOL

MATING: CROSSOVER AND MUTATE

REPLACEMENT: REPLACE ENTIRE POPULATION WITH OFFSPRING

end for

OTHER EXAMPLE PROBLEMS

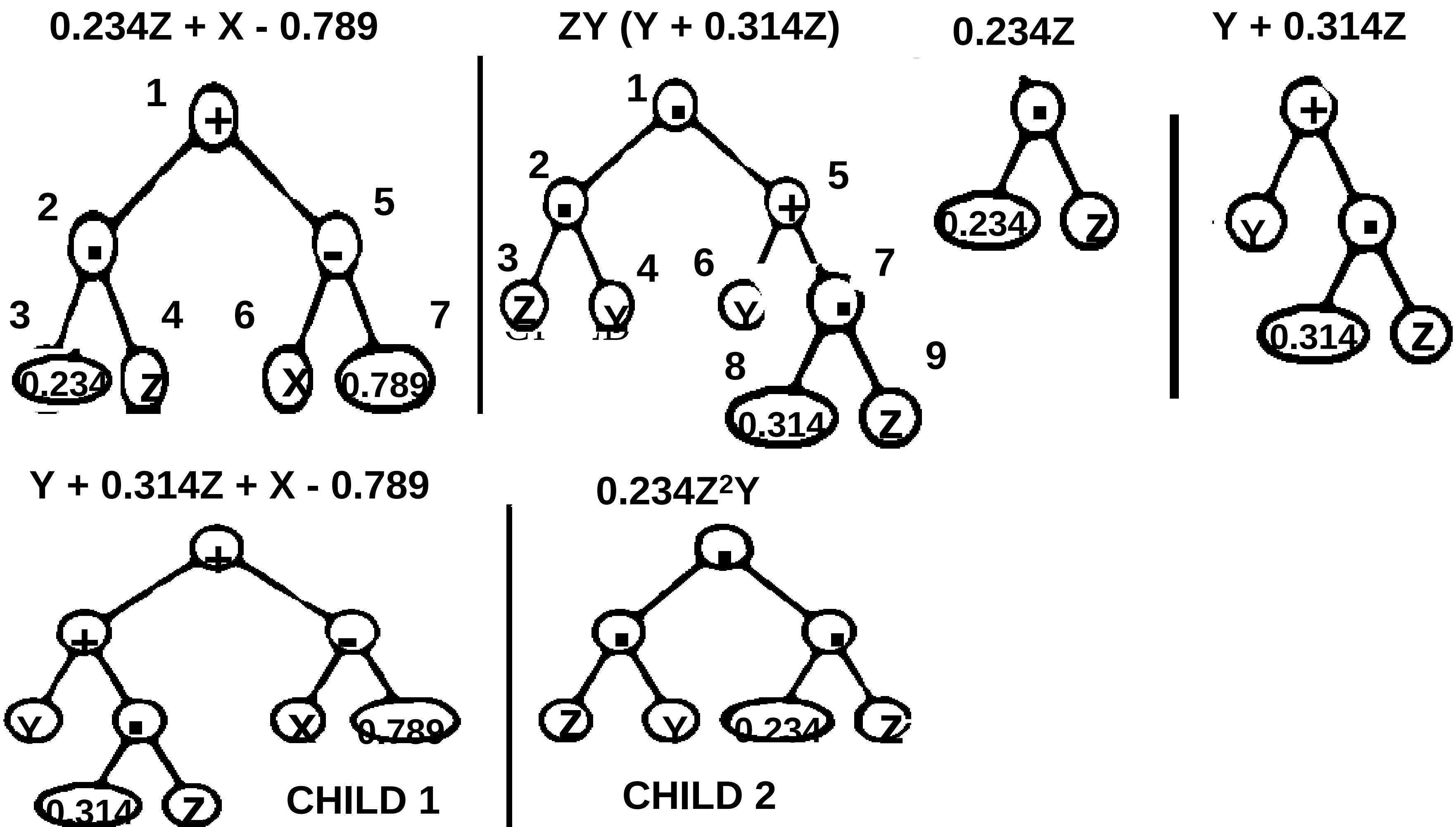
1. **Automatically Generating Computer Programs (Koza)**
2. **Prisoner's Dilemma (Axelrod)**
3. **Designing a Sorting Network Using Diploid Chromosomes (Hillis)**
4. **Prediction of Protein Secondary Structure (Koza)**
5. **Multi-objective Optimization (Cedeno and Vemuri)**
6. **DNA Fragment Assembly (Cedeno and Vemuri)**

WHAT IS GENETIC PROGRAMMING?

- THINK OF EACH SOLUTION IN THE GA AS A COMPUTER PROGRAM
- START WITH A FAMILY OF RANDOMLY GENERATED COMPUTER PROGRAMS
- RUN EACH PROGRAM AND DETERMINE HOW WELL IT SOLVES THE PROBLEM AT HAND
- CROSSOVER AND MUTATE PROGRAMS TO GENERATE OFFSPRING PROGRAMS
- AFTER A FEW GENERATIONS ENJOY A CORRECT COMPUTER PROGRAM

CROSSING COMPUTER PROGRAMS

PARENT 1 | PARENT 2 | CROSSOVER FRAGMENTS



PROTEIN STRUCTURE PROBLEM

- PROBLEM: LOCATE TRANSMEMBRANE SEGMENTS OF BACTERIORHODOPSIN
- TM PROTEINS CROSS CELL WALLS SEVERAL TIMES
- SHORT LOOPS ON EITHER SIDE OF THE MEMBRANE
- GOAL: IDENTIFY SEGMENTS THAT ARE WITHIN THE MEMBRANE, THE TRANSMEMBRANE DOMAINS
- OBJECTIVE: EVOLVE A COMPUTER PROGRAM FOR PREDICTING WHETHER OR NOT A GIVEN SEGMENT LIES IN TM DOMAIN

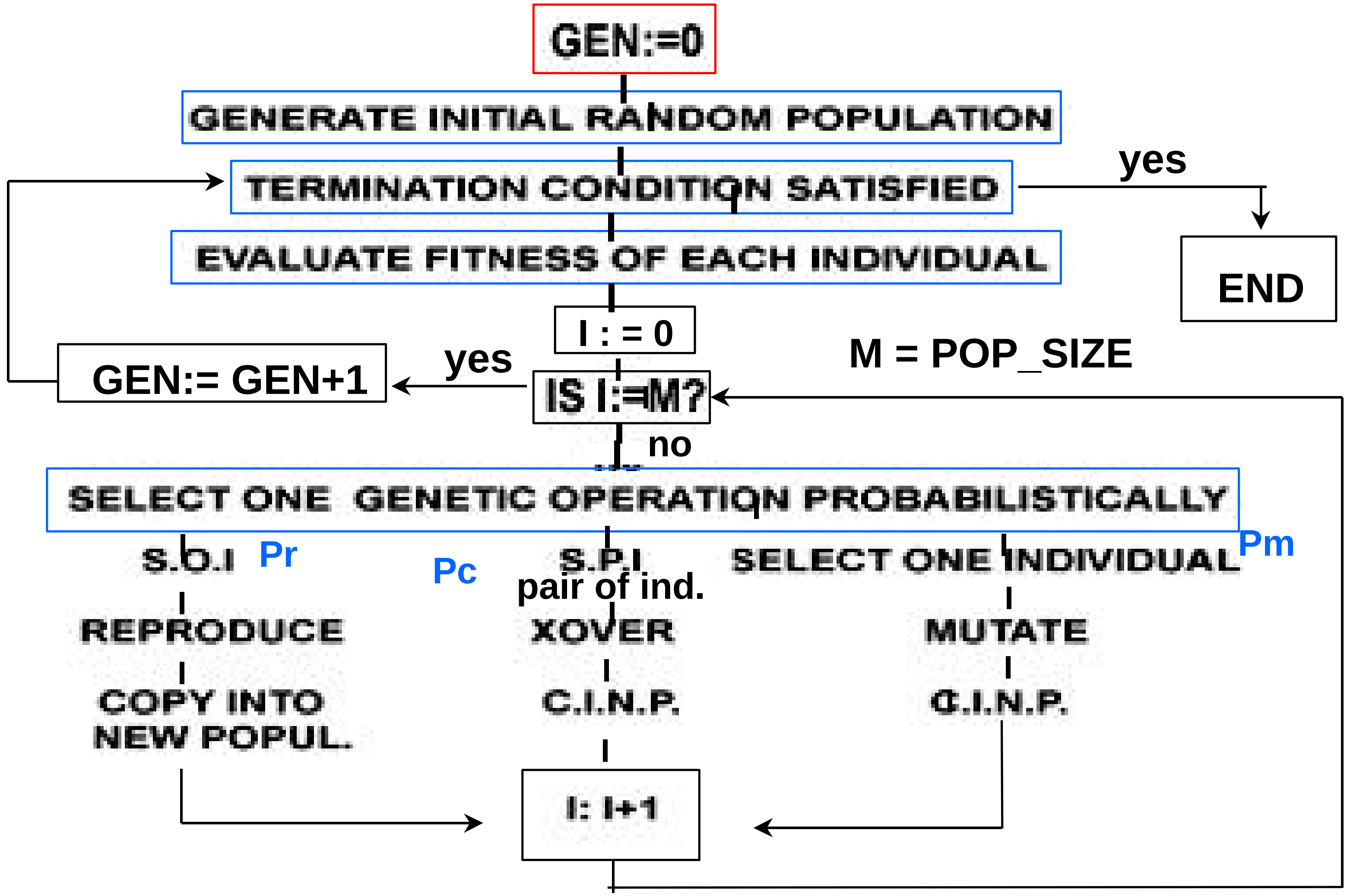
WINNING PROGRAM

```
(prog (looping-over-residues
        (SETM3 (- (+ (- (F?) (K?) )) (+ (-M3 (P?))
         (+ (I?) (SETM2 (SETM3 (L?)))))) (SETM2 (SETM2 (H?))))))
   (values (* (IFLTE (IFLTE (+ -5.606 M3) (* L M2) (% -2.786
          (IFLTE M1 M3 M2 M2)) (* M2  M0 ) (*% (+M2 M3)
          (+M3  L) (%M2  L) )) (* (+ (+M2 M1) (*M2 M0))
          (%M2 M2))))
```

PROJECTS AT DAS/LLNL

- GP WITH SISAL, A PARALLEL LANGUAGE
- GA TO TRAIN NEURAL NETS
- GA TO PROTEIN STRUCTURE PREDICTION
- GA TO DNA SEQUENCING ANALYSIS
- GA TO GROUND WATER REMEDIATION
- GA TO INFORMATION FILTERING
- GA TO COMMUNICATIONS AND NETWORKING

A FLOW CHART



GEN:=0
GENERATE INITIAL RANDOM POPULATION
TERMINATION CONDITION SATISFIED
yes
END
EVALUATE FITNESS OF EACH INDIVIDUAL
I : = 0
M = POP_SIZE
GEN:= GEN+1
yes
IS I:=M?
no
SELECT ONE GENETIC OPERATION PROBABILISTICALLY
S.O.I
Pr
Pc
S.P.I
pair of ind.
SELECT ONE INDIVIDUAL
Pm
REPRODUCE
XOVER
MUTATE
COPY INTO NEW POPUL.
C.I.N.P.
C.I.N.P.
I: I+1

SUMMARY

- GAs ARE RANDOMIZED OPTIMIZATION PROCEDURES
- GAs ARE EASY TO LEARN AND IMPLEMENT
- GAs CAN BE BE APPLIED TO A WIDE RANGE OF PROBLEMS
- GAs CAN BE COMBINED WITH NEURAL NETS
- GAs CAN BE USED TO AUTOMATICALLY GENERATE CORRECT COMPUTER PROGRAMS

REFERENCES

- **Holland, J. Adaptation in Natural and Artificial Systems, MIT Press, 1975**
- **Proceedings of the International Conferences on Genetic Algorithms, Lawrence Erlbaum Assoc. 1985, 87, 89, 91**
- **Davis, L. Genetic Algorithm and SimulAted Annealing, Pitman, 1987**
- **Goldberg, D. Genetic Algorithms in Optimization, Search, and Machine Learning, Addison Wesley, 1989**
- **Davis, L. Handbook of Genetic Algorithm, Van Nostrand.1990**
- **Koza, J. Genetic Programming: On the Programing of Computers by Means of Natural Selection, MIT Press**
- **Koza, J. Genetic Programming II: Automatic Discovery of Reusable Programs, MIT Press, 1994**