

On Extended Compact Genetic Algorithm (ECGA)

Kumara Sastry David E. Goldberg

Illinois Genetic Algorithms Laboratory
University of Illinois at Urbana-Champaign
<http://www-illigal.ge.uiuc.edu>



Genetic and Evolutionary Computation Conference
July 8-12, 2000
Las Vegas, Nevada

Outline

- Motivation
- Overview of ECGA
- Test problems: Onemax, Trap, Folded Trap
- Results on test problems
- A simple power cycle
- Results on power cycle optimization
- Conclusions

Motivation

- **Background**

- Increased interest in competent GAs
- An interesting competent GA is Harik's ECGA
 - * builds models of good data as linkage groups
- Under-investigated
 - * Does it match GA theory? Not tried on real problems.

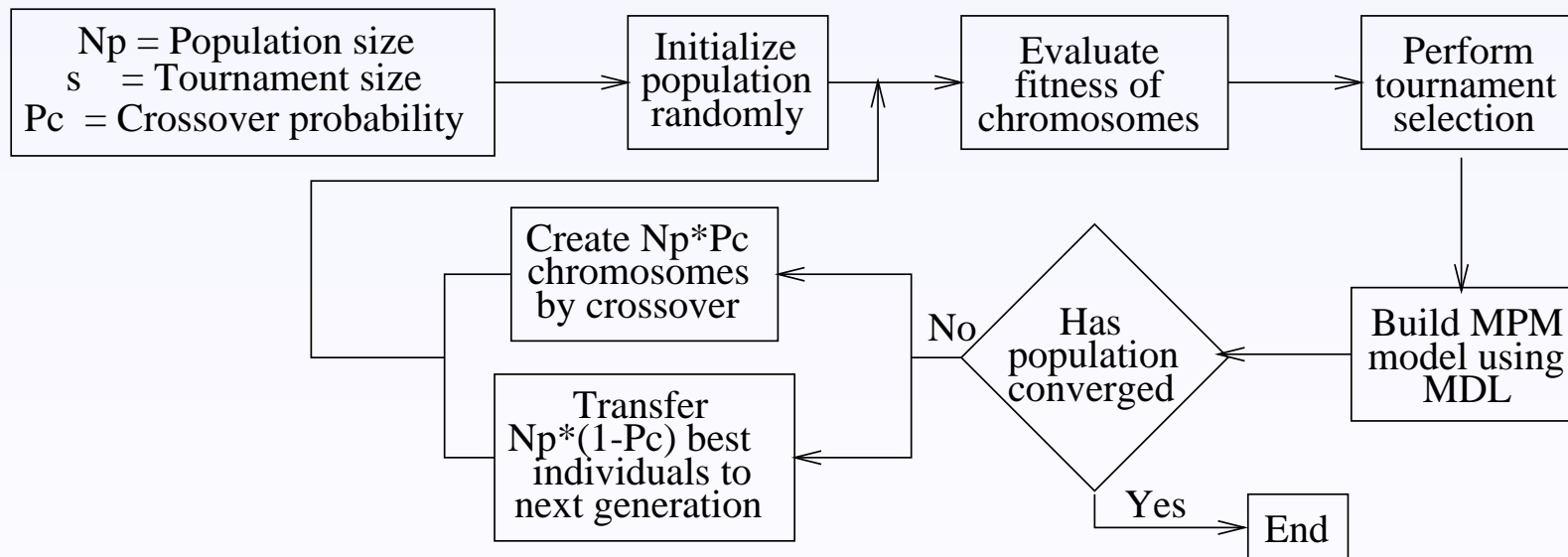
- **Rhetorical Purpose**

- Overview of ECGA
- Investigate connection to quality-duration theory
- Apply to practical problem

Overview of ECGA

- **Key Idea:** Good probability distribution \equiv *Linkage learning*
- Probability distribution: *Marginal Product Models* (MPM)
- Quantified based on *Minimum Description Length* (MDL)
- **MDL Concept:** Simpler distributions are better

Flowchart of ECGA



Marginal Product Models

- Product of marginal distributions on a partition of genes
- Similar to CGA (Harik et. al;1998) and PBIL(Baluja;1994)
- Represent more than one gene in a partition
- Make exposition simpler
- Gene partition maps to linkage groups

Minimum Description Length Models

- **Hypothesis:** Good distributions are those for which
 - Representation of the distribution is minimum (*model complexity*, C_m).
 - Representation of population compressed is minimum (*compressed population complexity*, C_p).
- Penalize complex models
- Penalize inaccurate models
- Combined complexity, $C_c = C_m + C_p$

Building MPM using MDL

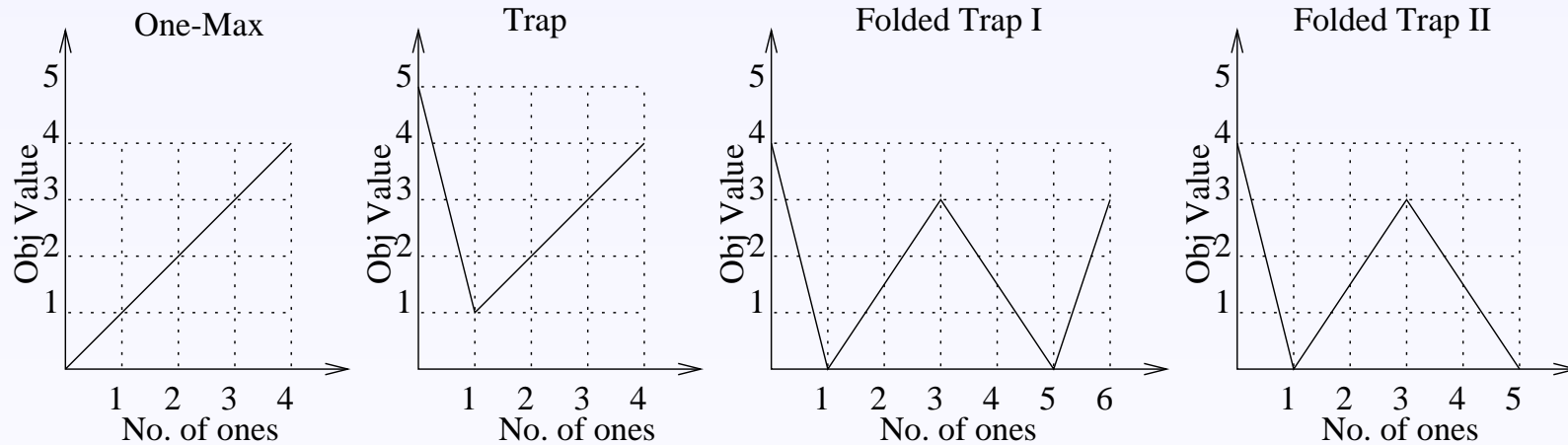
Uses a steepest ascent search:

1. Assume all the genes to be independent ($[1],[2],\dots,[L]$) and compute C_c .
2. Form all possible combinations ($N_{bb}(N_{bb}-1)/2$) of merging two subsets. eg., ($[1,2],[3],\dots,[L]$), \dots , ($[1],[2],[3],\dots,[L-1,L]$).
3. Select the set with minimum combined complexity (C'_c).
4. If $C_c > C'_c$ go to step 6.
5. Use the set with C'_c as the current MPM and go to step 2.
6. Merging is not possible, exit with set from step 2 as MPM.

Generation of New Population

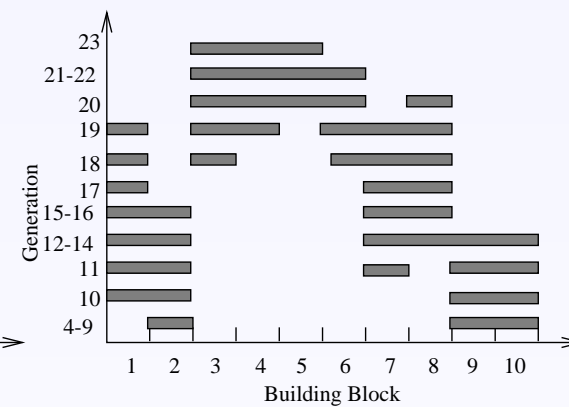
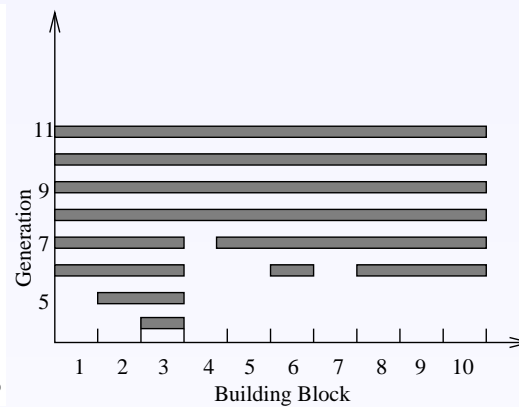
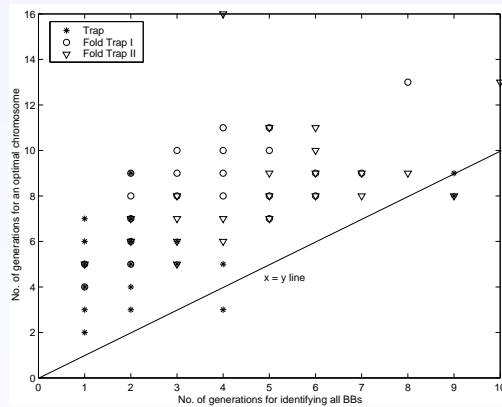
- Transfer $N_p * (1 - P_c)$ best individuals to the next generation
- The rest $N_p * P_c$ individuals are generated as follows:
 - Take each of the subset of MPM from one of the individuals.
 - Similar to multiple point crossover.
 - Number of crossover points = N_{bb} .
 - Instead of two parents we have N_{bb} parents.

Test Problems



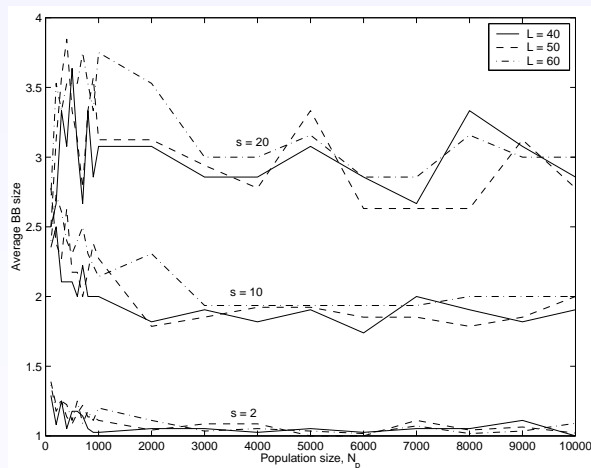
- Uniformly scaled, building blocks (BBs) are known
- Trap, FT-I, FT-II: 10 BBs, Onemax: 40,50,60 bits
- Tournament sizes: 2-20, Population Size = 50-25000

Results: Building Block Identification

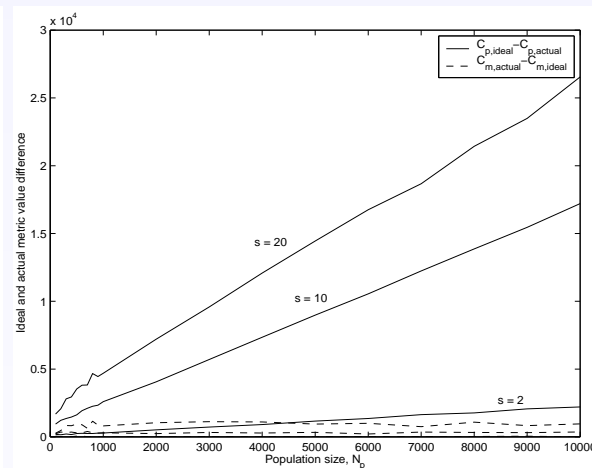


- BB identification is necessary
- BB identification modes: *Flash* and *Sequential*
- Flash: smaller BB size, large tournament sizes
- Sequential: larger BB size, small tournament sizes

Results: Onemax BB Identification



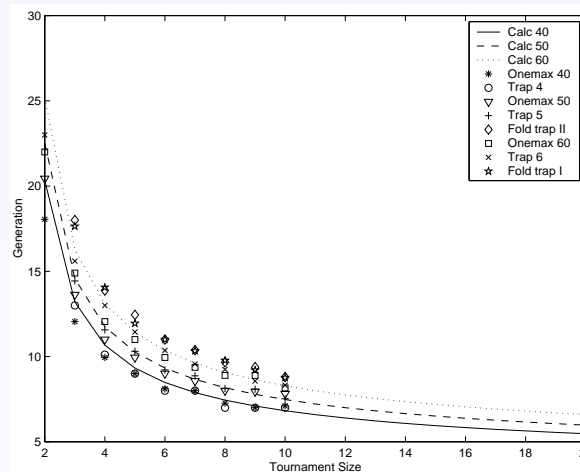
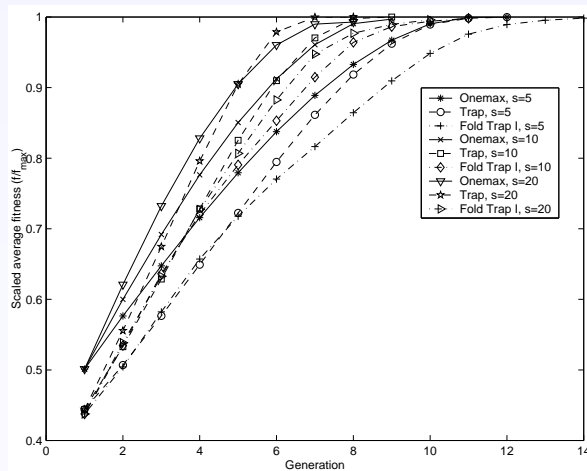
(a)



(b)

- *Apparent linkage or hitchhiking*
- A good random number generator reduces linkage

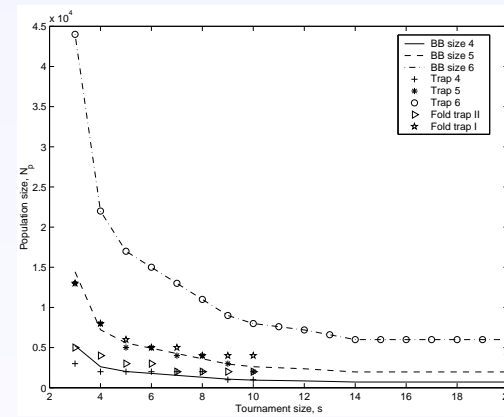
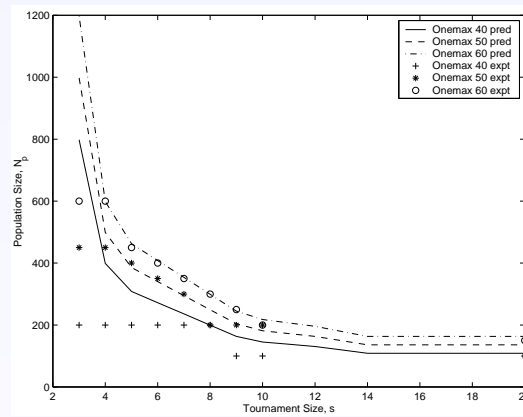
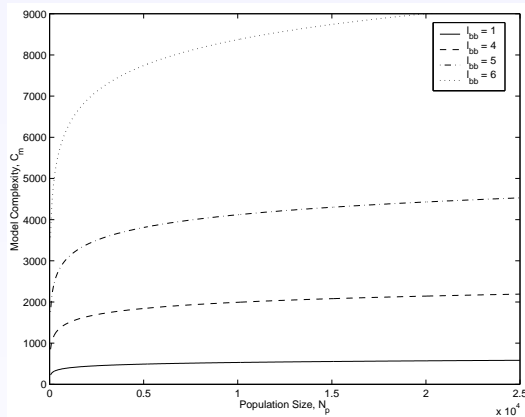
Results: Convergence Time



$$t_{\text{conv}} = 1 + k_c \left(\frac{\log_2 L + \log_2 (\log_2 L)}{\log_2 s} \right) \quad (L < 100)$$

$$t_{\text{conv}} = \frac{\pi \sqrt{L}}{2I} \quad (L > 100)$$

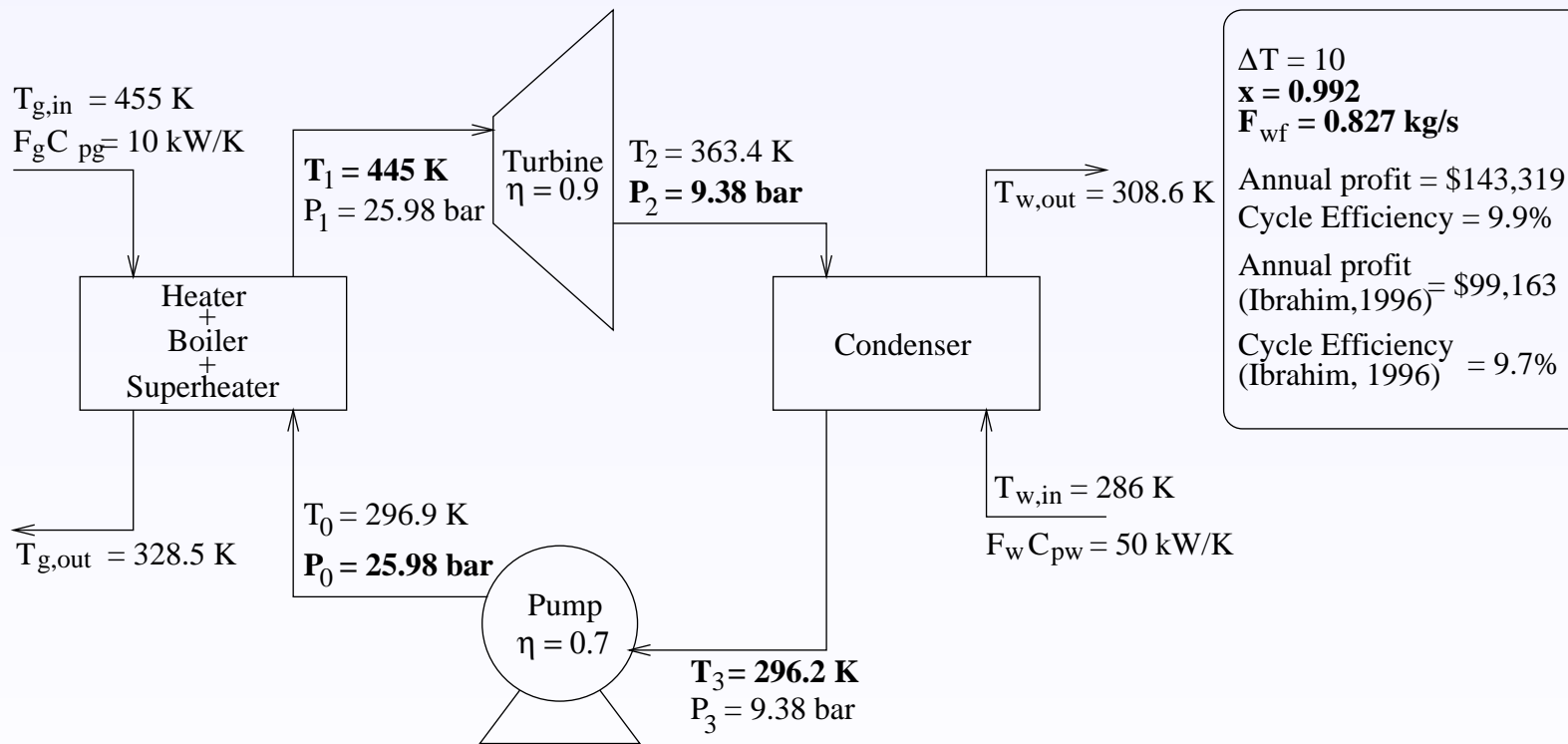
Results: Population Sizing



$$N_p = \frac{N_{bb} (2^{l_{bb}} - 1)}{\epsilon \log(2)} \quad \epsilon = \frac{\min \{0.1515, c_1 s + c_2\}}{3.5 + 0.5 (l_{bb} - 1)}$$

Similar to equations by (Goldberg et. al;1992, Pelikan et. al; 2000)

A Simple Power Cycle

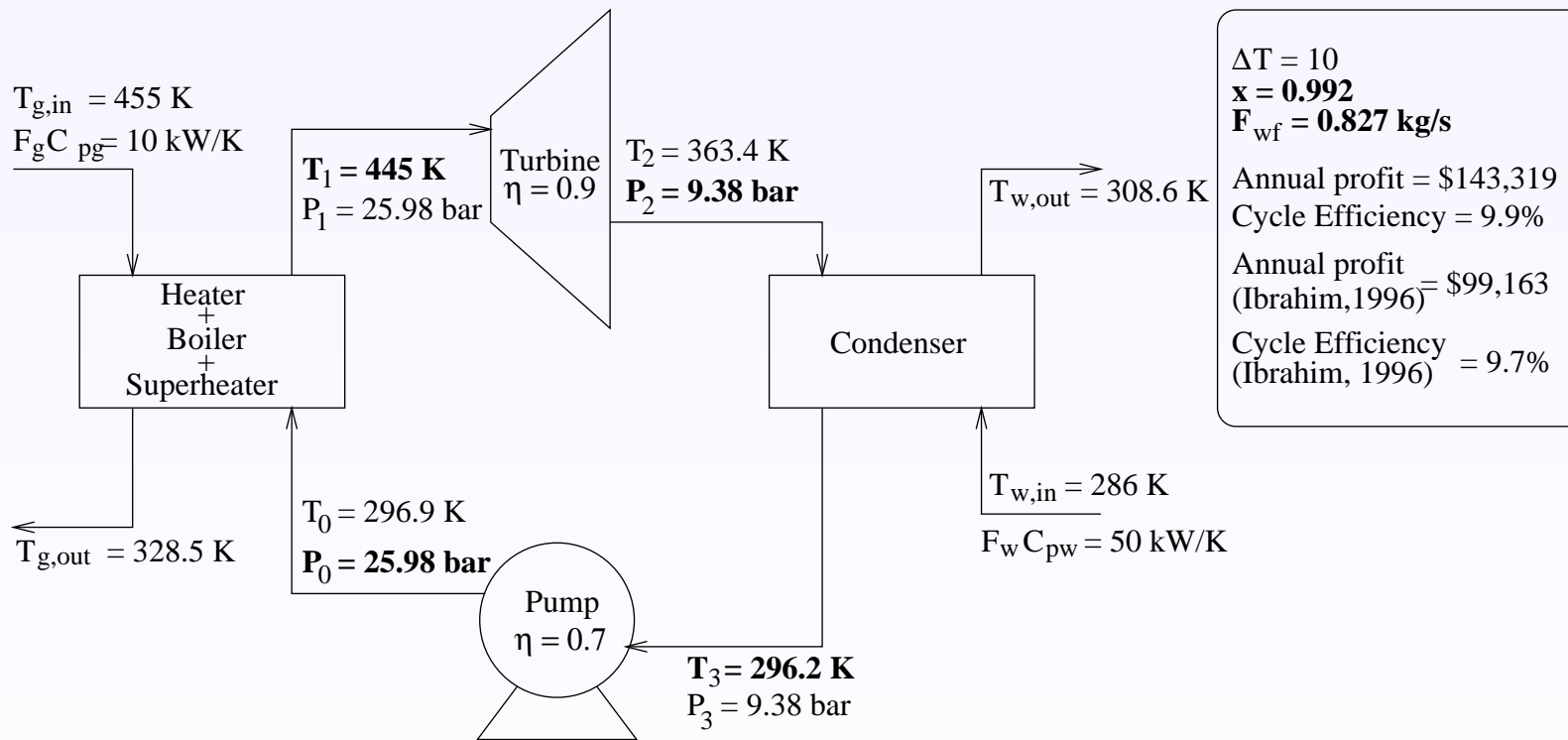


$\Delta T = 10$
 $x = 0.992$
 $F_{wf} = 0.827 \text{ kg/s}$
 Annual profit = \$143,319
 Cycle Efficiency = 9.9%
 Annual profit (Ibrahim, 1996) = \$99,163
 Cycle Efficiency (Ibrahim, 1996) = 9.7%

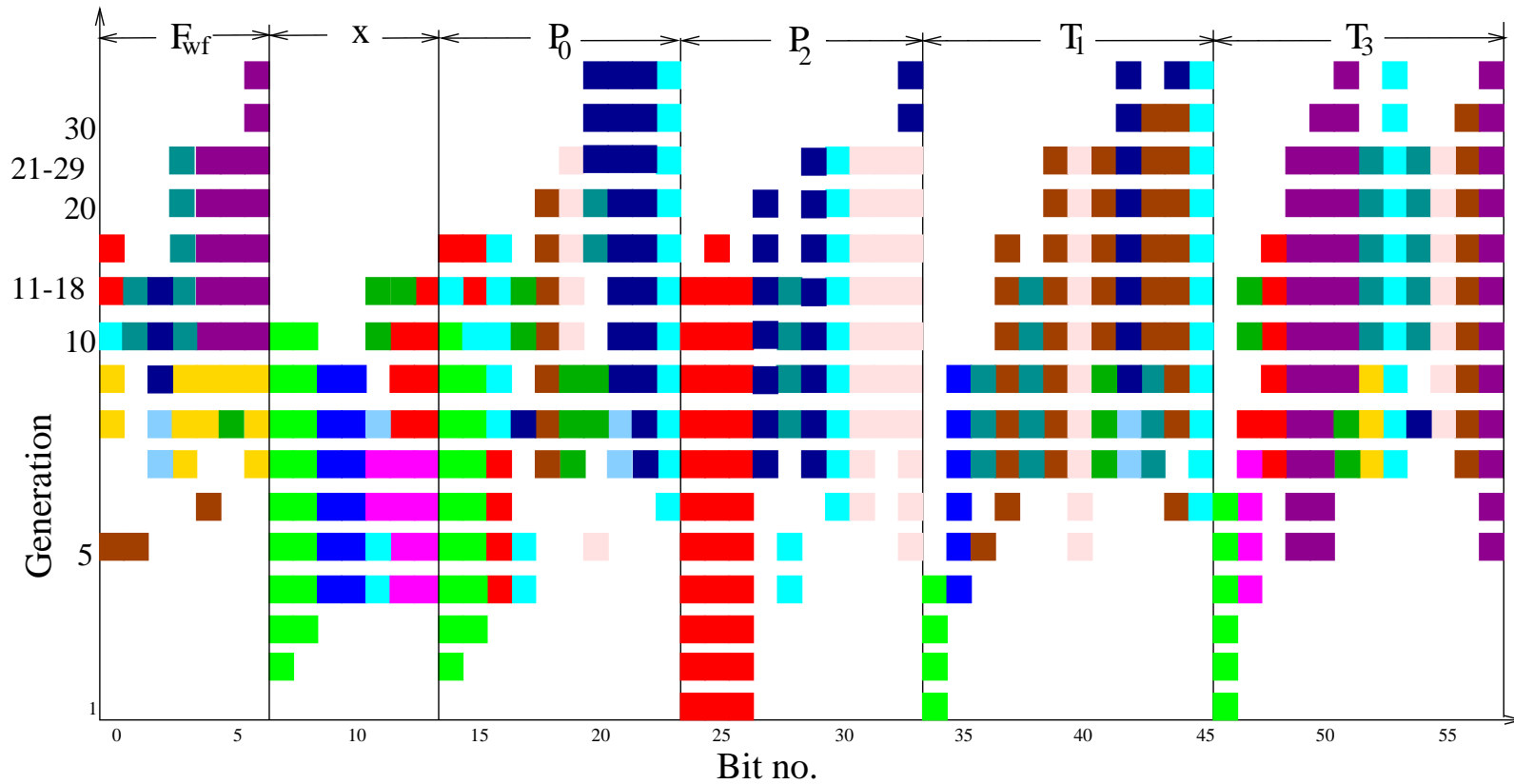
Power Cycle Problem Formulation

- Maximize profit: $C_t |\dot{W}_t| - C_p |\dot{W}_p| - C_w |\dot{Q}_c|$
- 2nd law of thermodynamics: $T_{cold} \leq T_{hot} - \Delta T_{min}$
- Fluid in/out of turbine **must** be a vapor
- Fluid out of condenser and in/out of pump **must** be a liquid
- Constraints are handled by penalty method
- Highly nonlinear & complex thermodynamics
(Ibrahim and Klein; 1993)

Results: Power Cycle Optimization



Results: Power Cycle Optimization



Conclusions

- Important insights into working of ECGA
- Empirical relations for convergence time and population sizing
- Effective in solving practical problem

Acknowledgments

- Air Force Office of Scientific Research, Air Force Materiel Command, USAF, under grants F49620-94-1-0103, F49620-95-1-0338, F49620-97-1-0050, and F49620-00-0163.
- National Science Foundation under grant DMI-9908252.
- U. S. Army Research Laboratory under the Federated Laboratory Program, Cooperative Agreement DAAL01-96-2-0003.