Study on Genetic Algorithm(GA) Improvement and Application

Yao Zhou

05/01/06

Advisor:
Prof. Kevin(Yiming) Rong

Outline

- Background & Motivation
- Objective
- Introduction & Literature Review
- GA Improvement
- GA Application in Computer-Aided Tolerancing
- Conclusion & Future Work

Background

- Design optimization is everywhere
  - Optimize materials composition for better properties
  - Optimize manufacturing processes for lower cost
  - Optimize network architecture for higher communication rate

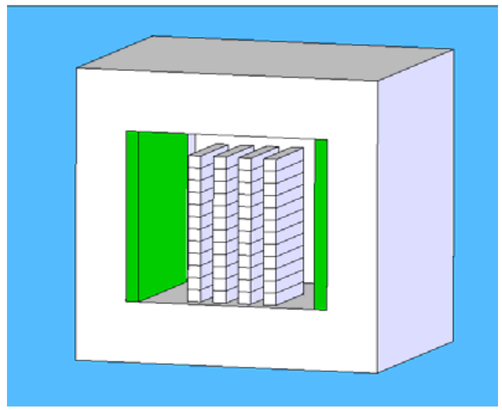
  ……

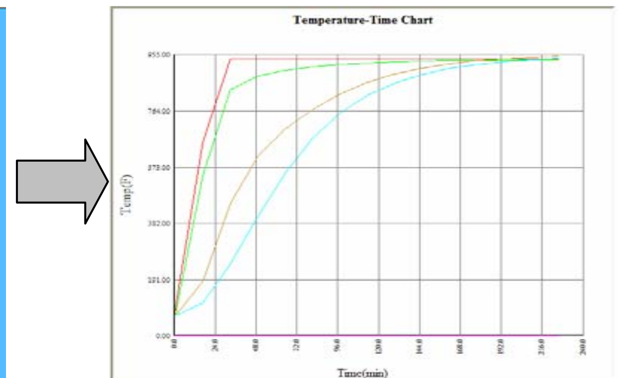
- Challenges in design optimization problems
  - Number of possible designs are infinite or huge
  - Relationship between design parameters ($x_1,\ldots,x_n$) and objective function $F(x_1,\ldots,x_n)$ is unknown or too complex

Background

Optimization of load arrangement in heat treating furnace

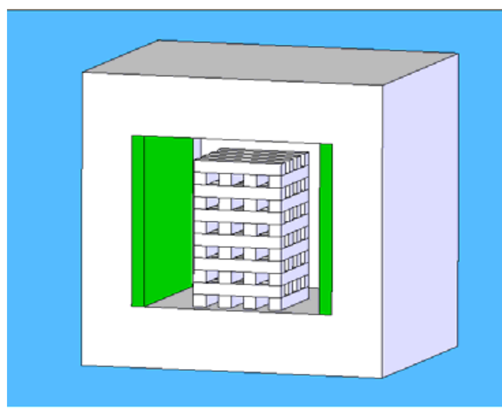
Objective:
Reduce
1. Cycle time
2. Energy consumption

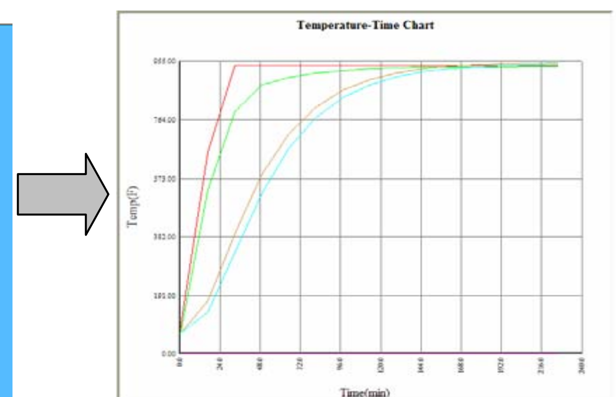




Load parameters:
1. Load pattern
2. Number of layers, rows and columns
3. Parts spacing
4. Parts orientation

……





Motivation

- **Challenges in load arrangement optimization**
  1. Number of possible load arrangements is huge
  2. Relationship between time/energy and load parameters is not explicit
  3. Computation time is very long for FDM-based simulation

Impossible for traditional optimization methods

- **Solutions**
  - For challenge 3, improvements need to be made on FD methods
  - **For challenges 1 & 2, alternative optimization method—GA—is studied**

Motivation

- Genetic Algorithm(GA)
  - Much more effective for large scale problems than traditional optimization methods
  - Able to handle implicit relationship between parameters and objective, which traditional methods are not

- Difficulties in Implementation
  - Run simulation thousands+ of times
  - FDM-based simulation time is too long

- However, GA is still applicable and powerful in many other manufacturing process optimization problems in general

Motivation

- For extremely large scale problems, standard GA is still not efficient enough.

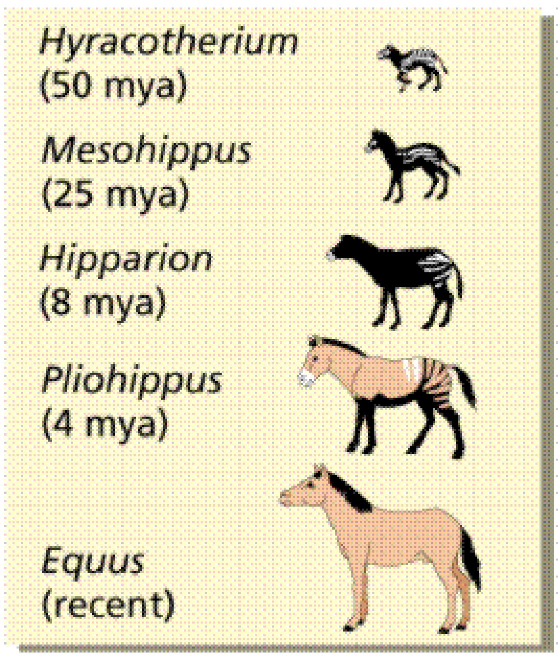
- To further reduce the computation time of optimization procedure, standard GA needs to be studied and improved.

Objectives

- Improve GA's computation efficiency
- Improve design of manufacturing processes using GA (Tolerance Assignment)

Introduction

- GA is a search and optimization algorithm inspired by the idea of natural selection, genetics and evolution
  - "favor of the fit individuals"



Introduction

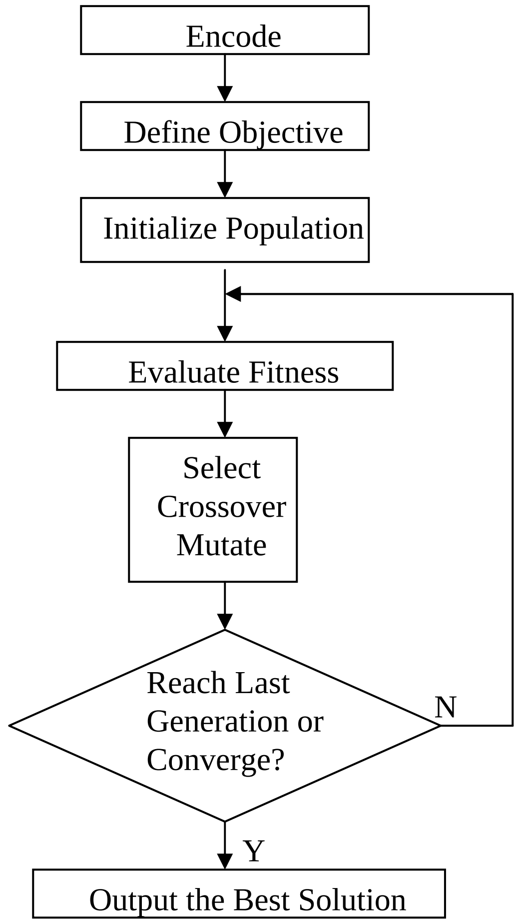
Natural Selection

- A population of organisms undergoes natural selection
- Natural selection identifies the fit and unfit individuals
- The fit thrive and have more offspring than the unfit
- Offspring inherit more genes from the fit, so offspring are also statistically fit
- The population evolve to be more and more fit

Genetic Algorithm

- A pool of candidate designs is evaluated by objective function
- Objective function evaluates and identifies good and bad designs
- Good designs thrive and more of their information is passed onto newly generated designs
- New designs inherit more information from the good designs, and new designs also statistically good
- The design pool evolve to be better

Introduction



Flowchart of standard GA

Introduction

GA Procedures

gene

chromosome

- Representation
  1. Encode designs $x=(010...110)$
  2. Define fitness function: $F(x) \leftrightarrow x$
- Initialization
  3. Initialize population: $P(x)=\{X_1, X_2 ... X_n\}$

Introduction

GA Procedures (cont.d)

- Iteration

4. Evaluate designs: F(X)=?

5. Selection: Adjust shares for x based on F(x)
(More shares for good chromosomes)

6. Crossover: 1001101 & 0101000

1001001 & 0101100

7. Mutation: 1011001→1010001

Literature Review

- Birth of GA (Holland 1975)

- Applications of GA
  - Job shop scheduling (Davern 1994)
  - Manufacturing cell design (Joine 1996)
  - Fixture design (Gold 1998)

  ……

Literature Review

- GA Improvement—What makes GA more efficient?
  - Parallel GA (Cantu-Paz 1999)
  - Incorporate problem-specific knowledge (Sinha 2002)
  - Choose population size (Srivastava 2002)
  - Approximate fitness functions (Jin 2000)
- Gene Interaction—What keeps GA from being efficient?
  - Define global gene interaction (Davidor 1990)
  - Linkage learning (Harik 1997)

Literature Review

- Summary

  – GA has proved successful for large scale design optimization problems in many applications, but it still needs improving

  – No methods have attempted to improve GA efficiency by <u>*predicting potentially good designs/chromosomes*</u>

  – A relatively simple method is needed to <u>*characterize the interaction between specific genes*</u>, and <u>*guide GA operators design*</u>

GA Improvement (I): Predicting Fit Chromosome

- Inspiration

  The evolution process of the nature takes millions of years. If we can predict the gene configuration of potentially fit individuals based on the information from current population, and create such individuals, they will thrive through natural selection and provide more fit genes to later generations, which will facilitate the evolution process toward the fit direction.

GA Improvement (I): Predicting Fit Chromosome

- Rationale

  1. ***Establish the relationship*** between the relative fitness of a gene segment and the change in its number of occurrence between generations

  2. ***Identify potentially fit gene segments*** by tracking changes in their number between generations

  3. ***Link the fit genes*** and we get potentially fit chromosome

Predicting Fit Chromosome

- Schema: A gene segment containing one or more genes

(1001000)→(***1***)   (1001000)→(**01*0*)

1. ***Schema Theorem*** (Holland 1975): Predict the number of a schema in the next generation based on
   – its number in this generation
   – its relative fitness
   – its length

$$E[N(S,t+1)] \geq \{\underbrace{1-\chi\frac{l(S)}{l-1}}_{\text{crossover effect}} - \underbrace{\mu k(S)}_{\text{mutation effect}}\}\underbrace{r(S,t)}_{\text{selection effect}}N(S,t)$$

-E[N(S,t+1)]: expected number of schema S at generation t+1

-N(S,t): number of schema S at generation t

-r(S,t): fitness of schema S relative to average fitness

-l(S), k(S): information regarding schema length

Predicting Fit Chromosome

When schema length=1 (single-digit gene), Schema Theorem is reduced to:

$$r(S,t) \approx \frac{N(S,t+1)}{N(S,t)\{1-\mu\}}$$

r(S,t): fitness of schema S relative to average fitness

N(S,t): number of schema S at generation t

N(S,t+1): number of schema S at generation t+1

μ : mutation rate, constant

Predicting Fit Chromosome

Relative fitness of a schema ⟺ Change in numbers of the schema between two generations

Observation: A potentially fit one-digit schema is the one whose number increases by the largest percentage from last generation

For the first digit, "0" is potentially better than "1" ⟺

generation t		generation t+1
(10100)		(00100)
(10010)	→	(01010)
(11001)		(11001)
(00100)		(00100)

Predicting Fit Chromosome

2. Building Block Hypothesis (Goldberg 1989)

   – A fit chromosome is constructed by combining fit and short schema

Summary of above analysis:

- From Schema Theorem, we can predict the potentially fit short schema by tracking their number of occurrence between generations

- From Building Block Hypothesis, we know linking short fit schema provides potentially fit chromosome

Evolution Direction Guided-GA

- Step 1: For each digit, find out the gene whose number increases the most from last generation
- Step 2: Create a chromosome, whose value at each digit is set according to the values found in step 1
- Step 3: Replace the worst chromosome in current generation with this chromosome
- Step 4: Continue other operations in GA

Test Function

- One-Max Problem (A Standard Test Function, DeJong 1975)
  Fitness value of a chromosome X= the number of “1”s in X, i.e. the more “1”s a chromosome has, the better it is

- 10-digit One-Max problem, Population=20
  - Theoretical optimum: (1111111111)

- Calculation Results
  Number of generations to find the optimum
  - Standard GA: 16 generations
  - EDG-GA: 5 generations

GA Improvement (II): Gene Interaction Characterization

- Problems with Gene Interaction
  - Standard GA chooses crossover point randomly
  - Interactive genes may be separated during crossover
    e.g.: Neither (1****) nor (*0***) is good, only (10***) is good, but (1|0***) happens during crossover
  - Separation of interactive genes will generate bad designs and slow down evolution.

- A method is needed to identify interdependent gene clusters and preserve them during crossover.

Gene Interaction Characterization

- Step1: Define gene's contribution to chromosome fitness

Gene-Chromosome Correlation (GCC) Function

$$GCC(i) = \frac{\sum_{X \in P(X)} [a_i(X) - \overline{a_i}] \cdot [f(X) - \overline{f}]}{\sigma[a_i(X)] \cdot \sigma[f(X)]}$$

- *GCC(i)*: correlation between the *i*th gene and the fitness of chromosome
- $a_i(X)$: gene value at location *i* in chromosome X, either 0 or 1
- $\overline{a_i}$: average gene value at location *i*, a value between 0 and 1
- $f(X)$: fitness function of chromosome X
- $\overline{f}$: average fitness of all the chromosomes in population P(X)
- $\sigma[a_i(X)]$: standard deviation of $a_i(X)$
- $\sigma[f(X)]$: standard deviation of $f(X)$

Gene Interaction Characterization

- Step2: Define gene interaction *R(i, j)* between genes at locations i and j

$$R(i, j) = \sigma[GCC(i)]|_{a(j)} / 2 \qquad (j \neq i)$$

$\sigma[GCC(i)]|_{a(j)}$ standard deviation of GCC(i) when gene j is set to different values

Meaning of *R(i, j)*: When another gene changes, how much this gene's contribution to the chromosome fitness (GCC) will change.

If *R(i, j)* is large, then the degree of interaction between genes at locations i and j is large.

Gene Interaction Characterization

- Step3: Create dependency matrix, and threshold the matrix to identify large interactions between genes

$$
\begin{array}{c|cccc|}
 & Gene1 & Gene2 & \cdots & Genel \\
\hline
Gene1 & 1 & R(1,2) & \cdots & R(1,l) \\
Gene2 & R(2,1) & 1 & \cdots & R(2,l) \\
\vdots & \vdots & \vdots & \vdots & \vdots \\
Genel & R(l,1) & R(l,2) & \cdots & 1
\end{array}
\qquad
\begin{bmatrix}
1 & & 1 & & 1 & & & 1 \\
 & 1 & & 1 & & & 1 & \\
1 & & 1 & & 1 & & & 1 \\
 & 1 & & 1 & & & 1 & \\
1 & & 1 & & 1 & & & 1 \\
 & & & & & 1 & & \\
 & 1 & & 1 & & & 1 & \\
1 & & 1 & & 1 & & & 1
\end{bmatrix}
$$

Threshold *R(i,j)*

to 0 and 1

Gene Interaction Characterization

- Step4: Identify interdependent gene clusters and choose crossover points outside clusters

$$
\begin{array}{c}
 \\
A \\ B \\ C \\ D \\ E \\ F \\ G \\ H
\end{array}
\begin{array}{c}
\begin{array}{cccccccc} A & B & C & D & E & F & G & H \end{array} \\
\begin{bmatrix}
1 & & 1 & & 1 & & & 1 \\
 & 1 & & 1 & & & 1 & \\
1 & & 1 & & 1 & & & 1 \\
 & 1 & & 1 & & & 1 & \\
1 & & 1 & & 1 & & & 1 \\
 & & & & & 1 & & \\
 & 1 & & 1 & & & 1 & \\
1 & & 1 & & 1 & & & 1
\end{bmatrix}
\end{array}
\qquad
\begin{array}{c}
 \\
B \\ D \\ G \\ A \\ C \\ E \\ H \\ F
\end{array}
\begin{array}{c}
\begin{array}{cccccccc} B & D & G & A & C & E & H & F \end{array} \\
\begin{bmatrix}
1 & 1 & 1 & & & & & \\
1 & 1 & 1 & & & & & \\
1 & 1 & 1 & & & & & \\
 & & & 1 & 1 & 1 & 1 & \\
 & & & 1 & 1 & 1 & 1 & \\
 & & & 1 & 1 & 1 & 1 & \\
 & & & 1 & 1 & 1 & 1 & \\
 & & & & & & & 1
\end{bmatrix}
\end{array}
$$

Matrix transformation

Identified interdependent gene clusters: (BDG),(ACEH),F

GA Improvement Summary

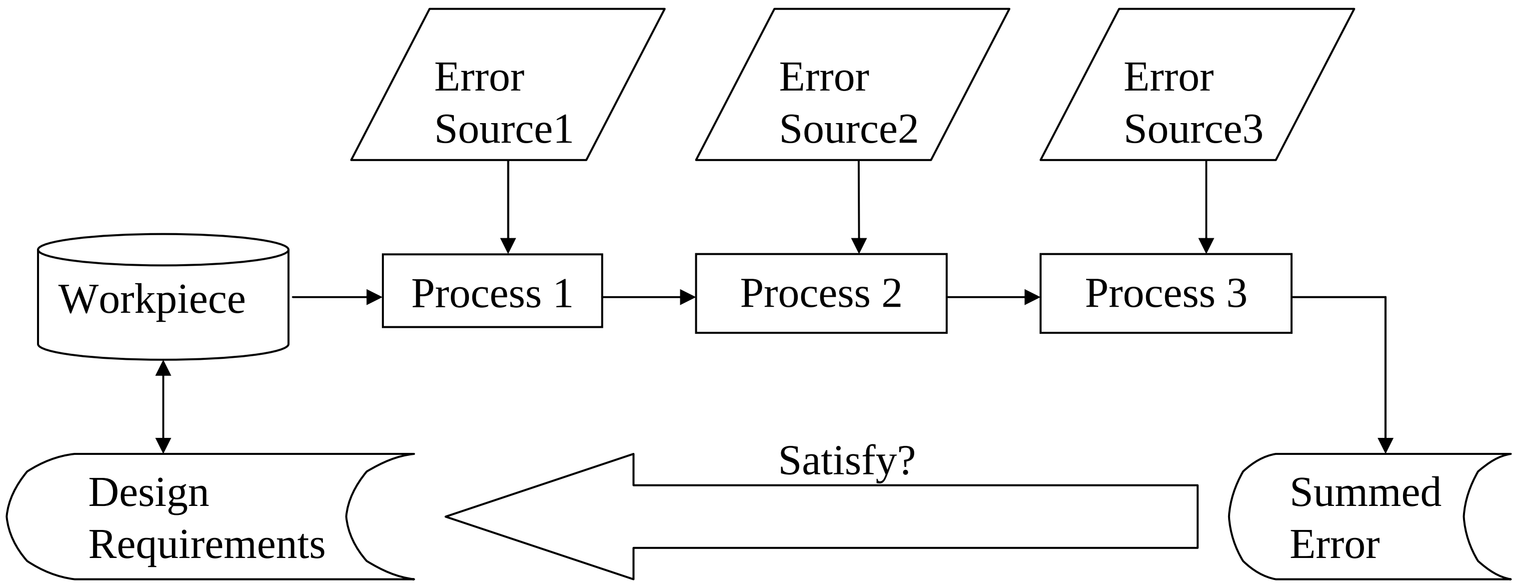
- Proposed EDG-GA is able to improve GA efficiency by *predicting potentially good designs based on non-problem-specific knowledge*
- A method is developed to *characterize the interaction between specific genes*, and *guide GA crossover operators design*

GA Application in Computer-Aided Tolerancing

- Tolerance: Permissible error range of dimension
- Design Tolerance: Error range allowed by design requirements
- Operational Tolerance: Error range allowed for a manufacturing process
- International Tolerance (IT) Grades: Operational tolerances are discretized to 18 grades.

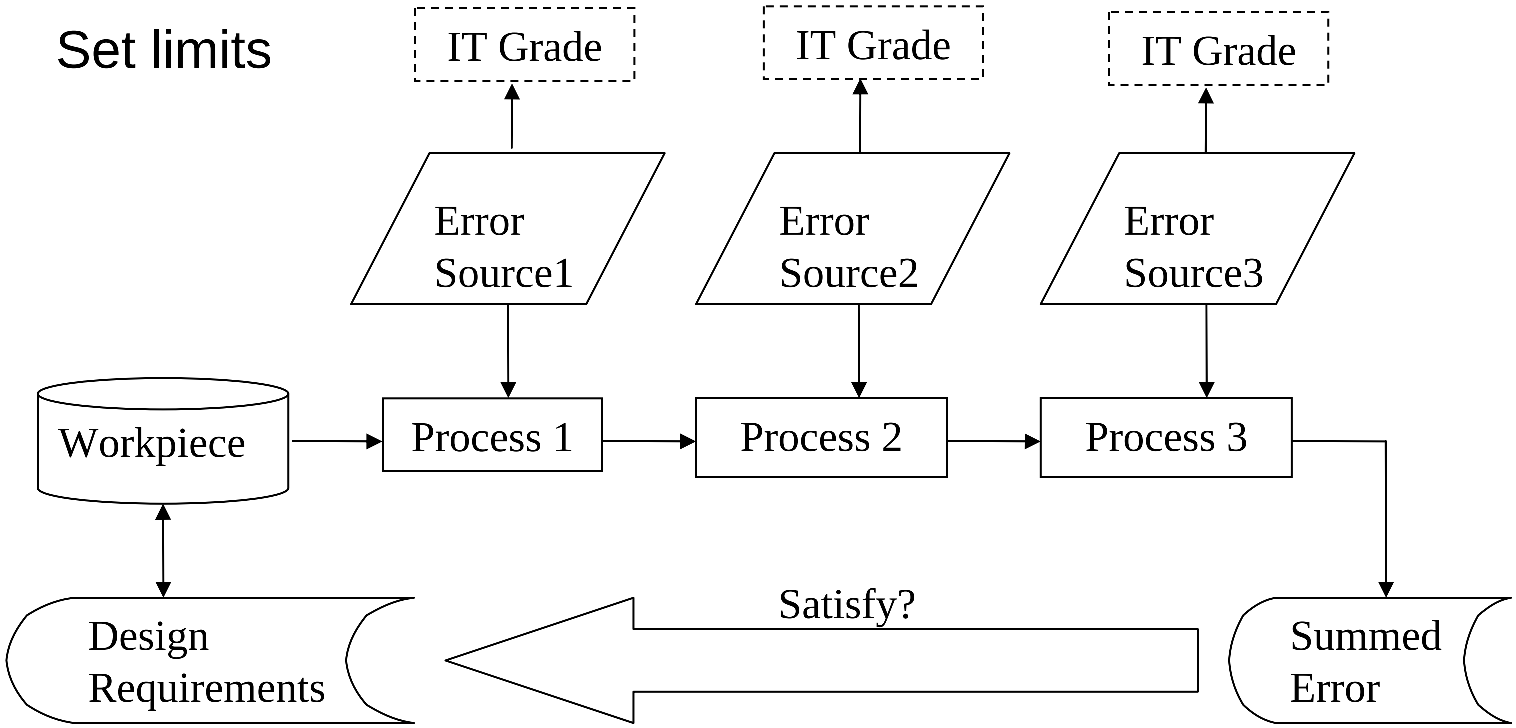
  (IT grade “1”—tightest; IT “grade” 18—loosest)

Tolerance Assignment



Tolerance Assignment

Set limits



Problem Definition

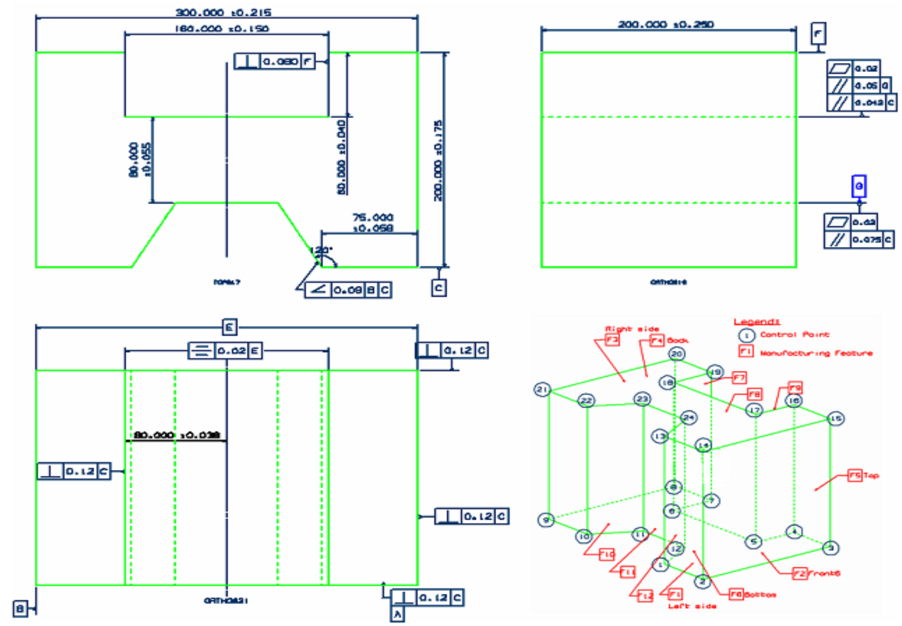
min Cost Function F[IT(1), IT(2)…IT(n)]

s.t.:

1. Operational tolerence stack-up <= design tolerance
2. The IT grades are feasible for the given processes

Case Study

- 12 processes,
  i.e. 12 IT grades to determine: IT(1),...,IT(12)
- Each IT grade is between IT grades 5 and 9,
  i.e. IT(i) $\in$ {5, 6, 7, 8, 9}
  $i = 1,2,\ldots,12$



- Cost Function:

$$C(F_i) = \alpha \cdot \beta \cdot f_1(F(i)) \cdot V(F(i)) \cdot \exp(a \cdot IT(i))$$

α material machinability factor

β feature complexity factor

a constant

$f_1(F(i))$ feature type factor

$V(F(i))$ feature size factor

GA Solution

- GA Representation
  1. Chromosome X=[IT(1), IT(2), ..., IT(12)],
     IT(i) $\in$ {5, 6, 7, 8, 9} $i = 1, 2, \ldots, 12$
  2. Fitness Function F(X): inverse Cost Function (fit when cost is low)
- Population Initialization
  3. Randomly create 50 chromosomes

GA Solution

- GA Iteration (20 generations)
  4. Evaluation: calculate F(X)
  5. Selection
  6. Crossover
  7. Mutation

generation t	evaluation	selection	crossover	mutation
(989797578867)	F(X)=0.9	(989797578867)	(989797578867)	(989887858867)
(698978756998)	F(X)=0.5	(989797578867)	(967887857569)	(989887878867)
(967887857569)	F(X)=0.6	(698978756998)	(989887858867)	
(785676758567)	F(X)=0.2	(967887857569)	(967797577569)	
……	……	……		

generation t+1 ……

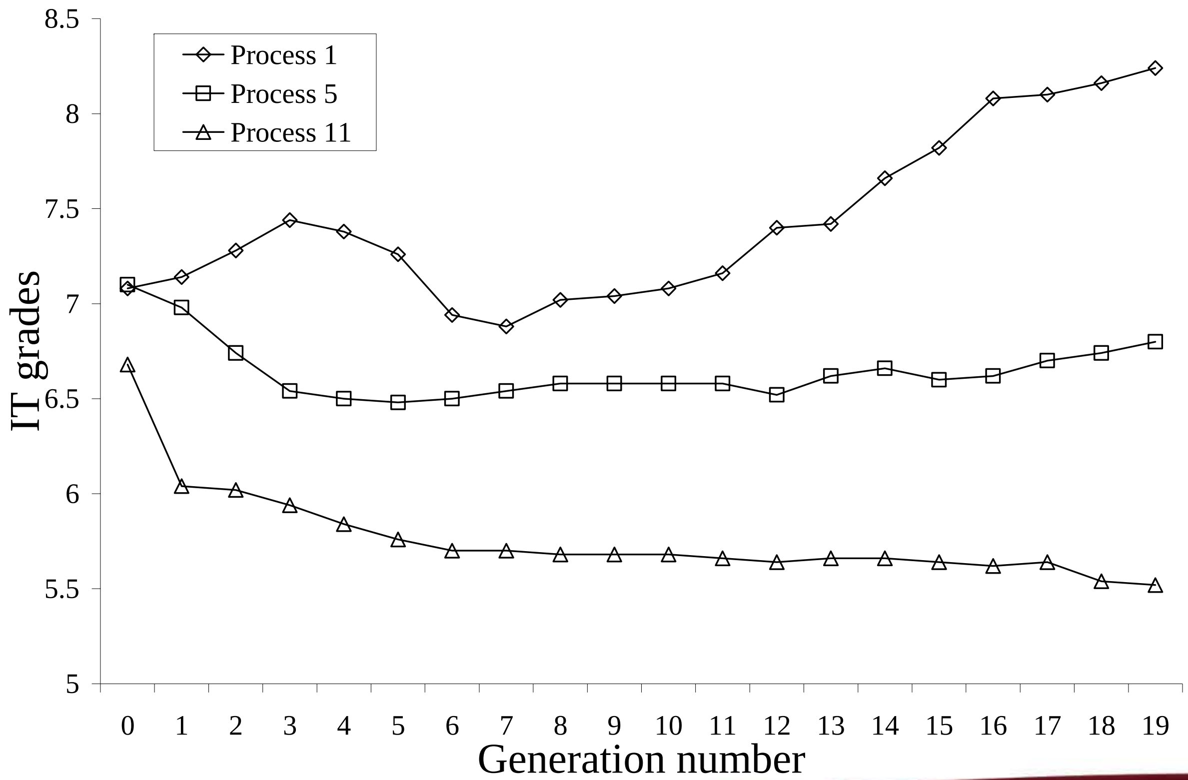
Results

- GA
  - Optimal IT Grades=(9, 6, 9, 8, 7, 6, 8, 9, 9, 8, 5, 9)
  - Cost=43.78

- Sensitivity Analysis
  - Optimal IT Grades=(9, 5, 9, 9, 6, 6, 6, 9, 9, 7, 6, 7)
  - Cost=45.29

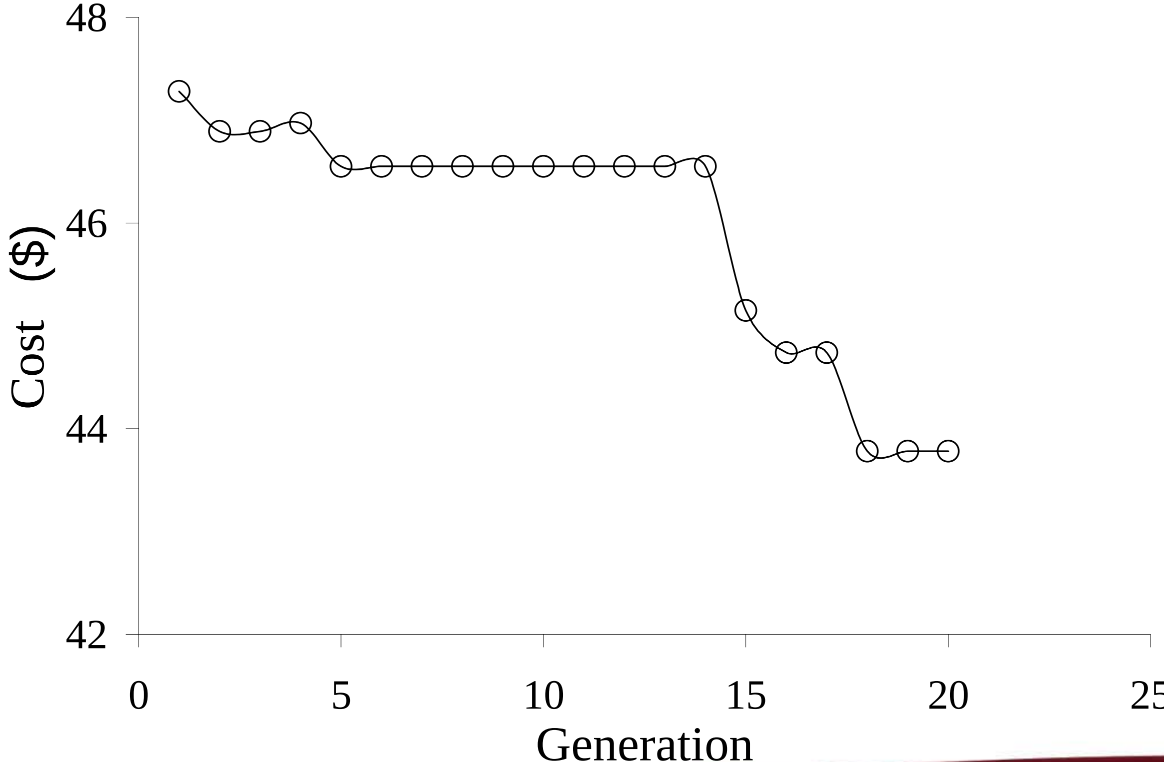
Results

- IT grades evolution over generations



Results

- Cost reduction procedure
  - Lowest cost in each generation



GA Application Summary

- GA is more effective than other optimization methods in tolerance assignment problem
- Using GA, we can identify the processes that allow loose tolerances and those require tight tolerances

Conclusion

- Improve GA’s computation efficiency
  - EDG-GA was developed to improve GA efficiency by *predicting potentially good designs based on non-problem-specific knowledge*
  - A method is proposed to *characterize the interaction between specific genes*, and *guide GA operators design*

- Improve manufacturing processes design using GA
  - GA was used to optimize tolerance assignment plan to reduce manufacturing cost

Future Work

- Proposed improving methods needs further justification through case studies
- Besides efficiency, reliability of GA also needs to be studied

Acknowledgements

- Professor Rong
- Professors Apelian and Sisson
- CHTE
- All CAM-Lab members
- Parents

Thanks

Questions?