A Multiobjective Genetic Algorithm for the Design of Asymmetric FIR Filters

Sabbir U. Ahmad and Andreas Antoniou

DSP Group

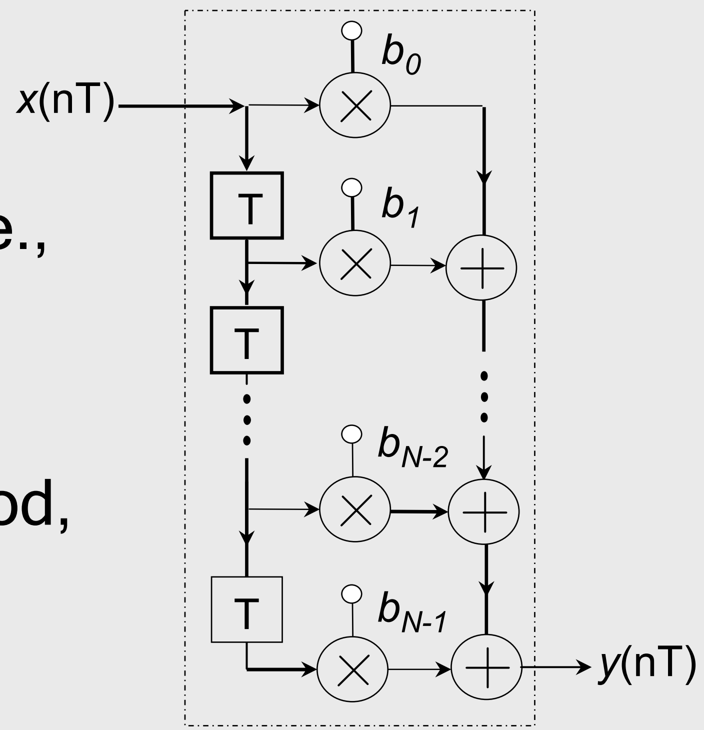
Dept. of Electrical and Computer Engineering
University of Victoria, Canada

OVERVIEW

- Introduction
- Classical vs. genetic optimization
- Multiobjective optimization
- The elitist nondominated sorting genetic algorithm approach
- Design example
- Conclusions

INTRODUCTION

- FIR filters are usually designed with symmetric coefficients to achieve linear phase response with respect to the baseband, i.e., $b_0 = b_{N-1}$, $b_1 = b_{N-2}$, etc.
  - Efficient design methods are available, e.g., window method, Remez algorithm
  - Large group delay



FIR Filter

INTRODUCTION (Cont'd)

- Filters with Asymmetric Coefficients
  - *Approximately* linear phase response in passband
  - Relatively small group delay
  - Can be designed by using classical optimization methods with a multiobjective formulation.
  - Can also be designed by using a multiobjective genetic algorithm (GA) known as the *elitist nondominated sorted GA* (ENSGA).

CLASSICAL OPTIMIZATION ALGORITHMS

- Fast and efficient
- Very good in obtaining local solutions
- Unbeatable for the solution of convex (concave) problems
- In multimodal problems, they tend to zoom to a solution in the locale of the initialization point.
- Not equipped to discard inferior local solutions in favour of better solutions.

GENETIC ALGORITHMS

- Are very flexible, nonproblem specific, and robust.
- Can explore multiple regions of the parameter space for solutions simultaneously.
- Can discard poor local solutions in favour of more promising subsequent local solutions.
- They are more likely to obtain better solutions for multimodal problems than classified methods.

GENETIC ALGORITHMS (Cont'd)

- Owing to the heuristic nature of GAs, arbitrary constraints can be imposed on the objective function without increasing the mathematical complexity of the problem.
- Multiple objective functions can be optimized simultaneously.
- They require a very large amount of computation.

MULTIOBJECTIVE OPTIMIZATION

- In many applications, several objective functions need to be optimized simultaneously.
- In classical optimization, multiple objective functions are used to construct a more complex unified objective function with or without constraints.
- On the other hand, with GAs multiple objective functions can be optimized directly to obtain a set of compromise solutions of the problem at hand.

MULTIOBJECTIVE OPTIMIZATION (Cont'd)

- A multiobjective optimization problem with *k* objective functions can be represented as:

> Minimize
>
> $$\mathbf{f}(\mathbf{x}) = [f_1(\mathbf{x}),\ f_2(\mathbf{x}),\ \ldots,\ f_k(\mathbf{x})]$$
>
> subject to $\mathbf{x} \in \mathbf{X}$
>
> where $\mathbf{X}$ is the solution space

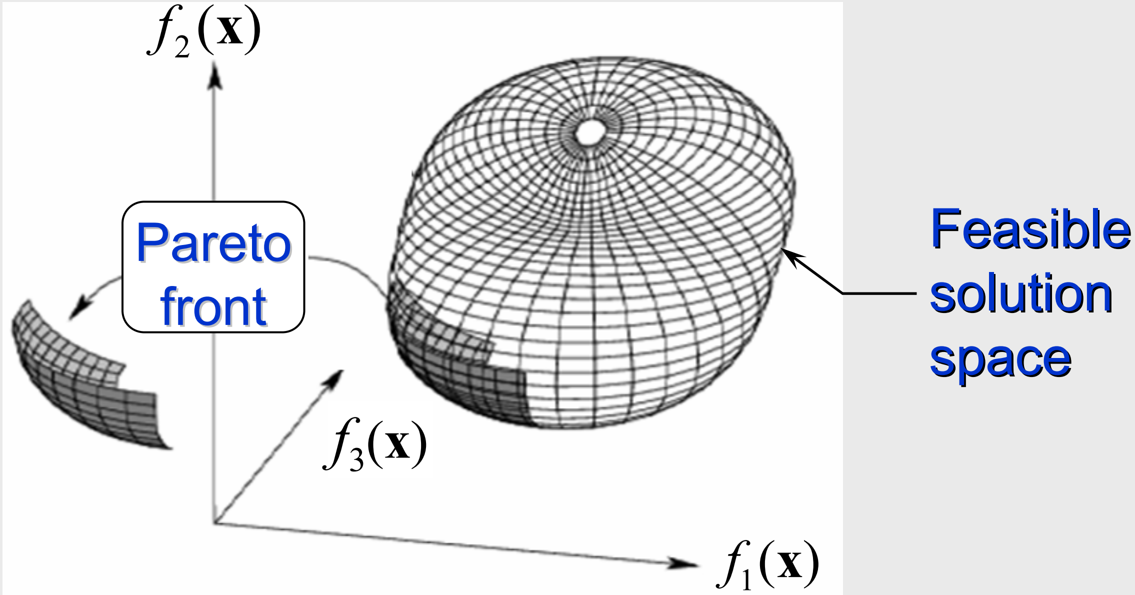
- A set of compromise solutions obtained by multiobjective approaches is known as a *Pareto optimal* solution set.
- The user can select the best compromise solution from a Pareto-optimal set.

PARETO OPTIMALITY

- A solution x_1 *is said to dominate a solution* x_2 if the following conditions hold:
  - Solution x_1 is no worse than x_2 for all objectives
  - Solution x_1 is strictly better than x_2 in at least one objective
- A solution that is not dominated by any other is said to be a *nondominated solution*.
- A set of nondominated solutions in the solution space is a *Pareto-optimal* solution set.

PARETO OPTIMALITY (Cont'd)

- The nondominated set of the entire feasible solution space is the *globally Pareto-optimal* set.
- The solution space corresponding to the Pareto optimal solution set is called the *Pareto front*.



ELITIST NONDOMINATED SORTING GA

- ENSGA introduces diversity in solutions by sorting the population according to the nondomination principle.

- classifies the population into a number of mutually exclusive classes

- assigns highest fitness to the members of the class that are closest to the Pareto-optimal front

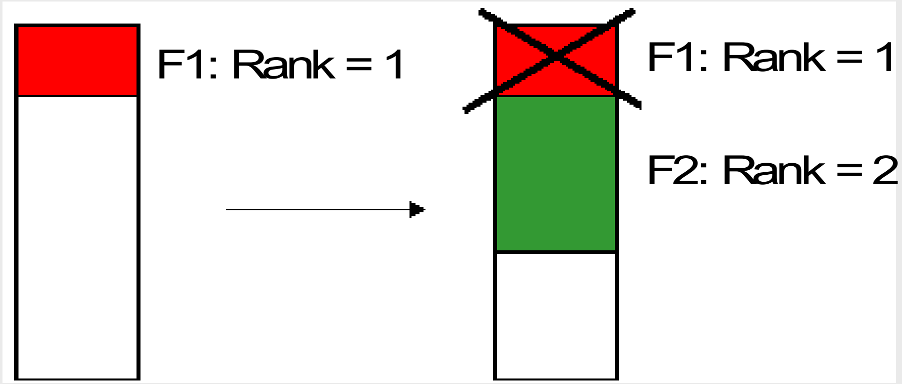
- uses the elitism principle to increase the number of Pareto solutions.

ENSGA STEPS

- Initialize population
- Evaluate fitness of individuals
- Select nondominated solutions
  - by using *nondominated sorting*
  - by assigning fitness values according to *crowding distance*
- Perform crossover to obtain new offsprings
- Mutate individuals in the offspring population
- Evaluate offspring solutions
- Perform *elitist replacement*

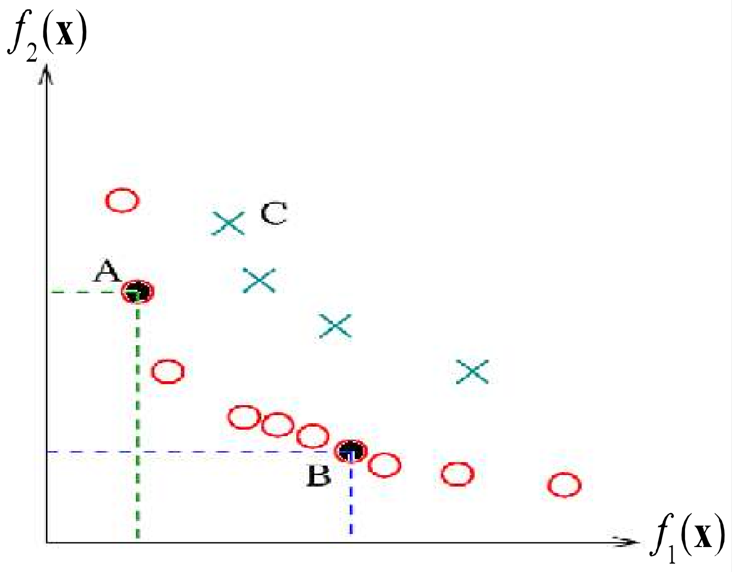
NONDOMINATED SORTING

- Identify the best nondominated set.
- Discard the nondominated solutions from the population temporarily.
- Identify the next best nondominated set.
- Continue till all solutions are classified.



CROWDING DISTANCE

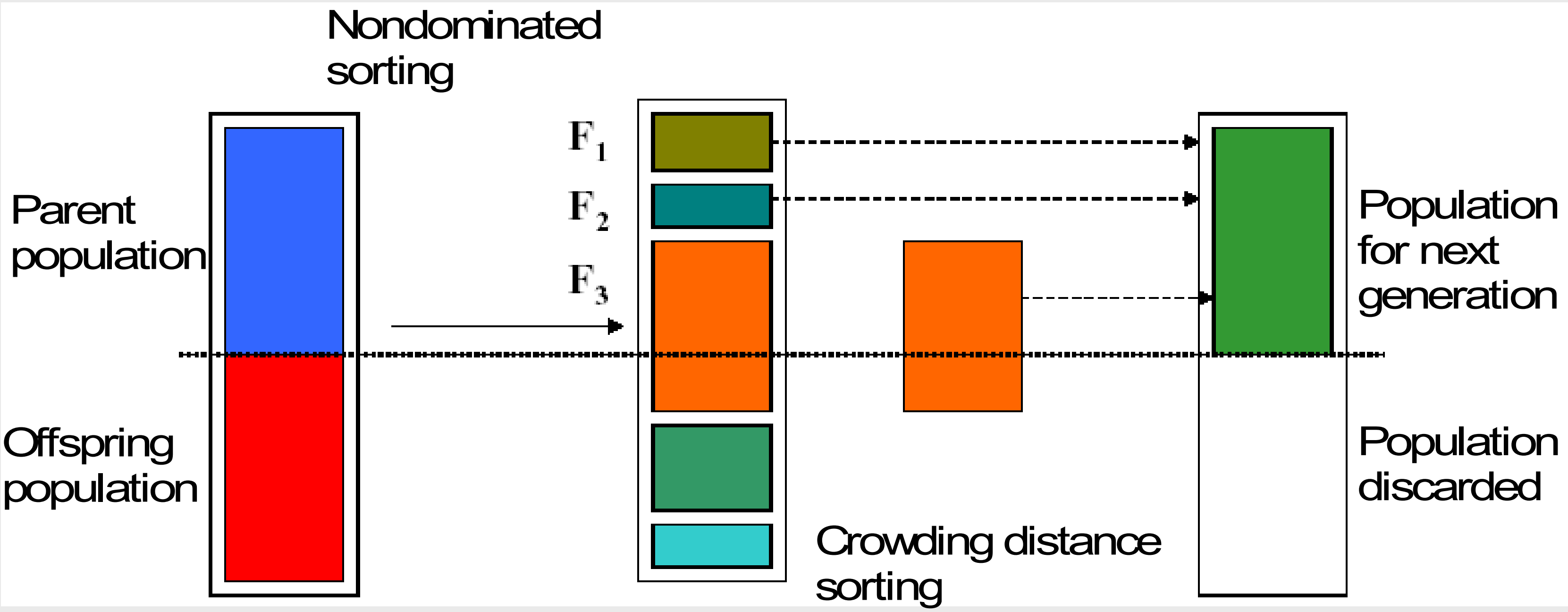
- Crowding distance is a diversity metric.
- Crowding distance is defined as the front density in a specific locale.
- Each solution is assigned a crowding distance.
- Solutions located in a less crowded space are preferred.



Solution **A** is located in a less crowded locale than **B**

ELITIST REPLACEMENT

- Combine parent and offspring populations.
- Select better ranking individuals and use crowding distance to break any ties.



DESIGN OF ASYMMETRIC FIR FILTERS

- A population of potential solutions is created from an initial least-squares solution.
- *Simulated binary crossovers* and *polynomial mutations* are applied according to predefined probabilities of occurrence, P_x and P_m, respectively.
- The objective functions used to evaluate the fitness of the individual solutions are based on the amplitude response and group-delay errors.
- The ENSGA is terminated after a prespecified number of generations.

CHROMOSOME STRUCTURE

- Chromosome (*candidate solution*) :
  - The coefficient vector of the FIR filter, $\mathbf{b}$, is used as the candidate solution.
  - To avoid very long binary strings, a floating-point representation is used in encoding the chromosomes.

OBJECTIVE FUNCTION

- Three objective functions have been used:
  - L_∞ norm of the passband amplitude-response error

$$F_p = \max \left|1 - H(e^{j\omega})\right| \qquad \text{for} \quad \omega \in \text{Passband}$$

  - L_2 norm of the stopband amplitude-response error with a constrained imposed on the peak error

$$F_a = \sum_{k=1}^{K_a} \left|H(e^{j\omega_k})\right|^2 \qquad \text{for} \quad \omega_k \in \text{Stopband}$$

$$\text{subject to } \min\left\{-20 \log_{10} |H(e^{j\omega_k})|\right\} \geq \delta_a \text{ dB}$$

  - A parameter Q which measures the degree of flatness of the passband group-delay characteristic

$$Q = \frac{100\ (\overset{\frown}{\tau} - \overset{\smile}{\tau})}{(\overset{\frown}{\tau} + \overset{\smile}{\tau})}$$

DESIGN EXAMPLE

- **Specifications:** Highpass FIR filter,

  $\omega_a = 0.4$, $\omega_p = 0.55$, $\omega_s = 1$ rad/s,

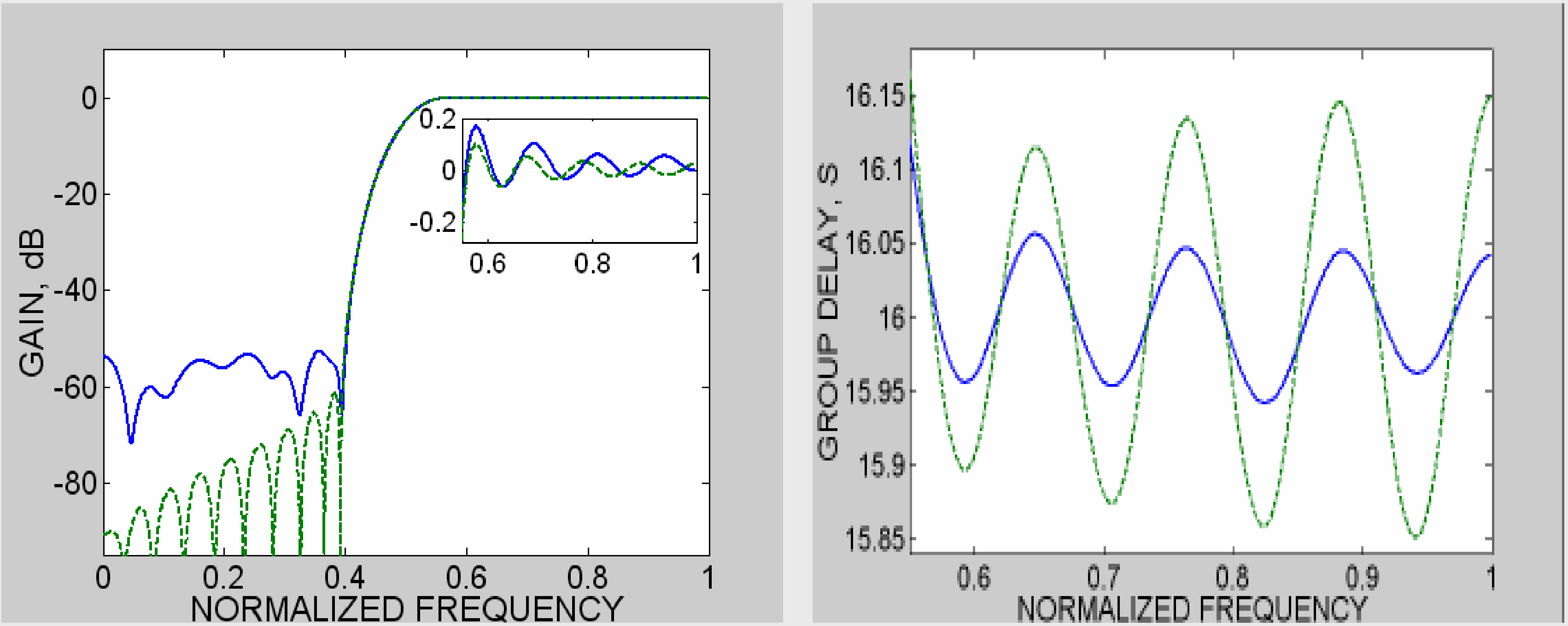
  $\delta_a = 52\ dB$, $\bar{\tau} = \dfrac{\overset{\frown}{\tau} + \overset{\smile}{\tau}}{2} \le 16.5$, $N = 35$

- **Initialization:** An FIR filter with nonsymmetric coefficients designed using a weighted least-squares method was used as the initial design.

- **Solutions:** The next two slides give the results for two solutions from the Pareto-optimal solution set.

DESIGN EXAMPLE (Cont'd)

Solution 1:



Amplitude response

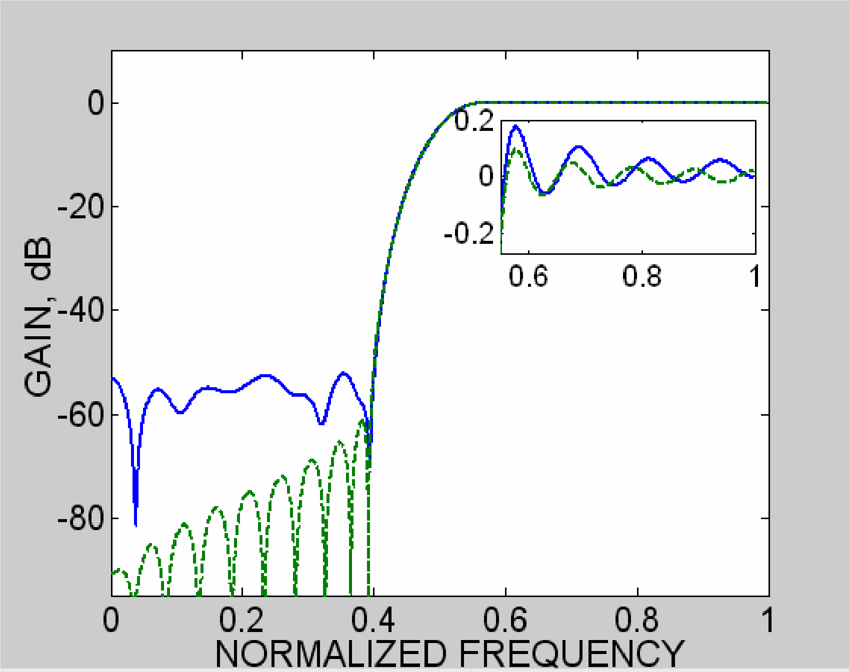
Delay characteristic

——— Design obtained with ENSGA

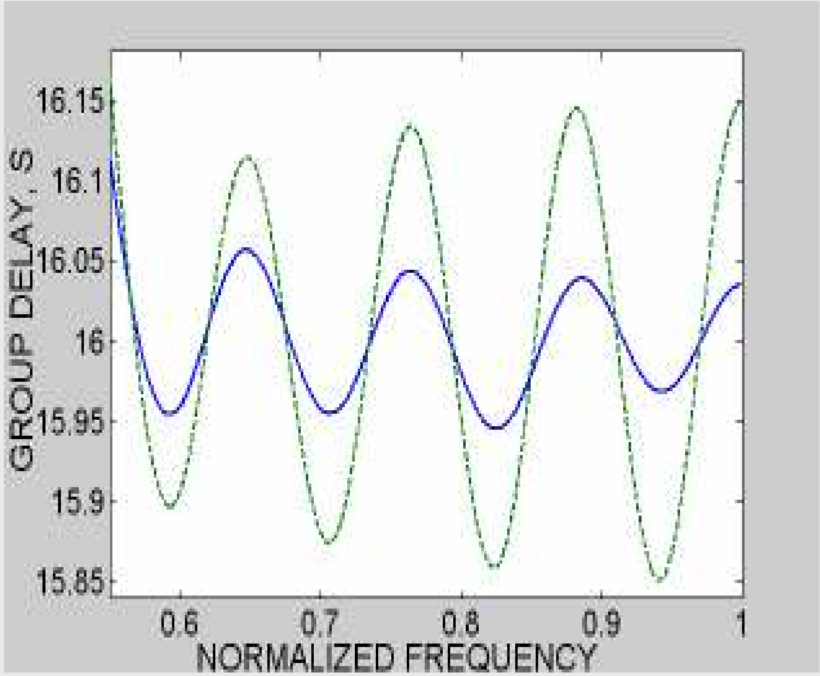
- - - - Initial design

University of Victoria

Solution 2:



Amplitude response

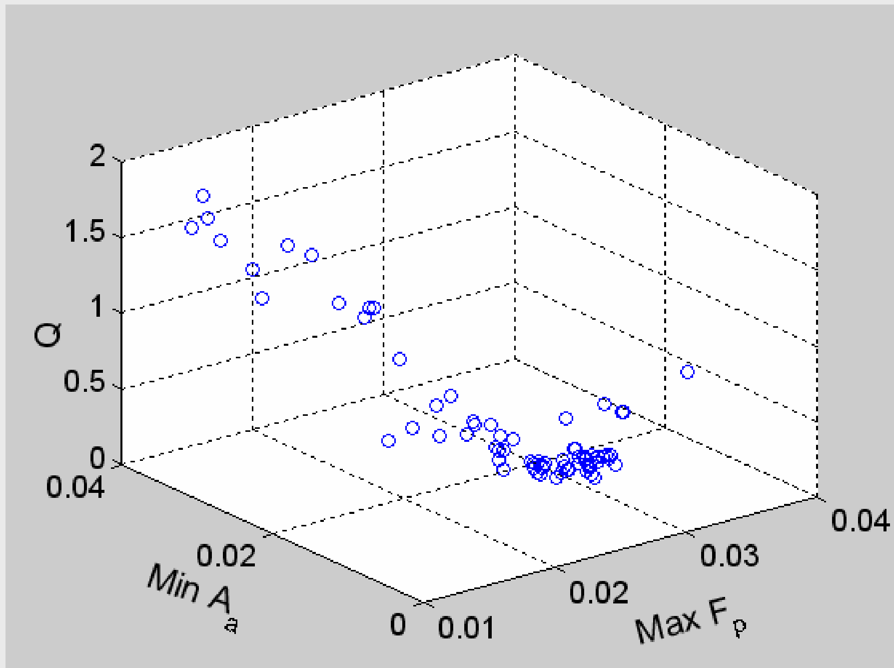


Delay characteristic

—— Design obtained with ENSGA - - - - Initial design

DESIGN EXAMPLE (Cont'd)

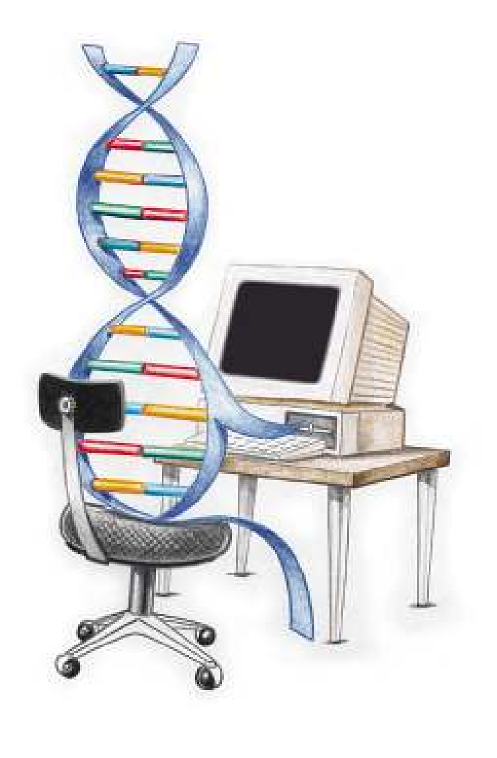
3-D scatter plot of the Pareto-optimal solutions obtained by using the ENSGA



$$A_a = -20\log_{10}|H(e^{j\omega})| \qquad \text{for} \quad \omega \in \text{Stopband}$$

CONCLUSIONS

- The ENSGA can be used to design nonsymmetric FIR filters that would satisfy multiple requirements imposed on the amplitude response and the delay characteristic.
- The approach yields an improved design with respect to the initial weighted least-squares design.
- The design that is best-suited to a specific application can be chosen from the set of Pareto-optimal solutions obtained.
- In common with other GAs, the ENSGA requires a large amount of computation. However, this is not a serious problem unless the filter design has to be carried out in real or quasi-real time.



Thank you