# Extending the bounds of the search space: A Multi-Population approach

**M. A. Beck**
Plymouth Engineering Design Centre
University of Plymouth
Drake Circus
Plymouth PL4 8AA UK
Phone: +44 (0)1752 233508,
Fax: +44 (0)1752 233529
e-mail: mbeck@soc.plym.ac.uk

**I.C. Parmee**
Plymouth Engineering Design Centre
University of Plymouth
Drake Circus
Plymouth PL4 8AA UK
Phone: +44 (0)1752 233508,
Fax: +44 (0)1752 233529
e-mail: iparmee@plymouth.ac.uk

## Abstract

This paper reports an investigation of GA based techniques that allow the search to move outside of its pre-defined limits. Using a hybrid coarse-grained / fine-grained GA, dynamic fitness sharing techniques are presented that encourage sub-populations to explore differing high performance areas of the search space. Such techniques when coupled with a migration typology provide a realistic mechanism for moving a GA based search to unspecified areas of the search space.

## 1 INTRODUCTION

The aim of the research underlying this paper is the development of adaptive search techniques to assist the designer at the conceptual stage of the design process. It is at this stage that neither the problem space and hence the solution space are fixed:

- Design objectives may change either through their weightings / importance or through the removal or addition of objectives.
- The initial limits assigned to the design variables are subject to change with the relaxation of variable limits and thus the necessity to search outside the initial pre-defined bounds.

Many researchers (i.e. Goldberg, 1989, Maher et al 1995) have recognised the power of GA based techniques for the exploratory nature of conceptual design, and more recent research (i.e. Gero and Kazakov, 1998, Parmee and Bonham, 1998) have employed GA based technique where the search progresses *outside* of the initial design variable limits. This paper builds on this research by investigating the use of a real-valued representation and an independent geographically structured multi-population technique.

Given that design spaces are high-dimensional and multi-modal, it is argued that a multi-search technique needs to be adopted whereby a collection of relatively independent sub-populations progressively evolve across differing areas of the search space, locating regions of potential high performance marking as found and then moving on to areas outside their current search limits. Thus it is envisaged that this is a dynamic situation and suitable adaptive search techniques need to be developed to meet this.

Initial research within a fixed design variable environment (Beck 1996, Beck and Parmee 1997) suggested that the natural emergent characteristics of an EcoGA (i.e. Davidor et al, 1993) could suit the problem at hand.

Essentially an EcoGA is a technique whereby a large population can be can be structured into a series of smaller sub-populations by placing one individual at each location on a toroidal 2-dimensional grid. With each individual assigned this way the grid locations are not necessarily related to the individuals' solutions rather they are arbitrary designations used to perform selection.

For the purposes of evolution an individual is randomly selected from the grid and a sub-population is defined as that individual plus its eight immediate neighbours. This sub-population is then evolved in similar fashion to that of a steady state GA (Syswerda 1989): Roulette wheel selection is used to select two individuals for mating and mutation, each resulting offspring then competes against an individual randomly selected from the parent sub-population for a place in that sub-population. The outcome of this ‘tournament’ is decided deterministically with the fittest individual winning.

Using an EcoGA many authors (i.e. Davidor et al 1993, McIlhagga et al, 1996, Parmee and Beck 1997) have reported the establishment of relatively stable genetic diversity through the emergence of *demes*: Clusters of genetically similar individuals. Moreover a local search occurs within a deme and new regions are explored through the recombination of individuals at the edges. These demes are not stable, and while uniform crossover with a high mutation rate may slow convergence (Spiessens and Manderick 1991), the tendency is towards single deme dominance.

To reduce the onset of single deme dominance, previous work (Beck 1996) maintained search diversity by detecting converging sub-populations and re-initialising when they reached a convergence criterion. The bounds of the converged region was then marked as tabu, and the fitness of individuals entering a tabu area was reduced.

The convergence criterion employed in both the previous work and in this paper is defined as some ratio $\lambda$, $\lambda \in (0, 1)$, of the mean distance between the individuals and their sub-population centroid at initialisation to their current distance. Thus if $dist_{t+1} <= \lambda * dist_0$ then the sub-population is said to have converged.

For the present problem it was intended to re-assign the converged sub-population an area outside the current search bounds. A difficulty with this however is that the sub-population consisted of only nine individuals and further individuals are needed to make a GA based search meaningful. Identifying candidate individuals that are close to the converged sub-population in terms of both geographical distance and phenotypic / genotypic distance was problematic.

To overcome this, a more structured EcoGA model is adopted in the form of a hybrid coarse / fine grained GA. As per the EcoGA a toroidal 2D grid is divided into a series of cells each with one individual occupying one cell. Cells are then designated as belonging to one sub-population such that the number of individuals / cells within each sub-population is equal. Thus figure 1, below, shows a part of a 16 x 16 grid of 4 x 4 sub-populations each with 16 individuals each.

This hybrid approach, termed here a DemeEcoGA, allows for two types of evolution: Firstly, sub-population evolution, in this instance only those individuals within the sub-population evolve. Secondly, boundary evolution, as individuals are geographically located only those individuals at the 'edges' of the sub-population are allowed to undergo recombination. This being illustrated by the greyed area of figure 1.

By utilising such a model it is reasoned that: Firstly, implementation of a steady-state GA within the sub-populations will lead to rapid exploitation of optima, and secondly, the extent to which sub-populations move to differing areas of the search space then an exploratory search is developed by means of the boundary evolution. Moreover the exploitation-exploration balance can, to some extent, be determined by the ratio of sub-population to boundary evolutions.

Whilst the sub-population evolution is fairly straightforward; it is a small population GA. The nature of the boundary evolution is more complex and proceeds as follows:

- Randomly select a sub-population (sub-pop1) and randomly select a parent individual (P1) from the bounds of sub-pop1. Now randomly select parent P2 from those individuals that are directly adjacent to P1. (see Figure 2)
- Recombine P1 & P2 to give C1 & C2.
- Although a fitness based tournament based replacement strategy is employed, whereby C1 & C2 compete for occupancy of the cells currently occupied by P1 & P2, a further criteria is added such that the child chromosomes are placed within the closest sub-population. Closeness being defined as the Euclidean distance from $C_i$ to the centroid of $Sub\text{-}pop_i$.

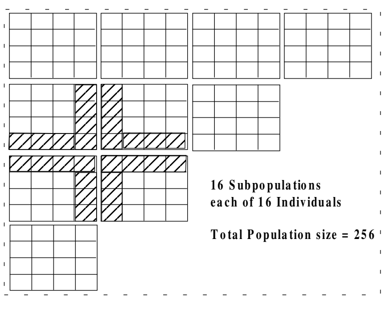


Figure 1: Structure of DemeEcoGA

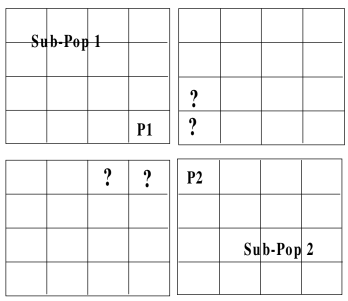


Figure 2: Selecting Parents for mating

## 2 EXPANDING THE SEARCH SPACE

Expanding the bounds of the search space raises two fundamental questions: Firstly, what should be the nature of the representation? and secondly, in which 'direction' should the search progress?

With regard to the first question, if this was a single population based search then the usual binary representation would present few problems: As the population reaches a convergence criteria the whole population can be reinitialised within a newly defined range. Unfortunately, however the nature of multi (sub) population based search *may* preclude such a binary representation. Thus if the representation of two individuals undergoing recombination is defined over differing but overlapping ranges then the number of bits they have in common may be limited and effectiveness of exchanging bit alleles may be of questionable utility.

A representation that does lend itself more naturally to the problem at hand is that of real valued chromosomes. While the initial population may be initialised within some pre-defined bounds, this representation does allow recombination of individuals of any value.

Using intermediate recombination (Muhlenbein. and Schlierkamp-Voosen, 1993) offspring are reproduced according to the rule:

$$C1 = P1 \text{ x } \alpha(P2 - P1)$$

Where: $\alpha$ is a scaling factor over some interval usually [-0.25, 1.25]

P1, P2 are the parent chromosomes

Moreover intermediate recombination is capable of producing new offspring within a slightly larger hyper-cube than that defined by the parents.

With regard to the second question, the work of Hajela and Lee (1996) may be a useful approach. In Hajela's work, a clustering algorithm is employed to group design regions with similar characteristics, a neural network is then trained using this region as training data. Finally by examining the weight matrices of such networks the effect of the input variables over the output objectives can be quantified. Thus the search should be expanded along those variables that are more likely to effect the design objectives.

While the above is both fascinating and relevant work, a more simpler approach is adopted here. As a sub-population reaches convergence then new search bounds are defined plus or minus some value of the current sub-population centre. Thus, to the extent that this convergence has occurred close to the current limits of the search space then a new search will be initialised outside the current search area *and* within a region of relative high performance. The technique adopted here for setting this new search range is to use the upper and lower limits of the initial search to set the bounds, and for the sub-population centre the centroid is used. Table 1 presents an example of this technique, where a converged sub-population is centred on (2.0, 2.6).

Table 1: Definition of new search bounds

| | **Initial Search** Centre : (0.0, 0.0) | | **New Search** Centre : (2.0, 2.6) | |
|---|---|---|---|---|
| | lower bounds | upper bounds | lower bounds | upper bounds |
| x1 | -3.0 | 3.0 | -3.0 + 2.0 | 3.0 + 2.0 |
| x2 | -3.0 | 3.0 | -3.0 + 2.6 | 3.0 + 2.6 |

### 2.1 FITNESS SHARING

While it is advanced that multi-population based searches will explore differing areas of the search space. In reality there is no guarantee that this will actually occur, and moreover as the search progresses genetic drift will pull the sub-populations to one optima. Hence to encourage an exploitatory based search of multiple-optima, two fitness sharing techniques are examined. For brevity these are termed $sharing_A$ and $sharing_B$ respectively.

The underlying rationale for both sharing techniques differs from the more usual approach of fitness sharing between *individuals*: In this study an individuals fitness is decreased by a function of its encroachment into each sub-populations current 'search area'. Thus the focus is upon fitness sharing between sub-populations, and it is the definition of a sub-populations 'search area' that distinguishes the two techniques under investigation. While the first, $sharing_A$, is similar to that of the classic work of Deb and Goldberg (1989), a major problem with sharing functions is the setting of an appropriate 'niche radius'. This being further complicated in this instance as sub-populations are reinitialised they are at differing stages of evolutionary search, and the 'niche radius' metric needs to take account of this.

Thus a *dynamic* niching technique is adopted: Each sub-populations 'search area' is defined as the mean Euclidean distance of all individuals within that sub-population to the sub-populations centroid. This mean distance acts as the niche radius, and an individuals fitness is reduced by the extent that it has entered the 'search area' of *any* sub-population. The consequence being that as a sub-population converges then the niche radius will also decrease.

More formally, the sharing function is now defined as:

$$f'(i) = \frac{f(i)}{\sum_{j=1}^{n} sh(d(i, Centroid_j))}$$

$$\mathrm{sh}(d(i,Centroid_j)) = \begin{cases} 1 - \left( \dfrac{d(i,Centroid_j)}{\sigma_{\mathrm{share}}} \right)^{\alpha} & \text{if } d(i,Centroid_j) < \sigma_{\mathrm{share}} \\ 0 & \text{otherwise} \end{cases}$$

Where:

d*(i, Centroid* $_j$*)* is the Euclidean distance between individual i and the centroid of sub-population j

$\sigma_{share}$ is the mean distance of all individuals of sub-population *j* to the centroid of sub-population *j*

While this mean distance decreases as a sub-population converges, upon re-initialisation the mean distance will be high and it is not appropriate to penalise possibly fit individuals simply because they occupy the niche of a randomly generated sub-population. Hence a linear scaling factor is introduced that reduces the niche radius for newly re-initialised sub-populations until they reach some stage in the evolutionary cycle. Thus $\sigma_{share}$ now becomes:

$$\sigma_{share} = \sigma_{share} * \text{Scaling}$$

$$Scaling = \begin{cases} (m * NEvol/MaxEvol) + c & \text{if } NEvol < MaxEvol \\ 1 & \text{otherwise} \end{cases}$$

Where:

NEvol = number of evolutions the sub-population has undergone since initialisation

MaxEvol = Pre-defined constant, currently set at 50

For the second fitness sharing technique, *sharing* $_B$, the 'search area' is defined as the sub-population centroid plus or minus one standard deviation of the position of the individuals in the search space. Again as in the case of the first approach the 'search area' is dynamic with the standard deviation of the position of the individuals decreasing as the sub-population evolves.

An individuals fitness is reduced by a linear decreasing function of its niche count, where 'niche count' is defined as the number of sub-population search areas that an individual occupies.

Fit(i)' = Fit(i) - Fit(i) * NCount/ NumSubPop

Where:

Fit(i) = Un-scaled fitness of individual i

NCount = Niche count

NumSubPop = total number of sub-populations

Thus this technique defines a hyper-cube, and is more concerned with a head count of individuals located anywhere inside the hyper-cube. Whereas the first technique will define a hyper-sphere, and once inside the hyper-sphere fitness will decrease by the extent of intrusion.

## 3 TESTING THE TECHNIQUES

To examine the effects of boundary evolution and both fitness-sharing techniques five tests are conducted across single and twin objective cases.

### 3.1 SINGLE OBJECTIVE

The first test function (F1) is a single-objective, two variable multi-modal function. A surface plot is shown in figure 3, and although the function is shown over the range [-10..10] for both x1 and x2, the initial search space is over a much small region of [-3..3], as illustrated by the dashed square area.

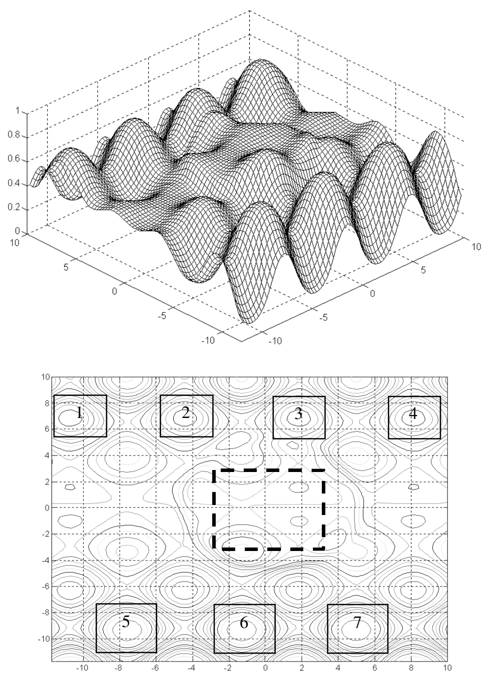


Figure 3: Surface plot of test function F1,top, and contour plot, lower, showing initial search area (dotted line) and the seven pre-defined numbered regions.

### 3.2 TWIN OBJECTIVE

The second test function (F2) is a twin objective three variable multi-modal function. this being constructed by combining two other two-variable functions, F2a and F2b. Surface plots and contour plots of these latter two functions are shown in figures 4 and 5. More specifically one variable, x1, is common to all functions, such that:

$$F2(x1, x2, x3) = F2a(x1, x2) + F2b(x1, x3)$$

As in the case of function F1 although figure 4 and 5 illustrates the functions over the range [-10..10] for x1, x2 and x3, the initial search space is over a much small region of [-3..3].

The aim of these functions is to provide not only an increase in dimensionality, but they are constructed so as to form a deceptive problem for the techniques under investigation. Thus while F2 has a global optimum ($F2_{opt}$) centred at (6.00, 6.50, 6.53), F2a and F2b have local optima running along a ridge illustrated by the dotted line in both contour plots (figures 4 & 5).

It is these local optima ridges which act as the deceptive part of the function which will tend to direct the search along the local optima and away from $F2_{opt}$ .The region of $F2_{opt}$ can be found however if the fitness sharing techniques moves the sub-populations along the ridges thus allowing cross-over between individuals of differing sub-populations to place new chromosomes with that region (see figure 6).

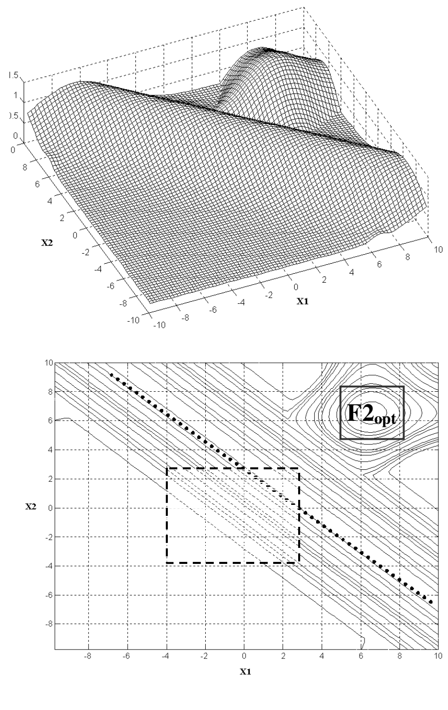


Figure 4: Surface plot of function F2a, top, and contour plot, lower, showing initial search area (dashed square) and local optima ridge (dotted line).

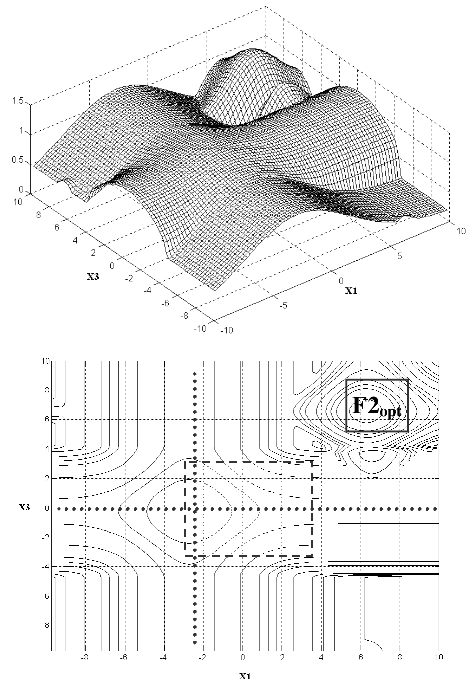


Figure 5: Surface plot of function F2b, top, and contour plot, lower, showing initial search area (dashed square) and local optima ridge (dotted line).

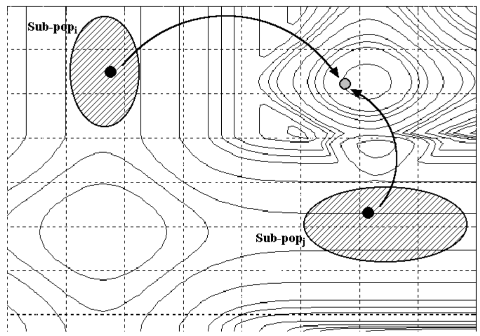


Figure 6: Recombination between individuals of dispersed sub-populations.

### 3.3 THE TESTS

Effects of both sharing techniques and boundary evolution are examined over five experimental conditions for each of the two test functions (F1 and F2). These are summarised in table 2, overleaf.

Table 2: Summary of tests used

| TEST NUMBER | | CONDITION | |
|---|---|---|---|
| | | **Sharing** | **Boundary** |
| 1.1 | 2.1 | $sharing_A$ | yes |
| 1.2 | 2.2 | $sharing_B$ | yes |
| 1.3 | 2.3 | None | yes |
| 1.4 | 2.4 | $sharing_A$ | no |
| 1.5 | 2.5 | $sharing_B$ | no |

The DemeEcoGA consists of 16 sub-populations of 16 individuals each, as illustrated in figure 1. Probability of cross-over is set at 0.7 and probability of mutation at 0.01. The convergence criterion, λ, described in section 1 is set at 0.4.

The performance metric employed is number of individuals that are located in the regions of the pre-defined high performance. Seven such regions are defined for function F1 (figure 3) and one for F2 (figures 4 & 5). Each test is run for a maximum of 15000 function evaluations (this being equivalent to 150 generations of a standard GA with population size of 100), and a region population count taken every 500 evaluations. The results are then averaged over 20 test runs.

# 4 RESULTS

## 4.1 SINGLE OBJECTIVE

Examining tests 1.1 and 1.2, the effects of fitness sharing combined with a boundary evolution, figures 7 and 8 illustrate that not only are all regions populated during the course of the run, but there is a characteristic rise and fall in regional population as regions are identified and the search progresses to other regions.

These results however cannot be attributable to either fitness sharing or boundary evolution alone. Thus in the condition of no fitness sharing, test 1.3, the results shown in figure 9 illustrate a dramatic rise in region 6 population with little or no corresponding population across the other regions. Similarly in the fitness sharing but no boundary conditions, tests 1.4 and 1.5, the results of figures 10 & 11 show that while there is a more diverse regional population, it does not significantly encompass all pre-defined regions. Moreover the shapes of the graphs suggest that the regions are beginning to be populated and thus the search has not progressed as far as that experienced in figures 7 & 8.

This latter observation is more succinctly illustrated when the mean total regional count is examined. Thus in figure 12 it can be seen that conditions 1.2 and 1.3 achieve a peak regional population within 7000 function evaluations whilst the other conditions are still climbing to a possible summit. Finally, examining results between the two sharing techniques it can be seen that the effects of $sharing_A$ is a more progressive region identification, with two regions being identified earlier in the search followed by the remaining five regions (figures 7 and 8). In contrast $sharing_B$ is characterised by a more uniform region identification although the number of individuals within those regions is reduced (figures 7, 8 and 12).

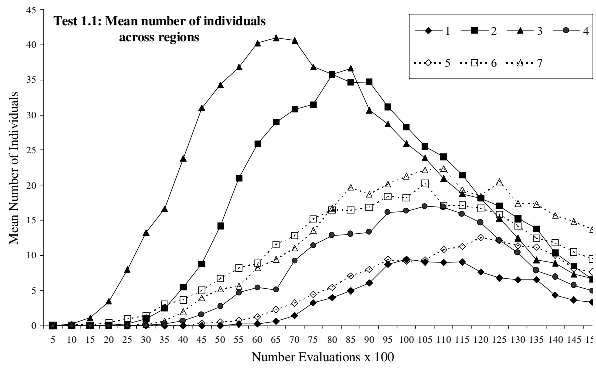


Figure 7: Test 1.1

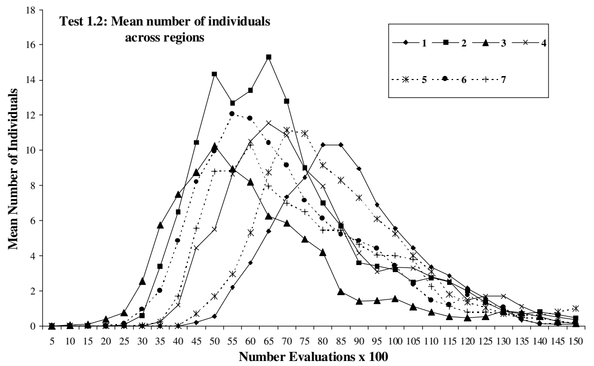


Figure 8: Test 1.2

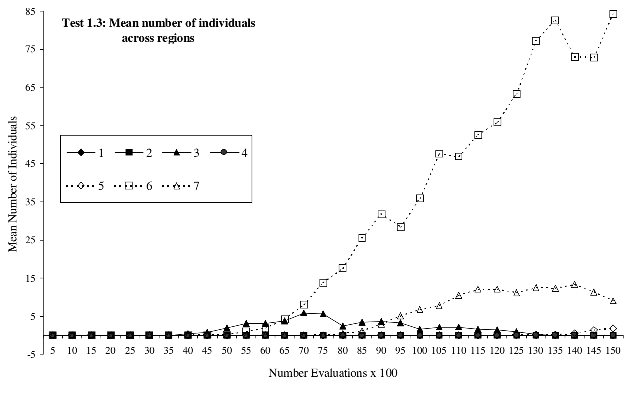


Figure 9: Test 1.3

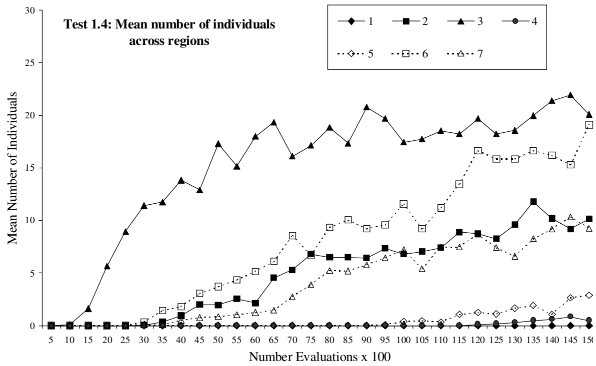


Figure 10: Test 1.4

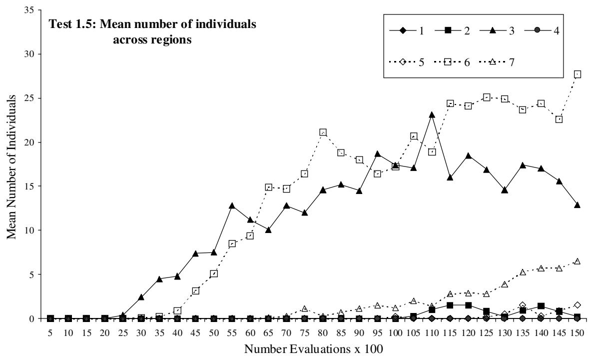


Figure 11: Test 1.5

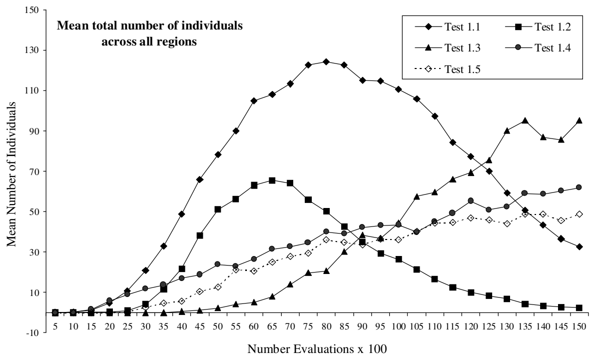


Figure 12: Mean total regional population across regions

## 4.2 TWIN OBJECTIVE

The results of the single objective case are echoed in the twin objective situation. Reference to figure 13 shows that in both conditions where a fitness sharing technique is coupled with a boundary evolution (tests 2.1 and 2.2) then the optima region of F2 is populated more densely than fitness sharing alone (tests 2.4 and 2.5).

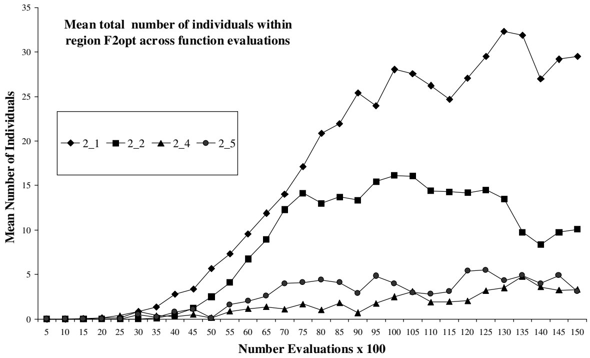


Figure 13: Tests 2.1 – 2.5

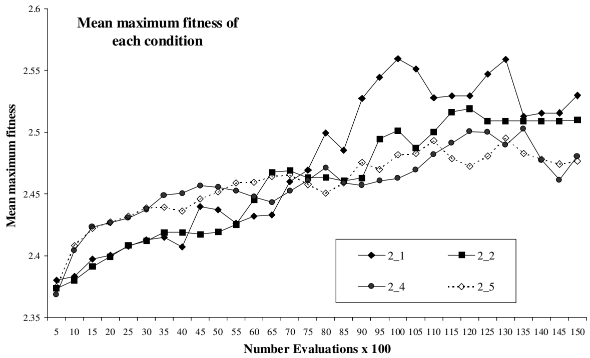


Figure 14: Mean maximium fitness across tests for function F2

The result of evolution without a sharing technique (test 2.3) is far more dramatic with $F2_{opt}$ not being located at all in *any* of the 20 trials, (consequently this has not been presented graphically).

Examining the effects between sharing techniques, similar results are found as those of function F1. Thus generally it can be seen that $sharing_A$, test 2.1, both identifies the region sooner and results in a higher mean regional population density than that of $sharing_B$ (figure 13).

The superiority of fitness sharing+boundary evolution is also reflected in the results of the mean maximum fitness obtained at each of the 500 function evaluation snapshots. Results of figure 14 show that although conditions without a boundary evolution achieve a higher mean maximum fitness at the start of the search, the sharing+boundary conditions results in a final superior fitness score through the location and population of region $F2_{opt}$.

This cross-over in fitness performance is directly attributable to the steady-state nature of sub-population evolution finding the local optima faster without being slowed by the exploratory pressure of the boundary evolutions.

## 5 CONCLUSION

Overall the results of the techniques are encouraging: Given that a search is initially defined within some pre-defined limits, then it is possible to move the bounds of the search to find possibly better regions outside of such limits.

In doing so however it has been shown necessary to incorporate techniques that facilitate this process. Thus while the exploratory nature of either a boundary evolution or a fitness sharing technique per se may to some extent promote the identification of high performance regions away from the initial search area, it is only through the utilisation of both techniques in tandem that acceptable results can be claimed. No technique, however, is implemented without some computational expense and in this light it is reassuring that the simpler niche technique achieved marginally better results in terms of a uniform regional identification.

There are however improvements that need to be made to the approach Thus for example re-initialising a converged sub-population around its current centroid is rather simplistic and makes no account of either its location in the search space nor the location and search areas of the other sub-populations. It is believed that by incorporating a more sophisticated information exchange between sub-populations an important reduction in the number of function evaluations will be achieved.


### Acknowledgements

The research has been carried out at The Plymouth Engineering Design Centre, at the University of Plymouth and has been supported by an EPSRC studentship award. We wish to thank these organisations for their continuing support.


### References


Beck, M. A. (1996). Parallel Genetic Algorithms: An Investigation of Coarse and Fine Grained approaches to Design Space Decomposition. MSc Thesis. University of Plymouth.

Davidor, Y., Yamada, T., Nakano, R. (1993). The ECOlogical framework II: Improving GA Performance at Virtually Zero Cost *In Forrest, S. (Ed.) Proceedings of the Fifth International Conference on Genetic Algorithms.* Morgan Kaufman, 171-175.

Deb, K., & Goldberg, D. E. (1989). An Investigation of Niche and Species Formation in Genetic Function Optimisation. *In Shaffer, J. D. (Ed.). Proceedings of the Third International Conference on Genetic Algorithms.* San Mateo, CA.

Gero, J and Kazakov, V (1998). Adapting Evolutionary computing in Creative Designing. *4th International Conference on Computational Models of Creative Design* Heron Island, Australia, 6-10 December.

Goldberg, D. E. (1989). Genetic Algorithms as a Computational Theory of Conceptual Design. In Rzevski, G. & Adeyi, R. A. (Eds) *Applications of Artificial Intelligence in Engineering*. Elsevier Science Publishing.

Hajela, P. and Lee, J (1996). Role of Emergent Computing Techniques in Multidisciplinary Rotor Blade Design. In Grierson, E. D. and Hajela, P. (eds.), *Emergent Computing Methods in Engineering Design: Applications of Genetic Algorithms and Neural Networks*: 162-187. Springer.

Maher. L. M., Poon, J. & Boulanger, S. (1995). Formalising Design Exploration as Co-Evolution: A combined gene approach. *Pre-prints of the Second W\IPIF WG5.2 Workshop on Formal Design Methods for CAD.*

McIlhagga, M., Husbands, P., Ives, R. (1996). A Comparison of Optimisation Techniques for Integrated Manufacturing Planning and Scheduling. *Parallel Problem Solving from Nature* vol. IV.

Muhlenbein, H. and Schlierkamp-Voosen (1993). Predictive models for the Breeder Genetic Algorithm: Continuous Parameter Optimisation. *Evolutionary Computation.* (1 )1: 25-49.

Parmee, I. C. and Beck, M. A. (1997). An Evolutionary, Agent-Assisted Strategy For Conceptual Design Space Decomposition. *In D. Corne and J. L. Shapiro (eds.), Evolutionary Computing: Selected Papers from the 1997 AISB International Workshop, Springer Lecture Notes in Computer Science*: 275-286.

Parmee, I. C. and Bonham, C. (1998). Supporting Innovative and Creative Design using Interactive Designer/Evolutionary Strategies. *4th International Conference on Computational Models of Creative Design* Heron Island, Australia, 6-10 December

Spiessens, P & Manderick, B. (1991). A Massively Parallel Genetic Algorithm: Implementation and First Analysis. *In Belew, R. and Booker, L. (eds.), Proceedings of the Fourth International Conference on Genetic Algorithms:* 279-286. Morgan Kaufman.

Syswerda, G. (1989). A study of reproduction in generational and steady state genetic algorithms. In G. Rawlings (ed.), *Foundation of Genetic Algorithms*: 94-101. Morgan Kaufman.