Chapter 5

Experimental Results

There are three kinds of lies:
lies, damn lies, and statistics.
Mark Twain, Autobiography

This chapter reports experimental results exploring the suitability of the author's genetic-algorithm (GA)/heuristic hybrid approach for multi-wavelength all-optical transport network design.

As a basis for the experimental work described here, five network design problems were generated using the approach described in §F.1. Each problem has fifteen nodes, placed within a 1,000 km $\times$ 1,000 km area using a minimum clear radius constant of 0.5, and is intended to carry an overall static traffic requirement of 1,500 Gbit/s. The node distributions of the five problems are illustrated in Fig. F.3 and the node coordinates and traffic matrices have been recorded in Tables F.1–F.10 in Appendix F.

The chapter is divided into four sections. Results for mesh network topology design are presented in §5.1, and for wavelength-path (WP) routing, fibre choice and wavelength allocation in §5.2. An investigation into NOMaD's operator-probability adaptation mechanism forms §5.3. Finally, the chapter is summarised in §5.4.

5.1 Mesh Network Topology Design

Network topology design, briefly introduced in §3.1, determines which node pairs in the network are to be directly connected given the node locations and static traffic requirements. For mesh network topology design, the choice of node pairs to interconnect is unconstrained, resulting in an arbitrary topology. In this thesis, the metric used for this is Network Cost Model A, outlined in §4.1.1 on p. 68. To summarise, this metric, given a network topology, routes all the primary and protection

GA	ML	SPCO	LGCO	DEG2	DEGD
1	0.35	0.45			
2	0.35		0.45		
3	0.25		0.35	0.10	0.10

Table 5.1: Initial operator probabilities – topology

paths, determines the link capacities, and then uses link and node cost models to estimate an overall cost for the network. It incorporates, where appropriate, penalties for either failing to route, or to provide a protection path, for each individual traffic requirement. However, while it does take into account the high relative cost of optical switching in selecting paths, it does not include fibre choice or wavelength allocation. Clearly, the overall design objective is to minimise this network cost.

5.1.1 Initial Topology Experiments

The twelve operators developed by the author for mesh network topology design within NOMaD were described in §4.2.2 on p. 80. In devising experiments based on these, it was decided initially to focus on comparing a more straightforward algorithm, using just single-link mutation (ML) and single-point crossover (SPCO)[1], first with one incorporating a problem-specific crossover instead, link-group crossover (LGCO), and then adding in two problem-specific heuristics, the degree-two (DEG2) and degree-decrement (DEGD) operators. A subset of these experiments was first published by the author at the 3rd Intl. Conf. on Artificial Neural Networks and Genetic Algorithms (ICANNGA'97) [34] (reprinted in Appendix A).

The three GAs employed each had a population size of 200, used tournament selection (of size 4), and were limited to a maximum of 5,000 trials (*i.e.* somewhat more than 25 generations). The operators used and their initial probabilities are given in Table 5.1. The latter were selected by a few test runs, aiming for settings that would remain approximately constant for the first few generations across all the network design problems tackled. It could be argued that such starting values best promote high-quality solutions by approximately matching operator probabilities to their initial productivities.

It should be emphasised that for all three GAs the operator-probability adaptation mechanism described in §4.2.4 on p. 87 was employed. Default values for this were used throughout *i.e.* credit was awarded for improvement over the best fitness found at the end of the previous generation; the credit constant, κ, was 0.5; adaptation occurred every $\epsilon = 4$ generations; adaptation weight was

[1]However, recall that this is not a traditional bit-string GA, as both an object-oriented representation and operator-probability adaptation are being employed.

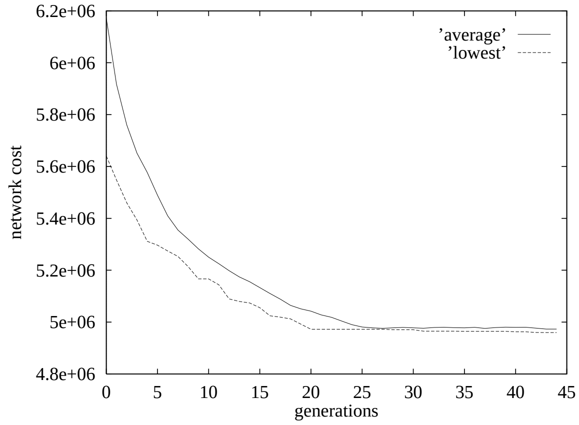
calculated *wrt.* children generated; the minimum probability level, $p_{\min}$, was set to 0.05; and the adaptation proportion, ψ, was taken to be 0.15.

The three GAs were each applied ten times with different pseudo-random number generator seeds to each of the five network design problems listed in §F.2. However, in assessing the designs, non-default values were used for certain parameters of the Network Cost Model A. The default values were originally selected by the author for design of the COST 239 European Optical Network [166], but this has both a larger number of nodes and covers a greater geographical area than the five problems considered here. Consequently, more appropriate parameter values needed to be selected. These, and the other non-default parameter values common to the three GAs are listed in Table F.11 in Appendix F.

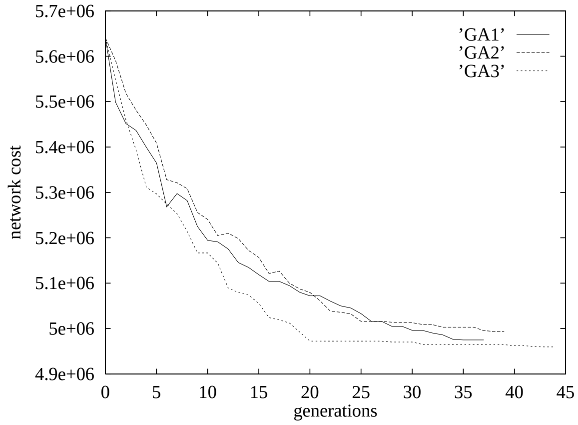
The progress of a single typical run of GA3 is given in Fig. 5.1(a), showing both the average network cost of the population and the lowest (*i.e.* best in population) cost, generation by generation. As can be seen from the average cost curve, after Generation 25 the population had substantially converged to the lowest network cost. The evolution of the best individual from the same run is illustrated in Fig. 5.2, which shows first the best individual in the initial random population (Generation 0), then at four subsequent points in the run, and finally the best overall (Generation 43). Considerable cost reductions were achieved in the first 13 generations, followed by more gradual refinement over the next 30. The final network obtained is close (at 4.960×10^6) to both the best obtained over ten runs each by the three GAs (4.953×10^6; not illustrated) and over five runs each (4.954×10^6), illustrated in Fig. 5.3(a) on p. 99. Comparing Figures 5.2(f) and 5.3(a), it can be seen that the two networks differ in only a few links, resulting in the small cost difference.

To compare the progress of typical runs from all three GAs, their lowest costs from individual runs all employing the same random seed, and thus starting with the same initial population, are given in Fig. 5.1(b). All three GAs employed the same maximum number of trials[2], but due to the different operators employed and their varying probabilities, the number of generations used necessarily differed. Clearly, at least for these three runs, GA3, with both problem-specific crossover and heuristics (in addition to mutation), made the most rapid progress and achieved the lowest cost. However, in this instance, the next lowest cost was achieved by GA1, just using mutation and single-point crossover. An additional point of interest in Fig. 5.1(b), is that the curve for GA1 clearly illustrates the non-elitist strategy employed by NOMaD: the lowest network cost *increases* from

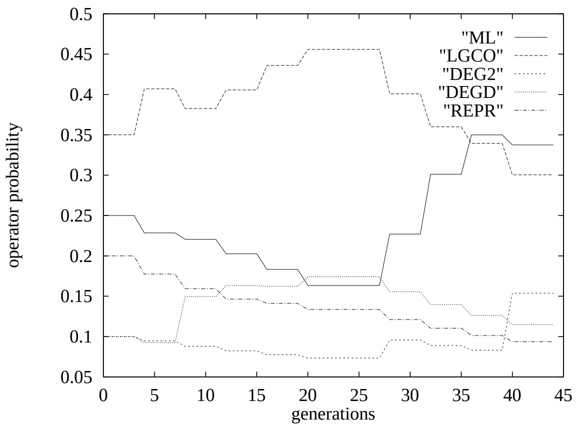
[2]Actual trials used will differ slightly, as here a GA is always allowed to complete the processing of its final generation, even if its maximum number of trials has been exceeded—see §3.9 and the description of the `GeneticAlgorithm` class on p. 74.



(a) Progress over a single GA3 run

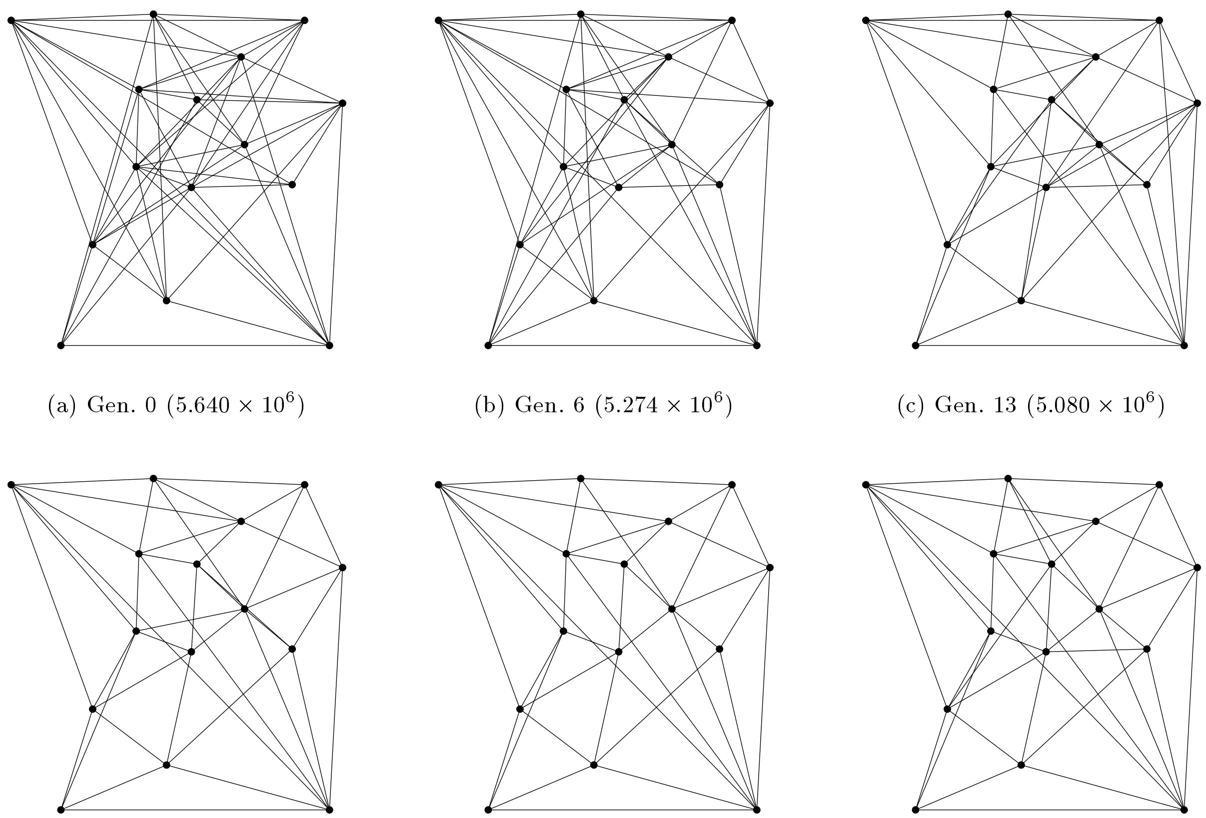


(b) Comparing lowest costs



(c) Operator-probability adaptation for GA3

Figure 5.1: Typical initial topology results (Problem 1, Run 1)



(a) Gen. 0 (5.640×10^6) (b) Gen. 6 (5.274×10^6) (c) Gen. 13 (5.080×10^6)

(d) Gen. 20 (4.972×10^6) (e) Gen. 31 (4.965×10^6) (f) Gen. 43 (4.960×10^6)

Figure 5.2: Best topologies over a single GA3 run (Problem 1, Run 1)

Problem	GA1 Best ($\times 10^6$)	GA1 Median ($\times 10^6$)	GA2 Best ($\times 10^6$)	GA2 Median ($\times 10^6$)	GA3 Best ($\times 10^6$)	GA3 Median ($\times 10^6$)	Significance Level (%) $s_{1,2}$	$s_{1,3}$	$s_{2,3}$
1	4.975	4.994	**4.953**	**4.975**	4.954	4.977	5.5	5.5	—
2	4.772	4.789	**4.751**	**4.764**	4.755	4.767	1.1	0.1	—
3	4.446	4.486	4.415	**4.421**	**4.404**	4.430	0.1	0.1	—
4	4.618	4.641	4.593	**4.612**	**4.591**	4.614	0.1	1.1	—
5	4.461	4.484	**4.424**	**4.450**	4.438	4.454	5.5	1.1	—

Table 5.2: Cost results for initial topology experiments

Generation 5 to 6. The lowest cost in the population is thus clearly distinct from the best so far.

The operation of the operator-probability adaptation mechanism for the same run of GA3 can be seen in Fig. 5.1(c). For the first 25 generations, while the population is converging—see Fig. 5.1(a), the operator probabilities remain substantially unchanged, with the majority of the work being undertaken by link-group crossover (LGCO). However, after this point, single-link mutation (ML) makes an increasing contribution, and link-group crossover less so, perhaps reflecting the need to introduce greater population diversity. It may be further speculated that the increase in probability of the degree-two operator (DEG2) at the end of the run results from an increased need to repair nodes almost or completely disconnected by destructive mutations.

In addition to the results provided in Fig. 5.1, the interested reader may like to note that further topology results, showing typical progress and operator-probability adaptation over single runs of both GA1 and GA2 on Problem 1, are illustrated in Fig. F.4.

Statistical comparison of the quality of results from the three GAs is given in Table 5.2. This records the best and median results from each group of ten runs, with both the best overall result and best median for each problem displayed in bold. In addition, as the same random seeds were used for each set of runs on a problem, a non-parametric sign test could be applied to matched pairs [307] from different GAs to establish if there were significant differences in their medians (for details, see §F.3.1 in Appendix F). The resulting significance levels are included in the table, with *e.g.* $s_{1,2}$ representing the significance level of the comparison of GA1 with GA2. For completeness, the full set of results from applying the GAs to all five of the network design problems, summarised here, is recorded in Tables F.12–F.14.

Surprisingly, the best medians for all five problems and the best overall result for three out of five were obtained by GA2, whereas GA3 only achieved the best result for two out of five problems. Clearly (for the problem sizes and types examined), adding two problem-specific heuristics to GA2, the degree-two and degree-decrement operators, so as to create GA3 has not resulted in the hoped for improvement in the final result quality. If anything, the reverse, with the necessary reduction

Network	1	2	3	4	5	6	7
Average node degree	5.33	5.07	5.07	4.93	4.80	4.00	3.33

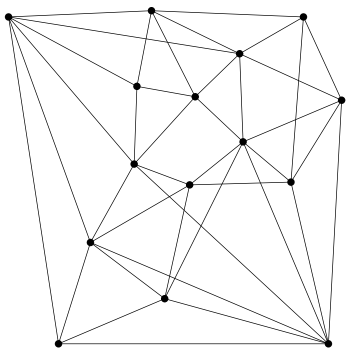
Table 5.3: Test network average node degrees

in link-group crossover probability to accommodate the two additional operators (see Table 5.1 on p. 93) perhaps responsible for the reduction in solution quality. However, it should be noted that these apparent differences between GA2 and GA3 are not reflected in a significant difference in their medians. In contrast, though, both GA2 and GA3 are significantly, and in some cases very highly significantly, better than GA1 in terms of median differences. Thus, as anticipated, employing link-group crossover provides a consistent improvement in solution quality over the use of single-point crossover.

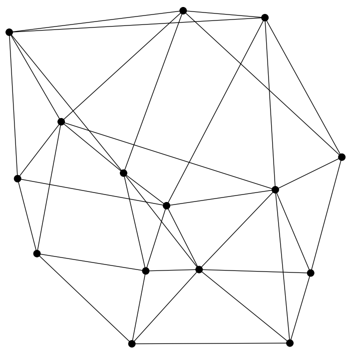
5.1.2 Test Networks

In addition to comparing different GAs in the initial experimental work, the opportunity was taken to select several topologies for use in subsequent wavelength allocation work. The initial choice of five networks was made when only five runs had been completed by each of the three GAs on the five network design problems introduced in §F.2. Consequently, it was the lowest-cost topologies at this stage, rather than the best of ten runs each, recorded in Table 5.2, that were selected. These five networks are illustrated in Fig. 5.3, and details of both their selection and their links are recorded in Tables F.20 and F.21, respectively, in Appendix F.

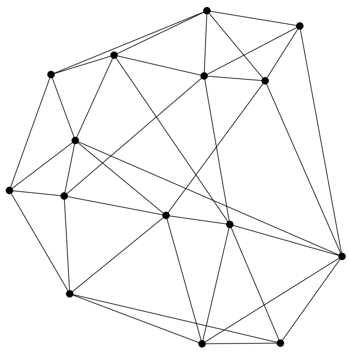
Inspection of the five test networks revealed, however, that they were all fairly similar and had relatively high average node degree, so to provide more diverse topologies, two further networks were generated from Problem 1. These have increasingly lower average node degrees, achieved by suitably modifying the parameters of Network Cost Model A. The actual parameter sets used are recorded in Tables F.16 and F.17. The first introduces a non-linear capacity dependency by lowering the value of α to 0.8; the second also increases the relative cost of network nodes by increasing K_0 and K_n. The results of applying GA2 and GA3 five times each, using the two parameter sets in turn, are given in Tables F.18 and F.19, respectively. The two networks selected (the lowest cost for each of the two parameter sets) are illustrated in Fig. 5.4, and then the average node degrees for all seven test networks are recorded in Table 5.3.



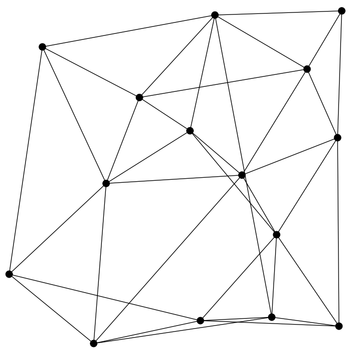
(a) Network 1



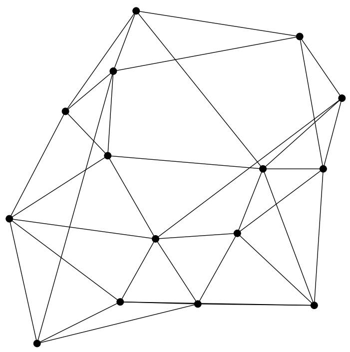
(b) Network 2



(c) Network 3

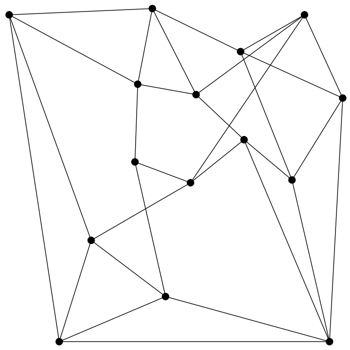


(d) Network 4

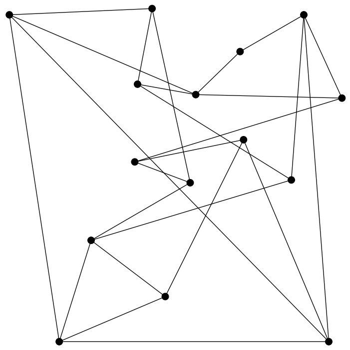


(e) Network 5

Figure 5.3: Test networks



(a) Network 6



(b) Network 7

Figure 5.4: Low-average-node-degree test networks

5.1.3 Local Search Operators

The second set of topology experiments focuses on a comparison between a straightforward GA employing just single-link mutation (ML) and link-group crossover (LGCO), and GAs which also include one or more local search heuristics. These are random add link (RAL), random remove link (RRL) and random change link (RCL; all three described in §4.2.2 on p. 81). These operators make a series of random link additions, removals or changes, respectively, until they either make an improvement or detect they are trapped in a local optimum.

In addition, as each of these operators can potentially use many trials whilst performing local search, it was felt appropriate to investigate the per-trial variant of the basic operator-probability adaptation mechanism (see §4.2.4 on p. 90). In the basic approach, operator credit is divided by the number of children each operator has generated before being used to reassign operator probabilities. However, if an operator uses more than one trial per child generated, this may unfairly favour operators that are expensive in their use of trials. An alternative is to use the number of trials itself as the divisor before probability reassignment. Thus an operator will have to provide benefit *proportional to the number of trials used* to gain probability share.

The nine GAs each had a population size of 200, used tournament selection (of size 4), and were limited to a maximum of 15,000 trials. The larger number of trials (*cf.* GA1–GA3) was discovered to be necessary (using a few test runs) to allow those GAs including local search operators to complete a reasonable number of generations (more than 25)[3]. The operators used and their initial probabilities are given in Table 5.4. Other non-default parameter values common to the nine GAs are listed in Table F.22 in Appendix F. Essentially, these GAs differ from those used in the initial topology experiments only in respect to: choice of operators and initial probabilities, increased run length, and whether per-child or per-trial operator-probability reassignment is used. In respect of the latter parameter, GA4, GA5, GA7, GA9 and GA11 all use child-based, whereas GA6, GA8, GA10 and GA12 use trial-based reassignment. In particular, GA4 is identical to the successful GA2 from the initial experiments, except for its larger maximum number of trials.

The nine GAs were each applied ten times with different random seeds to each of the five network design problems listed in §F.2. The progress of a single typical run of GA4 on Problem 1 is given in Fig. 5.5(a), showing both the average network cost of the population and the lowest (*i.e.* best in population) cost, generation by generation. The operator-probability adaptation curves for

[3]It could be argued that this consequently gives an unfair advantage to those GAs incorporating local search operators.

GA	ML	LGCO	RAL	RRL	RCL
4	0.35	0.45			
5/6	0.15	0.40	0.30		
7/8	0.15	0.40		0.30	
9/10	0.15	0.40			0.30
11/12	0.15	0.40	0.15	0.15	

Table 5.4: Initial operator probabilities – topology using local search operators

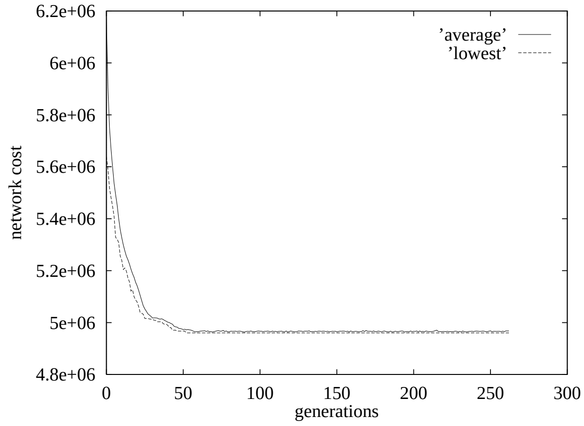
the same run are shown in Fig. 5.5(b). The adaptation curves can clearly be seen to have 'flat lined' after about Generation 50. The reason for this is that the adaptation mechanism is using the default credit-assignment approach, which rewards operators for improvements over the best so far individual. If no further improvement is made, then the operator probabilities are left unchanged, giving this 'flat line' appearance[4]. Thus, GA4 is hardly benefiting from the increased run length over the earlier GA2, which ran for nearly 40 generations (see Fig. F.4(c) on p. 310), with most evolution occurring in the first 30.

To provide a comparison between two otherwise identical GAs using either per-child or per-trial probability reassignment, Figures 5.5(c) and (d) show the adaptation curves for typical runs of GA9 and GA10, respectively, on Problem 1. Both GAs use the random change link operator (RCL), and in Fig. 5.5(c), with per-child reassignment, the random change link operator can be seen gaining probability largely at the expense of link-group crossover (LGCO). However, in Fig. 5.5(d), with per-trial allocation, a fairer assessment of the contribution of this multi-trial local search heuristic results in its probability declining, and link-group crossover earning an increasing share. Similar changes, resulting from using per-trial allocation, can be seen in comparing adaptation curves for typical runs of GA5 and GA6 (employing the random add link operator, RAL) and between GA7 and GA8 (with random remove link, RRL)—see Figures F.5 and F.6.

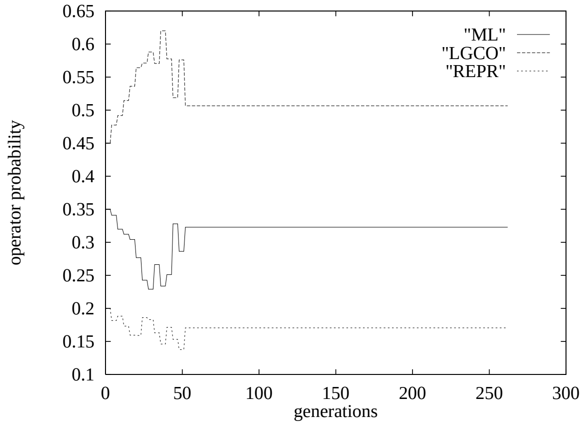
The lowest cost curves for five GAs, all but GA4 employing trial-based reassignment[5], and all using the same random seed and thus starting with the same initial population, are given in Figures 5.5(e) and (f). The former is plotted on a generation-by-generation basis, whereas the latter shows network cost against trials. As was noted *wrt.* the initial experiments, although all five GAs

[4] An investigation into *why* no further improvement occurred indicated that the population had almost completely converged, at least in fitness terms, and that single-link mutation was unable to discover further improvements over the best-so-far fitness. This was confirmed by increasing the level of mutation to 0.8, both with and without operator probability adaptation; in both cases a similar final outcome was observed. Thus, the use of a more disruptive mutation operator (changing more than one link at a time) may be beneficial, or perhaps the use of local search operators, the approach adopted here.

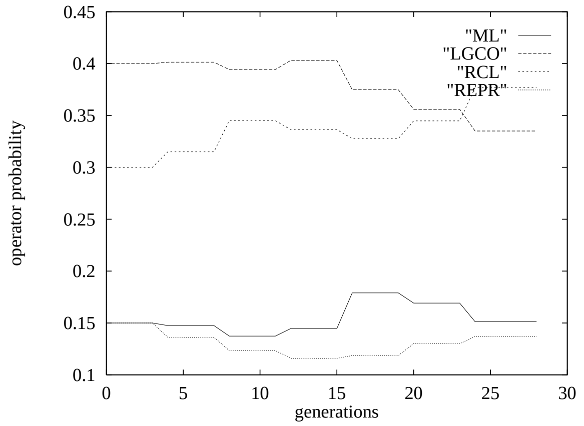
[5] However, as GA4 does not include any multi-trial operators, the run would have been (almost) identical using trial-based reassignment. The only differences would arise because of the different handling of the reproduction operator—see §4.2.4 on p. 90.



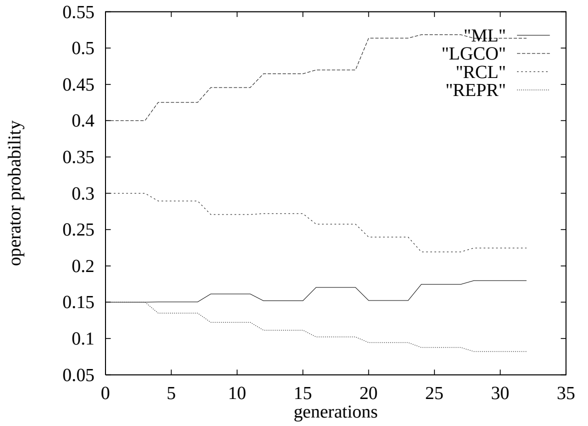
(a) Progress over a single GA4 run



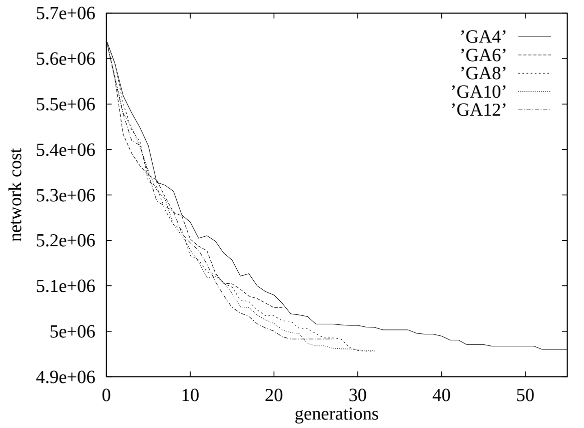
(b) Operator-probability adaptation for GA4



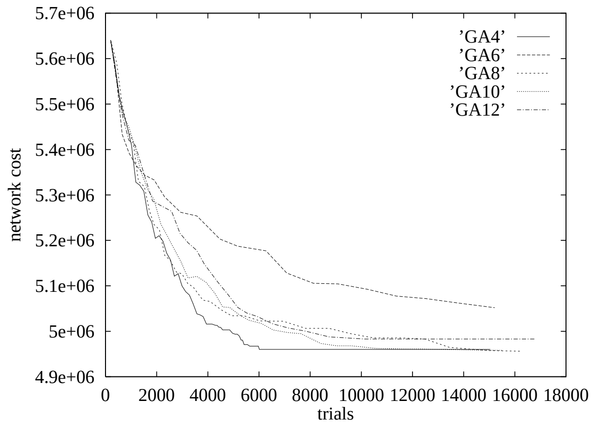
(c) Operator-probability adaptation for GA9



(d) Operator-probability adaptation for GA10



(e) Comparing lowest costs (*vs.* generations)



(f) Comparing lowest costs (*vs.* trials)

Figure 5.5: Typical topology results using local search operators (Problem 1, Run 1)

Problem	GA4 ($\times 10^6$)	GA5 ($\times 10^6$)	GA6 ($\times 10^6$)	GA7 ($\times 10^6$)	GA8 ($\times 10^6$)	GA9 ($\times 10^6$)	GA10 ($\times 10^6$)	GA11 ($\times 10^6$)	GA12 ($\times 10^6$)
1	4.947	4.988	4.987	4.953	4.956	4.947	**4.946**	4.957	4.951
2	4.746	4.838	4.809	4.747	**4.745**	4.747	4.748	4.750	4.753
3	4.409	4.474	4.466	4.418	4.407	**4.400**	4.408	4.447	4.415
4	4.591	4.692	4.616	4.592	4.598	4.594	**4.589**	4.619	4.593
5	4.424	4.513	4.452	4.435	4.434	**4.417**	4.426	4.439	4.423

Table 5.5: Best cost results for topology using local search operators

Problem	GA4 ($\times 10^6$)	GA5 ($\times 10^6$)	GA6 ($\times 10^6$)	GA7 ($\times 10^6$)	GA8 ($\times 10^6$)	GA9 ($\times 10^6$)	GA10 ($\times 10^6$)	GA11 ($\times 10^6$)	GA12 ($\times 10^6$)
1	4.958	5.062	5.021	4.980	4.969	4.962	**4.957**	4.986	4.975
2	**4.753**	4.871	4.840	4.766	4.763	4.758	4.755	4.774	4.759
3	**4.412**	4.539	4.496	4.425	4.420	4.421	4.414	4.463	4.430
4	**4.601**	4.718	4.672	4.607	4.608	4.604	4.602	4.639	4.616
5	**4.430**	4.569	4.527	4.452	4.452	4.440	4.438	4.473	4.459

Table 5.6: Median cost results for topology using local search operators

employed the same maximum number of trials, due to the different operators employed and their varying probabilities, the number of generations used necessarily differed. This effect is particularly marked here as, since GA4 does not employ local search operators, it uses considerably fewer trials per generation, and thus continues for more than 250 generations. Consequently, Fig. 5.5(f) provides the better basis for comparison. Overall, the best result for these runs is obtained by GA8, closely followed by GA10, however GA4 not only showed the most rapid progress, but also provided the best overall result up to 10,000 trials (although it seemed to make no further improvement after about the first 6,000 trials). Only by extending their runs beyond this point were two of the GAs incorporating local search operators able to produce (small) improvements in final solution quality over GA4.

Supplementing the results provided in Fig. 5.5, additional topology results, showing typical progress and operator-probability adaptation over single runs of GA5–GA12 on Problem 1, are provided in Figures F.5–F.8.

A summary of the results of applying the GAs is recorded in Tables 5.5 and 5.6, which list the best and median network costs, respectively, for each group of ten runs. In each table, either the best overall result or the best median for each problem are displayed in bold. Moreover, the full set of results from applying the GAs to all five of the network design problems is recorded in Tables F.23–F.31.

Of the best overall results, two were obtained by GA10 (random change link; trial-based), two by GA9 (random change link; child-based) and one by GA8 (random remove link; trial-based). However, four of the five best medians are due to GA4 (the straightforward GA), with only one

Problem	Significance level (%)							
	$s_{5,4}$	$s_{6,4}$	$s_{7,4}$	$s_{8,4}$	$s_{9,4}$	$s_{10,4}$	$s_{11,4}$	$s_{12,4}$
1	0.1	0.1	1.1	5.5	—	—	5.5	5.5
2	0.1	0.1	—	—	—	—	0.1	5.5
3	0.1	0.1	—	—	5.5	—	0.1	5.5
4	0.1	0.1	5.5	5.5	5.5	—	0.1	5.5
5	0.1	0.1	5.5	1.1	—	—	0.1	5.5

Table 5.7: Comparing GA4 with GAs using local search operators

best median obtained by GA10 (random change link; trial-based). Clearly, while some of the best results were obtained by those algorithms enhanced by local-search operators, the straightforward algorithm, with only single-link mutation and problem-specific crossover, performed well especially in median terms.

To assess the statistical significance of these differences, a non-parametric sign test was again applied to matched pairs from different GAs [307]. First, Table 5.7 shows comparisons between all those GAs using local search operators and GA4. Clearly, this more straightforward algorithm is significantly better for many of the problems, in terms of median difference, than all of the other GAs, with the single exception of GA10 (random change link; child-based), where the difference is not significant. These median differences are particularly marked for the two algorithms including the random add link operator, GA5 and GA6, all being very highly significant. Given that the best initial random networks are often more highly connected than those eventually evolved (*cf.* Figures 5.2(a) and (f) on p. 96), it is likely that using this local search operator, which can only *add* links, is counter-productive, serving only to limit the effectiveness of link-group crossover. The next two GAs, GA7 and GA8, use the random remove link operator, which by *removing* unwanted links has a greater chance of improving the initially over-connected networks. This is reflected in a less conclusive victory by GA4, with no significant median difference for two problems, and indeed a best overall result in one case for GA8 (trial-based). However, it would appear that the greater freedom from *both adding and removing* links, resulting from the use of the random change link operator in GA9 and GA10, leads to almost no significant median differences with GA4, and indeed the best overall results in four out of five problems. Interestingly, providing two separate operators to both add and remove links, as is done in GA11 and GA12, does not seem as successful as the use of the combined random change link operator, with significantly better medians obtained by GA4 over both GA11 and GA12 in all cases.

Next, Table 5.8 compares the otherwise identical child and trial-based credit division GAs directly. While for both GA6 over GA5 (random add link), and GA12 over GA11 (both random add

Problem	Significance level (%)			
	$s_{5,6}$	$s_{7,8}$	$s_{9,10}$	$s_{11,12}$
1	0.1	—	—	5.5
2	5.5	—	—	1.1
3	1.1	—	1.1	1.1
4	0.1	—	—	5.5
5	1.1	—	—	1.1

Table 5.8: Comparing child and trial-based credit division for GA5–GA12

Problem	Significance level (%)		
	$s_{6,10}$	$s_{8,10}$	$s_{12,10}$
1	0.1	—	5.5
2	0.1	5.5	5.5
3	0.1	—	1.1
4	0.1	5.5	1.1
5	0.1	—	5.5

Table 5.9: Comparing GA10 with other trial-based GAs using local search operators

and remove link), the trial-based algorithms are better than the child-based, only one significant median difference was found between the other two pairs of GAs. Nevertheless, given earlier reasoning, those median differences found, and that the best results were discovered by trial-based GAs for three out of the five problems, it seems reasonable to adopt trial-based probability reassignment for GAs incorporating such multi-trial operators.

Finally, Table 5.9 compares the other three trial-based credit division GAs with GA10 (random change link), which appears the strongest of those GAs incorporating local search operators. In all three cases, significant median differences were found, confirming this view, although this was perhaps less marked in the case of GA8 (random remove link), whose potential has already been noted above.

Overall, it would seem that GA4 (for the problem sizes and types examined), with just single-link mutation and link-group crossover, is a consistent algorithm producing good results. Only when given sufficient runs of adequate length might a better overall result be achieved with a GA incorporating a flexible local search operator, random-change link, as well as applying trial-based probability reassignment. This then confirms the strength of the basic GA4 algorithm, and in particular the usefulness of link-group crossover.

5.1.4 Iterative Heuristic Operators

The final set of topology experiments explores the use of the three iterative heuristic operators for mesh network topology design. These are random add links (RALS), random remove links (RRLS) and random change links (RCLS; all three described in §4.2.2 on p. 82). All of them are extensions

GA	ML	LGCO	RALS	RRLS	RCLS
4	0.35	0.45			
13	0.20	0.50	0.10		
14	0.20	0.50		0.10	
15	0.20	0.50			0.10

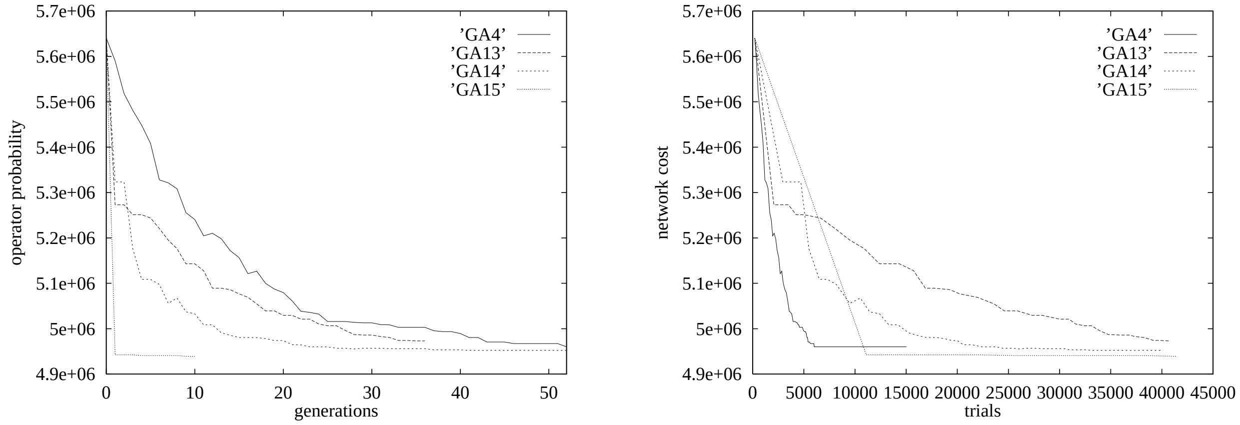
Table 5.10: Initial operator probabilities – topology using iterative heuristic operators

of the corresponding local search operators used in the previous section. However, rather than stopping after the first improvement is found, they each continue until a local optimum is detected. Thus each one is a complete iterative stochastic hill climber [29, 30]. Once again they will be compared against a more straightforward algorithm employing only single-link mutation (ML) and link-group crossover (LGCO).

The three GAs incorporating iterative heuristic operators (GA13, GA14 and GA15) each had a population size of 200, used tournament selection (of size 4), trial-based credit division, and were limited to a maximum of 40,000 trials. The further increase in the number of trials (*cf.* GA4–GA12) was again discovered to be necessary (using a few test runs) to allow the GAs, particularly GA15 (random change links), to complete a reasonable number of generations[6]. The operators used and their initial probabilities are given in Table 5.10. Other non-default parameter values common to the three GAs are listed in Table F.32 in Appendix F. Essentially, these GAs differ from those used in the local search experiments only in respect to: choice of operators and initial probabilities, increased run length, and that all three use per-trial operator-probability reassignment. However, GA4 is here identical to that used in the last section. As the algorithm makes no further progress after about 6,000 trials, and was in any case being run for 15,000, it seemed unnecessary to increase this to 40,000 trials and repeat the earlier runs.

The three GAs using iterative heuristic operators were each applied ten times with different random seeds to each of the five network design problems listed in §F.2. The lowest cost curves for these three GAs, as well as GA4, all using the same random seed and thus starting with the same initial population, are given in Figures 5.6(a) and (b); the former generation by generation, the latter *vs.* trials. Overall, the best result for these runs is obtained by GA15 (random change links), followed by GA14 (random remove links), however GA4 once again showed the most rapid progress, producing the best overall result up to 10,000 trials. Only beyond this point were two of the GAs incorporating iterative heuristic operators eventually able to overtake it. This again demonstrates the utility of the basic GA4 algorithm, incorporating just single-link mutation and

[6]Once again, this arguably gives an unfair advantage to those GAs incorporating iterative heuristic operators.



(a) Comparing lowest costs (*vs.* generations) (b) Comparing lowest costs (*vs.* trials)

Figure 5.6: Typical lowest costs using iterative heuristic operators (Problem 1, Run 1)

Problem	GA4		GA13		GA14		GA15	
	Best ($\times 10^6$)	Median ($\times 10^6$)	Best ($\times 10^6$)	Median ($\times 10^6$)	Best ($\times 10^6$)	Median ($\times 10^6$)	Best ($\times 10^6$)	Median ($\times 10^6$)
1	4.947	4.958	4.936	4.951	4.936	4.948	**4.935**	**4.937**
2	4.746	4.753	4.747	4.755	4.739	4.756	**4.736**	**4.742**
3	4.409	4.412	4.408	4.426	**4.400**	4.411	**4.400**	**4.402**
4	4.591	4.601	4.591	4.603	4.590	4.594	**4.585**	**4.587**
5	4.424	4.430	4.419	4.441	4.419	4.430	**4.416**	**4.418**

Table 5.11: Cost results for topology with iterative heuristic operators

link-group crossover.

Expanding on the results provided in Fig. 5.6, extra topology results, showing typical progress and operator-probability adaptation over single runs of GA13–GA15 on Problem 1, are given in Fig. F.9.

A summary of the results of applying the three GAs, as well as GA4, is recorded in Table 5.11, which lists the best and median network costs for each group of ten runs. As before, the best overall result and the best median for each problem are displayed in bold. In addition, the full set of results from applying the three GAs to all five of the network design problems is recorded in Tables F.33–F.35; the earlier GA4 results are listed in Table F.23.

Clearly, GA15 (random change links) is the best of the four algorithms, having obtained the best overall result and best medians for all five problems, although admittedly GA14 (random remove links) tied for best overall in one case. To assess the statistical significance of the median differences, a non-parametric sign test [307] was applied to compare matched pairs from the other algorithms with GA15. The results are listed in Table 5.12, which records that while there are almost no significant median differences between either GA13 or GA14 and GA4, GA15 is very

Problem	Significance level (%)		
	$s_{4,13}$	$s_{4,14}$	$s_{4,15}$
1	—	—	0.1
2	—	—	0.1
3	—	—	0.1
4	—	5.5	0.1
5	—	—	0.1

Table 5.12: Comparing GA4 with GAs using iterative heuristic operators

highly significantly better than GA4.

Overall, it would seem that for mesh network topology design (and the problem sizes and types examined) combining an iterative heuristic operator with the solid GA framework provided by GA4, including its use of single-link mutation and link-group crossover, consistently produces better results, although at the cost of extended run lengths (in terms of trials, rather than generations).

5.2 Routing, Fibre & Wavelength Allocation

Once a topology has been determined, the next problem is the network routing, fibre choice and wavelength allocation. The networks considered in the thesis are flat, all-optical, with arbitrary-mesh multiple-fibre-link topologies, and employ WP routing. Consequently, as was mentioned in §3.1, first the static traffic requirement between each node pair in the network has to be partitioned into an appropriate number of separate wavelength channels. Then, for each of these, a route through the network must be found, including which fibre and wavelength will be used on each link (cable) along the path. As WP routing is employed, each channel must use the same wavelength end-to-end, since no wavelength conversion is available, as well as obeying the usual constraint that on any one fibre, no wavelength is used by more than one channel. However, although restoration paths are supported during network topology design, NOMaD's network model provides no direct support for the fibre and wavelength allocation of restoration paths, and so these are not included in the allocation.

For the experiments in this section, the test networks described in §5.1.2 on p. 98 were used. Each link was taken to be a cable of two fibres, and the capacity of the individual wavelength channels set to 10 Gbit/s. The total wavelength channel requirements for the networks[7], calculated from the traffic matrices in §F.2, are given in Table 5.13.

[7]Recall that Networks 1, 6 and 7 were all designed for Problem 1, and so have identical traffic requirements.

Network	1/6/7	2	3	4	5
Channel requirement	268	248	258	262	268

Table 5.13: Total wavelength channel requirements of the test networks

GA	WKM	WKCO	WWRR	WWSO
16	0.25	0.50		
17	0.05	0.15	0.30	0.35

Table 5.14: Initial operator probabilities – minimum NWR

5.2.1 Minimum Network Wavelength Requirement

The first set of experiments employ the network wavelength requirement (NWR) metric outlined in §4.1.2 on p. 71. This modifies the actual NWR by adding a penalty for each unsatisfied traffic requirement (*i.e.* wavelength channel that remains unrouted) to avoid the false minimum of carrying no traffic at all, which would otherwise have an NWR of zero[8].

The four operators developed by the author for routing, fibre and wavelength allocation within NOMaD were described in §4.2.3 on p. 82. In devising experiments based on these, it was decided to compare, for two different GAs, the best of five runs with results obtained from three recent wavelength-allocation heuristics published in the literature[9]. The first GA is a more straightforward algorithm[10], which uses just wavelength-allocation k-shortest path mutation (WKM) and wavelength-allocation k-shortest path single-point crossover (WKCO), while the second GA also includes the two problem-specific heuristics, wavelength-allocation path reroute (WWRR) and wavelength-allocation path shift-out (WWSO). These experiments were first published by the author in *Electronic Letters* [35] (reprinted in Appendix A).

Both GAs had a population size of 500 and used tournament selection (of size 4); their initial operator probabilities are given in Table 5.14. Other non-default parameter values for GA16 and GA17 are listed in Tables F.36 and F.37 in Appendix F. It should be emphasised that they both used the operator-probability adaptation mechanism with default parameters throughout.

The two GAs were both applied five times with different random seeds to Networks 1–5. The progress of single typical runs of GA16 and GA17 on Network 1 are given in Figures 5.7(a) and (c), showing both the average NWR of the population and the lowest (*i.e.* best in population) NWR,

[8]The additional per-fibre penalty, described in §4.1.2, was not used in these experiments

[9]There are relatively few heuristics that tackle the specific allocation problem addressed here, and so those selected were simply those that the author was able to locate at the time. Additional heuristics discovered since will be included in the suggestions for further work (Ch. 6)

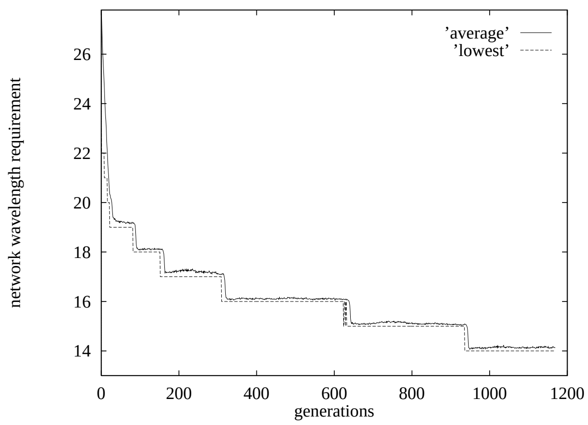
[10]As with GA1 for mesh network topology design, this is not a traditional bit-string GA, as both an object-oriented representation and operator-probability adaptation are being employed.

generation by generation. The operator-probability adaptation curves for the same runs are shown in Figures 5.7(b) and (d).

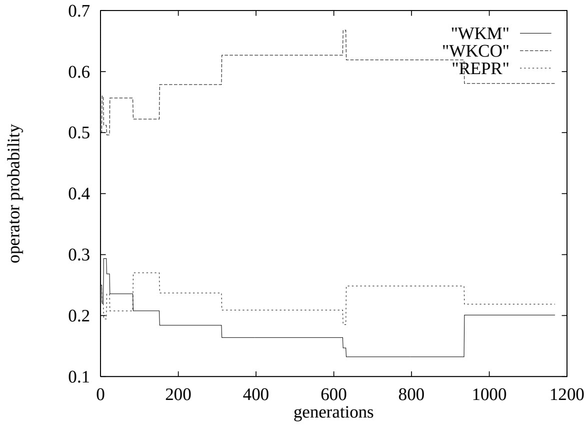
In contrast to most of the topology curves, all four figures show abrupt changes in value. This results from the integer-valued nature of NWR: whenever there is a new best individual in the population, its NWR is always at least a full unit lower than the previous best. In both Figures 5.7(a) and (c), this improvement appears to rapidly propagate through the population, as the average NWR swiftly drops to close to the best. In addition, as the operator-probability adaptation mechanism is driven by credit assigned when a new best individual is found, the integer-valued nature of NWR also results in fairly abrupt changes in operator probabilities, as can be observed in both Figures 5.7(b) and (d).

Turning to the NWR results, there is clearly a considerable difference in the progress of the two GAs. Whilst GA17 takes about 80 generations to obtain an NWR of 9, GA16 is only able to achieve an NWR of 14 using nearly 1200 generations (500,000 trials). This difference is even more apparent in Fig. 5.7(e) which directly compares the lowest NWR curves for the two GAs. Both these runs employed the same random seed, and thus started from the same initial population. From the operator-probability adaptation curves it appears that, at least for these runs, GA16 chiefly depends on the crossover operator (WKCO), with a modest contribution from mutation (WKM), whereas GA17 makes full use of both its problem-specific heuristic operators, shift-out (WWSO) and reroute (WWRR), with only minimal contributions from crossover and mutation.

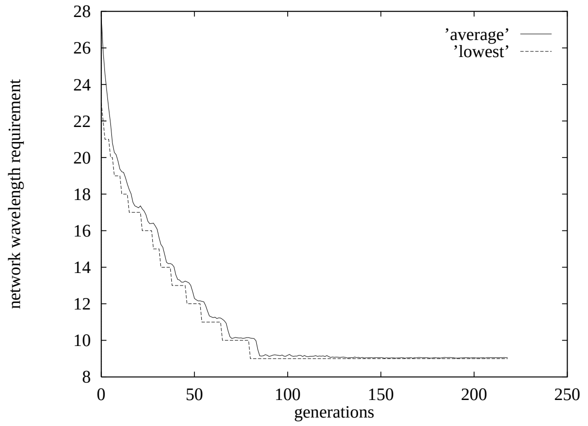
In addition to the two GAs, a wavelength-allocation heuristic by Wuttisittikulkij & O'Mahony and two by Wauters & Demeester were also applied to Networks 1–5. Wuttisittikulkij & O'Mahony's heuristic [298] (W&O) is simple, fast and non-iterative, working with k-lowest hop count paths (and multiple fibres) to obtain a low NWR. The first heuristic by Wauters & Demeester [306] (W&D1) is based on a version of Dijkstra's algorithm modified such that in "extend[ing] an already found part of a route, only those [cables] are considered which have a wavelength channel free which is available on all [cables] of the already found part of the route"; this algorithm produces a low-cost allocation, whilst also keeping NWR low. Their second algorithm (W&D2), is based on iterative application of a modified version of Yen's k-shortest path algorithm, and aims to produce the lowest cost allocation that still achieves minimum NWR. Description of all three heuristics is given in Appendix C, and their non-default parameters listed in Table F.40. Of the three algorithms, Wuttisittikulkij & O'Mahony's is non-deterministic and so, like the GAs, it was allowed the best of



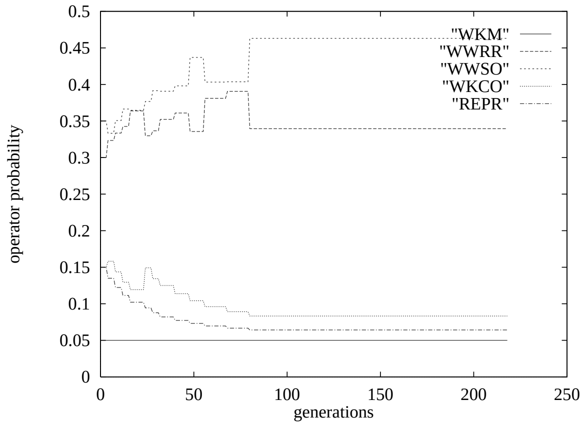
(a) Progress over a single GA16 run



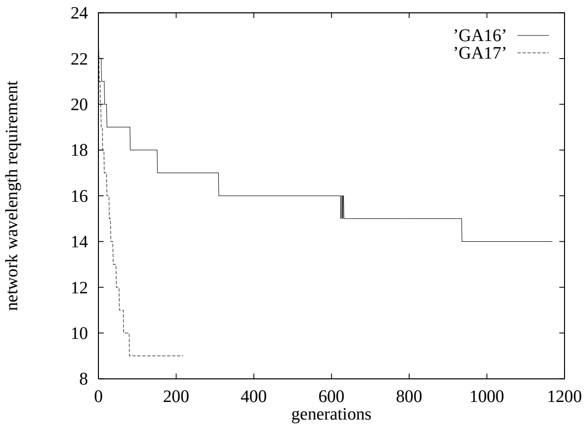
(b) Operator-probability adaptation for GA16



(c) Progress over a single GA17 run



(d) Operator-probability adaptation for GA17



(e) Comparing lowest NWR

Figure 5.7: Typical minimum NWR results (Network 1, Run 1)

Network	W&O	W&D1	W&D2	GA16		GA17	
	NWR	NWR	NWR	NWR	Trials	NWR	Trials
1	**8**	15	10	13	351,264	9	31,874
2	**7**	11	**7**	10	401,919	**7**	38,291
3	8	19	9	10	448,235	**7**	44,467
4	**8**	15	9	11	429,485	**8**	37,997
5	9	18	9	12	428,962	**8**	48,448

Table 5.15: Results for minimum NWR

five runs[11].

A summary of the results of applying the two GAs and three heuristics is recorded in Table 5.15, which lists the NWR, or the best NWR for five runs as appropriate, as well as the number of trials used to find the recorded NWR in the best GA run in each case[12]. As before, the best overall result for each network is displayed in bold. In addition, the full set of results from applying the two GAs and W&0 to all five of the networks is recorded in Tables F.38, F.39 and F.41, respectively.

Of the three heuristic algorithms, the poorest in this context is W&D1, which in aiming for low network cost, has obtained relatively high NWRs. Although the more recent W&D2 has clearly improved on this in NWR terms, the best heuristic overall is W&O, which despite its simplicity, achieves good NWR results. Although over the five test networks, GA16 consistently obtained better NWRs than W&D1, it was no match for the other two heuristics. However, GA17 equalled or exceeded the results of all three heuristics, with one exception, and using considerably fewer trials than GA16. Clearly, therefore, the addition of the two heuristic operators (reroute and shift-out), has led to a marked improvement of GA17 over GA16. In addition, despite these two additional operators being based on steps of the W&D2 heuristic, GA17 has in many cases exceeded the results of W&D2, managing to match the best heuristic, W&O.

Overall, these experimental results show that, at least for the five test networks, the GA/heuristic hybrid approach is as promising as has been hoped, as the blend of a GA with steps from Wauters & Demeester's HRWA heuristic [306] resulted in both improved performance (over the GA alone) and improved results (over the heuristic alone). In addition, the GA/heuristic hybrid results matched those of the best wavelength-allocation heuristic examined.

[11]The option to allocate the paths in order of hop count, largest first, was exercised, as this was both recommended by the heuristic's authors and found to obtain lower NWR results in trial runs.

[12]Note that as none of the heuristics make use of trials, as such, their comparative effort is zero!

5.2.2 Minimum Cost Allocation

The second set of experiments on routing, fibre choice and wavelength allocation used Network Cost Model B, outlined in §4.1.3 on p. 72. This changes the assessment from just looking at the NWR to an objective that consists of a simplified assessment of the cost of the whole network including, at an appropriate level, the cost of exceeding the minimum NWR. The cost model is parameterised, amongst others, by three constants, (α, β, γ). The first two, α and β, adjust the model of link cost in the network to allow for different dependencies on capacity. With $\alpha = \beta = 1$, the dependency is linear—doubling the capacity doubles the cost—whereas with $\alpha = \beta < 1$, the cost increases more slowly with capacity[13]. The remaining parameter, γ, allows the cost model to reflect the relative contributions to link and node costs of the capacity and link wavelength requirements. With $\gamma = 1$, the cost model only takes account of the actual capacity used; whereas, with $\gamma = 0$, the capacity is ignored, and the focus is entirely on the costs resulting from wavelength requirements. An intermediate value of γ, such as 0.5, is regarded as being more representative of the actual costs involved.

For these experiments, the the same four routing, fibre and wavelength allocation operators used in GA17, and described in §4.2.3 on p. 82, were used. Given the outcome of the experiments on minimum NWR design, it was decided to immediately employ all four operators within a GA, rather than further explore ineffective algorithms that did not include the problem-specific heuristic operators. Instead, the emphasis was on competing, on a network cost basis, with the same three overall wavelength-allocation heuristics used for the minimum NWR study. These experiments were first published at the 6th IEE Conf. on Telecommunications [36] and subsequently in *IEE Proc. Communications* [37] (both reprinted in Appendix A).

The three GAs used were identical except that they had different initial operator probabilities (given in Table 5.16). These were selected for problems with differing values of γ by using the usual approach of a few test runs, aiming for settings that would remain approximately constant for the first few generations across all the networks examined. All three GAs had a population size of 500, a maximum of 100,000 trials, and used tournament selection (of size 4). Other non-default parameter values for GA18–20 are listed in Table F.42 in Appendix F. It should be emphasised that they all used the operator-probability adaptation mechanism with default parameters throughout.

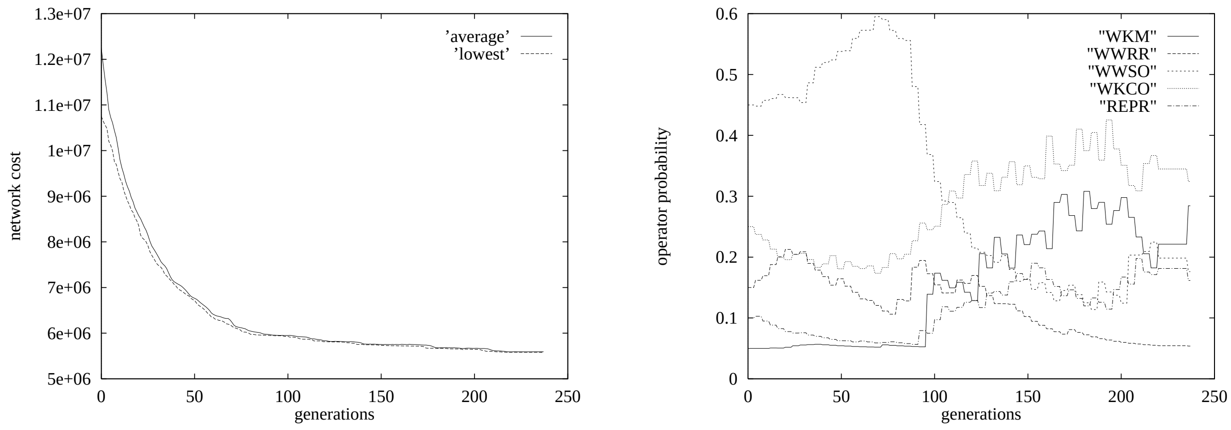
All three GAs were applied five times with different random seeds (*i.e.* best of five runs) to

[13] In these experiments, α and β were always given the same value, however, β is included in the model to fine-tune the dependency of cost on link wavelength requirement rather than just capacity alone

GA	WKM	WKCO	WWRR	WWSO
18	0.05	0.25	0.15	0.45
19	0.05	0.30	0.10	0.40
20	0.05	0.20	0.25	0.40

Table 5.16: Initial operator probabilities – minimum cost allocation

Networks 1–7, with γ set to 0.5, 1 and 0 for GA18, GA19 and GA20, respectively; first for $\alpha = \beta = 1$, then for $\alpha = \beta = 0.8$. For all three GAs (and the heuristics–see below), k was set to 8 for the Networks 1-5, and 4 for the remaining two, reflecting their lower average node degrees. The progress of single typical runs of GA18 on Network 1 with $(\alpha, \beta, \gamma) = (1, 1, 0.5)$ and $(0.8, 0.8, 0.5)$ are given in Figures 5.8(a) and 5.9(a), respectively, showing both the average network cost of the population and the lowest (*i.e.* best in population) cost, generation by generation. The operator-probability adaptation curves for the same runs are shown in Figures 5.8(b) and 5.9(b).

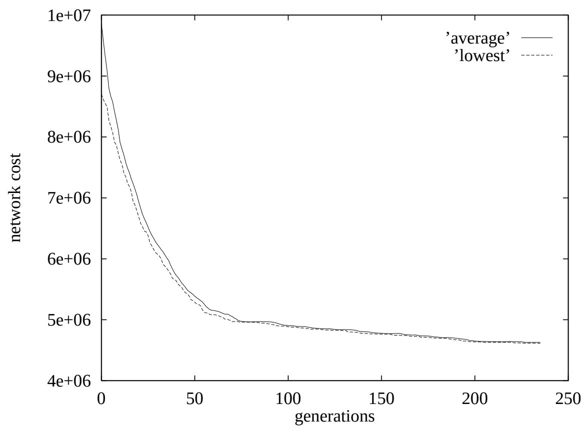


(a) Progress over a single run

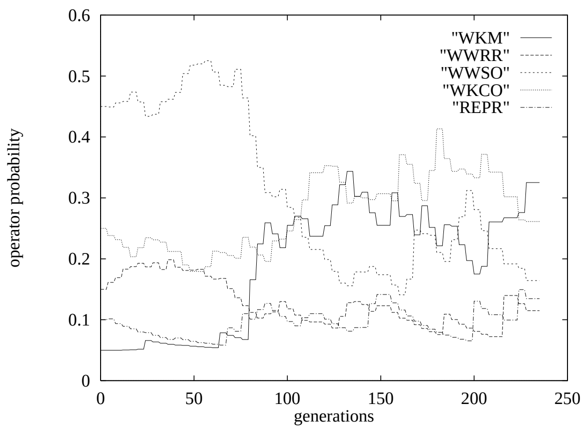
(b) Operator-probability adaptation

Figure 5.8: Typical results for GA18, $\alpha = \beta = 1$ and $\gamma = 0.5$ (Network 1, Run 1)

Both progress curves show a similar pattern of rapid initial reduction in network cost up to about generation 100, followed by more gradual improvement for the balance of the run. Examination of the operator probability adaptation curves indicates that, during this first phase of the run, the dominant operator is shift-out (WWSO), although this proves much less useful towards the end of the run. In contrast, however, after about generation 100, mutation (WKM), which started with a very low initial probability, becomes much more productive, as does crossover (WKCO). These changes probably reflect the far greater opportunity for local improvements by shift-out (and to a lesser extent, reroute) in the initial random individuals, whereas later in the run, crossover and mutation are needed to introduce diversity into the highly-evolved population, allowing further cost



(a) Progress over a single run



(b) Operator-probability adaptation

Figure 5.9: Typical results for GA18, $\alpha = \beta = 0.8$ and $\gamma = 0.5$ (Network 1, Run 1)

Network	W&O ($\times 10^6$)	W&D1 ($\times 10^6$)	W&D2 ($\times 10^6$)	GA18 ($\times 10^6$)
1	5.909	6.007	6.358	**5.579**
2	5.151	5.241	5.278	**4.886**
3	5.153	5.707	5.637	**4.976**
4	5.413	5.528	5.765	**5.117**
5	5.424	5.773	5.658	**5.108**
6	5.662	6.260	5.908	**5.344**
7	6.077	6.545	5.962	**5.651**

Table 5.17: Minimum cost allocation results for $\alpha = \beta = 1$ and $\gamma = 0.5$

reductions.

Supplementing the results provided in Figures 5.8 and 5.9, further minimum cost allocation results, showing typical progress and operator-probability adaptation over single runs of both GA19 and GA20 on Network 1 for both $(\alpha, \beta, \gamma) = (1, 1, 0.5)$ and $(0.8, 0.8, 0.5)$, are illustrated in Figures F.10–F.13.

In addition to the GAs, the three wavelength allocation heuristics by Wuttisittikulkij & O'Mahony (W&O) and Wauters & Demeester (W&D1 and W&D2) were also applied to Networks 1–7 for the same six combinations of (α, β, γ). As before, Wuttisittikulkij & O'Mahony's non-deterministic algorithm was allowed the best of five runs.

A summary of the results of applying all six GAs and heuristics is recorded in Tables 5.17–5.22, which lists the network cost, or the best cost for five runs, as appropriate. As before, the best overall result for each network, for each combination of (α, β, γ), is displayed in bold. In addition, the full set of results from applying the three GAs and W&0 to all seven of the networks is recorded in Tables F.43–F.48 and F.50–F.55.

Network	W&O ($\times 10^6$)	W&D1 ($\times 10^6$)	W&D2 ($\times 10^6$)	GA19 ($\times 10^6$)
1	5.704	5.172	5.896	**4.994**
2	5.081	4.589	5.046	**4.436**
3	5.051	4.625	5.446	**4.394**
4	5.313	4.808	5.388	**4.645**
5	5.302	4.806	5.326	**4.640**
6	5.516	5.010	5.282	**4.753**
7	5.907	5.380	5.606	**5.228**

Table 5.18: Minimum cost allocation results for $\alpha = \beta = 1$ and $\gamma = 1$

Network	W&O ($\times 10^6$)	W&D1 ($\times 10^6$)	W&D2 ($\times 10^6$)	GA20 ($\times 10^6$)
1	**6.036**	6.843	6.820	6.116
2	**5.213**	5.893	5.510	5.228
3	**5.229**	6.789	5.828	5.362
4	5.503	6.247	6.142	**5.499**
5	5.544	6.740	5.991	**5.523**
6	5.790	7.510	6.535	**5.781**
7	6.220	7.709	6.319	**6.001**

Table 5.19: Minimum cost allocation results for $\alpha = \beta = 1$ and $\gamma = 0$

Network	W&O ($\times 10^6$)	W&D1 ($\times 10^6$)	W&D2 ($\times 10^6$)	GA18 ($\times 10^6$)
1	4.653	4.890	5.103	**4.450**
2	4.089	4.238	4.212	**3.909**
3	4.138	4.656	4.559	**3.866**
4	4.312	4.474	4.608	**4.062**
5	4.313	4.684	4.518	**4.040**
6	4.246	4.842	4.482	**4.065**
7	4.203	4.674	4.178	**3.965**

Table 5.20: Minimum cost allocation results for $\alpha = \beta = 0.8$ and $\gamma = 0.5$

Network	W&O ($\times 10^6$)	W&D1 ($\times 10^6$)	W&D2 ($\times 10^6$)	GA19 ($\times 10^6$)
1	4.489	4.233	4.757	**4.014**
2	4.036	3.731	4.039	**3.531**
3	4.056	3.802	4.414	**3.541**
4	4.230	3.910	4.320	**3.716**
5	4.216	3.913	4.267	**3.716**
6	4.123	3.921	4.038	**3.633**
7	4.082	3.883	3.952	**3.686**

Table 5.21: Minimum cost allocation results for $\alpha = \beta = 0.8$ and $\gamma = 1$

Network	W&O ($\times 10^6$)	W&D1 ($\times 10^6$)	W&D2 ($\times 10^6$)	GA20 ($\times 10^6$)
1	**4.755**	5.546	5.448	4.895
2	**4.128**	4.745	4.384	4.249
3	**4.200**	5.510	4.704	4.271
4	**4.382**	5.039	4.896	4.444
5	4.409	5.455	4.769	**4.363**
6	**4.343**	5.762	4.926	4.344
7	4.306	5.464	4.404	**4.228**

Table 5.22: Minimum cost allocation results for $\alpha = \beta = 0.8$ and $\gamma = 0$

For $\gamma = 0.5$, a value regarded as being reasonably representative of actual costs, the best heuristic for both values of α,β was found to be W&O (except for Network 7), which is thus providing a good trade-off between capacity cost and link wavelength requirements. Nevertheless, for all seven networks the lowest network costs were actually obtained by the hybrid algorithm, GA18 (Tables 5.17 and 5.20).

With $\gamma = 1$, focusing the network cost on capacity rather than wavelength requirement, the best heuristic for both values of α,β was W&D1, and by a wide margin; this is perhaps to be expected, as W&D1 uses shortest paths (rather than k-lowest hop count or k-shortest paths, as do W&O and W&D2), thereby placing the emphasis on cost instead of wavelength requirement. However, once again, the hybrid algorithm, GA19, provided the best results for all seven networks (Tables 5.18 and 5.21).

Finally, setting $\gamma = 0$, so that the network cost depends entirely on wavelength requirement rather than capacity, the best heuristic for both values of α,β was W&O; this is no doubt due to the superiority of W&O, with respect to NWR, compared to the other two heuristics (see §5.2.1). Although for minimum NWR, GA17 matched the results of W&O, here, disappointingly, the GA20 hybrid algorithm was only able to obtain the best overall cost for four networks with $\alpha = \beta = 1$ and two networks with $\alpha = \beta = 0.8$, although its results were close to those of W&O on the other eight (Tables 5.19 and 5.22).

Overall, the experimental results show that, at least for the seven test networks, the GA/heuristic hybrid approach has again shown itself to be as promising as has been hoped, as in competition with three recent heuristics, it was able to obtain the lowest network costs for γ set to both 0.5 and 1; and in the case of $\gamma = 0$, was able to approach the results of the best of the allocation heuristics.

5.3 Operator-probability Adaptation

Complementing the earlier experimental work applying NOMaD's GA/heuristic hybrid approach to both mesh network topology design and routing, fibre & wavelength allocation, this section presents an investigation into NOMaD's operator-probability adaptation mechanism. Both the basic adaptation algorithm and its variants were described in §4.2.4, and some insight into the basic algorithm has already been provided from examining the operator-probability adaptation curves for the earlier experimental work.

In this study, it was decided to focus on the most difficult of the design problems examined so far, minimum cost allocation, which was nevertheless one in which the GA/heuristic hybrid approach performed well. To limit the amount of computation time required, only three of the test networks from §5.1.2 on p. 98 were used: Networks 1, 3 & 5. As before, each link was taken to be a cable of two fibres, and the capacity of the individual wavelength channels set to 10 Gbit/s.

For these experiments, all thirteen GAs employed the same four routing, fibre and wavelength allocation operators used in GA17–20, and described in §4.2.3 on p. 82. The metric for network assessment was Network Cost Model B, outlined in §4.1.3 on p. 72, with the default parameter values, including $(\alpha, \beta, \gamma) = (1, 1, 0.5)$. All the GAs were identical except for different settings of the adaptation algorithm parameters (see below). They had a population size of 500, a maximum of 100,000 trials, and used tournament selection (of size 4). Other non-default parameter values for GA21–25 and GA26–33 are listed in Tables F.56 and F.62 in Appendix F.

In applying NOMaD to specific design problems, initial operator probabilities have so far been selected by a few trial runs, aiming for settings that would remain approximately constant for the first few generations across all the networks examined. However, to avoid the potential bias of pre-tuned values, here the initial probabilities of all four operators plus the reproduction (*i.e.* null) operator, REPR, have been set to 0.2 in all the GAs.

5.3.1 Basic Adaptation Algorithm

The first set of experiments uses the basic adaptation algorithm, where operators are rewarded as a result of improvement over the fitness of the best individual found up to the end of the previous generation. These experimental results were first published by the author at the IEEE Intl. Conf. on Evolutionary Computation [33] (reprinted in Appendix A).

To explore the influence of the adaptation proportion, ψ, three values were selected, 0.15, 0.00 and 0.30, and combined with a credit constant, $\kappa = 0.5$, these are denoted GA21, GA22 and GA23 in Table 5.23. The levels in GA21 are those that have been used on the specific design problems; in GA22, the adaptation algorithm is completely deactivated, as with $\psi = 0.00$, the initial operator probabilities will remain unchanged for the entire run; GA23 represents a higher than usual rate of adaptation.

In examining the influence of the credit constant, the adaptation proportion was fixed to its usual value ($\psi = 0.15$), and κ set to 0.5, 0.0 and 1.0 in GA21, GA24 and GA25 respectively (Table 5.23). While the usual value of $\kappa = 0.5$ allows reduced credit to parent (0.5) and grandparent (0.25)

GA	ψ	κ
21	0.15	0.5
22	0.00	0.5
23	0.30	0.5
24	0.15	0.0
25	0.15	1.0

Table 5.23: Adaptation parameters for GAs using the basic adaptation algorithm

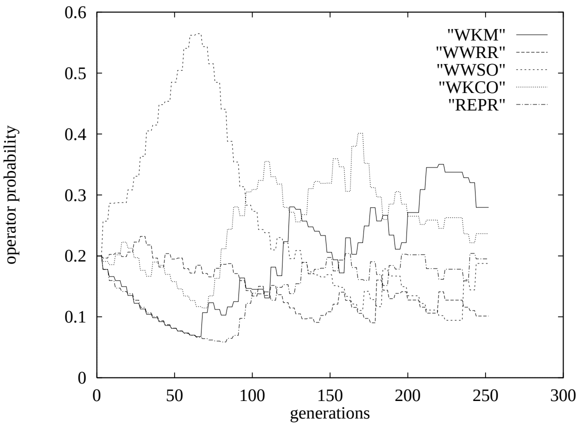
operators; in GA24, credit is only awarded to the actual operator responsible for an improvement; whereas in GA25, the operator, parent and grandparent operators are all rewarded equally.

All five GAs were then applied ten times with different random seeds to Networks 1, 3 & 5. The operator-probability adaptation curves for single typical runs of GA21, GA23, GA24 and GA25 are given in Fig. 5.10. As before, the graphs are for Network 1 with the same random seed, and hence have identical initial populations. However, the equivalent figure for GA22 is not shown, as all five curves remain unchanged at 0.2 throughout the run, since the adaptation mechanism is deactivated with $\psi = 0.00$. Corresponding graphs illustrating the progress of the runs for GA21–25, showing both the average network cost of the population and the lowest (*i.e.* best in population) cost, generation by generation, are given in Fig. F.14.

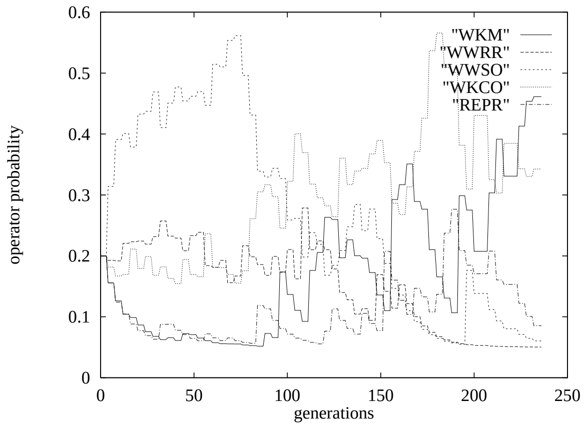
Inspecting GA21 in Fig. 5.10(a), the effect of the adaptation algorithm can be clearly seen. For example, the early productivity of the heuristic path shift-out operator, WWSO, results in its probability rising from 0.2 to nearly 0.6 in the first 50 generations, although it subsequently reduces to below 0.2. In contrast, the far simpler path mutation, WKM, falls from 0.2 to below 0.1 before generation 50, only then to rise to over 0.3 due to higher productivity in the latter part of the run.

Comparing GA21 and GA23, the impact of the higher adaptation proportion (ψ) in the latter is obvious in the greater variability of the probability curves (Fig. 5.10(b)). A similar effect is also apparent in Fig. 5.10(c), where setting the credit constant (κ) to 0.0 in GA24 resulted in wide fluctuations of operator probability, due to the hybrid's exclusive focus on operators producing direct improvements in fitness. In contrast, for GA25 (Fig. 5.10(d)), although there is some initial variation of probabilities, particularly for the productive path shift-out operator (WWSO), after 120 generations all the operators settle to an approximately equal share of probability. This would seem to indicate that with grandparent and parent operators receiving equal credit ($\kappa = 1.0$), the basic algorithm is unable to respond to all but the widest differences in relative productivity.

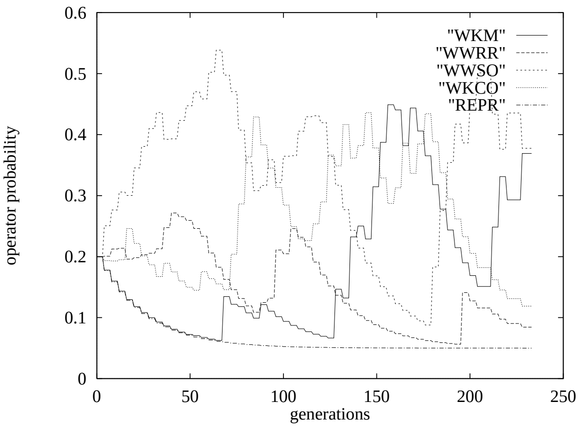
A summary of the results from applying the five GAs is given in Tables 5.24 and 5.25 (with GA21 included in both tables for ease of reference). These record the best and median results at both the end (100,000 trials) and the early part (25,000 trials) of ten runs on all three networks. In



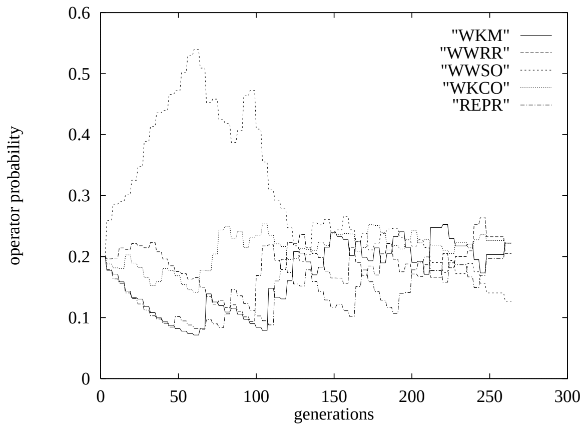
(a) GA21, $(\psi, \kappa) = (0.15, 0.5)$



(b) GA23, $(\psi, \kappa) = (0.30, 0.5)$



(c) GA24, $(\psi, \kappa) = (0.15, 0.0)$



(d) GA25, $(\psi, \kappa) = (0.15, 1.0)$

Figure 5.10: Typical operator-probability adaptation, using the basic adaptation algorithm, for GA21, GA23–GA25 (Network 1, Run 1)

Network	GA22 $\psi = 0.00$ Best ($\times 10^6$)	GA22 Median ($\times 10^6$)	GA21 $\psi = 0.15$ Best ($\times 10^6$)	GA21 Median ($\times 10^6$)	GA23 $\psi = 0.30$ Best ($\times 10^6$)	GA23 Median ($\times 10^6$)	Significance Level (%) $s_{22,21}$	$s_{23,21}$	$s_{22,23}$
				25,000 trials					
1	6.527	6.680	**6.222**	**6.517**	6.306	6.558	1.1	—	5.5
3	5.642	5.722	**5.468**	**5.604**	5.535	5.739	0.1	5.5	—
5	5.922	6.070	5.869	**5.967**	**5.784**	6.038	1.1	—	5.5
				100,000 trials					
1	5.621	5.681	**5.582**	5.659	5.609	**5.658**	—	—	—
3	4.861	**4.917**	4.879	4.945	**4.833**	4.924	—	—	—
5	**5.063**	5.171	5.081	5.144	5.090	**5.140**	—	—	—

Table 5.24: Basic adaptation algorithm results with adaptation proportion $\psi = 0.00$, 0.15 and 0.30

Network	GA24 $\kappa = 0.0$ Best ($\times 10^6$)	GA24 Median ($\times 10^6$)	GA21 $\kappa = 0.5$ Best ($\times 10^6$)	GA21 Median ($\times 10^6$)	GA25 $\kappa = 1.0$ Best ($\times 10^6$)	GA25 Median ($\times 10^6$)	Significance Level (%) $s_{24,21}$	$s_{25,21}$	$s_{24,25}$
				25,000 trials					
1	6.331	6.520	**6.222**	6.517	**6.222**	**6.476**	—	—	—
3	**5.460**	5.610	5.468	**5.604**	5.545	5.631	—	—	—
5	5.875	6.026	5.869	**5.967**	**5.819**	5.998	—	—	—
				100,000 trials					
1	5.613	5.659	5.582	5.659	**5.445**	**5.635**	—	—	5.5
3	4.887	4.932	4.879	4.945	**4.875**	**4.928**	—	—	—
5	**5.079**	5.158	5.081	**5.144**	**5.079**	5.145	—	—	—

Table 5.25: Basic adaptation algorithm results with credit constant $\kappa = 0.0$, 0.5 and 1.0

each table, both the best overall result and best median for a given network and number of trials are displayed in bold. Also, as the same random seeds were used for each set of runs on a network, a non-parametric sign test could be applied to matched pairs [307] from different GAs to establish if there were significant differences in their medians. The resulting significance levels are included in the tables, with *e.g.* $s_{22,21}$ representing the significance level of the comparison of GA22 with GA21. In addition, the full set of results from applying the five GAs to all three networks is recorded in Tables F.57–F.61.

Surprisingly, Table 5.24 at 100,000 trials shows that despite the wide differences in dynamic behaviour observed on individual runs, and discussed above, the final quality of results is unaffected by the choice of adaptation proportion (ψ). However, this can perhaps best be explained by also considering the relative performance in the early part of the same runs. The best results at this stage were achieved, as expected, by GA21 ($\psi = 0.15$), both in comparison with GA22 ($\psi = 0.00$; two highly and one very highly significant median differences) and GA23 ($\psi = 0.30$; a significant median difference). It would appear, therefore, that although at 25,000 trials, GA21 is out-performing the other two, by 100,000 trials the productivity of the combination of problem-specific operators used is such as to allow both GA22 (fixed probabilities) and GA23 (high adaptation rate) to obtain final results of equal quality to GA21.

Different values of credit constant (κ) also produced much less dramatic variations in solution quality than expected. Table 5.25 records that at the end of the runs, the best results were obtained by GA25 ($\kappa = 1.0$). However, there was little significant difference in medians, with only some indication that GA25 ($\kappa = 1.0$) was better than GA24 ($\kappa = 0.0$). Evidence from the early part of the runs is inconclusive, indicating little effect by the credit constant on the initial quality of results. Taking account of the final results though, there is nevertheless some indication both that crediting parent and grandparent operators is to be preferred over simply rewarding the operator responsible for an improvement, and also that a higher credit constant than the usual $\kappa = 0.5$ may be beneficial.

Overall, these experiments on the basic adaptation algorithm indicate that, at least for this application, with powerful problem-specific operators, the main benefits of operator-probability adaptation are in relieving the experimenter of the burden of setting initial probabilities and in the early performance of the GA/heuristic hybrid, rather than in improvements of the final solution quality obtained.

5.3.2 Variants on the Basic Algorithm

Complementing the basic adaptation algorithm, §4.2.4 describes three variants that assign credit to operators for improvements over: best-in-window fitness; the median fitness of the population at the end of the previous generation; or the fitness of the parent individual(s). This subsection explores these variants by repeating the experiments previously applied to the basic algorithm.

However, it would be unproductive, for this application, to use best-of-window fitness, as the (near) monotonic improvement in population fitness results means there is little difference between the fitness of the best individual found up to the end of the previous generation, as used by the basic algorithm, and best-in-window fitness. Consequently, only parental fitness and median fitness credit assignment will be considered.

In addition, there is no need to repeat the results for GA22, as with $\psi = 0.00$ the adaptation algorithm is completely deactivated, including credit assignment of whatever type. As a consequence, the GA22 results can be reused here.

The adaptation parameters (ψ, κ) used by the eight GAs are listed in Table 5.26. GA26-29 use parental credit assignment, and GA30–33 median. By comparing this with Table 5.23 on p. 119,

GA	ψ	κ
26/30	0.15	0.5
27/31	0.30	0.5
28/32	0.15	0.0
29/33	0.15	1.0

Table 5.26: Adaptation parameters for GAs using variant adaptation algorithms

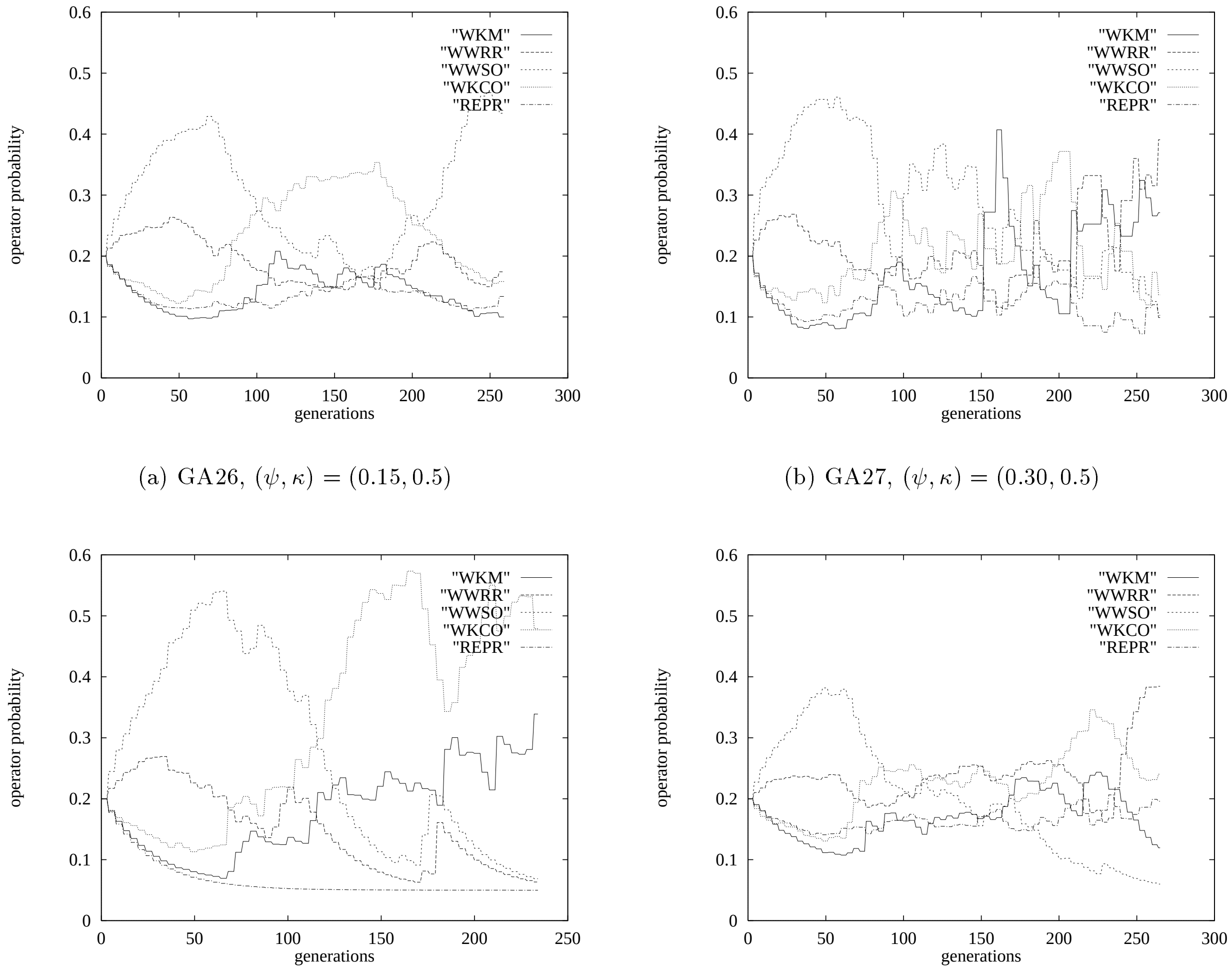
and recalling that there is no need to repeat the runs for GA22, the correspondence with GAs using the basic adaptation algorithm should be clear.

The four GAs using parental credit assignment were each applied ten times with different random seeds to Networks 1, 3 & 5. The operator-probability adaptation curves for single typical runs of GA26-29 are given in Fig. 5.11. In addition, corresponding graphs illustrating the progress of the runs for the four GAs are provided in Fig. F.15.

For GA26 in Fig. 5.11(a), the path shift-out operator (WWSO) is the most productive in the first 100 generations, then crossover (WKCO) for the next 100 generations, before a strong resurgence by shift-out at the end of the run. Compare this with GA21 in Fig. 5.10(a), whose adaptation parameters are identical to GA26, except that the former uses best-so-far, rather than parental, credit assignment. In that case, while first shift-out and then crossover were the most productive operators, it was mutation (WKM) that had highest probability at the end of the run. It would appear that, while shift-out struggles to make further improvements in best-so-far fitness in GA21, it is still able to improve the fitness of children *wrt.* their parents in GA26, even in the less genetically diverse conditions at the end of the run.

Although employing parental credit assignment here, altering the adaptation parameters showed similar effects on the operator-probability curves for the other three GAs as were previously observed for the basic adaptation algorithm. In Figures 5.11(b) and (c), both GA27 and GA28 show increased variability compared with GA26, doubtless for similar reasons as suggested above. However, this is particularly marked in Fig. 5.11(c) only for those operators, shift-out and crossover, that appear to perform well under parental credit assignment. This is to be expected, as with $\kappa = 0.0$, it is only the operator directly responsible for an improvement that is rewarded. For GA29 in Fig. 5.11(d), the same tendency to settle on equal operator probabilities of 0.2 can be seen as was noted for the equivalent GA using best-so-far credit assignment (GA25), although this is less marked towards the end of the run.

A summary of the results from applying the four GAs (plus GA22) is given in Tables 5.27 and 5.28 (with GA26 included in both tables for ease of reference). These record the best and



(a) GA26, $(\psi, \kappa) = (0.15, 0.5)$

(b) GA27, $(\psi, \kappa) = (0.30, 0.5)$

(c) GA28, $(\psi, \kappa) = (0.15, 0.0)$

(d) GA29, $(\psi, \kappa) = (0.15, 1.0)$

Figure 5.11: Typical operator-probability adaptation, using parental credit assignment, for GA26–GA29 (Network 1, Run 1)

Network	GA22 $\psi = 0.00$ Best $(\times 10^6)$	GA22 $\psi = 0.00$ Median $(\times 10^6)$	GA26 $\psi = 0.15$ Best $(\times 10^6)$	GA26 $\psi = 0.15$ Median $(\times 10^6)$	GA27 $\psi = 0.30$ Best $(\times 10^6)$	GA27 $\psi = 0.30$ Median $(\times 10^6)$	Significance Level (%) $s_{22,26}$	$s_{27,26}$	$s_{22,27}$
25,000 trials									
1	6.527	6.680	**6.371**	6.536	6.402	**6.474**	5.5	—	0.1
3	5.642	5.722	5.523	**5.602**	**5.470**	5.614	0.1	—	0.1
5	5.922	6.070	**5.707**	5.986	5.751	**5.948**	1.1	—	1.1
100,000 trials									
1	5.621	5.681	5.584	5.666	**5.576**	**5.646**	—	—	1.1
3	4.861	4.917	4.874	**4.907**	**4.840**	4.932	—	—	—
5	**5.063**	5.171	5.095	**5.150**	5.138	5.166	—	—	—

Table 5.27: Parental credit assignment results with adaptation proportion $\psi = 0.00$, 0.15 and 0.30

Network	GA28 $\kappa = 0.0$ Best $(\times 10^6)$	GA28 $\kappa = 0.0$ Median $(\times 10^6)$	GA26 $\kappa = 0.5$ Best $(\times 10^6)$	GA26 $\kappa = 0.5$ Median $(\times 10^6)$	GA29 $\kappa = 1.0$ Best $(\times 10^6)$	GA29 $\kappa = 1.0$ Median $(\times 10^6)$	Significance Level (%) $s_{28,26}$	$s_{29,26}$	$s_{28,29}$
25,000 trials									
1	6.332	6.491	6.371	6.536	**6.219**	**6.487**	—	—	—
3	5.536	5.621	5.523	5.602	**5.490**	**5.577**	—	—	5.5
5	5.760	**5.969**	**5.707**	5.986	5.765	6.013	—	—	—
100,000 trials									
1	**5.563**	**5.622**	5.584	5.666	5.568	5.634	—	—	—
3	4.884	4.941	**4.874**	**4.907**	4.893	4.968	—	5.5	—
5	5.091	**5.134**	5.095	5.150	**5.042**	5.164	—	—	—

Table 5.28: Parental credit assignment results with credit constant $\kappa = 0.0$, 0.5 and 1.0

median results at both the end (100,000 trials) and the early part (25,000 trials) of ten runs on all three networks. As before, in each table, both the best overall result and best median for a given network and number of trials are displayed in bold. The significance levels from a non-parametric sign test on median differences [307] is also included. The full set of results from the runs is recorded in Tables F.63–F.66.

Once again, after 100,000 trials (Table 5.27) there is little to choose between the final results, despite the different levels of adaptation proportion (ψ). However, at 25,000 trials, while both GA26 and GA27 produced very similar results, they were both quite significantly better than GA22, which with $\psi = 0.00$ had the adaptation algorithm deactivated.

Changing the value of the credit constant (κ) appeared to have little effect on either the early part of the run or the final results (Table 5.28). Only two significant median values were found, and the overall pattern of results did not seem to favour any particular κ setting.

Turning now to median credit assignment, the next four GAs (GA30–33) were each applied ten times with different random seeds to Networks 1, 3 & 5. The operator-probability adaptation curves for single typical runs of these are shown in Fig. 5.12. The corresponding graphs illustrating

Network	GA22 $\psi = 0.00$ Best ($\times 10^6$)	GA22 Median ($\times 10^6$)	GA30 $\psi = 0.15$ Best ($\times 10^6$)	GA30 Median ($\times 10^6$)	GA31 $\psi = 0.30$ Best ($\times 10^6$)	GA31 Median ($\times 10^6$)	Significance Level (%) $s_{22,30}$	$s_{30,31}$	$s_{22,31}$
25,000 trials									
1	6.527	6.680	6.396	6.442	**6.249**	**6.425**	0.1	5.5	0.1
3	5.642	5.722	**5.395**	5.590	5.420	**5.569**	0.1	—	1.1
5	5.922	6.070	5.871	5.946	**5.733**	**5.904**	5.5	—	1.1
100,000 trials									
1	5.621	**5.681**	**5.595**	5.706	5.637	5.712	—	—	—
3	**4.861**	**4.917**	4.919	4.971	4.890	4.925	(1.1)	—	—
5	**5.063**	5.171	5.113	**5.161**	5.107	5.179	—	—	—

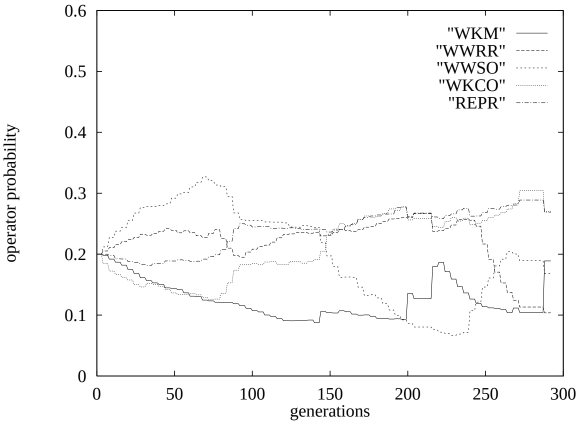
Table 5.29: Median credit assignment results with adaptation proportion $\psi = 0.00$, 0.15 and 0.30

the progress of the runs are given in Fig. F.16.

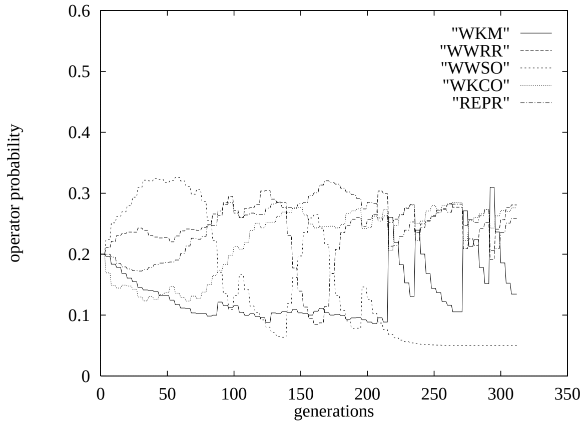
All four GAs seem to produce approximately the same overall pattern of probability curves. After some initial divergence from the common starting point of 0.2, the curves occupy the narrow band between 0.1 and 0.3. The shift-out operator (WKSO) perfoms well initially in all four GAs, but seems less productive later on, particularly for GA31 with its high adaptation proportion ($\psi = 0.30$). In contrast, mutation (WKM) appears to perfom poorly throughout in all four GAs, although this is less marked again in GA31. Overall, it may be that rewarding operators for improvements over median fitness is a relatively poor indicator of their productivity, as the result of a GA run is judged on the best individual found rather than the population median achieved. Consequently, basing credit assignment on median fitness is likely to give little guidance to the adaptation algorithm, resulting in the approximately equal apportionment of probability observed over the different adaptation parameter settings.

A summary of the results from applying the four GAs (plus GA22) is given in Tables 5.29 and 5.30 (with GA30 included in both tables). These record the best and median results at both the end (100,000 trials) and the early part (25,000 trials) of ten runs on all three networks. As usual, in each table, both the best overall result and best median for a given network and number of trials are displayed in bold. The significance levels from a non-parametric sign test on median differences [307] is also included. The full set of results from the runs is recorded in Tables F.67–F.70.

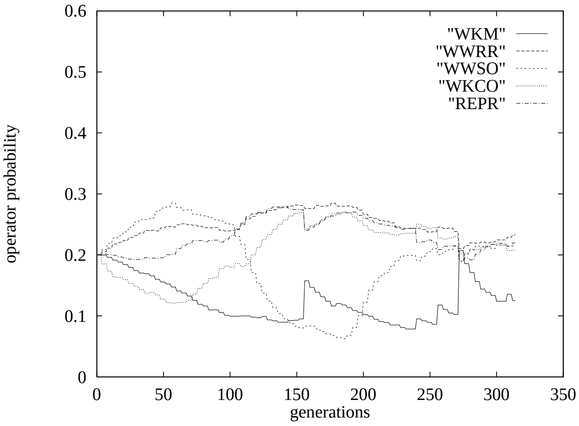
Considering first variations in the adaptation proportion (ψ) summarised in Table 5.29, the results at 100,000 trials surprisingly show GA22, with the adaptation algorithm deactivated, actually performing better than the other two GAs, although this is not a significant median difference except for one network. At 25,000 trials, however, it was GA31, with a high adaptation proportion ($\psi = 0.30$), which obtained the best results compared to GA22 (adaptation deactivated), although



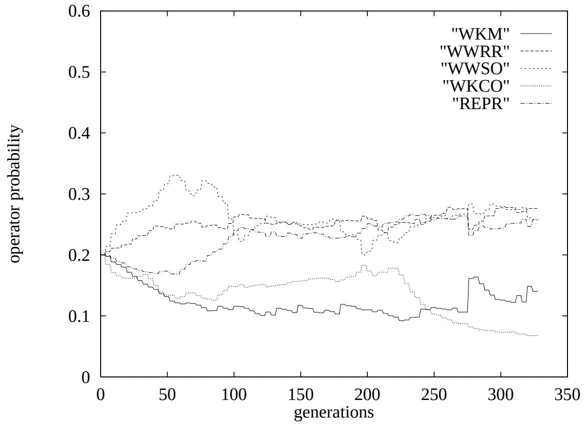
(a) GA30, $(\psi, \kappa) = (0.15, 0.5)$



(b) GA31, $(\psi, \kappa) = (0.30, 0.5)$



(c) GA32, $(\psi, \kappa) = (0.15, 0.0)$



(d) GA33, $(\psi, \kappa) = (0.15, 1.0)$

Figure 5.12: Typical operator-probability adaptation, using median credit assignment, for GA30–GA33 (Network 1, Run 1)

Network	GA32 $\kappa = 0.0$		GA30 $\kappa = 0.5$		GA33 $\kappa = 1.0$		Significance Level (%)		
	Best ($\times 10^6$)	Median ($\times 10^6$)	Best ($\times 10^6$)	Median ($\times 10^6$)	Best ($\times 10^6$)	Median ($\times 10^6$)	$s_{30,32}$	$s_{33,30}$	$s_{32,33}$
25,000 trials									
1	**6.305**	6.446	6.396	**6.442**	6.401	6.517	—	5.5	—
3	5.450	5.607	**5.395**	5.590	5.475	**5.575**	—	—	—
5	**5.791**	**5.933**	5.871	5.946	5.827	5.943	—	—	—
100,000 trials									
1	5.623	5.686	5.595	5.706	**5.547**	**5.669**	5.5	—	—
3	4.890	**4.956**	4.919	4.971	**4.840**	4.962	—	—	—
5	5.099	**5.134**	5.113	5.161	**5.075**	5.175	5.5	—	—

Table 5.30: Median credit assignment results with credit constant $\kappa = 0.0$, 0.5 and 1.0

GA30 was also better than GA22 with good significance levels. As usual, variations in credit constant (κ) had little effect, either in the early part or at the end of the run (Table 5.30).

Overall, it was not felt appropriate to judge which of the three variants of the adaptation algorithm examined was 'best' for minimum cost allocation, primarily because the initial operator probabilities were all artificially set to 0.2 in this study, rather than being tuned to improve algorithm performance. However, comparing GA21, GA26 and GA30, which all used the default adaptation parameter settings of $(\psi, \kappa) = (0.15, 0.5)$, there would appear to be little to choose between best-so-far and parental credit assignment, but that both would appear to be better than employing median fitness. In addition, the conclusions from the study of the basic adaptation algorithm also seem to apply here, namely that the main benefits of operator-probability adaptation are in relieving the experimenter of the burden of setting initial probabilities and in the early performance of the GA/ heuristic hybrid, rather than in improvements of the final solution quality obtained.

5.4 Summary

This chapter has provided a number of experiments exploring the suitability of the author's GA/ heuristic hybrid approach for multi-wavelength all-optical transport network design. Three studies exploring different approaches to mesh network topology design were reported: initial experiments established the superiority of a problem-specific form of crossover; work on local search operators showed the limited merits of incorporating a flexible local search operator combined with trial-based probability reassignment; finally, the power of iterative heuristic operators was demonstrated, although not without cost in computational terms. In addition, the initial topology approach was used to design the test networks that were to be used in the next group of experiments.

GA/heuristic hybrids for routing, fibre choice and wavelength allocation were then investigated. Two metrics for network assessment were used, one based on NWR, and the other on a simplified model of network cost. In both cases, the GAs were compared with recent wavelength-allocation heuristics published in the literature. For minimum NWR, the GA/heuristic hybrid approach was shown to be as promising as had been hoped, as the blend of a GA with steps from one of the heuristics resulted in both improved performance (over the GA alone) and improved results (over the heuristic alone). In addition, the GA results matched those of the best wavelength-allocation heuristic examined. Similar success was achieved with the minimum cost allocation, as in competition with three recent heuristics, the GA/heuristic hybrid approach was able to obtain the lowest network costs for two settings of the cost parameters (including the one judged most realistic), and in a third case, was able to approach the results of the best of the allocation heuristics.

The final investigation was of NOMaD's operator-probability adaptation mechanism, both in its basic form, where credit is assigned for improvements over the best-so-far individual, as well as two variants that reward improvements over parental and median fitness. At least in the application examined, minimum cost allocation which uses powerful problem-specific operators, the main benefits of operator-probability adaptation were found to be in relieving the experimenter of the burden of setting initial probabilities and in the early performance of the GA/heuristic hybrid, rather than in improvements of the final solution quality obtained.

In the next chapter, there will be more detailed reflection on the GA/heuristic hybrid approach, including the design of NOMaD, further discussion of the experimental results obtained, and consideration given to directions for further work.