Sporadic Model Building for Efficiency Enhancement of hBOA

Martin Pelikan
Kumara Sastry
David E. Goldberg

IlliGAL Report No. 2005026
December 2005

Illinois Genetic Algorithms Laboratory
University of Illinois at Urbana-Champaign
117 Transportation Building
104 S. Mathews Avenue Urbana, IL 61801
Office: (217) 333-2346
Fax: (217) 244-5705

Sporadic Model Building for Efficiency Enhancement of hBOA

Martin Pelikan
Dept. of Mathematics and Computer Science, 320 CCB
University of Missouri at St. Louis
One University Blvd.
St. Louis, MO 63121
pelikan@cs.umsl.edu

Kumara Sastry
Illinois Genetic Algorithms Laboratory, 117 TB
University of Illinois at Urbana-Champaign
104 S. Mathews Ave.
Urbana, IL 61801
kumara@illigal.ge.uiuc.edu

David E. Goldberg
Illinois Genetic Algorithms Laboratory, 117 TB
University of Illinois at Urbana-Champaign
104 S. Mathews Ave.
Urbana, IL 61801
deg@illigal.ge.uiuc.edu

December 4, 2005

Abstract

This paper describes and analyzes *sporadic model building*, which can be used to enhance the efficiency of the hierarchical Bayesian optimization algorithm (hBOA) and other estimation of distribution algorithms (EDAs). With sporadic model building, the structure of the probabilistic model is updated once every few iterations (generations), whereas in the remaining iterations only model parameters (conditional and marginal probabilities) are updated. Since the time complexity of updating model parameters is much lower than the time complexity of learning the model structure, sporadic model building decreases the overall time complexity of model building. The paper shows that for boundedly difficult nearly decomposable and hierarchical optimization problems, sporadic model building leads to a significant *model-building speedup* that decreases the asymptotic time complexity of model building in hBOA by a factor of $\Theta(n^{0.26})$ to $\Theta(n^{0.65})$, where n is the problem size. On the other hand, sporadic model building also increases the number of evaluations until convergence; nonetheless, for decomposable problems, the *evaluation slowdown* is insignificant compared to the gains in the asymptotic complexity of model building.

Keywords
Bayesian optimization algorithm, hierarchical BOA, estimation of distribution algorithms, efficiency enhancement, sporadic model building.

1 Introduction

The hierarchical Bayesian optimization algorithm (hBOA) (Pelikan & Goldberg, 2001; Pelikan & Goldberg, 2003; Pelikan, 2005) replaces standard variation operators of genetic and evolutionary algorithms (Holland, 1975; Goldberg, 1989; Rechenberg, 1973; Koza, 1992) by building a Bayesian network for selected solutions and sampling the built network to generate new candidate solutions. Additionally, hBOA uses restricted tournament replacement (Harik, 1995) to effectively maintain diversity and preserve alternative partial solutions. It was theoretically and empirically shown that hBOA can solve nearly decomposable and hierarchical optimization problems in a quadratic number of evaluations or faster (Pelikan, 2005; Pelikan, Sastry, & Goldberg, 2002). However, quadratic or subquadratic performance may be still insufficient for problems with thousands of decision variables or high-order interactions. Consequently, efficiency enhancement techniques (Goldberg, 2002; Sastry, 2001; Sastry, Pelikan, & Goldberg, 2004; Pelikan, 2005) may have to be incorporated into hBOA to make this algorithm practical even for extremely large and complex problems.

A number of efficiency enhancement techniques can be incorporated into hBOA and other genetic and evolutionary algorithms (Goldberg, 2002; Sastry et al., 2004; Albert, 2001; Sinha & Goldberg, 2001; Srivastava & Goldberg, 2001; Pelikan, 2005). This paper discusses *sporadic model building*, which can significantly speed up model building in hBOA and other similar estimation of distribution algorithms (EDAs). Specifically, with sporadic model building, hBOA updates the structure of the Bayesian network used to sample new solutions once every few iterations (generations); in the remaining iterations, the network structure from the previous iteration is used and only the parameters of the network (conditional and marginal probabilities) are updated based on the selected solutions. Since learning the structure is the most expensive component of model building, sporadic model building should lead to a significant speedup of model building.

This paper describes and analyzes sporadic model building in hBOA. The results indicate that sporadic model building leads to a significant speedup of model building that decreases the asymptotic complexity of model building in hBOA.

The paper is organized as follows. Section 2 describes hBOA and discusses its computational complexity. Section 3 presents sporadic model building and discusses its effects on the time complexity of hBOA. Section 4 presents a dimensional model that provides a bound on the structure-building period and discusses the expected optimal speedup. Section 5 presents experimental results. Finally, Section 6 summarizes and concludes the paper.

2 Hierarchical Bayesian Optimization Algorithm (hBOA)

Estimation of distribution algorithms (EDAs) (Mühlenbein & Paaß, 1996; Pelikan, Goldberg, & Lobo, 2002; Larrañaga & Lozano, 2002) evolve a population of candidate solutions to the given problem by building and sampling a probabilistic model of promising solutions. The hierarchical Bayesian optimization algorithm (hBOA) (Pelikan & Goldberg, 2001; Pelikan & Goldberg, 2003; Pelikan, 2005) is an EDA that uses Bayesian networks (BNs) to represent the probabilistic model and incorporates restricted tournament replacement (RTR) for effective diversity maintenance. This section describes hBOA and discusses time complexity of the methods for learning and sampling BNs used to guide exploration of the search space in hBOA.

2.1 Basic hBOA Procedure

hBOA evolves a population of candidate solutions represented by fixed-length strings over a finite alphabet (for example, binary strings). The initial population is generated at random according to a uniform distribution over the set of all potential solutions. Each iteration (generation) starts by selecting promising solutions from the current population using any standard selection method of genetic and evolutionary algorithms. For example, k-ary tournament selection can be used, which selects one solution at a time by first choosing a random subset of k candidate solutions from the current population and then selecting the best solution out of this subset; random tournaments are repeated until there are sufficiently many solutions in the selected population.

After selecting the promising solutions, hBOA builds a BN with local structures (Howard & Matheson, 1981; Pearl, 1988; Chickering, Heckerman, & Meek, 1997; Friedman & Goldszmidt, 1999) as a model for these solutions. More specifically, BNs with decision trees or graphs are used as probabilistic models. New solutions are generated by sampling the built network.

The new solutions are then incorporated into the original population using restricted tournament replacement (RTR) (Harik, 1995), which ensures effective diversity maintenance. RTR with window size $w > 1$ incorporates each new candidate solution X into the original population using the following three steps:

1. Randomly select a subset W of w candidate solutions from the original population.
2. Let Y be a solution from W that is most similar to X (based on genotypic distance).
3. Replace Y with X if X is better; otherwise, discard X.

A robust rule of thumb is to set $w = \min\{n, N/20\}$, where n is the number of decision variables in the problem and N is the population size (Pelikan, 2005).

The next iteration is executed unless some predefined termination criteria are met. For example, the run can be terminated when the maximum number of generations is reached or the entire population consists of copies of the same candidate solution. For more details about the basic hBOA procedure, see Pelikan and Goldberg (2001) or Pelikan (2005).

It is beyond the scope of this paper to provide details of methods for learning and sampling BNs with local structures; for more information about learning and sampling BNs with local structures, please see Chickering et al. (1997), Heckerman et al. (1994), and Pelikan (2005). Nonetheless, before discussing sporadic model building, it is important to understand the time complexity of the different components of hBOA variation and that is why the remainder of this section discusses the time complexity of learning and sampling BNs with decision trees.

2.2 Time Complexity of Learning and Sampling Bayesian Networks

Learning a BN with local structures consists of two steps (Heckerman et al., 1994):

(1) learn the structure and

(2) learn the parameters (conditional probabilities).

Assuming that the problem is decomposable into subproblems of order k, the time complexity of learning the structure of a BN with decision trees can be bounded by $O(kn^2N)$ (Pelikan, 2005). Under the same assumptions, the conditional probabilities can be computed in $O(knN)$ time steps.

Thus, the asymptotic computational complexity of learning the parameters is much lower than the asymptotic complexity of learning the network structure.

Assuming an order-k decomposable problem, sampling one candidate solution from a given BN can be done in $O(kn)$ steps. Consequently, the asymptotic complexity of sampling a population of N candidate solutions can be bounded by $O(knN)$.

Therefore, the asymptotic time complexity of building the network structure dominates the overall complexity of the variation operator that consists of building and sampling a BN. Similar behavior can be observed in other EDAs that use complex probabilistic models, for example, in the extended compact genetic algorithm (ECGA) (Harik, 1999; Sastry & Goldberg, 2000). The following section describes sporadic model building, which can be used to speedup BOA and other similar EDAs.

3 Sporadic Model Building (SMB)

The above section indicated that building the network structure is the most expensive part of hBOA variation. Consequently, for large problems, building the network structure may become a bottleneck. To speed up the structure-building procedure, one can exploit the fact that although the model structure does not remain the same throughout the run, the structures in consequent iterations of hBOA are usually very similar. *Sporadic model building* (SMB) exploits this behavior by building the structure once in every few iterations, while in the remaining iterations only the parameters are updated for the structure used in the previous iteration.

The remainder of this section discusses a simple schedule that can be used to control SMB; next, it discusses the effects of SMB on time complexity of hBOA.

3.1 Simple Periodic Schedule to Control SMB

One of the most important factors that influences the effectiveness of SMB is the *schedule* that determines when to build the structure and when to only update the parameters for the previous structure. This paper investigates a simple approach that uses a parameter t_{sb} called the *structure building period*, which denotes the period with which the structure is updated. For example, $t_{sb} = 1$ denotes the scenario in which the structure is built in every iteration, whereas $t_{sb} = 2$ denotes the scenario in which the structure is built only every other iteration. Of course, other types of schedules can be used; nonetheless, this paper indicates that even a simple schedule based on a fixed period yields asymptotic speedups of model building in hBOA and similar results can be expected for other similar EDAs.

3.2 Effects of SMB on BOA Time Complexity

There are two potential effects of using SMB in BOA: (1) speedup of model building and (2) slowdown of evaluation:

1. *Speedup of model building.* Since with SMB, the most expensive part of the model building procedure is run only each t_{sb}th iteration, increasing t_{sb} should lead to a speedup of model building. Ideally, the speedup of building the model structure would be linearly proportional to t_{sb}; nonetheless, SMB may lead to an increase of the population size and the number of iterations and, as a result, with SMB each model-building step may become more computationally expensive than without SMB. Furthermore, the actual speedup for a fixed problem size must be upper bounded by the expected number of iterations until convergence (the structure must be learned at least once).

Since the number of iterations is expected to grow at most as $O(n)$ (Mühlenbein & Schlierkamp-Voosen, 1993; Thierens, Goldberg, & Pereira, 1998), the speedup of model building is expected to be upper-bounded by $O(n)$. Another way to estimate an upper bound of the expected speedup is to argue that the asymptotic complexities of building the structure and learning the parameters differ by a factor of n and no matter whether the structure is built or not, the parameters must be updated in each iteration.

2. *Slowdown of evaluation.* SMB may lead to a decreased accuracy of the probabilistic models used to sample new candidate solutions. As a result, the population size N for building a sufficiently accurate model may increase with t_{sb}. Additionally, because of a less frequent adaptation of the model structure to the current population of promising solutions, SMB may slow down the convergence and the total number G of iterations may be expected to increase with t_{sb}.

   Consequently, an increase in the overall time spent in the evaluation of candidate solutions can be expected as a result of the increase of N and G, because the time spent in evaluation is linearly proportional to the number E of evaluations, where $E = N \times G$.

It can be expected that for each particular problem, there exists an optimal value of t_{sb}, which leads to the maximum overall speedup of hBOA on this problem. Nonetheless, to make our results independent of the time complexity of the evaluation function, which changes from problem to problem, we studied the above two effects in separation.

The only parameter required for using sporadic model building is the structure-building period t_{sb}, which determines the interval with which the structure is rebuilt. Since the speedup is always upper bounded by the value of t_{sb}, it is important to understand what are adequate choices of t_{sb}. The following section provides a simple theoretical model for estimating the growth of t_{sb}. Next, empirical results are presented and discussed.

4 Dimensional Model for the Structure-Building Period

In deriving the model for the structure building period t_{sb}, we assume that hBOA correctly models most of the subproblems in an appropriate problem decomposition and only a few subproblems are modeled incorrectly. We further assume that building an accurate model is essential for successfully converging to the global optima; in other words, the decision variables belonging to the incorrectly modeled subproblems are expected to converge to incorrect values. Therefore, we have to build the probabilistic model before the sub-optimal partial solutions take over the entire population. We also assume that the problem is an additively separable problem with m subproblems of order k. We further assume that there is no niching and, therefore, the model is directly applicable to BOA but for hBOA the bound provided by the model may be more restrictive than necessary.

Specifically, given an additively separable problem with m subproblems of order k, we assume that the model builder accurately captures $(m-1)$ subproblems, and fails to model one subproblem. We further assume that the variables belonging to the incorrectly modeled subproblem are modeled as independent of each other as shown in Figure 1. Since the lower-order models are misleading, inferior partial solutions (building blocks, BBs) get increased market share every time the model is sampled. For example, if the best BB is 1111, and the deceptive attractor is 0000, the average fitness of schema ***0 is higher than that of schema ***1 (Goldberg, 1987; Deb & Goldberg, 1994). Therefore, ***0 gets increased market share over ***1 even though 1111 is the best building block. We model the growth of inferior partial solutions using the gambler's ruin model used elsewhere for analyzing population sizing of selectorecombinative GAs (Harik, Cantú-Paz, Goldberg, & Miller,

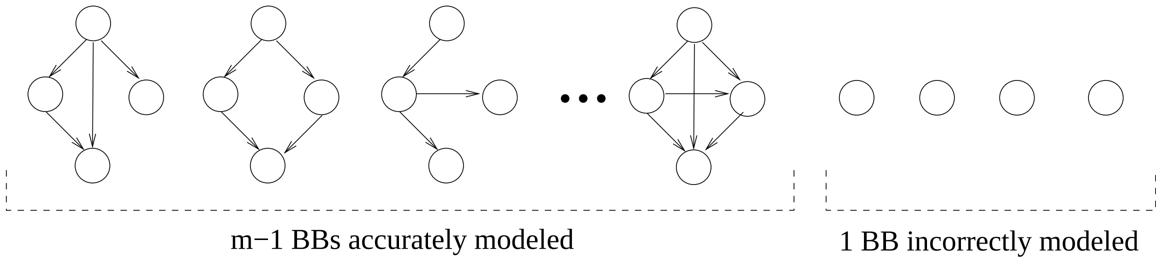


Figure 1: In deriving an upper bound on the structure building period, we assume that BOA accurately discovers $m-1$ out of m building blocks accurately and fails to discover one building block, variables of which are modeled as being independent of each other.

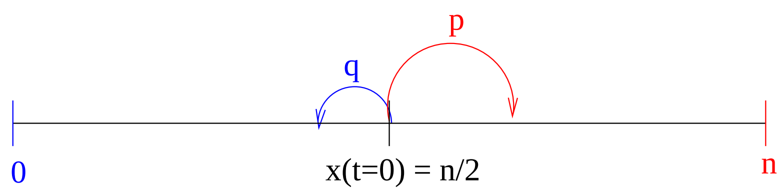


Figure 2: Gambler's ruin model illustration

1999). However, unlike the previous study, we estimate the time required for the sub-optimal building blocks to take over the entire population. This time to reach the absorbing state is an upper bound on the duration within which the probabilistic model should be rebuilt.

In the gambler's ruin problem, the gambler starts with some initial capital x_o, and competes with an opponent with an initial capital $N - x_o$. The gambler can reach one of the two absorbing states, one in which he loses all the money and the other in which the opponent loses all the money. At each step of the game, the gambler has a probability p of increasing his capital by one unit and a probability $1-p$ or losing one unit. Harik, Cantú-Paz, Goldberg, and Miller (1999) drew an analogy between the gambler's ruin problem and the growth in the market share of the best building block in competition with the second-best building block. Analogically to their model, in this study, we consider the decrease in the market share of schema ***0 (H_1) in competition with ***1 (H_2) in the partition belonging to the incorrectly modeled subproblem, as illustrated in Figure 2. That is, schema H_1 (the gambler) starts with an initial market share $x_o = N/2$, and competes with schema H_2 (the opponent) with a market share of $N - x_o = N/2$.

The probability of increasing the capital of H_1 by one unit is given by the probability that an individual with schema H_1 is better than an individual with schema H_2, called the decision-making probability (Goldberg, Deb, & Clark, 1992). Assuming additively separable problems, the fitness of the competing schemata H_1 and H_2 is $\mathcal{N}(\overline{f}_{H_1}, \sigma^2_{H_1})$ and $\mathcal{N}(\overline{f}_{H_2}, \sigma^2_{H_2})$, respectively, as shown in Figure 3. Where, $\overline{f}_{H_1}$ and $\overline{f}_{H_2}$ is the average fitness and $\sigma^2_{H_1}$ and $\sigma^2_{H_2}$ is the fitness variance of schema H_1 and H_2 respectively. The distance between average fitnesses of competing schemata is called the *signal* d:

$$d = \overline{f}_{H_1} - \overline{f}_{H_2}. \tag{1}$$

The probability of an individual with H_1 winning the competition over an individual with schema H_2 is equivalent to the probability that $f_{H_1} - f_{H_2} > 0$. Moreover, since f_{H_1} and f_{H_2} are normally distributed, $f_{H_1} - f_{H_2}$ is also normally distributed with mean d, and variance $\sigma^2_{H_1} + \sigma^2_{H_2}$:

$$f_{H_1} - f_{H_2} \sim \mathcal{N}\left(d, \sigma^2_{H_1} + \sigma^2_{H_2}\right). \tag{2}$$

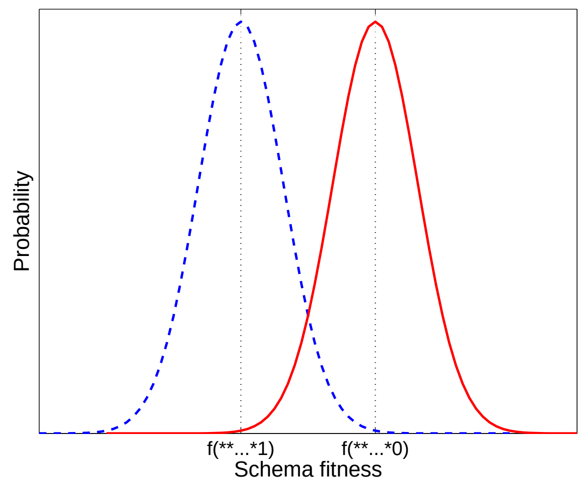


Figure 3: Decision making among competing building blocks

The probability of ***0 winning a tournament over ***1 is given by

$$p = \Phi\left(\frac{d}{\sqrt{\sigma^2_{H_1} + \sigma^2_{H_2}}}\right). \tag{3}$$

where, $\Phi(z)$ is the cumulative density function of a unit normal variate z. Since the fitness function is additively decomposable, $\sigma^2_{H_1}$ (and similarly $\sigma^2_{H_2}$) is the sum of the variance of the $m-1$ BBs that are correctly modeled and the variances of the $k-1$ alleles in the BB partition that is incorrectly modeled (Goldberg, Deb, & Clark, 1992). That is,

$$\sigma^2_{H_1} = \sigma^2_{H_2} = \sigma^2_H = (m-1)\sigma^2_{bb} + (k-1)\sigma^2_A, \tag{4}$$

where σ^2_{bb} is the fitness variance of a building block, σ^2_A is the fitness variance of a single variable (allele) in the incorrectly modeled subproblem, m is the number of subproblems, and k is the order of subproblems.

Given x_o and p, the total number of tournaments to reach an absorbing state, t_A, is given by (Feller, 1970)

$$t_A = \left(\frac{1}{1-2p}\right)\left[x_o - N\left(\frac{1-s^{x_o}}{1-s^N}\right)\right], \tag{5}$$

where $s = (1-p)/p$. Substituting for x_o and simplifying yields

$$t_A = -\frac{N}{2}\cdot\frac{1}{1-2p}\left[\frac{1-s^{N/2}}{1+s^{N/2}}\right]. \tag{6}$$

$$\begin{aligned} s &= \frac{1-p}{p}, \\ &= \frac{1-\Phi\left(\frac{d}{\sqrt{2}\sigma_H}\right)}{\Phi\left(\frac{d}{\sqrt{2}\sigma_H}\right)}. \end{aligned} \tag{7}$$

Approximating $\Phi(z)$ with first two terms of power series (Abramowitz & Stegun, 1972)

$$\Phi(z) \approx \frac{1}{2} + \frac{z}{\sqrt{2\pi}} \tag{8}$$

in Equation 7 and simplifying, we get

$$s = \frac{1 - \frac{d}{\sqrt{\pi}\sigma_H}}{1 + \frac{d}{\sqrt{\pi}\sigma_H}}. \tag{9}$$

Since $\sigma_H^2 \propto m\sigma_{bb}^2$, for moderate-to-large-sized problems ($m \geq 4$), $s \approx 1$. Using this approximation in Equation 6, and simplifying we get

$$t_A \approx \frac{N\sqrt{\pi}}{2}\left(\frac{\sigma_H}{d}\right). \tag{10}$$

It should be noted that the time to reach the absorbing state t_A is in units of number of tournaments. Since in a generation, there are N tournaments, the expected number of generations to reach an absorbing state is given by

$$t_{A,g} = \frac{t_A}{N} = \frac{\sqrt{\pi}}{2}\left(\frac{\sigma_H}{d}\right). \tag{11}$$

Recall that $\sigma_H^2 = (m-1)\sigma_{bb}^2 + (k-1)\sigma_A^2$. Since $m \gg k$ (bounded difficulty), σ_H^2 is approximately equal to $(m-1)\sigma_{bb}^2$. Substituting this approximation for σ_H^2, we get

$$t_{A,g} = \frac{\sqrt{\pi}}{2}\left(\frac{\sigma_{bb}}{d}\right) \cdot \sqrt{m}. \tag{12}$$

Therefore, the facetwise model for the structure-building period suggests that the update duration scales as square-root of the number of subproblems. Moreover, if the increase in the number of function evaluations due to stochastic model building is constant with the problem size, then the speed-up given by sporadic model building also scales as $\sqrt{m}$:

$$\eta \propto \sqrt{m}. \tag{13}$$

5 Experiments

This section presents experimental results. First, test problems are described and the experimental methodology is discussed. Next, the results are presented and discussed.

5.1 Test Problems

This section describes test problems used to test sporadic model building in hBOA. All test problems are nearly decomposable or hierarchical and assume that candidate solutions are represented by n-bit binary strings. Three test problems were used: (1) Dec-3, (2) Trap-5, and (3) hTrap:

1. *Dec-3: Concatenated 3-bit deceptive function.* In dec-3 (Deb & Goldberg, 1994), the input string is first partitioned into independent groups of 3 bits each. This partitioning is unknown to the algorithm and it does not change during the run. A 3-bit deceptive function is applied to each group of 3 bits and the contributions of all deceptive functions are added together to form the fitness. Each 3-bit deceptive function is defined as follows:

$$dec(u) = \begin{cases} 1 & \text{if } u = 3 \\ 0 & \text{if } u = 2 \\ 0.8 & \text{if } u = 1 \\ 0.9 & \text{if } u = 0 \end{cases}, \tag{14}$$

where u is the number of ones in the input string of 3 bits. The task is to maximize the function. An n-bit dec-3 function has one global optimum in the string of all ones and $2^{n/3} - 1$ other local optima. To solve dec-3, it is necessary to consider interactions among the positions in each partition because when each bit is considered independently, the optimization is misled away from the optimum (Thierens, 1995; Bosman & Thierens, 1999; Pelikan, Goldberg, & Cantú-Paz, 1999).

2. *Trap-5: Concatenated 5-bit trap.* Trap-5 can be defined analogically to dec-3, but instead of 3-bit groups, 5-bit groups are considered. The contribution of each group of 5 bits is computed as

$$trap_5(u) = \begin{cases} 5 & \text{if } u = 5 \\ 4 - u & \text{otherwise} \end{cases}, \tag{15}$$

   where u is the number of 1s in the input string of 5 bits. The task is to maximize the function. An n-bit trap5 function has one global optimum in the string of all ones and $2^{n/5} - 1$ other local optima. Traps of order 5 also necessitate that all bits in each group are treated together, because statistics of lower order are misleading.

3. *hTrap: Hierarchical Trap.* Dec-3 and trap-5 problems can be decomposed into separable subproblems of a fixed order. Nonetheless, hBOA can also solve problems that cannot be decomposed into subproblems of bounded order on a single level, but that must be solved hierarchically by building the solution from low-order building blocks over a number of levels of difficulty.

   An example hierarchical problem that cannot be efficiently solved using a single-level decomposition are hierarchical traps (hTraps) (Pelikan, 2002), created by combining trap functions of order 3 over multiple levels of difficulty. For hTraps, the string length should be an integer power of 3, that is, $n = 3^l$. On the lowest level, groups of 3 bits contribute to the overall fitness using generalized 3-bit traps defined as follows:

$$trap_3(u) = \begin{cases} f_{high} & \text{if } u = 3 \\ f_{low} - u\frac{f_{low}}{2} & \text{otherwise} \end{cases}, \tag{16}$$

   where $f_{high} = 1$ and $f_{low} = 1 + 0.1/l$.

   Each group of 3 bits corresponding to one of the traps is then mapped to a single symbol on the next level; a 000 is mapped to a 0, a 111 is mapped to a 1, and everything else is mapped to the null symbol '-'. The bits on the next level again contribute to the overall fitness using 3-bit traps defined above, and the groups are mapped to an even higher level. This continues until the top level is evaluated that contains 3 bits total. However, on the top level, a trap with $f_{high} = 1$ and $f_{low} = 0.9$ is applied. Any group of bits containing the null symbol does not contribute to the overall fitness. To make the overall contribution of each level of the same magnitude, the contributions of traps on ith level from the bottom are multiplied by 3^i.

   The task is to maximize the function. hTraps have many local optima, but only one global optimum in the string of all ones. Nonetheless, any single-level decomposition into subproblems of bounded order will lead away from the global optimum. That is why hTraps necessitate an optimizer that can build solutions hierarchically by juxtaposing good partial solutions over multiple levels of difficulty until the global optimum if found.

   For more details on hierarchical traps and solving hierarchical optimization problems, please see Pelikan (2002). Other hierarchical problems can be found in Watson, Hornby, and Pollack (1998).

5.2 Experimental Methodology

To analyze scalability of SMB, for each problem, we performed experiments for a range of problem sizes (n = 30 to 210 with step 15 for dec-3 and trap-5; n = 9, 27, 81, and 243 for hTrap). For each problem and problem size, hBOA with SMB was tested for t_{sb} from 1 to 20 with step 1. For each problem, problem size, and value of t_{sb}, a minimum population size required to find the global optimum in 10 out of 10 independent runs was determined using the bisection method (Pelikan, 2005). To reduce noise, for each setting, 10 bisection runs are performed. Therefore, for each problem, problem size, and value of t_{sb}, 100 successful independent runs are performed. Each run is terminated either when the global optimum is found or when hBOA has completed a large number of generations and it is unlikely that the algorithm to find the optimum in a reasonable time (the maximum number of iterations is determined based on preliminary experiments).

Binary tournament selection is used in all experiments and the window size for RTR is set as $w = \min\{n, N/20\}$ where n is the problem size and N is the population size. BDe metric with a penalty for complex networks as described in Pelikan (2005) is used in all experiments to measure quality of BNs with decision trees.

5.3 Results

The experimental results were processed to obtain the following quantities:

1. *Speedup of structure building for each problem, problem size, and structure building period.* The speedup is averaged over the 100 independent runs for each particular setting. To compute the speedup of model building, all factors have been incorporated into the theoretical model for computational complexity of hBOA, including the effects of SMB on the population size and the number of iterations until convergence, which are both measured empirically. The resulting speedup thus represents the actual speedup of the structure-building procedure of hBOA.

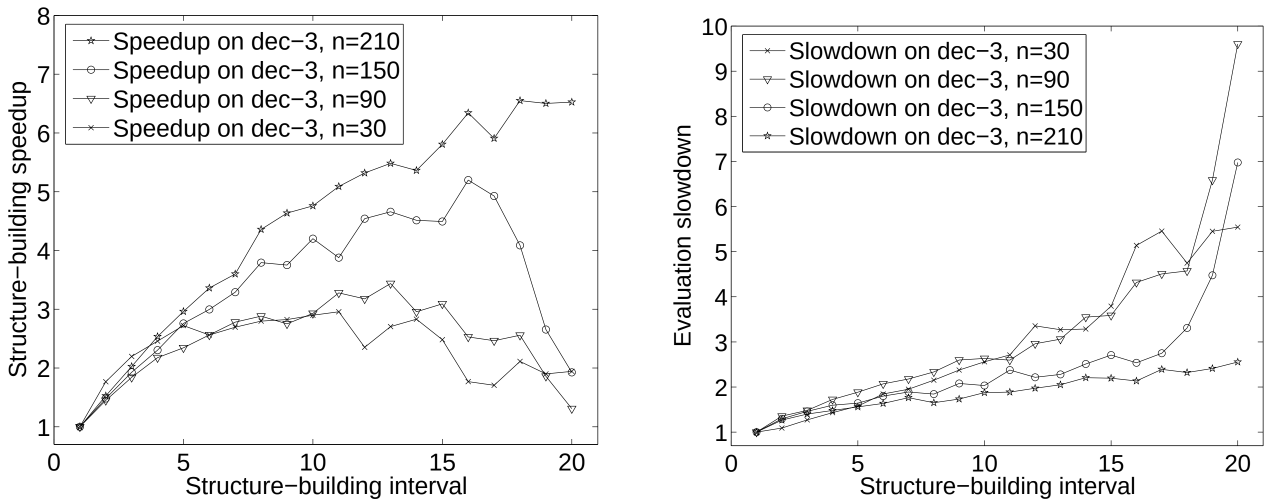
2. *Optimal speedup of structure building for each problem and problem size.* For each problem and problem size, an optimal speedup is determined empirically by using the value of t_{sb} that maximizes the speedup of structure building.

3. *Evaluation slowdown for optimal speedup of structure building.* For each problem and problem size, the slowdown of the evaluation time is determined for the best value of t_{sb}, which corresponds to the optimal speedup. The evaluation slowdown is defined as a factor by which the number of evaluations increases compared to the case with no SMB (that is, $t_{sb} = 1$).

4. *Scalability.* Scalability with various values of t_{sb} is evaluated to investigate the effects of SMB on the scalability of hBOA on decomposable problems. Specifically, scalability with $t_{sb} = 1$ and $t_{sb} = 10$ is evaluated and compared.

Figure 4 shows the structure-building speedup and the evaluation slowdown for several problem sizes of dec-3. Figure 5 shows the optimal speedup of structure building and the corresponding evaluation slowdown. Figure 6 shows the growth of the number of evaluations with problem size on dec-3 for $t_{sb} = 1$ (without SMB) and $t_{sb} = 10$ (with SMB where the structure is built in every 10th iteration).

Figure 7 shows the structure-building speedup and the evaluation slowdown for several problem sizes of trap-5. Figure 8 shows the optimal speedup of structure building and the corresponding evaluation slowdown. Figure 9 shows the growth of the number of evaluations with problem size on



(a) Speedup of structure building.

(b) Slowdown of evaluation.

Figure 4: The speedup of structure building and the slowdown of evaluation for dec-3 of 30, 90, 150 and 210 bits.

trap-5 for $t_{sb} = 1$ (without SMB) and $t_{sb} = 10$ (with SMB where the structure is built in every 10th iteration).

Figure 10 shows the structure-building speedup and the evaluation slowdown for several problem sizes of hTrap. Figure 11 shows the optimal speedup of structure building and the corresponding evaluation slowdown. Figure 12 shows the growth of the number of evaluations with problem size on hTrap for $t_{sb} = 1$ (without SMB) and $t_{sb} = 10$ (with SMB where the structure is built in every 10th iteration).

5.4 Discussion of Results

The results presented in figures 4, 7 and 10 indicate that for all test problems, the structure-building speedup grows with the structure building period t_{sb} until it reaches a local maximum at the optimal value of t_{sb} that maximizes the speedup; the rate of speedup growth increases with problem size. Additionally, these results indicate that the evaluation slowdown becomes less significant as the problem size increases and for larger problems, the evaluation slowdown remains nearly constant for all values of t_{sb}.

The results presented in figures 5, 8 and 11 indicate that the optimum speedup grows with problem size and its polynomial approximation grows as $\Theta(n^{0.26})$ to $\Theta(n^{0.65})$; that means that the maximum speedup of structure building is expected to grow with problem size and lead to a significant decrease of asymptotic complexity of model building in hBOA. On the other hand, the evaluation slowdown corresponding to the best value of t_{sb} is much less significant than the speedup of model building. In fact, the evaluation slowdown corresponding to the optimal speedup of model building decreases with problem size for both decomposable problems of bounded difficulty (dec-3 and trap-5), although it slowly increases for the hierarchical problem (hTrap).

The results presented in figures 6, 9 and 12 indicate that SMB does not negatively affect the polynomial growth of the number of evaluations until convergence with problem size; specifically, they show that with a fixed structure building period $t_{sb} = 10$, the number of evaluations appears to grow with a polynomial of the same order as without SMB or even better.

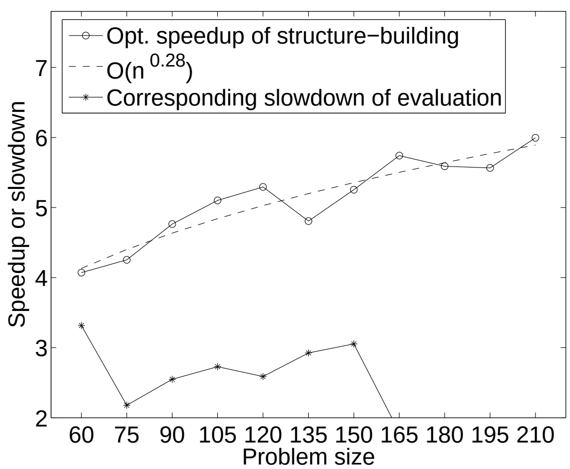


Figure 5: The optimal speedup of structure building and the corresponding evaluation slowdown for dec-3 of 60 to 210 bits.

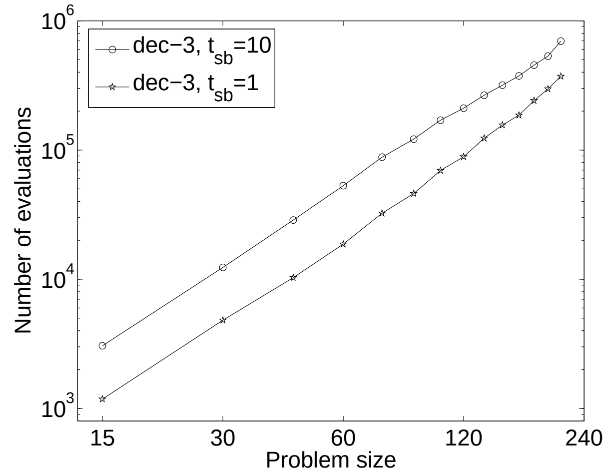


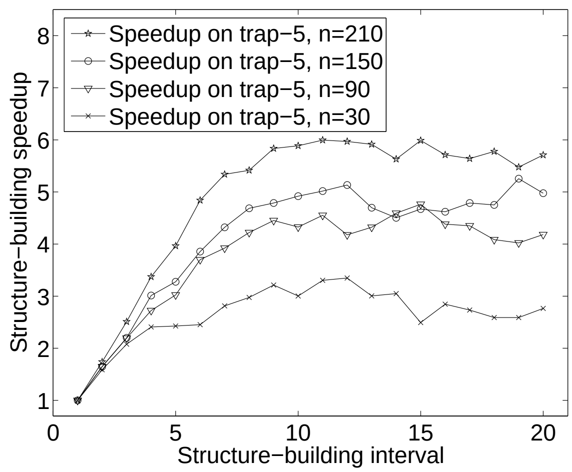
Figure 6: The scalability on dec-3 of 60 to 210 bits with $t_{sb} = 1$ (no SMB) and $t_{sb} = 10$ (with SMB at every 10th iteration).

Therefore, the results lead to four important observations:

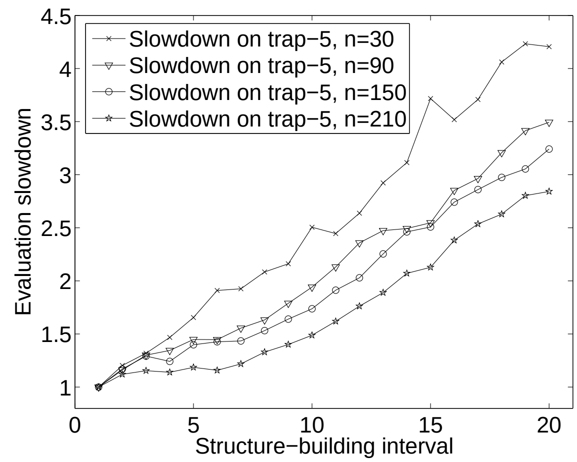
1. SMB leads to a significant decrease of the asymptotic time complexity of model building.
2. The optimal speedup obtained with SMB grows with problem size.
3. The factor by which SMB increases the number of evaluations is insignificant compared to the speedup of model building and it decreases with problem size for decomposable problems of bounded difficulty.
4. SMB does not lead to an increase of the asymptotic complexity of hBOA with respect to the number of evaluations until convergence.

6 Summary and Conclusions

This paper presented and analyzed an efficiency enhancement technique called *sporadic model building*, which can be used to speed up model building in BOA, hBOA and other similar EDAs. In sporadic model building, the structure of the probabilistic model is not updated in every iteration; instead, in some iterations only the parameters of the previous structure are updated with respect to the selected population of promising solutions. Building the model structure is the most computationally expensive part of hBOA variation and can become a computational bottleneck for large and complex problems. That is why sporadic model building represents an important efficiency enhancement technique for BOA, hBOA, and other similar EDAs. The paper proposed a simple schedule for sporadic model building, where the updates of the model structure are done with a fixed period t_{sb} called the structure-building period. For example, $t_{sb} = 1$ denotes the scenario where the structure is updated in every iteration, whereas $t_{sb} = 3$ denotes the scenario where the structure is updated every third iteration. The paper analyzed the proposed approach to sporadic model building and presented a dimensional theoretical model that can be used to guide the choice of an appropriate value of t_{sb} and to estimate the expected speedup of model building.



(a) Speedup of structure building.



(b) Slowdown of evaluation.

Figure 7: The speedup of structure building and the slowdown of evaluation for trap-5 of 60, 120 and 180 bits.

The experimental results presented in this paper lead to three important observations regarding the use of sporadic model building for solving nearly decomposable and hierarchical problems by hBOA:

1. Sporadic model building leads to significant improvements of asymptotic time complexity of hBOA model building and the optimal speedup of model building grows with problem size (here the optimal speedup grows as $\Theta(n^{0.26})$ to $\Theta(n^{0.65})$ depending on the problem). The speedup of hBOA variation is upper bounded by $\Theta(n)$ because of parameter updates and sampling.

2. Sporadic model building leads to an increase in the number of evaluations until convergence but the factor by which sporadic model building increases the overall number of evaluations is insignificant and it decreases with problem size for decomposable problems of bounded difficulty.

3. Sporadic model building does not lead to an increase of the asymptotic complexity of hBOA with respect to the number of evaluations until convergence and the number of evaluations with a fixed value of t_{sb} remains a low-order polynomial for nearly decomposable and hierarchical problems.

The dimensional theoretical model uses the gambler's ruin model to provide a method for setting t_{sb} and suggests that for nearly decomposable problems, t_{sb} should be proportional to a square root of the problem size, $t_{sb} \propto \sqrt{n}$. Given the empirical evidence that the factor by which the number of evaluations increases is upper bounded by a constant for each problem, the expected optimal speedup with sporadic model building can also be expected to be proportional to $\sqrt{n}$. Nonetheless, the theoretical model assumes that no niching is used and, therefore, for hBOA the derived bound may be more restrictive than necessary.

Acknowledgments

This work was partially supported by the Research Award and the Research Board at the University of Missouri. Most experiments were done using the hBOA software developed by Martin Pelikan

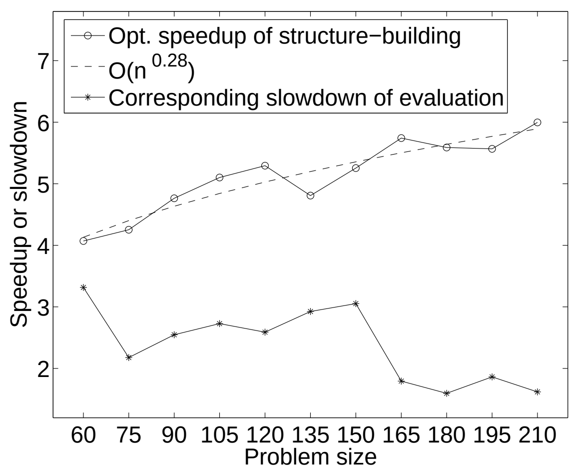


Figure 8: The optimal speedup of structure building and the corresponding evaluation slowdown for trap-5 of 60 to 210 bits.

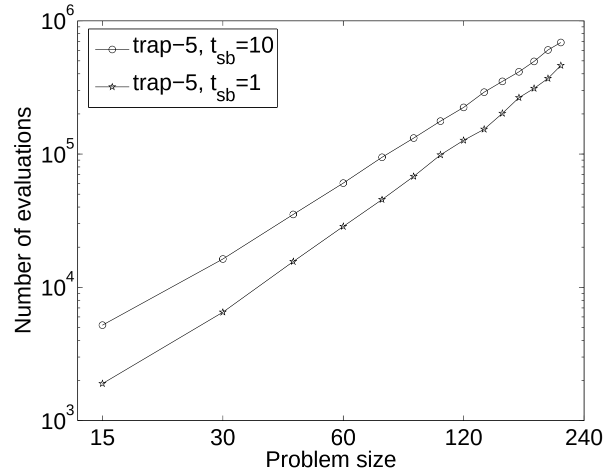
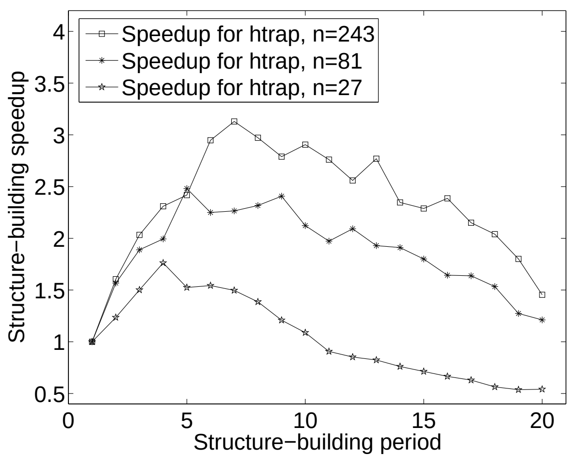


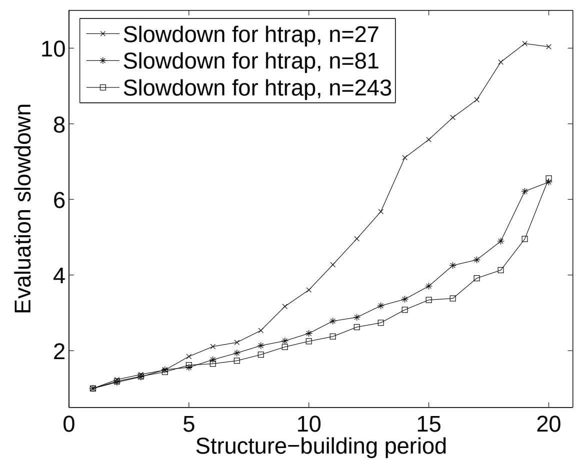
Figure 9: The scalability on trap-5 of 60 to 210 bits with $t_{sb} = 1$ (no SMB) and $t_{sb} = 10$ (with SMB at every 10th iteration).

and David E. Goldberg at the University of Illinois at Urbana-Champaign. Most experiments were completed at the Beowulf cluster at the University of Missouri at St. Louis.

This work was also sponsored by the Air Force Office of Scientific Research, Air Force Materiel Command, USAF, under grant F49620-03-1-0129, the National Science Foundation under ITR grant DMR-99-76550 (at Materials Computation Center), and ITR grant DMR-0121695 (at CPSD), and the Dept. of Energy under grant DEFG02-91ER45439 (at Fredrick Seitz MRL). The U.S. Government is authorized to reproduce and distribute reprints for government purposes notwithstanding any copyright notation thereon.



(a) Speedup of structure building.



(b) Slowdown of evaluation.

Figure 10: The speedup of structure building and the slowdown of evaluation for hTrap of 60, 120 and 180 bits.

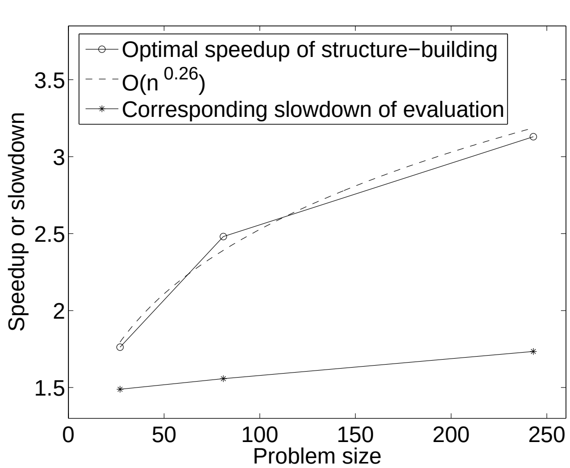


Figure 11: The optimal speedup of structure building and the corresponding evaluation slowdown for hTrap of 27 to 243 bits.

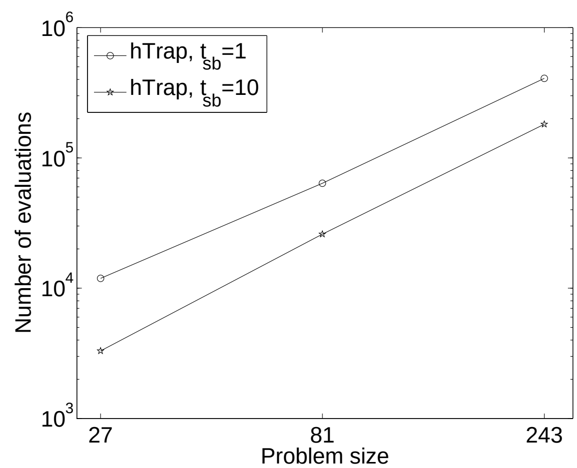


Figure 12: The scalability on hTrap of 27 to 243 bits with $t_{sb} = 1$ (no SMB) and $t_{sb} = 10$ (with SMB at every 10th iteration).

References

Abramowitz, M., & Stegun, I. A. (Eds.) (1972). *Handbook of mathematical functions with formulas, graphs, and mathematical tables*. Dover publications. Ninth printing.

Albert, L. A. (2001). *Efficient genetic algorithms using discretization scheduling*. Master's thesis, University of Illinois at Urbana-Champaign, Department of General Engineering, Urbana, IL.

Bosman, P. A. N., & Thierens, D. (1999). Linkage information processing in distribution estimation algorithms. *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO-99)*, *I*, 60–67.

Chickering, D. M., Heckerman, D., & Meek, C. (1997). *A Bayesian approach to learning Bayesian networks with local structure* (Technical Report MSR-TR-97-07). Redmond, WA: Microsoft Research.

Deb, K., & Goldberg, D. E. (1994). Sufficient conditions for deceptive and easy binary functions. *Annals of Mathematics and Artificial Intelligence*, *10*, 385–408.

Feller, W. (1970). *An introduction to probability theory and its applications*. New York, NY: Wiley.

Friedman, N., & Goldszmidt, M. (1999). Learning Bayesian networks with local structure. In Jordan, M. I. (Ed.), *Graphical models* (pp. 421–459). Cambridge, MA: MIT Press.

Goldberg, D. E. (1987). Simple genetic algorithms and the minimal, deceptive problem. In Davis, L. (Ed.), *Genetic Algorithms and Simulated Annealing* (Chapter 6, pp. 74–88). Los Altos, CA: Morgan Kaufmann.

Goldberg, D. E. (1989). *Genetic algorithms in search, optimization, and machine learning*. Reading, MA: Addison-Wesley.

Goldberg, D. E. (2002). *The design of innovation: Lessons from and for competent genetic algorithms*, Volume 7 of *Genetic Algorithms and Evolutionary Computation*. Kluwer Academic Publishers.

Goldberg, D. E., Deb, K., & Clark, J. H. (1992). Genetic algorithms, noise, and the sizing of populations. *Complex Systems*, *6*, 333–362.

Harik, G. (1999). *Linkage learning via probabilistic modeling in the ECGA* (IlliGAL Report No. 99010). Urbana, IL: University of Illinois at Urbana-Champaign, Illinois Genetic Algorithms Laboratory.

Harik, G., Cantú-Paz, E., Goldberg, D. E., & Miller, B. L. (1999). The gambler's ruin problem, genetic algorithms, and the sizing of populations. *Evolutionary Computation*, *7*(3), 231–253.

Harik, G. R. (1995). Finding multimodal solutions using restricted tournament selection. *Proceedings of the International Conference on Genetic Algorithms (ICGA-95)*, 24–31.

Heckerman, D., Geiger, D., & Chickering, D. M. (1994). *Learning Bayesian networks: The combination of knowledge and statistical data* (Technical Report MSR-TR-94-09). Redmond, WA: Microsoft Research.

Holland, J. H. (1975). *Adaptation in natural and artificial systems*. Ann Arbor, MI: University of Michigan Press.

Howard, R. A., & Matheson, J. E. (1981). Influence diagrams. In Howard, R. A., & Matheson, J. E. (Eds.), *Readings on the principles and applications of decision analysis*, Volume II (pp. 721–762). Menlo Park, CA: Strategic Decisions Group.

Koza, J. R. (1992). *Genetic programming: On the programming of computers by means of natural selection*. Cambridge, MA: The MIT Press.

Larrañaga, P., & Lozano, J. A. (Eds.) (2002). *Estimation of distribution algorithms: A new tool for evolutionary computation*. Boston, MA: Kluwer.

Mühlenbein, H., & Paaß, G. (1996). From recombination of genes to the estimation of distributions I. Binary parameters. *Parallel Problem Solving from Nature*, 178–187.

Mühlenbein, H., & Schlierkamp-Voosen, D. (1993). Predictive models for the breeder genetic algorithm: I. Continuous parameter optimization. *Evolutionary Computation*, *1*(1), 25–49.

Pearl, J. (1988). *Probabilistic reasoning in intelligent systems: Networks of plausible inference*. San Mateo, CA: Morgan Kaufmann.

Pelikan, M. (2002). *Bayesian optimization algorithm: From single level to hierarchy*. Doctoral dissertation, University of Illinois at Urbana-Champaign, Urbana, IL. Also IlliGAL Report No. 2002023.

Pelikan, M. (2005). *Hierarchical Bayesian optimization algorithm: Toward a new generation of evolutionary algorithms*. Springer-Verlag.

Pelikan, M., & Goldberg, D. E. (2001). Escaping hierarchical traps with competent genetic algorithms. *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO-2001)*, 511–518.

Pelikan, M., & Goldberg, D. E. (2003). A hierarchy machine: Learning to optimize from nature and humans. *Complexity*, *8*(5), 36–45.

Pelikan, M., Goldberg, D. E., & Cantú-Paz, E. (1999). BOA: The Bayesian optimization algorithm. *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO-99)*, *I*, 525–532.

Pelikan, M., Goldberg, D. E., & Lobo, F. (2002). A survey of optimization by building and using probabilistic models. *Computational Optimization and Applications*, *21*(1), 5–20.

Pelikan, M., Sastry, K., & Goldberg, D. E. (2002). Scalability of the Bayesian optimization algorithm. *International Journal of Approximate Reasoning*, *31*(3), 221–258.

Rechenberg, I. (1973). *Evolutionsstrategie: Optimierung technischer Systeme nach Prinzipien der biologischen Evolution*. Stuttgart: Frommann-Holzboog.

Sastry, K. (2001). *Evaluation-relaxation schemes for genetic and evolutionary algorithms*. Master's thesis, University of Illinois at Urbana-Champaign, Department of General Engineering, Urbana, IL. Also IlliGAL Report No. 2002004.

Sastry, K., & Goldberg, D. E. (2000). *On extended compact genetic algorithm* (IlliGAL Report No. 2000026). Urbana, IL: University of Illinois at Urbana-Champaign, Illinois Genetic Algorithms Laboratory.

Sastry, K., Pelikan, M., & Goldberg, D. E. (2004). Efficiency enhancement of genetic algorithms via building-block-wise fitness estimation. *Proceedings of the IEEE Conference on Evolutionary Computation*, 720–727.

Sinha, A., & Goldberg, D. E. (2001). Verification and extension of the theory of global-local hybrids. *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO-2001)*, 591–597. Also IlliGAL Report No. 2001010.

Srivastava, R., & Goldberg, D. E. (2001). Verification of the theory of genetic and evolutionary continuation. *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO-2001)*, 551–558. Also IlliGAL Report No. 2001007.

Thierens, D. (1995). *Analysis and design of genetic algorithms*. Doctoral dissertation, Katholieke Universiteit Leuven, Leuven, Belgium.

Thierens, D., Goldberg, D. E., & Pereira, A. G. (1998). Domino convergence, drift, and the temporal-salience structure of problems. *Proceedings of the International Conference on Evolutionary Computation (ICEC-98)*, 535–540.

Watson, R. A., Hornby, G. S., & Pollack, J. B. (1998). Modeling building-block interdependency. *Parallel Problem Solving from Nature*, 97–106.