Evolution Strategies for Mixed-Integer Optimization of Optical Multilayer Systems

Thomas Bäck and Martin Schütz

Abstract

An extension of the evolution strategy for mixed-integer optimization problems is introduced. The resulting generalized evolution strategy is applied to the problem of optical multilayer coating design and the results are compared with results obtained by standard methods. The generalized evolution strategy as a synthesis method does not require the existence of a starting design, and it competes well with refinement methods for the optimization of starting designs. The results are very encouraging and indicate that this method is a robust and helpful algorithm for optical multilayer design. Furthermore, the generalized evolution strategy is not a tailored heuristic but can be used for arbitrary mixed-integer optimization problems.

1 INTRODUCTION

Evolution strategies (ES) (Rechenberg 1973; Schwefel 1981) are a member of the class of so-called *evolutionary algorithms*, probabilistic search and optimization algorithms gleaned from organic evolution (see e.g. Bäck and Schwefel 1993). Evolution strategies work by imitating the biological principles of a *population* of *individuals* which undergo processes of *recombination*, *mutation*, and *selection*. Each individual is interpreted as a search point whose quality (*fitness*) can be measured according to the optimization problem the algorithm is applied to. The selection operator favors the individuals of higher quality to produce individuals for the next generation, such that the population evolves towards increasingly better regions of the search space.

Although some very early experiments with the evolution strategy dealt with discrete optimization problems using binomially distributed mutations (see Klockgether and Schwefel 1970; Lichtfuss 1965; Schwefel 1968), the modern (μ^{+},λ)-ES (the notation summarizes both the $(\mu+\lambda)$- and (μ,λ)-variant of the ES; see Bäck and Schwefel 1993) is designed for solving continuous parameter optimization problems of the form

$$\min\{f(\vec{x}) \mid \vec{x} \in M \subseteq I\!R^n\} \ . \tag{1}$$

Many application problems in industry, however, involve the simultaneous use of discrete and continuous object variables (Kelahan and Gaddy

1978). Therefore, a more general formulation of the global optimization problem as

$$\min\{f(\vec{x},\vec{d}) \mid \vec{x} \in M \subseteq I\!R^{n_x} \ , \ \vec{d} \in N \subseteq Z\!\!\!Z^{n_d}\} \tag{2}$$

has greater applicability, and it is highly desirable to develop an extension of the evolution strategy capable of yielding good solutions to nonlinear, mixed-integer optimization problems. There are many reasons to develop such an algorithm.

First, it is not sufficient to rely on the naive approach of embedding $Z\!\!\!Z$ in $I\!R$ and using the standard evolution strategy with rounding of the discrete object variables. For a discrete optimization problem, the optimum point obtained by rounding the result of the continuous optimization might be different from the true discrete optimum point even for linear objective functions with linear constraints (see e.g. Schütz 1994, p. 46). Furthermore, the evolution strategy mechanism to approximate an optimum with arbitrary precision can cause stagnation of the search in case of discrete optimization problems if the step sizes drop below one, because one is the smallest distance between two different points in $Z\!\!\!Z^{n_d}$.

Second, one might be tempted to optimize discrete and continuous parameters independently of each other as done in the *structure evolution* approach by Lohmann (1992). This method, however, assumes a separability between discrete and continuous optimization, which is not necessarily the case for highly nonlinear interactions between discrete and continuous variables. Additionally, structure evolution requires the development of specialized genetic operators rather than offering a general-purpose algorithm for mixed-integer optimization.

Third, practically no generally applicable algorithms for mixed-integer optimization are available so far, although many practical applications clarify the strong need for such an algorithm.

In this paper, an evolution strategy is generalized to a heuristic for mixed-integer optimization, and the algorithm is described in section 2. Section 3 demonstrates the practical application of the new algorithm to the synthesis of optical multilayer systems and shows the feasibility of the method. In section 4, some concluding remarks and an outline of further research are given.

2 THE GENERALIZED EVOLUTION STRATEGY

The *generalized evolution strategy* (GES) for mixed-integer optimization is described in this section by using a notation for evolutionary algorithms which was introduced in Bäck and Schwefel (1993) (see also Bäck and Hoffmeister 1994 for an overview of evolution strategies). The basic idea of the GES consists in the introduction of *mutation probabilities* $p_j \in (0,1)$ as *strategy parameters* that control the application of mutation to the discrete parameters. Assuming a dimension n_x of the continuous part of the search space and n_d of the discrete part, the space of individuals extends to:

$$I = I\!R^{n_x} \times Z\!\!\!Z^{n_d} \times A_s \ , \tag{3}$$

where $A_s = I\!R_+^{n_\sigma} \times (0,1)^{n_p}$ ($n_\sigma \in \{1,\ldots,n_x\}$, $n_p \in \{1,\ldots,n_d\}$) denotes the *strategy parameter space* (n_σ standard deviations $(\sigma_1,\ldots,\sigma_{n_\sigma})$ for the

normally distributed mutation of object variables x_i and n_p mutation probabilities $(p_1, \ldots, p_{n_p})$ for mutation of the object variables d_l).

The strategy parameters of a standard (μ,λ)-ES include up to $n_\sigma = n_x$ standard deviations in case of the standard mutation operator and additionally up to $n_x \cdot (n_x - 1)/2$ covariances in case of *correlated mutations* (for detailed information see Bäck and Schwefel 1993). Together, the set of strategy parameters attached to an individual describes a generalized n_x-dimensional normal distribution which controls the mutation of this individual. The strategy parameters are not controlled by an exogenous process but undergo mutation and recombination in a similar way as object variables do, such that the probability density functions for mutating individuals evolve over time according to the topological characteristics of the objective function. The selection operator favors advantageous strategy parameters by exploiting the implicit link between strategy parameters and fitness. For a more detailed introduction to the topic of self-adaptation, the reader is referred to Hoffmeister and Bäck (1992) and Schwefel (1992).

The GES discussed in this paper does not use correlated mutations, although the mechanism has been implemented. A description of the principles of correlated mutations can be found in Bäck and Hoffmeister (1994) and Rudolph (1992). The generalized evolution strategy self-adapts discrete mutation probabilities $\vec{p}$ in addition to standard deviations $\vec{\sigma}$.

Individuals as members of a population $P(t)$ at generation t are denoted

$$\vec{a}_i(t) = (\vec{x}_i(t), \vec{d}_i(t), \vec{\sigma}_i(t), \vec{p}_i(t)) \ .$$

The fitness $F : I \rightarrow I\!R$ of an individual is defined by means of the value of the objective function f, i.e., $F(\vec{a}(t)) = f(\vec{x}(t), \vec{d}(t))$. In the high-level pseudocode notation introduced in Bäck and Schwefel (1993), the generalized evolution strategy can be described as follows:

ALGORITHM 1 (outline of the GES)

$t := 0$;
initialize $P(0) := \{\vec{a}_1(0), \ldots, \vec{a}_\mu(0)\} \in I^\mu$
  where $I = I\!R^{n_x} \times Z\!\!Z^{n_d} \times I\!R_+^{n_\sigma} \times (0,1)^{n_p}$
  and $\vec{a}_i(t) = (\vec{x}_i(t), \vec{d}_i(t), \vec{\sigma}_i(t), \vec{p}_i(t))$
    $\forall i \in \{1, \ldots, \mu\}$;
evaluate $P(0)$: $\{F(\vec{a}_1(0)), \ldots, F(\vec{a}_\mu(0))\}$
  where $F(\vec{a}_i(0)) = f(\vec{x}_i(0), \vec{d}_i(0))$;
while $(\iota(P(t)) \neq$ **true**$)$ **do**
  recombine: $\vec{a}'_i(t) := r'_{\{r_x r_d r_\sigma r_p\}}(P(t))$
    $\forall i \in \{1, \ldots, \lambda\}$;
  mutate: $\vec{a}''_i(t) := m'_{\{\tau, \tau', \gamma\}}(\vec{a}'_i(t))$
    $\forall i \in \{1, \ldots, \lambda\}$;
  evaluate: $P''(t) := \{\vec{a}''_1(t), \ldots, \vec{a}''_\lambda(t)\}$:
    $\{F(\vec{a}''_1(t)), \ldots, F(\vec{a}''_\lambda(t))\}$
    where $F(\vec{a}''_i(t)) = f(\vec{x}''_i(t), \vec{d}''_i(t))$;
  select: $P(t+1) :=$ **if** *(μ,λ)-selection*
    then $s_{(\mu,\lambda)}(P''(t))$;
    else $s_{(\mu+\lambda)}(P(t) \cup P''(t))$;
  $t := t + 1$;
od

In comparison with the standard ES, extensions of the algorithm are almost exclusively related to the mutation operator. The selection operator remains completely unchanged ($s_{(\mu+\lambda)}$ deterministically selects the μ best individuals out of the union of parents and offspring to survive, while $s_{(\mu,\lambda)}$ selects the μ best out of the offspring population). Individuals are evaluated by applying the mixed-integer objective function f to the $\vec{x}$ and $\vec{d}$ components of individuals. The recombination operator is extended to the discrete object variables and the corresponding strategy parameters, and the implementation covers all possible recombination mechanisms, i.e.:

- *Discrete recombination*: For each variable of the offspring individual the variable is copied from either the first or the second (randomly chosen) parent with probability 1/2.
- *Intermediate recombination*: For each variable of the offspring individual the value is calculated as the average of the corresponding variables of the (randomly chosen) parent individuals. This can easily be generalized by allowing for arbitrary weighting factors $w \in [0,1]$ and $1-w$ instead of $w = 0.5$.

These operators can also be applied in their *global* form where one randomly chosen parent is held fixed and the second parent is randomly chosen anew for each single variable. Furthermore, different recombination operators are usually applied to the component vectors of an individual, such that a complete recombination operator is denoted $r'_{\{r_x r_d r_\sigma r_p\}}$ in the following. $r_x, r_\sigma, r_p \in \{-, d, D, i, I, r, R\}$ denote the possible choices of recombination for $\vec{x}$, $\vec{\sigma}$, and $\vec{p}$, while $r_d \in \{-, d, D\}$ characterizes the choices for $\vec{d}$. The codes $-, d, D, i, I, r, R$ specify no, discrete, global discrete, intermediate, global intermediate, generalized intermediate, and global generalized intermediate recombination as outlined above. More formally, $\forall c \in \{x, d, \sigma, p\}$ and $\forall i \in \{1, \ldots, n_c\}$, the recombination operator can be described as follows:

$$r'_{\{r_x r_d r_\sigma r_p\}}(P(t)) = \vec{a}' = (\vec{x}', \vec{d}', \vec{\sigma}', \vec{p}') \tag{4}$$

where

$$c'_i \quad = \quad \begin{cases} c_{S,i} & (-) \\ c_{S,i} \text{ or } c_{T,i} & (d) \\ c_{S,i} \text{ or } c_{T_i,i} & (D) \\ c_{S,i} + (c_{T,i} - c_{S,i})/2 & (i) \\ c_{S,i} + (c_{T_i,i} - c_{S,i})/2 & (I) \\ c_{S,i} + u \cdot (c_{T,i} - c_{S,i}) & (r) \\ c_{S,i} + u_i \cdot (c_{T_i,i} - c_{S,i}) & (R) \end{cases} \tag{5}$$

The indices S and T denote two parents selected at random from the population (the index i in T_i indicates T to be sampled anew for each value of i), $u \in [0,1]$ is a uniform random variable, sampled anew for each possible value of the counter i when used in the form u_i and sampled once per creation of one offspring individual when it is not indexed. "or" denotes a decision by a fair random coin toss. As indicated by the indices r_x, r_d, r_σ and r_p the complete recombination operator results from combining the component-wise operators. For example, the code r'_{IDIi} denotes a recombination operator which performs global intermediate recombination on $\vec{x}$

and $\vec{\sigma}$, intermediate recombination on $\vec{p}$, and global discrete recombination on $\vec{d}$.

The mutation operator $m'_{\{\tau,\tau',\gamma\}} : I \rightarrow I$ operates on continuous variables $\vec{x}$ and the corresponding standard deviations $\vec{\sigma}$ just as in case of the standard evolution strategy, i.e., ($\forall i \in \{1,\ldots,n_x\}$) (Bäck and Hoffmeister 1994):

$$\sigma'_i = \sigma_i \cdot \exp(\tau' \cdot N(0,1) + \tau \cdot N_i(0,1)) \quad (6)$$

$$x'_i = x_i + \sigma'_i \cdot N_i(0,1)\ . \quad (7)$$

Here, $N(0,1)$ denotes a realization of a normally distributed random variable with expectation zero and standard deviation one. The index i in $N_i(0,1)$ indicates that the random variable is sampled anew for each possible value of i. Schwefel (1981) suggests values

$$\tau \propto \frac{1}{\sqrt{2\sqrt{n_x}}} \quad , \quad \tau' \propto \frac{1}{\sqrt{2n_x}} \quad (8)$$

for the exogenous control parameters τ and τ' ("learning rates"). Usually, the constant of proportionality equals one in both cases.

For the mutation probabilities p_j, (with $j \in \{1,\ldots,n_p\}$), a mutation mechanism is searched for such that after mutation $p'_j \in (0,1)$ is still guaranteed. As usual, small changes should be more likely than large ones, and changes towards increasing p_j should occur with the same probability as changes towards decreasing p_j (i.e., the median equals p_j). Based on these requirements, Obalek (1994) proposed to use a logistic transformation of the form ($\forall j \in \{1,\ldots,n_p\}$):

$$p'_j = \left(1 + \frac{1-p_j}{p_j} \cdot \exp\left(-\gamma \cdot N_j(0,1)\right)\right)^{-1} \quad , \quad (9)$$

such that p'_j is distributed according to the logistic normal distribution with probability density function

$$f_{p'_j}(x) = \frac{1}{\sqrt{2\pi}\gamma x(1-x)} \exp\left(\frac{-\ln\left(\frac{1}{x(1-x)} - \zeta\right)^2}{2\gamma^2}\right) \quad , \quad (10)$$

where $\zeta = \ln \frac{p_j}{1-p_j}$. In the implementation, a transformation is used which guarantees that the p_j neither converge towards zero nor towards one due to the limited precision of the representation of numbers on a computer. The transformation simply resets $p'_j := \varepsilon_p$ if $p'_j < \varepsilon_p$ and $p'_j := 1 - \varepsilon_p$ if $p'_j > 1 - \varepsilon_p$, where ε_p is a small constant value (10^{-40} in the experimental results presented in section 3). This transformation has no impact on the probability density function (10) (see Schütz 1994, p. 60). In analogy with τ, the learning parameter γ is chosen according to the proportionality

$$\gamma \propto \frac{1}{\sqrt{2\sqrt{n_d}}}\ . \quad (11)$$

The experiments with this mutation mechanism clearly demonstrated that the learning process works well when $n_p = 1$ is chosen, but it has

some problems in the general case $n_p = n_d$ (see Schütz 1994, pp. 89–90). Therefore, the following experiments were performed with $n_p = 1$, and further tests of the mutation mechanism remain to be performed in future work.

The mutated mutation probability p' is used to determine whether the discrete object variables d_l ($l \in \{1, \ldots, n_d\}$) should undergo mutation according to the outcome of a sample of a uniform random variable $u \sim U(0,1)$:

$$d'_l = \begin{cases} d_l & , \quad u_l > p' \\ \widetilde{X}_l & , \quad u_l \leq p' \end{cases} \tag{12}$$

$\widetilde{X}$ indicates the outcome of sampling a random variable with an appropriate distribution over the discrete space, e.g., a distribution with maximum entropy. In the following, a uniform distribution over the set of possible values is assumed. This decision is justified because in case of the application to optical multilayer systems a semantically reasonable metrics does not exist on the discrete subspace $\{0, \ldots, n-1\} \subset \mathbb{Z}$ which is used to represent the available set of choices for discrete variables. In other words, all possible values are treated equally and it is not possible to define a distance or similarity measure between them. This equality of choices is taken into account by using a uniform distribution over $\{0, \ldots, n-1\}$. This argument also excludes the self-adaptation of step sizes of discrete probability distributions as introduced by Rudolph (1994) for application problems where the natural metrics on $\mathbb{Z}$ are meaningful.

To conclude this section, a notation for the generalized evolution strategy is defined that summarizes the most important parameters of the algorithm in the abbreviation

$$\mathrm{ES}(n_x, n_d, n_\sigma, n_p, r_{\{r_x r_d r_\sigma r_p\}}, s_{(\mu,\lambda)}) \quad .$$

This notation is used in the following section to describe the ES-variants applied to optical multilayer design problems.

3 SYNTHESIS OF OPTICAL MULTILAYER SYSTEMS

Optical multilayer coatings are of remarkable importance in a number of application fields such as optical and scientific instrumentation manufacturing, spectroscopy, medicine, and astronomy. An optical multilayer coating consists of a set of plane parallel isotropic layers separating two homogenous isotropic media (often these are air and a substrate). The permittivity and conductivity of the layers depend on one spatial coordinate perpendicular to the layer-media boundaries. A plane electromagnetic wave enters from the first medium (air) into the multilayer system and is partially reflected or transmitted at the borders between layers of different refractive indices.

The behavior of a multilayer system can be characterized by the *spectral reflectance profile* at the substrate, i.e., the dependence of reflectances on the wavelength. The spectral reflectance profile depends mainly on the number n of layers, their refractive indices $\vec{\eta} = (\eta_1, \ldots, \eta_n)$, and the thicknesses $\vec{d} = (d_1, \ldots, d_n)$ of the n layers. Given these parameters, several methods

are known to calculate the spectral reflectance profile of a multilayer system (see Furman and Tikhonravov 1992, chapter 1).

For the design of optical multilayer coatings, however, the inverse problem has to be solved: Given a desired spectral reflectance profile, find the parameters of an optimal multilayer coating that approximates the desired behavior as well as possible. To measure the difference between the target reflectance $\tilde{R}(\lambda)$ (λ denotes the wavelength) and the actual coating reflectance $R(\vec{d}, \vec{\eta}, \lambda)$ within the spectral band $[\lambda_d, \lambda_u]$ of interest, a merit function

$$f(\vec{d}, \vec{\eta}) = \int_{\lambda_d}^{\lambda_u} \left[R(\vec{d}, \vec{\eta}, \lambda) - \tilde{R}(\lambda) \right]^2 d\lambda \tag{13}$$

based on mean square estimation is used (Furman and Tikhonravov 1992, p. 106). Then, the design of optical multilayer coatings consists in solving the highly multimodal optimization problem of minimizing $f(\vec{d}, \vec{\eta})$. In practice, an approximation of equation (13) over a discrete number m of (usually equidistant) wavelength values λ_i ($i \in \{1, \ldots, m\}$) is used:

$$f(\vec{d}, \vec{\eta}) = \sum_{i=1}^{m} \left[R(\vec{d}, \vec{\eta}, \lambda_i) - \tilde{R}(\lambda_i) \right]^2 \quad . \tag{14}$$

Presently, *refinement* and *synthesis* methods for the design problem are distinguished. Refinement methods work by modifying the construction parameters of a given starting design, while synthesis methods do not require the input of a starting design. They generate a multilayer coating by themselves and often use a refinement method to improve its performance. Because the choice of a starting design is a time-consuming, difficult problem (especially for the design of antireflection coatings, neutral beam splitters and multilayer systems featuring non-standard spectral properties), the development of good synthesis methods is an important topic of research.

Normally, local optimization methods for the refinement of starting designs are indicated to yield (more or less) reasonable results (Furman and Tikhonravov 1992, p. 108). A large number of such refinement methods, including adaptive random search, damped least squares, a modified gradient method, golden section, the Hooke-Jeeves method, Powell's conjugate search, Rosenbrock's rotating coordinates, generalized simulated annealing, Monte Carlo simulated annealing, and a revised Nelder-Mead simplex method are compared by Dobrowolski and Kemp (1990). On three different problems, they were not able to identify a single best method, but recommended the choice of the Hooke-Jeeves or damped least squares method. The good performance of the Hooke-Jeeves algorithm corresponds well with the results reported in the reference article by Aguilera et al. (1988) that compares refinement software packages as used by nine different researchers. They define a reference problem which is also optimized by Dobrowolski and Kemp (1990) and therefore forms a good test case for the extended evolution strategy as described in the previous section.

First results concerning the application of *genetic algorithms* as synthesis methods have been published recently (Eisenhammer, Lazarov, Leutbecher, Schöffel, and Sizmann 1993; Martin, Brunet-Bruneau, Rivory, and

Schoenauer 1994). Genetic algorithms are evolutionary algorithms which were originally defined to operate on fixed-length binary strings with an emphasis on recombination and probabilistic selection and without the capability of self-adapting strategy parameters (Goldberg 1989; Holland 1975).

Eisenhammer et al. (1993) use a genetic algorithm with traditional binary encoding for the $2n$ parameters of silver-based heat mirrors. With 8-bit resolution per layer thickness (restricted to the range 20 – 530 μm and, due to the number of bits, a grid resolution of 2 nm) and 3-bit per layer material (sufficient to encode eight different materials), they synthesize a five-layer system and demonstrate the utility of genetic algorithms to generate a starting design "from scratch."

Martin et al. (1994) also refer to their evolutionary method as a genetic algorithm, but they work directly on a real-valued representation of layer thicknesses and refractive indices. In addition to proportional selection from genetic algorithms, they use a recombination operator similar to generalized intermediate recombination in evolution strategies and a mutation operator which adds normally distributed perturbations with expectation zero and a fixed standard deviation to the object variables. Refractive indices are forced to feasible values by rounding them to the nearest admissable value. After running the algorithm for 1600 generations, they apply a gradient refinement method to the resulting design.

Both articles demonstrate the robustness of evolutionary algorithms to find a solution even for quite difficult problems, if no starting design is given. The algorithms, however, are not sufficiently general to cover the properties of the synthesis problem, which — for reasons as outlined in the introduction — requires an algorithm for mixed-integer optimization.

The Reference Problem

Aguilera et al. (1988) proposed the design of an antireflection coating of germanium (Ge, $\eta = 4.2$) and zinc sulfide (ZnS, $\eta = 2.2$) in the wave length region 7.7 – 12.3 μm. Within this range, $m = 47$ wavelength values are defined according to $\lambda_1 = 7.7\mu$m and ($i \in \{1, \ldots, m\}$):

$$\lambda_i = \lambda_1 + \frac{i-1}{m-1} \cdot (12.3 - 7.7) \quad [\mu\text{m}] \quad . \tag{15}$$

The objective function is given by the *reflection mean square* (RMS) function, a slight modification of equation (14):

$$M(\vec{d}, \vec{\eta}) \quad = \quad \left(\frac{\displaystyle\sum_{i=1}^{m} v(\lambda_i) \left[R(\vec{d}, \vec{\eta}, \lambda_i) - \tilde{R}(\lambda_i) \right]^2}{\displaystyle\sum_{i=1}^{m} v(\lambda_i)} \right)^{1/2} \quad , \tag{16}$$

where all weighting factors $v(\lambda_i) = 1$ and the target reflectance $\tilde{R}(\lambda_i) = 0$ ($i \in \{1, \ldots, m\}$), i.e.,

$$M(\vec{d}, \vec{\eta}) = \sqrt{\frac{1}{m} \sum_{i=1}^{m} R(\vec{d}, \vec{\eta}, \lambda_i)^2} \; \rightarrow \; \min \quad . \tag{17}$$

According to the *matrix-method*, the reflection $R(\vec{d},\vec{\eta},\lambda_i)$ of a multilayer system at wavelength λ_i is given by (see Furman and Tikhonravov 1992, pp. 21–26):

$$R(\vec{d},\vec{\eta},\lambda_i) = |\eta_0 B - C|^2 \cdot D^{-1} \quad (18)$$

$$D = |\eta_0 B + C|^2 \quad (19)$$

where η_0 denotes the refractive index of the entrance medium air ($\eta_0 = 1$). B and C are obtained as follows:

$$\begin{pmatrix} B \\ C \end{pmatrix} = \mathbf{M}(\vec{d},\vec{\eta},\lambda_i) \begin{pmatrix} 1 \\ \eta_s \end{pmatrix} , \quad (20)$$

where η_s denotes the refractive index of the substrate germanium ($\eta_s = 4$) and

$$\mathbf{M}(\vec{d},\vec{\eta},\lambda_i) = \prod_{j=1}^{n} \begin{pmatrix} \cos\phi_{ij} & i\eta_j^{-1}\sin\phi_{ij} \\ i\eta_j \sin\phi_{ij} & \cos\phi_{ij} \end{pmatrix} \quad (21)$$

$$\phi_{ij} = \frac{2\pi\eta_j d_j}{\lambda_i} . \quad (22)$$

In order to get an impression of the topological characteristics of the objective function defined by equations (17)–(22), a three-dimensional plot of RMS-values for a two-layer filter with $\eta_1 = 2.2$, $\eta_2 = 4.2$, and d_1, d_2 varying in the range 0–20 μm is shown in figure 1. The landscape is characterized by

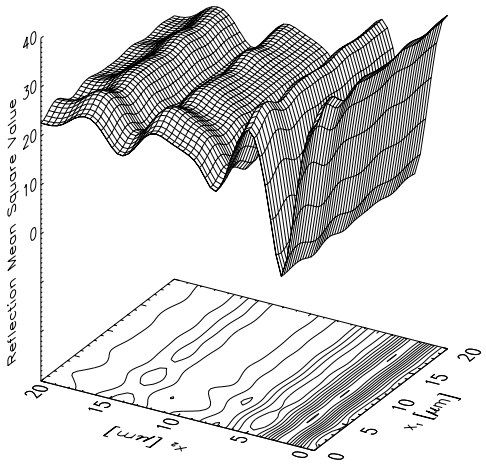

Figure 1: Topology of the RMS merit function in case of a fixed two-layer filter structure with $\eta_1 = 2.2$ and $\eta_2 = 4.2$. Optical thicknesses are varied in the range $0 - 20\,\mu$m.

parallel "waves," separated by valleys of increasing depth and decreasing width. It is obvious that optimization algorithms may stagnate within a suboptimal valley without ever surmounting the next wave and finding a better but narrower valley. Even within a valley, the landscape is characterized by local optima of different quality, located at the valley's bottom, and the algorithm would have to adapt to the valley direction in order to search for a better optimum within the valley. Evidently, the optimization of the RMS-function of an n-dimensional filter is an extremely complex, multimodal optimization problem, especially when no problem-specific knowledge — e.g., a starting design — is provided to the algorithm.

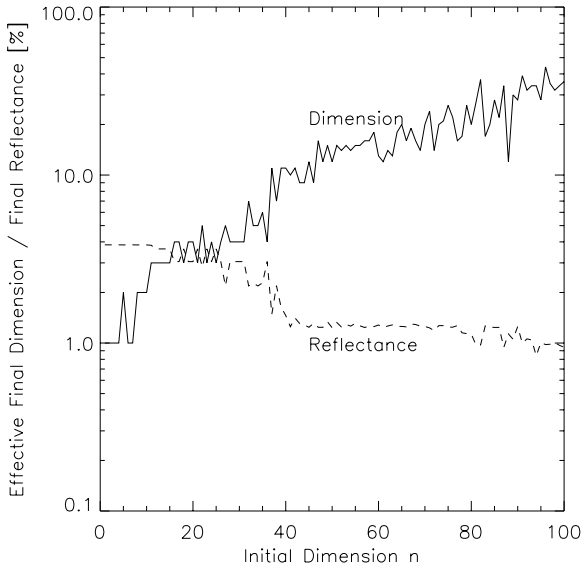

Figure 2: Final best objective function value and final effective filter dimension over the initial filter dimension for an ES$(n, n, n, 1, r_{IDII}, s_{(8,50)})$.

Experimental Results

The design of the evolution strategy follows the general principle that the algorithm is not tailored to the optimization of optical multilayer coatings but can be used for arbitrary mixed-integer optimization problems. Consequently, the algorithm presently does not incorporate the straightforward idea of combining adjacent layers of the same refractive index into a single layer. This operation is performed only at the end of an optimization run in order to determine the real structure of the filter. To get an impression of the dependence of the effective number of layers (after performing the final unification of adjacent layers of identical material) on the dimension $n = n_x = n_d$ which was used for initialization of the algorithm, single optimization runs with the reference problem were performed for initial dimensions in the range $n \in \{1, \ldots, 100\}$. An (8,50)-ES with global

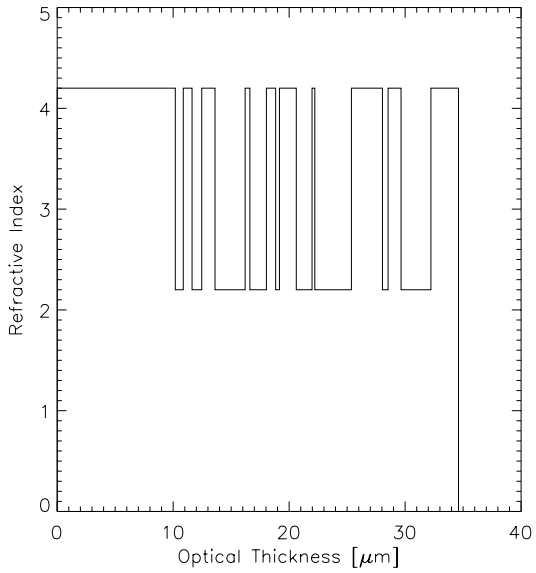

Figure 3: Refractive-index profile of the best antireflection coating found by the extended evolution strategy. The refractive indices of the layers are plotted over the corresponding optical thicknesses. The plot can be interpreted as a cut through the filter.

discrete recombination on integer variables and global intermediate recombination on all further variables (an ES$(n, n, n, 1, r_{IDII}, s_{(8,50)})$) was run for 3000 generations, and the final best objective function value as well as the final effective filter dimension are plotted as a function of the initial filter dimension n in figure 2.

The results clearly demonstrate the improvement of the final reflectance value with growing initial filter dimension n, corresponding to a growing effective filter dimension. In comparison to the initial filter dimension, the effective filter dimension reduces to about one fourth during the optimization runs.

The largest improvement of reflectance values is observed for effective filter dimensions in the range 4–10. Beyond an effective dimension of 10, reflectance values in the range 0.94–1.63% are consistently found by the evolution strategy, and almost all values are better than 1.35%, the second best value found by refinement methods for a predefined starting design (Aguilera et al. 1988). The best value reported by Aguilera et al. (1988) also achieves a quality of 0.94% as obtained by the evolution strategy. These results indicate the robustness of an evolution strategy as a synthesis method.

A further experimental test was performed to investigate the impact of the values of learning rates $\tau \propto 1/(\sqrt{2\sqrt{n_x}})$ and $\gamma \propto 1/(\sqrt{2\sqrt{n_d}})$ on the final reflectance values. The test was based on an ES$(70, 70, 70, 1, r_{IDII}, s_{(15,100)})$, running for 3000 generations for each single value of τ and γ (varying in the range 0.01–0.5 in steps of 0.01). An initial dimension $n_x = n_d = 70$

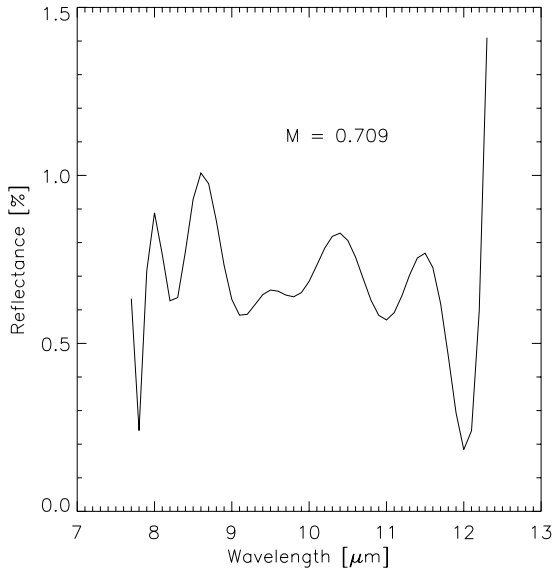

Figure 4: Spectral reflectance profile of the best antireflection coating found by the extended evolution strategy. The reflectances in % are plotted over the wavelength. This elucidates the characteristics of the filter over the wavelength.

was chosen because the best result reported by Aquilera et al. (1988) was achieved for a starting design with 69 layers. The experimental results confirm the choice of τ, while the test runs suggest a correction of γ by a factor of 3/5 (see Schütz 1994, pp. 118–120).

Using these parameter values for τ and γ and the evolution strategy parameters as indicated above, a single run of 25,000 generations yields a final reflectance value of 0.709% for a filter with 20 effective layers. The refractive-index profile of this filter is shown in figure 3; its spectral reflectance profile is shown in figure 4. Column A of table 1 specifies the exact filter structure in comparison with the starting design given by Aquilera et al. (1988) (column 1B) and the overall best results reported there (column 3F), which was obtained from the special starting design with 69 layers. Taking into account that the evolution strategy finds the solution of quality 0.709% without requiring an initial design, which is normally well thought out, the quality of this solution is remarkable.

When the evolution strategy is applied with fixed dimension to the starting design given in column 1B of table 1, a result of quality 1.287% as reported in column B of table 2 is obtained (this value reflects the best result found within 10 runs). The corresponding coating consists of only 13 effective layers and still yields a reasonable and useful reflectance value (which ranks second-best when compared to the results given in (Aguilera et al. 1988), table II).

	1B Starting		A ES		3F (Tab. III) (Aguilera et al. 1988)	
Layer	η	ηd	η	ηd	η	ηd
Substrate	4.0		4.0		4.0	
1	2.2	0.7150	4.2	9.7505	2.2	0.1753
2	4.2	0.7675	2.2	0.4323	4.2	5.7821
3	2.2	0.7300	4.2	0.6883	2.2	1.3845
4	4.2	0.7825	2.2	0.7574	4.2	0.2953
5	2.2	0.7300	4.2	0.8351	2.2	2.7938
6	4.2	0.7675	2.2	1.1462	4.2	2.8657
7	2.2	0.7150	4.2	2.5962	2.2	0.3023
8	4.2	0.7675	2.2	0.4108	4.2	1.3117
9	2.2	1.1000	4.2	1.4234	2.2	2.7810
10	4.2	0.1450	2.2	0.7966	4.2	1.1075
11	2.2	1.1000	4.2	0.3279	2.2	0.4671
12	4.2	0.3050	2.2	1.4493	4.2	2.7636
13	2.2	1.6450	4.2	1.3594	2.2	2.7489
14	4.2	0.3050	2.2	0.2400	4.2	0.4351
15	2.2	1.1000	4.2	3.1560	2.2	1.1792
16	4.2	0.1450	2.2	2.6750	4.2	2.5351
17	2.2	4.3250	4.2	0.4876	2.2	2.3890
18	4.2	0.5825	2.2	1.1183		
19	2.2	6.0125	4.2	2.5736		
20	4.2	1.3875	2.2	2.3804		
21	2.2	2.7975				
Air	1.0		1.0		1.0	
$\sum \eta d\,[\mu\text{m}]$	26.93		34.60		31.32	
Reflection %	10.6		0.709		0.66	

Table 1: A comparison of different multilayer filter solutions found.

Comparison with a Genetic Algorithm

To assess its qualities on a test problem for which results from an other evolutionary heuristic are known, the generalized evolution strategy is compared in this section with the "genetic algorithm" proposed by Martin et al. (1994) (for a brief discussion of their algorithm see section 3).

The problem presented by Martin et al. (1994) consists in designing an antireflection coating in the region of $500 - 1000$nm. For this problem, the available materials are MgF_2, ZnS, SiO_2, Si_3N_4 and Al_2O_3, the incident medium is air ($\eta_0 = 1.0$), and the substrate is glass ($\eta_s = 1.5$). To take the dispersiveness of the materials into consideration, a *Cauchy law* is assumed to calculate the refractive index

$$\eta(\lambda) = A + \frac{B}{\lambda^2} + \frac{C}{\lambda^4}$$

for each material (the material-dependent values of the constants A, B, and C are given in table 2 of (Martin et al. 1994)). As the materials are only

	B ES		C ES	
Layer	η	ηd	η	ηd
Substrate	4.0		1.5	
1	2.2	0.6381	1.378	0.044717
2	4.2	1.1226	1.606	0.040172
3	2.2	1.7249	1.378	0.103689
4	4.2	1.1539	1.448	0.033130
5	2.2	0.7761	1.606	0.005947
6	4.2	3.3985	2.255	0.000520
7	2.2	0.3484	1.606	0.039084
8	4.2	5.9490	1.378	0.075274
9	2.2	5.4071	1.606	0.421092
10	4.2	0.8947	1.378	0.035882
11	2.2	5.7535	1.606	0.428359
12	4.2	1.1886	1.378	0.082064
13	2.2	2.8530	1.606	0.231126
14			2.255	0.106371
15			1.606	0.048374
16			2.255	0.006470
17			2.0	0.046954
18			2.255	0.064731
19			1.606	0.000515
20			1.378	0.185529
Air	1.0		1.0	
$\sum \eta d\,[\mu\text{m}]$	31.22		2.0	
Reflection %	1.287		0.163	

Table 2: A comparison of different multilayer filter solutions found.

slightly dispersive, however, it is possible to define an average refractive index for each material, a method which yields the values $\eta_{MgF_2} = 1.378$, $\eta_{Al_2O_3} = 1.606$, $\eta_{ZnS} = 2.255$, $\eta_{SiO_2} = 1.448$, and $\eta_{Si_3N_4} = 2.0$. In contrast to Martin et al. (1994) where the wavelength dependent refractive index values are used, the results presented here are obtained on the basis of the averaged index values as given above. The numerical difference between both methods is negligible.

To formulate the objective function as a minimization problem (in contrast to Martin et al. (1994) in which the inverse of the sum of squared reflectance values is maximized), the negative sum over a total of $m = 26$ equidistant wavelengths λ_i is minimized here:

$$f(\vec{d}, \vec{\eta}, n) = -\left(\sum_{i=1}^{m} R(\vec{d}, \vec{\eta}, \lambda_i)^2 \right)^{-1} \rightarrow \min \quad . \tag{23}$$

Martin et al. (1994) synthesized a 40-layer system of 2000 nm ($= 2\mu$m) total optical thickness. This can be achieved by scaling (multiplying) each

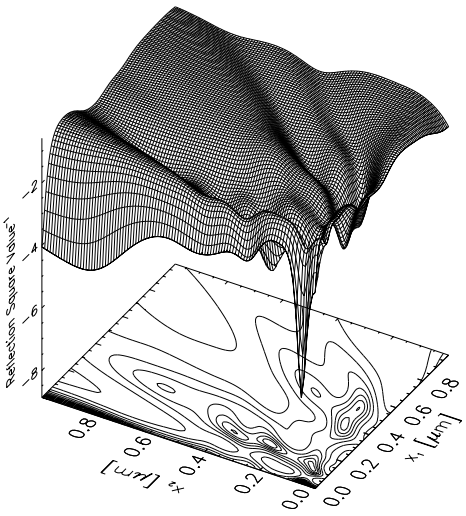

Figure 5: Topology of the merit function in case of a fixed five-layer filter structure with $\eta_1 = \eta_0 = 1.0$, $\eta_2 = 1.606$, $d_2 = 100$nm, $\eta_3 = 2.255$, $d_3 = 100$nm, $\eta_4 = 1.448$, $d_4 = 100$nm and $\eta_5 = \eta_s = 2.0$. Optical thicknesses of the first and fifth layer are varied in the range $0 - 1\mu$m.

thickness by the constant

$$c = \frac{2000}{\vec{d}\vec{\eta}^t} \quad .$$

Figure 5 displays the topology of the objective function of a five dimensional system. While the layers two, three, and four are fixed ($\eta_2 = 1.606$, $d_2 = 100$nm, $\eta_3 = 2.255$, $d_3 = 100$nm, $\eta_4 = 1.448$, $d_4 = 100$nm), the thicknesses of layer one ($\eta_1 = \eta_0 = 1.0$) and layer five ($\eta_5 = \eta_s = 2.0$) are varied in the range $0 - 1000$nm. In contrast to the previous reference problem a distinct, sharp global minimum can be identified in the plot of this simplified version of the merit function.

Again, the generalized evolution strategy yields remarkably good results: Similar to the GA offered in Martin et al. (1994), the GES identifies an antireflection coating with an average reflectance of 0.2% after 1600 generations (i.e., the averaged reflection over all wavelengths amounts to 0.2%). After executing the algorithm for an additional 3400 generations the average reflectance of the system decreases to 0.163% with 20 layers (the spectral profile of the filter resulting after these 5000 generations is presented in figure 6). This value is nearly as good as the 0.15% system (with 29 layers) generated by the refinement method used by Martin et al. to get a better performance for the coating found by their genetic algorithm.

The corresponding refractive index profile is shown in figure 7, and column C of table 2 lists the structure of the system.

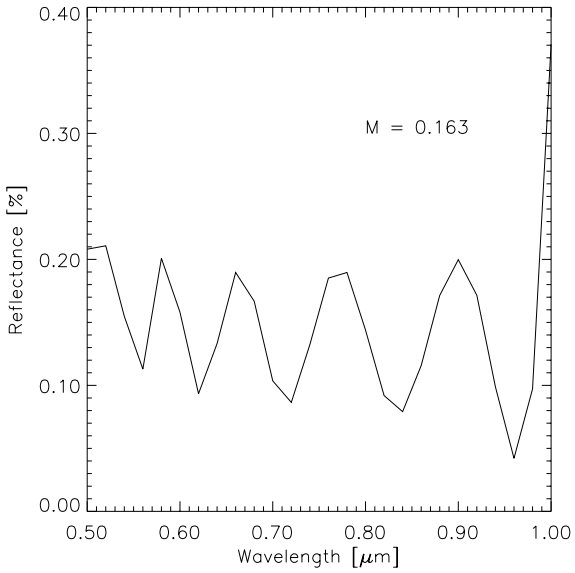

Figure 6: Spectral profile of the best antireflection coating found after 5000 generations.

4 CONCLUSIONS

The results presented for the design of optical multilayer coatings demonstrate that the generalized evolution strategy is a useful new algorithm for the synthesis of such systems. For a reference problem by Aguilera et al. (1988), the results obtained by the evolution strategy ranked second-best in comparison with refinement methods using a starting design, and for a reference problem optimized by Martin et al. (1994), with an evolutionary heuristic on the basis of a real-valued genetic algorithm, the evolution strategy achieved comparable results while not needing a refinement method to improve its solutions.

Although these results are encouraging, the algorithm still offers some obvious directions for further investigations, especially regarding the following questions:

- The test runs indicated just *one* mutation probability to yield best performance of the algorithm. This mutation probability decreases relatively quickly, such that the search converges towards an optimum in the discrete subspace faster than in the continuous subspace. It is important to investigate the conditions for self-adaptation of more than one mutation probability, e.g., by means of increasing the selective pressure.

- The dimension of the objective function might also be variable for optimization problems such as the optical multilayer design. Operators such as *gene deletion* and *gene duplication* are helpful to incorporate this

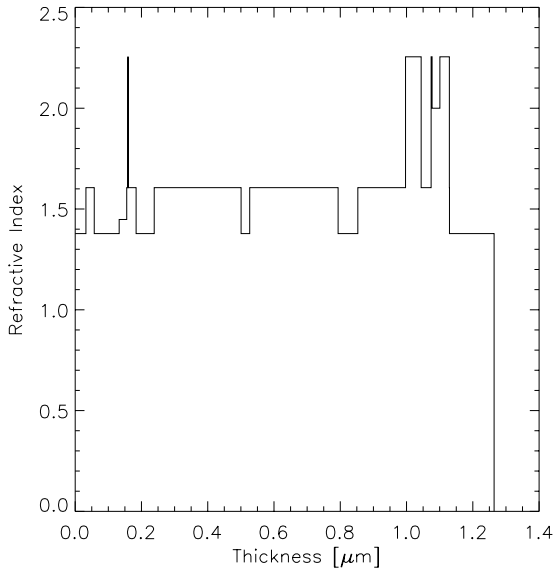

Figure 7: Refractive index profile of the best antireflection coating found after 5000 generations. The profile is shown from the substrate (left) to air (right).

into the generalized evolution strategy. The design of the recombination operator, however, forms a major problem in this case (Schütz 1994).

- As for all synthesis methods, the generalized evolution strategy can be combined with refinement methods that perform local optimization of the final design or of intermediate solutions obtained during a run.
- Correlated mutations (see e.g. Bäck and Hoffmeister 1994) are not tested within the generalized evolution strategy so far. It is expected, however, that they are able to yield a further improvement of the results because correlations might allow for a learning of the direction of valleys. Consequently, with correlated mutations the algorithm might be able to find better optima within such a valley (cf. figure 1).

It is important to recall that the generalized evolution strategy is designed as a general-purpose algorithm for mixed-integer optimization which does not rely on any knowledge which is problem-specific for optical coating design. Renunciating such knowledge, the algorithm still yields solutions of remarkable quality and promises successful applicability to a large number of important industrial problems.

Acknowledgements

The authors gratefully acknowledge support by the BMFT–project EVOALG, grant 01 IB 403 A (Thomas Bäck) and the DFG, grant Schw 361/5-2

(Martin Schütz). We would also like to thank David B. Fogel for his valuable comments on a draft version of this paper and Horst Greiner for helpful technical discussions about optical multilayer systems.

References

Aguilera, J. A., J. Aguilera, P. Baumeister, A. Bloom, D. Coursen, J. A. Dobrowolski, F. T. Goldstein, D. E. Gustafson, and R. A. Kemp (1988). Antireflection coatings for germanium IR optics: A comparison of numerical design solutions. *Applied Optics 27*(14), 2832–2840.

Bäck, T. and F. Hoffmeister (1994). Basic aspects of evolution strategies. *Statistics and Computing 4*(2), 51–63.

Bäck, T. and H.-P. Schwefel (1993). An overview of evolutionary algorithms for parameter optimization. *Evolutionary Computation 1*(1), 1–23.

Dobrowolski, J. A. and R. A. Kemp (1990). Refinement of optical multilayer systems with different optimization procedures. *Applied Optics 29*(19), 2876–2893.

Eisenhammer, T., M. Lazarov, M. Leutbecher, U. Schöffel, and R. Sizmann (1993). Optimization of interference filters with genetic algorithms applied to silver-based heat mirrors. *Applied Optics 32*(31), 6310–6315.

Furman, S. A. and A. V. Tikhonravov (1992). *Basics of Optics of Multilayer Systems*. Basics of. Gif-sur-Yvette Cedex – France: Editions Frontières.

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

Hoffmeister, F. and T. Bäck (1992). Genetic self–learning. In *Proceedings of the First European Conference on Artificial Life*, Eds. F. J. Varela and P. Bourgine, pp. 227–235. Cambridge, MA: The MIT Press.

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

Kelahan, R. C. and J. L. Gaddy (1978). Application of the adaptive random search to discrete and mixed integer optimization. *International Journal for Numerical Methods in Engineering 12*, 289–298.

Klockgether, J. and H.-P. Schwefel (1970). Two–phase nozzle and hollow core jet experiments. In *Proc. 11th Symp. Engineering Aspects of Magnetohydrodynamics*, Ed. D. Elliott, pp. 141–148. Pasadena, CA: California Institute of Technology.

Lichtfuss, H. J. (1965). Evolution eines Rohrkrümmers. Diplomarbeit, Technische Universität Berlin.

Lohmann, R. (1992). Structure evolution in neural systems. In *Dynamic, Genetic, and Chaotic Programming*, Eds. B. Soucek and the IRIS Group, pp. 395–411. New York: Wiley.

Martin, S., A. Brunet-Bruneau, J. Rivory, and M. Schoenauer (1994). Darwinian evolution of homogenous multilayer systems: A new method for optical coatings design. In *Proceedings of the International Symposium on Optical Interference Coatings*.

Martin, S., J. Rivory, and M. Schoenauer (1994). Simulated Darwinian evolution of homogenous multilayer systems: a new method for optical coatings design. *Optics Communications 110*, 503–506.

Obalek, J. (1994). Rekombinationsoperatoren für Evolutionsstrategien. Diplomarbeit, Universität Dortmund, Fachbereich Informatik.

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

Rudolph, G. (1992). On correlated mutations in evolution strategies. In *Parallel Problem Solving from Nature 2*, Eds. R. Männer and B. Manderick, pp. 105–114. Amsterdam: Elsevier.

Rudolph, G. (1994). An evolutionary algorithm for integer programming. In (Eds.), *Parallel Problem Solving from Nature — PPSN III International Conference on Evolutionary Computation*, Eds. Y. Davidor, H.-P. Schwefel, and R. Männer, Volume 866 of *Lecture Notes in Computer Science*. pp. 139–148. Berlin: Springer.

Schütz, M. (1994). Eine Evolutionsstrategie für gemischt-ganzzahlige Optimierungsprobleme mit variabler Dimension. Diplomarbeit, Universität Dortmund, Fachbereich Informatik.

Schwefel, H.-P. (1968). Projekt MHD-Staustrahlrohr: Experimentelle Optimierung einer Zweiphasendüse, Teil I. Technischer Bericht 11.034/68, 35, Berlin: AEG Forschungsinstitut.

Schwefel, H.-P. (1981). *Numerical Optimization of Computer Models*. Chichester: Wiley.

Schwefel, H.-P. (1992). Imitating evolution: Collective, two-level learning processes. In U. Witt (Ed.), *Explaining Process and Change — Approaches to Evolutionary Economics*, pp. 49–63. Ann Arbor, MI: The University of Michigan Press.