

---

# Design and Analysis of Schemes for Adapting Migration Intervals in Parallel Evolutionary Algorithms

**Andrea Mambrini**

Department of Computer Science, University of Sheffield, Sheffield S1 4DP, UK

**Dirk Sudholt**

Department of Computer Science, University of Sheffield, Sheffield S1 4DP, UK

doi:10.1162/EVCO\_a\_00153

---

## Abstract

The migration interval is one of the fundamental parameters governing the dynamic behaviour of island models. Yet, there is little understanding on how this parameter affects performance, and how to optimally set it given a problem in hand. We propose schemes for adapting the migration interval according to whether fitness improvements have been found. As long as no improvement is found, the migration interval is increased to minimise communication. Once the best fitness has improved, the migration interval is decreased to spread new best solutions more quickly. We provide a method for obtaining upper bounds on the expected running time and the communication effort, defined as the expected number of migrants sent. Example applications of this method to common example functions show that our adaptive schemes are able to compete with, or even outperform, the optimal fixed choice of the migration interval, with regard to running time and communication effort.

## Keywords

Parallel evolutionary algorithms, island model, migration interval, runtime analysis, theory.

## 1 Introduction

Evolutionary algorithms (EAs) have given rise to many parallel variants (Luque and Alba, 2011; Tomassini, 2005) fuelled by the rapidly increasing number of CPU cores and the ready availability of computation power through GPUs and cloud computing. Parallelization provides a cost-effective approach to solving problems in real time and for tackling large-scale problems.

There are many variants of parallel evolutionary algorithms, from parallelising function evaluations on multiple processors to fine-grained models such as cellular evolutionary algorithms and coarse-grained models such as island models (Tomassini, 2005; Luque and Alba, 2011). In the latter approach, multiple populations evolve independently for a certain period of time. Every  $\tau$  generations, for some fixed parameter  $\tau$  called *migration interval*, individuals migrate between these islands to coordinate searches on the island. Communication takes place according to a spatial structure, a topology connecting populations. Common topologies include rings, two-dimensional grids or toroids, hypercubes, and complete graphs with all possible connections.

Island models are popular optimisers for several reasons:

- Multiple communicating populations can make the same progress as a single population in a fraction of the time, speeding up computation.
- Smaller populations can be simulated faster than large populations, effectively reducing the execution time on each processor (Alba, 2002).
- Periodic communication only requires small bandwidth if the migration interval is not very small, leading to low communication costs.
- Solution quality is improved as different populations can explore different regions of the search space.

The usefulness of parallel populations has been demonstrated in thousands of successful applications ranging from language tagging, circuit design, scheduling and planning to bioinformatics (Luque and Alba, 2011; Alba et al., 2013).

However, designing an effective parallel evolutionary algorithm can be challenging, as the method and amount of communication need to be tuned carefully. Too frequent communication leads to high communication costs, and it can compromise exploration. Too little communication means that the populations become too isolated and unable to coordinate their searches effectively. There is agreement that even the effect of the most fundamental parameters on performance is not well understood (Luque and Alba, 2011; Alba et al., 2013).

We make a contribution toward finding good values for the migration interval, the parameter describing the frequency of migration. We propose adaptive schemes that adjust the migration interval depending on whether islands have managed to find improvements during the last migration interval or not. The goal is to reduce communication while not compromising the exploitation of good solutions. The main idea of our schemes is that if an island has managed to improve its current best fitness, migration should be intensified to spread this solution to other islands. Otherwise, islands decrease the frequency of migration to avoid large communication costs.

Two different adaptive schemes are proposed, inspired by previous work (Lässig and Sudholt, 2011a). In both of them islands have individual migration intervals that are adapted throughout the run. In Scheme A if an island has not improved its current best fitness during the last migration interval, its migration interval is doubled. Once an improvement is found, the migration interval is set to 1 to communicate this new solution quickly. In Scheme B an island also doubles the migration interval when no improvement was found, while when an improvement is found, it halves  $\tau$  at the end of the current migration interval.

We show that doubling the migration interval guarantees for elitist EAs that the number of migrations from an island is logarithmic in the time this island spends on a certain fitness level, for any value of the current best fitness.

We contribute a rigorous analytical framework that yields upper bounds on the expected optimisation time and the expected communication effort, defined as the total number of migrants sent. This is described for fixed migration intervals in Section 3, Scheme A in Section 4, and Scheme B in Section 5. This framework can be applied to a range of evolutionary algorithms; we demonstrate its application for a simple island model called parallel (1+1) EA (Lässig and Sudholt, 2014b). Our adaptive schemes are then compared in Section 6 against the best fixed values of the migration interval for classical test problems. The results reveal that our adaptive schemes are able to match

or even outperform the best fixed migration intervals with regard to upper bounds on the expected parallel time and the expected communication effort.

Note that our methodology only provides upper bounds on the expected parallel times and expected communication efforts. Hence performance comparisons between migration schemes or topologies are based on comparisons of upper bounds. A better upper bound translates to a better expectation in case the respective upper bounds are tight, in an asymptotic sense. For the application to the parallel (1+1) EA and two of our test problems, OneMax and LO, we know from general lower bounds for all mutation-based evolutionary algorithms<sup>1</sup> (Sudholt, 2013) that all bounds for expected parallel times are tight. For fixed migration intervals also, the stated bounds on the expected communication effort are tight (see Theorem 1 in Section 3). For other problems or problem classes, we can only compare upper bounds.

This paper is based on an extended abstract published at GECCO 2014 (Mambrini and Sudholt, 2014), where some proofs were omitted. The present manuscript contains all proofs and several extensions, most notably a discussion in Section 7 about the balance between exploration and exploitation in the light of our adaptive schemes.

## 1.1 Related Work

This paper is in line with recent theoretical research on the running time of parallel EAs. Lässig and Sudholt (2010) presented a method for analysing speedups in island models, with applications to a range of combinatorial problems (Lässig and Sudholt, 2011b, 2014a). Neumann et al. (2011) considered the benefit of using crossover during migration for artificial problems and instances of the vertex cover problem. Mambrini et al. (2012) studied the running time and communication effort of homogeneous and heterogeneous island models for finding good solutions for the NP-hard set cover problem.

Different migration policies were compared by Araujo and Merelo Guervós (2011). Bravo et al. (2012) studied the effect of the migration interval when tackling dynamic optimization problems.

Skolicki and De Jong (2005) investigated the impact of the migration interval and the number of migrants on performance. They found that the dynamic behaviour of the algorithm is not just the result of the number of exchanged individuals, but it results from several phenomena. For frequent migrations the effect of varying the migration interval are much stronger than that of varying the number of migrants. Performance degrades when the number of migrants approaches the population size of islands. And performance may degrade in the presence of large migration intervals if the algorithm stops prematurely.

Alba and Luque (2005) analysed growth curves and takeover times based on migration intervals and migration topologies, showing how quickly good solutions spread in an island model. Theoretical analyses of this takeover time were presented by Rudolph (2006).

Lässig and Sudholt (2013) presented a theoretical analysis and a problem where island models excel over both panmictic populations as well as independent runs. This requires a delicate choice of the migration interval, and performance degrades

---

<sup>1</sup>The class of mutation-based evolutionary algorithms describes all algorithms starting with a population of individuals picked uniformly at random, and afterwards only using standard bit mutation as variation operator. The parallel (1+1) EA considered in Section 6 fits in this framework, regardless of the topology and migration policy used, as standard bit mutation is the only variation operator.

drastically when suboptimal parameter values are being used. This again emphasises the importance of this parameter for the performance of island models.

Hong et al. (2007) and Lin et al. (2012) presented a fitness-based adaptive migration scheme. Each island compares its increase of its best fitness over the last migration interval with the same quantity from the migration interval before. If the new fitness increase is larger than the old one, the migration interval is increased by some constant value. Otherwise, it is decreased by the same constant value. A preliminary experimental study on random 0/1 knapsack problem instances showed that the adaptive scheme can lead to a better final fitness (Hong et al., 2007). The difference in our scheme is that we adapt the migration interval in opposite directions: we decrease the migration interval in case of good fitness gains and increase it otherwise. Furthermore, our changes to the migration interval are more drastic than their additive changes.

Osorio et al. (2011; 2013) presented adaptive schemes for the migration interval, which aim for convergence at the end of the run (for runs of fixed length). The migration interval is set according to growth curves of good individuals and the remaining number of generations; migration intensifies toward the end of a run. They obtained competitive performance results, compared to optimal fixed parameters, for MAX-SAT instances (Osorio et al., 2013). Our perspective is different, as we do not optimise for fixed-length runs.

Finally, Lässig and Sudholt (2011a) presented schemes for adapting the number of islands during a run of an island model. The same schemes also apply to offspring populations in a  $(1+\lambda)$  EA as a special case. Scheme A doubles the number of islands in case no improvement has been found in one generation. Otherwise, the number of islands is reduced to one island. Scheme B also doubles the number of islands when no improvement is found, and halves it otherwise. Both schemes achieve optimal or near-optimal parallel running times while not increasing the total number of function evaluations by more than a constant factor. Our schemes for adapting the migration interval are inspired by this work.

## 2 Preliminaries

We define the parallel EAs considered in this work, which contain our adaptive schemes. Our analytical framework is applicable to all elitist EAs: EAs that do not lose their current best solution. We define our schemes for maximisation problems.

Scheme A (see Algorithm 1) maintains a migration interval  $\tau_i$  for each island. As soon as the current best fitness on an island has improved through evolution, the island communicates this solution to its neighbouring islands. In this case, or when the best fitness increases after immigration, the migration interval for that island drops to 1. This implies that copies of a high-fitness immigrant are propagated to all neighbouring islands in the next generation. If no improvement of the current best fitness is found after  $\tau_i$  generations, the migration interval  $\tau_i$  doubles.

For the purpose of a theoretical analysis, we assume that all islands run in synchronicity: the  $t$ th generation is executed on all islands at the same time. However, this is not a restriction of our adaptive scheme, as it can be applied in asynchronous parallel architectures using message passing for implementing migration.

Inspired by Lässig and Sudholt (2011a), we also consider a Scheme B (see Algorithm 2), where the migration interval is halved (instead of being set to 1) once an improvement has been detected. In contrast to Scheme A, this change is not implemented immediately but only after the current migration period has ended. A flag “success<sub>*i*</sub>” is used to indicate whether a success on island  $i$  has occurred in the current

**Algorithm 1** Elitist parallel EA with adaptive Scheme A

---

```

1: Initialize  $\lambda$  islands  $P_1^1, \dots, P_1^\lambda$  uniformly at random and let  $\tau_i := 1$  and  $u_i := 1$  for
   all  $1 \leq i \leq \lambda$ .
2: Let  $f_i^*$  be the best fitness value of island  $P_1^i$ , for all  $1 \leq i \leq \lambda$ .
3: for  $t := 1$  to  $\infty$  do
4:   for all  $1 \leq i \leq \lambda$  in parallel do
5:     Simulate one generation of the EA and create a new population  $P_{t+1}^i$ .
6:     Update  $P_{t+1}^i$  by selecting a new population from the union of  $P_t^i$  and  $P_{t+1}^i$ ,
       keeping a best individual. Let  $f_i^{*'} be the best fitness value in  $P_{t+1}^i$ .
7:     if  $u_i \geq \tau_i$  or  $f_i^{*'} > f_i^*$  then
8:       Send a copy of a fittest offspring in  $P_{t+1}^i$  to all neighbouring islands.
9:     end if
10:    Update  $P_{t+1}^i$  by selecting a new population from the union of  $P_{t+1}^i$  and all
      immigrants, keeping a best individual. Let  $f_i^{*''}$  be the best fitness value in  $P_{t+1}^i$ .
11:    if  $u_i \geq \tau_i$  or  $f_i^{*''} > f_i^*$  then
12:      if  $f_i^{*''} > f_i^{*'}$  then
13:        Let  $\tau_i := 1$  and  $f_i^* := f_i^{*'}$ 
14:      else
15:        Let  $\tau_i := \tau_i \cdot 2$ 
16:      end if
17:      Let  $u_i := 0$ 
18:    end if
19:     $u_i := u_i + 1$ 
20:  end for
21: end for$ 
```

---

migration period. The advantage of Scheme B is that it uses less communication than Scheme A, and if there is a good region in the parameter space of  $\tau$ , our hope is that it will maintain a good parameter value in that region over time.

We provide general methods for analysing the expected parallel time and the expected communication effort for arbitrary elitist EAs that migrate copies of selected individuals (the original individuals remain on their island). The parallel time is defined as the number of generations until a global optimum is found and denoted  $T^{\text{par}}$ . The communication effort  $T^{\text{com}}$  is defined as the total number of individuals migrated until a global optimum is found. For simplicity and ease of presentation, we assume that each migration only transfers one individual; if  $\nu > 1$  individuals migrate, the communication effort has to be multiplied by  $\nu$ .

In order to demonstrate and illustrate this approach, we consider one simple algorithm in more detail: following Lässig and Sudholt (2011a), the parallel (1+1) EA is a special case where each island runs a (1+1) EA.

In terms of communication topologies, for Scheme A we consider general graphs on  $\lambda$  vertices as well as the following common special cases. A unidirectional ring is a ring with edges going in the same direction. A grid graph contains undirected edges with vertices arranged on a two-dimensional grid. A torus can be regarded as a grid where edges wrap around horizontally and vertically. A hypercube graph of dimension  $d$  contains  $2^d$  vertices. Each vertex has a  $d$ -bit label, and vertices are neighboured if and only if their labels differ in exactly one bit. The complete graph contains all possible

---

**Algorithm 2** Elitist parallel EA with adaptive Scheme B

---

```

1: Initialize  $\lambda$  islands  $P_1^1, \dots, P_1^\lambda$  uniformly at random and let  $\tau_i := 1$ ,  $u_i := 1$ , and
   success $_i := \text{false}$  for all  $1 \leq i \leq \lambda$ .
2: Let  $f_i^*$  be the best fitness value of island  $P_1^i$ , for all  $1 \leq i \leq \lambda$ .
3: for  $t := 1$  to  $\infty$  do
4:   for all  $1 \leq i \leq \lambda$  in parallel do
5:     Simulate one generation of the EA and create a new population  $P_{t+1}^i$ .
6:     Update  $P_{t+1}^i$  by selecting a new population from the union of  $P_t^i$  and  $P_{t+1}^i$ ,
       keeping a best individual.
7:     if  $u_i \geq \tau_i$  then
8:       Send a copy of a fittest offspring in  $P_{t+1}^i$  to all neighbouring islands.
9:     end if
10:    Update  $P_{t+1}^i$  by selecting a new population from the union of  $P_{t+1}^i$  and all
      immigrants, keeping a best individual. Let  $f_i^{*'}$  be the best fitness value in  $P_{t+1}^i$ .
11:    if  $f_i^{*'} > f_i^*$  then
12:      Let  $f_i^* := f_i^{*'}$  and success $_i := \text{true}$ 
13:    end if
14:    if  $u_i \geq \tau_i$  then
15:      if success $_i$  then
16:        Let  $\tau_i := \lceil \tau_i / 2 \rceil$ 
17:      else
18:        Let  $\tau_i := \tau_i \cdot 2$ 
19:      end if
20:      Let  $u_i := 0$  and success $_i := \text{false}$ 
21:    end if
22:     $u_i := u_i + 1$ 
23:  end for
24: end for

```

---

edges. For Scheme B we consider only complete topologies. Notice that in this situation the individual migration intervals of each island will all be equal.

The diameter  $\text{diam}(G)$  of a graph  $G$  with  $\lambda$  vertices is defined as the largest number of edges on a shortest path between any two vertices. The unidirectional ring has the largest diameter of  $\lambda$ . The diameter of any grid or torus graph with side lengths  $\sqrt{\lambda} \times \sqrt{\lambda}$  is at most  $2\sqrt{\lambda}$ . The diameter of a  $(\log \lambda)$ -dimensional hypercube is  $\log \lambda$ , and that of a complete topology is 1.

### 3 Fixed Migration Intervals

In order to compare our adaptive schemes against fixed migration intervals, we first need to investigate the latter. For fixed migration intervals, every period of  $\tau$  generations leads to one migration. This simple argument shows that the parallel time and the communication effort are related as follows.

**THEOREM 1:** *Consider an island model with an arbitrary communication topology  $G = (V, E)$  and a fixed migration interval  $\tau$ . Then the communication effort  $T^{\text{com}}$  is related to the parallel optimization time  $T^{\text{par}}$  as follows:*

$$\left\lceil \frac{T^{\text{par}}}{\tau} \right\rceil \leq \frac{T^{\text{com}}}{|E|} \leq \frac{T^{\text{par}}}{\tau}.$$

In order to bound the (expected) communication effort from above or below, it is therefore sufficient to bound the (expected) parallel time.

Lässig and Sudholt (2010; 2013) presented general upper bounds for the parallel optimisation time of island models with different topologies. Their method is based on the so-called fitness level method, also known as fitness-based partitions (Wegener, 2002; Lehre, 2011).

Throughout this work, we use a special case of this method. Without loss of generality, consider a problem with fitness values  $1, \dots, m$ . Consider fitness level sets  $A_1, \dots, A_m$  such that  $A_i$  contains all points with fitness  $i$ . In particular,  $A_m$  contains all global optima. We further assume that if the current best individual of a population is in  $A_i$ , there is a lower bound  $s_i$  for the probability of finding a higher fitness level  $A_{i+1} \cup \dots \cup A_m$  in one generation through variation operators used in the algorithm (e.g., mutation, recombination). It is easy to show that then  $\sum_{i=1}^{m-1} 1/s_i$  is an upper bound for the expected running time of an elitist EA.

Lässig and Sudholt (2010; 2011b) showed how upper bounds on the parallel optimisation time can be derived from functions of the success probabilities  $s_1, \dots, s_{m-1}$ . They considered migration in every generation ( $\tau = 1$ ) (Lässig and Sudholt, 2011b) as well as probabilistic migration, where every island independently decides for each neighbouring island whether migration occurs, and the probability for a migration is a fixed parameter  $p$  (Lässig and Sudholt, 2013). The following theorem is an adaptation of the latter, which is valid for periodic migration with migration interval  $\tau$ . The results for the expected communication effort on a topology with edge set  $E$  follow from multiplying the expected parallel time by  $|E|/\tau$ , as this term reflects the average number of migrated individuals across the topology in one generation. The upper bounds on the expected parallel time can be derived as in Lässig and Sudholt (2013).

**THEOREM 2:** *Consider an island model with  $\lambda$  islands, each running an elitist EA. Every  $\tau$  iterations each island sends a copy of its best individual to all neighbouring islands. Each island incorporates the best out of its own individuals and its immigrants. For fitness level sets  $A_1, \dots, A_m$ ,  $A_i$  containing all points of the  $i$ th fitness value, let  $s_i$  be a lower bound on the probability that in one generation an island in  $A_i$  finds a point in  $A_{i+1} \cup \dots \cup A_m$ . At most, expected parallel optimization time  $E[T^{par}]$  and the expected communication effort  $E[T^{com}]$  are*

$$E[T^{par}] \leq 3\tau^{1/2} \sum_{i=1}^{m-1} \frac{1}{s_i^{1/2}} + \frac{1}{\lambda} \sum_{i=1}^{m-1} \frac{1}{s_i}$$

$$E[T^{com}] \leq 3\lambda\tau^{-1/2} \sum_{i=1}^{m-1} \frac{1}{s_i^{1/2}} + \frac{1}{\tau} \sum_{i=1}^{m-1} \frac{1}{s_i}$$

for every unidirectional ring,

$$E[T^{par}] \leq 3\tau^{2/3} \sum_{i=1}^{m-1} \frac{1}{s_i^{1/3}} + \frac{1}{\lambda} \sum_{i=1}^{m-1} \frac{1}{s_i}$$

$$E[T^{com}] \leq 12\lambda\tau^{-1/3} \sum_{i=1}^{m-1} \frac{1}{s_i^{1/3}} + \frac{4}{\tau} \sum_{i=1}^{m-1} \frac{1}{s_i}$$

for an undirected grid or torus with side lengths  $\sqrt{\lambda} \times \sqrt{\lambda}$ ,

$$E[T^{par}] \leq 2\tau m + \tau \sum_{i=1}^{m-1} \log\left(\frac{1}{\tau s_i}\right) + \frac{1}{\lambda} \sum_{i=1}^{m-1} \frac{1}{s_i}$$

$$E[T^{com}] \leq 2\lambda(\log \lambda)m + \lambda(\log \lambda) \sum_{i=1}^{m-1} \log\left(\frac{1}{\tau s_i}\right) + \frac{\log \lambda}{\tau} \sum_{i=1}^{m-1} \frac{1}{s_i}$$

for the  $(\log \lambda)$ -dimensional hypercube, and

$$E[T^{par}] \leq m\tau + m + \frac{1}{\lambda} \sum_{i=1}^{m-1} \frac{1}{s_i}$$

$$E[T^{com}] \leq \lambda(\lambda - 1)m + \frac{\lambda(\lambda - 1)m}{\tau} + \frac{\lambda - 1}{\tau} \sum_{i=1}^{m-1} \frac{1}{s_i}$$

for the complete topology  $K_\lambda$ .

PROOF: The statements on communication times follow from the upper bounds on the parallel time, using the second inequality from Theorem 1. The number of (directed) edges in the topology,  $|E|$ , is  $\lambda$  for a unidirectional ring, at most  $4\lambda$  for the stated grids and tori,  $\lambda \log \lambda$  for the hypercube, and  $\lambda(\lambda - 1)$  for the complete topology. So the upper bounds on the expected communication effort follow from multiplying upper bounds on  $E[T^{par}]$  by  $|E|/\tau$ . In the following, we only show the statements on the parallel time.

Consider the time for leaving fitness level  $i$ . For the ring we note that after  $(k - 1)\tau$  generations, for any integer  $1 \leq k \leq \lambda$ ,  $k$  islands will be on fitness level  $i$ , and then the probability of finding a better fitness level is at least

$$1 - (1 - s_i)^k \geq 1 - e^{-ks_i} \geq 1 - \frac{1}{1 + ks_i} = \frac{ks_i}{1 + ks_i},$$

where we have used  $e^x \geq 1 + x$  in both inequalities. So, for every  $1 \leq k \leq \lambda$ , the expected number of generations on level  $i$  is at most

$$(k - 1)\tau + \frac{1 + ks_i}{ks_i} \leq k\tau + \frac{1}{ks_i}.$$

If  $\tau^{-1/2}s_i^{-1/2} < 1$ , we use the trivial upper bound

$$\frac{1}{s_i} < \tau^{1/2}s_i^{-1/2}.$$

Otherwise, if  $1 \leq \tau^{-1/2}s_i^{-1/2} \leq \lambda$ , we pick  $k := \lceil \tau^{-1/2}s_i^{-1/2} \rceil$  such that  $1 \leq k \leq \lambda$  (since  $\lambda$  is an integer) and hence  $k \leq 2\tau^{-1/2}s_i^{-1/2}$ , which yields the upper bound

$$k\tau + \frac{1}{ks_i} \leq 2\tau^{1/2}s_i^{-1/2} + \tau^{1/2}s_i^{-1/2} = 3\tau^{1/2}s_i^{-1/2}.$$

If  $\tau^{-1/2}s_i^{-1/2} > \lambda$  we set  $k := \lambda$  and get

$$\lambda\tau + \frac{1}{\lambda s_i} < \tau^{1/2}s_i^{-1/2} + \frac{1}{\lambda s_i}.$$

The maximum over all these cases is at most

$$3\tau^{1/2}s_i^{-1/2} + \frac{1}{\lambda s_i},$$

and summing over all  $1 \leq i \leq m - 1$  yields the claimed bound for the ring.

Likewise, for torus and grid graphs within  $(k - 1)\tau$  generations, for any integer  $1 \leq k \leq \sqrt{\lambda}$ , at least  $k^2$  islands will be on the current best fitness level, as this time is sufficient to cover a rectangle of side lengths  $k \times k$ . As before, we get a time bound of

$$(k - 1)\tau + \frac{1 + k^2s_i}{k^2s_i} \leq k\tau + \frac{1}{k^2s_i}.$$

If  $\tau^{-2/3}s_i^{-2/3} < 1$ , we again use the trivial upper bound

$$\frac{1}{s_i} \leq \tau^{2/3}s_i^{-1/3}.$$

Otherwise, if  $1 \leq \tau^{-2/3}s_i^{-2/3} \leq \lambda$ , we pick  $k := \lceil \tau^{-1/3}s_i^{-1/3} \rceil$  such that  $1 \leq k \leq \sqrt{\lambda}$  (as  $\sqrt{\lambda}$  is an integer) and hence  $k \leq 2\tau^{-1/3}s_i^{-1/3}$ , which yields the upper bound

$$k\tau + \frac{1}{k^2s_i} \leq 2\tau^{2/3}s_i^{-1/3} + \tau^{2/3}s_i^{-1/3} = 3\tau^{2/3}s_i^{-1/3}.$$

Finally, for  $\tau^{-2/3}s_i^{-2/3} > \lambda$ , we get for  $k := \sqrt{\lambda}$ ,

$$\sqrt{\lambda}\tau + \frac{1}{\lambda s_i} \leq \tau^{2/3}s_i^{-1/3} + \frac{1}{\lambda s_i}.$$

The maximum over all these cases is at most

$$3\tau^{2/3}s_i^{-1/3} + \frac{1}{\lambda s_i},$$

and summing over all  $1 \leq i \leq m - 1$  yields the claimed bound for the torus.

For the hypercube, after  $(k - 1)\tau$  generations, for any integer  $1 \leq k \leq \log \lambda$ ,  $\sum_{i=0}^k \binom{\log \lambda}{i} \geq 2^k$  islands will be on fitness level  $i$ . So the expected time on fitness level  $i$  is at most

$$(k - 1)\tau + \frac{1}{2^k s_i} + 1 \leq k\tau + \frac{1}{2^k s_i}.$$

If  $\tau^{-1}s_i^{-1} \leq 1$ , we get

$$\frac{1}{s_i} < \tau.$$

If  $\tau^{-1}s_i^{-1} > \lambda$ , we set  $k := \log \lambda$  and get

$$(\log \lambda)\tau + \frac{1}{\lambda s_i} \leq \tau \log \left( \frac{1}{\tau s_i} \right) + \frac{1}{\lambda s_i}.$$

Otherwise, if  $1 < \tau^{-1}s_i^{-1} \leq \lambda$ , putting  $k := \lceil \log(\frac{1}{\tau s_i}) \rceil$  (implying  $k \leq \log \lambda$ , since  $\log \lambda$  is an integer) and using  $k \leq \log(\frac{1}{\tau s_i}) + 1$  gives an upper bound of

$$k\tau + \frac{1}{2^k s_i} \leq \tau \log \left( \frac{1}{\tau s_i} \right) + 2\tau.$$

The maximum over all these cases is at most

$$2\tau + \tau \log\left(\frac{1}{\tau s_i}\right) + \frac{1}{\lambda s_i},$$

and summing over all  $1 \leq i \leq m - 1$  yields the claimed bound for the hypercube.

Finally, for the complete graph, if  $\tau^{-1}s_i^{-1} < 1$ , we use the upper bound

$$\frac{1}{s_i} < \tau,$$

and otherwise after  $\tau$  generations all  $\lambda$  islands are on fitness level  $i$ , yielding the upper bound

$$\tau + \frac{1}{\lambda s_i} + 1.$$

Summing over all  $1 \leq i \leq m - 1$  yields the claimed bound for the complete graph.  $\square$

The upper bounds from Theorem 2 match the ones from Lässig and Sudholt (2014b) for the case of probabilistic migration if we compare  $\tau$  against a migration probability of  $p = 1/\tau$ ; the constant factors here are even better. The constants for probabilistic migration are higher to account for the variation in the spread of information. Periodic migration is more reliable in this respect, since information is guaranteed to be spread every  $\tau$  generations.

#### 4 Adaptive Scheme A

In this section we analyse Scheme A on different topologies, including those from Theorem 2. Note that whenever an island improves its current best solution, a copy of this solution is spread to all neighbouring islands immediately. Thus, good fitness levels spread in the same way as migrating in every generation would do, i.e., using a global parameter  $\tau = 1$ . This means that the upper bounds from Theorem 2 apply for  $\tau = 1$ .

**THEOREM 3:** *For Scheme A on topologies from Theorem 2, the expected parallel optimisation time is bounded from above as in Theorem 2 with  $\tau = 1$ .*

Note that the bounds on the expected parallel time from Theorem 2 are minimised for  $\tau = 1$ . This implies that we get upper bounds on the expected parallel time equal to the best upper bounds for any fixed choice of the migration interval. In case these bounds are asymptotically tight, this means that our adaptive Scheme A never increases the expected parallel running time asymptotically.

The intended benefit of Scheme A comes from a reduced communication effort, as all islands decrease communication, while no improvement is encountered through either variation or immigration. The expected communication effort is bounded from above in the following theorem. The main observation is that for each fitness level, the number of migrations from an island is logarithmic in the time it remains on that fitness level. For an upper bound we consider the expected worst-case time spent on a fitness level  $A_i$ , where the worst case is taken over all populations with their best individual in  $A_i$ .

**THEOREM 4:** *Consider Scheme A on an arbitrary communication topology  $G = (V, E)$  with diameter  $\text{diam}(G)$ . Let  $E[T_i^{\text{par}}]$  be (an upper bound on) the worst-case expected number of generations during which the current best search point in the island model is on fitness level  $i$ .*

Then the expected communication effort is at most

$$E[T^{com}] \leq |E| \sum_{i=1}^{m-1} \log(E[T_i^{par}] + \text{diam}(G)).$$

PROOF: Initially, and after improving its current best fitness, an island will double its migration interval until its current best fitness improves again. If the current best fitness does not improve for  $t$  generations, the island will perform at most  $\log t$  migrations.

Consider an island  $v$  after reaching fitness level  $i$  for the first time, either through variation or immigration. If no other island has found a better fitness level, the random parallel time for some island finding such an improvement is given (or bounded) by  $T_i^{par}$ . Then this solution (or another solution of fitness better than  $i$ ) will be propagated through the topology, advancing to neighbouring islands in every generation. Hence, some solution on a better fitness level than  $i$  will eventually reach  $v$  within the following  $\text{diam}(G)$  generations. The latter also holds if some island has already found a better fitness level than  $i$  at the time  $v$  reaches fitness level  $i$ . In any case, the total time  $v$  will spend on fitness level  $i$  is at most  $T_i^{par} + \text{diam}(G)$ .

An island  $v$  in the topology with outdegree  $\text{deg}^+(v)$  will send  $\text{deg}^+(v)$  individuals during each migration. Hence, the total number of migrated solutions in  $t$  generations on fitness level  $i$  is at most

$$\sum_{v \in V} \text{deg}^+(v) \cdot \log t = |E| \log t.$$

The expected communication effort, therefore, is at most

$$\begin{aligned} E[T^{com}] &\leq |E| \sum_{i=1}^{m-1} E[\log(T_i^{par} + \text{diam}(G))] \\ &\leq |E| \sum_{i=1}^{m-1} \log(E[T_i^{par} + \text{diam}(G)]) \\ &= |E| \sum_{i=1}^{m-1} \log(E[T_i^{par}] + \text{diam}(G)), \end{aligned}$$

where the inequality follows from Jensen's inequality and the fact that  $\log$  is a concave function.  $\square$

The communication effort is proportional to the logarithm of the expected time spent on each fitness level. For functions that can be optimised by the island model in expected polynomial time, and for polynomial  $\lambda$  (note  $\text{diam}(G) \leq \lambda$ ), this logarithm is always at most  $O(\log n)$ . Then Theorem 4 gives the following upper bound.

COROLLARY 5: Consider a fitness function  $f$  with  $m$  fitness values, such that  $f$  is being optimised by the island model in an expected polynomial number of generations, for every initial population. If also  $\lambda = n^{O(1)}$ , we have

$$E[T^{com}] \leq O(|E|m \cdot \log n).$$

The expected parallel time on a fitness level can be smaller than a polynomial. If sufficiently many islands are being used, and the topology spreads information quickly enough, it can be logarithmic or even constant. For specific topologies we get

the following results by combining Theorem 4 with bounds on the parallel time from Theorem 2 for  $\tau = 1$ , as established by Theorem 3.

**THEOREM 6:** *Given success probabilities  $s_1, \dots, s_{m-1}$  as in Theorem 2, the expected communication effort for Scheme A is bounded from above for certain topologies with  $\lambda$  islands as follows:*

- $\lambda \sum_{i=1}^{m-1} \log \left( \frac{3}{s_i^{1/2}} + \frac{1}{\lambda s_i} + \lambda \right)$  for a unidirectional ring
- $4\lambda \sum_{i=1}^{m-1} \log \left( \frac{3}{s_i^{1/3}} + \frac{1}{\lambda s_i} + 2\sqrt{\lambda} \right)$  for every undirected grid or torus with side lengths  $\sqrt{\lambda} \times \sqrt{\lambda}$
- $\lambda(\log \lambda) \sum_{i=1}^{m-1} \log \left( 2 + \log \left( \frac{1}{s_i} \right) + \frac{1}{\lambda s_i} + \log \lambda \right)$  for the  $(\log \lambda)$ -dimensional hypercube
- $\lambda(\lambda - 1) \sum_{i=1}^{m-1} \log \left( 2 + \frac{1}{\lambda s_i} \right)$  for the complete graph

We demonstrate the application of this theorem in Section 6.

## 5 Adaptive Scheme B

Scheme A resets the migration interval of an island to 1 every time an improvement is found. We propose Scheme B, which halves this value instead. This may be advantageous if there is a Goldilocks region of good values for the migration interval across several subsequent fitness levels. In contrast to Scheme A, Scheme B should be able to maintain a value in that region over several fitness levels.

When improvements are being found with probability  $p$ , good parameter values are close to  $\tau \approx 1/(\lambda p)$ , as then we find one improvement in expectation between any migration. If the current migration interval is much smaller, chances of finding an improvement are small, and  $\tau$  is likely to increase. Likewise, if  $\tau$  is large, the island will find an improvement with high probability and halve  $\tau$ . Thus, the migration interval will reach an equilibrium state close to  $1/(\lambda p)$ .

Scheme B might have a smaller communication effort as it does not reset the migration interval to 1, thus leading to a chain of frequent migrations. This, however, only holds if Scheme B does not lead to an increase in the parallel time. In fact, for sparse topologies  $G$ , such as the unidirectional ring, there is a risk that improvements may take a very long time to be communicated. If, say, all islands had the same migration interval  $\tau$ , and one island found a new best solution, it might take up to  $\text{diam}(G) \cdot \tau$  generations for this information to arrive at the last island. Scheme B therefore makes more sense for dense topologies such as the complete graph, where  $\text{diam}(G) = 1$ , and decreases of  $\tau$  quickly propagate to all islands.

We follow the analysis of Scheme B for adapting the number of islands in Lässig and Sudholt (2011a) and note the following similarities. In both approaches, over some time span, a resource is doubled and halved depending on whether an improvement of the best fitness is found. In Lässig/Sudholt, this resource is the number of islands and hence the number of function evaluations executed in one generation. Here it is the number of generations within one migration period. The Lässig/Sudholt time span is just one generation, leading to the parallel time as performance measure. In our case the time span is the current migration interval, leading to the communication effort as performance measure.

For  $\lambda = 1$  the parallel time in our work equals the sequential time in Lässig/Sudholt, and the communication effort in our work equals the parallel time in their paper. However, a difference emerges for  $\lambda > 1$ , as in our scenario an island has  $\lambda\tau$  trials to find an improvement, so the resources for finding improvements are by a factor of  $\lambda$  higher compared to Lässig/Sudholt.

We adapt the analysis from Lässig and Sudholt to accommodate this difference. The following lemma is a straightforward adaptation of parts of their Lemma 1. It estimates the expected communication effort for finding a single improvement, based on a given initial migration interval  $\tau_0$ . We abbreviate  $\max\{x, 0\}$  by  $(x)^+$ .

LEMMA 7: Assume an island model with a complete topology starts with migration interval  $\tau_0$  and that in each generation each island finds an improvement over the current best individual with probability at least  $p$ . Let  $T^{com}$  be the random number of individuals migrated between islands until an improvement is found on any island. Let  $\tau^*$  be the migration interval at this time (before it is halved). Then for every  $\alpha \in \mathbb{N}_0$

1.  $\Pr(\log(\tau^*) - \log(\tau_0) > (\log(1/(\lambda p)) - \log(\tau_0))^+ + \alpha) \leq \exp(-2^\alpha)$ ,
2.  $E[T^{com}] \leq \lambda(\lambda - 1) \cdot ((\log(1/(\lambda p)) - \log(\tau_0))^+ + 2)$ .

PROOF: The proof closely follows the proof of Lässig and Sudholt (2011a), Lemma 1. The condition

$$\log(\tau^*) - \log(\tau_0) > (\log(1/(\lambda p)) - \log(\tau_0))^+ + \alpha$$

requires that no island finds an improvement before  $\log(\tau^*) - \log(\tau_0)$  exceeds the stated threshold. The latter is equivalent to

$$\begin{aligned} \log(\tau^*) &\geq (\log(1/(\lambda p)) - \log(\tau_0))^+ + \alpha + \log(\tau_0) \\ &\geq \log(1/(\lambda p)) + \alpha. \end{aligned}$$

The number of generations between migrations is then at least

$$\tau^* \geq 2^{\log(1/(\lambda p)) + \alpha} = 1/(\lambda p) \cdot 2^\alpha.$$

In order for  $\log(\tau^*) - \log(\tau_0)$  to exceed the threshold, we must not have a success on any island during these generations. The number of trials for obtaining a success on any island is by a factor of  $\lambda$  larger than the number of generations. Hence

$$\begin{aligned} \Pr(\log(\tau^*) - \log(\tau_0) > (\log(1/(\lambda p)) - \log(\tau_0))^+ + \alpha) &\leq (1 - p)^{1/p \cdot 2^\alpha} \\ &\leq \exp(-2^\alpha). \end{aligned}$$

To bound the expectation, we observe that the first statement implies  $\Pr(\log(\tau^*) - \log(\tau_0) \geq (\log(1/(\lambda p)) - \log(\tau_0))^+ + \alpha + 1) \leq \exp(-2^\alpha)$ . The number of migrants sent in one migration is  $\lambda \cdot (\lambda - 1)$ . Since  $E[T^{com}]$  is non-negative, we have

$$\begin{aligned} &\frac{E[T^{com}]}{\lambda \cdot (\lambda - 1)} \\ &= 1 + E[\log(\tau^*) - \log(\tau_0)] \\ &= 1 + \sum_{t=1}^{\infty} \Pr(\log(\tau^*) - \log(\tau_0) \geq t), \end{aligned}$$

splitting the sum at  $(\log(1/(\lambda p)) - \log(\tau_0))^+$  and bounding probabilities for small  $t$  by 1,

$$\begin{aligned} &\leq (\log(1/(\lambda p)) - \log(\tau_0))^+ + 1 + \sum_{\alpha=0}^{\infty} \Pr(\log(\tau) \geq (\log(1/(\lambda p)) - \log(\tau_0))^+ + \alpha + 1) \\ &\leq (\log(1/(\lambda p)) - \log(\tau_0))^+ + 1 + \sum_{\alpha=0}^{\infty} \exp(-2^\alpha) \\ &< (\log(1/(\lambda p)) - \log(\tau_0))^+ + 2, \end{aligned}$$

as the last sum is less than 1. □

The expected number of migrations is expressed as the difference between logarithms of the ideal value  $1/(\lambda p)$  and the initial value  $\tau_0$ . If the initial migration interval is larger,  $\tau_0 \geq 1/(\lambda p)$ , the expected number of migrations is just 2.

This fact is reflected in the following theorem. The upper bound on the communication effort only contains values  $\log(1/(\lambda s_j))$  when the migration interval needs to be increased, i.e.  $s_j < s_{j-1}$ . For the special case where fitness levels get progressively harder,  $s_1 \geq s_2 \geq \dots \geq s_{m-1}$ , the bound simplifies significantly.

**THEOREM 8:** *Given success probabilities  $s_1, \dots, s_{m-1}$  as in Theorem 2, for Scheme B we have*

$$\begin{aligned} E [T_B^{par}] &\leq 3m + \frac{3}{2\lambda} \sum_{i=1}^{m-1} \frac{1}{s_i} \quad \text{and} \\ E [T_B^{com}] &\leq \lambda(\lambda - 1) \cdot \left( 3m + \log\left(\frac{1}{\lambda s_1}\right) + \sum_{j=2}^{m-1} \left( \log\left(\frac{1}{\lambda s_j}\right) - \log\left(\frac{1}{\lambda s_{j-1}}\right) \right)^+ \right). \end{aligned}$$

For  $s_1 \geq s_2 \geq \dots \geq s_{m-1}$  the latter simplifies to

$$E [T_B^{com}] \leq \lambda(\lambda - 1) \cdot \left( 3m + \log\left(\frac{1}{\lambda s_{m-1}}\right) \right).$$

The upper bound for the expected parallel time is only by a factor of 3/2 larger than the upper bound for Scheme A. Hence, both upper bounds are asymptotically equal. In other words, the reduced communication in Scheme B does not worsen the running time, for problems where the upper bounds for Schemes A and B are asymptotically tight.

We give an informal argument to convey the intuition for this result. Assume that Scheme B has raised the migration interval from 1 to some large value  $\tau^* = 2^i$  before an improvement is found. Then Scheme B has spent  $1 + 2 + 4 + \dots + 2^i = 2^{i+1} - 1$  generations leading up to this value. In the worst case, the new migration interval  $\tau^*/2 = 2^{i-1}$  is too large: improvements are found easily, and then the algorithm has to idle for nearly  $\tau^*/2$  generations before being able to communicate the current best fitness. The total time spent is around  $3/2 \cdot (2^{i+1} - 1)$ , whereas Scheme A in the same situation would spend  $2^{i+1} - 1$  generations. So, in this setting Scheme B needs at most 3/2 as many generations as Scheme A.

A formal argument was given in Lässig and Sudholt (2011a) to derive an upper bound for the parallel time of Scheme B, which is 3/2 that for Scheme A. The bound on the expected communication effort follows from similar arguments and applying our refined Lemma 7. We refrain from giving a formal proof here as it can be obtained with straightforward modifications from the proof of Theorem 3 in Lässig and Sudholt (2011a).

## 6 Performance on Common Example Functions

The analytical frameworks for analysing fixed migration intervals and our two adaptive schemes can be applied by simply using lower bounds on success probabilities for improving the current fitness. We demonstrate this approach by analysing the parallel time and the communication effort on common test problems.

In the following we provide an analysis for the maximisation of the same pseudo-Boolean test functions investigated in Lässig and Sudholt (2014b). For a search point  $x \in \{0, 1\}^n$ ,  $x = x_1, \dots, x_n$ , we define  $\text{OneMax}(x) := \sum_{i=1}^n x_i$  as the number of 1s in  $x$ , and  $\text{LO}(x) := \sum_{i=1}^n \prod_{j=1}^i x_j$  as the number of leading 1s in  $x$ . We also consider the class of unimodal functions taking  $d$  fitness values  $f_1, \dots, f_d$ . A function is called unimodal if every nonoptimal search point has a Hamming neighbour with strictly larger fitness. Finally for  $1 \leq k \leq n$  we consider

$$\text{Jump}_k := \begin{cases} k + \sum_{i=1}^n x_i, & \text{if } \sum_{i=1}^n x_i \leq n - k \text{ or } x = 1^n, \\ \sum_{i=1}^n (1 - x_i) & \text{otherwise.} \end{cases}$$

The name comes from the fact that typically at the some point in the evolution a mutation flipping  $k$  specific bits (a “jump”) has to be made. The parameter  $k$  tunes the difficulty of this function.

### 6.1 Fitness Partition and Success Probabilities

In order to apply Theorems 2, 6, 8, it is just necessary to define the probability  $s_i$  to move from the fitness level  $A_i$  to a better one. Recall that  $s_i$  is a lower bound on the probability of one island finding a search point of strictly higher fitness, given that the population has current best fitness  $i$ .

For the simple (1+1) EA, these values are easy to derive:

- For OneMax, a search point with  $i$  1s has  $n-i$  zeros. The probability of a mutation flipping only one of these zeros is  $s_i \geq (n-i)/n \cdot (1-1/n)^{n-1} \geq (n-i)/(en)$ .
- For LO, it is sufficient to flip the first 0-bit, which has probability  $s_i \geq 1/n \cdot (1-1/n)^{n-1} \geq 1/(en)$ .
- For unimodal functions, the success probability on each fitness level is at least  $s_i \geq 1/n \cdot (1-1/n)^{n-1} \geq 1/(en)$ , as for any nonoptimal point there is always a better Hamming neighbour.
- For  $\text{Jump}_k$  with  $k \geq 2$ , it is possible to find an improvement to an individual having up to  $n-k$  1-bits just increasing the number of 1s, thus the  $s_i$  with  $i < n-k$  are equal to the 1s for OneMax. A similar argument applies to levels  $n-k < i < n$ . Once we have obtained an individual with  $n-k$  1-bits, an improvement is found by generating as offspring a specific bit string having Hamming distance  $k$  from the parent, which has probability  $s_{n-k} \geq \binom{1}{n}^k \cdot (1-\frac{1}{n})^{n-k} \geq (\frac{1}{n})^k \cdot (1-\frac{1}{n})^{n-1} \geq \frac{1}{en^k}$ .

## 6.2 Fixed Scheme

Given Theorem 2, it is possible to bound the parallel time and the communication effort for fixed migration intervals similarly to Lässig and Sudholt (2014b). For example, for OneMax and the complete topology, we get an expected parallel time of  $E[T^{\text{par}}] = O(n\tau + \frac{n \log n}{\lambda})$  and an expected communication effort of  $E[T^{\text{com}}] = O(\lambda^2 n + \frac{\lambda n \log n}{\tau})$ .

For fixed  $\lambda$ , the value of  $\tau$  yields a trade-off between upper bounds for the parallel time and that of the communication effort. In our example, for, say,  $\lambda = O(1)$ , we get  $E[T^{\text{par}}] = O(n\tau + n \log n)$  and  $E[T^{\text{com}}] = O(n + \frac{n \log n}{\tau})$ . We can notice how a large  $\tau$  always minimises the bound for the communication effort, while a small one (i.e.,  $\tau = 1$ ) minimises the bound for the parallel time.

Given a fixed number of islands  $\lambda$ , we define the best  $\tau$  as the largest  $\tau$  that does not asymptotically increase the bound for the parallel time (compared to  $\tau = 1$ ). This assures good scalability while minimising the communication effort. For the example proposed, the best  $\tau$  is  $\tau = \Theta(\frac{\log n}{\lambda})$ . This leads to  $E[T^{\text{par}}] = O(\frac{n \log n}{\lambda})$  and  $E[T^{\text{com}}] = O(\frac{|E|}{\tau} \cdot E[T^{\text{par}}]) = O(\frac{\lambda^2}{\tau} \cdot E[T^{\text{par}}]) = O(\lambda^2 n)$ . The results for other topologies and problems are summarised in Table 1. Notice that fixing  $\tau$  to its best value is possible, provided that the number of islands is small enough. Particularly for the example proposed the number of islands must be  $\lambda = O(\log n)$  in order for the best  $\tau$  to be defined ( $\tau \geq 1$ ).

For OneMax and LO, lower bounds on the expected parallel times in Table 1 follow from general lower bounds on the class of mutation-based evolutionary algorithms (Sudholt, 2013). Every mutation-based algorithm needs at least  $\Omega(n \log n)$  function evaluations on OneMax and at least  $\Omega(n^2)$  function evaluations on LO. The parallel (1+1) EA with  $\lambda$  islands makes  $\lambda$  evaluations in one generation; hence the bounds translate to  $E[T^{\text{par}}] = \Omega(\frac{n \log n}{\lambda})$  for OneMax and  $E[T^{\text{par}}] = \Omega(\frac{n^2}{\lambda})$  for LO. Lower bounds for the communication efforts for fixed migration intervals come from the lower bound in Theorem 1. The floor function can be ignored in the asymptotic notation, as in all cases of Table 1,  $E[T^{\text{par}}]/\tau = \Omega(1)$ , implying  $E[\lfloor T^{\text{par}}/\tau \rfloor] \geq E[T^{\text{par}}/\tau - 1] = \Omega(E[T^{\text{par}}/\tau])$ .

## 6.3 Adaptive Scheme

In order to calculate the parallel time for the adaptive Scheme A, we can refer to the results for the fixed scheme when  $\tau = 1$ , as shown in Theorem 3. For example, our Scheme A running on a complete topology solves OneMax in time  $O(\frac{n \log n}{\lambda} + n)$ . Scheme B has asymptotically the same parallel time of Scheme A, as Theorem 8 shows.

We only consider values of  $\lambda$  that lead to a linear speedup as defined in Lässig and Sudholt (2013): the expected parallel time for  $\lambda$  islands is by a factor of  $\lambda$  smaller than the expected optimisation time of a single island, in an asymptotic sense. In this setting an island model thus makes the same expected number of function evaluations compared to a single island. In the example proposed, linear speedup is achieved for a number of islands up to  $\lambda = O(\log n)$ ; in fact for a larger number of islands the upper bound on the parallel time would be  $O(n)$  regardless of  $\lambda$ . The bounds on  $\lambda$  limit the best upper bound on the parallel time achievable with our analytical framework. Table 1 shows the bound on  $\lambda$  for different problems and topologies and the best achievable parallel time bound. For OneMax and LO, lower bounds on the expected parallel times in Table 1 follow from general lower bounds on the class of mutation-based evolutionary algorithms (Sudholt, 2013).

In order to calculate the communication effort we use Theorems 6 and 8 for Schemes A and B, respectively. We first get a general bound for every  $\lambda$  included

in the linear speedup bound, and then we calculate it for the maximum value of  $\lambda$ , thus providing the communication effort for the value of  $\lambda$  leading to the best achievable parallel time. Table 1 shows all results derived in this manner. In the following we provide an example of this calculation for Scheme A on LO.

### 6.3.1 Example: Communication Effort of Scheme A for LO

We provide details on how to calculate bounds on the expected communication effort for Scheme A using Theorem 6, choosing LO as an example. Calculations for other test functions are similar. The purpose is to illustrate how we derived the results in Table 1.

In the following,  $\lambda$  is restricted to the cases leading to linear speedup, as stated in the following and in Table 1. The calculations often use that  $\log(a + b + c) \leq \log(3 \max(a, b, c))$ .

- For the complete topology ( $\lambda = O(n)$ ),

$$\begin{aligned} E(T_A^{\text{com}}) &\leq \lambda(\lambda - 1) \sum_{j=0}^{n-1} \log\left(2 + \frac{en}{\lambda}\right) \\ &\leq \lambda(\lambda - 1) \sum_{j=0}^{n-1} \log\left(\frac{2en}{\lambda}\right) = O\left(\lambda^2 n \log\left(\frac{n}{\lambda}\right)\right). \end{aligned}$$

If we set  $\lambda = \Theta(n)$ ,

$$E(T_A^{\text{com}}) = O(n^3).$$

- For ring ( $\lambda = O(n^{1/2})$ ),

$$\begin{aligned} E(T_A^{\text{com}}) &= \lambda \sum_{i=0}^{n-1} \log\left(\frac{3}{s_i^{1/2}} + \frac{1}{\lambda s_i} + \lambda\right) \\ &\leq \lambda \sum_{i=0}^{n-1} \log\left(3(en)^{1/2} + \frac{en}{\lambda} + \lambda\right) \\ &\leq \lambda \sum_{i=0}^{n-1} \log\left(\frac{9en}{\lambda}\right) \\ &= O\left(\lambda n \log\left(\frac{n}{\lambda}\right)\right). \end{aligned}$$

If we set  $\lambda = \Theta(n^{1/2})$ , we obtain

$$E(T_A^{\text{com}}) = O(n^{3/2} \log n).$$

- For the grid ( $\lambda = O(n^{2/3})$ ), we get

$$\begin{aligned} E(T_A^{\text{com}}) &= 4\lambda \sum_{i=1}^{n-1} \log\left(\frac{3}{s_i^{1/3}} + \frac{1}{\lambda s_i} + 2\sqrt{\lambda}\right) \\ &= O\left(\lambda n \log\left(\frac{n}{\lambda}\right)\right). \end{aligned}$$

If we set  $\lambda = \Theta(n^{2/3})$ ,

$$E(T_A^{\text{com}}) = O(n^{5/3} \log n).$$

Table 1: Expected parallel optimisation times and expected communication efforts for  $\lambda$  islands running a  $(1+1)$  EA. The table shows restrictions on  $\lambda$  to yield linear speedups and fixed values for  $\tau$ , leading to the best upper bounds for the communication effort while not increasing the parallel running time. For both parallel time and communication effort, we show bounds for general  $\lambda$  in the realm of linear speedups, and the best parallel time  $E[T_{\text{best}}^{\text{par}}]$  achieved by using the largest such  $\lambda$ , along with the communication effort  $E[T_{\text{best}}^{\text{par}}]$  for the same  $\lambda$ .

Topology and Scheme	$\lambda$	$E[T_{\text{best}}^{\text{par}}] \rightsquigarrow$	$E[T_{\text{best}}^{\text{com}}] \rightsquigarrow$
OneMax			
Complete, Scheme A	$O(\log n)$	$\Theta\left(\frac{n \log n}{\lambda}\right) \rightsquigarrow \Theta(n)$	$O(\lambda^2 n) \rightsquigarrow O(n \log^2(n))$
Complete, Scheme B	$O(\log n)$	$\Theta\left(\frac{n \log n}{\lambda}\right) \rightsquigarrow \Theta(n)$	$O(\lambda^2 n) \rightsquigarrow O(n \log^2(n))$
Complete, $\tau = \frac{\log n}{\lambda}$	$O(\log n)$	$\Theta\left(\frac{n \log n}{\lambda}\right) \rightsquigarrow \Theta(n)$	$\Theta(\lambda^2 n) \rightsquigarrow \Theta(n \log^2(n))$
Ring, Scheme A	$O(\log n)$	$\Theta\left(\frac{n \log n}{\lambda}\right) \rightsquigarrow \Theta(n)$	$O(n \lambda \log \lambda) \rightsquigarrow O(n \log(n) \log \log(n))$
Ring, $\tau = \left(\frac{\log n}{\lambda}\right)^2$	$O(\log n)$	$\Theta\left(\frac{n \log n}{\lambda}\right) \rightsquigarrow \Theta(n)$	$\Theta\left(\frac{\lambda^2 n}{\log n}\right) \rightsquigarrow \Theta(n \log(n))$
Grid, Scheme A	$O(\log n)$	$\Theta\left(\frac{n \log n}{\lambda}\right) \rightsquigarrow \Theta(n)$	$O(n \lambda \log \lambda) \rightsquigarrow O(n \log(n) \log \log(n))$
Grid, $\tau = \left(\frac{\log n}{\lambda}\right)^{3/2}$	$O(\log n)$	$\Theta\left(\frac{n \log n}{\lambda}\right) \rightsquigarrow \Theta(n)$	$\Theta\left(\frac{\lambda^{3/2} n}{\log^{1/2}(n)}\right) \rightsquigarrow \Theta(n \log(n))$
Hypercube, Scheme A	$O(\log n)$	$\Theta\left(\frac{n \log n}{\lambda}\right) \rightsquigarrow \Theta(n)$	$O(n \lambda \log(\lambda) \log \log(\lambda)) \rightsquigarrow O(n \log(n) \log \log(n) \log \log \log(n))$
Hypercube, $\tau = \frac{\log n}{\lambda}$	$O(\log n)$	$\Theta\left(\frac{n \log n}{\lambda}\right) \rightsquigarrow \Theta(n)$	$\Theta(n \lambda \log \lambda) \rightsquigarrow \Theta(n \log(n) \log \log(n))$
LO			
Complete, Scheme A	$O(n)$	$\Theta\left(\frac{n^2}{\lambda}\right) \rightsquigarrow \Theta(n)$	$O(\lambda^2 n \log\left(\frac{n}{\lambda}\right)) \rightsquigarrow O(n^3)$
Complete, Scheme B	$O(n)$	$\Theta\left(\frac{n^2}{\lambda}\right) \rightsquigarrow \Theta(n)$	$O(\lambda^2 n) \rightsquigarrow O(n^3)$
Complete, $\tau = \frac{n}{\lambda}$	$O(n)$	$\Theta\left(\frac{n^2}{\lambda}\right) \rightsquigarrow \Theta(n)$	$\Theta(\lambda^2 n) \rightsquigarrow \Theta(n^3)$
Ring, Scheme A	$O(n^{1/2})$	$\Theta\left(\frac{n^2}{\lambda}\right) \rightsquigarrow \Theta(n^{3/2})$	$O(\lambda n \log\left(\frac{n}{\lambda}\right)) \rightsquigarrow O(n^{3/2} \log n)$
Ring, $\tau = \frac{n}{\lambda^2}$	$O(n^{1/2})$	$\Theta\left(\frac{n^2}{\lambda}\right) \rightsquigarrow \Theta(n^{3/2})$	$\Theta(\lambda^2 n) \rightsquigarrow \Theta(n^2)$
Grid, Scheme A	$O(n^{2/3})$	$\Theta\left(\frac{n^2}{\lambda}\right) \rightsquigarrow \Theta(n^{4/3})$	$O(\lambda n \log\left(\frac{n}{\lambda}\right)) \rightsquigarrow O(n^{5/3} \log n)$
Grid, $\tau = \frac{n}{\lambda^{3/2}}$	$O(n^{2/3})$	$\Theta\left(\frac{n^2}{\lambda}\right) \rightsquigarrow \Theta(n^{4/3})$	$\Theta(n \lambda^{3/2}) \rightsquigarrow \Theta(n^2)$
Hypercube, Scheme A	$O\left(\frac{n}{\log n}\right)$	$\Theta\left(\frac{n^2}{\lambda}\right) \rightsquigarrow \Theta(n \log n)$	$O(\lambda \log(\lambda) n \log\left(\frac{n}{\lambda}\right)) \rightsquigarrow O(n^2 \log \log n)$
Hypercube, $\tau = \frac{n}{\lambda \log n}$	$O\left(\frac{n}{\log n}\right)$	$\Theta\left(\frac{n^2}{\lambda}\right) \rightsquigarrow \Theta(n \log n)$	$\Theta(\lambda \log(\lambda) n \log(n)) \rightsquigarrow \Theta(n^2 \log n)$

Table 1: Continued.

Topology and Scheme	$\lambda$	$E[T^{\text{par}}] \rightsquigarrow E[T^{\text{par}}_{\text{best}}]$	$E[T^{\text{com}}] \rightsquigarrow E[T^{\text{com}}_{\text{best}}]$
Unimodal, $d f$ -values	Complete, Scheme A	$O\left(\frac{dn}{\lambda}\right) \rightsquigarrow O(d)$	$O(\lambda^2 d \log\left(\frac{n}{\lambda}\right)) \rightsquigarrow O(n^2 d)$
	Complete, Scheme B	$O\left(\frac{dn}{\lambda}\right) \rightsquigarrow O(d)$	$O(\lambda^2 d) \rightsquigarrow O(n^2 d)$
	Complete, $\tau = \frac{n}{\lambda}$	$O\left(\frac{dn}{\lambda}\right) \rightsquigarrow O(d)$	$O(\lambda^2 d) \rightsquigarrow O(n^2 d)$
	Ring, Scheme A	$O\left(\frac{dn}{\lambda}\right) \rightsquigarrow O(dn^{1/2})$	$O(\lambda d \log\left(\frac{n}{\lambda}\right)) \rightsquigarrow O(dn^{1/2} \log n)$
	Ring, $\tau = \frac{n}{\lambda^2}$	$O\left(\frac{dn}{\lambda}\right) \rightsquigarrow O(dn^{1/2})$	$O(\lambda^2 d) \rightsquigarrow O(dn)$
	Grid, Scheme A	$O\left(\frac{dn}{\lambda}\right) \rightsquigarrow O(dn^{1/3})$	$O(\lambda d \log\left(\frac{n}{\lambda}\right)) \rightsquigarrow O(dn^{2/3} \log n)$
	Grid, $\tau = \frac{n}{\lambda^{3/2}}$	$O\left(\frac{dn}{\lambda}\right) \rightsquigarrow O(dn^{1/3})$	$O(\lambda^{3/2} d) \rightsquigarrow O(dn)$
	Hypercube, Scheme A	$O\left(\frac{n}{\log n}\right) \rightsquigarrow O(d \log n)$	$O(\lambda \log(\lambda) d \log\left(\frac{n}{\lambda}\right)) \rightsquigarrow O(dn \log \log n)$
	Hypercube, $\tau = \frac{n}{\lambda \log n}$	$O\left(\frac{n}{\log n}\right) \rightsquigarrow O(d \log n)$	$O(\lambda \log(\lambda) d \log(n)) \rightsquigarrow O(dn \log n)$
	$\text{Jump}_{k'}$ $3 \leq k = O(n/(\log n))$	Complete, Scheme A	$O\left(\frac{n^k}{\lambda}\right) \rightsquigarrow O(n)$
Complete, Scheme B		$O\left(\frac{n^k}{\lambda}\right) \rightsquigarrow O(n)$	$O(\lambda^2 n) \rightsquigarrow O(n^{2k-1})$
Complete, $\tau = \frac{n^{k-1}}{\lambda}$		$O\left(\frac{n^k}{\lambda}\right) \rightsquigarrow O(n)$	$O(\lambda^2 n) \rightsquigarrow O(n^{2k-1})$
Ring, Scheme A		$O\left(\frac{n^k}{\lambda}\right) \rightsquigarrow O(n^{k/2})$	$O(n\lambda \log \lambda) \rightsquigarrow O(kn^{k/2+1} \log n)$
Ring, $\tau = \frac{n^k}{\lambda^2}$		$O\left(\frac{n^k}{\lambda}\right) \rightsquigarrow O(n^{k/2})$	$O(\lambda^2) \rightsquigarrow O(n^k)$
Grid, Scheme A		$O\left(\frac{n^k}{\lambda}\right) \rightsquigarrow O(n^{k/3})$	$O(n\lambda \log \lambda) \rightsquigarrow O(kn^{2k/3+1} \log n)$
Grid, $\tau = \frac{n^k}{\lambda^{3/2}}$		$O\left(\frac{n^k}{\lambda}\right) \rightsquigarrow O(n^{k/3})$	$O(\lambda^{3/2}) \rightsquigarrow O(n^k)$
Hypercube, Scheme A		$O\left(\frac{n^k}{\lambda}\right) \rightsquigarrow O(n)$	$O(n\lambda \log(\lambda) \log \log(\lambda)) \rightsquigarrow O(kn^k \log(n) \log \log(n^{k-1}))$
Hypercube, $\tau = \frac{n^{k-1}}{\lambda}$		$O\left(\frac{n^k}{\lambda}\right) \rightsquigarrow O(n)$	$O(n\lambda \log \lambda) \rightsquigarrow O(kn^k \log n)$

- For the hypercube ( $\lambda = O(\frac{n}{\log n})$ ),

$$\begin{aligned}
 E(T_A^{\text{com}}) &= \lambda(\log \lambda) \sum_{i=1}^{n-1} \log \left( 2 + \log \left( \frac{1}{s_i} \right) + \frac{1}{\lambda s_i} + \log \lambda \right) \\
 &\leq \lambda(\log \lambda) \sum_{i=1}^{n-1} \log \left( 2 + \log(en) + \frac{en}{\lambda} + \log \lambda \right) \\
 &= O \left( \lambda(\log \lambda)n \log \left( \frac{n}{\lambda} \right) \right).
 \end{aligned}$$

If we set  $\lambda = \Theta(\frac{n}{\log n})$ , we get

$$E(T_A^{\text{com}}) = O(n^2 \log \log n).$$

### 6.4 Evaluation of Results

Recall that Table 1 only shows results for linear speedups, hence all (upper bounds on) parallel times are equal, but the range of  $\lambda$  values varies between topologies.

Table 2 compares upper bounds from Table 1 on the communication efforts for the best fixed value of  $\tau$  against our adaptive schemes. For OneMax on all topologies, the upper bound on the communication effort is by a small  $O(\log \log n)$  term larger for the adaptive schemes compared to the best fixed  $\tau$ . The latter varies according to the topology: it is  $\tau = ((\log n)/\lambda)^2$  for the ring,  $\tau = ((\log n)/\lambda)^{3/2}$  for the grid, and  $\tau = (\log n)/\lambda$  for the hypercube and the complete graph. So, the additional  $O(\log \log n)$  factor is a small price to pay for the convenience of adapting  $\tau$  automatically.

For LO, Scheme A on the ring has a communication effort of  $O(n^{3/2} \log n)$  compared to  $\Theta(n^2)$ , and for the grid it is  $O(n^{5/3} \log n)$  versus  $\Theta(n^2)$ . Since the bounds for fixed  $\tau$  are tight, we see an improvement of  $\Omega(n^{1/2}/\log n)$  and  $\Omega(n^{1/3}/\log n)$ , respectively. These significant improvements show that decreasing the migration interval is an effective strategy in lowering the communication costs without harming the parallel running time. For the hypercube the communication effort is lower by a factor of  $\Omega(\log(n)/\log \log n)$ , whereas for the complete graph no differences are visible in the upper bounds.

Table 2: Comparison of upper bounds on the communication efforts. The table shows the asymptotic ratio of upper bounds on the communication effort from the rightmost column of Table 1 for the best fixed choice of  $\tau$  and the best adaptive scheme based on bounds from Table 1. A value less than 1 indicates that the best fixed  $\tau$  leads to better bounds, and a value larger than 1 indicates a decrease of the upper bound on the communication effort by the stated factor. In all cases  $\lambda$  was chosen as the largest possible value that guarantees a linear speedup according to the upper bounds.

	OneMax	LeadingOnes	Unimodal	Jump <sub>k</sub>
Complete	1	1	1	1
Ring	$\frac{1}{\log \log n}$	$\frac{n^{1/2}}{\log n}$	$\frac{n^{1/2}}{\log n}$	$\frac{n^{k/2-1}}{k \log n}$
Grid/Torus	$\frac{1}{\log \log n}$	$\frac{n^{1/3}}{\log n}$	$\frac{n^{1/3}}{\log n}$	$\frac{n^{k/3-1}}{k \log n}$
Hypercube	$\frac{1}{\log \log \log n}$	$\frac{\log n}{\log \log n}$	$\frac{\log n}{\log \log n}$	$\frac{1}{\log \log(n^{k-1})}$

For general bounds for unimodal functions, Scheme A on the ring has a communication effort of  $O(n^{3/2} \log n)$  compared to  $O(n^2)$ , and for the grid it is  $O(n^{5/3} \log n)$  versus  $O(n^2)$ . This means that the adaptive scheme can guarantee a lower upper bound for the communication effort. For the hypercube the upper bound on the communication effort is lower by a factor of  $O(\log(n)/\log \log n)$ , whereas for the complete graph no differences are visible in the upper bounds.

For Jump, with regard to comparing upper bounds, there are no differences for the complete graph, while on the hypercube Scheme A is by a  $O(\log \log(n^{k-1}))$  factor worse than the best fixed value. For rings and grids the adaptive scheme is better; the performance gap even grows with  $k$  and hence the difficulty of the function.

Comparing Schemes A and B in Table 1, both achieve similar results. For LO we see an advantage of Scheme B over Scheme A: the general bound for the communication effort of Scheme A is  $O(\lambda^2 n \log(n/\lambda))$ , whereas that for Scheme B is  $O(\lambda^2 n)$ . This makes sense, as the probability for finding an improvement in one generation is of order  $\Theta(\lambda/n)$  for the considered  $\lambda$ , and the ideal value for the migration interval is in the region of  $\Theta(n/\lambda)$ . Scheme A needs to increase the migration interval around  $\log(n/\lambda)$  times to get into this range, which is precisely the performance difference visible in our upper bounds. The difference disappears for  $\lambda = \Theta(n)$ .

The same argument also applies to the more general function class of unimodal functions.

## 7 Discussion

The adaptive schemes presented here were designed to reduce the communication effort without compromising exploitation. The results from Section 6.3 have demonstrated that this goal is achieved for functions that require a high degree of exploitation and little or no exploration. In settings where more exploration and less exploitation is needed, our schemes may not be able to find optimal or near-optimal migration intervals.

As mentioned in Section 1.1, the function class LOLZ from Lässig and Sudholt (2013) was the first constructed example where island models excel over panmictic populations and independent runs. Its structure is similar to the LO problem in a sense that bits have to be fixed to their correct values from left to right. In addition, LOLZ contains a number of traps that islands may fall into. For any island that has not gotten stuck in a trap, the probability of finding a fitness improvement is always at least  $1/(en)$  as for LO.

Solving LOLZ efficiently requires a delicate choice of the migration interval. For  $\tau = n^{5/3}$ , the parallel (1+1) EA finds global optima on LOLZ efficiently with overwhelming probability, given appropriate parameter settings for the islands model (number of islands and communication topology) (Lässig and Sudholt, 2013, Theorem 3). This value of  $\tau$  is large enough to allow islands to explore different regions of the search space independently. But it is also small enough to propagate solutions of islands that are on target to finding the optimum, which then take over islands that have gotten stuck in local optima.

Let us consider the performance of our schemes with a complete topology, using nonrigorous arguments that could likely be turned into a rigorous (but lengthy) analysis. As argued in the beginning of Section 5, Scheme B will be drawn toward an equilibrium state for  $\tau$  close to  $1/(\lambda p)$ , where  $p = \Theta(1/n)$  as for LO. This holds so long as not all islands have gotten stuck in a trap. So for most of the time, we will have  $\tau = O(n/\lambda)$ , as deviations to larger values are very unlikely (see the first statement of Lemma 7). The same also holds for Scheme A.

However, as the diameter of the complete topology  $K_\lambda$  is 1, we are then in a realm where  $\tau = O(n/(\lambda \cdot \text{diam}(K_\lambda)))$ , and the island model is known to be inefficient (Lässig and Sudholt, 2013, Theorem 1). The reason is that migration happens too frequently, and the island model behaves similarly to a panmictic population, which is prone to getting stuck in a trap. Here our schemes, for complete topologies, would focus too much on exploitation, and not give the necessary degree of exploration needed to optimise LOLZ.

Lässig and Sudholt (2014a, sec. 6.1) also studied an instance of the Eulerian cycle problem, where a large migration interval is beneficial, as it leads to a better exploration. The same arguments as before lead to the same conclusion: our schemes typically will not give the necessary degree of exploration, leading to suboptimal performance for complete topologies.

For sparse topologies like rings, however, exploration is increased and the situation may improve. It is still unlikely that the recommended choice of  $\tau = n^{5/3}$  for LOLZ is being reached, as migration intervals are likely to remain in the  $O(n)$  range for most of the time (by similar arguments as for the first statement of Lemma 7). But we conjecture that for sufficiently large  $\lambda$  and sparse topologies, a sufficient degree of exploration can be achieved. A detailed, formal analysis is beyond the scope of this paper and is left as an open problem for future work.

## 8 Conclusions and Future Work

We have presented two adaptive schemes for adapting the migration interval in island models. Both schemes have been accompanied by analytical frameworks that yield upper bounds on the expected parallel time and the expected communication effort, based on probabilities of fitness improvements in single islands running elitist EAs. The results show that our schemes are able to decrease the upper bound on the communication effort without significantly compromising exploitation. For arbitrary topologies, we got upper bounds on the expected parallel time that are asymptotically no larger than those for maximum exploitation, that is, migration in every generation.

Example applications to the parallel (1+1) EA on common example functions revealed that in the realm of linear speedups and comparing against the best fixed choice of the migration interval, the upper bound on the expected communication effort was larger by a tiny  $O(\log \log n)$  term for OneMax and similarly for the hypercube on Jump, but significantly lower for a general analysis of unimodal functions and for rings and grids on Jump. For LO, the adaptive Scheme A on grid and ring topologies can even guarantee an upper bound on the communication effort that is polynomially lower than the lower bound for the best fixed choice of the migration interval.

One avenue for future work is to evaluate our adaptive schemes empirically on parallel hardware and for real-world problems, to assess the practicality of our approach outside the scope of examples covered here. Another avenue is formalising the ideas from Section 7, leading to a rigorous analysis of our schemes with different topologies on examples that require a certain degree of exploration. This might also address the challenging question of how to choose the right communication topology for a given problem. And it might be possible to further refine our schemes to allow an explicit tuning of the balance between exploitation and exploration.

Finally, this work presents mostly *upper bounds* for the expected running time and just an upper bound for the expected communication effort of the adaptive schemes. Obtaining corresponding *lower bounds* would help to identify what performance can be achieved, and assist in the search for provably optimal communication strategies. A

promising direction is using black box complexity (Droste et al., 2006; Lehre and Witt, 2012), which describes universal lower bounds on the expected (worst-case) running time of *every* black box algorithm on a given class of functions. Recent advances toward a black box complexity for parallel and distributed black box algorithms have been made (Badkobeh et al., 2014; 2015), which include island models using mutation for variation.

## Acknowledgments

The authors would like to thank Joseph Kempka and the participants of Dagstuhl seminar 13271 “Theory of Evolutionary Algorithms” for fruitful discussions, and the anonymous reviewers for their constructive comments. The work of Andrea Mambrini was partly funded by EPSRC (Grant No. EP/I010297/1). The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement 618091 (SAGE).

## References

- Alba, E. (2002). Parallel evolutionary algorithms can achieve super-linear performance. *Information Processing Letters*, 82(1): 7–13.
- Alba, E., and Luque, G. (2005). Theoretical models of selection pressure for dEAs: Topology influence. In *Proceedings of the IEEE Congress on Evolutionary Computation*, pp. 214–221.
- Alba, E., Luque, G., and Nesmachnow, S. (2013). Parallel metaheuristics: Recent advances and new trends. *International Transactions in Operational Research*, 20(1): 1–48.
- Araujo, L., and Merelo Guervós, J. J. (2011). Diversity through multiculturalism: Assessing migrant choice policies in an island model. *IEEE Transactions on Evolutionary Computation*, 15(4): 456–469.
- Badkobeh, G., Lehre, P. K., and Sudholt, D. (2014). Unbiased black-box complexity of parallel search. In *Parallel Problem Solving from Nature*, pp. 892–901. Lecture Notes in Computer Science, vol. 8672.
- Badkobeh, G., Lehre, P. K., and Sudholt, D. (2015). Black-box complexity of parallel search with distributed populations. In *Proceedings of the Conference on Foundations of Genetic Algorithms*, pp. 3–15.
- Bravo, Y., Luque, G., and Alba, E. (2012). Influence of the migration period in parallel distributed GAs for dynamic optimization. In *Proceedings of the Conference on Learning and Intelligent Optimization*, pp. 343–348.
- Droste, S., Jansen, T., and Wegener, I. (2006). Upper and lower bounds for randomized search heuristics in black-box optimization. *Theory of Computing Systems*, 39(4): 525–544.
- Hong, T.-P., Lin, W.-Y., Liu, S.-M., and Lin, J.-H. (2007). Experimental analysis of dynamic migration intervals on 0/1 knapsack problems. In *Proceedings of the IEEE Congress on Evolutionary Computation*, pp. 1163–1167.
- Lässig, J., and Sudholt, D. (2010). General scheme for analyzing running times of parallel evolutionary algorithms. In *Parallel Problem Solving from Nature*, pp. 234–243. Lecture Notes in Computer Science, vol. 6238.
- Lässig, J., and Sudholt, D. (2011a). Adaptive population models for offspring populations and parallel evolutionary algorithms. In *Proceedings of the Conference on Foundations of Genetic Algorithms*, pp. 181–192.
- Lässig, J., and Sudholt, D. (2011b). Analysis of speedups in parallel evolutionary algorithms for combinatorial optimization. In *Proceedings of the International Symposium on Algorithms and Computation*, pp. 405–414. Lecture Notes in Computer Science, vol. 7074.

- Lässig, J., and Sudholt, D. (2013). Design and analysis of migration in parallel evolutionary algorithms. *Soft Computing*, 17(7): 1121–1144.
- Lässig, J., and Sudholt, D. (2014a). Analysis of speedups in parallel evolutionary algorithms and  $(1+\lambda)$  EAs for combinatorial optimization. *Theoretical Computer Science*, 551:66–83.
- Lässig, J., and Sudholt, D. (2014b). General upper bounds on the running time of parallel evolutionary algorithms. *Evolutionary Computation*, 22(3): 405–437.
- Lehre, P. K. (2011). Fitness-levels for non-elitist populations. In *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO)*, pp. 2075–2082.
- Lehre, P. K., and Witt, C. (2012). Black-box search by unbiased variation. *Algorithmica*, 64(4): 623–642.
- Lin, W.-Y., Hong, T.-P., Liu, S.-M., and Lin, J.-H. (2012). Revisiting the design of adaptive migration schemes for multipopulation genetic algorithms. In *Proceedings of the Conference on Technologies and Applications of Artificial Intelligence*, pp. 338–343.
- Luque, G., and Alba, E. (2011). *Parallel genetic algorithms: Theory and real world applications*. Studies in Computational Intelligence, vol. 367. Berlin: Springer.
- Mambrini, A., and Sudholt, D. (2014). Design and analysis of adaptive migration intervals in parallel evolutionary algorithms. In *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO)*, pp. 1047–1054.
- Mambrini, A., Sudholt, D., and Yao, X. (2012). Homogeneous and heterogeneous island models for the set cover problem. In *Parallel Problem Solving from Nature*, pp. 11–20. Lecture Notes in Computer Science, vol. 7491.
- Neumann, F., Oliveto, P. S., Rudolph, G., and Sudholt, D. (2011). On the effectiveness of crossover for migration in parallel evolutionary algorithms. In *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO)*, pp. 1587–1594.
- Osorio, K., Alba, E., and Luque, G. (2013). Using theory to self-tune migration periods in distributed genetic algorithms. In *Proceedings of the IEEE Congress on Evolutionary Computation*, pp. 2595–2601.
- Osorio, K., Luque, G., and Alba, E. (2011). Distributed evolutionary algorithms with adaptive migration period. In *Proceedings of the International Conference on Intelligent Systems Design and Applications*, pp. 259–264.
- Rudolph, G. (2006). Takeover time in parallel populations with migration. In *Proceedings of the International Conference on Bioinspired Optimization Methods and Their Applications*, pp. 63–72.
- Skolicki, Z., and De Jong, K. (2005). The influence of migration sizes and intervals on island models. In *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO)*, pp. 1295–1302.
- Sudholt, D. (2013). A new method for lower bounds on the running time of evolutionary algorithms. *IEEE Transactions on Evolutionary Computation*, 17(3): 418–435.
- Tomassini, M. (2005). *Spatially structured evolutionary algorithms: Artificial evolution in space and time*. Berlin: Springer.
- Wegener, I. (2002). Methods for the analysis of evolutionary algorithms on pseudo-Boolean functions. In R. Sarker, X. Yao, and M. Mohammadian (Eds.), *Evolutionary optimization*, pp. 349–369. New York: Springer.