Algorithmica (2019) 81:886–915
https://doi.org/10.1007/s00453-018-0445-2

Island Models Meet Rumor Spreading

Benjamin Doerr[1] · Philipp Fischbeck[2] · Clemens Frahnow[2] · Tobias Friedrich[2] · Timo Kötzing[2] · Martin Schirneck[2]

Received: 31 August 2017 / Accepted: 18 April 2018 / Published online: 30 April 2018

Abstract Island models in evolutionary computation solve problems by a careful interplay of independently running evolutionary algorithms on the island and an exchange of good solutions between the islands. In this work, we conduct rigorous run time analyses for such island models trying to simultaneously obtain good run times and low communication effort. We improve the existing upper bounds for both measures (i) by improving the run time bounds via a careful analysis, (ii) by balancing individual computation and communication in a more appropriate manner, and (iii) by replacing the usual communicate-with-all approach with randomized rumor spreading. In the latter, each island contacts a randomly chosen neighbor. This epidemic communication paradigm is known to lead to very fast and robust information dissemination in many applications. Our results concern island models running simple $(1+1)$ evolutionary algorithms to optimize the classic test functions OneMax and LeadingOnes. We investigate binary trees, d-dimensional tori, and complete graphs as communication topologies.

Keywords Evolutionary algorithm · Run time analysis · Communication costs · Island model · Rumor spreading

An extended abstract of this work was presented at the *2017 Genetic and Evolutionary Computation Conference* (GECCO) [8].

✉ Philipp Fischbeck
philipp.fischbeck@hpi.de

1 Laboratoire d'Informatique (LIX), École Polytechnique, Palaiseau, France

2 Hasso Plattner Institute, Potsdam, Germany

1 Introduction

To speed up evolutionary algorithms, *island models* can be used as a means of distributing the work load over many computing nodes. Each island runs a simple evolutionary algorithm, occasionally sharing information with other nodes. One of the most common ways of sharing information is to send a copy of the best-so-far solution to other islands, a process called *migration*. Many applications of this paradigm are known to be successful [1,2,6].

One main choice for designing an efficient algorithm following the island model is to choose the way migration is carried out. For example, the islands can be equipped with a neighborhood structure, determining for each island to which other islands to migrate its individuals. This is referred to as the *migration topology*. Dense migration topologies, such as the complete graph, lead to a fast spread of good solutions at the price of a high communication overhead. The impact of the migration topology on algorithm performance has been analyzed both experimentally [23] and theoretically [17]. Another choice lies in the frequency of migration. A frequent approach is to introduce a parameter τ indicating that once every τ generations all islands engage in communication with all neighbors. A high value of τ can thus save on the communication overhead, at the price of delays in the spread of new good individuals. Setting the migration interval correctly is a challenge for designing efficient island algorithms [20]. It has been noted that island models are also particularly useful for dynamic optimization problems [19] and when employing crossover [22]. An overview of practical concerns of research in the area of island models can be found in [2]; for an overview of theoretical work, see [25].

In this work, we will consider λ islands running a $(1+1)$ EA, a standard evolutionary algorithm (EA) considered in theoretical analyses [13]. For various migrations topologies (such as d-dimensional tori and the complete graph) and migration intervals τ, we are interested in the expected time until some island evolves the optimal solution for the given fitness function, of which we consider the two standard functions OneMax and LeadingOnes. It is not surprising that in this simple setting of unimodal fitness functions, fast migration topologies, such as the complete graph, perform best in terms of the number of generations while performing badly in terms of communication [17]. We improve the analysis especially pertaining to the *combined costs* (number of generations plus number of communications per island) in the following ways.

First, we analyze the run time of the island models carefully. We see that, for the number of generations that any of the topologies require on OneMax, the dependence on τ is not linear, but, surprisingly, logarithmic. For the complete topology we further improve the bounds on the number of generations by making a detailed analysis of different optimization phases and employing a variable drift theorem; we show this analysis to be tight by providing matching lower bounds. Second, we use the parameter τ to avoid communication overhead. By finding the right balance between individual computation of the islands and spreading the information to the neighbors, we see that the combined costs for the complete graph on OneMax are as low as $\mathrm{O}(n \log\log n)$, using $\lambda = \Theta(\log n)$ islands and a migration interval of $\tau = \Theta(\log n)$. Similarly, we obtain combined costs for the complete graph on LeadingOnes of $\mathrm{O}(n^{3/2}) \cap$

$\Omega(n^{3/2}/\log n)$, using $\lambda = \Theta(\sqrt{n})$ islands and a migration interval of $\tau = \Theta(\sqrt{n})$ for the upper bound. Finally, we question the method of broadcasting the information to all available neighbors. Instead, we propose to employ the *push protocol*, known from the area of *epidemic algorithms* or *rumor spreading*, where in each communication round each island chooses one neighbor uniformly at random to send the best individual to. It is known that for the complete topology the process requires logarithmically many communication rounds until all islands are informed [9,11]. This is significantly faster than the ring and torus topologies considered previously (and also faster than d-dimensional tori in general), while the communication overhead is still constant per island and communication round (compared with the linear overhead of complete topologies). By proving lower bounds on the performance of the complete topology, we show that the push protocol is superior even to broadcast communication in some settings. Additionally, we generalize the findings for the push protocol to broadcast communication in any migration topology where the number of informed islands has an exponential growth in the number of communication rounds, in particular the complete binary tree.

Table 1 gives an overview of our results. Section 2 formally introduces the island models and test functions, and lists relevant tools from the literature. In Sect. 3 the push protocol is examined. For comparison, Sect. 4 gives run time bounds for the broadcast communication model on multidimensional tori. Section 5 treats the complete topology. The work is concluded in Sect. 6. This article improves upon its extended abstract [8] in several ways. We employ a correspondence between the push protocol and the information dissemination in complete binary trees to extend our method to those topologies as well, see Corollary 12. In Corollary 16 we show a universal upper bound on the run time and communication costs of any island model on connected graphs optimizing arbitrary unimodal functions. Also, the lower bounds in Theorem 24 regarding the LeadingOnes fitness function have been improved.

2 Island Models

In this paper we examine the maximization of unimodal, pseudo-Boolean functions $f\colon \{0,1\}^n \to \mathbb{R}_0^+$ on bit strings $\mathbf{x} = \mathbf{x}_1\mathbf{x}_2\cdots\mathbf{x}_n$ of length n. We interpret the value $f(\mathbf{x})$ as the *fitness* of the individual $\mathbf{x}$. A fitness function is called *unimodal* if every non-optimal bit string has a Hamming-neighbor of higher fitness. We investigate

$$\textsc{OneMax}(\mathbf{x}) = \sum_{i=1}^{n} \mathbf{x}_i, \quad \textsc{LeadingOnes}(\mathbf{x}) = \sum_{i=1}^{n} \prod_{j=1}^{i} \mathbf{x}_j$$

as prototypes of unimodal functions with $n+1$ different values. The main difference between these two functions is the number of improving Hamming-neighbors. While every bit string $\mathbf{x}$ with $\textsc{OneMax}(\mathbf{x}) = i < n$ has $n-i$ neighbors of higher fitness, the improving neighbor with respect to LeadingOnes is unique.

We employ the *island model* as a common framework for distributed evolutionary computation, cf. [17,22,23]. Assume an undirected graph $G = (V, E)$, the *migration*

Table 1 Overview of results. The expected optimization times of island models on λ islands, with migration interval τ, working on bit strings of length n. The objective functions are OneMax, LeadingOnes, and general unimodal functions with ν values. Shown are the run times for the push protocol on complete graphs as well as broadcast communication on complete binary trees, d-dimensional tori, and complete graphs. A universal upper bound for any connected migration topology is given. The minimal optimization time (min. opt. time) is the parallel run time for the optimal choice of parameters λ and τ. The minimal combined costs (min. comb. costs) refer to the parameter setting that minimizes the sum of the optimization time and communication costs

Objective	Model	Optimization time	Min. opt. time	Min. comb. costs
OneMax	Push	$\mathrm{O}\left(\frac{n}{\lambda}\log n + n\log\tau\right)$	$\mathrm{O}(n)$	$\mathrm{O}(n)$
	Binary tree	$\mathrm{O}\left(\frac{n}{\lambda}\log n + n\log\tau\right)$	$\mathrm{O}(n)$	$\mathrm{O}(n)$
	d-Torus	$\mathrm{O}\left(\frac{n}{\lambda}\log n + n\log\tau\right)$	$\mathrm{O}(n)$	$\mathrm{O}(n)$
	Complete	$\Theta\left(\frac{n}{\lambda}\log n + \frac{\tau n}{\log\lambda}\log\left(\frac{\log\lambda}{\tau}\right)\right)$, if $\tau = \mathrm{o}(\log\lambda)$ $\Theta\left(\frac{n}{\lambda}\log n + n\right)$, if $\tau = \Theta(\log\lambda)$ $\Theta\left(\frac{n}{\lambda}\log n + n\log\left(\frac{\tau}{\log\lambda}\right)\right)$, if $\tau = \omega(\log\lambda)$	$\Theta\left(n\,\frac{\log\log n}{\log n}\right)$	$\Theta(n\log\log n)$
LeadingOnes	Push	$\mathrm{O}\left(\frac{n^2}{\lambda} + \tau n\log n\right)$	$\mathrm{O}(n\log n)$	$\mathrm{O}(n\log n)$
	Binary tree	$\mathrm{O}\left(\frac{n^2}{\lambda} + \tau n\log n\right)$	$\mathrm{O}(n\log n)$	$\mathrm{O}(n\log n)$
	d-Torus	$\mathrm{O}\left(\frac{n^2}{\lambda} + \tau^{\frac{d}{d+1}} n^{\frac{d+2}{d+1}}\right)$	$\mathrm{O}\left(n^{\frac{d+2}{d+1}}\right)$	$\mathrm{O}\left(n^{\frac{d+2}{d+1}}\right)$
	Complete	$\Omega\left(\frac{n^2}{\lambda} + \frac{n}{\log(\lambda/n)} + \frac{\tau n}{\log^2 n}\right)$, $\mathrm{O}\left(\frac{n^2}{\lambda} + \tau n\right)$	$\Omega\left(\frac{n}{\log n}\right)$, $\mathrm{O}(n)$	$\Omega\left(\frac{n^{3/2}}{\log n}\right)$, $\mathrm{O}\left(n^{3/2}\right)$
Unimodal	Connected	$\mathrm{O}\left(\frac{\nu n}{\lambda} + \nu\sqrt{\tau n}\right)$	$\mathrm{O}\left(\nu\sqrt{n}\right)$	$\mathrm{O}\left(\nu\sqrt{n}\right)$

topology, on $\lambda = |V|$ vertices to be given. Every vertex, called an *island*, represents an independent instance of the (1 + 1) EA using standard bit mutation. Prior to the first iteration all islands are initialized uniformly at random, after that they operate in lockstep. Occasionally, governed by a *migration protocol*, the islands share copies of their currently best solutions along the edges of G. A migrant replaces the solution on the receiving island if the fitness of the former is not smaller than that of the latter. Ties among incoming migrants of maximal fitness are broken uniformly at random. We use periodic migration every τ rounds, the *migration interval*. The simplest migration protocol is a broadcast to all neighboring islands. This leads to Algorithm 1. Here $\mathbf{x}^{(j)}$ denotes the best individual on island j.

We are mainly interested in two measures of complexity. The first is the number of generations until an optimal individual is sampled for the first time on any island. We call this random variable (of the random choices of the algorithm) the *optimization time*, denoted by T. Additionally, we keep track of the messages sent during the migration phases leading up to an optimal solution, see line 9 of Algorithm 1. We adopt an amortized view on the *communication costs* as we account for the *average* number of messages per island. Let C denote the number of messages sent over the whole optimization time. Observe that C is a random variable even in the case of a deterministic migration protocol usually strongly correlated with T. We refer to the sum $T + C$ as the *combined costs*. This implicitly assumes that the expenses to generate and evaluate a new individual as well as sending a message to one neighboring island are within constant factors. However, as all theorems quantify the two measures separately, they can easily be extended to larger weights.

In this work we derive asymptotic bounds on the expectations $E[T]$ and $E[C]$ for several migration topologies and protocols. To distinguish the two fitness functions also in notation, we let $E[T_{\text{OM}}]$ and $E[C_{\text{OM}}]$ stand for the respective cost measures when optimizing OneMax, and $E[T_{\text{LO}}]$ and $E[C_{\text{LO}}]$ for LeadingOnes. All bounds are in terms of n, λ, and τ simultaneously. More formally, we regard $\lambda = \lambda(n)$ and $\tau = \tau(n)$ as positive, non-decreasing, integer-valued functions and calculate the univariate asymptotics of the expected costs with respect to n for arbitrary choices of λ and τ.

2.1 The Spreading Time

Our main tool to establish the results below is a fitness level argument, cp. [26]. We say that Algorithm 1 is on *fitness level* i if the maximum fitness over all islands equals i. OneMax and LeadingOnes both induce $n + 1$ fitness levels. Due to the elitist selection the level never decreases and T measures the number of rounds until the algorithm enters level n. We split T into partial optimization times $T_0, T_1, \ldots, T_{n-1}$, where T_i is the time needed to leave level i. The sum over the expected partial waiting times is then an upper bound on $E[T]$.

Algorithm 1: Island model with migration topology $G = (V, E)$ on λ islands and migration interval τ.

$t \leftarrow 0$;
for $1 \leq j \leq \lambda$ **in parallel do**
  $\mathbf{x}^{(j)} \leftarrow$ solution drawn uniformly at random from $\{0, 1\}^n$;
repeat
  $t \leftarrow t + 1$;
  for $1 \leq j \leq \lambda$ **in parallel do**
    $\mathbf{y}^{(j)} \leftarrow$ flip each bit of $\mathbf{x}^{(j)}$ independently with probability $1/n$;
    if $f(\mathbf{y}^{(j)}) \geq f(\mathbf{x}^{(j)})$ **then** $\mathbf{x}^{(j)} \leftarrow \mathbf{y}^{(j)}$;
    if $t \bmod \tau = 0$ **then**
      Send $\mathbf{x}^{(j)}$ to all islands k with $\{j, k\} \in E$;
      $N = \{\mathbf{x}^{(i)} \mid \{i, j\} \in E\}$;
      $M = \{\mathbf{x}^{(i)} \in N \mid f(\mathbf{x}^{(i)}) = \max_{\mathbf{x} \in N} f(\mathbf{x})\}$;
      $\mathbf{y}^{(j)} \leftarrow$ solution drawn uniformly at random from M;
      if $f(\mathbf{y}^{(j)}) \geq f(\mathbf{x}^{(j)})$ **then** $\mathbf{x}^{(j)} \leftarrow \mathbf{y}^{(j)}$
until *termination condition met*;

The value of T_i crucially depends on the number of islands whose individual has the currently best fitness i. By preferring fitter individuals, migration helps to spread good solutions so that more islands can effectively contribute to the overall progress. The ability of a topology to speed up computation through migration is quantified in the notion of the *spreading time*.

Definition 1 Let $G = (V, E)$ be a migration topology, $v \in V$ a vertex, and $1 \leq k \leq \lambda$ an integer. For a given (potentially randomized) communication protocol, let $X_v(k)$ denote the random number of communication steps until at least k islands are informed for the first time, starting from v. The *spreading time* of G is the function $S(k) = \max_{v \in V} E[X_v(k)]$, if it exists.

The spreading time $S(k)$ is thus the *worst-case* expected number of steps needed to inform k islands. In this form, the definition can be applied to arbitrary topologies and protocols. However, if G is not connected, variable $X_v(k)$ may be unbounded with positive probability, leaving the spreading time undefined. In this work we almost exclusively investigate regular graphs in which the expectations $E[X_v(k)]$ are equal for all v. The two exceptions are binary trees in Sect. 3 and general connected graphs in Sect. 4. Whenever the communication scheme is deterministic, so is $X_v(k)$. Hence, $S(k)$ simplifies to $\max_{v \in V} X_v(k)$. In the case of broadcast communication, the spreading time is the smallest integer s such that the s-neighborhood of *any* vertex has size at least k.

Communication steps happen only during migration phases, once every τ rounds. The number of generations that pass until a good solution is sufficiently widespread is thus by a factor τ larger than the spreading time. The following lemma was also given implicitly by Lässig and Sudholt [17, Lemma 1].

Lemma 2 *Let p_i be (a lower bound on) the probability that the* (1 + 1) EA *samples an individual of fitness larger than i from one of fitness exactly i and $1 \le \lambda_i \le \lambda$ an arbitrary positive integer. Then, we have*

$$E[T_i] \le 1 + \tau\, S(\lambda_i) + \frac{1}{p_i\, \lambda_i}.$$

Proof After an expected number of $\tau\, S(\lambda_i)$ iterations at least λ_i islands have adopted an individual of maximum fitness i via migration. We pessimistically ignore the effect of independent improvements during that phase. The expected waiting time until one of the λ_i "informed" islands creates a solution of larger fitness gives an upper bound on $E[T_i]$. The probability of not finding a better solution in one round is at most $(1 - p_i)^{\lambda_i}$, thus an upper bound on the waiting time is

$$\frac{1}{1 - (1 - p_i)^{\lambda_i}} \le 1 + \frac{1}{p_i\, \lambda_i}.$$

□

The spreading time S is non-decreasing, so it worsens the estimate when λ_i gets larger. On the other hand, if more islands share a good solution, the probability to complete the current level increases. The extreme value of $\lambda_i = 1$ completely eliminates the influence of the spreading time (as $S(1) = 0$), but it also bars migration from contributing to the optimization process. As a corollary of Lemma 2, we can choose λ_i independently for every fitness level to balance out these two opposing trends.

Corollary 3 *Let $(\lambda_0, \lambda_1, \ldots, \lambda_{n-1})$ be any sequence of positive integers not larger than λ, then*

$$E[T] \le n + \sum_{i=0}^{n-1} \left(\tau\, S(\lambda_i) + \frac{1}{p_i\, \lambda_i} \right).$$

Corollary 3 is formulated for the case of $n+1$ fitness levels (with the last one containing the optimal solutions), the extensions to arbitrary fitness landscapes and even more general definitions of "levels" are immediate. The fact that sequence $(\lambda_i)_i$ can be chosen freely, lends great versatility to above result. In the remaining sections we repeatedly exploit this to bound the run times and communication costs of island models on several topologies.

2.2 Additional Tools

Beside the spreading time technique we employ several other tools, most notably in the analysis of broadcast protocol on the complete topology in Sect. 5. To bound the tail of the binomial distribution we use Chernoff bounds in multiplicative form, see for example the textbook by Mitzenmacher and Upfal [21].

Theorem 4 (Chernoff bounds, cf. [21]) *For a probability p, let $X \sim \mathrm{Bin}(n, p)$ be a binomially distributed random variable with parameters n and p. Then, for any $\delta > 0$,*

$$\Pr[X \geq (1+\delta)pn] \leq \left(\frac{e^{\delta}}{(1+\delta)^{(1+\delta)}}\right)^{pn} \leq \exp\left(-\ln\left(\frac{1+\delta}{e}\right)(1+\delta)pn\right).$$

In particular, if $\delta \geq 1$, we have

$$\Pr[X \geq (1+\delta)pn] \leq e^{-\frac{\delta}{3}pn},$$

and, for $c \geq 2e$,

$$\Pr[X \geq cpn] \leq 2^{-cpn}.$$

In some cases we use drift analysis to augment the fitness level approach. Hereby, the inner state of the algorithm in question is mapped to a real number via a *potential function* whose development is interpreted as a random process. The time needed to hit a certain value then translates back to the run time of the algorithm. The arguably most general statements to bound such hitting times are the Variable Drift Theorems.

Theorem 5 (Variable Drift Theorem for upper bounds [16]) *Let $(X^{(t)})_{t\in\mathbb{N}}$ be a sequence of random variables over $\mathbb{R}_0^+$ and $T = \min\{t \geq 0 \mid X^{(t)} \leq s_{\min}\}$ the random variable denoting the earliest point in time such that the sequence falls below some prescribed value $s_{\min} \geq 0$. Suppose there is a monotonically increasing function $h\colon \mathbb{R}_0^+ \to \mathbb{R}^+$ such that $1/h$ is integrable and*

$$E\left[X^{(t)} - X^{(t+1)} \,\middle|\, T > t,\ X^{(t)} = s\right] \geq h(s).$$

Then, for all $s_0 \in \mathbb{R}_0^+$ we have

$$E\left[T \,\middle|\, X^{(0)} = s_0\right] \leq \frac{1}{h(s_{\min})} + \int_{s_{\min}}^{s_0} \frac{\mathrm{d}s}{h(s)}.$$

Theorem 6 (Variable Drift Theorem for lower bounds [14]) *Let $s_{\max} \geq s_{\min} > 0$ be two positive real numbers, $S = 0 \cup [s_{\min}, s_{\max}] \subseteq \mathbb{R}^+$, and $(X^{(t)})_{t\in\mathbb{N}}$ a sequence of random variables over S observing $X^{(t+1)} \leq X^{(t)}$ for all t. $T = \min\{t \geq 0 \mid X^{(t)} = 0\}$ is the hitting time. Let $h, \xi\colon [s_{\min}, s_{\max}] \to \mathbb{R}^+$ be two functions such that h is* *right-continuous, monotonically increasing, and $1/h$ is integrable. If there exists a positive real number $\beta > 0$ such that for all $s_1 \in S$,*

$$E\left[X^{(t)} - X^{(t+1)} \;\middle|\; T > t,\ X^{(t)} = s_1\right] \leq h(\xi(s_1))$$

and

$$\Pr\left[X^{(t+1)} < \xi(X^{(t)}) \;\middle|\; T > t,\ X^{(t)} = s_1\right] \leq 1 \Big/ \left(\beta\left(\frac{s_{\min}}{h(s_{\min})} + \int_{s_{\min}}^{s_1} \frac{\mathrm{d}s}{h(s)}\right)\right),$$

then, for all $s_0 \in S \setminus \{0\}$,

$$E\left[T \;\middle|\; X^{(0)} = s_0\right] \geq \frac{\beta}{1+\beta}\left(\frac{s_{\min}}{h(s_{\min})} + \int_{s_{\min}}^{s_0} \frac{\mathrm{d}s}{h(s)}\right).$$

More specific bounds can be asserted under additional assumptions.

Theorem 7 (Additive Drift Theorem [15]) *Let $S \subseteq \mathbb{R}^+$ be a finite set of positive numbers with minimum $s_{\min}$ and $(X^{(t)})_{t \in \mathbb{N}}$ a sequence of random variables over S, $T = \min\{t \geq 0 \mid X^{(t)} \leq s_{\min}\}$. Suppose there is a positive real number $\delta > 0$ such that*

$$E\left[X^{(t)} - X^{(t+1)} \;\middle|\; T > t\right] \geq \delta.$$

Then, for all $s_0 \in S$,

$$E\left[T \;\middle|\; X^{(0)} = s_0\right] \leq \frac{s_0 - s_{\min}}{\delta}.$$

Conversely, if $E\left[X^{(t)} - X^{(t+1)} \mid T > t\right] \leq \delta$, then $E\left[T \mid X^{(0)} = s_0\right] \geq (s_0 - s_{\min})/\delta$.

Theorem 8 (Multiplicative Drift Theorem for lower bounds [27]) *Let $S \subseteq \mathbb{R}^+$ be a finite set of positive numbers with minimum 1 and $(X^{(t)})_{t \in \mathbb{N}}$ a sequence of random variables over S such that $X^{(t+1)} \leq X^{(t)}$ for all t. Target value $s_{\min} > 0$ is arbitrary and $T = \min\{t \geq 0 \mid X^{(t)} \leq s_{\min}\}$ as above. Suppose there are real numbers $1 \geq \delta, \beta > 0$ such that for all $s \in S$,*

$$E\left[X^{(t)} - X^{(t+1)} \;\middle|\; T > t,\ X^{(t)} = s\right] \leq \delta s \quad \textit{and}$$
$$\Pr\left[X^{(t)} - X^{(t+1)} \geq \beta s \;\middle|\; T > t, X^{(t)} = s \neq 1\right] \leq \frac{\beta\delta}{\ln s}.$$

Then, for all $s_0 \in S$,

$$E\left[T \;\middle|\; X^{(0)} = s_0\right] \geq \frac{1-\beta}{1+\beta} \cdot \frac{\ln(s_0) - \ln(s_{\min})}{\delta}.$$

3 Push Protocol

We start the analysis with a probabilistic approach to migration, namely, the *push protocol*. The migration interval is fixed at value τ, but the transmission itself is randomized. Every island chooses a neighbor uniformly at random and sends its currently best solution to it. Using probabilistic communication is a robust way to save on the communication costs, even in densely connected migration topologies. We prove this exemplarily for the complete graph K_λ on λ vertices. The push protocol is well-analyzed in literature. To bound the expected spreading time we use an adaption of a more general result by Doerr and Künnemann [11]. Symbol $\operatorname{ld} x$ stands for the base-2 logarithm.

Lemma 9 ([11, Lemma 3.3]) *Consider the push protocol on the complete graph K_λ as migration topology. There exists a constant $c \geq 1$ such that $E[S(k)] \leq \operatorname{ld} k + 2$ for all $1 \leq k \leq \lambda/c$.*

Recall that $E[T_{\mathrm{OM}}]$ and $E[C_{\mathrm{OM}}]$ denote the expected optimization time and communication costs, respectively, when optimizing the OneMax fitness function; $E[T_{\mathrm{LO}}]$ and $E[C_{\mathrm{LO}}]$ correspond to LeadingOnes.

Theorem 10 *Consider an island model using the push protocol on the complete graph K_λ to optimize* OneMax. *The expected optimization time and communication costs are*

$$E[T_{\mathrm{OM}}] = \mathrm{O}\left(\frac{n}{\lambda}\log n + n\,\log\tau\right) \quad \textit{and} \quad E[C_{\mathrm{OM}}] = \mathrm{O}\left(\frac{n}{\tau\lambda}\log n + \frac{n}{\tau}\log\tau\right)$$

Proof The bound on $E[C_{\mathrm{OM}}]$ can be obtained from $E[T_{\mathrm{OM}}]$ and the fact that every island sends exactly one message every τ generations. We are thus left with bounding the optimization time. A standard computation shows that the probability p_i (formally defined in Lemma 2) of the (1 + 1) EA finding an improving Hamming-neighbor of an individual $\mathbf{x}$ with $\text{OneMax}(\mathbf{x}) = i$ is at least $(n-i)/(en)$. We want to use the randomized version of Corollary 3 in the proof and thus define a sequence $(\lambda_i)_{0\leq i<n}$. Its members represent the minimum target number of islands to which we want to distribute the best solution. Let constant c be as in Lemma 9 and set

$$\lambda_i = \begin{cases} 1, & \text{if } i < (n - n/\tau); \\ n/(\tau(n-i)), & \text{if } (n - n/\tau) \leq i < (n - cn/(\tau\lambda)); \\ \lambda/c, & \text{otherwise.} \end{cases}$$

We tacitly assume $\lambda/c > 1$; otherwise, λ is a constant and we get the usual $\mathrm{O}(n\log n)$ bound [13]. Let $L_1 = n - n/\tau$ denote the lower limit and $L_2 = n - cn/(\tau\lambda)$ the upper limit of the range of i specified in the second case of above equation. By Lemma 9 and the simple fact that it takes no time to inform yourself, we get the following bounds on the partial spreading times,

$$S(\lambda_i) \leq \begin{cases} 0, & \text{if } i < L_1; \\ \mathrm{ld}(n/(\tau(n-i))) + 2, & \text{if } L_1 \leq i < L_2; \\ \mathrm{ld}(\lambda/c) + 2, & \text{otherwise.} \end{cases}$$

Intuitively speaking, while the fitness $i < L_1$ is small, a single island is capable of making significant progress on its own and does not require any migration. The middle range is chosen in such a way that the value of

$$\tau\, S(\lambda_i) + \frac{1}{p_i\, \lambda_i}$$

stemming from Lemma 2 is minimized (up to constant factors). This balances the time needed to spread good solutions with the waiting time to complete the level. If $i \geq L_2$ is already quite large, we need a lot of generations to make further progress, it is thus beneficial to inform a constant fraction of the islands in the meantime.

Applying Corollary 3 to the sequence $(\lambda_i)_i$ gives

$$E[T_{\mathrm{OM}}] \leq n + \sum_{i=0}^{L_1-1} \left(\frac{en}{n-i}\right) + \sum_{i=L_2}^{n-1} \left(\tau\ \mathrm{ld}\left(\frac{\lambda}{c}\right) + 2\tau + \frac{cen}{\lambda(n-i)}\right)$$
$$+ \sum_{i=L_1}^{L_2-1} \left(\tau\ \mathrm{ld}\left(\frac{n}{\tau(n-i)}\right) + 2\tau + \frac{en}{n-i}\left(\frac{\tau(n-i)}{n}\right)\right).$$

If one of the ranges i is empty, the corresponding partial sum evaluates to 0. In more detail, the following analysis assumes $0 < L_1 < L_2 < n-1$, in addition to $\lambda/c > 1$ this implies $1 < \tau$ and $\tau\lambda < cn$. Observe that even if these constraints are violated, we get an upper bound on the run time. We handle the terms separately.

$$P_1 = \sum_{i=0}^{L_1-1} \left(\frac{en}{n-i}\right) = en \left(\sum_{j=n-L_1+1}^{n} \frac{1}{j}\right) \leq en \left(\ln\left(\frac{n}{n-L_1}\right) + 1\right).$$

The last inequality is due to the estimate $\sum_{j=k}^{n} 1/j \leq \ln(n/(k-1)) + 1$ of the harmonic series for $k > 1$. Substituting $n - L_1 = n/\tau$ gives

$$P_1 \leq en\,(\ln\tau + 1).$$

Regarding the second term, we have

$$P_2 = \sum_{i=L_2}^{n-1} \left(\tau\ \mathrm{ld}\left(\frac{\lambda}{c}\right) + 2\tau + \frac{cen}{\lambda(n-i)}\right)$$

$$= \tau(n - L_2)\left(\mathrm{ld}\left(\frac{\lambda}{c}\right) + 2\right) + \frac{cen}{\lambda}\left(\sum_{j=1}^{n-L_2}\frac{1}{j}\right)$$
$$\leq \tau(n - L_2)\left(\mathrm{ld}\left(\frac{\lambda}{c}\right) + 2\right) + \frac{cen}{\lambda}(\ln(n - L_2) + 1).$$

With $n - L_2 = cn/(\tau\lambda)$ we get

$$P_2 \leq \frac{cn}{\lambda}\left(\mathrm{ld}\left(\frac{\lambda}{c}\right) + e\ \ln\left(\frac{cn}{\tau\lambda}\right) + e + 2\right).$$

Finally, the third term equals

$$P_3 = \sum_{i=L_1}^{L_2-1}\left(\tau\ \mathrm{ld}\left(\frac{n}{\tau(n-i)}\right) + 2\tau + \frac{en}{n-i}\left(\frac{\tau(n-i)}{n}\right)\right)$$
$$= \tau\left(\sum_{j=n-L_2+1}^{n-L_1}\mathrm{ld}\left(\frac{n}{\tau j}\right)\right) + (e+2)\tau(L_2 - L_1).$$

For any integrable decreasing function g, the sum $\sum_{j=a}^{b} g(j)$ can be upper bounded by the integral $\int_{a-1}^{b} g(j)\ \mathrm{d}j$. This is applied to $g(j) = \ln(n/(\tau j))$. Observe that the lower limit of the interval of integration must be by 1 smaller than that of the sum, i.e., $n - L_2 = cn/(\tau\lambda)$. This accounts for the first summand. The integral is computed using the antiderivative $\int g\ \mathrm{d}j = j\ \ln(n/(\tau j)) + j$. Finally, we insert $L_2 - L_1 = (1 - c/\lambda) \cdot n/\tau \leq n/\tau$ to arrive at

$$P_3 \leq \tau\left(\int_{cn/\tau\lambda}^{n/\tau}\mathrm{ld}\left(\frac{n}{\tau j}\right)\mathrm{d}j\right) + (e+2)\,n = \frac{1}{\ln 2}\,n - \frac{cn}{\lambda}\,\mathrm{ld}\left(\frac{\lambda}{c}\right) - \frac{1}{\ln 2}\,\frac{cn}{\lambda} + (e+2)\,n$$
$$\leq (e+4)\,n - \frac{cn}{\lambda}\left(\mathrm{ld}\left(\frac{\lambda}{c}\right) + 1\right).$$

As a result, the $(cn/\lambda) \cdot \mathrm{ld}(\lambda/c)$ terms in P_2 and P_3 cancel out and we get

$$E[T_{\mathrm{OM}}] \leq n + P_1 + P_2 + P_3 \leq en\ \ln\tau + \frac{cn}{\lambda}\left(e\ \ln\left(\frac{cn}{\tau\lambda}\right) + e + 1\right) + (2e+5)\,n$$
$$= \mathrm{O}\left(\frac{n}{\lambda}\log n + n\ \log\tau\right).$$

□

Bounds on the expected optimization time and communication costs yield some insights in how to choose the parameters of the island model. For the push protocol on a complete graph a linear parallel optimization time can be enforced by setting $\lambda = \Omega(\log n)$ and τ a constant. This also minimizes the expected combined costs $E[T_{\mathrm{OM}} + C_{\mathrm{OM}}]$. Observe that the communication costs are always dominated by the optimization time. When choosing a large migration interval of $\tau = \Omega(n)$, for example to reduce communication costs even more, the island model behaves like a single (1 + 1) EA and the influence of migration is diminished.

Lässig and Sudholt [17] consider, in the place of migration intervals, a *migration probability* p with which any two neighboring islands communicate in any given round. When comparing the expected optimization times of both systems, this corresponds to a migration interval of $\tau = 1/(p(\lambda-1))$. Theorem 18 in [17] gives a bound on $E[T_{\mathrm{OM}}]$ that has a linear dependence on $1/p$, Theorem 10 above improves this to logarithmic in τ.

Theorem 11 *Consider an island model using the push protocol on the complete graph K_λ to optimize* LeadingOnes. *The expected optimization time and communication costs are*

$$E[T_{\mathrm{LO}}] = \mathrm{O}\left(\frac{n^2}{\lambda} + \tau n \log n\right) \quad \textit{and} \quad E[C_{\mathrm{LO}}] = \mathrm{O}\left(\frac{n^2}{\tau\lambda} + n \log n\right).$$

Proof It is clearly enough to show the bound on the optimization time. The proof follows the same ideas as that of Theorem 10 but is somewhat simpler. The reason is that for LeadingOnes the bound on the probability p_i does not depend on the current level, and it is always at least $1/en$. Let again $c \geq 1$ be such that $S(k) \leq \mathrm{ld}\, k + 2$ whenever $k \leq \lambda/c$. We choose $\lambda_i = \min\{n/\tau, \lambda/c\}$ for all $0 \leq i \leq n-1$. Assume for the moment that the λ_i defined that way are all at least 1. We split the analysis into two cases. First, suppose that the minimum is obtained at n/τ. From Corollary 3 we get

$$E[T_{\mathrm{LO}}] \leq n + n\left(\tau\, S\left(\frac{n}{\tau}\right) + \frac{en}{n/\tau}\right) \leq n + n\tau\left(\mathrm{ld}\left(\frac{n}{\tau}\right) + 2 + e\right) = \mathrm{O}\left(n\tau \log n\right).$$

Conversely, if $\lambda_i = \lambda/c$, then

$$E[T_{\mathrm{LO}}] \leq n + n\left(\tau\, S\left(\frac{\lambda}{c}\right) + \frac{en}{\lambda/c}\right) \leq n + n\tau\left(\mathrm{ld}\left(\frac{\lambda}{c}\right) + 2\right) + \frac{cen^2}{\lambda}.$$

Using the assumption $\lambda/c \leq n/\tau \leq n$ gives the claimed bound.

If $\min\{n/\tau, \lambda/c\}$ is smaller than 1, we choose $\lambda_i = 1$ instead for all i. Note that this implies $S(\lambda_i) = 0$ as no communication steps are needed. Corollary 3 gives a bound of order $\mathrm{O}(n^2)$, which is proportional to the run time of a single (1 + 1) EA [24]. This can happen in two cases. If $\lambda \leq c$ is a constant, the first term of the bound claimed in the theorem is quadratic. Similarly, if the migration interval $\tau = \Omega(n)$ is too large to benefit the optimization, $\mathrm{O}(n^2)$ is subsumed by $\mathrm{O}(\tau n \log n)$. That completes the proof. □

An expected optimization time of $\mathrm{O}(n \log n)$ can be achieved for LeadingOnes by setting the number of islands to $\lambda = \Omega(n/\log n)$ with a constant migration interval τ. The same parameter setting minimizes the combined costs. The bound given in

Theorem 11 and the run time of $O(n^2/\lambda + n\tau \log \lambda)$ proven in [17] are asymptotically equivalent for all reasonable settings, i.e., when λ is polynomially bounded in n.

The reasoning above holds true for all migration topologies in which the *growth rate* [4] is exponential. That means, the number of informed nodes increases by a constant factor in each round. This connection is independent of the means of communication, randomized or deterministic, and the density of the underlying graph. As an extreme example, we show the following corollary regarding island models on trees.

Corollary 12 *Consider an island model using broadcast communication on the complete binary tree as migration topology. For* OneMax *the expected optimization times are* $E[T_{\mathrm{OM}}] = O\left(\frac{n}{\lambda}\log n + n\log\tau\right)$ *and communication costs* $E[C_{\mathrm{OM}}] = O\left(\frac{n}{\tau\lambda}\log n + \frac{n}{\tau}\log\tau\right)$. *For* LeadingOnes *it is* $E[T_{\mathrm{LO}}] = O\left(\frac{n^2}{\lambda} + \tau n\log n\right)$ *as well as* $E[C_{\mathrm{LO}}] = O\left(\frac{n^2}{\tau\lambda} + n\log n\right)$.

Proof To establish this result, it is enough to show a logarithmic spreading time until at least a constant fraction of the nodes in the tree are informed. In the worst case a leaf node has the initial rumor. Assume the information has traveled for t steps, where t is smaller than the diameter of the tree. After $t/2$ steps an island j of height $t/2$ has been reached, j in turn roots a complete subtree consisting of $2\cdot 2^{t/2} - 1$ nodes which are informed in the remaining $t/2$ steps. Therefore, at least $(\sqrt{2})^t$ islands obtain the rumor in t iterations. For any $1 \le k \le \lambda/2$ we have $S(k) \le \log_{\sqrt{2}} k$, this is within a constant factor of the bound given in Lemma 9. □

4 Multidimensional Tori

In this section we investigate broadcast communication in island models a bit further. Every τ generations each island sends its best solution to *all* of its neighbors simultaneously. The spreading time is now a deterministic function. Also, the communication costs are functionally determined by the optimization time and the structure of the underlying graph. Let $\deg_{av}(G)$ denote the average degree of a given migration topology G, then random variables C and T differ by a factor $\deg_{av}(G)/\tau$. Consequently, we focus on bounding the optimization time.

As a proof of concept, we consider the d-dimensional torus as the migration topology. It can be constructed from a d-grid by connecting the outermost vertices with wrapping edges. More formally, fix two integers $d, \ell \ge 1$ and define the vertex set $V = \{0, \ldots, \ell - 1\}^d$. $\binom{V}{2}$ is the collection of all unordered pairs of vertices. We define the edge set to be $E = \{\{u, v\} \in \binom{V}{2} \mid \exists i : (u_i - v_i = 1 \mod \ell) \wedge (\forall j \ne i : u_j = v_j)\}$. Here, u_i denotes the i-th component of vector u. In one dimension the construction gives a bidirectional ring and in two dimensions the usual torus. A characteristic property of d-tori is the spreading time being in the order of the d-th root.

We would like to point out that throughout this section, d is regarded as a constant independent of n.

Lemma 13 *For broadcast communication on a d-dimensional torus as the migration topology, the spreading time is at most* $S(k) \le d^2 \sqrt[d]{k}$.

Proof In a first step, we show that the spreading time is bounded by $S(k) \leq d\sqrt[d]{k}$ for all $1 \leq k \leq \lambda/(2d)^d$, this is surely enough to establish the claimed looser bound in this regime. To that end, let the *growth rate* $F(t)$ be the total number of informed nodes after t communication steps [4]. We bound F from below in order to get an upper bound on the spreading time. Let d, ℓ be as defined above and fix the value of t. As long as no wrapping edges are involved, the collection of informed nodes make up a d-dimensional diamond shape containing $F(t)$ nodes. W.l.o.g. this polytope is centered at node $\mathbf{0} \in V$. It is bounded by 2^d $(d-1)$-dimensional faces consisting of the islands that have still uninformed neighbors. These are the only ones contributing to the rumor spreading in the next round. To avoid double counting, we only consider a single face, namely, the one pointing in the direction of the all-ones vector $\mathbf{1}$. After t communication steps this face consists precisely of the points $(a_1, a_2, \ldots, a_d) \in V$ that satisfy $a_1 + a_2 + \cdots + a_d = t$. Basic combinatorics tells us that there are $\binom{d+t-1}{d-1}$ many of them. Considering all communication steps leading up to t, we get

$$F(t) \geq \sum_{i=0}^{t} \binom{d+i-1}{d-1} = \binom{d+t}{d} \geq \left(\frac{t}{d}\right)^d.$$

This implies $S(k) = \min\{t \mid F(t) \geq k\} \leq d\sqrt[d]{k} \leq d^2\sqrt[d]{k}$. The condition of not using the wrapping edges is satisfied for the first $\ell/2$ steps. The number of nodes that can be informed in this phase is at least

$$F\left(\frac{\ell}{2}\right) \geq \left(\frac{\ell}{2d}\right)^d = \frac{\lambda}{(2d)^d}.$$

The equality $\ell^d = \lambda$ is due to the construction of graph G.

Now observe that the spreading time $S(\lambda)$ to reach *all* nodes is bounded by the graph diameter $d\ell/2$, and that function S is non-decreasing. Thus, for $\lambda/(2d)^d < k \leq \lambda$, we have

$$S(k) \leq d\,\frac{\ell}{2} = d\,\frac{\sqrt[d]{\lambda}}{2} < d\,\frac{\sqrt[d]{(2d)^d k}}{2} = d^2\sqrt[d]{k}.$$

□

Theorem 14 *Consider an island model using broadcast communication on a d-dimensional torus to optimize* OneMax. *The expected optimization time and communication costs are*

$$E[T_{\mathrm{OM}}] = \mathrm{O}\left(\frac{n}{\lambda}\log n + n\,\log\tau\right) \quad \textit{and} \quad E[C_{\mathrm{OM}}] = \mathrm{O}\left(\frac{n}{\tau\lambda}\log n + \frac{n}{\tau}\log\tau\right).$$

Proof The communication costs are by a factor $2d/\tau$ larger than the optimization time as the d-torus is a $2d$-regular graph. Recall that d is a constant.

We now prove the claimed bound on $E[T_{\mathrm{OM}}]$. The proof follows the same structure as the one of Theorem 10. Lemma 13 asserts a spreading time of at most $d^2\sqrt[d]{k}$. Using

$p_i \geq (n-i)/(en)$, we choose the sequence $(\lambda_i)_i$ as

$$\lambda_i = \begin{cases} 1, & \text{if } i < (n - n/\tau); \\ (n/(\tau(n-i)))^{\frac{d}{d+1}}, & \text{if } (n-n/\tau) \leq i \leq \left(n - (n/\tau)\cdot(1/\lambda)^{\frac{d+1}{d}}\right); \\ \lambda, & \text{otherwise.} \end{cases}$$

Let L_1 denote the lower limit and L_2 the upper limit of the middle range above. We get the following upper bounds on the (deterministic) spreading time.

$$S(\lambda_i) \leq \begin{cases} 0, & \text{if } i < L_1; \\ d^2 \sqrt[d+1]{n/(\tau(n-i))}, & \text{if } L_1 \leq i \leq L_2; \\ d^2 \sqrt[d]{\lambda}, & \text{otherwise.} \end{cases}$$

Corollary 3 now yields

$$E[T_{\mathrm{OM}}] \leq n + \sum_{i=0}^{L_1-1}\left(\frac{en}{n-i}\right) + \sum_{i=L_2}^{n-1}\left(d^2\tau\sqrt[d]{\lambda} + \frac{en}{\lambda(n-i)}\right)$$
$$+ \sum_{i=L_1}^{L_2-1}\left(d^2\tau\left(\frac{n}{\tau(n-i)}\right)^{\frac{1}{d+1}} + \frac{en}{n-i}\left(\frac{\tau(n-i)}{n}\right)^{\frac{d}{d+1}}\right).$$

The partial sums can again be bounded by elementary means.

$$P_1 = \sum_{i=0}^{L_1-1}\left(\frac{en}{n-i}\right) \leq en\left(\ln\left(\frac{n}{n-L_1}\right)+1\right) = en\,(\ln\tau + 1) \quad \text{and}$$
$$P_2 = \sum_{i=L_2}^{n-1}\left(d^2\tau\sqrt[d]{\lambda} + \frac{en}{\lambda(n-i)}\right) \leq (n-L_2)\, d^2\tau\sqrt[d]{\lambda} + \frac{en}{\lambda}(\ln(n-L_2)+1).$$

With $n - L_2 = (n/\tau)\cdot(1/\lambda)^{\frac{d+1}{d}}$ we get

$$P_2 \leq \frac{d^2 n}{\lambda} + \frac{en}{\lambda}\left(\ln\left(\frac{n}{\tau}\right) + \frac{d+1}{d}\ln\left(\frac{1}{\lambda}\right) + 1\right) \leq \frac{n}{\lambda}(e\,\ln n + d^2 + 1).$$

In the simplification of the last sum, we benefit greatly from the fact that the λ_i can be chosen freely. They are such that in the middle range $\tau\, S(\lambda_i)$ and $1/(p_i\lambda_i)$ differ only a constant factor. Namely, the former has coefficient d^2, while it is e for the latter.

$$P_3 = \sum_{i=L_1}^{L_2-1}\left(d^2\tau\left(\frac{n}{\tau(n-i)}\right)^{\frac{1}{d+1}} + \frac{en}{n-i}\left(\frac{\tau(n-i)}{n}\right)^{\frac{d}{d+1}}\right)$$

$$= (d^2 + e)\, \tau^{\frac{d}{d+1}} n^{\frac{1}{d+1}} \left(\sum_{i=L_1}^{L_2-1} (n-i)^{-\frac{1}{d+1}} \right)$$
$$= (d^2 + e)\, \tau^{\frac{d}{d+1}} n^{\frac{1}{d+1}} \left(\sum_{j=n-L_2+1}^{n-L_1} j^{-\frac{1}{d+1}} \right).$$

The terms of the last sum decrease in j, thus we use an integral to bound it.

$$P_3 \le (d^2 + e)\, \tau^{\frac{d}{d+1}} n^{\frac{1}{d+1}} \left(\int_{\frac{n}{\tau}\left(\frac{1}{\lambda}\right)^{\frac{d+1}{d}}}^{\frac{n}{\tau}} j^{-\frac{1}{d+1}} \,\mathrm{d}j \right)$$
$$= (d^2 + e)\, \tau^{\frac{d}{d+1}} n^{\frac{1}{d+1}} \left(\left(\frac{n}{\tau}\right)^{\frac{d}{d+1}} \cdot \frac{d+1}{d} \left(1 - \frac{1}{\lambda}\right) \right)$$
$$\le (d^2 + d + 2e)\, n.$$

The three estimates together finally yield the claimed run time,

$$E[T_{\mathrm{OM}}] \le n + P_1 + P_2 + P_3 = \mathrm{O}\left(\frac{n}{\lambda} \log n + n \log \tau\right).$$

□

Theorem 15 *Consider an island model using broadcast communication on a d-dimensional torus to optimize* LeadingOnes*. The expected optimization time and communication costs are*

$$E[T_{\mathrm{LO}}] = \mathrm{O}\left(\frac{n^2}{\lambda} + \tau^{\frac{d}{d+1}} n^{\frac{d+2}{d+1}}\right) \quad \textit{and} \quad E[C_{\mathrm{LO}}] = \mathrm{O}\left(\frac{n^2}{\tau\lambda} + \tau^{-\frac{1}{d+1}} n^{\frac{d+2}{d+1}}\right).$$

Proof The proof follows the same line as that of Theorem 11. Recall that $p_i \ge 1/en$ holds for all fitness levels i. We choose $\lambda_i = \min\{(n/\tau)^{\frac{d}{d+1}}, \lambda\}$ for all i. If $\lambda_i = (n/\tau)^{\frac{d}{d+1}}$, we get

$$E[T_{\mathrm{LO}}] \le n + n \left(d^2 \tau \left(\frac{n}{\tau}\right)^{\frac{1}{d+1}} + en \left(\frac{\tau}{n}\right)^{\frac{d}{d+1}} \right) = n + (d^2 + e)\, \tau^{\frac{d}{d+1}} n^{\frac{d+2}{d+1}}.$$

In case of $\lambda_i = \lambda$,

$$E[T_{\mathrm{LO}}] \le n + n \left(d^2 \tau \sqrt[d]{\lambda} + \frac{en}{\lambda} \right).$$

Note that $\lambda_i = \lambda$ implies $\lambda \le (n/\tau)^{\frac{d}{d+1}}$, by the definition of the λ_i. Using this bound twice gives

$$\tau \sqrt[d]{\lambda} \le \tau \left(\frac{n}{\tau}\right)^{\frac{1}{d+1}} = n \left(\frac{n}{\tau}\right)^{-\frac{d}{d+1}} \le \frac{n}{\lambda}.$$

Finally, we substitute this estimate into our bound on $E[T_{\text{LO}}]$,

$$E[T_{\text{LO}}] \leq n + \frac{(d^2 + e)\, n^2}{\lambda}.$$

If the minimum above were to be smaller than 1, we choose $\lambda_i = 1$ instead, which gives a bound of $\mathrm{O}(n^2)$. Necessarily, we have $\tau \geq n$ or $\lambda = 1$, whence $E[T_{\text{LO}}]$ is of order $\mathrm{O}\left(n^2/\lambda + \tau^{\frac{d}{d+1}}\, n^{\frac{d+2}{d+1}}\right)$ in this case as well. □

When the goal is to minimize the optimization time, it is best to choose $\lambda = \Omega(\log n)$ for OneMax, which gives $E[T_{\text{OM}}] = \mathrm{O}(n)$, and $\lambda = \Omega(n^{d/(d+1)})$ for LeadingOnes to reach $E[T_{\text{LO}}] = \mathrm{O}(n^{(d+2)/(d+1)})$. In both cases one should set the length τ of the communication interval to a constant. These assertions also extend to the respective combined optimization costs.

Theorem 14 improves on the bounds by Lässig and Sudholt for OneMax, after equating τ with $1/(2dp)$, in showing that the dependency of the optimization time on τ is logarithmic instead of $\sqrt{\tau}$ on a ring topology or $\sqrt{\tau^3}$ on a 2-dimensional torus [17]. Theorem 15 reproves their results for LeadingOnes on the ring and torus, and generalizes it to arbitrary dimensions $d \geq 1$.

It seems to be counter-intuitive at first that the dimension does not influence the (asymptotic) optimization time for OneMax compared to the push protocol on a complete graph, although the growth rate is only polynomial instead of exponential. The reason is that the time the algorithm spends in the middle range, which is influenced by d the most, is dominated by the early levels in which the islands try to find independent improvements and the later stages where a coordinated effort is necessary for any progress. Also, a careful analysis shows that many complementary effects of the dimension on the middle range cancel each other out, e.g. the number of levels and the time spend in each of them. For LeadingOnes this separation into ranges does not exist since each level is equally likely to leave. Here, the influence of the dimension is much stronger throughout the whole optimization process. Since more nodes can be reached in the same number of communication steps, the influence of the problem size n on the runtime decreased when d grows larger. On the contrary, the waiting time between these steps, τ, emerges as a bottle-neck in higher dimensions.

In every connected topology, while not all nodes are informed, the rumor is spread to at least one new island in every round. This translates to a spreading time of $S(k) \leq k$ which is equal to the bound given in Lemma 13 for $d = 1$. Moreover, for every unimodal function the number of increasing Hamming-neighbors is at least as large as in the case of LeadingOnes. Hence, the bound given in Theorem 15 extends to all unimodal functions when adapted to the number of possible fitness values. Similar observations were made by Lässig and Sudholt [17] as well as Badkobeh et al. [4].

Corollary 16 *Let $f : \{0,1\}^n \to \mathbb{R}_0^+$ be a unimodal function with $|\mathrm{img}(f)| = v$ fitness values. Consider an island modal using broadcast communication on a connected graph on λ vertices to optimize f. The expected optimization time and communication* *costs are*

$$E[T] = \mathrm{O}\left(\frac{vn}{\lambda} + v\sqrt{\tau n}\right) \quad \textit{and} \quad E[C] = \mathrm{O}\left(\frac{vn}{\tau\lambda} + v\sqrt{\frac{n}{\tau}}\right).$$

Under the respective optimal parameter choices, $E[T]$ *and* $E[T + C]$ *are both of order* $\mathrm{O}(v\sqrt{n})$.

5 The Complete Graph

Finally, we discuss the special case of broadcast communication on a complete graph as the migration topology. This way we reach the maximum number of nodes in one communication step. Conversely, the spreading time degenerates into a step function. In the majority of iterations the islands compute their local improvements in total isolation. However, periodically *all* islands are updated to a maximum-fitness solution in a network-spanning communication effort only to be left alone for another phase of τ generations.

If the migration interval is chosen too large, namely, if $\tau > en$, one can expect a significant portion of the islands to find an improving Hamming-neighbor without migration (especially on OneMax-like functions, where there are many of them). Communication between nodes becomes obsolete, eroding the characteristics of an island model. This can also be seen from the Theorems 10 through 14. For $\tau = \Omega(n)$ the parallel optimization time of any island model is in the same complexity as the run time of a simple (1 + 1) EA on the same fitness function. Consequently, we assume $\tau \leq en$ throughout this section. For the other extreme of $\tau = 1$, it has been pointed out that the island model using broadcast on a complete graph shares many traits with the (1+λ) Evolutionary Algorithm [17]. The only distinction is that the island model can store different solutions of the same maximum fitness. Tight run time bounds for the (1+λ) EA on OneMax are known [12]. Hence, by characterizing the optimization time of the island model we can precisely quantify the influence of the migration interval τ.

In this section we give tight bounds for the expected time to optimize OneMax as well as upper and lower bounds for LeadingOnes. The expected communication costs can be obtained from this value by multiplying it with a factor $(\lambda - 1)/\tau$ since every islands sends a message to each neighbor every τ iterations. Although these bounds will be proven for the K_λ as migration topology, the lower bounds extend to any connected graph. This is due to the fact that additional informed islands can only benefit the optimization and no topology spreads solutions faster than the complete graph.

5.1 OneMax

Theorem 17 *Consider an island model using broadcast communication with a migration interval* $\tau \leq en$ *on the complete graph* K_λ *to optimize* OneMax. *The optimization time depends on the relation between* τ *and* λ.

If $\tau = \mathrm{o}(\log \lambda)$*,*

$$E[T_{\mathrm{OM}}] = \Theta\left(\frac{n}{\lambda}\log n + \frac{\tau n}{\log \lambda}\log\left(\frac{\log \lambda}{\tau}\right)\right) \quad \textit{and}$$
$$E[C_{\mathrm{OM}}] = \Theta\left(\frac{n}{\tau}\log n + \frac{\lambda n}{\log \lambda}\log\left(\frac{\log \lambda}{\tau}\right)\right).$$

If $\tau = \Theta(\log \lambda)$*,*

$$E[T_{\mathrm{OM}}] = \Theta\left(\frac{n}{\lambda}\log n + n\right) \quad \textit{and}$$
$$E[C_{\mathrm{OM}}] = \Theta\left(\frac{n}{\tau}\log n + \frac{\lambda n}{\tau}\right).$$

If $\tau = \omega(\log \lambda)$*,*

$$E[T_{\mathrm{OM}}] = \Theta\left(\frac{n}{\lambda}\log n + n\,\log\left(\frac{\tau}{\log \lambda}\right)\right) \quad \textit{and}$$
$$E[C_{\mathrm{OM}}] = \Theta\left(\frac{n}{\tau}\log n + \frac{\lambda n}{\tau}\,\log\left(\frac{\tau}{\log \lambda}\right)\right).$$

In the remainder of this section we prove the various bounds on the expected optimization time given in the theorem, the communication costs follow from this. We start with the upper bounds. We need the following two lemmas. The first one is another tail bound on the binomial distribution specifically tailored to our application. The second lemma by Doerr [7] asserts that the (1 + 1) EA is unlikely to make any large jumps when optimizing unimodal functions.

Lemma 18 *Let* d *be a positive integer and* $\varepsilon > 0$ *a constant. For a probability* $0 < p < 1$*, let* $X \sim \mathrm{Bin}(d, p)$ *be a binomially distributed random variable with parameters* d *and* p*. Set* $q = (1-p)^d$*. If* $\ln(q/\varepsilon) \geq epd$*, then*

$$\Pr\left[X \geq \frac{\ln(q/\varepsilon)}{\ln\left(\frac{\ln(q/\varepsilon)}{pd}\right)}\right] \geq \varepsilon.$$

Proof Set $r = \ln(q/\varepsilon)/(pd)$ and $t = \ln(q/\varepsilon)/\ln(r)$. We get the following useful identity,

$$\frac{pd}{t} = \frac{pd \cdot \ln r}{\ln(q/\varepsilon)} = \frac{\ln r}{r}.$$

To prove the lemma, we have to bound the probability $\Pr[X \geq t]$ from below.

$$\Pr[X \geq t] \geq \binom{d}{t} p^t (1-p)^{d-t} \geq q\left(\frac{pd}{t}\right)^t = q\,\exp\left(t\,\ln\left(\frac{pd}{t}\right)\right)$$

Substituting above identity and the definition of t into this bound yields

$$\begin{aligned}\Pr[X \geq t] &\geq q\ \exp\left(\frac{\ln(q/\varepsilon)}{\ln r} \cdot \ln\left(\frac{\ln r}{r}\right)\right) = q\ \exp\left(\frac{\ln(q/\varepsilon)}{\ln r}(\ln\ln r - \ln r)\right)\\ &\geq q\ \exp\left(-\ln(q/\varepsilon)\right) = \varepsilon.\end{aligned}$$

For the last estimate, observe that $\ln(q/\varepsilon) \geq epd$ implies $\ln\ln r \geq 0$. □

Lemma 19 ([7, Proposition 9]) *For a bit string* $\mathbf{x} \in \{0,1\}^n$*, let* $|\mathbf{x}|_0 = n - \textsc{OneMax}(\mathbf{x})$ *denote the number bits set to* 0 *and* $\mathbf{x}'$ *an offspring obtained from* $\mathbf{x}$ *by standard bit mutation. If* $k \leq 3n/5$ *is a positive integer, then*

$$\Pr\left[|\mathbf{x}'|_0 \leq \frac{k}{2}\ \middle|\ |\mathbf{x}|_0 = k\right] = \exp(-\Omega(k)).$$

Proof of the upper bounds of Theorem 17 Since we aim to optimize OneMax, we call the number $|\mathbf{x}|_0$ of remaining 0-bits (formally defined in Lemma 19) the *distance* to the optimum. We divide the optimization process into three phases depending on the minimal distance over all islands. The cutoff points between the phases are at $d_0 = \min\{n,\ n \ln\lambda/(2e\tau)\}$ and $d_1 = n/(\tau \ln\lambda)$. Note that the first phase is only relevant if $2e\tau > \ln\lambda$. In a first step, we estimate the time it takes for Algorithm 1 to achieve a distance of at most d_0. It is bounded above by the time a single island needs to reach this distance. Thus, we can resort to the well-known run time bounds for the (1 + 1) EA, cf. e.g. [10]. They imply that, if $2e\tau > \ln\lambda$, the length of the first phase is of order

$$\mathrm{O}\left(n\ \log\left(\frac{n}{d_0}\right)\right) = \mathrm{O}\left(n\ \log\left(\frac{\tau}{\log\lambda}\right)\right).$$

Before considering the second phase between distances d_0 and d_1, we analyze the third phase of optimization from d_1 all the way to the optimum. To that end, suppose one island has a non-optimal solution of distance $d \leq d_1$. Using Corollary 3 adapted to the landscape of the d_1 fitness levels in question, we see that the expected remaining optimization time is at most

$$\sum_{d=1}^{d_1}\left(1 + \tau + \frac{en}{d\lambda}\right) \leq d_1(1+\tau) + \frac{en}{\lambda}(\ln d_1 + 1) = \mathrm{O}\left(\frac{n}{\log\lambda} + \frac{n}{\lambda}\log n\right).$$

Now we turn to the more involved phase of optimization between distances d_0 and d_1. Let $d_0 > d > d_1$ be the maximum fitness. When considering a point in the optimization where migration just occurred, all islands have an individual of fitness d, and τ rounds of evolution without migration will follow. The probability to gain a specific 1-bit one a single island in any given iteration is at least $1/(en)$. The probability that it is *not* gained during τ iterations is at most $(1 - 1/(en))^\tau \leq 1 - \tau/(2en)$. We used $\tau \leq en$ here. For technical reasons that will be explained later, we pessimistically consider only $d_{\min} = \min\{d, 2en/\tau\}$ bits as candidates for improvement. The random number of gained bits, out of the $d_{\min}$ advantageous ones, during τ iterations first-order stochastically dominates the binomial variable $X \sim \mathrm{Bin}(d_{\min}, \tau/(2en))$. We employ Lemma 18 to derive a lower bound on X and hence on the possible improvements. We set the threshold ε to $1/\lambda$. This way, with constant probability at least one island in the whole model that has made the progress described by the lemma. Namely,

$$\ln(q\lambda) \Big/ \ln\left(\frac{2en\ \ln(q\lambda)}{\tau\ d_{\min}}\right) \quad \text{where} \quad q = \left(1 - \frac{\tau}{2en}\right)^{d_{\min}}.$$

Note that $(1 - \tau/(2en))^d$ is sub-constant if $d > 2en/\tau$ is too large; thus, our choice of considering at most $2en/\tau$ many missing bits ensures that q is bound below by a positive constant. The upper bound $q \leq 1$ is trivial. Lemma 18 additionally requires $\ln(q\lambda) \geq \tau\ d_{\min}/(2n)$, which is provided by $d_{\min} \leq d_0 \leq n\ \ln\lambda/(2e\tau)$.

Next, we derive a lower bound on the progress in terms of d, instead of $d_{\min}$. Observe that $d \leq d_0$ implies both $d_{\min} \geq d/\ln\lambda$ and $n\ \ln\lambda/(\tau d) \geq 2e$. Hence,

$$\ln\left(\frac{2en\ \ln(q\lambda)}{\tau\ d_{\min}}\right) \leq \ln\left(\frac{2en\ \ln^2\lambda}{\tau d}\right) \leq \ln(2e) + 2\ \ln\left(\frac{n\ \ln\lambda}{\tau d}\right) \leq 3\ \ln\left(\frac{n\ \ln\lambda}{\tau d}\right).$$

Finally, we have proven the existence of a positive constant C such that the *expected* progress over all islands in τ iterations between subsequent migrations starting from distance d is at least $C \cdot h(d)$, where

$$h(d) = \ln\lambda \Big/ \ln\left(\frac{n\ \ln\lambda}{\tau d}\right)$$

Since we are only interested in asymptotic bounds, we will omit the constant C in what follows.

We now bring in the drift analysis mentioned in the preliminary Sect. 2.2. The distance naturally lends itself as the potential function. Observe that the expected drift h of the potential is monotonically increasing in d and $1/h$ is integrable over the interval $[d_0,\ d_1]$. The Variable Drift Theorem for upper bounds (Theorem 5) yields an expected optimization time of at most

$$\begin{aligned} \tau\left(\frac{1}{h(d_1)} + \int_{d_1}^{d_0} \frac{\mathrm{d}x}{h(x)}\right) &= \frac{2\tau\ \ln\ln\lambda}{\ln\lambda} + \frac{\tau}{\ln\lambda}\left(\int_{d_1}^{d_0} \ln\left(\frac{n\ln\lambda}{\tau x}\right)\mathrm{d}x\right) \\ &= \frac{2\tau\ \ln\ln\lambda}{\ln\lambda} - \frac{\tau}{\ln\lambda}\left(\int_{d_1}^{d_0} \ln\left(\frac{\tau x}{n\ln\lambda}\right)\mathrm{d}x\right). \end{aligned}$$

For the moment, we focus on the second term and employ integration by substitution with $y(x) = \tau x/(n\ \ln\lambda)$,

$$-\frac{\tau}{\ln\lambda}\left(\int_{d_1}^{d_0}\ln\left(\frac{\tau x}{n\ln\lambda}\right)\mathrm{d}x\right) = -\frac{\tau}{\ln\lambda}\frac{n\ln\lambda}{\tau}\left(\int_{\tau d_1/(n\ \ln\lambda)}^{\tau d_0/(n\ \ln\lambda)}\ln y\ \mathrm{d}y\right)$$
$$= -n\left(\int_{1/\ln^2\lambda}^{\tau d_0/(n\ \ln\lambda)}\ln y\ \mathrm{d}y\right).$$

An antiderivative of $\ln y$ is $y(\ln(y)-1)$. In the case of $2\tau \le \ln\lambda$, we get $d_0 = n$ and thus resolve the integral to

$$-n\Big[y(\ln(y)-1)\Big]_{1/\ln^2\lambda}^{\tau/\ln\lambda} = \frac{n}{\ln\lambda}\left(\tau\left(\ln\left(\frac{\ln\lambda}{\tau}\right)+1\right)-\frac{2\ln\ln\lambda+1}{\ln\lambda}\right).$$

Also, the additive term $\tau/h(d_1) \le \ln\ln\lambda$ is small enough. Putting together the bounds on the second and third phase (the first one is irrelevant in this case) proves the run time claimed in the first case of Theorem 17. If $2\tau > \ln\lambda$, then $d_0 = n\,(\ln\lambda)/2\tau$ and $\tau/h(d_1) \le 2\tau$. For the integral we get

$$-n\Big[y(\ln(y)-1)\Big]_{1/\ln^2\lambda}^{1/2} = n\left(\frac{1}{2}(1+\ln 2)-\frac{2\ln\ln\lambda+1}{(\ln\lambda)^2}\right),$$

which is of order $\mathrm{O}(n)$. Combined with the bounds on all three phases this gives the last case of the theorem. If τ and $\ln\lambda$ are within constant factors of each other, then the asymptotic bounds proven above coalesce and the middle case follows as well. □

We turn to the lower bounds of Theorem 17. The next lemma prepares an inequality we use in their proof.

Lemma 20 *Let* $r(d) = \ln(n\ \ln\lambda/\tau d)$ *and* $h(d) = \ln\lambda/r(d)$ *be two functions. If* $0 < d < n\,(\ln\lambda)/\tau$, *then*

$$\ln\left(\frac{n\,h(d)}{\tau d}\right) \ge \frac{r(d)}{2}.$$

Proof The restrictions on d are such that $r(d)$ and $h(d)$ exist and are positive. Hence, $\ln(r(d)) \le r(d)/2$. Inserting this into the quantity at hand gives

$$\ln\left(\frac{n\,h(d)}{\tau d}\right) = \ln\left(\frac{n\ln\lambda}{\tau d\,r(d)}\right) = \ln\left(\frac{n\ \ln\lambda}{\tau d}\right)+\ln\left(\frac{1}{r(d)}\right) = r(d) - \ln(r(d)) \ge \frac{r(d)}{2}.$$

□

Proof of the lower bounds of Theorem 17 It is straightforward to prove a lower bound of $\Omega(n\,(\log n)/\lambda)$ from the observation that the *unary unbiased black-box complexity* of OneMax is $\Omega(n\ \log n)$ [18]. The (1 + 1) EA, as any unbiased black-box algorithm, needs at least an expected number of $\Omega(n\ \log n)$ fitness evaluations to optimize OneMax. So λ copies of it need $\Omega(n\,(\log n)/\lambda)$ generations to provide this many evaluations.

For the remaining terms we examine partial phases of the optimization process which together have the claimed run time as a lower bound. In order to see that none of these phases is skipped, we use Lemma 19. First, assume $\tau = \omega(\log \lambda)$. We show that the expected time it takes until any island samples a solution with fewer than $n\,(\mathrm{ld}\,\lambda)/\tau$ bits set to 0 for the first time is of order $\Omega(n\ \log(\tau/\log \lambda))$. This establishes the bound in this case. Again, consider the progress of a single island in the τ iterations between migrations. Suppose the current solution has distance $d \geq n\,(\mathrm{ld}\,\lambda)/\tau$ to the optimum. Witt has shown that the expected progress in one generation is at most d/n [27, Lemma 6.7]. Therefore, within τ iterations the expected progress is at most $\tau d/n \geq \mathrm{ld}\,\lambda$. Let $C \geq 2e$ be a constant. By the Chernoff bound in Theorem 4, the probability of a progress of at least $C\tau d/n$ is at most $2^{-C\tau d/n} \leq 2^{-C\,\mathrm{ld}\,\lambda} = \lambda^{-C}$. In particular, the probability that there is an island that makes this much progress is smaller than λ^{-C+1}. This can be seen by a simple union bound over the λ islands. Hence, the *maximum* expected progress of the whole island model between migrations is $c\tau d/n$, for some constant $c > 0$.

We now invoke the Multiplicative Drift Theorem for lower bounds (Theorem 8) with parameters $s_0 = 2n/5$, $s_{\min} = n\,(\mathrm{ld}\,\lambda)/\tau$, $\delta = c\tau/n$, and $\beta = 1/2$. It has several prerequisites that need to be checked. By Theorem 4, a randomly initialized bit string has at least s_0 bits set to 0 with probability exponentially close to 1. The definition of the distance ensures that the random process over $S = \{s_{\min}, s_{\min} + 1, \ldots, s_0\}$ is non-increasing. Finally, Lemma 19 implies that the condition on the probability of large jumps is met. We get that the expected number of iterations the island model takes to optimize a random bit string into one with at most $s_{\min}$ 0-bits is at least

$$\begin{aligned}\tau \cdot \frac{1-\beta}{1+\beta} \cdot \frac{\ln(s_0) - \ln(s_{\min})}{\delta} &= \tau \cdot \frac{1}{3} \cdot \frac{n}{c\tau} \cdot \left(\ln\left(\frac{2n}{5}\right) - \ln\left(\frac{n\ \mathrm{ld}\,\lambda}{\tau}\right)\right) = \frac{n}{3c}\ \ln\left(\frac{2\tau}{5\ \mathrm{ld}\,\lambda}\right)\\ &= \Omega\left(n\ \log\left(\frac{\tau}{\log\lambda}\right)\right).\end{aligned}$$

Next we consider the case of $\tau = \Theta(\log \lambda)$, where we want to show a bound of $\Omega(n)$. To that end, we measure the time Algorithm 1 takes to get from distance $n\,(\mathrm{ld}\,\lambda)/\tau$ to distance $n\,(\mathrm{ld}\,\lambda)/2\tau$ from the optimum. The reasoning is similar as above. Suppose the number of bits set to 0 in the currently best individual is still $d \geq n\,(\mathrm{ld}\,\lambda)/2\tau$. In expectation the fitness on this island improves by at most $d\tau/n \geq (\mathrm{ld}\,\lambda)/2$ within τ rounds and for $C \geq 2e$, the probability that the island makes progress of at least $C\tau d/n$ is at most $2^{-C\tau d/n} \leq \lambda^{-C/2}$. Using the assumption $d \leq n\,(\mathrm{ld}\,\lambda)/\tau$, we get a maximum expected progress of $c\,\mathrm{ld}\,\lambda$, $c > 0$ a constant, over all λ islands and τ iterations. The lower bound of the Additive Drift Theorem (Theorem 7) implies $\Omega(n)$ generations are needed to find a solution of distance at most $n\,(\mathrm{ld}\,\lambda)/2\tau$.

Finally, assume $\tau = o(\log \lambda)$. We show a lower bound of $\Omega(\frac{n\tau}{\log\lambda}\log(\frac{\log\lambda}{\tau}))$. This time we consider the range between distance $n\,(\ln \lambda)/\tau$ and n/τ to any optimum. Suppose the current best individual of an island has $d \geq n/\tau$ 0-bits. We define

$$h(d) = \ln\lambda \Big/ \ln\left(\frac{n\ \ln\lambda}{\tau d}\right).$$

Let $r(d)$ abbreviate $\ln(n(\ln\lambda)/\tau d)$ such that $h(d) = (\ln\lambda)/r(d)$. These definitions are chosen such that they meet the requirements of Lemma 20. In this case, we choose constant C to be at least e. The expected progress of a single island is at most $\tau d/n$ in τ iterations. The first clause of Theorem 4 with $(1+\delta) = Cn\,h(d)/(\tau d)$ shows that the probability of any island making progress of more than $C\,h(d)$ is at most

$$\exp\left(-\ln\left(\frac{Cn\,h(d)}{e\tau d}\right) C\,h(d)\right) \leq \exp\left(-\ln\left(\frac{n\,h(d)}{\tau d}\right) C\,h(d)\right).$$

Lemma 20 gives a bound in terms of λ only,

$$\exp\left(-\ln\left(\frac{n\,h(d)}{\tau d}\right) C\,h(d)\right) \leq \exp\left(-\frac{r(d)}{2}\cdot\frac{C\,\ln\lambda}{r(d)}\right) = \exp\left(-\frac{C}{2}\,\ln\lambda\right) = \lambda^{-\frac{C}{2}}.$$

Once again we conclude that the maximum progress of λ islands in τ iterations is of order $\mathrm{O}(h(d))$. With the Variable Drift Theorem for lower bounds (Theorem 6) we employ the same integration method as for the upper bound to get a matching run time. This completes the proof of Theorem 17. $\square$

The tight bounds on the expected optimization time translate to the following optimal parameter setting. To minimize the parallel optimization time one should choose $\lambda = n^{\Theta(1)}$ and τ a constant to obtain $E[T_{\mathrm{OM}}] = \Theta(n\,(\log\log n)/\log n)$. Note that artificially small *parallel* optimization times could be achieved by choosing λ super-polynomially large in the solution size n. However, we discard this choice as this would nullify any hope of implementing the island model on any practical computation system. Furthermore, the communication overhead and the sequential run time would also grow super-polynomially. To optimize for the combined costs $E[T_{\mathrm{OM}} + C_{\mathrm{OM}}]$ is thus much more reasonable. Here, the ideal choice is $\lambda = \Theta(\log n)$ and $\tau = \Theta(\log n)$ which leads to a common bound of $\Theta(n\,\log\log n)$. Combined costs this low are only possible employing the third case of the theorem, where this parameter setting yields the best possible combined costs.

Corollary 21 *Let λ be bounded above by a polynomial in n. Consider an island model using broadcast communication with a migration interval $\tau \leq en$ on the complete graph K_λ to optimize* OneMax. *The best-possible expected optimization time is $E[T_{\mathrm{OM}}] = \Theta\left(n\frac{\log\log n}{\log n}\right)$. The best expected combined costs are $E[T_{\mathrm{OM}} + C_{\mathrm{OM}}] = \Theta(n\,\log\log n)$.*

5.2 LeadingOnes

In the remainder of this paper we derive upper and lower bounds on the optimization time and communication costs in the case of LeadingOnes on the complete graph. The bounds are within poly-logarithmic factors of each other and hence give a good picture of the run time complexity in this setting. Following the discussion above, we only consider the case in which λ is polynomially bounded. Note that Lässig and Sudholt showed the same upper bounds in [17].

Theorem 22 *Consider an island model using broadcast communication with a migration interval $\tau \leq en$ on the complete graph K_λ to optimize* LEADINGONES. *The expected optimization time and communication costs are*

$$E[T_{\mathrm{LO}}] = \mathrm{O}\left(\frac{n^2}{\lambda} + \tau n\right) \quad \textit{and} \quad E[C_{\mathrm{LO}}] = \mathrm{O}\left(\frac{n^2}{\tau} + \lambda n\right).$$

Proof As $p_i = 1/en$ regardless of the fitness i, we choose $(\lambda_i)_{0 \leq i < n}$ all equal to λ and get $S(\lambda_i) = 1$ for the spreading time. The theorem now follows from Corollary 3. □

As a sidenote, the proof above also holds for $\tau > en$, but in this case using a single $(1+1)$ EA is better than any island model since a quadratic run time is never worse than $\mathrm{O}(n\tau)$.

We proceed to prove a lower bound on the expected optimization time. The main difficulty in applying fitness-level arguments to *lower* bounds is the possibility that the optimization process may skip several levels with a single improvement. In particular, this is a problem on the complete graph since a large number of computational nodes search for improvements in parallel and then, after migration, the *globally* best solution is provided to all islands. We handle this issue by combining a poly-logarithmic number of consecutive levels to one block; skipping a block then is unlikely. The same technique has been used by Badkobeh et al. [4].

Lemma 23 *Consider an island model using broadcast communication with a migration interval $\tau \leq en$ on the complete graph K_λ to optimize* LEADINGONES. *If λ is bounded above by a polynomial in n, then there exists a constant $c > 0$ such that the probability of any island finding a fitness improvement of more than $c \ln^2 n$ between consecutive migrations is* $\mathrm{o}(1/n)$.

Proof Let the natural number k be such that $\lambda \leq n^k$. As no migration takes place we can focus on the optimization process on a single island. We call an iteration in which the $(1+1)$ EA increases the fitness of the currently best solution an *essential step*. Suppose individual $\mathbf{x}$ is of fitness LEADINGONES$(\mathbf{x}) = i$. It is well known that the bits $\mathbf{x}_{i+2}$ through $\mathbf{x}_n$ form a random substring distributed uniformly over $\{0,1\}^{n-i-1}$. This implies that, in any essential step and for any $1 \leq j \leq n-i$, the probability of a fitness improvement of j is 2^{-j} [5,13]. Hence, there is a constant $c_1 > 0$, namely, $c_1 = (k+2+\varepsilon)/\ln 2$ for some $\varepsilon > 0$, such that the probability of an improvement of more than $c_1 \ln n$ is in $\mathrm{o}(1/n^{k+2}) = \mathrm{o}(1/(\lambda n^2))$.

Observe that the probability of failure is small enough such that the assumption that none of the improvements is larger than $c_1 \ln n$ in λn essential steps over the whole island model fails only with probability in $\mathrm{o}(1/n)$. We assume that this does not happen during the optimization.

Working on solution $\mathbf{x}$, the probability of an essential step is at most $1/n$, independently of the fitness, as the prominent 0-bit at position $\mathbf{x}_{i+1}$ must be flipped. There are $\tau/n \leq e$ essential steps in τ iterations in expectation. By Theorem 4, there is a constant $c_2 > 0$ such that the probability of overshooting this expectation by more than a factor $c_2 \ln n$ is $\mathrm{o}(1/(\lambda n))$. By the above assumption we need at least $c_2 \ln n$ essential steps in τ iterations to improve the fitness by more than $c_1 c_2 \ln^2 n$. Choosing $c = c_1 c_2$ and taking a union bound over all λ islands thus implies the lemma. □

Theorem 24 *Let λ be bounded above by a polynomial in n. Consider an island model using broadcast communication with a migration interval $\tau \leq en$ on the complete graph K_λ to optimize* LeadingOnes. *The expected optimization time and communication costs are*

$$E[T_{\mathrm{LO}}] = \Omega\left(\frac{n^2}{\lambda} + \frac{n}{\log(\lambda/n)} + \frac{\tau n}{\log^2 n}\right) \quad \textit{and}$$
$$E[C_{\mathrm{LO}}] = \Omega\left(\frac{n^2}{\tau} + \frac{\lambda n}{\tau\log(\lambda/n)} + \frac{\lambda n}{\log^2 n}\right).$$

Proof A lower bound of order $\Omega(n^2/\lambda + n/\log(\lambda/n))$ follows from the λ-*parallel unbiased black-box complexity* of LeadingOnes [3]. The third term stems from the time the islands spend on finding independent improvements between migrations. We prove this bound by coupling several random variables.

Let the constant $c > 0$ be as in Lemma 23 and, for integers $0 \leq j \leq (n+1)/(c\ \ln^2 n)$, define the j*-th fitness block* as the fitness levels from $j(c\ \ln^2 n)$ to $(j+1)(c\ \ln^2 n)-1$. We say that the optimization process is in the j-th block, if the maximum fitness over all islands is in that block. Note that optimization starts in block 0 with probability superpolynomially close to 1, by Chernoff bounds. We define a simplified random process of *block discovery* that models the computation on a single island but, at the same time, also incorporates the beneficial influence of migration. There are two ways for the island to discover a block in this process. Every essential step, flipping the left-most 0-bit and none of the leading 1s, reveals a whole new block. Also, every τ iterations grants another one, simulating the benefits of migration.

Let T' stand for the random variable denoting the number of rounds the process needs to discover the entirety of $(n+1)/(c\ \ln^2 n)+1$ blocks. Lemma 23 implies that the probability of the original optimization skipping a full block in τn iterations is o(1) as none of the islands makes a progress of more than $c\ln^2 n$. Hence, T_{LO} first-order stochastically dominates T' and any lower bound on $E[T']$ extends to $E[T_{\mathrm{LO}}]$ (as long as it does not exceed $n\tau$).

To establish $E[T'] = \Omega(n\tau/\log^2 n)$, it is enough to prove the existence of a constant $c' > 0$ such that

$$\lim_{n\to\infty} P\left[T' \leq c'\frac{\tau n}{\ln^2 n}\right] = 0.$$

We fix some positive real number $c' \leq 1/(c\,(2e+1))$. The reason for this choice will become apparent in the following discussion. For the moment, it is sufficient to ensure $C = (1/c) - c' > 0$.

Set $t = c'\tau n/\ln^2 n$. The definition of the discovery process guarantees at least $t/\tau = c'n/\ln^2 n$ blocks in t iterations unconditionally. For $T' \leq t$ to hold, the essential steps have to make up for the remaining blocks. Let X denote the number of essential steps during t rounds. X, in turn, is dominated by a binomially distributed variable

$Y \sim \mathrm{Bin}(t, 1/n)$.

$$P[T' \le t] \le P\left[X > \frac{n+1}{c\,\ln^2 n} - \frac{t}{\tau}\right] \le P\left[Y > C\frac{n}{\ln^2 n}\right].$$

It is best to split the remaining argument into two cases depending on the limit behavior of $E[Y] = t/n = c'\tau/\ln^2 n$. First, assume the expectation is bounded for all n, then $\mathrm{Var}[Y] = E[Y](1 - 1/n)$ is bounded as well. By Chebyshev's inequality,

$$\lim_{n\to\infty} P\left[Y > C\frac{n}{\ln^2 n}\right] \le \lim_{n\to\infty} \frac{\mathrm{Var}[Y]}{\left(C\frac{n}{\ln^2 n} - E[Y]\right)^2} = 0.$$

In case $E[Y]$ diverges, we define

$$1 + \delta = \frac{C\,n}{E[Y]\,\ln^2 n} = \frac{\frac{1}{c} - c'}{c'} \cdot \frac{n}{\tau}.$$

The assumptions $\tau \le en$ and $c' \le 1/(c\,(2e+1))$ together now imply $\delta \ge 1$. Using Theroem 4, we finally arrive at

$$P\left[Y > C\frac{n}{\ln^2 n}\right] = P\big[Y > (1+\delta)\,E[Y]\big] \le \exp\left(-\frac{\delta}{3}E[Y]\right).$$

The right member of the inequality converges to 0. □

The upper and lower bounds for LeadingOnes yield the following optima.

Corollary 25 *Consider an island model using broadcast communication with a migration interval $\tau \le en$ on the complete graph K_λ to optimize* LeadingOnes. *The best-possible expected optimization time $E[T_{\mathrm{LO}}]$ is in $\Omega\left(\frac{n}{\log n}\right) \cap \mathrm{O}(n)$. The best expected combined costs $E[T_{\mathrm{LO}} + C_{\mathrm{LO}}]$ are in $\Omega\left(\frac{n^{3/2}}{\log n}\right) \cap \mathrm{O}\left(n^{3/2}\right)$.*

Proof To minimize the upper bound on $E[T_{\mathrm{LO}}]$ set $\lambda = \Omega(n)$ and a constant τ, resulting in $\mathrm{O}(n)$. For every polynomial number of islands, the the lower bound is of order $\Omega(n/\log n)$, regardless of τ. For the combined costs choose both λ and τ in $\Theta(\sqrt{n})$ to get an upper bound of $\mathrm{O}(n^{3/2})$. The lower bound is minimized by choosing them in $\Theta(\sqrt{n}\,\log n)$ instead, which gives $\Omega(n^{3/2}/\log n)$. □

6 Conclusion

In this paper we derived upper and lower bounds on the parallel run time of distributed evolutionary algorithms, called *island models*, for different migration topologies and means of communication. We introduced a general approach to bound the optimization times based on a fitness-level argument. Our versatile method has been proven successful in a variety of settings and is powerful enough to improve existing run time bounds.

At the same time, we analyzed the communication effort in these distributed computing systems. We investigated the combined costs of computation and communication as a much more realistic cost measure. In particular, we showed that the complete graph with broadcast communication does not yield the best performance due to its high communication overhead. It turned out that the randomized communication policy of *rumor spreading* achieves a much better balance between a fast run time and low network traffic.

Acknowledgements This work has been partially supported by the *Direction générale l'armement* of the French Ministry of Defense under the X-DGA Contract and by the *German Research Foundation* under Grant Agreement FR2988 (TOSU). We would like to thank the anonymous reviewers as well as the guest editors, Carola Doerr and Dirk Sudholt, for their constructive feedback, which significantly improved the quality of the paper.

References

1. Alba, E.: Parallel evolutionary algorithms can achieve super-linear performance. Inf. Process. Lett. **82**, 7–13 (2002)
2. Alba, E., Luque, G., Nesmachnow, S.: Parallel metaheuristics: recent advances and new trends. Int. Trans. Oper. Res. **20**, 1–48 (2013)
3. Badkobeh, G., Lehre, P.K., Sudholt, D.: Unbiased black-box complexity of parallel search. In: Proceedings of the 13th International Conference on Parallel Problem Solving from Nature (PPSN), pp. 892–901 (2014)
4. Badkobeh, G., Lehre, P.K., Sudholt, D.: Black-box complexity of parallel search with distributed populations. In: Proceedings of the 13th Workshop on the Foundations of Genetic Algorithms (FOGA), pp. 3–15 (2015)
5. Böttcher, S., Doerr, B., Neumann, F.: Optimal fixed and adaptive mutation rates for the leadingones problem. In: Proceedings of the 11th International Conference on Parallel Problem Solving from Nature (PPSN), pp. 1–10 (2010)
6. Cantú-Paz, E.: A survey of parallel genetic algorithms. IlliGAL Report 97003 (1997)
7. Doerr, B.: Optimal parameter settings for the $(1 + (\lambda, \lambda))$ genetic algorithm. CoRR arxiv:1604.01088 (2016)
8. Doerr, B., Fischbeck, P., Frahnow, C., Friedrich, T., Kötzing, T., Schirneck, M.: Island models meet rumor spreading. In: Proceedings of the 2017 Genetic and Evolutionary Computation Conference (GECCO), pp. 1359–1366 (2017)
9. Doerr, B., Fouz, M.: Asymptotically optimal randomized rumor spreading. In: Proceedings of the 38th International Colloquium on Automata, Languages and Programming (ICALP), pp. 502–513 (2011)
10. Doerr, B., Johannsen, D., Winzen, C.: Multiplicative drift analysis. Algorithmica **64**, 673–697 (2012)
11. Doerr, B., Künnemann, M.: Tight analysis of randomized rumor spreading in complete graphs. In: Proceedings of the 2014 Meeting on Analytic Algorithmics and Combinatorics (ANALCO), pp. 82–91 (2014)
12. Doerr, B., Künnemann, M.: Optimizing linear functions with the $(1+\lambda)$ evolutionary algorithm—different asymptotic runtimes for different instances. Theor. Comput. Sci. **561**, 3–23 (2015)
13. Droste, S., Jansen, T., Wegener, I.: On the analysis of the $(1+1)$ evolutionary algorithm. Theor. Comput. Sci. **276**, 51–81 (2002)
14. Gießen, C., Witt, C.: Optimal mutation rates for the $(1 + \lambda)$ EA on OneMax. In: Proceedings of the 2016 Genetic and Evolutionary Computation Conference (GECCO), pp. 1147–1154 (2016)
15. He, J., Yao, X.: A study of drift analysis for estimating computation time of evolutionary algorithms. Nat. Comput. **3**, 21–35 (2004)
16. Johannsen, D.: Random combinatorial structures and randomized search heuristics. Ph.D. thesis, Universität des Saarlandes (2010)

17. Lässig, J., Sudholt, D.: General upper bounds on the runtime of parallel evolutionary algorithms. Evol. Comput. **22**, 405–437 (2014)
18. Lehre, P.K., Witt, C.: Black-box search by unbiased variation. Algorithmica **64**, 623–642 (2012)
19. Lissovoi, A., Witt, C.: On the utility of island models in dynamic optimization. In: Proceedings of the 2015 Genetic and Evolutionary Computation Conference (GECCO), pp. 1447–1454 (2015)
20. Mambrini, A., Sudholt, D.: Design and analysis of schemes for adapting migration intervals in parallel evolutionary algorithms. Evol. Comput. **23**, 559–582 (2015)
21. Mitzenmacher, M., Upfal, E.: Probability and Computing: Randomized Algorithms and Probabilistic Analysis. Cambridge University Press, Cambridge (2005)
22. Neumann, F., Oliveto, P.S., Rudolph, G., Sudholt, D.: On the effectiveness of crossover for migration in parallel evolutionary algorithms. In: Proceedings of the 2011 Genetic and Evolutionary Computation Conference (GECCO), pp. 1587–1594 (2011)
23. Ruciński, M., Izzo, D., Biscani, F.: On the impact of the migration topology on the island model. Parallel Comput. **36**, 555–571 (2010)
24. Rudolph, G.: Convergence Properties of Evolutionary Algorithms. Verlag Dr. Kovač, Hamburg (1997)
25. Sudholt, D.: Parallel evolutionary algorithms. In: Springer Handbook of Computational Intelligence, pp. 929–959. Springer, Berlin (2015)
26. Wegener, I.: Theoretical aspects of evolutionary algorithms. In: Proceedings of the 28th International Colloquium on Automata, Languages and Programming (ICALP), pp. 64–78 (2001)
27. Witt, C.: Tight bounds on the optimization time of a randomized search heuristic on linear functions. Comb. Probab. Comput. **22**, 294–318 (2013)