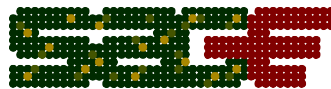


Small or medium-scale focused research project (STREP)

Full proposal

ICT FET Open Call

FP7-ICT-2013-C



Speed of Adaptation in Population Genetics and Evolutionary Computation (SAGE)

Date of preparation: 15th April 2013

Name of short proposal this full proposal refers to: Speed of Adaptation in Population Genetics and Evolutionary Computation (SAGE 332681)

Type of funding scheme: Small or medium-scale focused research project (STREP)

Work programme topics addressed: ICT-2013.9.3: FET Young Explorers

Name of coordinating person: Per Kristian Lehre

List of participants:

Participant no.	Participant organisation name	Short name	Country
1	The University of Nottingham	UNOTT	UK
2	Institute of Science and Technology Austria	IST Austria	AT
3	Friedrich-Schiller-Universität Jena	UNI-JENA	DE
4	The University of Sheffield	USFD	UK

Proposal Abstract

Biological evolution has produced an extraordinary diversity of organisms, even the simplest of which is highly adapted, with multiple complex structures. Evolutionary computation has found that many innovative solutions to optimisation and design problems can be achieved by artificial evolution via random variation and selection.

Despite the centrality of evolution to biology and the usefulness of evolutionary algorithms in optimisation, the dynamics of evolution are not well understood. Consequently, population genetics theory can only make quantitative predictions about short-term, simple biological evolution, and the design and parameter tuning of evolutionary algorithms is mostly done ad-hoc in a laborious and cost-intensive process.

Both fields have studied the speed of adaptation independently, and with orthogonal approaches. Our project brings together an interdisciplinary consortium of ambitious researchers from the theory of evolutionary computation and theoretical population genetics to synergise these complementary approaches and to create the foundation of a unified quantitative theory describing the speed of adaptation in both biological and artificial evolution.

The transformative impact of this unified theory will lie in enabling long-term predictions about the efficiency of evolution in settings that are highly relevant for both fields and related sciences. Our approach will reveal how this efficiency is fundamentally determined by evolutionary and environmental parameters. Tuning these parameters will allow researchers from biology and computation to increase the efficiency of evolutionary processes, revolutionising applications ranging from evolutionary algorithms to experimental evolution and synthetic biology.

Table of Contents

1	Scientific and/or technical quality, relevant to the topics addressed by the call	1
1.1	Targeted breakthrough and its relevance towards a long-term vision	1
1.2	Novelty and foundational character	4
1.3	S/T methodology	5
2	Implementation	29
2.1	Management structure and procedures	29
2.2	Individual participants	34
2.3	Consortium as a whole	38
2.4	Resources to be committed	39
3	Impact	41
3.1	Transformational impact on science, technology and/or society	41
3.2	Contribution at the European level	47
3.3	Dissemination and use of project results	47
4	Ethical issues	51
A	Appendix	56

1 Scientific and/or technical quality, relevant to the topics addressed by the call

1.1 Targeted breakthrough and its relevance towards a long-term vision

Living organisms show an extraordinary variety of complex adaptations, ranging from the basic molecular mechanisms that allow accurate replication and translation of DNA, through the regulatory programs that reliably develop multicellular organisms, to the cognitive abilities that allow us to begin to understand our own evolution.

Understanding and exploiting the power of adaptation through evolution is at the heart of two scientific disciplines. **Population genetics (PG)** is the discipline that studies the adaptation of populations in biological evolution, and how it is shaped by basic forces like mutation, recombination, selection, and migration between sub-populations [77]. **Evolutionary computation (EC)** studies, develops, and applies **evolutionary algorithms (EAs)** (also called **genetic algorithms**), powerful and versatile optimisation techniques using an artificial, virtual evolution of candidate solutions for solving a broad range of problems in optimisation and design [20, 35, 127].

Even though both disciplines share a common inspiration, they have developed in almost complete isolation. In recent years both fields have independently uncovered the **speed of adaptation** as a key concept for understanding evolution. It describes the time evolution needs to assemble a complex adaptation (in natural evolution) and to evolve a high-quality solution (through an evolutionary algorithm). It thus describes the **efficiency of biological and artificial evolution**, and how it depends on parameters of evolution (number of genes, population size, mutation rate, etc.) and parameters of the environment [6, 84].

The speed of adaptation is a key indicator as it reveals how complex adaptations have evolved in the past, and it can predict how quickly future adaptations can happen. This is highly relevant in many areas of biological evolution such as synthetic biology, experimental evolution and preventing the evolution of drug resistances. In evolutionary computation the speed of adaptation is the primary concern in the design and application of EAs; there is tremendous and ever increasing demand for evolving better solutions and solving larger problems with limited computational resources.

PG and EC each have developed analytical techniques for quantifying the speed of adaptation, with complementary strengths and weaknesses. PG is able to make short-term predictions about how genes evolve and spread throughout large populations, but is lacking techniques for making long-term predictions and handling large numbers of genes. EC naturally deals with large-scale problems involving many genes and has developed a quantitative theory for making long-term predictions. However, this theory is largely restricted to trivially small populations, lacking the technology for predicting evolution in large populations as found in most practical applications of EAs.

The limited understanding of evolution implies that population genetics cannot explain recent developments in gene sequencing, showing that evolution acts on a large number of genes, and the poor understanding of EAs used in practice represents a major obstacle for their applicability in industry [48]. The SAGE project is designed to **synergise population genetics and evolutionary computation through unifying their complementary approaches** in order to create a **unified quantitative theory for the speed of adaptation in biological and artificial evolution**. To this end we have assembled an **interdisciplinary consortium** of excellent young researchers from the theory of evolutionary computation (UNOTT, UNI-JENA, USFD) as well as theoretical population geneticists from the world-leading group at IST Austria, all of which have the necessary skills, expertise, and determination for achieving this breakthrough.

The transformative impact of this unified theory will lie in enabling long-term predictions about the efficiency of evolution in settings that are highly relevant for both PG and EC. Our approach will reveal how this efficiency is fundamentally determined by evolutionary and environmental parameters. Tuning these parameters will allow researchers from biology and computation to increase the efficiency of evolutionary processes, revolutionising applications ranging from evolutionary algorithms to experimental evolution and synthetic biology.

Population genetics (PG)

Population genetics (PG) studies how evolution is shaped by basic forces such as mutation, selection, recombination, migration among sub-populations, and stochasticity; it forms the core of the modern understanding of evolutionary theory (the so called “modern synthesis”). PG has a long tradition of mathematical

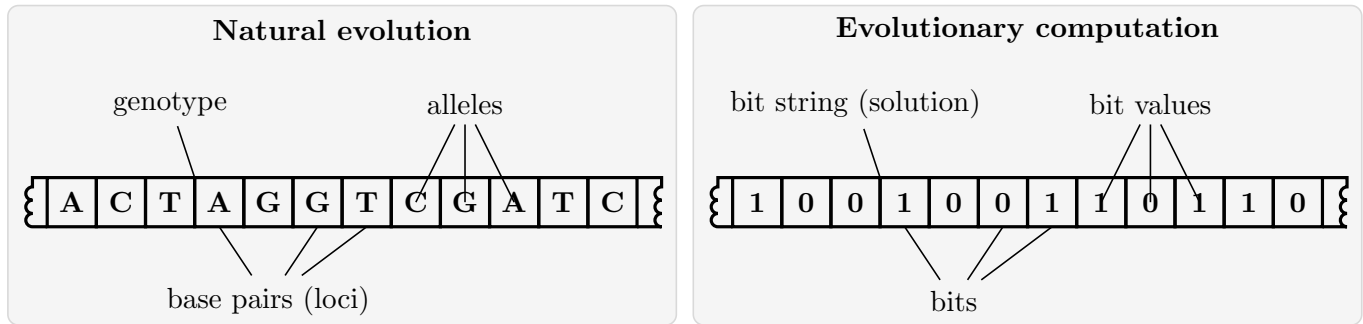


Figure 1.1a: Basic terminology in natural and artificial evolution. Evolution takes place on *genotypes* and *bit strings*, respectively; the latter are also called (*candidate*) *solutions*. Genotypes consist of *loci* or *genes* and bit strings consist of *bits*. Each locus carries one *allele* from the set $\{A, C, G, T\}$ and each bit has a *bit value* from $\{0, 1\}$. Note that bit strings are just one of many common representations for candidate solutions in evolutionary computation.

modelling, starting in the 30s with the pioneering work of Fisher, Wright, Haldane and others, and is now a highly sophisticated field in which mathematical analysis plays a central role [16]. Early work focused on simple deterministic models with small numbers of loci (Figure 1.1a clarifies basic terminology), aiming at understanding how the change in genotype frequencies in a population was affected by basic evolutionary forces. It has since branched out to investigate topics such as the evolution of sexual reproduction, the role of environmental fluctuations in driving genetic change, and how populations evolve to become independent species. Recently, the explosion of data from new techniques for genetic sequencing and manipulation has made it clear that selection acts at a large number of interacting genes, creating a need for more realistic and ambitious models of evolution. An especially pressing problem is understanding how the speed of adaptation scales with population and genome size, and with the basic forces mentioned above, all of which vary over many orders of magnitude in natural populations. This problem is especially challenging because all of these factors, particularly the recombination rate, can interact in subtle ways with the “fitness landscape”, the function that describes the mapping between an organism’s genetic makeup and its fitness, the quantity relevant for natural selection. Almost all current PG models are restricted to the simplest fitness landscapes. Since natural fitness landscapes are likely to be far more complicated, indeed too complicated to ever be measured completely, there is a need for a theory that describes the speed of adaptation over a broad range of landscapes in terms of just a few key features.

Evolutionary computation (EC)

The astonishing success of natural evolution has inspired evolutionary computation (EC), which artificially evolves candidate solutions to solve practical design and optimisation problems. Such evolutionary algorithms (EAs) maintain a population of candidate solutions (individuals) whose quality is measured by a problem-dependent objective function called the fitness function. Each generation of evolution consists of recombining and mutating good solutions from a population and using selection to allow only the “fittest” solutions to survive (see Figure 1.1b for illustrations of mutation and recombination). Countless variations of this principle have been successfully applied in numerous domains, such as scheduling [48], logistics [44], drug design [105], and the automotive industry [34] to the open-ended evolution of designs for engineering problems, like evolving radio antennas for satellites sent into space by NASA [49].

The main obstacle when applying EAs is that in most applications their efficiency depends crucially,

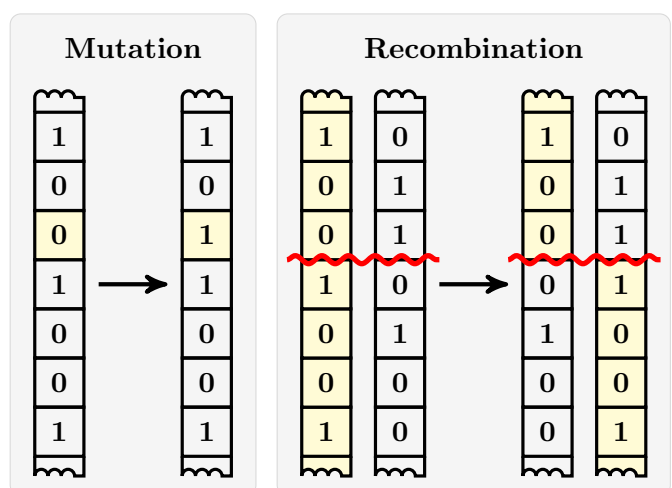


Figure 1.1b: Illustration of asexual reproduction (mutation) and sexual reproduction (recombination). Mutation changes the values of alleles at random loci. Recombination acts by exchanging genetic information between the two parent individuals, forming new genetic combinations that did not exist in the parents.

and often unpredictably, on design choices such as choosing the right operators, representations of solutions, and parameter settings for the problem at hand. This includes the question of whether and how to use recombination operators and how to implement parallel variants of evolutionary algorithms that evolve populations in parallel, with individuals migrating between them. The challenges in understanding artificial evolution in EAs are similar to those faced in natural evolution. To date, design and parameter tuning of EAs is still error-prone and costly, which limits the applicability of EAs. A 2007 survey amongst EC practitioners identified “poor understanding” as the biggest obstacle for industry uptake of EAs [48] (39.7% of respondents).

In recent years runtime analysis has emerged as a successful theoretical approach, taking an algorithmic point of view to study the speed of adaptation in EAs [6, 84]. Runtime analysis describes how the time needed to find a solution of satisfactory quality depends on parameters of the problem and the EA. In evolutionary terms, this corresponds to the number of generations needed to adapt the population to a new environment, i.e. the speed of adaptation. Many studies have explained the impact of design choices of EAs on performance, in the light of various problem features [32, 84, 87]. However, most studies were constrained to EAs with trivial populations, evolving just one or two candidate solutions [52]. Further progress is impeded by the lack of methods and techniques for dealing with realistic populations.

Breakthrough and long-term vision

While PG and EC are asking many similar questions, they have developed in almost complete isolation from each other, and have consequently led to separate approaches, each with their own strengths and limitations. In evolutionary computation, while current analytical techniques consider a wide range of possible fitness landscapes, the EAs that are analysed are far less sophisticated than practically used EAs. EAs that are successful in practical applications often have sexual reproduction and non-overlapping generations, while theoretical analyses are mostly restricted to a population of one individual evolving by mutation only. Conversely, while many population genetics models incorporate quite complicated populations, most studies focus on extremely simple fitness landscapes, often ignoring complicating effects such as epistasis¹.

This project’s **breakthrough** will be to give a **unified, quantitative theory for the speed of adaptation** that enables **inter-disciplinary** studies of **artificial and biological evolution** by combining the complementary approaches from population genetics and evolutionary computation.

Synergising these two fields has potential for transformative impact in both fields, revolutionising their understanding of evolution and opening new possibilities for applications involving evolutionary processes. In particular, our quantitative theory will highlight how the speed of adaptation depends on parameters of evolution as well as parameters of the fitness landscape a population is adapting to. Thereby, it will reveal what variants and parameters of evolutionary processes are most effective for particular fitness landscapes, which is of high relevance for both fields.

By focusing on the speed of adaptation, SAGE will provide new tools for researchers trying to predict the future adaptation of natural populations. This will have significant impact on a series of challenges PG is currently confronted with, as well as allowing it to address new problems:

- better methods for **inferring evolutionary parameters from observed genetic changes**, enabling better inferences about the evolutionary history of natural populations
- designing better selection schemes, allowing for more effective strategies against the **evolution of drug and pesticide resistance**, improving the **in vitro evolution of enzymes**, quantifying the lifetime of engineered biological circuits, such as in **synthetic biology**, and increasing the efficiency of **animal and crop breeding programs**.
- more efficient methods for the **exploration and identification of features of fitness landscapes**, allowing for the elucidation of many fundamental questions in PG.
- a more general understanding of evolution, allowing PG to investigate **non-genetic forms of evolution**, such as cultural evolution, and the spread of technological innovations

Regarding EC, our framework will allow for quantifying the performance of population-based EAs with regard to their speed of adaptation, and how performance depends on parameters of the evolutionary algorithm and the optimisation task. This understanding will enable the identification of optimal parameters

¹*Epistasis* describes a phenomenon where the effect of a gene depends on other genes. In other words, genes are not independent.

and design choices for EAs, thus significantly reducing the labour- and cost-intensive process of designing and tuning EAs by trial and error. This will alleviate the biggest obstacle to industry uptake of EAs and contribute to improving the effectiveness of EAs in several ways:

- improving adaptation of populations to generate **better quality solutions** in applications of EAs,
- speeding up adaptation, hence generating high-quality solutions in **less time**,
- addressing the emerging opportunities and challenges posed by designing effective parallel EA variants on **massively parallel architectures**,
- improving the scalability of EAs: how their running time depends on the problem size, thus enabling to **solve larger problems** with limited or costly computational resources.

All of these points have enormous potential for long-lasting impact in many areas of science and technology. They are instrumental for keeping up with the rapidly increasing size and complexity of modern-day design and optimisation problems from many areas including telecommunications, drug design, energy-efficiency, and computational biology.

Objectives

SAGE will lay the foundation towards achieving this vision through the following **objectives**:

- O1: to synergise models and methods from PG and EC, thus enabling to quantify the speed of adaptation of **realistic populations** on **realistic fitness landscapes** with many loci (Work Package 1, Deliverables D 1.1, D 1.2, D 1.3, D 1.4).
- O2: to identify **key features of fitness landscapes** and to quantify their effect on the speed of adaptation in artificial and biological evolution with asexual reproduction (Work Package 2, Deliverables D 2.1, D 2.2, D 2.3).
- O3: to develop a quantitative theory of **sexual evolution**, allowing to quantify the **benefit of using recombination** in artificial and biological evolution, and how recombination can be **used most effectively** in EAs (Work Package 3, Deliverables D 3.1, D 3.2, D 3.3).
- O4: to develop a quantitative theory of **parallel evolution in structured populations**, which is instrumental in using **parallel implementations of EAs** in many-core architectures most effectively (Work Package 4, Deliverables D 4.1, D 4.2, D 4.3).

These deliverables will be published as reports and in high-impact venues in both fields. They will represent proof-of-concept regarding the targeted breakthrough, and stimulate further research in a novel research area unifying artificial and biological evolution.

1.2 Novelty and foundational character

Theoretical population genetics and evolutionary computation have undergone rapid development in recent years. In PG, the explosion of data from new techniques for genetic sequencing and experimental evolution has not only created a need for more sophisticated models of evolution, it has also for the first time made it realistic to hope to test such models. A necessary component of more realistic models will be complex fitness landscapes, which have been explored in detail in EC. For the particular question of the speed of adaptation, initial discussions have shown that techniques from runtime analysis can provide a valuable new perspective.

The runtime analysis of EAs has started off in the late 1990s considering simple EAs on simple test problems [31]. Researchers from the theory of algorithms, mathematics, machine learning, and probability theory have contributed methods and techniques that have been applied successfully for the analysis of EAs. Recent years have seen the field mature, culminating in numerous studies of EAs on many problems ranging from function optimisation to computationally hard combinatorial problems [6, 84, 87]. However, it is becoming obvious that the theory of EAs is starting to stagnate on EAs with trivially small population sizes, thus falling short of explaining the success of population-based EAs in many practical applications. PG provides sophisticated methods for handling large populations, thus promising to fill a crucial gap in the toolbox currently available to EC and allowing the theory of EC to grow much further for many years to come.

Novelty and timeliness

Even though EC is inspired by biological evolution, this inspiration is only very loosely based on natural processes. EC has therefore developed almost independently from PG, and there have only been few, sporadic cases where EC researchers have integrated PG theory [82, 89].

SAGE represents the first concerted effort for the interdisciplinary study of evolution between EC and PG, involving researchers from both fields. Our focus is on the speed of adaptation as a unifying research question that has been recognised as a key topic in both fields. In EC, the speed of adaptation, or runtime analysis, has become one of the best established research fields in the theory of EAs [6, 84]. It has now reached a degree of maturity, where for the first time a unified theory of evolution can become a reality. The situation in PG is very similar; the speed of adaptation has also gained traction in PG, and stochastic methods of the same nature as in EC have been put forward as the right framework to study this central question in evolution. This highlights that the time is ripe for an integration of the two fields: both fields converged on the same questions and on tools that are reconcilable. This has been recognised in both fields: for example, researchers from IST Austria have very recently reached out to the theory of EC as a precursor to the SAGE project [7, 91]. This has also been recognised by others: other initiatives are under way in third-party countries, such as the US. For example, the UC Berkeley Theory of Computation Institute recently announced a programme for “Evolutionary Biology and the Theory of Computing” aiming at promoting “the interaction between theoretical computer scientists and researchers from evolutionary biology”. The scope of this initiative is different from SAGE as the former aims at synergising theoretical computer science and biology, but it highlights that new perspectives are required in order to further the understanding of evolution. This renewed interest from both fields shows that this is a perfect time for integrating the two disciplines of EC and PG, and our consortium is uniquely prepared to do so.

Challenging current thinking

Due to its fundamentally interdisciplinary nature, SAGE is designed to challenge current thinking in both PG and EC, combining techniques from both areas to overcome the limitations of each field and to extend the range of evolutionary processes that are amenable to mathematical analysis. In particular, SAGE will develop general mathematical models covering a wide range of evolutionary regimes and evolutionary algorithms, along with mathematical techniques that precisely describe how populations adapt. These models and techniques will address the simplifying assumptions that either field make, such as overly simplistic fitness landscapes (PG), or trivially small population sizes (EC). Furthermore, SAGE will focus on the speed of adaptation, i. e., on future adaptation. This will lead to a shift of the current focus in PG, from inferring historical evolution to making predictions about the future. This will have the potential impact of enabling the use of evolutionary theory as a technology to predict and direct future evolution.

The common scientific foundation, which will be one of the central results of the SAGE project, is outlined in the first work package in the next section.

All work packages will provide further details regarding the state of the art in EC and PG and the novelty of our approach in their respective context.

1.3 S/T methodology

Overall strategy

Our work programme comprises four technical work packages WP 1–4 and two management work packages WP 5–6. It is designed to develop synergies between PG and EC in a well-structured, guided, and systematic manner, as shown in Figure 1.3a. Work packages WP 1–4 will specifically target our objectives O1–4, respectively.

The first step is to build a common scientific foundation towards a quantitative theory of evolution (Objective O1). Work package WP 1 directs and focuses efforts in this direction through building a common language to bridge the cultural divide, reconciling models and analytical techniques for studying evolutionary processes, transferring knowledge between the two fields, and developing novel analysis methods combining the strengths of current approaches. A particular focus regarding knowledge transfer is how to analyse large populations. PG has developed sophisticated techniques in this respect, particularly regarding asexual populations, whereas such techniques are sorely missing in EC. Transferring and integrating these techniques will enable us to analyse realistic asexual populations in natural and artificial evolution and prove a vital tool for the remainder of the work programme.

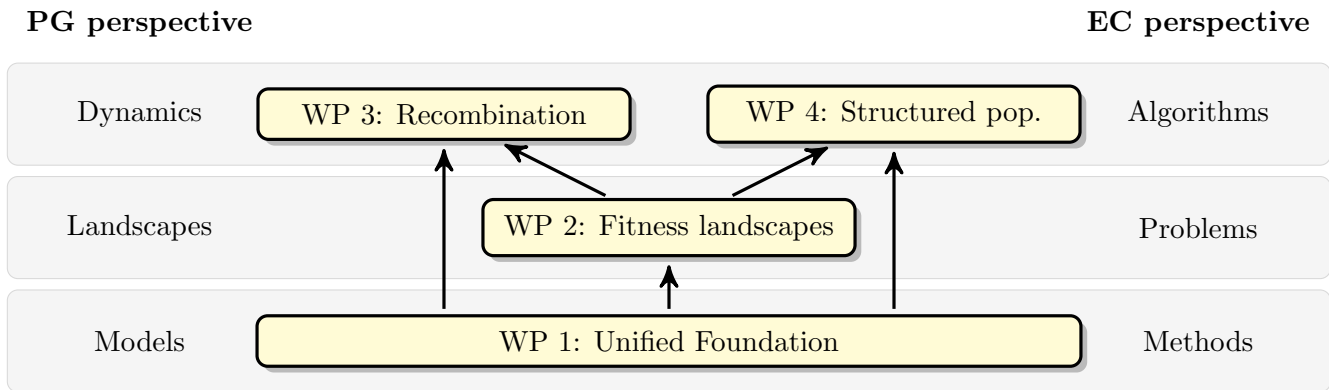


Figure 1.3a: Structure of technical work packages WP 1–4. The labels on either side describe the aspects being investigated, using terminology from PG (left) and EC (right), respectively. Unifying *models* and *methods* from PG and EC will provide a solid foundation, enabling the study of evolution on fitness *landscapes* and optimisation *problems*, respectively. Both WP 1 and WP 2 then allow us to study two vital aspects of evolutionary *dynamics* and evolutionary *algorithms*: recombination and parallel evolution in structured populations.

One of the most important factors for evolution is the structure of the fitness landscape, which describes the relationship between an organism's genetic makeup and its fitness (Objective O2). EC has studied a large variety of fitness landscapes as EAs have been applied to a plethora of optimisation problems with different characteristics. In WP 2 we will investigate how the structure of the fitness landscape affects the speed of adaptation, based on a systematic study of relevant problem features such as multiple peaks (multiple local optima), plateaus or neutral regions (areas of equal fitness), where evolution is not guided by fitness, and dynamically changing fitness landscapes as frequently occurring in real-world optimisation and natural evolution. Thereby we will use techniques for the analysis of large asexual populations from WP 1.

Work packages WP 3 and WP 4 are laid out to extend our theory towards more complex evolutionary dynamics that are highly relevant for both natural and artificial evolution.

WP 3 will extend our quantitative theory towards sexual reproduction (Objective O3). Most natural evolutionary processes use sexual reproduction, and understanding why sex evolved is one of the most important open problems in biology. Similarly, most EAs use recombination, but the reasons when and why recombination is useful remain elusive. We will tackle this fundamental challenge head on by synergising knowledge and methods from PG and EC. This will generate novel insights about the advantages of sexual reproduction, leading to practical recommendations on how to use recombination operators in EC.

Finally, WP 4 will extend our theory to parallel, structured populations as found in nature and in many common parallel implementations of evolutionary algorithms (Objective O4). In parallel EAs, parallel populations evolve on different processors, periodically communicating genetic information through processes that reflect the migration of organisms or pollination in plants. Communication takes place on a spatial structure connecting these populations (see Figure 1.3b on page 16), and the size and shape of this structure, as well as the amount, method and frequency of communication, all have a strong impact on evolutionary dynamics. Using parallel EAs is a highly active research area, fuelled by the rapidly increasing number of CPU cores. However, there is little guidance as to how to design effective parallel EAs as the parallel evolution of structured population is poorly understood. Our work will provide a sorely needed theoretical foundation, through exploring synergies with relevant models from PG. This will enable users of parallel EAs to make informed design choices and to use parallelism more effectively.

Achieving these ambitious aims requires a strong, coherent, and efficient management as well as an effective strategy for maximising dissemination and exploitation of our results. To reflect these key points in our work plan, we further implement designated work packages for coordination and management (WP 5, detailed in Section 2.1) as well as dissemination and exploitation (WP 6, detailed in Section 3.3).

The remainder of this section gives more details on the technical work packages WP 1–4.

WP 1 Unified foundation

The PG and EC communities have been evolving independently for many years, with only a few isolated efforts to bridge the gap between them. Both areas have developed a wealth of results unknown to the other area. Some of these results are likely to have relevant equivalents in the other field, but a systematic effort

to translate and compare these results has never before been taken. The initial phase of the project will be to establish a common mathematical methodology that encompasses the techniques and results of both PG and EC.

WP 1.1 A unified model of evolution

PG and EC have developed different terminology, specialised to the respective field. This represents a major barrier in communicating results from one field to the other, and it is part of the reason why these two fields have been separated for such a long time.

Our first step in this work package will be to develop a common foundation by unifying terminology. We will create an online wiki that clarifies terms and concepts from both PG and EC. A wiki provides an easy means for cross-referencing between related concepts in the other field. It has built-in functionality to allow remote collaborations between all members of the project, and furthermore it will be made accessible to a world-wide audience, thus benefiting the wider communities of PG and EC and attracting further participants in this research direction.

Next, we will establish the basic components of a unified abstract model of evolution, suitable for comparing models/evolutionary regimes, tools and techniques. This model will have concrete instantiations which will enable direct comparisons between the two fields. The basic ingredients of such a general model for evolution are the concepts of population, inheritance, variation and selection. By classifying the different models of both fields in these terms, we will harmonise our terminology, allowing us to identify results obtained for similar problems, and to ease the application of the techniques of each field to the opposite field.

Deliverable D 1.1: An abstract general model of evolution, based on a unified terminology, that allows for direct comparisons between EC algorithms and PG evolutionary regimes. A translation of terminology from PG and EC into our unified model.

WP 1.2 Translating results between PG and EC

Having built a common terminology, we will start to translate results between the two fields. All PIs have a full command of analysis techniques used in their respective field (EC: UNOTT, UNI-JENA, USFD; PG: IST Austria), expertise in interdisciplinary work, and an extensive knowledge of relevant literature, which makes our consortium perfectly placed for a systematic interdisciplinary knowledge transfer.

To exemplify this reconciliation, we discuss a striking similarity between the strong selection weak mutation (SSWM) regime of PG and the evolutionary algorithm (1+1) EA [32]. The SSWM regime and the (1+1) EA are two of the best-understood models in PG and EC, respectively, but the results from the two fields have never before been compared, making this an excellent starting point for building a common foundation towards a unified theory.

The SSWM regime describes a population in which mutations occur so rarely that their fate is determined and resolved before a new mutation arises [45]. More specifically, this means that in the time between two mutations a mutation either spreads throughout the whole population (i. e. every member of the population carries the mutation), or it disappears from the population. Under this assumption, for most of the time the population contains identical genotypes. This implies that information about a single genotype is sufficient information in order to describe the current state of the population.

Evolutionary algorithms consist of a process in which trial solutions are generated by mutation from a resident genotype and their fate is decided based on some selection procedure and the difference in fitness of the mutant with respect to the resident genotype. The (1+1) EA represents a bare-bones version of this fundamental mechanism of mutation and selection. It maintains a single resident genotype and in each iteration it generates a single offspring by mutation. The offspring replaces its parent as the resident genotype, unless the offspring's fitness is worse than that of its parent. Because of its simplicity, it has been studied intensively from a theoretical perspective.

Both the SSWM and the (1+1) EA work on similar state spaces – natural evolution works on genotypes with letters A, T, C, G , whereas evolutionary algorithms such as the (1+1) EA work on discrete spaces such as bit strings with letters 0 and 1, which encode a candidate solution for an optimisation problem in hand (cf. Figure 1.1a). Mutation in both SSWM and the (1+1) EA works similarly, by allowing each locus to be mutated with a certain probability, called *mutation rate*.

In EC, performance or runtime analyses of the (1+1) EA have analysed the expected number of iterations until a genotype with maximal fitness is discovered, when starting from a random initial genotype. PG has studied the number of mutations required until evolution reaches a local or global optimum in the given

fitness landscape. Both fields also ask what is the probability of finding a global optimum vs. reaching only local optima. These questions are almost identical in both fields, apart from the initialisation (evolutionary algorithms typically initialise genotypes uniformly at random; this is different in nature). Also EC considers the time for generating mutations, whereas PG only counts those generations where mutations do happen. These different time scales do not pose a significant problem as analytical results can be transferred between those two time scales with moderate effort.

EC has provided a plethora of performance analyses of the (1+1) EA for various fitness landscapes, from illustrative objective functions to many problems from combinatorial optimisation [84, 87]. Translating these results into the language of SSWM will yield immediate results on how evolution can or cannot cope with various fitness landscapes. Conversely, adaptation under the SSWM regime has been thoroughly analysed in PG, for example, in landscapes of varying ruggedness [58]. Again, translating these results to EC will provide information about runtime in very general landscapes, albeit for a specific evolutionary algorithm.

Another example of our approach are the typical landscape models that the two fields use. Both fields make use of stereotypical models of typical features of fitness landscapes in order to understand how evolutionary dynamics are affected by specific features of these landscapes. Some of these stereotypical models are analogous between the two fields, making them an ideal baseline to be used in our unified theory of evolution. For example, the following two models are frequently used in PG:

- additive models: in which each locus contributes a certain amount to the overall fitness, depending on its allele and the importance (weight) of the locus. In EC this class of problems is known as *linear functions*; formally, the fitness f is defined as

$$f_{\text{linear}}(x_1 \dots x_n) = \sum_{i=1}^n w_i x_i,$$

where $w_1, \dots, w_n \in \mathbb{R}$ are real-valued weights and $x_i \in \{0, 1\}$ represent all bit values. These constitute an example of problems that do not contain epistasis.

- stabilising selection: where each locus contributes to some quantity, as in linear models, but the objective is to achieve a certain target value z_{opt} for this quantity, lying between the minimum and maximum possible values for this quantity. This is typically formalised in PG as

$$f_{\text{stab}}(x_1 \dots x_n) = - \left(\sum_{i=1}^n w_i x_i - z_{\text{opt}} \right)^2.$$

The underlying problem of approximating a target value via selecting a subset of additive values is at the heart of many fundamental computationally hard (NP-hard) problems in computer science and operational research such as SUBSET SUM, KNAPSACK, and PARTITION. These problems have already been studied in EC in the context of the (1+1) EA [110, 124, 128], establishing a direct link to models used in PG.

We will exploit the similarities and differences between these stereotypical models of fitness landscape features to advance theory in both fields.

Deliverable D 1.2: Correspondence and translation of results between the different evolutionary regimes (such as the SSWM) and different algorithms (such as the (1+1) EA), to identify synergies between PG and EC, and to initiate a systematic knowledge transfer.

WP 1.3 Integrating and developing methodology

The stochastic processes underlying mathematical models of evolution are tremendously complex, and exact analyses are rarely possible. While the fields have used some similar techniques, such as Markov chains and branching processes, their approaches are substantially different.

EC has drawn on techniques from the analysis of algorithms, discrete mathematics, and stochastic processes. Often upper and lower bounds are derived for (expected) times to reach a solution of satisfactory quality. This usually only takes into account the mean dynamics. The obtained bounds hold with mathematical rigour, but there might be a gap between the upper and lower bounds. Results generally hold for arbitrary problem sizes, and they come with explicit preconditions under which the results hold. This means that we get precise mathematical truths holding in a well-defined context.

PG makes regular use of techniques from statistical physics, following a different philosophy. Instead of aiming at rigorous statements, PG approximates the underlying stochastic processes by more mathematically tractable diffusion processes. This means that the predicted dynamics need to be validated, for instance through simulations and statistical evaluations, but often the approximations are able to give accurate predictions of the real dynamics. In contrast to EC, this approach is able to deal with more complicated dynamic processes [17].

We expect that methods from PG and EC will prove useful in the other respective field. Extending both communities' toolboxes of analysis methods is likely to extend the range of evolutionary dynamics that can be analysed by currently available methods. In addition, we hope to combine the advantages of different methods to create even more powerful tools.

One particular example where a transfer of methods might prove very beneficial is the diffusion approximation, which is regularly used in PG. It represents a form of the Fokker-Planck equation, or forward Kolmogorov equation, that describes the time evolution of a probability distribution of some stochastic process. Exploiting this analogy, it is feasible to adapt the equation to the stochastic models typically used in EC. This would allow for a description of the steady states distribution of these processes, among other results that have proved pivotal in PG. In particular, the adjoint of the forward equation, the backward equation, allows for the calculation of the probability of fixation of an allele under general schemes of selection, mutation and finite population size sampling. In order to accomplish this generalisation, one needs only to characterise the different evolutionary/genetic algorithms in terms of their expected change in mean and variance. This will allow a quantitative characterisation of many properties, such as steady state distributions or probabilities of fixation, of arbitrary genetic or evolutionary algorithms.

This quantitative understanding is a vital stepping stone towards understanding the dynamics of natural and artificial evolution. Techniques integrating methods from PG and EC will be invaluable for the quantitative studies in the following work packages.

Deliverable D 1.3: New methods for analysing the speed of adaptation, obtained by transferral or recombination of existing methods in the different fields, such as applying the diffusion approximation to different evolutionary algorithms, enabling quantitative performance analyses of natural and artificial evolution.

WP 1.4 Black-box complexity: limits to the speed of adaptation for all evolutionary processes

This work package will also be used to establish limits on the speed of adaptation in general evolutionary processes. Evolutionary algorithms are often applied to complicated optimisation and design problems where knowledge on the problem is so limited that it is regarded a *black box*. The only way of getting knowledge about the problem is by evaluating candidate solutions.

The *black-box complexity* [33] is a concept that defines the minimum number of candidate solutions that must be evaluated in order to locate a global optimum (or, more generally, any set of target points). The only assumption is that the problem belongs to a given class of problems, e.g. being an instance of a travelling salesperson problem, a convex optimisation problem, etc. Black-box complexity is a powerful concept as it establishes fundamental limits on the speed of adaptation for *every search process*, including evolutionary processes found in evolutionary algorithms and natural evolution. As the black-box complexity describes the minimum number of solution evaluations, this constitutes a lower bound on the number of genotypes that need to be evolved before a global optimum can be found.

The black-box complexity model from [33] covers arbitrary search processes, which may be tailored to the problem class in hand and exhibit working principles very different from evolutionary processes. Lehre (Nottingham PI) and Witt [71] very recently presented a black-box theory tailored to evolutionary processes. In their model search processes are restricted to using “blind” mutation operators, in a sense that mutations are random and do not exhibit any bias towards favouring particular genotypes. This covers mutation operators in evolutionary algorithms as well as mutation operators that have yet to be invented. It also covers mutation in natural evolution as mutations are thought to be unbiased in nature. Therefore, this so-called *unbiased black-box complexity* exhibits limits to the speed of adaptation in *every asexual evolutionary process*. Further recent studies extended this model towards search processes using mutation and combinations of multiple solutions, hence encompassing mutation and recombination [28]. These black-box complexity models therefore give limits to the speed of adaptation in *any model of evolution*.

This perspective has recently emerged in evolutionary computation, and is completely novel in population genetics. We will translate existing results to our unified model of evolution from D 1.1 and estimate the black-box complexity in the above-mentioned models for further fitness landscapes investigated in work

packages WP 2, WP 3, and WP 4. Further, we will extend black-box complexity towards a model of structured, communicating populations to explore the limits of speed of adaptation in structured populations. This will enable a number of applications in both fields:

1. The black-box complexity of particular classes of fitness landscapes is interesting in its own right as it tells how many generations evolutionary processes need at the least to evolve a particular feature (see WP 2). Black-box complexity thus serves as a baseline for quantitative performance analyses of evolutionary processes.
2. Comparing the black-box complexity of several classes of fitness landscapes helps to determine the difficulty of such landscapes for evolutionary processes, which is vital for WP 2.
3. Comparing the black-box complexity across different black-box models, such as those only allowing for mutations vs. those that allow combinations of several solutions, helps to determine how powerful recombination is, feeding into WP 3.
4. Comparing the black-box complexity of structured populations with that of a single population helps to determine speedups achieved by multiple parallel populations and parallel implementations of EAs, feeding into WP 4.
5. Black-box complexity provides limits on the efficiency of *every evolutionary process*, thus forming a vital stepping stone towards our vision of a unified theory of evolution.

Deliverable D 1.4: A black-box complexity theory for models of natural of evolution, showing the inherent limitations of speed of adaptation on given landscapes. This will serve as a baseline for future quantitative performance analyses for all evolutionary models, by itself and when comparing performance across different fitness landscapes (WP 2), asexual vs. sexual reproduction (WP 3), and structured populations (WP 4).

Summary of WP 1

WP 1 will provide the foundation for bringing together PG and EC. Building a common language and a unified model of evolution will help to tear down the cultural divide between the two fields. We will enable synergies through knowledge transfer regarding existing results and techniques, starting with the similar models of SSWM and the (1+1) EA. Further, black-box complexity will provide general limits to the speed of adaptation of arbitrary evolutionary processes.

In addition to the above plans, further opportunities for unifying and integrating methods and models from PG and EC will be explored as they unfold throughout the course of the project.

Novelty and impact: This work package represents a concerted effort to remove barriers between the two fields. A systematic attempt at a unified quantitative theory has not been made before. In particular, to our knowledge a link between the SSWM domain in PG and the (1+1) EA in EC has not been made before. Also black-box complexity is unknown to PG. WP 1 will provide both fields with new results and tools, which will have a long lasting impact in each respective opposite field. Furthermore, this work package will provide us with the tools to tackle the rest of the work packages.

Expertise/track record: All PIs have previous experience with interdisciplinary work, see Section 2.2. All PIs have extensive experience with fundamental analysis methods of the two fields, having actively developed key methods. USFD introduced lower bounds via artificial fitness levels [111]. UNOTT pioneered unbiased black box complexity [28, 71], and introduced new techniques for population-based EAs (population drift, and fitness levels for populations [68, 69]). UNI-JENA introduced new techniques to analyse distributed random processes [27, 41] and extensively studied the computational complexity of multi-objective evolutionary algorithms [14, 15]. IST Austria has worked on the theory of dynamical systems and is trained as a physicist, giving ample background on the continuous methods used in PG. He has made contributions to many topics in theoretical biology and has an extensive background in interdisciplinary work [18, 23, 92, 93].

WP 2 Features of fitness landscapes and the speed of adaptation

The relationship between the set of all genotypes and their fitnesses is often described in terms of a *fitness landscape* [125]. This term is common to both PG and EC, and in both areas the fitness landscape has a decisive impact on the efficiency of natural evolution as well as evolutionary algorithms.

In this work package we identify three features of fitness landscapes that are critical for our understanding of evolutionary dynamics and explore them using the combined tools of both fields, as well as the new methods arising from WP D 1.3. We will translate results from models that are similar, attack models that

are dissimilar with tools from the opposite field and translate techniques and models from both fields in order to gain a deeper understanding of how features of the fitness landscape affect the speed of adaptation.

WP 2.1 Shifts between multiple peaks

Real-world optimisation and design problems are often difficult to solve because they exhibit a number of local optima, that is, peaks in the fitness landscape with unsatisfactory quality. One key strength of evolutionary algorithms lies in their ability to explore multiple peaks at the same time through a population of candidate solutions. However, it may happen that the population becomes trapped on one particular peak, and then has to rely on variation in order to reach a better peak. The process of how populations explore multiple peaks, and how they move through “fitness valleys” to get from one peak to another, is not well understood. Yet this understanding is key for explaining the success of evolutionary algorithms in many practical applications on complex fitness landscapes. Also natural populations are possibly faced with the same problem, but in natural populations it is virtually impossible to determine if local peaks exist. This is because of the enormous number of genotypes that would be needed to evaluate. Hence it is in practice impossible to rule out the existence of mutational paths connecting all high fitness genotypes [123]. However, the fact that speciation occurs and that different species do not interbreed provides a good argument for the existence of at least large fitness valleys separating different species.

The problem of peak shifts and fitness valley crossings has been studied in EC for nature-inspired search heuristics evolving a single solution, such as the (1+1) EA and simulated annealing [55], and in the context of explicit diversity-preserving mechanisms [42] or hybridisations of EAs with local search [110], involving members of our consortium. But studies for realistic populations are missing.

PG has considered peak shifts in greater detail, with regard to asexual [121] and sexual populations [122], with explicit formulas for the frequency of peak shifts [8], as a function of many evolutionary parameters, like population size and mutation rates. Peak shifts form a vital part of Wright’s *Shifting Balance Theory*, a foundational and highly debated theory in PG, which reasons that peaks shifts are a key factor in natural evolution. We deal with this in more detail in WP 4. Even though theory in PG is extensive, it is tailored specifically for a special kind of sampling process, and selection mechanism. In order for it to be useful in the analysis of generic genetic and evolutionary algorithms, it needs to be generalised.

We will develop a quantitative theory describing how quickly populations are able to cross fitness landscapes in various fitness landscapes with multiple peaks. Our methods will be inspired by the branching process methods in PG but generalised so that they can deal with arbitrary sampling methods and selection schemes. This will allow us to answer questions such as what is the expected time to cross a fitness valley, given an arbitrary sampling and selection scheme (arbitrary EA)? and how does this expected time depend on the dimensionality of the fitness landscape?

Deliverable D 2.1: A quantitative theory of speed of adaptation, with particular focus on how quickly populations are able to perform peak shifts on complex fitness landscapes.

WP 2.2 Neutrality in fitness landscapes

Another feature of fitness landscapes which can be challenging for natural and artificial evolution is regions of genotypes that all have indistinguishable fitness. Such regions of equal fitness values are called *plateaus* or *neutral networks* [102], and the term *neutrality* is used to reflect the fact that fitness is “neutral” in these regions. Neutrality can arise from redundancy in encodings which seems to be ubiquitous both in biological and artificial systems: the genetic code, that specifies how proteins are coded in the DNA, displays a significant proportion of redundancy in the assignment from triplets of base pairs to amino acids [39]; RNA folding, the base pairing mechanism that maps RNA sequences to secondary structures also exhibits vast neutral networks that span the genotype network [57]; and gene regulatory networks, at a higher description level, are also believed to be highly redundant [21, 90]. Since the fitness landscape of EAs is largely determined by the way in which candidate solutions are encoded, understanding whether adaptation is facilitated by neutrality would help guide the choice of solution encodings. Understanding the impact of neutrality on evolution is therefore a highly significant problem, both in evolutionary computation and population genetics.

Neutrality is highly debated both within evolutionary computation and population genetics as, depending on the situation, it can have detrimental or beneficial effects. The neutrality debate was launched by Kimura who, based on the expected rate of mutation in natural populations, argued that most evolutionary change should come from neutral or nearly neutral mutations, and that the role of selection would be merely to purge deleterious mutations (purifying selection) [59]. These neutral mutations do not have an effect on

fitness and thus cannot directly contribute to adaptation. If the selective force is absent from evolution, then it degrades to a random process. However, neutral mutations may change the fitness effects of subsequent mutations, potentially paving the way to a new adaptation. Thus, while neutrality can slow adaptation by reducing the immediate effectiveness of selection, neutral mutations can also enormously increase the speed of adaptation if they allow greater exploration in productive directions. Recently, it has been proposed that in biological systems these large neutral networks can promote “evolvability”, defined as the ability of a population to generate adaptive genetic variation [117]. Intuitively, it is clear that a more robust genotype will be less evolvable, since less of its neighbours will have a different selective value. However, the presence of these neutral mutations means that evolutionary change can still happen, potentially finding mutational paths that lead to further fitness increases. For example, it was found, in the RNA secondary structure example mentioned above, that all frequent structures are within a small mutational neighbourhood of any arbitrarily chosen sequence [38]. In contrast, in the absence of neutrality or neighbours of higher fitness, this genotype would be trapped at a local peak. Furthermore, when we consider a population evolving in one of these neutral networks, the existence of a large set of neutral neighbours allows for the rapid exploration of the local phenotypic neighbourhood. Neutrality can only affect evolvability if there is epistasis between the loci, since these neutral mutations need to enable adaptive mutations [116]. Thus, epistasis may significantly impact neutral network topology, and thereby population dynamics. While it seems then that neutrality can increase the potential for further adaptation, it is currently an open question how neutrality impacts the speed of adaptation.

There is little contention in the EC community that the presence of neutral networks impacts the speed of adaptation in evolutionary algorithms. However, it is poorly understood how neutrality changes the behaviour of EAs, and the existing results seem contradictory [22, 25, 40, 126]. Through empirical simulations and theoretical studies, it was found that arbitrarily adding neutrality can either increase or decrease the speed of adaptation. Further factors may therefore have to be taken into account, such as the shapes of neutral networks, how the neutral networks are interconnected or the mode of reproduction of the population [90]. All in all, the consensus is that the role of neutrality in EC is unclear and that rigorous theoretical studies are needed [43].

Population genetics and evolutionary computation have been prevented from making theoretical progress in this area for different reasons. The theory of EC has been developed by the need to understand the behaviour of EAs on practical optimisation problems, and have therefore developed techniques for handling complex fitness landscapes with many loci. However, EC has often studied overly simplified models of evolution, only taking into account a population of size one. Such simple models cannot provide any insight into population diversity on neutral networks, which is essential to fully understand evolutionary dynamics. PG has developed techniques for dealing with large populations which can be used to quantify the amount of neutral diversity that can be accumulated under different evolutionary regimes [60]. However, PG has mainly studied single-, two-, or three-loci models, and multi-locus models are less common [103]. The fitness landscapes that arise from models with few loci are too small and simple to contain the large and intricate neutral networks that are relevant to study.

Combining the complementary strengths of the two fields therefore presents a unique opportunity to make a significant breakthrough in the understanding of neutrality. EC will contribute with methods for handling complex fitness landscapes, with intricate neutral network topologies, and PG will contribute with methods for understanding the dynamics of large populations, for example by quantifying neutrality in a population using the diffusion approximation. As such, the extension of this technique for arbitrary selection and sampling processes we propose in D 1.3 will allow us to tackle questions such as:

- How does neutral network topology impact the expected time for a population to traverse it?
- How does the expected time depend on the specific sampling mechanism used?
- What is the population structure of a population traversing a neutral network? Does it proceed in a single cluster of individuals, or as a set of clusters?

Deliverable D 2.2: A quantitative theory on how neutrality affects the speed of adaptation, specifically on how the topology of a neutral network impacts population diversity and the speed of adaptation in evolutionary algorithms.

WP 2.3 Dynamic fitness landscapes and changing environments

Natural populations have to cope with a changing environment, not just because the environment itself is dynamic, but also because of the ecological feedback that results from their own, as well as other species’,

effects on the environment. For example, a dynamic environment can result from random environmental fluctuations (for example due to the intrinsic noise of gene expression [92]) or from frequency dependent selection (for example, in sex-ratio evolution [93]).

The same holds true for many real-world optimisation problems: customer demands can change unexpectedly, which may imply a need to refine the optimisation criteria. Constraints as well as the objective function may change over time.

Existing runtime analyses in EC have widely neglected this problem; only very initial analyses for simple populations are known so far [30, 53, 99]. We will extend our quantitative theory of peak shifts from D 2.1 to dynamic fitness landscapes, targeting settings where peaks move in genotype space, or landscapes where peaks change in height, so that peak shifts may become necessary in order to jump from a deteriorating peak to a better one.

In order to reach this goal, we will extend the techniques used to obtain the frequencies of peak shifts in PG for arbitrary stochastic processes. This will involve the extensions to the diffusion approximation, obtained in D 1.3, as well as extending the branching processes techniques for more general selection and sampling processes.

Deliverable D 2.3: A quantitative theory of the speed of adaptation in dynamically changing fitness landscapes based on D 2.1, paving the way for identifying EA designs that maximise adaptation, and recommendations on how to design EAs for dynamic optimisation problems.

Summary of WP 2

WP 2 will provide the foundations of a quantitative theory of speed of adaptation in natural and artificial evolution, and how it depends on features of the fitness landscape. Multiple peaks and neutrality are important concepts, and a better understanding of how these features affect evolution has high significance for theory and practice of EC. It will provide a foundation on which to build a classification of problems in order to better match algorithms and problems. Studying dynamic fitness landscapes will foster EC and PG's understanding of how organisms evolve in a changing environment.

All these insights will establish a solid foundation for the investigation of recombination (WP 3) and parallel, structured populations (WP 4).

Novelty and impact: The novelty of our approach lies in analysing realistic populations on multi-locus fitness landscapes, combining techniques from EC and PG. Studying dynamic fitness landscapes is a highly novel topic, which has not been explored so far.

Expertise/track record: UNOTT studied how runtime depends on accessibility between neutral networks [70], and on neutrality arising from complex genotype-phenotype mappings [67]. UNI-JENA and USFD previously collaborated on fitness-valley crossings through explicit diversity mechanisms [42] and both have further relevant publications on neutrality (UNI-JENA, [40]) and fitness-valley crossings by hybrid EAs and simulated annealing (USFD, [110]). IST Austria has recently started a research programme on the impact of epistasis on the evolutionary process and has provided fundamental insights on how recombination interacts with neutral networks [90]. Finally, UNOTT provided a pioneer study of dynamic evolutionary optimisation [99].

WP 3 Recombination and the speed of adaptation

WP 3 is laid out to extend our quantitative theory of speed of adaptation towards sexual reproduction, thereby addressing the fundamental question when and how using sexual reproduction, i. e., recombination, increases the speed of adaptation. Understanding why recombination improves evolution is a major open problem in both PG and EC. In PG, recombination, or sexual reproduction, is a famous open question and called “the queen of problems in evolutionary biology” by Graham Bell [11] and others. Asexual reproduction permits efficient fitness increases and seems to adapt faster [9], but can also reduce genetic variation, which potentially decreases fitness of the offspring [106]. Sexual reproduction allows more variation, but is considered more costly. However, all known organisms engage in some form of sexual recombination. Even very simple asexual life forms like diatoms (which are unicellular photosynthetic organisms and part of the plankton community) can temporarily switch to sexual reproduction when the size of a cell is less than a critical threshold [50]. There appears to be a fundamental advantage of sexual reproduction in nature, which is still not fully understood.

Also in EC the advantage of sexual reproduction is not very well understood. There it is known that complex real-world optimisation problems can benefit from using recombination [20] (also called **crossover**), but theoretical results and explanations on recombination are still scarce. There are some results in pseudo-

Boolean optimisation which construct artificial test functions where it can be proven that a genetic algorithm with recombination asymptotically outperforms an evolutionary algorithm without recombination [54, 56, 61, 72, 85, 107, 118] and the other way around [98]. These test functions are typically tailored to the specific algorithm and proof technique, so that their mathematical proofs give only little insight in the intuitive advantage of sexual reproduction. There are also rigorous proofs that recombination is beneficial for some problem-specific algorithms and representations, namely colouring problems inspired by the Ising model [109] and the all-pairs shortest path problem [29]. In this work package we use and extend the techniques to handle populations developed in WP 2, to study the behaviour of sexual reproduction and its benefits for EC.

WP 3.1 Recombination in different fitness landscapes

The first step in understanding the benefit of recombination will be to extend our results from WP 2 regarding the speed of adaptation of asexual populations on fitness landscapes with multiple peaks (deliverable D 2.1) and neutrality (deliverable D 2.2). The main question here is whether introducing recombination will facilitate adaptation, and if so, to quantify this benefit for common recombination mechanisms.

Evolutionary processes with recombination are much harder to analyse than those with mere asexual reproduction because of the complex interaction between the stochastic processes [5, 96, 97]. We will draw on existing results in PG theory, mapping simple, already-understood PG scenarios to EAs, in a similar spirit to WP 2. This will have the immediate benefit of gaining a deeper understanding of how and under which circumstances recombination provides a benefit for a given EA, at least for simple fitness landscapes. We will then extend this analysis to the more complicated landscapes and broader range of diversity mechanisms found in EC. We will also incorporate new, integrated methods developed in WP 1 (deliverable D 1.3).

Deliverable D 3.1: A quantitative theory of the speed of adaptation in evolutionary processes with recombination on fitness landscapes investigated in WP 2, extending our theory for asexual populations (deliverables D 2.1 and D 2.2). This theory will lead to a better understanding of this genetic operator and enable the design of better recombination operators.

WP 3.2 Investigating the Hill-Robertson effect

In PG, there are several potential explanations for the advantage of sex. The currently favoured explanations revolve around the Hill-Robertson effect, the fact that alleles simultaneously spreading in the same population can interfere with each other and reduce the speed of adaptation [100]. This can be best understood with a simple example: Due to finite population sizes, random sampling of individuals will generate random associations between a newly generated allele and an already spreading allele. If these associations are positive, meaning that the two beneficial alleles are on the same individual genotype, this will result in a rapid fixation of these alleles. However, if the associations are negative, meaning that the two alleles are in different individuals, they will compete with each other, leading to a decrease in the speed of adaptation. This can be ameliorated by an increase in recombination rate, leading to a faster coupling of the two alleles. This latter effect is dominant because in the situations where negative associations are generated they tend to last longer than ones where the two alleles are in the same individual. This effectively constitutes a limit for the speed of adaptation [120].

We will extend the analysis of the Hill-Robertson effect to scenarios dealing with different mechanisms for maintaining diversity or sampling individuals in a recombining population. This will provide a central piece for the understanding of the speed of adaptation in evolutionary algorithms with recombination. In return, this will allow for testing the relevance of this effect in more general settings. Furthermore, we will extend the classic analysis of the Hill-Robertson effect to more general fitness landscapes, including neutrality, fitness valleys, and epistasis. This will increase the understanding of both fields about the importance of recombination for the speed of adaptation.

Characterising the Hill-Robertson effect, that is the interference between mutations simultaneously spreading in a population, under more general sampling schemes, is crucial for the effective use of populations in EAs. It will enable a more general theory for how sex affects the speed of adaptation on a wide range of fitness landscapes.

Deliverable D 3.2: An analysis of the Hill-Robertson effect in general sampling and variation schemes, thus increasing the understanding of the importance of this effect in natural population and leading to a better prediction of the speed of adaptation in EAs.

WP 3.3 Optimal parameters in the presence of recombination

Another observation in EC is that there the benefits of recombination seem to be linked to the specific selection scheme and the mutation rate. For elitist selection, where the best among parents and offspring survive, there are functions where adding recombination implies that the optimal mutation rate increases [108]. On the other hand, in generational systems, where only the offspring survive, there are other examples where adding recombination implies that the optimal mutation rate decreases [86]. The reason is supposedly that recombination increases the variance in the search process in terms of the fitness of offspring. Confirming this and comparing it with models from PG is an open problem. Answering it will increase our understanding of the benefits of sexual reproduction. We also want to study how recombination impacts population diversity in evolutionary algorithms, both in terms of fitness diversity and genotypic diversity.

A more ambitious direction that may be explored is using adaptive parameter settings, which are adjusted automatically according to the progress of the evolutionary process. Even self-adaptive processes might be analysed – where evolutionary parameters like the recombination rate or the mutation rate are evolved alongside the original genotypes. This will be achieved by adapting recombination modifiers methods typically used in PG [100]. These methods assess the short-term fate of a locus that does not contribute to fitness but that changes the recombination rate. Such a procedure will enable assaying whether increasing the recombination rate will be favoured, given the current state of the population. This will allow the use of EAs that effectively adapt their recombination rate to the selection pressures the population is experiencing.

Deliverable D 3.3: Analysis of the interplay of recombination rates with other parameters: mutation rates and selection intensity. Quantifying the co-dependence of these crucial parameters will enable users of EAs to set parameters in an informed fashion, increasing the efficiency and accessibility of EAs.

Summary of WP 3

Even though sexual reproduction has proven very beneficial in both EC and in natural evolution, the fundamental reasons are not well understood. WP 3 will bring together insights and methods from PG and EC to develop a quantitative theory of sexual reproduction and deepen our understanding of this “queen of problems”. In addition to better understanding natural evolution, this work package will enable the use of recombination operators in EC in two ways: identifying and quantifying co-dependencies between crucial design parameters in EAs, and through designing more powerful recombination operators.

Novelty and impact: New methods to design recombination operators in a problem specific way. Evaluating long standing theories in more general settings, (e.g. Hill Robertson with epistasis). New knowledge and insight about the evolutionary process about one of the main open problems in both fields (new limits to the speed of adaptation, coming from phenomena in PG, best strategies to overcome these limits.)

Expertise/track record: UNI-JENA worked on diversity of populations [42]. USFD has worked on the usefulness of recombination, having presented the first non-constructed example where recombination leads to exponential performance gains [109] as well as studies on popular test functions [61, 112]. IST Austria has been investigating the effect of recombination on neutral landscapes, under the usual PG assumptions, both in explicit genotype-phenotype maps [90] and more general mutational landscapes [93]. UNOTT looked at recombination operators in software testing [72].

WP 4 Parallel evolution in structured populations

Evolutionary algorithms have given rise to many parallel variants [3, 79, 83, 113] 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 is needed to satisfy increasing demands in a modern world requiring larger and larger problems to be solved efficiently.

There are several ways of parallelizing EAs. One is to use a master-slave architecture, where a master process coordinates many slave processes on different processors who execute evolutionary operators in parallel, like creating offspring through selection and variation, and evaluating their fitness. However, this requires intensive communication between the master and all its slaves.

Another, very popular approach is to parallelize evolution by evolving different populations that are run on different processors [79]. These populations run independently most of the time, but they periodically communicate genetic information to coordinate search. Communication takes place according to a spatial structure, a topology connecting populations. Common topologies include rings, two-dimensional grids or toroids, hypercubes, or complete graph with all possible connections. Figure 1.3b illustrates some of these topologies. The process of communication is called *migration* or *pollination*, where in the former case

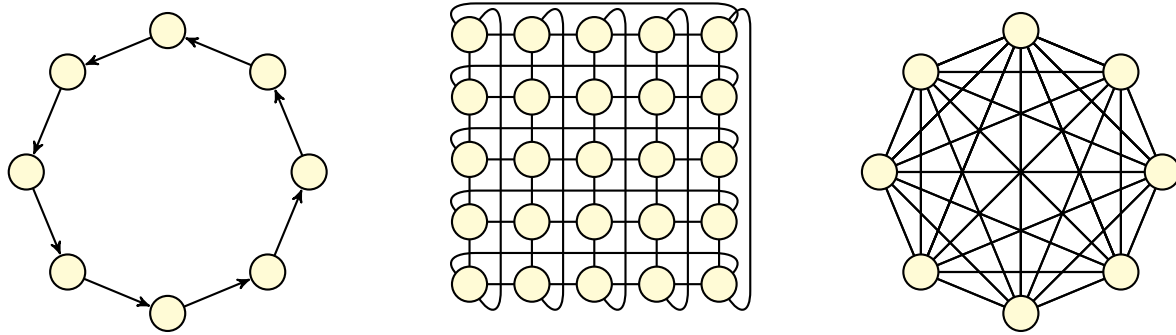


Figure 1.3b: Structured populations in parallel EAs exemplifying three common topologies: unidirectional ring, toroid, and complete graph. Migration is performed along connections between populations. These connections can be unidirectional or bidirectional.

individuals are removed from the original population and in the latter copies of those individuals are being sent.

The main advantages of evolving parallel populations in this way are:

- 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 [2].
- Periodic communication only requires small bandwidth, leading to low communication costs.
- Solution quality is improved as different populations can effectively 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 [4, 79].

However, designing an effective parallel evolutionary algorithm can be challenging as the method and amount of communication needs to be tuned carefully. Too frequent communication leads to high communication costs, and it can overwhelm the parallel system, thus compromising exploration. Too little communication means that the populations become too isolated and unable to coordinate their searches effectively. Leading researchers on parallel evolutionary algorithms acknowledge that even the effect of the most fundamental parameters on performance is not well understood [4, 79].

Many questions remain open, such as: when and why does communication between spatially distributed populations speed up adaptation? How does the spatial structure of distributed populations and the amount and strategy for communication impact the speed of adaptation? How to find optimal communication and parameter setups to most effectively exploit multi-core and parallel architectures in evolutionary computation?

Answering these questions is an ambitious task that requires a novel interdisciplinary approach. We will build on our foundational methods from WP 1 and our quantitative theory on different fitness landscapes from WP 2 to provide a theory of speed of adaptation in structured populations. In particular, results on the black-box complexity of parallel populations (Deliverable D 1.4) will be used as a baseline for measuring the speedup obtained by evolving multiple populations instead of just a single population.

WP 4.1 Wright's shifting balance theory for parallel EAs

The main obstacle in understanding how parallel EAs work is that they combine two levels of dynamic processes: *intra-population dynamics* describe the evolution within individual populations, and *inter-population dynamics* describe communication between parallel populations.

A parallel EA is analogous to isolated biological sub-populations that diverge and adapt in different ways. The theory in PG is extensive and has considered structured populations since its classic foundations. It has been found that the degree of population structure can have a large impact on the probability that a variant invades (see, for example, [23]). Most importantly, it already provides a framework for an integrated analysis of parallel communicating populations.

Wright's Shifting Balance Theory (SBT) is one of the most foundational and debated evolutionary theories, yet it provides an integrated analysis which may prove very beneficial in the context of parallel

EAs, as discussed in the following. SBT considers populations divided into small sub-populations adapting on a fitness landscape with multiple peaks. The theory has been quantified mathematically with stochastic methods; it represents a cornerstone in evolutionary theory as it represented an alternative to Fisher's view of adaptive evolution, in which adaptation would come about solely through mass selection. Many of the mathematical and conceptual foundations of modern evolutionary theory were obtained in the context of SBT, including the very concept of a fitness landscape and the island model of population subdivision. SBT provides an integrated view of evolution, since it combines several evolutionary forces and features of the genetic architecture: drift, migration, selection and epistasis all play a major role in SBT.

According to Sewall Wright, adaptation happens through the interplay of several evolutionary forces, each with special relevance at the different stages of a three phase process:

- During Phase I genetic drift dominates, spreading populations across the fitness landscape, away from the current peak and through fitness valleys, towards new peaks.
- In Phase II, selection dominates, pulling populations towards new fitness peaks
- Finally in Phase III, populations at different fitness peaks compete with each other through migration, leading to the spread of the fittest gene combinations.

Phases I and II act within each population, and occur independently across the several populations, while Phase III is where populations interact with each other through migration. So, SBT involves a phase of parallel search across the fitness landscape (Phases I and II) and a phase of communication between populations. Wright formalised the different stages of the process, leading to several foundational mathematical models of PG. For example, he introduced the deme/island model of migration, calculated the probability of shifts between fitness peaks, and derived the distribution of allele frequencies under selection, mutation and drift.

Due to its integrated perspective on evolution, SBT makes for an ideal tool for the analysis of parallel EAs. The following arguments present criticism raised against the SBT from the perspective of natural populations and at the same time reveal why this critique does not apply to EC.

- Wright's SBT assumes that all three phases take place at the same time. This makes it hard to derive a complete mathematical description. However, parallel EAs are, by nature, periodic systems and naturally imply a subdivision into phases: migration and periods of independent evolution. Phases I and II occur between migrations, and Phase III precisely describes the process of migration.
- In order for Phase III to occur as Wright envisaged it, fitter populations must spread across the rest of the species. Wright envisaged that populations of higher fitness would produce more migrants, so that they would spread across the other demes. However, in natural populations the effect of selection is generally too weak to allow fitter individuals to spread efficiently. In contrast, migration policies in parallel EAs are much more effective in spreading high-fitness individuals. Instead of moving high-fitness individuals between populations, often copies thereof are being sent, so that the original individuals remain on the original island, and the number of high-fitness individuals increases. This process is often called *pollination* (though we will stay with the more general term *migration* in the remainder). In addition, parallel EAs provide means for tuning the migrant pool to ensure that only high-fitness individuals are being sent, and that high-fitness immigrants are included on target islands' populations. This contributes to decoupling the number of migrants from either population density or fitness.
- In natural populations Phases I and II require a delicate and sometimes paradoxical balance of exploration and exploitation. On the one hand, evolution must be able to cross through fitness valleys under the action of drift (Phase I). On the other hand, evolution must be able to climb peaks (Phase II). This requires a delicate balance between parameters such as population size, selection strength, and fitness differences between the peaks and the valley, which seems implausible in natural populations. However, a designer of a parallel EA has control over many parameters, including dynamics of population size, selection strength, migration rate, etc. By carefully tuning these parameters, one can enable populations to cross fitness values and climb better peaks during a period of independent evolution. In addition, and where finding the right balance proves tricky, there is a wealth of more explicit mechanisms that can encourage the desired behaviour:

- **Time-dependent parameters:** Time-varying selection mechanisms can be used to gradually increase the selection strength during such a period. This is analogous to the strategy used in simulated annealing, but with much better analytical understanding.
- **Diversity mechanisms:** EC provides a number of explicit diversity-preserving mechanisms that encourage populations to spread out. Mechanisms like niching, crowding, and fitness sharing all refine selection to favour individuals that are dissimilar to other members of the population [42]. This makes it easier for individuals to leave a low-fitness peak and to traverse fitness valleys.
- **Heterogeneous populations:** Parallel EAs can use heterogeneous population models, where different populations use different parameters [79]. If different selection strengths are being used, this increases the chance that some populations have the right exploration-exploitation balance for any particular stage of the optimisation process. This even applies if the optimal balance changes during the course of the artificial evolution.

This exposition shows that the way how parallel EAs are designed effectively alleviates the major shortcomings of the SBT, implying that this cornerstone of evolutionary theory seems better suited for parallel EAs than it is for natural populations themselves.

Early work by Oppacher and Wineberg [88, 89] has explored this idea, leading to the *Shifting Balance Genetic Algorithm* (SBGA), which incorporates some of the above ideas. The SBGA evolves a large *core* population alongside smaller *colony* populations. An explicit diversity mechanism drives colony populations away from the core population (Phase I), enabling them to find new peaks (Phase II). Colony populations send migrants to the core population, enabling it to implement Phase III. Empirical results for the SBGA were promising in that the SBGA performed better than traditional genetic algorithms [88]. However, the SBGA is not well suited for implementation on many-core architectures as the large core population represents a bottleneck for parallelization.

We aim at providing a powerful theory of speed of adaptation in structured populations, focussing on analytical results with a high degree of rigour, as it will be based on a solid theoretical foundation developed in WP 1 and WP 2, particularly the study of peak shifts from Deliverable D 2.1.

Thereby we will study homogeneous or heterogeneous populations that require the same computational effort and hence are suitable for implementation in many-core architectures. By combining the SBT with recent runtime analyses of simple parallel EAs, we will be able to quantify performance of parallel EAs that more closely reflect parallel EAs used in practice, and to build a solid foundation for the theory of parallel EAs. Studying the speed of adaptation will allow for predictions on the performance of parallel EAs in illustrative settings. As a result, we envisage recommendations on optimal designs and parameter choices.

To achieve this ambitious goal, we will first derive a basic integrated model of evolution in structured populations. This will be done by tailoring and formalising the SBT to match common setups in parallel EAs. In this first step we will retain the original selection and migration models from SBT to build a first model for quantitative performance studies. This also implies that migration can take place between any two populations, referred to in EC as a complete topology.

Deliverable D 4.1: A quantitative theory of speed of adaptation in structured populations integrating the analysis of parallel EAs with Wright’s Shifting Balance Theory (SBT), enabling quantitative performance studies of parallel EAs with specific selection and migration models, restricted to complete topologies.

WP 4.2 Extension to common topologies, selection and migration policies

The basic integrated model from D 4.1 will be extended in various ways, to enhance its applicability to parallel EAs.

One extension is to take into account different communication topologies, that is, restricting migration to only occur between islands that are neighbouring in the topology. We will incorporate ring topologies, two-dimensional grids, hypercubes, and complete topologies. Our models of migration have to be adapted to these specific topologies. UNI-JENA has extensive experience in stochastic communication processes, where information is being communicated in a network of nodes [26, 27]. This is highly relevant for this research as high-fitness individuals represent genetic information, which is spread probabilistically through a network of distributed populations.

Furthermore, we will incorporate common selection mechanisms and migration policies used in EC. This framework will allow us to perform quantitative studies of the performance of parallel EAs with realistic populations, which goes far beyond USFD’s pioneering work regarding parallel EAs evolving single individuals [62, 63, 64, 65, 66, 80, 85]. Wherever appropriate, our framework may also include some of

the mentioned design features of parallel EAs such as time-varying parameters, diversity mechanisms and heterogeneous models.

Deliverable D 4.2: An extension of the quantitative theory of speed of adaptation in structured populations (D 4.1) towards common selection mechanisms and migration policies used in EC. This theory will enable quantitative performance studies regarding specific migration models, across different communication topologies and common selection mechanisms used in EC.

WP 4.3 Application to parallel EAs

Our integrated theory will be put to the test by performing quantitative performance studies for parallel EAs, demonstrating its practicability and power. We will target problems from combinatorial optimisation (including, but not limited to those studied in [66, 84, 85]) and address the following questions, which are crucial for the design of parallel EAs:

- what **speedups** can be obtained when using multiple identical populations instead of a single one?
- when are **many small parallel populations** more effective than a **single large population**?
- **how many parallel populations** shall be used in particular applications?
- what is the most effective **frequency and structure of communication** between populations?
- when and why are **heterogeneous populations** better than homogeneous ones?

Answers to these questions will develop our understanding of how parallel EAs work, and how to design effective EAs for a variety of problems.

Deliverable D 4.3: Quantitative performance analyses of parallel EAs in combinatorial optimisation, obtained through our quantitative theory (D 4.2), and recommendations on how to design parallel EAs with regard to choosing communication topologies, selection mechanisms, and migration policies.

Summary of WP 4

The goal of WP 4 is to provide a powerful integrated model of evolution in structured populations, combining existing analyses of parallel EAs with Wright's Shifting Balance Theory (SBT). SBT represents one of the cornerstones and most debated theories in PG, which turns out to be ideally suited for analysing parallel EAs. Our integrated model will be applied to problems from combinatorial optimisation and result in recommendations on how to design effective parallel EAs.

Novelty and impact: The theory of parallel EAs is in its very infancy, lacking a fundamental understanding of how design parameters affect performance. The only existing analyses consider parallel populations each evolving a single individual. SAGE will provide the first performance analyses of realistic populations with communication, leading to recommendations that will enable the design of effective parallel EAs.

The potential impact is enormous due to the development in computer architectures. The International Roadmap for Semiconductors estimates that the number of cores in modern devices will double roughly every two years, leading to 1 000 CPU cores in 10 years time [51]. Our quantitative theory of evolution in structured populations will provide insight and guidance for the design of effective EAs on massively parallel hardware, satisfying increasing demands for better solution quality, increasing problem sizes, and optimisation in real time.

Expertise/track record: USFD has pioneered the runtime analysis of parallel evolutionary algorithms with each parallel population evolving one individual [62, 63, 64, 66] and has presented analyses and design recommendations for various problems, including combinatorial optimisation [65, 80, 85]. UNI-JENA provides expertise on probabilistic communication processes in spatial structures [26, 27]. IST Austria has studied how population structure and migration rates affect the evolution of toxin-antitoxin systems [23].

Risk assessment

Our proposed work is of a fundamental and interdisciplinary nature and as such, it is risky. Here we address potential pitfalls and difficulties.

- **Incompatible mathematical descriptions.** Some of the tools we plan to integrate are clearly related: drift analysis and the diffusion approximation draw from the same theory of analysis of stochastic processes. Other approaches do not seem to have an immediate equivalent in the other field. However, the mere translation of these methods into each others language will undoubtedly prove useful, opening new areas of research in both fields.
- **Usefulness of models.** There is the possibility that the models that we will generate are not appropriate for the other field, limiting their usefulness for real world applications. For example, it could be that certain EAs are not meaningful in a natural setting. However, even these unconventional algorithms may find applications in some subfields of biology, such as artificial selection. Similarly, certain models of natural evolution may not be well suited for optimisation. Therefore, it may be necessary to abstract from biological models and to extend theoretical models of PG towards such more abstract settings. The enormous success and popularity of EAs provides convincing evidence that taking inspiration from natural evolution can yield powerful tools for optimisation.
- **Limits to generality.** One inherent risk of any project aiming at a unifying framework is that it may not be possible to reach the desired level of generality. When aiming at quantitative analyses of the speed of adaptation in evolution, the risk is not whether such an analysis can be achieved, but to what degree this can be done. In particular, the risk is not whether it is possible to develop a model that unifies some form of biological and artificial evolution (as exemplified by the similarity between the (1+1) EA and the SSWM), but rather how general such a unified model of evolution can be. There are low-risk options such as studying simple models of evolution on simple problems. These options will be used as proofs of concepts en route to a more general and more powerful theory. The level of ambition, difficulty and risk can be tuned in several ways:
 - by aiming at more sophisticated and more general models of evolution
 - by analysing more complex problems and problem classes and
 - by aiming at a high degree of precision and accuracy in our quantitative studies vs. giving loose upper and lower bounds on the speed of adaptation

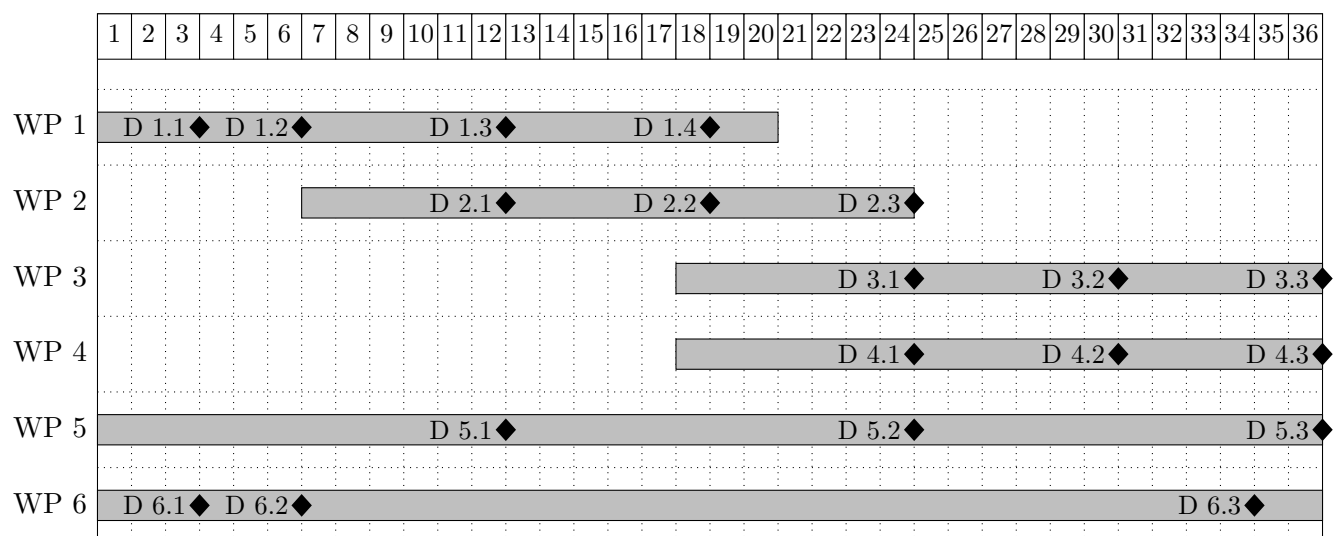
When encountering difficulties in establishing a general and precise quantitative theory, straightforward contingency plans are to focus on simpler evolutionary processes, simpler or more concrete problems, models of settings, or to derive less precise quantitative results, for instance through loose upper and lower bounds on the speed of adaptation.

This will have the side effect of illuminating what the limitations of existing techniques are. Moreover, finding new fundamental limitations to the generality of theories of evolution would be an important result in itself and would establish the need for further research.

- **Balance between PG and EC.** SAGE brings together three sites from EC (UNOTT, UNI-JENA, USFD) and one site from PG (IST Austria), so a potential risk is that sufficient expertise from PG may not be available at all times to produce all deliverables on time and to a high standard. IST Austria will therefore employ two full-time project members, twice as many as the partners from EC, to achieve a ratio of 3:2 full-time project members in EC and PG, respectively. Our project partners at IST Austria are embedded in a thriving and world-leading group from PG, hosting a number of postdocs and PhD students with highly relevant expertise, thus providing an excellent pool for recruitment, contingency plans for unforeseen circumstances, as well as potential collaborators. The composition of our consortium as well as required and provided expertise is discussed in greater detail in Section 2.3.

Tables

Time line for the project. The following Gantt Chart shows the time line for the project, measured in months. Diamonds indicate deliverables as stated in the work programme and work package description.



Work package interdependencies. The following Pert Chart shows the dependencies between the technical deliverables. Arrows denote input from one deliverable into another one. To increase readability, we only show immediate dependencies. This means if there are dependencies $a \rightarrow b$ and $b \rightarrow c$, there is also a dependency $a \rightarrow c$, which is not explicitly shown. The set of all dependencies can therefore be derived from the transitive closure of the shown digraph.

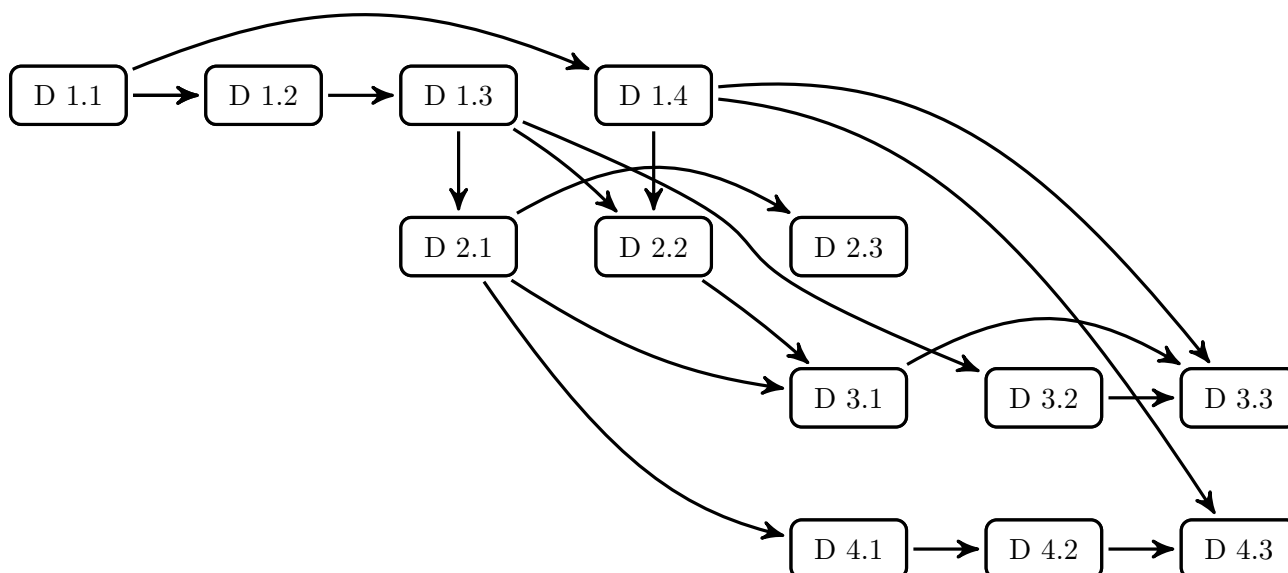


Table 1.3a: Work package list

WP no.	Work package title	Type of activity	Lead pt. no.	Lead pt. short name	Person-months	Start month	End month
WP 1	Unified model of evolution	RTD	2	IST Austria	54	1	20
WP 2	Features of fitness landscapes and the speed of adaptation	RTD	1	UNOTT	43	7	24
WP 3	Recombination and the speed of adaptation	RTD	3	UNI-JENA	46	18	36
WP 4	Parallel evolution in structured populations	RTD	4	USFD	50	18	36
WP 5	Coordination and management	MGT	1	UNOTT	8	1	36
WP 6	Dissemination and exploitation	MGT	1	UNOTT	7	1	36
		TOTAL			208		

Table 1.3b: List of deliverables

Del. no.	Deliverable name	WP no.	Nature	Dissemination level	Delivery date
D 1.1	Unified model of evolution	1	R	PU	3
D 1.2	Translation of results	1	R	PU	6
D 1.3	Integrating and developing methodology	1	R	PU	12
D 1.4	Black-box complexity: limits of evolution	1	R	PU	18
D 2.1	Shifts between multiple peaks	2	R	PU	12
D 2.2	Neutrality in fitness landscapes	2	R	PU	18
D 2.3	Dynamic fitness landscapes and changing environments	2	R	PU	24
D 3.1	Recombination in different fitness landscapes	3	R	PU	24
D 3.2	Investigating the Hill-Robertson effect	3	R	PU	30
D 3.3	Optimal parameters in the presence of recombination	3	R	PU	36
D 4.1	Wright's shifting balance theory for parallel EAs	4	R	PU	24
D 4.2	Extension to topologies, selection and migration policies	4	R	PU	30
D 4.3	Application to parallel EAs	4	R	PU	36
D 5.1	First year scientific report	5	R	RE	12
D 5.2	Second year scientific report	5	R	RE	24
D 5.3	Third year scientific report	5	R	RE	36
D 6.1	Kick-off workshop	6	O	RE	3
D 6.2	Website and Wiki	6	O	PU	6
D 6.3	Final workshop	6	O	PU	34

Table 1.3c: List of milestones

SAGE will follow an integrative and iterative process developing a unified quantitative theory of evolution and achieving the listed objectives. Thus, all milestones are listed as deliverables, see Table 1.3b.

Table 1.3d: Work package description (WP 1)

Work package number	1	Start date:			Month 1
Work package title	Unified model of evolution				
Activity type	Research and technological development (RTD)				
Participant number	1	2	3	4	Total
Participant short name	UNOTT	IST Austria	UNI-JENA	USFD	
Person months	10	20	12	12	54
Objectives	The first work package is designed to provide the foundation that will allow a sustainable collaboration between population genetics and evolutionary computation, and the development of a new scientific field. Results and techniques from both fields will be transferred via a unified model and terminology. New analytical techniques for deriving results about artificial and natural evolution will be developed. Fundamental limitations to the speed of evolution will be explored.				
Description of work	<div>D 1.1 Unified model of evolution<ul style="list-style-type: none">Description: An abstract general model of evolution, based on a unified terminology, that allows for direct comparison between EC algorithms and PG regimes. A translation of terminology from PG and EC into our unified model. See page 7 for details.Primary site: IST AustriaContributing sites: UNOTT, UNI-JENA, USFD</div> <div>D 1.2 Translation of results<ul style="list-style-type: none">Description: A translation of results between the different evolutionary regimes intended as a tool to identify synergies between PG and EC, and the benefits resulting from knowledge transfer. See page 7 for details.Primary site: UNOTTContributing sites: IST Austria, UNI-JENA, USFD</div> <div>D 1.3 Integrating and developing methodology<ul style="list-style-type: none">Description: New methods for analysing the speed of adaptation, obtained by transferal or recombination of existing methods in the different fields, such as applying the diffusion approximation to different evolutionary algorithms, enabling quantitative performance analyses of natural and artificial selection. See page 8 for details.Primary site: IST AustriaContributing sites: UNOTT, USFD</div> <div>D 1.4 Black-box complexity: limits of evolution<ul style="list-style-type: none">Description: A black-box complexity theory for models of natural evolution, showing how the inherent limitations of speed of adaptation on a given landscape. See page 9 for details.Primary site: IST AustriaContributing sites: UNOTT, USFD</div>				
Deliverables	Month 3: D 1.1 Unified model of evolution Month 6: D 1.2 Translation of results Month 12: D 1.3 Integrating and developing methodology Month 18: D 1.4 Black-box complexity: limits of evolution				

Table 1.3d: Work package description (WP 2)

Work package number	2	Start date:			Month 7
Work package title	Features of fitness landscapes and the speed of adaptation				
Activity type	Research and technological development (RTD)				
Participant number	1	2	3	4	Total
Participant short name	UNOTT	IST Austria	UNI-JENA	USFD	
Person months	18	16	3	6	43
Objectives	The second work package is designed to uncover how the speed of adaptation in an asexually evolving population depends on characteristics of the fitness landscape. The results will make it easier to identify which algorithms are appropriate for a given fitness landscape. The results will serve as a foundation for the analysis of more complex evolutionary models in WP3 and WP4.				
Description of work	<p>D 2.1 Shifts between multiple peaks</p> <ul style="list-style-type: none">• Description: A quantitative theory of speed of adaptation, with particular focus on how quickly populations are able to perform peak shifts on complex fitness landscapes. See page 11 for details.• Primary site: UNOTT• Contributing sites: IST Austria, USFD <p>D 2.2 Neutrality in fitness landscapes</p> <ul style="list-style-type: none">• Description: A quantitative theory on how neutrality affects the speed of adaptation, specifically how the topology of a neutral network impacts population diversity and the speed of adaptation in evolutionary algorithms. See page 11 for details.• Primary site: UNOTT• Contributing site: IST Austria <p>D 2.3 Dynamic fitness landscapes and changing environments</p> <ul style="list-style-type: none">• Description: A quantitative theory of the speed of adaptation in dynamically changing fitness landscapes based on D 2.1. See page 12 for details.• Primary site: UNOTT• Contributing site: IST Austria				
Deliverables	<p>Month 12: D 2.1 Shifts between multiple peaks</p> <p>Month 18: D 2.2 Neutrality in fitness landscapes</p> <p>Month 24: D 2.3 Dynamic fitness landscapes and changing environments</p>				

Table 1.3d: Work package description (WP 3)

Work package number	3	Start date:			Month 18
Work package title	Recombination and the speed of adaptation				
Activity type	Research and technological development (RTD)				
Participant number	1	2	3	4	Total
Participant short name	UNOTT	IST Austria	UNI-JENA	USFD	
Person months	6	17	17	6	46
Objectives	This work package will describe what role recombination has in evolution, in particular with respect to the speed of adaptation, and how it interacts with other aspects of the evolutionary process, such as mutation and selection intensity. This knowledge will lead to recommendations on how to design more powerful recombination operators.				
Description of work	<div>D 3.1 Recombination in different fitness landscapes</div> <div><ul style="list-style-type: none">Description: A quantitative theory of the speed of adaptation in evolutionary processes with recombination on fitness landscapes investigated in WP 2, extending our theory for asexual populations (deliverables D 2.1 and D 2.2). See page 14 for details.Primary site: UNI-JENAContributing sites: IST Austria, USFD</div> <div>D 3.2 Investigating the Hill-Robertson effect</div> <div><ul style="list-style-type: none">Description: An analysis of the Hill-Robertson effect in general sampling and variation schemes, thus increasing the understanding of the importance of this effect in natural populations and leading to a better prediction of the speed of adaptation in EAs. See page 14 for details.Primary site: IST AustriaContributing sites: UNI-JENA, UNOTT</div> <div>D 3.3 Optimal parameters in the presence of recombination</div> <div><ul style="list-style-type: none">Description: Analysis of the interplay of recombination rates with other parameters, including mutation rates and selection intensity. See page 15 for details.Primary site: USFDContributing sites: IST Austria, UNI-JENA</div>				
Deliverables	<div>Month 24: D 3.1 Recombination in different fitness landscapes</div> <div>Month 30: D 3.2 Investigating the Hill-Robertson effect</div> <div>Month 36: D 3.3 Optimal parameters in the presence of recombination</div>				

Table 1.3d: Work package description (WP 4)

Work package number	4	Start date:			Month 18
Work package title	Parallel evolution in structured populations				
Activity type	Research and technological development (RTD)				
Participant number	1	2	3	4	Total
Participant short name	UNOTT	IST Austria	UNI-JENA	USFD	
Person months	6	17	8	19	50
Objectives	The goal of this work package is to provide a powerful and integrated model of evolution in structured populations. It will combine existing analyses of parallel EAs with the Shifting Balance Theory (SBT) in population genetics. This novel perspective will allow us to study the speed of adaptation of EAs on combinatorial optimisation problems. The insights from these analyses will lead to recommendations on how to design effective parallel EAs.				
Description of work	<div>D 4.1 Wright’s shifting balance theory for parallel EAs<ul style="list-style-type: none">• Description: A quantitative theory of speed of adaptation in structured populations integrating the analysis of parallel EAs with Wright’s Shifting Balance Theory (SBT), enabling quantitative performance studies of parallel EAs with specific selection and migration models, restricted to complete topologies. See page 16 for details.• Primary site: IST Austria• Contributing sites: USFD, UNOTT</div> <div>D 4.2 Extension to topologies, selection and migration policies<ul style="list-style-type: none">• Description: An extension of the quantitative theory of speed of adaptation in structured populations (D 4.1) towards common selection mechanisms and migration policies used in EC. See page 18 for details.• Primary site: USFD• Contributing sites: IST Austria, UNI-JENA</div> <div>D 4.3 Application to parallel EAs<ul style="list-style-type: none">• Description: Quantitative performance analyses of parallel EAs in combinatorial optimisation, obtained through our quantitative theory (D 4.2). Recommendations on how to design parallel EAs with regard to choosing communication topologies, selection mechanisms, and migration policies. See page 19 for details.• Primary site: USFD• Contributing sites: IST Austria, UNI-JENA</div>				
Deliverables	<div>Month 24: D 4.1 Wright’s shifting balance theory for parallel EAs</div> <div>Month 30: D 4.2 Extension to topologies, selection and migration policies</div> <div>Month 36: D 4.3 Application to parallel EAs</div>				

Table 1.3d: Work package description (WP 5)

Work package number	5	Start date:			Month 1
Work package title	Coordination and management				
Activity type	Management of the consortium (MGT)				
Participant number	1	2	3	4	Total
Participant short name	UNOTT	IST Austria	UNI-JENA	USFD	
Person months	5	1	1	1	8
Objectives	This work packages covers the coordination of four technical work packages, overall management of the project and conflict resolution. Our management plan is detailed in Section 2.1. We will finalise and sign a consortium agreement based on the DESCAs template before the signature of the contract; see page 33 for details. The deliverables are annual scientific reports on the progress of the project.				
Description of work	<p>D 5.1 First year scientific report</p> <ul style="list-style-type: none">• Description: Scientific progress report on first year activities.• Primary site: UNOTT• Contributing sites: IST Austria, UNI-JENA, USFD <p>D 5.2 Second year scientific report</p> <ul style="list-style-type: none">• Description: Scientific progress report on second year activities.• Primary site: UNOTT• Contributing sites: IST Austria, UNI-JENA, USFD <p>D 5.3 Third year scientific report</p> <ul style="list-style-type: none">• Description: Scientific progress report on third year activities.• Primary site: UNOTT• Contributing sites: IST Austria, UNI-JENA, USFD				
Deliverables	<p>Month 12: D 5.1 First year scientific report</p> <p>Month 24: D 5.2 Second year scientific report</p> <p>Month 36: D 5.3 Third year scientific report</p>				

Table 1.3d: Work package description (WP 6)

Work package number	6	Start date:			Month 1
Work package title	Dissemination and exploitation				
Activity type	Management of the consortium (MGT)				
Participant number	1	2	3	4	Total
Participant short name	UNOTT	IST Austria	UNI-JENA	USFD	
Person months	3	1	2	1	7
Objectives	This work package will implement our strategy for dissemination and exploitation of our achievements, including the organisation of workshops and annual meetings. See Section 3.3 for details.				
Description of work	<p>D 6.1 Kick-off workshop</p> <ul style="list-style-type: none">• Description: The consortium board organises a kick-off workshop at IST Austria in spring 2014. The scientific advisory board and associated industrial partners are invited. All PIs will give tutorials on techniques from PG (e. g. diffusion approximation) and EC (e. g. drift analysis and black-box complexity). The expected outcome is a draft of the unified framework (deliverable D 1.1). See Section 3.3 for details.• Primary site: IST Austria• Contributing sites: UNOTT, UNI-JENA, USFD <p>D 6.2 Website and Wiki</p> <ul style="list-style-type: none">• Description: Setting up a web site and a Wiki with a public and a private part accessible only for the project partners. The public part of the web site informs about the projects goals, activities and progress and also serves as outreach to the community. See Section 3.3 for details.• Primary site: UNI-JENA• Contributing sites: IST Austria, UNOTT, UNI-JENA, USFD <p>D 6.3 Final workshop</p> <ul style="list-style-type: none">• Description: Organising a final workshop in summer 2016 in a distinct location like Dagstuhl. Researchers from both fields, the scientific advisory board, and the associated industrial partners are invited. The outcomes of the project are presented, follow-up research is planned, and further dissemination to industry is organised. See Section 3.3 for details.• Primary site: UNOTT• Contributing sites: IST Austria, UNI-JENA, USFD				
Deliverables	<p>Month 3: D 6.1 Kick-off workshop</p> <p>Month 6: D 6.2 Website and Wiki</p> <p>Month 34: D 6.3 Final workshop</p>				

Table 1.3e: Summary of effort

The person-months of work-package leaders are shown in bold for each WP.

Partic. no.	Partic. short name	WP 1	WP 2	WP 3	WP 4	WP 5	WP 6	Total person months
1	UNOTT	10	18	6	6	5	3	48
2	IST Austria	20	16	17	17	1	1	72
3	UNI-JENA	12	3	17	8	1	2	43
4	USFD	12	6	6	19	1	1	45
Total		54	43	46	50	8	7	208

2 Implementation

2.1 Management structure and procedures

SAGE is an ambitious and highly innovative research project spanning four sites in three European countries and two disciplines. As such, it requires a strong, coherent and efficient management structure to ensure the achievement of its aims.

Consortium board. The central authority within SAGE is the consortium board, comprising the coordinator and all site managers:

- *Per Kristian Lehre*, Coordinator and site manager UNOTT
- *Tiago Paixão*, Site manager IST Austria
- *Tobias Friedrich*, Site manager UNI-JENA
- *Dirk Sudholt*, Site manager USFD

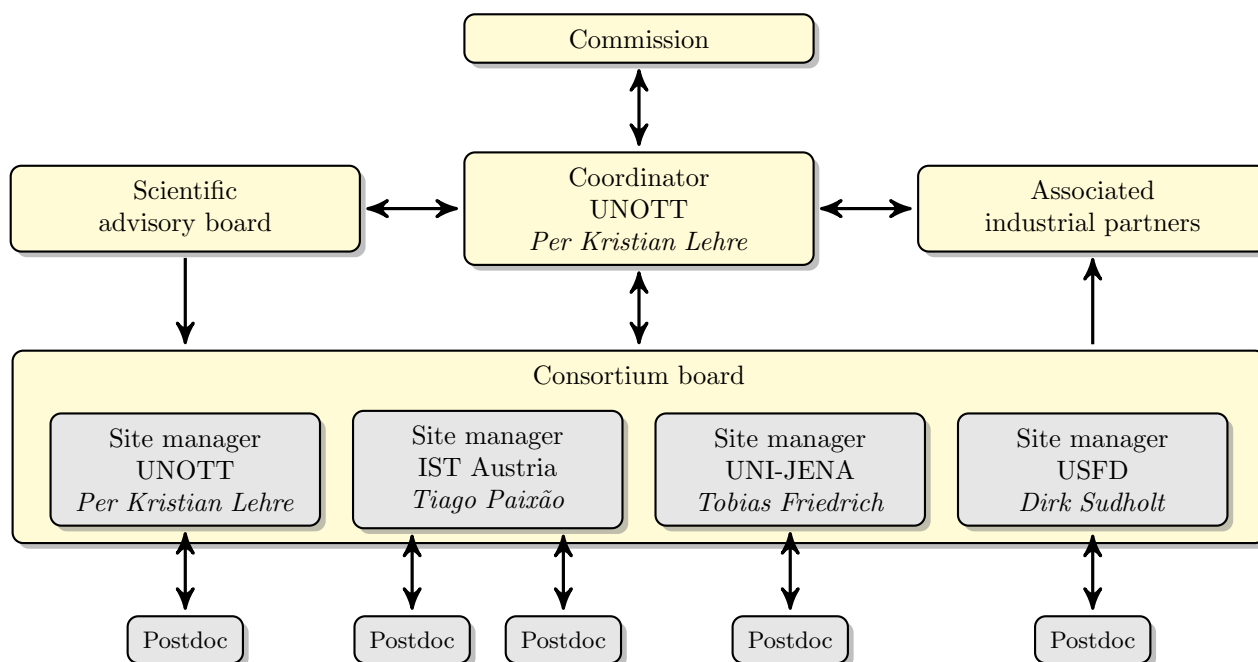
The board will continuously monitor the progress of the project, and decide about all matters that affect the course of the project so as to ensure that all activities are coordinated with one another. In particular, the consortium board makes final decisions should disagreements arise amongst the partners.

Coordinator. The consortium board is chaired by the coordinator. The coordinator will act as the intermediary with the European Commission, as defined in the Rules for Participation. The coordinator's responsibilities include

- managing and controlling the consortium as a whole,
- managing and chairing meetings of the consortium board,
- reporting to the European Commission authorities,
- preparing the consortium agreement regulating the internal organisation and management of the consortium prior to the signature of the grant agreement with the EC,
- approving the transmission of deliverables after consultation with the consortium board,
- managing and administrating finances after consultation with the EC and the consortium board,
- monitoring and keeping updated accounts,
- monitoring the preparation of cost statements and audit certificates,
- monitoring any ongoing ethical considerations of individual research activities,
- monitoring recruitment of women and ensuring equality of treatment and opportunity,
- acting as the focal point for all external and internal requests.

Site managers. The task of the site managers is to take care that each site produces its contributions to the objectives of the work packages they are involved in. They ensure that high quality deliverables are produced in a timely fashion and within the budget. The site managers are responsible for the operations of the project at the site they represent, and must carry out all activities decided upon by the consortium board. As shown in Table 1.3a, each work package is coordinated by one site manager. The site managers are responsible for the deliverables of the work package they are coordinating, and for the dissemination of results.

Management structure. The following diagram shows the management structure of SAGE.



Scientific advisory board. The consortium board is advised by the scientific advisory board, which consists of internationally leading experts and scientists from PG, EC and related academic communities.

The scientific advisory board is informed annually about the status of the project by the coordinator through scientific reports (deliverables D 5.1–5.3). It will constructively but critically review the work plan and the progress of each WP and the whole programme, providing independent review, advice and guidance. Many of its members are accomplished pioneers in their respective fields. Their expertise in founding new research fields and championing new technologies will provide an invaluable resource for managing an ambitious and highly innovative project like SAGE.

The following researchers enthusiastically agreed to serve on the scientific advisory board:

- *Lee Altenberg* is an associate professor at the University of Hawaii at Manoa and served as Associate Editor of Biosystems and on the editorial board of Genetic Programming and Evolvable Machines. He was also part of the founding board of directors of the International Society for Artificial Life. He is one of the few researchers that has been publishing in both evolutionary computation and population genetics.
- *Nick Barton* is full professor at IST Austria. He is one of the internationally leading evolutionary biologists and co-author of an influential textbook on evolution. He has developed a mathematical framework for the study of multilocus genetics which he has used to make ground-breaking contributions to the dynamics of hybrid zones, the evolution of sex, speciation, and towards understanding the limits to selection. He is a Fellow of the Royal Society and a Fellow of the Royal Society of Edinburgh. In 2008 he was awarded the Darwin-Wallace Medal, which is given every 50 years by the Linnean Society of London.
- *Thomas Bäck* is head of the Natural Computing Research Group at the Leiden Institute of Advanced Computer Science (LIACS). He is co-editor of the Handbook of Evolutionary Computation and editorial board member of several journals on evolutionary computation. He is also President and CEO of Divis Intelligent Solutions GmbH.
- *Chris Cannings* is professor emeritus in the School of Mathematics and Statistics at the University of Sheffield. His expertise covers the deterministic and stochastic modelling in evolutionary biology, population genetics, and stochastic processes. He is the editor-in-chief of *Applied Mathematics*.
- *Bernard Chazelle* is full professor of computer science and mathematics at Princeton University. He invented the concept of “Natural Algorithms” with a focus on collective behaviour and influence systems, which recently raised a lot of attention and gained the best paper award at the leading algorithms conference in 2009 [19].

- *Paulien Hogeweg* is a full professor at the Utrecht university, where she heads the Bioinformatics Group. She is known for having coined the term Bioinformatics to describe her research which focuses on the dynamics of information in biotic systems and the evolution of complexity. Her specific research areas focus on evolution of gene regulation, and multi-level selection.
- *Holger H. Hoos* is full professor in the computer science department of the University of British Columbia, Canada. He is an expert on solving hard combinatorial optimisation problems with stochastic search algorithms with a focus on automatic parameter tuning, for which he won several awards and competitions.
- *Zbigniew Michalewicz* is an internationally renowned optimisation expert who has published over 250 research articles and 25 books. He is the co-founder and chief scientist of SolveIT Software, which sells supply and demand optimisation software based on evolutionary algorithms and other meta-heuristics.
- *Frank Neumann* heads the Evolutionary Computation Group at the University of Adelaide. He is a leading figure in the theory of evolutionary computation and co-authored a recent book on combinatorial optimisation with evolutionary algorithms [84].
- *Una-May O'Reilly* is director of the Evolutionary Design and Optimisation Group at the Computer Science and Artificial Intelligence Laboratory (CSAIL) of the Massachusetts Institute of Technology (MIT). She serves on the executive board of the ACM Sig-EVO.
- *Jonathan E. Rowe* is Professor of Natural Computation at the University of Birmingham, where he currently heads the School of Computer Science. He has made many fundamental contributions to the theory of genetic algorithms, including one of the first books on the subject.
- *Hans-Paul Schwefel* is professor emeritus at the Dortmund University of Technology. He is the co-inventor of evolution strategies, one of the major branches in EC. Schwefel received a Honorary Doctor of Science from the University of Birmingham and the IEEE Frank Rosenblatt award “for pioneering contributions to evolutionary computation through the theory and application of evolution strategies” in 2011. He is also co-founder of one of the leading conferences on evolutionary computation.
- *Günter Wagner* is the Alison Richard Professor of Ecology and Evolutionary Biology at Yale University. His research interests revolve around the evolution of development, and how it constrains or potentiates morphological evolution. He has been awarded the MacArthur Prize in 1992 and the Humboldt Prize in 2005. He is also a corresponding Member of Austrian Academy of Sciences, a Fellow of the American Association for the Advancement of Science, and a Fellow of the American Academy of Arts and Sciences.
- *Alden Wright* is emeritus and research professor at the University of Montana. He is a pioneer of evolutionary computation and former chair of the conference series on foundations of genetic algorithms (FOGA). He is also one of the few researchers who have contributed to both EC and PG.
- *Xin Yao* is full professor and head of the Natural Computation group at the University of Birmingham. He is an IEEE Fellow, former editor-in-chief of *IEEE Transactions on Evolutionary Computation* and holder of the IEEE Computational Intelligence Society EC pioneer award.

Associated industrial partners. The associated industrial partners will be informed about new techniques developed in the project through annual scientific project reports released by the coordinator (deliverables D 5.1–5.3). They will be invited to the kick-off workshop (deliverable D 6.1) to pose questions relevant for their specific applications and will take part in the final workshop (deliverable D 6.3) in order to generate short-term industrial impact. We also expect some of our undergraduate and graduate students to spend time at these industrial partners. This allows fast dissemination of knowledge on newest evolutionary optimisation techniques. All associated industrial partners are very interested in the project and confirmed their support. Letters of support can be found in the appendix.

- *SolveIT* is a subsidiary of the French Fortune Global 500 company Schneider Electric, and specialises in supply & demand optimisation and predictive modelling.
- *divis intelligent solution GmbH* is a German company providing services in data analysis and optimisation for the automotive and other industries.

- *EventMAP* a company operating from Belfast, Northern Ireland, is a specialist in resource planning, such as exam scheduling, for the university sector.
- *Exergy Studios* is a company operating from Bratislava, Slovakia, which specialises in energy certification for buildings and provides design optimisation through integrated design and building simulation.

Communication. Much information exchange will be done electronically, but it is clear that difficult issues must be solved at meetings. The consortium board communicates monthly via Skype about all ongoing matters. Central to the electronic communication are also the project Wiki, Google Docs, and the Subversion revision control system (svn) which have been already used for preparing the proposal and which can be adapted also for external communication and dissemination. The external part of the Wiki should also collect translations of terminology, models, and results.

Meetings. The consortium board meets twice a year in person at one of the participating sites. At the beginning of the project we hold a kick-off workshop at IST Austria, and in the final year we organise a final workshop. The scientific advisory board will be invited to both these workshops. The consortium board also holds additional monthly Skype meetings when no physical meetings take place. A consortium board meeting will be deemed quorate if at least 3 out of 4 site managers are present. The aim of the meetings of the consortium board is to manage deliverables, dissemination, internal and external communications, finances, collaborations, and proactive services such as web site and press releases. In each meeting, the consortium board will monitor all current research activities by reviewing targets and deliverables, and promoting synergies between different activities. The board will identify technical and scientific problems and assess the risk of ongoing and upcoming tasks and formulate contingency plans (based on feedback from the scientific advisory board whenever appropriate).

Risk management. Such an interdisciplinary project naturally involves tasks that at current knowledge have to be considered risky. We are aware of potential scientific and organisational risks, as detailed in Section 1.3, page 20. The consortium board will regularly monitor all potential risks during monthly project meetings to identify potential problems proactively and to formulate contingency plans.

Community building. For the long-term success of the project it is important that both communities communicate with each other more effectively. We therefore plan tutorials or workshops at the leading conferences on evolutionary computation (GECCO, PPSN, IEEE CEC) and evolutionary biology (ESEB). The PhD candidates and postdocs will work as seed for more interaction between EC and PG. We additionally plan to have a program of open activities including a summer school and workshop intended mostly for academia but also open for industry at a distinct location like Cumberland Lodge, Schloss Dagstuhl, MF Oberwolfach, or Lorentz-Center Leiden.

Overcoming cultural differences. A major challenge in such an interdisciplinary project is that the two communities, EC and PG, have different standards, perspectives, and expectations regarding scientific work. For example, in EC conferences provide a major venue for publications, and submitted papers are expected to be original, novel, and rigorous. In PG conferences are informal meetings with much more lax standards, implying that conference papers are not regarded as highly as in EC. These cultural differences need to be addressed in order to reach out from one community to the other.

The first step towards overcoming these differences is to acknowledge these differences and making them explicit. The PIs have a high sense of awareness in this respect, based on extensive discussions during the preparation of this proposal. All PIs are experienced in interdisciplinary work, having worked at the interface of at least two fields (IST Austria: physics and biology, UNI-JENA: discrete mathematics, computational complexity, and evolutionary computation, UNOTT: operations research, heart research, SEBASE: theoretical computer science vs. software engineering, and evolutionary computation, USFD: algorithmic analysis and evolutionary computation, involvement in collaborative research center). We will use the first project meeting to feed on these results to all members of the project, so that all members become aware of these cultural differences.

We anticipate that many publications will be coauthored by members of both communities, in which case coauthors belonging to the targeted audience will ensure that publications are addressing their cultural expectations. In rare cases where no member of the targeted audience is coauthoring a paper, the authors will actively seek advice from a PI of the other community. In all cases, whenever appropriate we will further seek the advisory board's advice on strategic questions for overcoming cultural differences.

Dispute resolution. The following procedure will govern the resolution of conflicts between the different sites. A conflict is defined as a severe disagreement between at least two sites about a technical or managerial aspect of either a particular work-task or about the project as a whole. The term conflict includes identified areas of possible future conflict, which may be avoidable.

A possible conflict may be raised by any site; it requires that a written statement signed by the site manager of the site involved be sent to the coordinator. This written statement should identify the problem. The coordinator will decide on the appropriate action to be taken, where he considers the following options: (i) he tries to resolve the conflict, and/or (ii) he puts the conflict on the agenda for the next meeting of the consortium board, or (iii) he calls a special meeting of the consortium board to consider and resolves the conflict. Information about the conflict which is presented at the meeting of the consortium board must be circulated to all members of the consortium board at least one week before the meeting. The conflict will be debated by the consortium board according to the following rules. The two or more alternatives will be presented and clearly stated by one of its supporters. Sufficient time for discussion must follow the presentation of each alternative at the meeting. Following this discussion, the board will take a decision. Each member of the board has a vote. A simple majority suffices for all decisions. Ties are broken by the coordinator. The coordinator is responsible for ensuring that appropriate notification of the decision is given to each partner.

Consortium agreement. We will finalise and sign a consortium agreement before the signature of the contract. We use the template of the DESCAs model consortium agreement, which addresses the following points:

- Rules and regulations concerning access rights of the partners to pre-existing knowhow, and to all kind of intellectual property, generated during the contract and not existing before the start of the project;
- Arrangements for the distribution of the EU-contribution among participants and among activities;
- A description of the internal organisation of the consortium, including conflict resolution and decision making, in line with this section of the proposal.

2.2 Individual participants

The University of Nottingham (short name: UNOTT)

Organisation. The University of Nottingham is a founding member of the Russell group of leading research-intensive universities in the UK. The School of Computer Science was ranked 8th from 81 returns in Computer Science in the UK in the most recent Research Assessment Exercise rankings, with 80% of all research given the highest rating of “internationally excellent”. The Automated Scheduling, Optimisation and Planning (ASAP) Research Group is located in the School of Computer Science and carries out research into models, heuristics and algorithms for automatically producing high-quality solutions to a variety of real-world applications and optimisation problems. The group comprises 8 members of academic staff, 9 research assistants, 34 PhD students and 1 administrative staff. The ASAP research group alone has been holding funds from a variety of national and international sources, totalling up to approximately £ 11.5M since 2009.

Key person. *Per Kristian Lehre* obtained MSc and PhD degrees in Computer Science from the Norwegian University of Science and Technology (NTNU) in 2003 and 2006, respectively. He has significant international experience, having held research positions in Norway (NTNU), Denmark (Technical University of Denmark), the UK (University of Birmingham, and University of Nottingham), and several longer research visits (e.g., TU Dortmund and Max-Planck-Institut Saarbrücken).

He has worked for a decade on theoretical aspects of evolutionary computation and has made fundamental contributions, leading to four best paper awards in major conferences. He is regularly invited to give talks about evolutionary computation, such as in international summer schools in the UK, France, and Estonia. He is also giving tutorials in major conferences and workshops on theoretical aspects of evolutionary computation, such as about drift analysis (GECCO, CEC, and ThRaSH), and runtime analysis of EAs (GECCO). He serves on the editorial board of the journal *Evolutionary Computation*, and guest edited the special issue of *Theoretical Computer Science* on “Theoretical Foundations of Evolutionary Computation”. He will also guest edit a forthcoming special issue of *IEEE Transactions on Evolutionary Computation* on theory of evolutionary computation. Lehre is also vice-chair of the IEEE Task Force on Theoretical Foundations of Bio-inspired Computation.

Since September 2011, he is a lecturer on operational research and computer science in the School of Computer Science at the University of Nottingham. He supervises two PhD students and one postdoctoral researcher.

Expertise and role in project. His PhD thesis combined concepts from theoretical biology and theoretical computer science, notably in analysing the structure of neutral networks in complex genotype-phenotype mappings, and how this structure influences the runtime of evolutionary algorithms [67].

He has made many fundamental contributions to the theory of evolutionary computation. In particular, he introduced the unbiased black-box model [71] which since then has made black-box complexity a very active research area in theory of evolutionary computation. His expertise in this area will be important for WP 1. He also pioneered runtime analysis of evolutionary algorithms with large populations [68, 69, 73].

He has extensive previous experience with interdisciplinary research, including bio-medical research [10], software engineering, theoretical computer science, and operational research.

UNOTT is the main coordinator of the project, as well as leader of WP 2. Lehre’s membership on the management team of the large EPSRC-funded SEBASE project has given him extensive experience in managing large, multi-institutional, and interdisciplinary research projects.

Institute of Science and Technology Austria (short name: IST Austria)

Organisation. The Institute of Science and Technology Austria (IST Austria) is a new research institute and graduate school in the natural and mathematical sciences. IST Austria was established in 2006, the campus opened for theoretical scientists in 2009, and the first laboratory building for experimental scientists opened in 2010. In June 2013, the Institute will have 28 research groups, which represent evolutionary, cell, plant, and neurobiology, computer science, mathematics, and physics. The Institute has been given two consecutive 10-year budgets which will allow it to grow to 90–100 research groups by 2026. The structures at the Institute have been set up to allow IST Austria to compete with the top institutions in the world for professors, postdocs, and PhD students. The culture at the Institute emphasises inter- and cross-disciplinary science as well as pure, curiosity-driven research. In support of these goals, IST Austria aims to put up as few internal boundaries and hierarchies as possible. There are no departments and there is a single, multi-disciplinary PhD program. Office and laboratory space, scientific services, and institutional research funding are set up to encourage the interaction between research groups and to ensure the most efficient use of resources. The Austrian Federal Law which established IST Austria specifies that the scientific board of the institute shall ensure that the quality of its science is world-class. The scientific board set up in-depth area evaluations by international experts in order to assess the scientific quality of a research area and its professors, and to give recommendations for the future development of the area.

Key person. *Tiago Paixão* obtained his degree in Physics Engineering from Instituto Superior Tecnico, in Lisbon in 2001. After spending one year as a research assistant at the same school, working at the non-linear dynamics group, he then started his PhD at in the area of stochastic gene expression at the Theoretical Immunology group at Instituto Gulbenkian de Ciencia. Here, he developed collaborations with groups working on experimental biology, namely on the expression dynamics of the cytokine IL-10, developing both statistical analysis and mechanistic models of gene expression dynamics. He was awarded his PhD degree in Biomedical Sciences in 2007, by the Universidade do Porto, which was awarded 3rd place for the Reinhart Heinrich Doctoral Thesis Award for best European Doctoral Thesis in theoretical biology. He then moved to Houston, Texas, for his postdoctoral work in Ricardo Azevedo's Lab, where he worked on many topics in evolutionary theory, including the theory of sex ratio, dynamics in neutral networks, and the evolution of gene networks. Here, he collaborated with distinct groups such as the Zufall group, an experimental group working on *Tetrahymena Thermophila*, where he developed models to study the dynamics of sex ratio evolution in this organism, and Tim Cooper's lab, another experimental group working on the dynamics of evolution, where he examined the influence of spatial structure on the evolution of Toxin-Antitoxin systems, and developed new methods to inspect the repeatability of evolutionary trajectories on experimentally determined fitness landscapes. In 2011, he joined Nick Barton's group at IST Austria, where he has been working on the consequences of epistasis for evolution, using a quantitative genetics framework. He has been an invited reviewer for *PLoS Computational Biology*, *PLoS Biology*, *Evolution*, *Genetics*, *Proceedings of the Royal Society B* and *Journal of Theoretical Biology*.

Expertise and role in project. Tiago Paixão has a background in physics, granting him extensive experience in continuous analytical methods which are crucial for many of the techniques used in PG. Since then, he has been working on theoretical biology, gaining substantial expertise in many topics in this field. He made contributions to many topics, including the dynamics of stochastic gene expression, the role of recombination on population dynamics in neutral networks, and the evolution of under frequency-dependent selection. His skills will be required throughout the project, contributing to all work packages, and coordinating the efforts on the PG field.

Friedrich-Schiller-Universität Jena (short name: UNI-JENA)

Organisation. Friedrich-Schiller-Universität Jena, founded in 1558, is one of the oldest universities in Europe, being a member in the COIMBRA group, a network of prestigious, traditional European universities. It is organized into ten faculties with overall 20 000 students and about 350 professors. UNI-JENA has experience in managing more than 70 European research contracts. The faculty of mathematics and computer science employs 30 full time faculty members. Members of the department of computer science currently coordinate two other FP7 FET-projects: “Artificial wet neuronal networks from compartmentalised excitable chemical media” (Peter Dittrich) and “Computational geometric learning” (Joachim Giesen).

The Chair of Theoretical Computer Science is the oldest chair at the department of computer science. It currently consists of Tobias Friedrich (full professor), Jörg Vogel (senior researcher), Timo Kötzing (postdoc) and Anton Krohmer (PhD student). The main focus of the group is the analysis of random structures and algorithms with a strong emphasise on bio-inspired algorithms and their applications. More information is found at <http://www.theinf.uni-jena.de>.

Key person. *Tobias Friedrich* graduated from the University of Sheffield, UK, with a MSc degree in Computer Science in 2003 and a Diploma degree in mathematics in 2005 from the University of Jena, Germany. In 2007 he obtained his PhD in computer science from Saarland University, Germany. His thesis ‘Use and Avoidance of Randomness’ was supervised by Prof. Kurt Mehlhorn. After staying at ICSI Berkeley, Max-Planck-Institute for Computer Science Saarbrücken, and the Cluster of Excellence on Multimodal Computing and Interaction, Germany, in 2012 he joined the department of computer science at UNI-JENA as full professor and chair of theoretical computer science.

Friedrich co-chaired the theory track of the leading conference in evolutionary computation (GECCO) in 2012 and 2013 and was member of the program committee of the leading conference in random structures and algorithms (RANDOM) in 2012 and the leading conference in experimental algorithmics (SEA) in 2013. In the past six years he was nominated five times for a best paper award at GECCO, which he won three times. He also gave invited tutorials or talks at almost all relevant meetings on evolutionary computation (GECCO, PPSN, CEC, ThRaSH).

Expertise and role in project. Friedrich’s background is discrete mathematics and theoretical computer science, where he made fundamental contributions to the theory of (deterministic) random walks (e.g. [24]) and to distributed computing (e.g. [27, 41]). Within EC, he worked on properties of bio-inspired search heuristics, e.g. on neutrality [40] and diversity [42]. He also extensively studied the complexity of evolutionary multi-objective optimisation (e.g. [14, 15]). His unique experience within the consortium, which is required in WP 4, is his substantial background in distributed computation (load-balancing, information dissemination, rumour spreading).

The University of Sheffield (short name: USFD)

Organisation. The University of Sheffield is one of the original red brick universities and a member of the Russell Group of leading research-intensive universities in the UK. It has nearly 25 000 students from 128 countries and over 5 500 staff members. The University has been named UK University of the Year in the 2011 Times Higher Education Awards. It was ranked 72nd among universities worldwide in the 2011 QS World University. The Department of Computer Science has an international reputation for the quality of its research and teaching. It hosts a number of research groups that will support and benefit from the SAGE project, most notably the Behavioural and Evolutionary Theory Lab (headed by Dr. James Marshall), investigating emergent behaviours in natural evolution, and the Verification and Testing group, applying evolutionary algorithms to optimisation problems in software engineering (Prof. Mike Holcombe, Dr. Phil McMinn, Dr. Gordon Fraser).

The department also has an outstanding record for coordinating collaborative research, particularly in European programmes. Since 2001 it participated in over 40 EC-funded consortia, coordinating 10 of them.

Key person. *Dirk Sudholt* obtained his Diploma degree in 2004 and his PhD in Computer Science in 2008, both from the Technische Universität Dortmund, Germany. During this time he was working at a large interdisciplinary collaborative research centre, coordinating 40 staff as executive officer in 2008/09. He has held postdoc positions at the International Computer Science Institute in Berkeley, California, working with Prof. Richard M. Karp and at the University of Birmingham, UK, working with Prof. Xin Yao. Since January 2012 he is a Lecturer at the University of Sheffield. Sudholt is currently supervising two PhD students working on the theory of parallel evolutionary algorithms: Joseph Kempka and Andrea Mambrini (joint supervision with the University of Birmingham). He is a regular member of program committees for all major conferences in EC (GECCO, PPSN, CEC, FOGA), a member of the editorial board of the Evolutionary Computation journal, and he has given invited tutorials at GECCO and ThRaSH. His work has received 6 best paper awards at the two leading conferences in the field, GECCO and PPSN.

Expertise and role in project. Sudholt has a background in theoretical computer science, specifically algorithmic analysis and computational complexity. He has made many fundamental contributions to the theory of nature-inspired search heuristics like evolutionary algorithms, ant colony optimisation, particle swarm optimisation, and hybridisations with local search.

Most relevant to this proposal is his work on the usefulness of recombination in evolutionary algorithms [61, 112], which includes the first non-constructed problem where recombination speeds up optimisation exponentially [109]. He further pioneered the runtime analysis of parallel EAs with structured populations, in the context of pseudo-Boolean optimisation [62, 63, 65] as well as combinatorial optimisation [66, 80].

USFD will lead WP 4 on parallel evolution in structured populations, building on Sudholt's expertise regarding parallel EAs. USFD will also contribute to WP 1 since developing a unified foundation is key for SAGE. Further contributions include the study of peak shifts (D 2.1) as this is highly relevant for WP 4, and analyses of sexual reproduction (D 3.1 and D 3.3), contributing expertise on the analysis of EAs with recombination.

2.3 Consortium as a whole

SAGE is an ambitious and highly innovative interdisciplinary project, designed to bring together two different research fields. However, it was also recognised that most of the societal impacts that SAGE will bring forth are enabled by the contributions it will make to the design of new and improved optimisation techniques. As such, it is vital to maintain a good balance across these two fields and to ensure at the same time that all required skills are represented.

Our consortium was designed to provide high degrees of complementarity as well as coherence. Three sites have a background on the theory of evolutionary algorithms (UNOTT, UNI-JENA, USFD), and all site managers have a full command of common analytical techniques used in EC. This provides excellent common ground to tackle the highly ambitious task of synergising with PG.

Moreover, each of these sites has specialised in areas that will be key for this project: UNOTT has extensive experience with studying the speed of adaptation of EAs in different fitness landscapes, particularly neutrality and dynamic optimisation, making it the perfect leader for WP 2. UNI-JENA has a strong background in the analysis of random processes with stochastic dependencies, which is highly needed in WP 3 for studying recombination. USFD has pioneered the runtime analysis of parallel EAs, so naturally USFD will lead WP 4.

Each of the above sites will fund one postdoctoral researcher who will participate in the SAGE project. This ensures full exploitation of our results with regard to developing novel and improved designs for evolutionary algorithms.

IST Austria represents the ideal member from population genetics as it hosts a world-leading group in PG and has strong ties to other world leading researchers in theoretical PG (Reinhardt Bürger, Joachim Hermisson, Karl Sigmund), all based in Vienna. Furthermore, IST Austria hosts a major community in theoretical computer science. This is the ideal environment where to develop such a project, since all extra expertise will be at hand in a thriving scientific environment. IST Austria will fund two postdoctoral researchers in SAGE in order to obtain an ideal balance needed to achieve the objectives of SAGE: it gives IST Austria enough man-power to ensure a strong contribution from PG in all work packages to develop synergies and achieve the objectives of SAGE. Having two postdocs at the same institution also facilitates communication and coordination, providing coherence between the project members from PG.

In order to provide further detail, we list required areas of expertise for each deliverable. These areas cover analysis techniques in EC and PG and important topics relevant for one area or jointly for both. In the table we indicate strong experience (++) and experience (+) with respect to different areas of expertise.

Expertise	WP 1				WP 2			WP 3			WP 4		
	1.1	1.2	1.3	1.4	2.1	2.2	2.3	3.1	3.2	3.3	4.1	4.2	4.3
EC techniques													
• Runtime analysis of EAs	+	++	++	+	+	++	++	++	+	++	+	++	+
• Drift analysis		+	++		++		++	+			+		+
• Artificial fitness levels		+	++		+		+	+					+
• Black-box complexity		+		++	+	+		+			+		+
PG techniques													
• Math. population genetics	++	++	++		++	++	++	++	++	+	++	+	+
• Theory of dynamical systems	++	++	++										
• Diffusion approximation		+	++		++		+	+			+		+
Joint topics													
• Population diversity		++	++		+	++	++	++	+	++	++	++	+
• Neutrality		+				++	+	++	++	+			
• Sexual reproduction		++						++	++	++			
• Parallelism in evolution		++		+							++	++	++

The following table describes the expertise of the participating sites, thus indicating that on the one hand there is a good complementarity between the partners, while on the other the partners form a coherent consortium.

Expertise	UNOTT	IST Austria	UNI-JENA	USFD
EC techniques				
• Runtime analysis of EAs	++		++	++
• Drift analysis	++		+	+
• Artificial fitness levels	+		+	++
• Black-box complexity	++			+
PG techniques				
• Mathematical population genetics		++		
• Theory of dynamical systems		++		
• Diffusion approximation		++		
Joint topics				
• Population diversity	+	++	+	+
• Neutrality	++	++	++	
• Sexual Reproduction	+	++	++	++
• Parallelism in evolution		+	+	++

2.4 Resources to be committed

Since SAGE is a foundational project creating a theoretical framework, its main costs are due to salaries of the research team. We have budgeted for five postdoctoral researchers, one for each of USFD, UNOTT, and UNI-JENA and two for IST Austria, reflecting the balance proposed in Section 2.3. Due to the ambitious nature of the project, these postdocs will be committed for 3 years.

In evolutionary computation, the publishing venues that have a broader base and impact are conferences. However, in population genetics the typical venues are journals, which incur in publication costs and page charges. We estimate that the project will release 6 publications in population genetics, and we budgeted accordingly at the typical rates for high-profile journals in this field.

The balance in terms of person-months between RTD (WP 1–4, 92.8%) and non-RTD activities – management and dissemination – (WP 5–6, 7.2%) will ensure that the project is adequately managed. The allocation of 3.4% of the person months to dissemination (WP 6) activities will secure an adequate dissemination of the project.

From the point of view of main costs chapters, the initial estimated breakdown is the following:

Type of costs	Euros	%
Direct	1 282k €	63%
Indirect	766k €	37%
Total	2 048k €	100%

Table 2.4a: Costs Budget Breakdown by type of costs.

From the point of view of main Direct Costs chapters, the initial estimated breakdown is the following:

Type of activities	Euros	%
Personnel	1 126k €	88%
Travel & subsistence	144k €	11%
Other costs	12k €	1%
Total	1 282k €	100%

Table 2.4b: Costs Budget Breakdown by main direct costs chapters

The main assumptions adopted for estimating cost budget are as follows:

Direct costs

- **Personnel costs:** Personnel costs have been calculated on the basis of person months agreed to be allocated to the Project by the partners and the corresponding average person-months rates of each of the partners.
- **Other costs:** Since the Researchers will undertake bilateral visits and conference / workshop attendance for dissemination activities, funds for one shared project high-powered laptop is requested per site, excluding IST Austria. Publication costs have also been included in this heading based on a typical cost for an open access journal and we have assumed six publications.
- **Travel and subsistence:** As a project aiming at bringing together two scientific fields, we anticipate that substantial physical interaction will be required. To this end we have budgeted travel costs allowing for bilateral visits and interchange of PhD students. These costs cover all participants travel and daily allowance and also factor in costs for Consortium-wide meetings and working sessions, as well as participation in several workshops as mentioned in Section 2. We have also included a budget for 2 international conferences per year. It is worth noting that this section has been reviewed and minimised as much as possible.
- **Subcontracting:** Within this chapter only the costs of Audit Certificates on the financial statements have been included. For IST Austria, who receive an European Commission contribution of more than 375k €, one certificate has been calculated.

Indirect costs

- **Overheads:** Total Overheads costs arise from adding overheads costs of each of the partners, calculated according to their respective costs models.

Project costs & co-financing

In order to contribute to finance the Project costs budget, the grant to the budget from the European Commission has been estimated, using the established rules on this matter, as 2.05M € for the full Project costs. Breakdown of the grant by type of activities is as follows:

Type of activities	Budget		EU Contribution	
RTD	1 879k €	92%	1 409k €	89%
Management	169k €	8%	169k €	11%
Total	2 048k €	100%	1 578k €	100%

Table 2.4c: Project Budget and European Commission grant to the budget

While the postdoctoral researchers funded by the project will develop the new models and methods described in the work programme, all participating research groups additionally have PhD students (paid from base funding or stipends) who will work on related topics and use the tools developed in the project. This will further increase the impact of the project and help having a long-lasting interaction between population genetics and evolutionary computation. The unfunded Ph.D. students will also take part in the project workshops.

3 Impact

3.1 Transformational impact on science, technology and/or society

SAGE is designed to provide a breakthrough in our understanding of evolutionary processes through combining the complementary approaches from population genetics and evolutionary computation. The primary **transformational impact** of the project lies in breaking down the barriers between population genetics and evolutionary computation, in order to enable and promote synergies between the two fields and to seed a new research field on the quantitative analysis and design of evolutionary processes.

Our work programme is laid out to create a unified quantitative theory of speed of adaptation applicable to both biological and artificial evolution (WP 1). This theory will reveal how the efficiency of evolutionary processes depends on parameters of evolution as well as environmental parameters in terms of the fitness landscape (WP 2), and uncover the effect of recombination (WP 3) and parallel evolution in structured populations (WP 4).

Creating such a unified theory is an ambitious goal that **requires a concerted effort on a European level**, as no single country in Europe possesses the necessary skills and expertise to synergise population genetics and evolutionary computation. Our interdisciplinary consortium has gathered a unique team of excellent young researchers from all across Europe that is both complementary and coherent. All project members have the ambition, drive, skills, expertise, and commitment to make SAGE a success.

Successfully achieving the objectives of SAGE will lead to a breakthrough in the scientific understanding of evolution, transforming the way evolutionary processes are understood in nature, and how evolutionary processes are exploited in many areas of science. On a longer term, these developments will lay the foundation of a **new research field** unifying evolutionary computation and population genetics.

SAGE is performing **basic research** due to its foundational nature, hence the impact of SAGE is largely **independent from external factors**. The expected impact of basic research comes from *new insights into the basics mechanisms of nature and society* [1]. Such insight is a prerequisite for making innovative advances in applied research that can further lead to product research and development.

SAGE will not only provide immediate and profound scientific impact leading to a new research field, but it has also – due to the tremendously wide-spread desire in the industry to exploit the optimisation capability of evolutionary processes – a spectacular opportunity for tremendous medium- to long-term impact on technology and society.

Transformational impact on science

The European Commission acknowledges that an advanced industrial society needs a strong basic research program as *basic research is necessary for the active participation in the international scientific community to give and take part in the exchange of knowledge that leads to new insights and inventions*. [1].

New insights will be generated as a direct result from achieving the objectives of our work programme:

- Systematically synergising models and methods from PG and EC, thus **enabling long-term quantitative predictions** on the speed of adaptation of realistic populations on realistic fitness landscapes with many loci (Objective O1).
- Identifying **key features of fitness landscapes** and to quantify their effect on the speed of adaptation in artificial and biological evolution with asexual reproduction (Objective O2).
- Developing a quantitative theory of **sexual evolution**, allowing to quantify the **benefit of using recombination** in artificial and biological evolution, and how recombination can be **used most effectively** in EAs (Objective O3).
- Developing a quantitative theory of **parallel evolution in structured populations**, which is instrumental in using **parallel implementations of EAs** in many-core architectures most effectively (Objective O4).

Sharing this knowledge with academic beneficiaries will have immediate impact in strengthening the health of related scientific disciplines.

Theoretical biology will clearly benefit from the new set of tools that SAGE will bring, allowing it to tackle old questions in new ways and expand its reach to new problems. The flood of data in many areas of biology raises the exciting possibility of observing evolutionary dynamics in real time, as it has the potential to provide insight into many fundamental questions. For example, how frequent are the different modes of

selection, how relevant is genetic drift or what are the prevalent patterns of epistasis, are just a few of the questions that these data promise to uncover. However, in order to make sense of these data we require better models for the dynamics of evolution. The combined efforts of WPs 1 and 2 will be instrumental in bringing forth these new tools. A more general theory of evolution will also allow researchers to deploy a new set of tools to extend the reach of evolutionary theory to more societal problems, like cultural evolution and the spread and origin of technological innovation.

Natural computation and computational intelligence are umbrella terms for using natural processes in technical systems. SAGE will foster the strong presence of evolutionary algorithms in these areas. Specifically, researchers on the theory of evolutionary algorithms will benefit from an increased visibility of runtime analysis as one of the major cornerstones of SAGE, new foundational methods (deliverables D 1.3 and D 1.4), and exposing highly relevant research questions addressed in work packages WP 2, 3, and 4.

The scope of **operational research** (OR) is on solving real-world problems in numerous disciplines in science, engineering, and medicine. Evolutionary algorithms are popular tools for solving these problems as many of these problems cannot be solved by conventional exact techniques. Any progress in making EAs more efficient will thus have an immediate impact on researchers in OR. Furthermore, the need for developing new theory has been expressed strongly in OR [36].

SAGE will benefit the **theory of computation** by providing a computational view on evolution in the context of speed of adaptation. The algorithmic-mathematical analysis methods developed and refined in this project, particularly through deliverable D 1.3 and stochastic communication processes (deliverable D 4.2), may also find applications for studying the efficiency (computational complexity) of randomised algorithms and probabilistic processes such as information propagation in networks [27], queuing theory, and combinatorial optimisation.

Synthetic biology is an emerging discipline that will reach maturity in the near future [12]. It aims at engineering biological agents for various tasks, including beneficial biological compounds or degrading pollutants from the environment (bioremediation). However, one of its challenges is how to maintain the engineered circuits in the face of natural selection [104]. Without selection pressure to maintain these engineered circuits, the cost imposed by the burden of having to express foreign genes will inevitably lead to their degradation by mutation. The methods obtained through deliverables D 2.1 and D 2.2 will be a stepping stone towards rational biological circuit design, and to give predictive estimates on the lifetime of these circuits. Moreover, in situations where these engineered biological agents are released in the wild (for example, when trying to clear a lake of some pollutant), our results are bound to be useful in providing estimates to the rate at which they will adapt to these new environments.

Improving the efficiency of evolutionary algorithms also has potential impact in **data analysis in biological research** where evolutionary algorithms have been used as part of the research methodology en route to new scientific discoveries, such as in data analysis and computer assisted modelling. They are also used in classification of high-throughput data from the biological sciences, like data from micro-arrays or large numbers of full genome sequences [114]. A further application area is the estimation of phylogenetic trees from genomic data [13, 74, 76].

The consortium board has an extensive network of scientific collaborators in all mentioned areas and will seek opportunities for dissemination and developing synergies.

Technological and economical impact

The SAGE Consortium foresees immediate impact in many scientific disciplines. However, in order to develop its full potential in transforming the current state-of-the-art and enabling future and emerging technologies, it is paramount to also develop a long-lasting technological and economical impact.

The European Commission makes the point that *deep understanding of the fundamental principles of a domain is a prerequisite for radical proactive innovations and for the improvements of existing products and services* [1].

There is wide-spread recognition that evolutionary computation is currently suffering from a significant **lack of understanding**. This has led to significant impediments for industrial applications. A major obstacle in applying evolutionary algorithms in the industry is that their efficiency crucially, and often unpredictably, depends on design choices like the encoding of solutions, the design of recombination and mutation operators, or parameters such as the population size, the strength of mutation or the amount of recombination. The impact of these parameters is not well understood, and as a result, designing and parameter tuning for EAs is costly and time-consuming; leading experts estimate that around **10%–25% of development time is spent on parameter tuning** [47].

In many cases the resulting EAs are far from exploiting their full potential as efficient optimisers, and potential users of EAs are being deterred. A recent survey among users of evolutionary algorithms, found that **“poorly understood”** was the primary (39.7% of respondents) obstacle for acceptance of evolutionary computation in the industry [48].

There is therefore a demonstrable and urgent need for foundational research, and this is addressed head on by SAGE in its goal of creating a quantitative theory for the speed of adaptation. This theory will contribute to remedy the lack of understanding and to unfold medium- and long-term impact for a wide range of potential applicants. The SAGE consortium has received interest from a number of industries and has established associated industrial partnerships with selected companies that are keen to exploit fundamental advances in their applications (see support letters in the appendix).

While basic research often leads to industrial and societal impact in unforeseen ways, we have identified **three pathways to medium- and long-term industrial impact:**

1. facilitate a wide-spread adoption of EC techniques in the industry by alleviating the main obstacle which is “poor understanding” [48] and at the same time reduce development costs for parameter tuning
2. a more effective exploitation of EAs in industries which have already adopted EC, by improving the efficiency of EAs through optimal parameter settings, recommendations on the design of effective EAs, and through exploitation of parallel hardware
3. entirely new and innovative applications of evolutionary processes enabled by the solid, theoretical foundation of evolution provided by SAGE

Facilitating uptake and usage of evolutionary algorithms

The *first pathway* to impact consists in eliminating the main barrier to industrial uptake, which has been “poor understanding” [48]. Many fundamental issues related to evolutionary algorithms are poorly understood, such as for which problem types evolutionary techniques are effective, which EA is most appropriate for a given problem, and the amount of computational resources the EA will consume to find solutions of satisfactory quality. As a consequence of this poor understanding, the return on investment by using evolutionary computation cannot easily be predicted, and as a result many industries, especially SMEs, are reluctant to adopt EAs.

Most of our deliverables in the work programme contribute directly to an improved understanding of evolutionary algorithms. For example, the black-box complexity results from deliverable D 1.4 will provide much needed insights into the inherent potential and limitations of evolutionary algorithms with regards to their speed of adaptation. This insight is crucial when deciding whether evolutionary approaches can be effective for a given optimisation problem.

The study of evolution in various fitness landscapes (WP 2) will shape an understanding of what characteristics of optimisation problems contribute to their hardness. This is instrumental for enabling users to customise evolutionary algorithms to the optimisation problem in hand, if some knowledge on the problem’s characteristics is available. The results from WP 3 and WP 4 will explain how design choices, such as the use of recombination operators or spatially structured populations, influence the speed of adaptation. The overall theme of understanding speed of adaptation will allow predictions of the amount of computational resources EAs consume to find solutions of satisfactory quality. The impact from SAGE will be that the evolutionary techniques become more predictable and less risky, thus promoting stronger industrial uptake.

More effective evolutionary algorithms

The *second pathway* to industrial impact consists in the improvements to the parameter-settings of EAs, and the design of improved EAs that a stronger theory allows (deliverables D 3.3 and D 4.3 in particular).

The effectiveness of EAs can be improved in several ways:

- improving adaptation of populations to generate **better quality solutions** in applications of EAs,
- speeding up adaptation, hence generating high-quality solutions in **less time**,
- addressing the emerging opportunities and challenges posed by designing effective parallel EA variants on **massively parallel architectures**,
- improving the scalability of EAs: how their running time depends on the problem size, thus enabling to **solve larger problems** with limited or costly computational resources.

The whole work programme strongly supports this direction, particularly through recommendations on optimal parameter settings in EAs (deliverable D 3.3) and recommendations on how to design EAs with structured populations for many-core parallel architectures (deliverable D 4.3).

These points bring about tremendous potential for economic impact as in many applications better solution quality translates into better products and services and/or cost savings for manufacturing and operation. An increased efficiency of EAs may translate to immediate impact on existing industrial applications of the algorithms. Scalability through decreased running times and/or parallelization is a key point as many industries such as telecommunication, medical applications, and computational biology are faced with larger and larger problems that must be solved with existing computational resources.

In order to convey the enormous scale of potential cost-savings and the resulting economic impact, we review selected major industries in the EU economy, with regard to how they employ evolutionary algorithms as a key technology.

The EU is the world's largest manufacturer of motor vehicles. As such, the **automotive industry** is central for European productivity. The German company BMW uses evolutionary algorithms to optimise car bodies with respect to crash, noise, vibration, and harshness [34]. In the US, EAs are used to plan the distribution of used cars to automobile auctions [81]. Our industrial partner divis intelligent solution GmbH has strong ties with the automotive industries.

The **pharmaceutical industry** is another major industry in Europe. In 2011, the production from this industry in Europe was 205 000M €, and it employed 660 000 people [37]. This industry has the highest ratio of R&D investment to net sales (EUROSTAT). Due to the high cost of developing new drugs, (the R&D cost to develop a new drug is estimated to 1 059M €), it is essential to use effective computer software. Evolutionary algorithms are important in computational drug design. A central task in drug discovery is molecular docking, which is the modelling and prediction of the orientation in which two molecules form a stable complex. This task is carried out using so-called docking tools. The premier docking tool, AutoDock [105]², is based on an evolutionary algorithm, and so are other popular docking tools, such as GOLD³. Improvements to the EAs used in these tools will have substantial benefits.

Logistics involve a wide range of optimisation problems, and finding better solutions have important consequences, both for reducing the environmental impact, as well as reducing the cost. Metaheuristics (an umbrella term) represent the state-of-the art for solving large scale vehicle routing problems with time windows [44]. Our associated industrial partner SolveIT has extensive experience in optimisation of supply chain networks for the mining industry and other industries. Further example applications of EAs include automated antenna design, such as antennas with optimal radiation patterns that were evolved by NASA and sent into space [49].

The particular focus on parallel evolution in structured populations (Objective O4, WP 4) addresses the highly important and rapidly emerging topic on how to design **EAs for massively parallel hardware**. According to the International Roadmap for Semiconductors⁴, the number of cores in modern architectures will rise exponentially, increasing each year by a factor of 1.4, reaching 1 000 cores in 10 years time [51, Figure SYSD2]. This brings about both tremendous opportunities and challenges: opportunities for speeding up computation through parallel hardware, allowing for increased scalability, but also challenges in designing structured populations that use parallelism in evolution most effectively. A 2013 survey paper on parallel EAs and other parallel metaheuristics by leading figures in the field [4] lists 130 successful high-impact applications of these algorithms to real-world problems in various domains: automation and robotics; bioinformatics; engineering design; hydraulic engineering; information processing, classification, and data mining; manufacturing and industrial applications; routing, logistics and vehicle planning; scheduling; software engineering and software development; telecommunications; energy and power network optimisation; health and medicine; and many more. Thereby they acknowledge the need for a solid theory of parallel metaheuristics, saying that “the more we know on the parallel algorithm and the problem landscape the better we can build new operators and collaborative algorithms” and that “developments in theory on parallel metaheuristics is a must” [4, page 31].

SAGE will provide such a theory (deliverable D 4.2), including recommendations on how to design parallel EAs for practical problems (deliverable D 4.3). These contributions will enable researchers in

²According to ISI Web of Science, AutoDock was the most cited of all docking programs in 2009.

³http://en.wikipedia.org/wiki/Searching_the_conformational_space_for_docking

⁴The International Roadmap for Semiconductors is sponsored by the five leading chip manufacturing regions in the world: Europe, Japan, Korea, Taiwan, and the United States.

applied research to make informed design choices, and assist in developing novel forms of parallel EAs and increasing the efficiency of parallel EAs in practice.

Finally, making EAs more effective opens new possibilities, for instance inclusion of secondary optimisation criteria such as minimising the environmental impact of vehicle routing schedules with evolutionary algorithms [95].

Novel applications of evolutionary processes

The *third pathway to impact* comes through novel and innovative uses of evolutionary processes. These cannot be foreseen in their entirety at this stage, but a few representative promising examples are given. It has been predicted that EAs can become a central ingredient in future **self-optimising mobile networks** [101]. Mobile networks form a central and critical component of European infrastructure, and such an application would therefore put a strong demand on the dependability of the evolutionary process. The key to guarantee dependability of such a system is a strong theoretical foundation.

The field of **evolutionary medicine** promises to provide insight into common modern diseases by elucidating our evolutionary history [78]. Conditions such as asthma, autoimmune disease and cancer can result from mismatches between ancient adaptations and modern environments. Likewise, the pathogenicity of virus and bacteria infecting human populations require an evolutionary explanation. These insights can provide a path towards better therapeutic strategies, and a better understanding of the dynamics of evolution, such as the one aimed at by SAGE, will be instrumental in achieving this.

Another potential field of impact is the management of **biodiversity in natural habitats**. The impending threat of climate change stresses the need for better strategies for the preservation of biodiversity. Current models rarely take into account evolutionary processes, but this shortcoming has been recently recognised [46]. In order to achieve effective predictions of the dynamics of genetic diversity, models of adaptive dynamics that are realistic enough will be required. These models will have to be able to deal with multiple loci and to be flexible enough to include effects of a changing environment (deliverable D 2.3). The results arising from SAGE will be an important step towards this goal.

Impact on society

While the short and medium term impact of SAGE will mostly be felt in science and technology, the project will have significant positive long-term effects, such as a stronger economy and improved health.

A consequence of the industrial impact will be a stronger and more competitive European industry, which will contribute to creating new jobs and strengthening the economy. Other benefits will come from better medical services and products, including **more effective drug design**. Evolutionary algorithms used in the AutoDock software [105] were instrumental in discovering the first clinically-approved antiviral drug for HIV (Isentress) by Merck Pharmaceutical Company⁵. It is therefore not inconceivable that downstream impact from SAGE will contribute to developing novel drugs and better therapeutic strategies for cancer, which would have immediate economical and societal benefits. Another example is **prevention of drug resistance**. Drug resistance has become an increasingly important problem in the UK and in other countries. For example, the cases of Carpanem (an important class of antibiotics) resistance increased from less than five cases a year in 2007 to more than 600 in 2011⁶ in the UK. These strains are evolving under selection pressure originating from the widespread use of antibiotics. This represents a relatively simple case of biological adaptation [75]. Recently, there has been considerable effort from the scientific community to identify the mutations that cause this resistance as it represents one of the most pressing issues in evolutionary biology [94]. It was found that these mutations are typically found in the same genes and that the order in which these mutations fix conditions the probability of acquiring resistance [119]. This means that a thorough understanding of the dynamics of evolution and a focus on future prediction, such as the one our project proposes, will be a major factor for the effective design of therapeutic strategies for the use of antibiotics that slow down the evolution of resistance.

Finally, SAGE may also impact **policy making** with regard to bio-technological developments such as genetically modified organisms (GMOs). The EU has one of the most stringent regulations on GMOs in the world. All new GMOs are approved on a case by case, based on risk assessment to human health or environmental damage, all of these based on scientific evidence. One of the biggest concerns is that GMO genes start to introgress with native species by gene flow. For this, special buffer zones are set in place

⁵<http://autodock.scripps.edu/news/autodocks-role-in-developing-the-first-clinically-approved-hiv-integrase-inhibitor>

⁶<http://www.hpa.org.uk/hpr/archives/2011/news2411.htm>

where hybrids can co-exist. However, legislation on the size of these buffer zones varies wildly from country to country. SAGE considers evolutionary dynamics in spatially distributed populations (WP 4), which can be applied to this setting and provide a means of calculating effective buffer zones to prevent this gene-flow.

Maintaining impact after the project

The work programme of SAGE constitutes seminal work towards a unified quantitative theory of evolution as well as proof-of-concept that synergies between PG and EC can achieve the desired breakthrough in our understanding of natural and artificial evolution. As explained above, even these seminal results will have immediate and lasting impact.

In addition, we will ensure that SAGE has a lasting impact after the project has ended. Lasting scientific impact will be achieved by creating a vibrant and sustainable community; technological and societal impact will be ensured by proper dissemination of results, and by engaging with leading companies that use evolutionary processes.

Our long-term plan is to stimulate cross-talk between the fields by co-locating workshops on evolutionary computation in theoretical biology conferences, as well as workshops on population genetics in conferences on evolutionary computation. One of the main goals of these workshops will be to communicate the open problems in each field to the other field within the common framework provided by WP 1, to entice researchers from both fields to study problems that are significant for the other field.

It is essential that the new scientific field becomes sustainable. This will be achieved by blurring the lines between population genetics and evolutionary computation. An important vehicle for this process will be the common terminology and model introduced in WP 1 (deliverable D 1.1). This new research direction also requires a common forum for research output. We will aim to establish a journal that focuses on disseminating high-quality theoretical results that cover both artificial and natural evolution.

An important ingredient for sustainability is to ensure a steady recruitment of young researchers to the new field. At the undergraduate and postgraduate taught level, this will be achieved by the introduction of courses that cover both evolutionary algorithms and population genetics. These will initially be launched by the institutions and universities participating in SAGE. The undergraduate and postgraduate students trained by the project will bring the developed techniques into industrial applications.

Related international research activities

The SAGE project is unique in its mission to synergise population genetics and evolutionary computation. However, there are a number of related international research activities that highlight the world-wide interest in, and potential of studying evolutionary processes. This includes initiatives to synergise biology with wider areas of theory of computing and stochastic processes, respectively.

The Simons Foundation has donated \$ 60M in funding to UC Berkeley to establish a Theory of Computing Institute. The goal of the institute is to identify programs that can lead to substantial advances in theoretical computer science. One of their programs will be “Evolutionary Biology and the Theory of Computing”. Other institutions have also started research in this direction. The Princeton Center for Computational Intractability is a joint venture of Princeton University, Institute for Advanced Study, Rutgers University and New York University, which is funded by a \$ 10M Expeditions Grant of the National Science Foundation (NSF). They currently organise a workshop on “Natural Algorithms and the Sciences” which will also bring together researchers from computer science and biology. This shows that there is a significant interest in an ICT perspective of evolution, and the EU is already losing some leading researchers in Europe to this institute.

In 2011, the German research council (DFG) funded a six year Priority Programme on Probabilistic Structures in Evolution (SPP 1590), with fifteen German institutions participating and a total budget of 5.3M €. The programme is focussed on the theoretical study of stochastic processes in population genetics and evolutionary game theory. The topic does not significantly overlap with SAGE as SPP 1590 neither deals with evolutionary algorithms, nor does it consider the speed of adaptation. However, it shows that the biggest economy in Europe sees the value of providing significant funding to study evolution mathematically.

Further evidence at the European level is the ERC Advanced Grant 250152 SelectionInformation led by Nick Barton which aims at bridging computer science and biology. We will exploit synergies between this project and SAGE through knowledge exchange and potential collaborations in the context of our work programme.

The consortium board will keep track of these research activities as well as emerging ones and actively explore synergies and potential collaborations.

3.2 Contribution at the European level

SAGE is inherently an interdisciplinary effort, based on strong collaborations, and is designed to foster future collaborations between researchers in evolutionary computation and population genetics. For the first time, these two complementary fields will be brought together, resulting in a much extended research community, and in a wealth of new results for both fields. This will translate in increased attention from both fields to publications in the opposite field and cross-publications. Some of the members of our consortium have already begun to spark this interdisciplinary interest [7, 91].

Over time, these efforts will develop into long-lasting collaborations between the researchers in both fields, helping to make Europe a destination for world-class researchers and increasing the attractiveness of the EU as a research location, by exploiting cross-border synergies. This will certainly be true for the members of the consortium, who are investing their careers in this line of research, and are already engaged in interdisciplinary collaborations. They will serve as role-models for young scientists, inspiring them to tackle ambitious research projects. The students and postdocs exposed to this project will be acquainted with a rich and solid research methodology and the inner workings of evolutionary processes, both in nature and in the context of optimisation methods. They will further be acquainted with rich knowledge on their strengths and shortcomings and carry this knowledge with them, as they move into industry. Hence they become catalysts of increased adoption of these methods, and of innovation as they apply them to solving new and unexpected problems.

The initial injection of new results and tools in both fields that SAGE will challenge current thinking in both fields, leading to a profound impact on how research is performed in both fields. It will introduce new tools that will enable the analysis of problems that were intractable before, such as the analysis of population EAs in EC and long-term predictions involving many genes in PG.

Furthermore, this project will translate into a series of technological advances, with a strong potential for impacting on all areas that require optimisation methods. In turn, this will potentiate future projects that require robust optimisation techniques, by enabling the tools required by ambitious projects, allowing for a cross-cutting impact across specific strategic priority areas of EU challenges.

Specifically, SAGE will have a significant contribution to projects dealing with *smart cities*, where many optimisation problems need to be solved. Given the complexity and breadth of a city, analytical models are unlikely to be useful, rendering traditional optimisation techniques useless. SAGE will contribute to the development of efficient evolutionary techniques for solving these kinds of optimisation problems. The same can be said about projects that aim at technologies for *clean and efficient energy*: optimisation techniques that do not depend on the availability of analytical models are required for their success. This vision is already becoming reality as evolutionary methods have recently been used for optimising wind farms, with involvement from our consortium [115]. SAGE will also contribute toward stopping the increasing threat of *antimicrobial resistance*. The acquisition of resistance by biological agents is inherently an evolutionary process. SAGE will allow for a more thorough understanding of the evolutionary process, consequently enabling better strategies for the prevention of the evolution of resistance.

SAGE will also contribute to the future growth of the EU by empowering innovative small and medium sized enterprises (SMEs). Several SMEs have already expressed interest in SAGE's expected outcomes and are part of this proposal as associated industrial partners. These are SMEs that make use of search-based heuristics for optimisation and for whom a better theoretical understanding of evolutionary processes can translate into higher efficiency.

These impacts will translate in a significant macro-economic impact over the long run on European society, leading to a sustainable increase in jobs and economic growth in a knowledge-based society, and in developing a more open and competitive European Research Area (ERA), making it more attractive as a research location.

3.3 Dissemination and use of project results

An important goal of the SAGE project is the dissemination of our results, since the impact of our project crucially depends on its widespread adoption by researchers and practitioners.

The dissemination strategy targets three separate groups of beneficiaries. The first target group are the academic communities that study or use evolutionary processes. This target group can be separated into two sub-groups. The first sub-group are researchers who have evolutionary processes as their primary object of study, for example theoreticians and practitioners in evolutionary computation, population genetics,

and other fields. The second sub-group consists of other scientific and engineering disciplines which use evolutionary processes as a method to achieve some other goal, for example structural optimisation. The second target group are companies, industrialists, and practitioners who employ evolutionary processes for commercial purposes. The third target group are the general public and the public media.

SAGE will employ all necessary communication channels to ensure that these target groups are reached effectively, and that the maximum impact is achieved. The central communication channel will be the SAGE project website, which will be developed and hosted at UNI-JENA (deliverable D 6.2). The website will have an internal part available only to consortium members and project partners, serving as a primary communication channel within the project. The external website will have three main sections, tailored to the three target groups described above. The SAGE project website will be enriched with a steady stream of project-related information from the first day of the project, to guarantee that it will be visited frequently by media and other target groups. Social media technology will be integrated to further strengthen the SAGE website. The design and development of the SAGE website and use of accompanying social media technology will be supported by the Communications teams of UNI-JENA and UNOTT who have extensive professional expertise in this area.

Dissemination to the academic community

Dissemination to the *academic community* will be via several channels, including peer-reviewed scientific publications, the SAGE website, internal training of interdisciplinary researchers, tutorials, workshops, and the scientific advisory board. The primary channel will be research reports and peer-reviewed papers in high-quality conference proceedings and scientific journals on evolutionary computation (e.g. Evolutionary Computation, IEEE Transactions on Evolutionary Computation, GECCO, PPSN, CEC) and high-quality scientific journals on population genetics (e.g. PNAS, Genetics, Evolution, Journal of Theoretical Biology). Preprints of reports and papers will be publicly available on the SAGE website, along with a complete bibliography of the publications that have arisen from the project. We will also organise tutorials and workshops about the outcomes of the project at the leading conferences on evolutionary computation (GECCO, PPSN, CEC) and evolutionary biology (ESEB). The scientific advisory board as well as associated industrial partners will receive new results in a timely manner, hence targeting high-profile beneficiaries in all relevant academic and industrial areas.

The wiki produced in deliverable D 6.2 will be a central component of the SAGE website targeting the academic community. The wiki will provide a comprehensive comparison and translation of concepts and terminology from evolutionary computation and population genetics, thus becoming a reference point for researchers in both fields. To further support the community, each wiki article will have comment fields that will stimulate discussions among researchers on the relevant topic. The public website will also list all publications and reports from the project, along with a news section that informs about relevant activities and progress. The SAGE website will therefore not only have a community building role within the new scientific discipline, but will also generate a source of traffic to the parts of the website that describe the results from SAGE, thus increasing the impact of SAGE.

The project provides ideal opportunities for final year projects as well as involving PhD students to assist in the work programme. This is explicitly recognised by the European Commission as a feature of basic research projects: *the unity of research and education provides an optimal environment for PhD and master students to learn about the limits of the current state of our knowledge. The significant contributions of basic research to advanced education cannot be overestimated* [1].

The project members as well as associated undergraduate and postgraduate students will establish a network with mutual visits between the participating institutions to encourage mobility within the EU. They will be a highly efficient mean to disseminate our results as they will move on to new places and career challenges, carrying with them what they have learned during their involvement with SAGE. Furthermore, maintaining and expanding active collaborations with leading researchers in application areas will also help in disseminating our work.

To cement the connections between population genetics and evolutionary computation that will have been established during the first years of the project, a workshop will be organised towards the end of the project (deliverable D 6.3). Leading researchers from the two fields as well as selected industrialists will be invited. We plan to hold the workshop at the Schloss Dagstuhl – Leibniz Center for Informatics in Germany, which is considered as a world leading venue for computer science. The venue is ideal for SAGE, as the centre especially welcomes interdisciplinary topics with relevance to industry. The focus of the workshop will be on generating further ideas and collaboration, thus stimulating further activity, and ensuring that SAGE

will have a long lasting impact. We will invite the scientific advisory board and the associated industrial partners to increase dissemination of the research outcomes.

Dissemination to industry

Dissemination to *industry* will be conveyed via a range of channels, and will be supported by the Business Engagement & Innovation Services at the University of Nottingham, and similar services in the other institutions. Our associated industrial partners (see Section 2.1 at page 31) form a direct communication channel to industries that employ evolutionary algorithms and other evolutionary meta-heuristics. They are very interested in the outcomes of SAGE (see letters of support in the appendix) and will be the first users of new technology. We will engage with our associated industrial partners, and expand to other industries during the course of the project. A dedicated section of the SAGE website will provide information and best-practice relevant to the industry, for example the recommendations for how to use recombination operators arising from WP 3. The associated industrial partners will also be invited to our workshops (see WP 6).

Given the ubiquitous use of evolutionary techniques in industrial optimisation (cf. Section 3.1), there is strong demand for skilled engineers and trained professionals with a background in these techniques. Besides the postdoctoral researches funded by the SAGE project, the involved researchers will train a significant number of undergraduate and postgraduate students in these advanced optimisation techniques. As many of these students will not stay in academia, but move on to European industries, this knowledge will be timely disseminated and applied.

All PIs from EC are members of the *IEEE Task Force on Theoretical Foundations of Bio-inspired Computation*⁷, which was instantiated by IEEE to promote the theory of evolutionary algorithms and related optimisation techniques to industry. This provides an excellent venue for disseminating our results to a potential audience of more than 425 000 IEEE members in more than 160 countries.

Dissemination to the general public

Dissemination to the *general public* will follow several channels, including traditional media channels, social media channels, school outreach activities, “open days” at the involved institutions as well as the project website. Engagement with media in Europe and internationally will be supported by the Communications team at the University of Nottingham. This team is staffed by media professionals with past experience from major media companies such as the BBC. They have extensive experience in communicating research results, both to traditional media channels such as newspapers and TV channels, as well as to social media. The international profile of the University of Nottingham allows news outlets beyond Europe to be reached, notably in China and Malaysia. The Communications team will assist us in preparing press releases, targeting specific media outlets, fine-tuning the social media strategy, as well as providing general media and marketing advice. IST Austria will contribute with expertise in promoting science.

A range of **outreach activities** about SAGE and evolutionary theory will be organised to raise awareness of the research to the general public. In particular, we will visit several high schools. These activities will be initiated in the UK, and will be coordinated with Louise Orpin from the OR Society in the UK. The events will be promoted via the Learning about OR website (www.learnaboutor.co.uk), and similar web sites aimed at school outreach events in the STEM disciplines. Dr Vince Knight, the chair of the OR in School’s task force in the UK, as well as Prof Paul Harper, who has extensive experience in outreach activities, will provide further assistance. The experience from the outreach events in UK schools will be used to organise similar events in Austria and Germany.

Further outreach activities include a series of events at university open days. In particular, the SAGE project will set up a “research island” at the IST Austria Open Campus Day, which is a popular annual event attracting more than 1 500 school children and other visitors from nearby regions.

Another important channel for outreach will be the SAGE website targeting the general public and the media will contain press releases, popular science descriptions of key results, supporting material for school outreach activities, and further relevant material.

Exploitation of project results

The work programme has been designed to maximise the level of *exploitation*, not only in the academic community, but also in industries which exploit evolutionary processes for commercial purposes.

The totality of deliverables constitutes a powerful analytical framework for quantifying the speed of adaptation in biological evolution and evolutionary algorithms. This represents an ideal platform for further

⁷<http://www.cs.bham.ac.uk/~olivetps/CISTheory.html>

work in extending, generalising and refining this framework and applying it in quantitative studies of concrete evolutionary processes or algorithms. This opportunity presents itself to SAGE project members beyond the duration of the project as well as to other researchers or beneficiaries collaborating with our consortium or building on our work on their own.

There is huge potential for industrial exploitation in terms of obtaining quantitative results on the efficiency of evolutionary algorithms tailored to specific problems or problem classes prevalent in particular applications. This is highly relevant for many engineering and other applied disciplines which represent end-users of EAs. Our dissemination activities to academic beneficiaries as well as industry will advertise the scientific platform SAGE provides, in order to bring about maximum exploitation.

The SAGE work programme also provides high potential for exploration in academia. The unified model accessible both to population genetics and evolutionary computation will be a strong incentive for researchers in both fields to express their results within this framework. Analytical tools, such as diffusion approximation and drift analysis, are general mathematical results that explain how various fundamental properties of evolutionary processes relate to each other. For example, drift analysis explains how the long-term changes in an evolutionary process depend on its short-term changes. Such analytical tools play a fundamental role as building-blocks in population genetics and evolutionary computation. The project's focus on stronger and more general analytic techniques maximises the possibilities for exploitation by other researchers. The wiki (deliverable D 1.1) will be a widely accessible and useful resource. The results from SAGE will therefore be widely used in the academic community, long after the end of the project.

Finally, the work programme is also designed to facilitate exploitation in terms of recommendations for how to configure and parameter-tune evolutionary algorithms that are easy to use. The industrial partners will be used as one dissemination channel to the industry. As discussed in the previous subsection, improved optimisation algorithms will have a tremendous impact on the cost and quality of products from a wide range of industries. The economic impact of the easy-to-exploit results generated from this project will be significant. Several associated industrial partners, ranging from small to multinational companies (see the appendix), will facilitate the dissemination and exploitation to the industry.

In order to achieve maximum exploitation, all *intellectual property* generated from the project will be made available in the public domain. Research results in the form of preprints and reports will be made publicly available via the SAGE website. Any source code arising from the project will be made available. Support from the Engagement & Innovation Services at the University of Nottingham will be sought if particular questions arise regarding commercial exploitation of intellectual property.

The consortium board will closely monitor the dissemination and exploitation activities, and evaluate their progress annually. The scientific advisory board as well as associated industrial partners will receive new results in a timely manner.

4 Ethical issues

SAGE is a foundational project aiming at establishing an analytical framework for understanding evolutionary processes. Evolutionary computation is based on simulating a purely virtual evolution of candidate solutions, with no ethical issues arising from it. With regard to natural evolution, the project will make use of mathematical models from population genetics and does not require experimental manipulation involving biological organisms or samples. As such, we do not foresee any ethical issues that may arise in the project as can be seen also from the following table. Should any ethical questions arise, we will seek advice from the ethical committees of the participating institutions.

Ethical Issues Table

	YES	PAGE
Informed Consent		
• Does the proposal involve children?	No	
• Does the proposal involve patients?	No	
• Does the proposal involve persons not able to give consent?	No	
• Does the proposal involve adult healthy volunteers?	No	
Biological research		
• Does the proposal involve Human Genetic Material?	No	
• Does the proposal involve Human biological samples?	No	
• Does the proposal involve Human biological data collection?	No	
• Does the proposal involve Human Embryos?	No	
• Does the proposal involve Human Foetal Tissue or Cells?	No	
• Does the proposal involve Human Embryonic Stem Cells?	No	
Privacy		
• Does the proposal involve processing of genetic information or personal data (e.g. health, sexual lifestyle, ethnicity, political opinion, religious or philosophical conviction)	No	
• Does the proposal involve tracking the location or observation of people without their knowledge?	No	
Research on Animals		
• Does the proposal involve research on animals?	No	
• Are those animals transgenic small laboratory animals?	No	
• Are those animals transgenic farm animals?	No	
• Are those animals cloned farm animals?	No	
• Are those animals non-human primates?	No	
Research Involving Developing Countries		
• Is any part of the research carried out in countries outside of the European Union and FP7 Associated states?	No	
Dual Use		
• Does the research have direct military applications?	No	
• Does the research have the potential for terrorist abuse?	No	
ICT Implants		
• Does the proposal involve clinical trials of ICT implants?	No	
I CONFIRM THAT NONE OF THE ABOVE ISSUES APPLY TO MY PROPOSAL	Yes	

Bibliography

We have marked own publications with “•” on the left margin. Additionally, all entries are crossreferenced with the pages where they are cited.

- [1] Towards Horizon 2020 – Recommendations of ISTAG on FP7 ICT work program 2013. ISTAG, 2012. (pages 41, 42, 48).
- [2] E. Alba. Parallel evolutionary algorithms can achieve super-linear performance. *Information Processing Letters*, 82:7–13, 2002. (page 16).
- [3] E. Alba. *Parallel Metaheuristics: A New Class of Algorithms*. Wiley-Interscience, 2005. (page 15).
- [4] E. Alba, G. Luque, and S. Nesmachnow. Parallel metaheuristics: Recent advances and new trends. *International Transactions in Operational Research*, 20:1–48, 2013. (pages 16, 44).
- [5] S. Arora, Y. Rabani, and U. V. Vazirani. Simulating quadratic dynamical systems is PSPACE-complete. In *Proceedings of the 26th ACM Symposium on the Theory of Computing (STOC)*, pp. 459–467, 1994. (page 14).
- [6] A. Auger and B. Doerr, editors. *Theory of Randomized Search Heuristics – Foundations and Recent Developments*. Number 1 in Series on Theoretical Computer Science. World Scientific, 2011. (pages 1, 3, 4, 5).
- [7] N. Barton and T. Paixão. Can quantitative and population genetics help us understand evolutionary computation? In *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO)*, 2013. in press. (pages 5, 47).
- [8] N. Barton and S. Rouhani. The frequency of shifts between alternative equilibria. *Journal of Theoretical Biology*, 125:397–418, 1987. (page 11).
- [9] N. H. Barton and B. Charlesworth. Why sex and recombination? *Science*, 281:1986–1990, 1998. (page 13).
- [10] V. Beisvåg, P. K. Lehre, H. Midelfart, H. Aass, O. Geiran, A. K. Sandvik, A. Lægreid, J. Komorowski, and Ø. Ellingsen. Aetiology-specific patterns in end-stage heart failure patients identified by functional annotation and classification of microarray data. *European Journal of Heart Failure*, 8:381–389, 2006. (page 34).
- [11] G. Bell. *The masterpiece of nature the evolution and genetics of sexuality*. Univ. of California Press, 1982. (page 13).
- [12] S. A. Benner and A. M. Sismour. Synthetic biology. *Nature Reviews Genetics*, 6:533–543, 2005. (page 42).
- [13] M. J. Brauer, M. T. Holder, L. A. Dries, D. J. Zwickl, P. O. Lewis, and D. M. Hillis. Genetic algorithms and parallel processing in maximum-likelihood phylogeny inference. *Molecular Biology and Evolution*, 19:1717–1726, 2002. (page 42).
- [14] K. Bringmann and T. Friedrich. Approximation quality of the hypervolume indicator. *Artificial Intelligence*, 195:265–290, 2013. (pages 10, 36).
- [15] D. Brockhoff, T. Friedrich, N. Hebbinghaus, C. Klein, F. Neumann, and E. Zitzler. On the effects of adding objectives to plateau functions. *IEEE Transactions on Evolutionary Computation*, 13:591–603, 2009. (pages 10, 36).
- [16] R. Bürger. *The Mathematical Theory of Selection, Recombination, and Mutation*. Wiley, 1 edition, 2000. (page 2).
- [17] J. Carneiro, T. Paixão, D. Milutinovic, J. Sousa, K. Leon, R. Gardner, and J. Faro. Immunological self-tolerance: Lessons from mathematical modeling. *Journal of Computational and Applied Mathematics*, 184:77–100, 2005. (page 9).
- [18] J. Carneiro, K. Leon, I. Caramalho, C. Van Den Dool, R. Gardner, V. Oliveira, M.-L. Bergman, N. Sepúlveda, T. Paixão, J. Faro, and J. Demengeot. When three is not a crowd: a crossregulation model of the dynamics and repertoire selection of regulatory CD4+ T cells. *Immunological Reviews*, 216:48–68, 2007. (page 10).
- [19] B. Chazelle. Natural algorithms. In *Proceedings of the 20th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pp. 422–431. SIAM, 2009. (page 30).
- [20] R. Chiong, T. Weise, and Z. Michalewicz, editors. *Variants of Evolutionary Algorithms for Real-World Applications*. Springer, 2012. (pages 1, 13).
- [21] S. Ciliberti, O. C. Martin, and A. Wagner. Innovation and robustness in complex regulatory gene networks. *Proceedings of the National Academy of Sciences of the United States of America*, 104:13591–6, 2007. (page 11).
- [22] M. Collins. Finding needles in haystacks is harder with neutrality. *Genetic Programming and Evolvable Machines*, 7:131–144, 2006. (page 12).
- [23] T. F. Cooper, T. Paixão, and J. A. Heinemann. Within-host competition selects for plasmid-encoded toxin-antitoxin systems. *Proceedings of the Royal Society B: Biological Sciences*, 277:3149–3155, 2010. (pages 10, 16, 19).
- [24] B. Doerr and T. Friedrich. Deterministic random walks on the two-dimensional grid. *Combinatorics, Probability and Computing*, 18:123–144, 2009. (page 36).
- [25] B. Doerr, M. Gnewuch, N. Hebbinghaus, and F. Neumann. A rigorous view on neutrality. In *IEEE Congress on Evolutionary Computation*, pp. 2591–2597, 2007. (page 12).
- [26] B. Doerr, T. Friedrich, and T. Sauerwald. Quasirandom rumor spreading. In *Proceedings of the 19th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pp. 773–781. SIAM, 2008. (pages 18, 19).

- [27] B. Doerr, M. Fouz, and T. Friedrich. Social networks spread rumors in sublogarithmic time. In *Proceedings of the 43rd ACM Symposium on the Theory of Computing (STOC)*, pp. 21–30, 2011. (pages 10, 18, 19, 36, 42).
- [28] B. Doerr, D. Johannsen, T. Kötzing, P. K. Lehre, M. Wagner, and C. Winzen. Faster black-box algorithms through higher arity operators. In *Proceedings of the 11th International Workshop on Foundations of Genetic Algorithms (FOGA)*, pp. 163–172, 2011. (pages 9, 10).
- [29] B. Doerr, E. Happ, and C. Klein. Crossover can provably be useful in evolutionary computation. *Theoretical Computer Science*, 425:17–33, 2012. (page 14).
- [30] S. Droste. Analysis of the (1+1) EA for a dynamically bitwise changing onemax. In *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO)*, pp. 909–921. 2003. (page 13).
- [31] S. Droste, T. Jansen, and I. Wegener. A rigorous complexity analysis of the (1+1) evolutionary algorithm for separable functions with Boolean inputs. *Evolutionary Computation*, 6:185–196, 1998. (page 4).
- [32] S. Droste, T. Jansen, and I. Wegener. On the analysis of the (1+1) evolutionary algorithm. *Theoretical Computer Science*, 276:51–81, 2002. (pages 3, 7).
- [33] S. Droste, T. Jansen, and I. Wegener. Upper and lower bounds for randomized search heuristics in black-box optimization. *Theory of Computing Systems*, 39:525–544, 2006. (page 9).
- [34] F. Duddeck. Multidisciplinary optimization of car bodies. *Structural and Multidisciplinary Optimization*, 35:375–389, 2008. (pages 2, 44).
- [35] A. E. Eiben and J. E. Smith. *Introduction to Evolutionary Computing*. Natural Computing Series. Springer, corr. 2nd printing edition, 2007. (page 1).
- [36] EPSRC/ESRC. Review of research status of operational research in the UK, 2004. (page 42).
- [37] European Federation of Pharmaceutical Industries and Associations (EFPIA). The pharmaceutical industry in figures, 2012. (page 44).
- [38] W. Fontana and P. Schuster. Shaping space: the possible and the attainable in RNA genotype-phenotype mapping. *Journal of Theoretical Biology*, 194:491–515, 1998. (page 12).
- [39] S. J. Freeland and L. D. Hurst. The genetic code is one in a million. *Journal of Molecular Evolution*, 47:238–248, 1998. (page 11).
- [40] T. Friedrich and F. Neumann. When to use bit-wise neutrality. *Natural Computing*, 9:283–294, 2010. (pages 12, 13, 36).
- [41] T. Friedrich and T. Sauerwald. Near-perfect load balancing by randomized rounding. In *Proceedings of the 41st ACM Symposium on the Theory of Computing (STOC)*, pp. 121–130. ACM Press, 2009. (pages 10, 36).
- [42] T. Friedrich, P. S. Oliveto, D. Sudholt, and C. Witt. Analysis of diversity-preserving mechanisms for global exploration. *Evolutionary Computation*, 17:455–476, 2009. (pages 11, 13, 15, 18, 36).
- [43] E. Galván-López, R. Poli, A. Kattan, M. O’Neill, and A. Brabazon. Neutrality in evolutionary algorithms... what do we know? *Evolving Systems*, 2:145–163, 2011. (page 12).
- [44] M. Gendreau and C. D. Tarantilis. Solving large-scale vehicle routing problems with time windows: The state-of-the-art. Technical Report CIRRELT-2010-04, CIRRELT, 2010. (pages 2, 44).
- [45] J. H. Gillespie. Molecular evolution over the mutational landscape. *Evolution*, 38:1116–1129, 1984. (page 7).
- [46] A. A. Hoffmann and C. M. Sgró. Climate change and evolutionary adaptation. *Nature*, 470:479–485, 2011. (page 45).
- [47] H. H. Hoos. Personal communication, 2013. (page 42).
- [48] G. S. Hornby and T. Yu. EC practitioners: results of the first survey. *SIGEVolution*, 2:2–8, 2007. (pages 1, 2, 3, 43).
- [49] G. S. Hornby, A. Globus, D. S. Linden, and J. D. Lohn. Automated antenna design with evolutionary algorithms. Technical report, American Institute of Aeronautics and Astronautics, 2006. (pages 2, 44).
- [50] R. Horner. *A taxonomic guide to some common marine phytoplankton*. Biopress Ltd, 2002. (page 13).
- [51] ITRS. System drivers. In *International Technology Roadmap for Semiconductors*. 2011 edition. <http://www.itrs.net/Links/2011ITRS/2011Chapters/2011SysDrivers.pdf>. (pages 19, 44).
- [52] T. Jansen. *Analyzing Evolutionary Algorithms*. Natural Computing Series. Springer, 2013. (page 3).
- [53] T. Jansen and U. Schellbach. Theoretical analysis of a mutation-based evolutionary algorithm for a tracking problem in the lattice. In *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO)*, pp. 841–848, 2005. (page 13).
- [54] T. Jansen and I. Wegener. Real royal road functions—where crossover provably is essential. *Discrete Applied Mathematics*, 149:111–125, 2005. (page 14).
- [55] T. Jansen and I. Wegener. A comparison of simulated annealing with a simple evolutionary algorithm on pseudo-Boolean functions of unitation. *Theoretical Computer Science*, 386:73–93, 2007. (page 11).
- [56] T. Jansen and I. Wegener. The analysis of evolutionary algorithms – a proof that crossover really can help. *Algorithmica*, 34:47–66, 2002. (page 14).
- [57] T. Jorg, O. Martin, and A. Wagner. Neutral network sizes of biological RNA molecules can be computed and are not atypically small. *BMC Bioinformatics*, 9:464, 2008. (page 11).

- [58] S. Kauffman and S. Levin. Towards a general theory of adaptive walks on rugged landscapes. *Journal of Theoretical Biology*, 128:11–45, 1987. (page 8).
- [59] M. Kimura. Evolutionary rate at the molecular level. *Nature*, 217:624–626, 1968. (page 11).
- [60] M. Kimura. Genetic variability maintained in a finite population due to mutational production of neutral and nearly neutral isoalleles. *Genetics Research*, 11:247–270, 1968. (page 12).
- [61] T. Kötzing, D. Sudholt, and M. Theile. How crossover helps in pseudo-boolean optimization. In *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO)*, pp. 989–996, 2011. (pages 14, 15, 37).
- [62] J. Lässig and D. Sudholt. The benefit of migration in parallel evolutionary algorithms. In *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO)*, pp. 1105–1112, 2010. (pages 18, 19, 37).
- [63] J. Lässig and D. Sudholt. General scheme for analyzing running times of parallel evolutionary algorithms. In *Proceedings of the 11th International Conference Parallel Problem Solving From Nature (PPSN)*, pp. 234–243, 2010. (pages 18, 19, 37).
- [64] J. Lässig and D. Sudholt. Experimental supplements to the theoretical analysis of migration in the island model. In *Proceedings of the 11th International Conference Parallel Problem Solving From Nature (PPSN)*, pp. 224–233, 2010. (pages 18, 19).
- [65] J. Lässig and D. Sudholt. Adaptive population models for offspring populations and parallel evolutionary algorithms. In *Proceedings of the 11th International Workshop on Foundations of Genetic Algorithms (FOGA)*, pp. 181–192, 2011. (pages 18, 19, 37).
- [66] J. Lässig and D. Sudholt. Analysis of speedups in parallel evolutionary algorithms for combinatorial optimization. In *Proceedings of the 22nd International Symposium Algorithms and Computation (ISAAC)*, pp. 405–414, 2011. (pages 18, 19, 37).
- [67] P. K. Lehre. *Complexity and Geometry in Artificial Development*. PhD thesis, Department of Computer and Information Science, Norwegian University of Science and Technology, 2006. (pages 13, 34).
- [68] P. K. Lehre. Negative drift in populations. In *Proceedings of the 11th International Conference Parallel Problem Solving From Nature (PPSN)*, pp. 244–253, 2011. (pages 10, 34).
- [69] P. K. Lehre. Fitness-levels for non-elitist populations. In *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO)*, pp. 2075–2082, 2011. (pages 10, 34).
- [70] P. K. Lehre and P. C. Haddow. Accessibility and runtime between convex neutral networks. In *Proceedings of the 6th International Conference on Simulated Evolution and Learning (SEAL)*, pp. 734–741, 2006. (page 13).
- [71] P. K. Lehre and C. Witt. Black-box search by unbiased variation. *Algorithmica*, pp. 1–20, 2012. (pages 9, 10, 34).
- [72] P. K. Lehre and X. Yao. Crossover can be constructive when computing unique input-output sequences. *Soft Computing*, 15:1675–1687, 2011. (pages 14, 15).
- [73] P. K. Lehre and X. Yao. On the impact of mutation-selection balance on the runtime of evolutionary algorithms. *IEEE Transactions on Evolutionary Computation*, 16:225–241, 2012. (page 34).
- [74] A. R. Lemmon and M. C. Milinkovitch. The metapopulation genetic algorithm: An efficient solution for the problem of large phylogeny estimation. *Proceedings of the National Academy of Sciences*, 99:10516–10521, 2002. (page 42).
- [75] B. R. Levin, V. Perrot, and N. Walker. Compensatory mutations, antibiotic resistance and the population genetics of adaptive evolution in bacteria. *Genetics*, 154:985–997, 2000. PMID: 10757748. (page 45).
- [76] P. O. Lewis. A genetic algorithm for maximum-likelihood phylogeny inference using nucleotide sequence data. *Molecular Biology and Evolution*, 15:277–283, 1998. (page 42).
- [77] R. C. Lewontin. *The genetic basis of evolutionary change*. Columbia University Press, 1974. (page 1).
- [78] J. B. Losos, S. J. Arnold, G. Bejerano, E. D. Brodie, D. Hibbett, H. E. Hoekstra, D. P. Mindell, A. Monteiro, C. Moritz, H. A. Orr, D. A. Petrov, S. S. Renner, R. E. Ricklefs, P. S. Soltis, and T. L. Turner. Evolutionary biology for the 21st century. *PLoS Biol*, 11:e1001466, 2013. (page 45).
- [79] G. Luque and E. Alba. *Parallel Genetic Algorithms—Theory and Real World Applications*, Vol. 367 of *Studies in Computational Intelligence*. Springer, 2011. (pages 15, 16, 18).
- [80] A. Mambrini, D. Sudholt, and X. Yao. Homogeneous and heterogeneous island models for the set cover problem. In *Proceedings of the 12th International Conference Parallel Problem Solving From Nature (PPSN)*, pp. 11–20, 2012. (pages 18, 19, 37).
- [81] Z. Michalewicz, M. Schmidt, M. Michalewicz, and C. Chiriach. Case study: An intelligent decision-support system. *IEEE Intelligent Systems*, 20:44–49, 2005. (page 44).
- [82] H. Mühlenbein and D. Schlierkamp-Voosen. Predictive models for the breeder genetic algorithm, I: Continuous parameter optimization. *Evolutionary Computation*, 1:25–49, 1993. (page 5).
- [83] N. Nedjah, L. de Macedo Mourelle, and E. Alba. *Parallel Evolutionary Computations*. Springer, 2006. (page 15).
- [84] F. Neumann and C. Witt. *Bioinspired Computation in Combinatorial Optimization – Algorithms and Their Computational Complexity*. Springer, 2010. (pages 1, 3, 4, 5, 8, 19, 31).
- [85] F. Neumann, P. S. Oliveto, G. Rudolph, and D. Sudholt. On the effectiveness of crossover for migration in parallel evolutionary algorithms. In *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO)*,

- pp. 1587–1594, 2011. (pages 14, 18, 19).
- [86] G. Ochoa, I. Harvey, and H. Buxton. On recombination and optimal mutation rates. In *Proceedings of Genetic and Evolutionary Computation Conference*, pp. 488–495. Morgan Kaufmann, 1999. (page 15).
 - [87] P. S. Oliveto, J. He, and X. Yao. Time complexity of evolutionary algorithms for combinatorial optimization: A decade of results. *International Journal of Automation and Computing*, 4:100–106, 2007. (pages 3, 4, 8).
 - [88] F. Oppacher and M. Wineberg. The shifting balance genetic algorithm: Improving the GA in a dynamic environment. In *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO)*, pp. 504–510. Morgan Kaufmann, 1999. (page 18).
 - [89] F. Oppacher and M. Wineberg. Reconstructing the shifting balance theory in a GA: Taking Sewall Wright seriously. In *Proceedings of the IEEE Congress on Evolutionary Computation (CEC)*, pp. 219–226, 2000. (pages 5, 18).
 - [90] T. Paixão and R. B. R. Azevedo. Redundancy and the evolution of cis-regulatory element multiplicity. *PLoS Computational Biology*, 6:e1000848, 2010. (pages 11, 12, 13, 15).
 - [91] T. Paixão and N. Barton. A variance decomposition approach to the analysis of genetic algorithms. In *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO)*, 2013. in press. (pages 5, 47).
 - [92] T. Paixão, T. P. Carvalho, D. P. Calado, and J. Carneiro. Quantitative insights into stochastic monoallelic expression of cytokine genes. *Immunology and Cell Biology*, 85:315–22, 2007. (pages 10, 13).
 - [93] T. Paixão, S. S. Phadke, R. B. R. Azevedo, and R. A. Zufall. Sex ratio evolution under probabilistic sex determination. *Evolution*, 65:2050–2060, 2011. (pages 10, 13, 15).
 - [94] S. R. Palumbi. Humans as the world’s greatest evolutionary force. *Science*, 293:1786–1790, 2001. (page 45).
 - [95] J. Partyka and R. Hall. 2012 vehicle routing software survey. *ORMS-Today*, 39, 2012. (page 45).
 - [96] Y. Rabani, Y. Rabinovich, and A. Sinclair. A computational view of population genetics. *Random Structures and Algorithms*, 12:313–334, 1998. (page 14).
 - [97] Y. Rabinovich, A. Sinclair, and A. Wigderson. Quadratic dynamical systems. In *Proceedings of the 33rd IEEE Annual Symposium on Foundations of Computer Science (FOCS)*, pp. 304–313, 1992. (page 14).
 - [98] J. N. Richter, A. Wright, and J. Paxton. Ignoble trails – where crossover is provably harmful. In *Proceedings of the 10th International Conference Parallel Problem Solving From Nature (PPSN)*, pp. 92–101, 2008. (page 14).
 - [99] P. Rohlfshagen, P. K. Lehre, and X. Yao. Dynamic evolutionary optimisation: an analysis of frequency and magnitude of change. In *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO)*, pp. 1713–1720, 2009. (page 13).
 - [100] D. Roze and N. H. Barton. The hill–robertson effect and the evolution of recombination. *Genetics*, 173:1793–1811, 2006. (pages 14, 15).
 - [101] O. Sallent, J. Pérez-Romero, J. Sanchez-Gonzalez, R. Agustí, M. Á. Díaz-Guerra, D. Henche, and D. Paul. A roadmap from UMTS optimization to LTE self-optimization. *IEEE Communications Magazine*, 49:172–182, 2011. (page 45).
 - [102] P. Schuster, W. Fontana, P. F. Stadler, and I. L. Hofacker. From sequences to shapes and back: a case study in RNA secondary structures. *Proceedings. Biological Sciences / The Royal Society*, 255:279–284, 1994. (page 11).
 - [103] M. Shpak and S. Gavrillets. *Population Genetics: Multilocus*. John Wiley & Sons, Ltd, 2001. (page 12).
 - [104] S. C. Sleight, B. A. Bartley, J. A. Lieviant, and H. M. Sauro. Designing and engineering evolutionary robust genetic circuits. *Journal of Biological Engineering*, 4:12, 2010. PMID: 21040586. (page 42).
 - [105] S. F. Sousa, P. A. Fernandes, and M. J. Ramos. Protein–ligand docking: Current status and future challenges. *Proteins: Structure, Function, and Bioinformatics*, 65:15–26, 2006. (pages 2, 44, 45).
 - [106] C.-P. Stelzer, J. Schmidt, A. Wiedlroither, and S. Riss. Loss of sexual reproduction and dwarfing in a small metazoan. *PLoS ONE*, 5:e12854, 2010. (page 13).
 - [107] T. Storch and I. Wegener. Real royal road functions for constant population size. *Theoretical Computer Science*, 320:123–134, 2004. (page 14).
 - [108] D. Sudholt. Crossover speeds up building-block assembly. In *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO)*, pp. 689–702, 2012. (page 15).
 - [109] D. Sudholt. Crossover is provably essential for the Ising model on trees. In *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO)*, pp. 1161–1167, 2005. (pages 14, 15, 37).
 - [110] D. Sudholt. Hybridizing evolutionary algorithms with variable-depth search to overcome local optima. *Algorithmica*, 59:343–368, 2011. (pages 8, 11, 13).
 - [111] D. Sudholt. General lower bounds for the running time of evolutionary algorithms. In *Proceedings of the 11th International Conference Parallel Problem Solving From Nature (PPSN)*, pp. 124–133, 2010. (page 10).
 - [112] D. Sudholt. Crossover speeds up building-block assembly. In *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO)*, pp. 689–696, 2012. (pages 15, 37).
 - [113] M. Tomassini. *Spatially Structured Evolutionary Algorithms: Artificial Evolution in Space and Time*. Springer, 2005. (page 15).
 - [114] K. Treier, A. Berg, P. Diederich, K. Lang, A. Osberghaus, F. Dimer, and J. Hubbuch. Examination of a genetic algorithm for the application in high-throughput downstream process development. *Biotechnology Journal*, 7:

- 1203–1215, 2012. (page 42).
- [115] K. Vladislavleva, T. Friedrich, F. Neumann, and M. Wagner. Predicting the energy output of wind farms based on weather data: Important variables and their correlation. *Renewable Energy*, 50:236–243, 2013. (page 47).
- [116] A. Wagner. Neutralism and selectionism: a network-based reconciliation. *Nature Reviews Genetics*, pp. 965–974, 2008. (page 12).
- [117] A. Wagner. Robustness and evolvability: a paradox resolved. *Proceedings of the Royal Society B: Biological Sciences*, 275:91–100, 2008. (page 12).
- [118] R. A. Watson and T. Jansen. A building-block royal road where crossover is provably essential. In *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO)*, pp. 1452–1459, 2007. (page 14).
- [119] D. M. Weinreich, N. F. Delaney, M. A. DePristo, and D. L. Hartl. Darwinian evolution can follow only very few mutational paths to fitter proteins. *Science*, 312:111–114, 2006. (page 45).
- [120] D. B. Weissman and N. H. Barton. Limits to the rate of adaptive substitution in sexual populations. *PLoS Genetics*, 8:e1002740, 2012. (page 14).
- [121] D. B. Weissman, M. M. Desai, D. S. Fisher, and M. W. Feldman. The rate at which asexual populations cross fitness valleys. *Theoretical Population Biology*, 75:286–300, 2009. (page 11).
- [122] D. B. Weissman, M. W. Feldman, and D. S. Fisher. The rate of fitness-valley crossing in sexual populations. *Genetics*, 186:1389–1410, December 2010. (page 11).
- [123] M. C. Whitlock, P. C. Phillips, F. B.-G. Moore, and S. J. Tonsor. Multiple fitness peaks and epistasis. *Annual Review of Ecology and Systematics*, 26:601–629, 1995. (page 11).
- [124] C. Witt. Worst-case and average-case approximations by simple randomized search heuristics. In *Proceedings of the 22nd Annual Symposium on Theoretical Aspects of Computer Science (STACS)*, pp. 44–56, 2005. (page 8).
- [125] S. Wright. The roles of mutation, inbreeding, crossbreeding and selection in evolution. In *Proceedings of the 6th International Congress on Genetics*, Vol. 1, pp. 356–366, 1932. (page 10).
- [126] T. Yu and J. Miller. Finding needles in haystacks is not hard with neutrality. In *Proceedings of the 5th European Conference on Genetic Programming (EuroGP)*, pp. 13–25, 2002. (page 12).
- [127] T. Yu, L. Davis, C. M. Baydar, and R. Roy, editors. *Evolutionary Computation in Practice*, Vol. 88 of *Studies in Computational Intelligence*. Springer, 2008. (page 1).
- [128] Y. Zhou, Z. Guo, and J. He. On the running time analysis of the (1+1) evolutionary algorithm for the subset sum problem. In K. Li, M. Fei, G. W. Irwin, and S. Ma, editors, *Bio-Inspired Computational Intelligence and Applications*, pp. 73–82. Springer Berlin Heidelberg, 2007. (page 8).

A Appendix

Enclosed: Letters of Support and PhD Certificates



9 April 2013

Dear Per Kristian,

Many thanks for informing us about your project "Speed of Adaptation in Population Genetics and Evolutionary Computation (SAGE)", and for the invitation to take part. We are pleased to support your research programme.

As you know, Schneider Electric is a global specialist in energy management with operations in more than 100 countries. We offer integrated solutions across multiple market segments, including leadership positions in Utilities & Infrastructures, Industries & Machine manufacturers, Non-residential buildings, Data centers & Networks, and in Residential. Focused on making energy safe, reliable, efficient, productive and green, the Group's 140,000 plus employees achieved sales of 24 billion Euros in 2012, through an active commitment to help individuals and organizations make the most of their energy.

Our subsidiary SolveIT Software is specialising in supply & demand optimisation and predictive modelling. Its award-winning applications are based on advanced optimisation and prediction algorithms. These applications support what-if analysis and optimise production & supply chain activities on both a local (plant) and global (network) level. Our mining software applications cover mine planning & scheduling as well as mine-to-port logistics optimisation.

Evolutionary algorithms form a core component in our proprietary SolveIT Optimisation Platform. This software is highly customizable and used by our customers in a wide range of industries, including Rio Tinto Iron Ore, BHP Iron Ore, Viterro, Pacific National Coal, Incitec Pivot, Foster's Group, Nufarm, Liquor Marketing Group, PFD Food Services, Pernod Ricard, Fortescue Metals Group, Xstrata Copper/Zinc/Coal, BMA Coal, Hancock Prospecting, ElectraNet, Crossland Resources, and CBH Group.

As a company with many years of experience in applying evolutionary algorithms, we recognize the need for a better theoretical foundations. Your proposed work plan on building a stronger theory for evolutionary algorithms and evolution in general is important needed, and we share your vision that SAGE may increase the effectiveness and useability of evolutionary algorithms. We expect that companies like ours will benefit from the results of this project as we may integrate the findings into our software.

Recognising the fundamental importance of this research, we are keen to support your project. SolveIT Software is committed to

- provide guidance on real-world applications on evolutionary algorithms,
- discuss how your research findings can be integrated into our algorithms, thus not only benefitting our company, but also all our customers,
- participate in key meetings (if possible),
- nominate senior scientist as contact person.

We look forward to working with you.

Yours faithfully,

Zbigniew Michalewicz
Chief Scientist
SolveIT Software



divis GmbH, Joseph-v.-Fraunhofer-Str. 20, 44227 Dortmund

Prof. Dr. Tobias Friedrich
Friedrich-Schiller-Universität Jena
Fakultät für Mathematik und Informatik
Lehrstuhl für Theoretische Informatik I
Ernst-Abbe-Platz 2
07743 Jena
Germany

Fax

Email

baeck@divis-gmbh.de

Ihr Zeichen, Ihre Nachricht vom

Unser Zeichen, unsere Nachricht vom

TB

Telefon, Name
0231 - 9700341

Thomas Bäck

Datum

12.04.2013

Dear Professor Friedrich,

Thank you very much for contacting us regarding the EU project "Speed of Adaptation in Population Genetics and Evolutionary Computation (SAGE)". We are very pleased to support your research program and wish to express our full support for the grant proposal to the European Commission.

divis intelligent solution GmbH provides state-of-the-art technologies for the analysis and optimization of products and processes. We apply nonlinear data analysis and global optimization techniques based on evolutionary algorithms and other heuristics in order to help our clients to automatically generate models and knowledge from their data. We are working for global companies in Europe, North America and Australia for more than 12 years. These companies come from a wide range of industries such as automotive industry, cosmetic and detergent industry, and food and beverage industry; including for example Air Liquide, BMW, Boehringer Ingelheim, Daimler, Ford, Henkel, Honda, Unilever and Volkswagen.

As we are using evolutionary algorithms for difficult optimization problems, we are very interested in new developments and fundamental research on evolutionary processes. We could highly benefit from new results on designing good crossover operators as well as guidance for automatic parameter tuning of our optimizers.

We look forward to continue working with your team in the future.

A handwritten signature in black ink, appearing to read "T. Bäck", is written over a light blue horizontal line.

Yours sincerely,
Prof. Dr. Thomas Bäck, CEO



Dear Dr. Paixão

Thank you inviting us to be involved with your research project proposal "Speed of Adaptation in Population Genetics and Evolutionary Computation (SAGE)". We would be very happy to support your research program.

EXERGY Studios is a SME registered in Slovakia and operating in both in Slovakia and Austria. EXERGY specializes in building energy certification and optimization. We use an integrated approach to building and systems design by combining expertise in construction physics and heating, cooling and ventilation systems and we rely heavily on thermodynamic simulations of building design.

We have developed in-house software to provide optimization solutions for building design to our clients. This software is mainly based on meta-heuristic optimization methods, and depends on Evolutionary and Genetic Algorithms. As such, we fully recognize the need for a better theoretical understanding of these tools and are confident that the results of this project will be crucial in achieving precisely that.

We would be happy to stay in contact with this project and provide guidance on real-world problems we are faced with, and to discuss potential implementations of the results coming out of SAGE in our internal software methods.

We look forward to working with you and best regards,

Marcus Jones
M.Sc., LEED AP
Director, EXERGY Studios s.r.o.
+43 (0)664 104 5356



EventMAP Building,
33 Richmond Park
Stranmillis
Belfast

Vat number: 798 8124 66

14 April 2013

Re: Letter of support for FET Open proposal SAGE

Per Kristian
School of Computer Science
University of Nottingham
Jubilee Campus
Nottingham

Dear Per Kristian,

Letter of support for FET Open proposal SAGE

We have read with great interest your proposal on "Speed of Adaptation in Population Genetics and Evolutionary Computation" and would like to express our strongest support.

Many large educational institutions suffer significant losses every year from poor and cost-ineffective planning of resources and > events. EventMAP is the leading provider of software and consultancy > services for resource planning and timetabling in the university sector. Educational institutions worldwide have used our solutions to reduce their costs in timetabling and exam scheduling, including University of Helsinki, the Swiss Education Group, and the University of Sussex,

The EventMAP software suite is based on state of the art optimisation algorithms which has been developed in close collaboration with researchers from the ASAP research group at the University of Nottingham, and researchers from Queen's University of Belfast. Our

company has benefited significantly from its close ties with academia, and we are very keen to continue this collaboration. Where possible, we will make use of the results and insights generated from this project to improve our software suite.

We are therefore pleased to commit the following support to the SAGE project.

I will personally work with you on the project allowing direct access to descriptions of the typical problem characteristics in real-world timetabling problems. I commit to attending attend meeting and workshops when possible

Best regards,

Yours Sincerely

A handwritten signature in black ink, appearing to read 'Barry McCollum'.

Dr Barry McCollum
Managing Director
EventMAP Limited.



Norges teknisk-naturvitenskapelige universitet
Fakultet for informasjonsteknologi, matematikk og elektroteknikk

tildeler

Per Kristian Lehre

født 19. januar 1977

graden

Philosophiae doctor

i Datateknikk og informasjonsvitenskap

Graden er oppnådd 28. august 2006

Tittel på avhandlingen:

Complexity and Geometry in Artificial Development

Trondheim, 11. januar 2007

A handwritten signature in blue ink, reading 'Per Kristian Lehre', written over a horizontal line.

Dekanus

En komplett oversikt over hvilke emner dette studiet inneholder og hvilke karakterer kandidaten har fått, fins i karakterutskriften som er utstedt sammen med dette vitnemålet.



CERTIDÃO

Grau de Doutor

ANTÓNIO PEREIRA DE BASTOS, Chefe de Divisão do Serviço Académico da Secretaria Geral da Universidade do Porto: Certifico em face dos dados constantes do respectivo processo, que

TIAGO MATIAS MACHADO DOS SANTOS SEARA PAIXÃO

titular do bilhete de identidade nº 11299728, de nacionalidade portuguesa, concluiu, no Instituto Ciências Biomédicas de Abel Salazar, desta Universidade, no dia um de Outubro de dois mil e sete, o doutoramento no ramo de conhecimento em Ciências Biomédicas, tendo ficado aprovado. _____

A presente certidão vai firmada com o selo branco desta Universidade.

Porto e Secretaria Geral da Universidade, 01 de Abril de 2008

O Chefe de Divisão

António Pereira de Bastos

Emol. € 15,00

Emit. / DC

Conf. *Am25*

UNIVERSITÄT DES SAARLANDES

Naturwissenschaftlich-Technische Fakultät I

Mathematik und Informatik

U R K U N D E

DIE FAKULTÄT PROMOVIERT

Herrn Tobias FRIEDRICH

geboren am 28. Juli 1980 in Neustrelitz

zum

DOKTOR DER NATURWISSENSCHAFTEN

(doctor rerum naturalium; Dr. rer. nat.)

NACHDEM ER IM ORDENTLICHEN
PROMOTIONSVERFAHREN DURCH DIE DISSERTATION

USE AND AVOIDANCE OF RANDOMNESS

BEGUTACHTET VON

Herrn Priv.-Doz. Dr. Benjamin Doerr

Herrn Professor Dr. Kurt Mehlhorn

Herrn Professor Dr. Ulrich Meyer

SOWIE DURCH DAS ABSCHLIESSENDE KOLLOQUIUM AM 13. DEZEMBER 2007
SEINE WISSENSCHAFTLICHE QUALIFIKATION NACHGEWIESEN HAT

MIT DER GESAMTNOTE

AUSGEZEICHNET

(SUMMA CUM LAUDE)


Saarbrücken, den 13. Dezember 2007

Der Universitätspräsident



Professor Dr. Volker LINNEWEBER

Der Dekan



Professor Dr.-Ing. Thorsten HERFET

Die

Technische Universität Dortmund

verleiht den Grad

**Doktor der Naturwissenschaften
(Dr. rer. nat.)**

an Herrn

Dirk Sudholt

geboren am 12. Oktober 1979 in Stadtlohn

nachdem er in ordnungsgemäßem Promotionsverfahren an der
Fakultät für Informatik durch seine Dissertation

**Computational Complexity of Evolutionary
Algorithms, Hybridizations, and Swarm
Intelligence**

sowie durch die mündliche Prüfung seine wissenschaftliche
Befähigung erwiesen und dabei das Gesamturteil

ausgezeichnet

erhalten hat.

Dortmund, den 15. Dezember 2008

