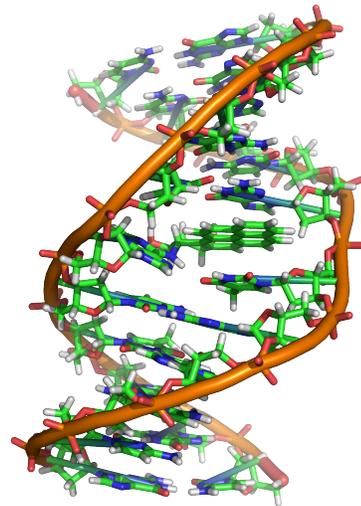


Theory of Crossover-Based Optimization

Motivation An important component of most large software systems is the optimization of individual parts and their interconnection. For the task of solving hard optimization problems, Genetic Algorithms (GAs) have been used successfully for more than 50 years, tackling an ever growing range of problems. In fact, some researchers believe that GAs "are the next step forward from deep learning: the form of AI that can **think outside the box**. And it is this kind of **creativity** that we need to advance AI beyond its current achievements." [1, highlights added]

Classical approaches towards optimization, for example (integer) linear programming, require the objective function (or some sufficient approximation) to be explicitly available. However, for complex optimization problems, the (possibly high-dimensional) objective function is typically not given as a mathematical function, but merely accessible indirectly as a "black box" which can be queried for the quality of a proposed solution. The first, simple approaches to black-box optimization are concerned with **local search**: iteratively improve a solution by making local (small) changes. Building on these initial ideas and on inspiration from nature (in particular evolution), modern black-box optimization algorithms transcend the abilities of mere local search by employing a wealth of different strategies. In various applications, these nature-inspired algorithms have proven their worth: successful examples include the optimization of antennas for NASA, for the turbine geometry of the Boeing 777 GE and the first clinically approved anti-viral drug for HIV.

One key way to bypass the drawbacks of local search algorithms (which get stuck in local optima) is to work with a diverse set of solutions and combining these to derive entirely novel candidates for approaching the global optimum. Combining multiple solutions is called a **crossover** operation (just as in biology where this refers to the combination of two parent genomes, but with the option of having more than two parents). The use of good crossover operations is typically essential for the practical success of Genetic Algorithms.



Purpose of the project With this project we want to introduce you to [all parts of research](#) in theoretical computer science. We start with a literature search, followed by creative white-board work and verification of ideas with formal proofs, concluded by the actual writing of a paper. As for the content, we want to formally analyze the effect of crossover operations in black-box search. We want to consider functions

$$f : \{0, 1\}^n \rightarrow \mathbb{R}, x \mapsto \sum_{i=1}^n w_i x_i,$$

where $(w_i)_{i \leq n}$ is a sequence of weights. Such functions f are called linear pseudo-Boolean functions, and their optimization is a good testbed for black-box search. If we impose a [cardinality constraint](#) (i.e., for some $k < n$, only those bit strings with at most k many 1s are acceptable solutions), the problem quickly approaches the well-known NP-hard knapsack problem. However, with this simple cardinality constraint good solutions can still be found in time about $\Theta(n^2)$ by simple local search-like genetic algorithms [2]. We will introduce to you advanced tools from probability theory (including drift theory) which make such optimization processes amenable for formal analysis and which can be used to turn expected progress per expectation into expected optimization times. We believe that with these tools a tremendous speedup to $O(n \log n)$ can be shown when employing good crossover operations with appropriate diversity mechanisms (see [3]).

What we expect from you You should bring the [curiosity](#) and willingness to delve into an interesting research topic within Theoretical Computer Science. Our main goal is a [mathematical understanding](#) of crossover operations and we expect you to contribute theoretical results. While the ability to reason formally is expected, no particular mathematical knowledge is expected.

What you can expect from us We provide material that will gently introduce you to the field (including some basics of probability) and accompany you all along this interesting journey. This will be a [team effort](#), and we aim at publishing our results at a renowned international conference.

How to contact us Feel free to send your questions to us via e-mail:
Timo.Koetzing@hpi.de

“The chicken is only an egg’s way
for making another egg.”



– Richard Dawkins

[1] <https://venturebeat.com/2018/05/17/evolutionary-computation-will-drive-the-future-of-creative-ai/>

[2] DOI: [10.1016/j.tcs.2018.04.051](https://doi.org/10.1016/j.tcs.2018.04.051)

[3] DOI: [10.1145/2908812.2908956](https://doi.org/10.1145/2908812.2908956)