

# Solving combinatorial problems with SAT

**Doctoral Thesis**

**Author(s):**

Chebiryak, Yury

**Publication date:**

2012

**Permanent link:**

<https://doi.org/10.3929/ethz-a-007608469>

**Rights / license:**

In Copyright - Non-Commercial Use Permitted

DISS. ETH Nr. 20737

# Solving Combinatorial Problems with SAT

ABHANDLUNG  
zur Erlangung des Titels

DOKTOR DER WISSENSCHAFTEN

der

ETH ZÜRICH

vorgelegt von

Yury Chebiryak

M. Sc. Universität des Saarlandes, Saarbrücken

geboren am 24. November 1981

von Wladiwostok, Russland

Angenommen auf Antrag von

Prof. Dr. Jürg Gutknecht

Prof. Dr. Ivo Sbalzarini

Prof. Dr. Daniel Kroening

Dr. Jürg Nievergelt

2012

## Abstract

Combinatorics is a branch of pure mathematics concerning the study of discrete (and usually finite) objects. Recently, D. E. Knuth applied a specialized data structure, Binary Decision Diagrams (BDD), to solve some problems of enumerative combinatorics [85]. This thesis presents how certain kinds of such objects, combinatorial Gray codes, can be efficiently obtained using a reduction to a propositional satisfiability formula. Solving such a formula with state-of-the-art tools can be more efficient than competitive algorithms (such as genetic algorithms that were used in the last decade to find longest snake-in-the-box codes).

Genetic Algorithms are the state-of-the-art for finding snake-in-the-box and cyclic codes (coils). For finding circuit codes (a generalized variation of snake-in-the-box), the hand-crafted constructions were leading the competition. Our SAT-based approach did not terminate for the record lengths of snakes (within a timeout of 24 hours), but for the circuit codes we report 19 new codes.

Both SAT- and BDD-based algorithms have proved their efficiency in the field of Software Model Checking. When coping with combinatorial problems, these techniques have similar advantages over, for instance, genetic algorithms. In particular, both methods return a definite answer to the question of whether a specified combinatorial object exists upon termination and allow for enumeration.

The combinatorial Gray codes that we have been able to find using SAT solvers are used in fields of information theory and computational biology. In the latter case, additional linear constraints should be added to guarantee certain stability properties of a biological system. We can incorporate these constraints in a Satisfiability Modulo Theory formula and solve it using a corresponding SMT-solver. An alternative (and more efficient) solution uses a SAT solver to find Gray codes of interest and then checks stability properties.

We have obtained new combinatorial codes that are of interest in information theory: a 9-bead binary necklace containing 60 code words, 19 circuit codes, and two distance preserving codes. Four of the new circuit codes have been shown to be optimal: the UNSAT results for larger lengths indicate that these bounds are tight. We enumerated Gray codes with respect to the distribution of weights for the hypercubes of dimension up to 6.

This thesis provides details about optimizations, both in propositional SAT encoding and combinatorial aspects, which enabled these results.

## Zusammenfassung

Die Kombinatorik ist ein Gebiet der reinen Mathematik, welches sich mit diskreten und im Normalfall endlichen Strukturen befasst. Vor kurzem verwendete D. E. Knuth eine spezielle Datenstruktur - so genannte binäre Entscheidungsbäume (engl. BDD) - um Probleme der abzählbaren Kombinatorik zu lösen. In dieser Arbeit wird beschrieben, wie eine spezielle Untergruppe dieser Strukturen, nämlich kombinatorische Gray-Codes, effizient berechnet werden können. Dies geschieht durch eine Reduktion auf eine aussagenlogische Formel. Die Lösung dieser Formeln mit der Hilfe von Entscheidungsprozeduren für das aussagenlogische Erfüllbarkeitsproblem (SAT) ist oftmals effizienter als die Standard-Algorithmen zur Bestimmung von Gray-Codes (wie z.B. genetische Algorithmen, die im letzten Jahrzehnt die Berechnung von Snake-in-the-Box Codes dominiert haben).

Sowohl auf SAT als auch auf BDD aufbauende Algorithmen haben ihre Effizienz im Bereich Software Model Checking bewiesen. Bei kombinatorischen Problemen haben diese Algorithmen ähnliche Vorteile gegenüber anderen Methoden, wie z.B. gegenüber genetischen Algorithmen. Besonders hervorzuheben ist die Eigenschaft, dass beide Methoden nach Programmende die Frage, ob eine bestimmte kombinatorische Struktur existiert oder nicht, definitiv beantworten können. Ausserdem erlauben sie es, alle Lösungen aufzuzahlen.

Die kombinatorischen Gray-Codes, die wir mittels SAT-Solvern erhalten haben, werden in der Informationstheorie und der computergestützten Biologie eingesetzt. Im zweiten Fall müssen noch zusätzliche, lineare Randbedingungen eingeführt werden, um bestimmte Stabilitätseigenschaften zu garantieren. Wir können SAT-Formeln konstruieren, die zu Lösungen führen, die diese Randbedingungen erfüllen und die sich dann mit Hilfe von entsprechenden SAT-Solvern lösen lassen. Eine alternative und effizientere Methode besteht darin, zuerst Gray-Codes mittels SAT-Solvern zu bestimmen und erst dann die Stabilitätskriterien zu prüfen. Wir haben folgende neue kombinatorische Codes erzeugt, die für die Informationstheorie von Interesse sind: binäre Perlenketten der Länge 9 (engl. 9-bead binary necklace), die 60 Codewörter beinhalten, 19 circuit Codes und zwei distanzerhaltende Codes. Weiterhin haben wir die Gray-Codes entsprechend ihrer Gewichtverteilung für Hyperwürfel bis zu der Dimension 6 klassifiziert.

Diese Doktorarbeit beschreibt die Details des Optimierungsvorganges, sowohl für das aussagenlogische Erfüllbarkeitsproblem als auch die kombinatorischen Aspekte, welche diese Ergebnisse ermöglicht haben.