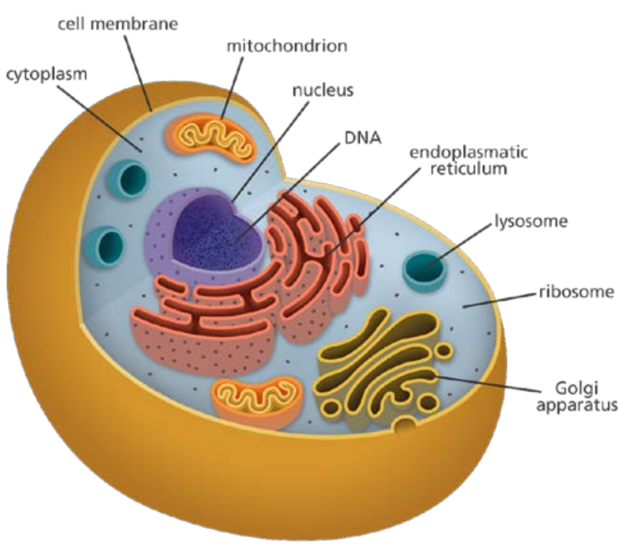
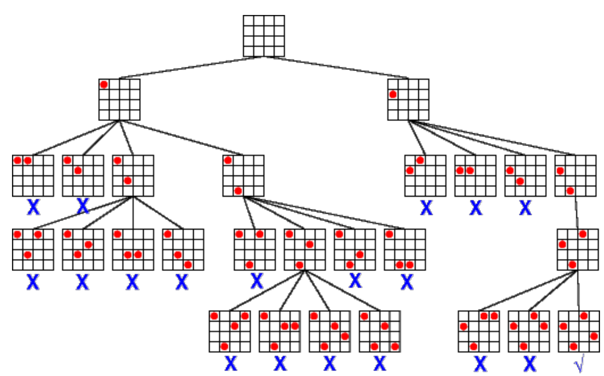
MDSC: Modèles Discrets pour les Systèmes Complexes

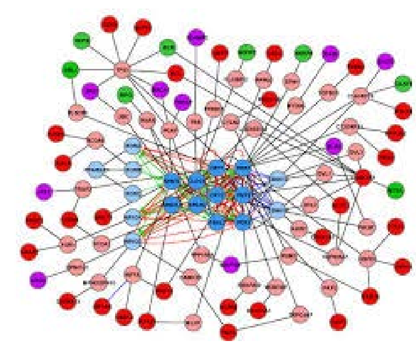
Jean-Charles Régin

Complex Systems

Simple objects that interact each other and produce a macroscopic **complex behavior**







Impacts

2012 Turing Scholarship (Simon Martiel)

2013 Research Excellence Award (J.-C. Régin)

2017 Research Excellence Award (E. Formenti)

2013 AAAI (Jean-Charles Régin), classic paper

2013 CSBio (Jean-Paul Comet)

2015 MCU (Simon Martiel)

CP 2011

AUTOMATA 2013

UCNC 2013

CPAIOR 2014

CMSB 2015 et 2016

Eric Goles (Chili)

- Articles
- Projets communs
- Echange PhD
- Séjours invités
- Organisation de Conf. Int.

Laurent Michel (USA)

- Organisation de CI
- Séjours invités
- Intégration de logiciels

Bernard P. Zeigler (USA)

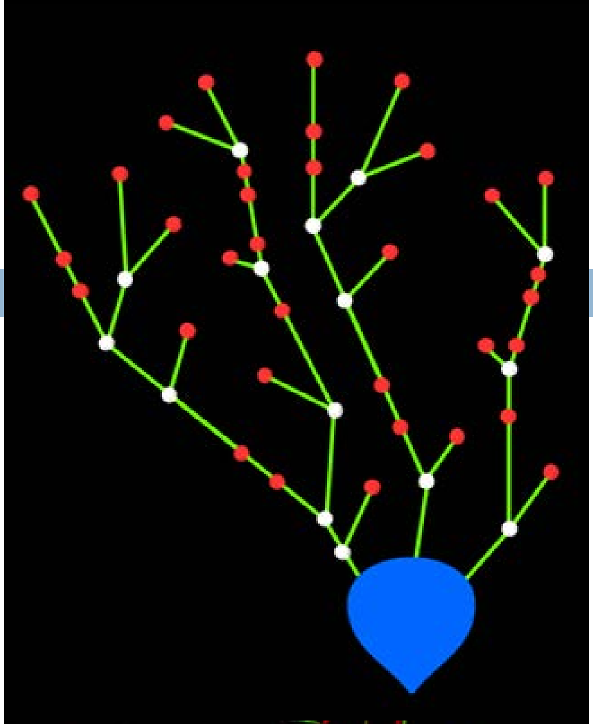
- Articles
- Projets communs
- Echange PhD
- Séjours invités
- Organisation de Confs/Workshops

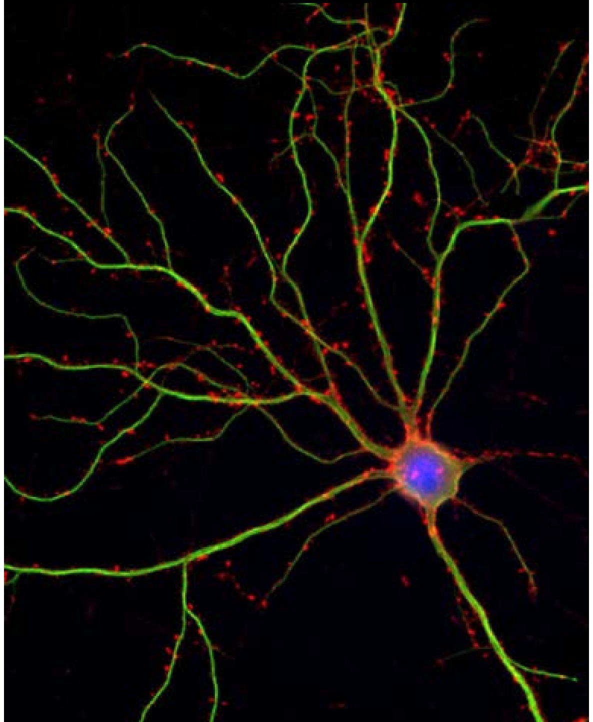
Petr Kurka (Rép. Tchèque)

- Articles
- Projets communs
- Echange PhD
- Séjours invités

CMB

- Neurocognition++
- Toxicology
- Genes network: Hoare logic and parametrisation



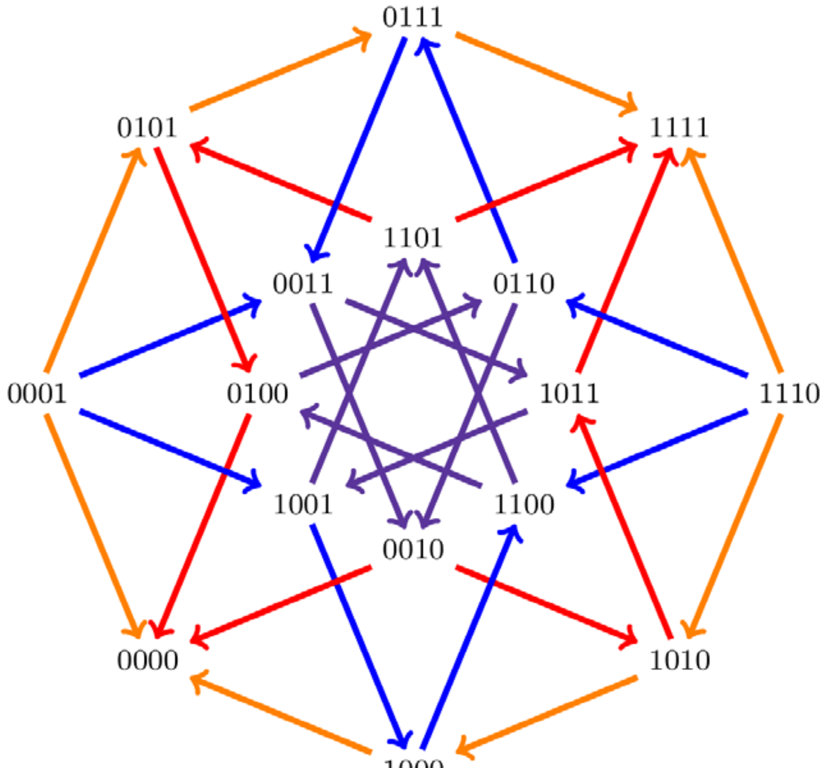


CMB

- Application of formal methods to biology, more specifically to biology of systems (circadian cyle, breathing,…)
- Modelization and simulation of biological neurons and neurosciences in general
- Recent result:
  - proof of some conjectures between the dynamical behavior of gene networks and the positive or negative cycles in their interaction graph

MCS

- Finite vs. Infinite
- Asynchronism
- Finite Discrete Dynamic System
- Formal Languages



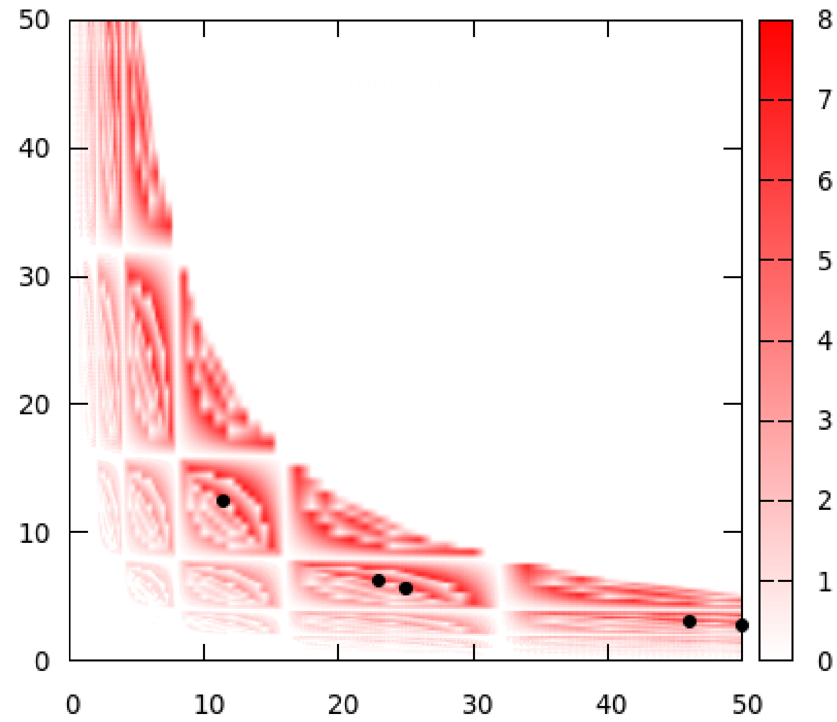
MCS

- Modeling of complex systems coming from physic, biology, chemistry, etc.
- Use of discrete dynamic systems as a main tool for analyzing and modeling problems

- Recent result:
  - proof that the distance between two genes is related to the strength of their interaction (we have also proved a new theorem in the context of stochastic process which is of some interest in its own) - this work was put in the highlights of CNRS
  - method for transporting dynamical properties of symbolic dynamical systems in high dimensions into lower dimensions in order to exploit known low dimension results (this work was in the highlights for the 40 years of TCS international journal)

CSOPV

- Path Problems
- Multi-Valued Decision Diagrams
- Embarrasingly Parallel Search
- Abstract Interpretation and Constraint Programming
  Floatting point operations



Combinatorial Optimization Problems

- Ubiquitous in real world
  - Rostering, resource allocation, scheduling, production planning, automatic text generation…
- Some advances of computers are translating into computation time improvements
  - Increase of CPUs frequency $\longrightarrow$ accelerate the resolution
  - Increase of the number of cores $\longrightarrow$ we split the problems into parts and solve them in // $\longrightarrow$ accelerate the resolution
- What's about memory?

Memory

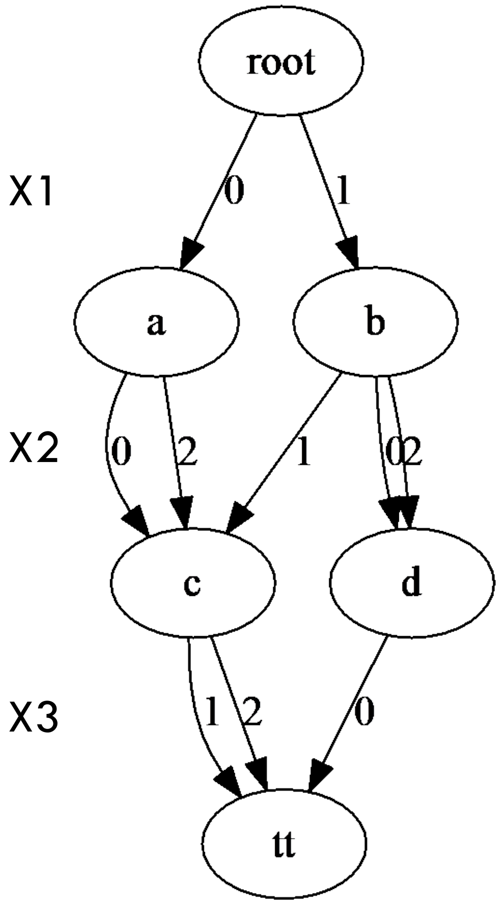
- Until today: memory was a constraint
  - In 1996: Windows 3.11: 4Mo
  - This is no more true. We have a lot of available GB
  - With NVMe SSD disk accessible like RAM (factor 3 slower)
    - equivalent of Terabytes of RAM for 500€
- Why memory is not seen as a chance?
- How to improve the solving time if you have more memory?

Multivalued Decision Diagram

- BDD generalization

$$f : \{0 \ldots d-1\}^r \rightarrow \{true, false\}$$

- Each **level** represents a **variable**
- Each **path** from node ***root*** to node ***tt*** represents a **valid assignment**
- The MDD models **all the tuples** satisfying the constraint
- Compression may gain an **exponential factor**



MDD: creation

- MDD **can be created without enumerating the solution set** (e.g. from dynamic programming like for the sum)
- MDD **can be combined without decompressing them**

Compression gain

- Compression may gain an exponential factor
- It often does!
- Example (see later)
  - MDD requiring 600,000 edges for representing 10^90 solutions, that is a compression factor of 10^86.
- Sometimes it can be subtle

MaxOrder

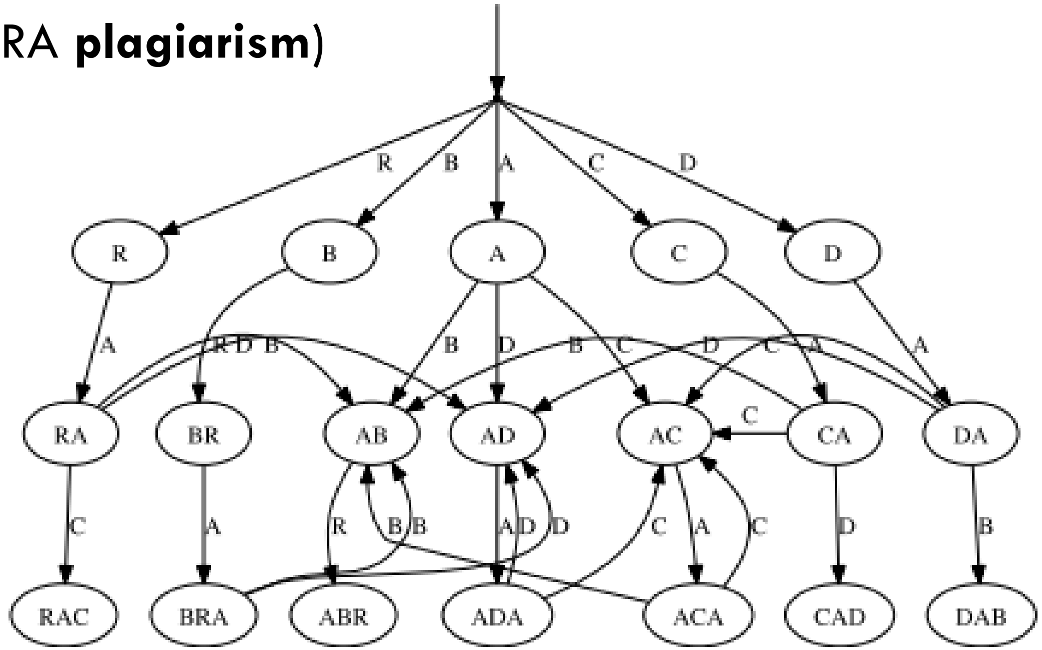
- Problem:
  - A. Papadopoulos & al (AAAI-14)
  - Avoiding plagiarism in Markov sequence generation

- Example of goal:
  - Based on a corpus
    - Books, music…
  - Generate sequences of 20 words
    - sequence generation
  - **All** the subsequences of **size 2 belong** to the corpus
    - **Markov** sequence generation
  - **None** of the subsequences of **size 4 belong** to the corpus
    - **Avoiding plagiarism**

MaxOrder

- Letter succession « ABRACADABRA »
- plagiarism = 4 letters
  - ACACABR is ok
  - ABACA is not ok (BA **doesn't exist**)
  - DABRACA is not ok (ABRA **plagiarism**)

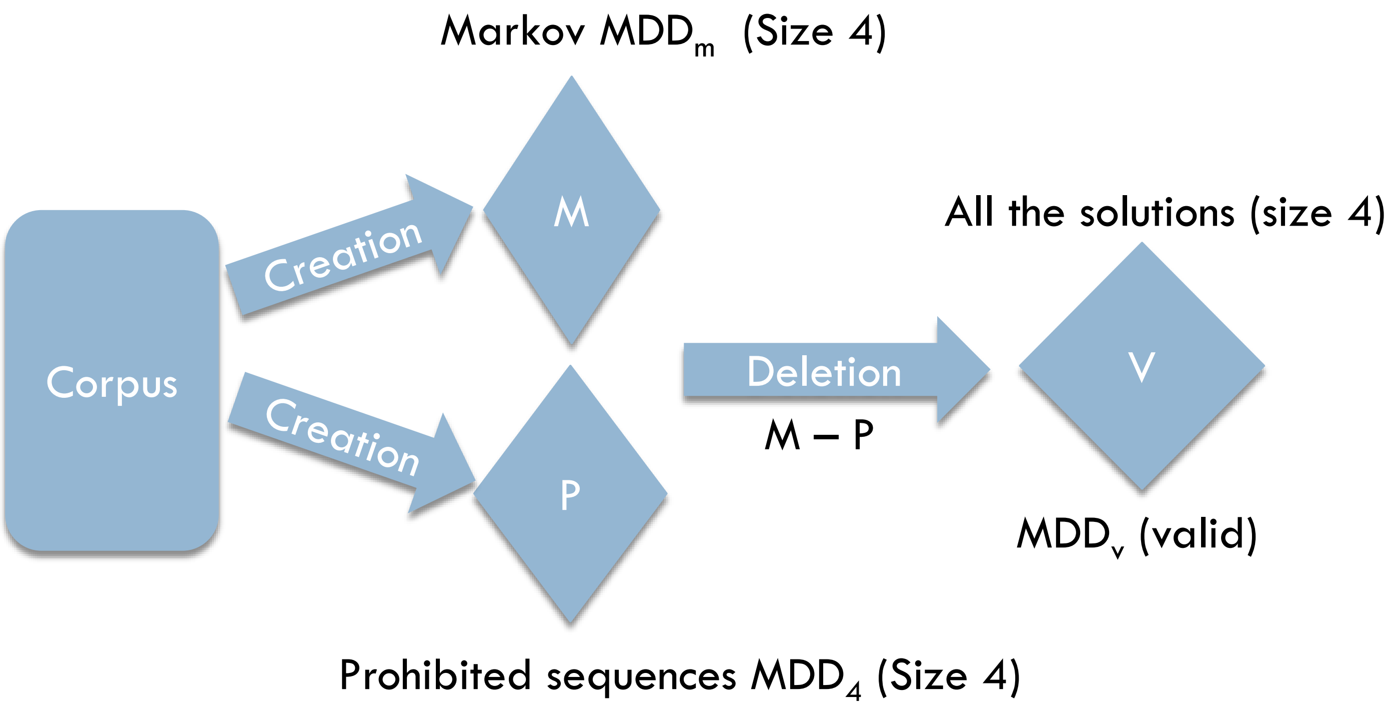
- *ad-hoc* Algorithm
  - Automata
  - Regular constraint



MaxOrder

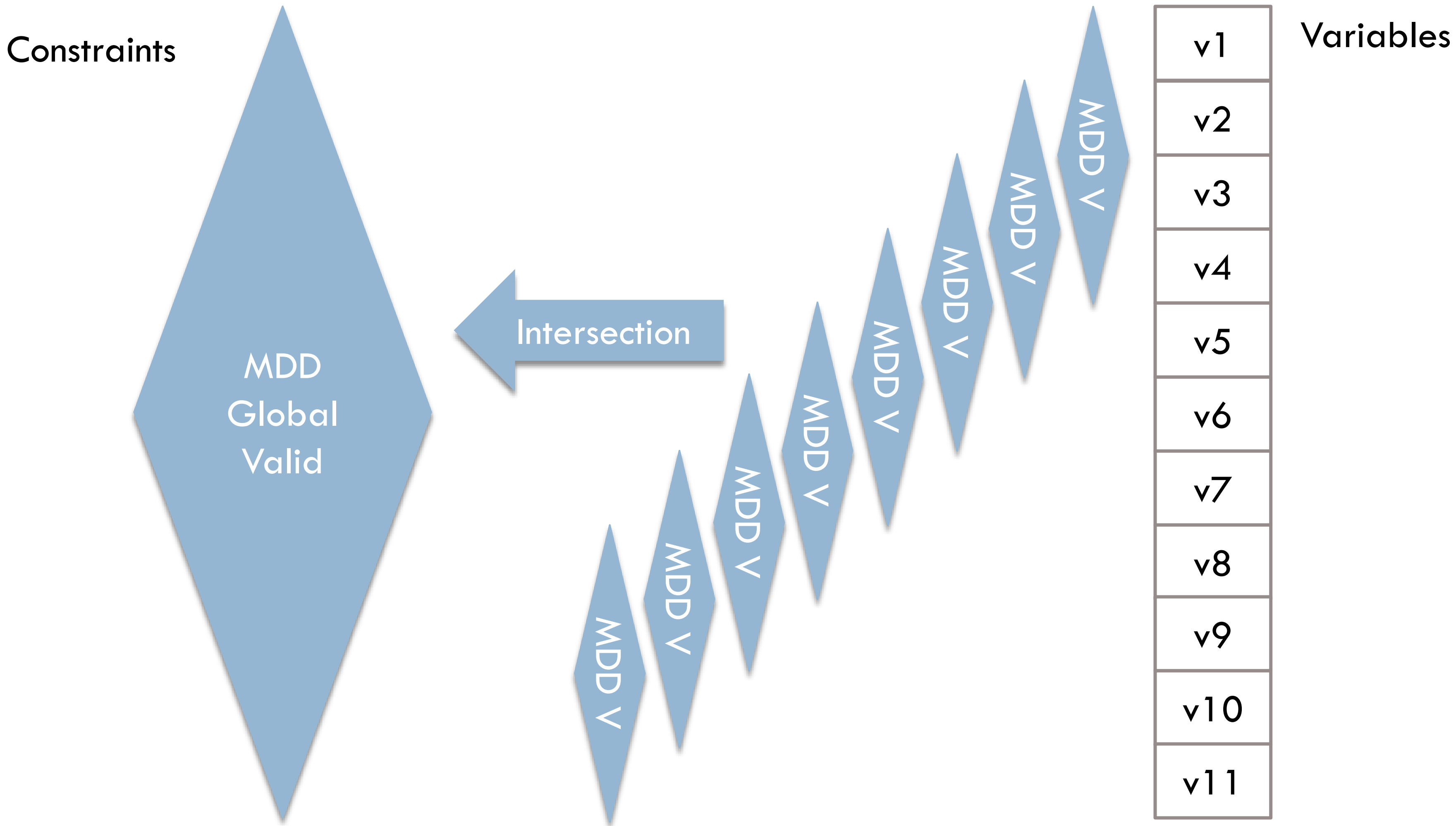
- Definitions
  - Mdd_4 : All the sequences of length 4 from the corpus
    - Plagiarism sequences
  - Mdd_m : All the Markov sequences of length 4
    - Each subsequences of size 2 belong to the corpus
    - Should not be represented by a table
  - Mdd_a : Defined by $Mdd_m - Mdd_4$
    - All the sequences of size 4 allowed
  - Mdd_r
    - Definition of 16 $Mdd_{s[i]} = Mdd_a$
    - $Mdd_{s[i]}$ is define for the variables $\{x_i, x_{i+1}, x_{i+2}, x_{i+3}\}$
    - Mdd_r defined by the successive intersection of the $Mdd_{s[i]}$
    - Represents all the solutions

MaxOrder

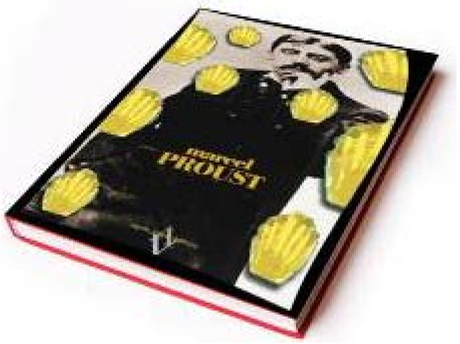


Remark: P ⊆ M

MaxOrder

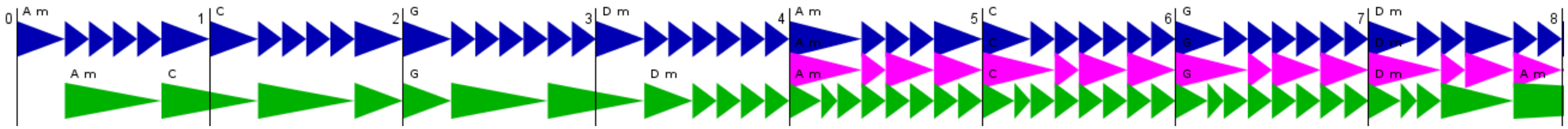

Constraints
MDD
Global
Valid
Intersection
MDD V
MDD V
MDD V
MDD V
MDD V
MDD V
MDD V
MDD V
v1
v2
v3
v4
v5
v6
v7
v8
v9
v10
v11
Variables

MaxOrder



- **Proust**
  - "À la recherche du temps perdu"
- Size
  - Mdd_4 : 56 225 nodes & 127 786 edges
    - Reduction: 123,025 nodes to 56,225
  - Mdd_m : 15 950 nodes & 129 465 edges
  - Mdd_r : 1 208 219 nodes & 188 035 203 edges
- #solutions 2.2*10^35 tuples
- **All the solutions** of the problem
- **Competitive** with the ad hoc algorithm!

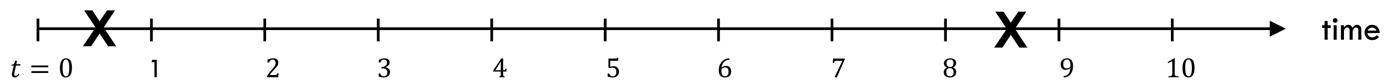
Audio Multi-Tracks



- Audio multi-track synchronization problem
- Concatenate audio chunks to make tracks
- Synchronize several tracks
  - Synchronization points = DATES
  - Specific constraints at sync. points
- See videos

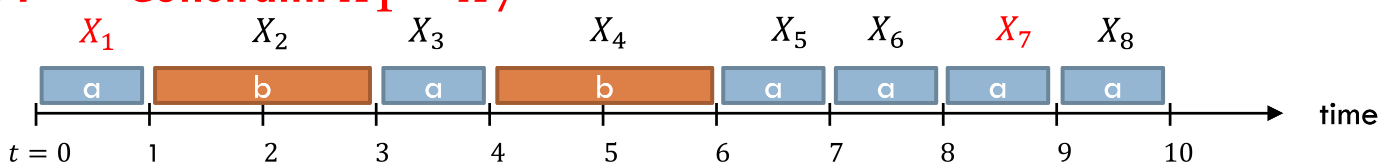
Relations between Temporal Positions

Constraint: Same item at $t = 0$ and at $t = 8$

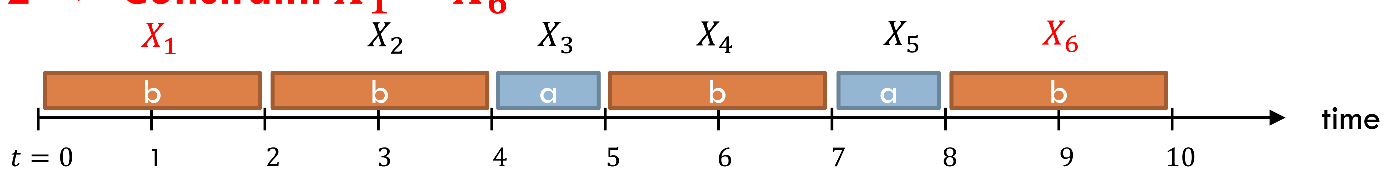


HOW CAN WE DEFINE SUCH A CONSTRAINT?

Sol. #1 => Constraint $X_1 = X_7$



Sol. #2 => Constraint $X_1 = X_6$



Sol #1 and sol #2 correspond to different CSPs

MDD Compression

- Time complexity: linear in the size of the MDD
- Compression:
  - 111 variables
  - 12726 noeuds, 231957 edges, 12726 instance
  - 305282122988337896711788854927396642159525357491709373229145150273524687 14593 solutions
  - **2.10^5 edges vs 3.10^73 solutions**

Generation of synchronized audio multi-tracks

Fails to solve small problems

2 synchronization points: 8.4 seconds (40 variables globally)

3 sync. points: timeout (+1800 seconds) (60 variables globally)

n	MDD size (#Vertices, #Edges)						Time (ms)
	Guitar		Bass		Drum		
6	2382	41k	848	13667	1864	73k	2301
8	4199	74k	1493	24k	3817	156k	7219
10	6530	117k	2388	39k	6513	275k	23k
12	9374	169k	3623	61k	9957	429k	57k
14	12k	231k	5085	87k	14k	617k	112k

Solves 14 sync. points in less than 2 minutes => 14 bars, roughly 30 seconds of music

Simple model fails to solve 3 sync. points in 30 minutes

Future

- Continue to work on MDD
  - Relaxation
  - Consider more problems (multi dimensional bin packing, path, knowledge compilation, robust optimization…)
- Invent new data structures
  - Compressed data structure
  - Integerable into a solver
  - Manipulable
  - Non determinism for having a better compression

Travels of my last phD Student

- At least one week
- **France**: Lille, Lyon (int'l conf CP-14) ,Paris (Sony, Google), Toulouse (int'l conf, CP16), Iles Porquerolles (Summer School 17)
- **Belgium**: Louvain-la-Neuve
- **Canada**: Baanf (Calgary; int'l conf CPAIOR-16)
- **USA**: San Francisco (int'l conf AAAI-16)
- **Argentina**: Buenos Aires (int'l conf, IJCAI-15)
- **Australia**: Melbourne (int'l conf, CP-17)
- **Ireland**: Cork (Summer School, 16)

- Currently in postdoc at Cornell Univ. (USA)