

# Scalable Coupling of Deep Learning with Logical Reasoning

Accepted at IJCAI2023  
PFIA'23

Marianne Defresne (TBI, MIAT), Thomas Schiex (MIAT)  
and Sophie Barbe (TBI)  
marianne.defresne@insa-toulouse.fr

July, 5th 2023

**ANITI**

ARTIFICIAL & NATURAL INTELLIGENCE  
TOULOUSE INSTITUTE



**RÉPUBLIQUE  
FRANÇAISE**

*Liberté  
Égalité  
Fraternité*

**INRAE**

**INSA**

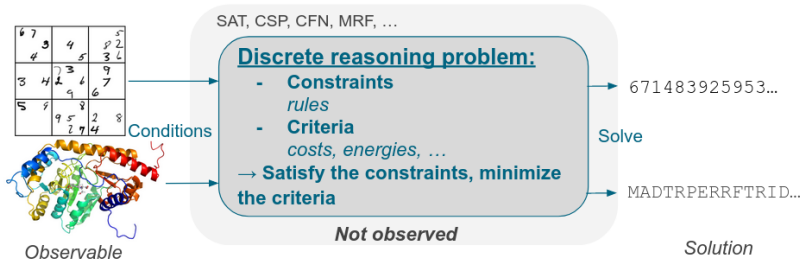
TOULOUSE

**MA**  
TOULOUSE

**tbi**  
Toulouse Biotechnology Institute  
Bio & Chemical Engineering

**BioEco**  
Circular Economy  
Governmental Building Bio-Based Economy

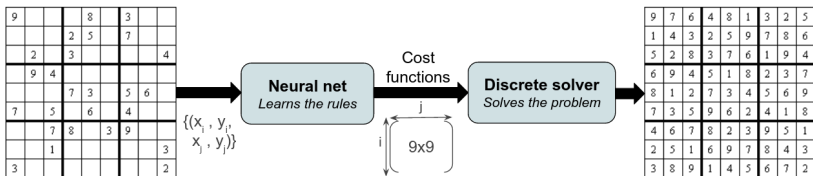
# General context: learning how to reason



- | **Goal:** solve new natural-input instances without access to the discrete model parameters
  - > Learn to predict the underlying constraints & criteria
  - > Decision-focused learning
- | **How?** By interfacing two branches of AI:
  - > Deep Learning (DL)
  - > Discrete reasoning (Weighted Constraint Satisfaction Problem, WCSP)

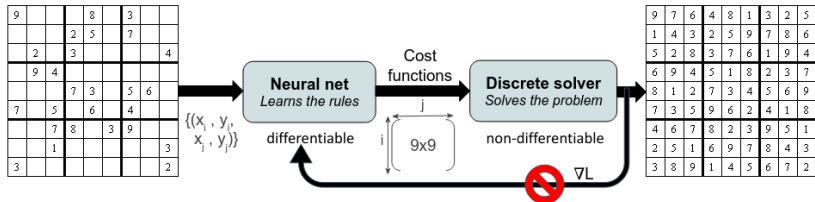
# Zoom on the Sudoku toy problem *(Brouard, Givry, and Schiex 2020)*

- | Aim: learning a representation of the Sudoku rules
  - > Data: (initial grid, solved grid)
  - > Rules (cost functions) are unknown



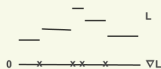
# Zoom on the Sudoku toy problem (Brouard, Givry, and Schiex 2020)

- Aim: learning a representation of the Sudoku rules
  - > Data: (initial grid, solved grid)
  - > Rules (cost functions) are unknown



## Aiming to minimize the decision error

$$L = \text{Hamming}(y; \hat{y}) = \frac{1}{81} \sum_{i=1}^8 1[y_i \neq \hat{y}_i]$$



- > Difficulty: **discrete objective** vs **gradient descent**
  - o  $rL$  is either 0 or non-existent

# The solver embedded as a neural layer

## Extracting meaningful gradients

- > Differentiable & informative upper bound of  $L$ : Hinge loss (*Altun, McAllester, and Belkin 2005*)
- > Continuous interpolation of  $L$ : Blackbox (*Pogancic et al. 2019*)
- | **Exact solving** during inference
- | **Training cost**: each instance is a NP-hard problem

## Continuous relaxation of the problem

- > SATNet (*Wang et al. 2019*)
- | **Tractable** differentiable optimization layer
- | **Approximate** solving

# Illustration on the Sudoku

Approach	Characteristic	Acc.	Grids	Training set
RRN	Pure DL	96.6%	Hard	180,000
SATNet	Relaxation	99.8%	Easy	9,000
Hinge	Extract gradients	<b>100%</b>	Hard	1,000

Recurrent Relational Net (*Palm, Paquet, and Winther 2018*)

# Illustration on the Sudoku

Approach	Characteristic	Acc.	Grids	Training set
RRN	Pure DL	96.6%	Hard	180,000
SATNet	Relaxation	99.8%	Easy	9,000
Hinge	Extract gradients	<b>100%</b>	Hard	1,000

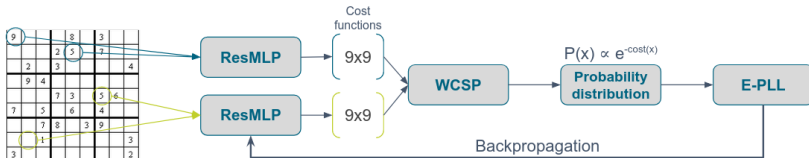
Recurrent Relational Net (*Palm, Paquet, and Winther 2018*)

## Training with the Hinge loss

- |  $rL(\hat{y}, y)$  (*Sahoo et al. 2023*)
  - >  $\hat{y}$  solution of the predicted discrete problem
- | Tuning the solver is challenging
  - > L1 regularization on costs
  - > Random initialization ! random discrete problems
    - | Easier problem (20 variables to predict) on first epochs
- | 2-3 days of training ! intractable on bigger instances

# 2-stage approach: learning before optimizing

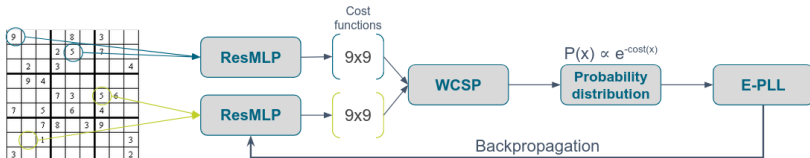
- | How to assess the learned discrete problem without solving it?
  - > **Pseudo-log likelihood** (Besag 1975):  $\prod_i \log P(y_i | y_{-i})$





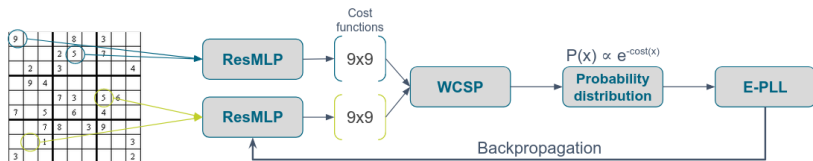
# 2-stage approach: learning before optimizing

- | How to assess the learned discrete problem without solving it?
  - > **Pseudo-log likelihood** (Besag 1975):  $\sum_i \log P(y_i | y_{-i})$



- | Fails on 100% of test grids
  - > Interpreting the learned model: partial constraints are learned

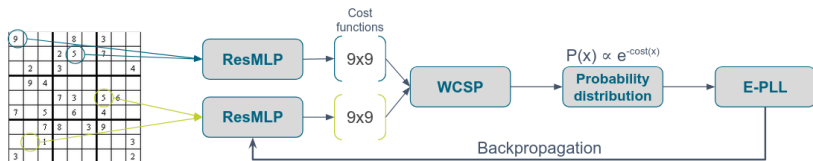
# 2-stage approach with the Emmental-PLL (E-PLL)



- PLL enhanced to learn constraint (Defresne, Barbe, and Schiex 2023)<sup>1</sup>
- > **E-PLL**:  $\prod_i \log P(y_i | y_{i \in M(i)})$

<sup>1</sup>Marianne Defresne, Sophie Barbe, and Thomas Schiex (2023). "Scalable Coupling of Deep Learning with Logical Reasoning". In: *Thirty-second International Joint Conference on Artificial Intelligence, IJCAI'2023*.

# 2-stage approach with the Emmental-PLL (E-PLL)



- PLL enhanced to learn constraint (Defresne, Barbe, and Schiex 2023)<sup>1</sup>
- > **E-PLL**:  $\sum_i \log P(y_i | y_{i \in M(i)})$

Approach	Acc.	Train set	Train time	Redundant constraints
Embedded solver	100%	1000	2-3 days	no
E-PLL	100%	200	15 min	yes

- Restricted usage: solver after neural layers & fixed loss

<sup>1</sup>Marianne Defresne, Sophie Barbe, and Thomas Schiex (2023). "Scalable Coupling of Deep Learning with Logical Reasoning". In: *Thirty-second International Joint Conference on Artificial Intelligence, IJCAI'2023*.

- | **Dataset:** Sudoku grids with multiple solutions
  - > Incomplete information: at most 5 solutions are observed
- | **Task:** Predicting one of the solutions
- | Pure DL: which loss?

Approach	SelectR <sup>1</sup>	E-PLL
Accuracy	86.7%	<b>100%</b>

- | With the E-PLL, the correct rules are learned
  - > All the solutions can be enumerated by the solver

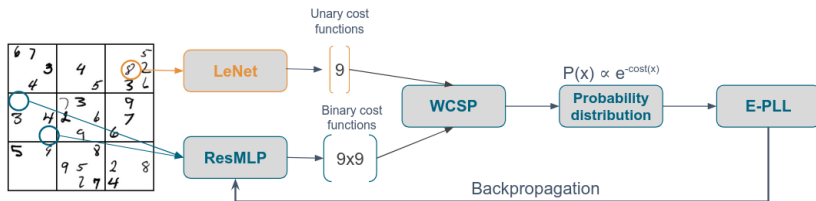
---

<sup>2</sup>Yatin Nandwani et al. (2021). "Neural Learning of One-of-Many Solutions for Combinatorial Problems in Structured Output Spaces". In: *International Conference on Learning Representations, ICLR'21*. url: <https://openreview.net/forum?id=ATp1nW2FuZL>.

# Application on natural-input problems

## Visual sudoku

> Learn to play Sudoku & to recognize digit



| Solver able to correct digit mis-classification

SATNet	Theoretical (no corrections)	Ours
63.2 %	74.2%	<b>94.1</b> 0.8%

## Learning the laws of protein design

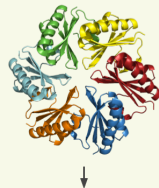
Cost function = pairwise interaction score (*Traore et al. 2013*)

> Main changes:

- Train set up to 10,000 variables, variable size
- Energy conditioned by the input structure
- One-of-Many solution setting

> Intractable inference ! use an approximate solver (*Durante, Katsirelos, and Schiex 2022*)

> Outperforms existing decomposable score functions



SSNAIGLIETKGYVAA...

	Rosetta <sup>1</sup>	Our
Similarity (↑)	17.9%	<b>27.8%</b>

<sup>1</sup>Park et al. 2016

## Hybridizing automated reasoning and ML vs. pure DL

- > Data-efficiency
- > Interpretability
- > *A posteriori* control (adding constraints or criteria)

## Perspective: protein design

- > Scalable method required
- > Applied projects

# Acknowledgment




- | The **organizers** of PFIA'23
- | My **PhD supervisors**: Sophie Barbe and Thomas Schiex
- | My teams in **TBI** and **MIAT**
- | **CALMIP** & **IDRIS** & **GenoToul** for computational resources

**Thanks for your attention!**








This work has been supported by the Agence Nationale de la Recherche (ANR) [grant ANR-18-EURE-0021]



-  Altun, Yasemin, David McAllester, and Mikhail Belkin (2005). “Maximum margin semi-supervised learning for structured variables”. In: *Advances in neural information processing systems* 18.
-  Besag, Julian (1975). “Statistical Analysis of Non-Lattice Data”. In: *Journal of the Royal Statistical Society: Series D (The Statistician)* 24.3, pp. 179–195.
-  Brouard, Céline, Simon de Givry, and Thomas Schiex (2020). “Pushing Data into CP Models Using Graphical Model Learning and Solving”. In: *International Conference on Principles and Practice of Constraint Programming*. Springer, pp. 811–827.

-  Defresne, Marianne, Sophie Barbe, and Thomas Schiex (2023). “Scalable Coupling of Deep Learning with Logical Reasoning”. In: *Thirty-second International Joint Conference on Artificial Intelligence, IJCAI'2023*.
-  Durante, Valentin, George Katsirelos, and Thomas Schiex (July 2022). “Efficient low rank convex bounds for pairwise discrete Graphical Models”. In: *Thirty-ninth International Conference on Machine Learning*.
-  Nandwani, Yatin et al. (2021). “Neural Learning of One-of-Many Solutions for Combinatorial Problems in Structured Output Spaces”. In: *International Conference on Learning Representations, ICLR'21*. url: <https://openreview.net/forum?id=ATp1nW2FuZL>.

-  Palm, Rasmus, Ulrich Paquet, and Ole Winther (2018). “Recurrent Relational Networks”. In: *Advances in Neural Information Processing Systems*. Ed. by S. Bengio et al. Vol. 31. Curran Associates, Inc.
-  Park, Hahnbeom et al. (2016). “Simultaneous Optimization of Biomolecular Energy Functions on Features from Small Molecules and Macromolecules”. In: *Journal of Chemical Theory and Computation* 12.12, pp. 6201–6212.
-  Pogančić, Marin Vlastelica et al. (2019). “Differentiation of blackbox combinatorial solvers”. In: *International Conference on Learning Representations*.

-  Sahoo, Subham Sekhar et al. (2023). “Backpropagation through Combinatorial Algorithms: Identity with Projection Works”. In: *Proc. of ICLR'23*. url: <https://arxiv.org/abs/2205.15213>.
-  Traoré, Seydou et al. (July 2013). “A new framework for computational protein design through cost function network optimization”. In: *Bioinformatics* 29.17, pp. 2129–2136. issn: 1367-4803.
-  Wang, Po-Wei et al. (2019). “SATNet: Bridging deep learning and logical reasoning using a differentiable satisfiability solver”. In: *Proceedings of the 36th International Conference on Machine Learning*. Vol. 97. Proceedings of Machine Learning Research. PMLR, pp. 6545–6554.