

Projet ANR-18-CE40-0011

PING/ACK

Appel à projets générique 2018

A	IDENTIFICATION	2
B	RESUME CONSOLIDE PUBLIC	2
	B.1 Résumé consolidé public en français	2
	B.2 Résumé consolidé public en anglais	4
C	MEMOIRE SCIENTIFIQUE	6
	C.1 Résumé du mémoire	6
	C.2 Enjeux et problématique, état de l'art.....	6
	C.3 Approche scientifique et technique	6
	C.4 Résultats obtenus	6
	C.5 Exploitation des résultats.....	10
	C.6 Discussion	10
	C.7 Conclusions.....	10
	C.8 Références.....	10
D	LISTE DES LIVRABLES	11
E	IMPACT DU PROJET	11
	E.1 Indicateurs d'impact.....	11
	E.2 Liste des publications et communications	11
	E.3 Liste des éléments de valorisation	14
	E.4 Bilan et suivi des personnels recrutés en CDD (hors stagiaires)	16

A IDENTIFICATION

Acronyme du projet	PING/ACK
Titre du projet	<i>Preprocessing Information for Nontrivial Goals / Advanced Compilation of Knowledge</i>
Coordinateur du projet (société/organisme)	Pierre Marquis Univ. Artois, CNRS, CRIL
Période du projet (date de début – date de fin)	20 mars 2019 - 20 septembre 2023 (compte-tenu des 6 mois de prolongation attribués à cause de la pandémie COVID-19)
Site web du projet, le cas échéant	http://www.cril.univ-artois.fr/pingack/

Rédacteur de ce rapport	
Civilité, prénom, nom	Pierre Marquis
Téléphone	03 21 79 17 86
Adresse électronique	marquis@cril.univ-artois.fr
Date de rédaction	Août 2023

Si différent du rédacteur, indiquer un contact pour le projet	
Civilité, prénom, nom	
Téléphone	
Adresse électronique	

Liste des partenaires présents à la fin du projet (société/organisme et responsable scientifique)	CRIL GREYC IRIT LaBRI
---	--------------------------------

B RESUME CONSOLIDE PUBLIC

B.1 RESUME CONSOLIDE PUBLIC EN FRANÇAIS

Compilation avancée de connaissances : nouveaux résultats, nouvelles applications.

L'objectif général du projet PING/ACK fut le développement d'avancées scientifiques de divers types dans le domaine de la compilation de connaissances.

La compilation de connaissances est une approche pour la résolution de problèmes s'appuyant à la fois sur des informations (ou « connaissances ») ne variant pas ou variant peu et d'informations plus volatiles. Il s'agit typiquement de résoudre plusieurs instances (Σ, α_i) du problème considéré, partageant les informations qui ne varient pas. Le principe est de pré-traiter (« compiler ») une fois seulement ces informations Σ , lors d'une phase de calcul hors

ligne et d'exploiter une forme compilée de ces informations lorsque les instances entières (Σ , α_i) du problème sont connues de façon à les résoudre plus efficacement (en terme de temps de calcul) lors d'une phase en ligne. Des bénéfices sont possibles quand le temps de compilation (celui consommé lors de la phase hors ligne) reste « suffisamment restreint », ce qui demande en particulier que la taille de forme compilée produite $\text{comp}(\Sigma)$ soit « suffisamment petite ». Toutefois, cela n'est pas toujours possible.

Dans le projet PING/ACK, nous nous sommes efforcés d'élargir la portée scientifique de la compilation de connaissances, à la fois du côté théorique et du côté pratique, en mettant en avant des résultats utiles pour le raisonnement automatisé et plusieurs domaines d'application de celui-ci.

Les résultats obtenus se sont appuyés sur des éléments méthodologiques classiquement exploités en compilation de connaissances.

Ceux-ci font intervenir les notions de langages de formes compilées et d'efficacité spatiale et d'efficacité temporelle de tels langages. La résolution de problèmes à partir de formes compilées passe, en effet, par la composition de traitements élémentaires, incluant des requêtes (c'est-à-dire l'extraction d'informations à partir d'une forme compilée) et des transformations (la modification d'une forme compilée, par exemple pour y intégrer de l'information supplémentaire). L'efficacité temporelle d'un langage de formes compilées peut alors être définie comme l'ensemble des traitements élémentaires qui peuvent être réalisés efficacement (en temps polynomial) à partir de formes compilées (et éventuellement d'informations additionnelles). L'efficacité spatiale d'un langage de formes compilées précise la capacité de ce langage à encoder en utilisant une quantité d'espace restreinte une information représentée dans un autre langage. Le choix d'un langage de formes compilées adaptée à la résolution d'un problème passe par la comparaison des efficacités spatiales et temporelles des langages candidats, visant à dégager un meilleur compromis. L'évaluation multicritère de langages candidats s'appelle une carte de compilation.

Les principaux résultats obtenus dans le cadre du projet PING/ACK sont, d'une part, la réalisation de nouvelles cartes de compilation adaptées à des langages de représentation plus expressifs (ou plus succincts) que ceux considérés jusque-là. Ainsi, des langages de représentation de préférences, mais aussi des langages de représentation de plans d'action (politiques décisionnelles) ou encore des langages de représentations de modèles probabilistes ont-ils été considérés. D'autre part, des formes de compilation approchée ont été étudiées. Par ailleurs, nous avons montré que l'efficacité spatiale de certains langages est liée à la complexité des preuves que l'on peut produire dans ces langages. Enfin, du côté des applications, nous avons montré que la compilation de connaissances peut être utile au développement d'approches pour l'IA de confiance.

La production scientifique du projet PING/ACK a été soutenue. Les revues et conférences internationales du meilleur rang (A* ou A selon les classements ERA et CORE : <http://portal.core.edu.au/>) ont été ciblées. Un aspect marquant de cette production est la variété des sujets abordés mais aussi la nature de la production, qui va d'articles décrivant des résultats à visée théorique jusqu'à des développements logiciels (conception et évaluation de programmes utilisés pour produire les résultats empiriques fournis dans diverses

publications). On gardera en tête qu'en IA, la pratique de publication est de diffuser ses résultats plutôt dans des actes de conférence que dans des revues. Enfin, on notera les deux thèses financées dans le cadre du projet ont été soutenues dans les délais requis et ont été à l'origine de plusieurs publications.

$$\begin{array}{ccc} \Sigma & \longrightarrow & \text{comp}(\Sigma) \\ & & \downarrow \\ \alpha_i & \longrightarrow & f(\text{comp}(\Sigma), \alpha_i) \end{array}$$

La figure ci-dessus synthétise la résolution de problèmes s'appuyant sur la compilation de connaissances : il s'agit de calculer efficacement les solutions $f(\Sigma, \alpha_i)$ de plusieurs instances (Σ, α_i) d'un même problème, en remplaçant Σ par une forme compilée partagée entre les instances.

Le projet PING/ACK fut un projet de recherche fondamentale coordonné par le Centre de Recherche en Informatique de Lens (CRIL). Il a associé aussi les laboratoires GREYC, IRIT et LaBRI. Le projet a commencé en mars 2019 et a duré 54 mois. Il a bénéficié d'une aide ANR de 375 990,77 € pour un coût global de l'ordre de 1 550 640,65 €.

B.2 RESUME CONSOLIDE PUBLIC EN ANGLAIS

Advanced compilation of knowledge: new results, new applications.

The general objective of the PING/ACK project was the development of scientific advances of various types in the field of knowledge compilation.

Knowledge compilation is an approach to problem solving based both on information (or "knowledge") that does not vary (or varies little), and more volatile information. It typically involves solving several instances (Σ, α_i) of the same problem, sharing pieces of information Σ that do not vary. The principle is to pre-process ("compile") Σ once, during an off-line compilation phase and to use a compiled form of it when the entire instances (Σ, α_i) of the problem are known in order to solve them more efficiently (in terms of computation time) during an online phase. Benefits are possible when the compilation time (consumed during the offline phase) remains "small enough", which requires in particular that the size of the compiled form $\text{comp}(\Sigma)$ that is generated is "small enough". However, this is not always possible.

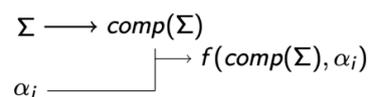
In the PING/ACK project, we have endeavored to broaden the scientific scope of knowledge compilation, both theoretically and practically, by highlighting useful results for automated reasoning and several areas of application thereof.

The results obtained were based on methodological elements traditionally used in knowledge compilation. These involve the notions of compilation languages and the spatial and temporal efficiency of such languages. Solving problems from compiled forms typically goes through the composition of elementary processing, including queries (i.e., the extraction of information from a compiled form) and transformations (a modification of a compiled form, for example to include additional information). The temporal efficiency of a compilation language can then

be defined as the set of elementary operations that can be performed efficiently (in polynomial time) from compiled forms (and possibly additional information). The spatial efficiency of a compilation language specifies the ability of this language to encode using a limited amount of space information represented in another language. The choice of a compilation language adapted to problem solving requires the comparison of the spatial and temporal efficiencies of the candidate languages, aiming to find a better compromise. The multi-criteria evaluation of candidate languages is called a compilation map.

The main results obtained within the framework of the PING/ACK project are, on the one hand, the creation of new compilation maps adapted to more expressive (or more succinct) representation languages than those considered so far. Thus, languages for representing preferences, but also languages for representing action plans (decision policies) and languages for representing probabilistic models have been considered. On the other hand, forms of approximate compilation have been studied. In addition, we have shown that the spatial efficiency of certain languages is linked to the complexity of the proofs that can be produced in these languages. Finally, from the application side, we have shown that knowledge compilation can be useful for the development of approaches for trustworthy AI.

The scientific production of the PING/ACK project was sustained. Best-ranked international journals and conferences (A* or A according to the ERA and CORE rankings: <http://portal.core.edu.au/>) were targeted. A striking aspect of this production is the variety of topics covered but also the nature of the production, which ranges from articles describing theoretical results to software developments (program design and evaluation, used to get the empirical results presented in many papers). It should be kept in mind that in AI, the practice of publication is to disseminate its results rather in conference proceedings than in journals. Finally, it should be noted that the two theses financed within the framework of the project were defended within the required deadlines and were the source of several publications.



The figure above summarizes the resolution of problems based on a knowledge compilation approach: the approach consists in computing the solutions $f(\Sigma, \alpha_i)$ of several instances (Σ, α_i) of the same problem, replacing Σ with a compiled form of it, shared between the instances.

The PING/ACK project was a fundamental research project coordinated by the Lens Computer Science Research Center (CRIL). It also associated the GREYC, IRIT and LaBRI laboratories. The project started in March 2019 and lasted 54 months. It received ANR funding of €375,990.77 for an overall cost of around €1,550,640.65.

C MEMOIRE SCIENTIFIQUE

Mémoire scientifique confidentiel : non

C.1 RESUME DU MEMOIRE

L'objectif général du projet PING/ACK fut le développement d'avancées scientifiques de divers types dans le domaine de la compilation de connaissances.

C.2 ENJEUX ET PROBLEMATIQUE, ETAT DE L'ART

Dans le projet PING/ACK, nous nous sommes efforcés d'élargir la portée scientifique de la compilation de connaissances, à la fois du côté théorique et du côté pratique, en mettant en avant des résultats utiles pour le raisonnement automatisé et plusieurs domaines d'application de celui-ci. Cela a nécessité, en particulier, l'introduction et l'étude de nouveaux langages de formes compilées, plus expressifs que ceux de la famille DNNF [Darwiche, 2001] considérée jusque-là, mais aussi de nouvelles requêtes adaptées aux domaines d'application visés.

C.3 APPROCHE SCIENTIFIQUE ET TECHNIQUE

Les résultats obtenus se sont appuyés sur des éléments méthodologiques classiquement exploités en compilation de connaissances. Ceux-ci ont fait intervenir les notions de langages de formes compilées et d'efficacité spatiale et d'efficacité temporelle de tels langages [Darwiche and Marquis, 2002].

C.4 RESULTATS OBTENUS

Les résultats obtenus dans le cadre du projet PING/ACK sont de nature variée. Les principaux résultats peuvent être répartis grossièrement en quatre catégories :

1. la réalisation de nouvelles cartes de compilation adaptées à des langages de représentation plus expressifs ou succincts que ceux considérés jusque-là.
2. la définition et l'analyse de formes de compilation approchée.
3. l'étude de liens entre compilation de connaissances et les systèmes de preuve.
4. l'utilisation de la compilation de connaissances pour le développement d'approches pour l'IA de confiance.

Pour des raisons d'espace, nous décrivons ici uniquement les principaux résultats obtenus.

1. Cartes de compilation

Carte de compilation pour les préférences. Les préférences conditionnelles ont été utilisées pour représenter de manière compacte des préférences sur des domaines combinatoires. Elles sont au cœur des CP-nets et de leurs généralisations, mais aussi des arbres de préférences lexicographiques. Plusieurs travaux ont abordé la complexité de certaines requêtes (optimisation, dominance notamment). Nous nous sommes attachés à étendre certains de ces résultats, et étudier d'autres requêtes et certaines transformations qui n'ont pas été abordées jusqu'à présent, comme l'équivalence, les coupures, les marginalisations, contribuant ainsi à une carte de compilation pour les langages de représentation des préférences à base de préférence conditionnelles. Un article décrivant les résultats obtenus a été accepté à la conférence internationale AAMAS'21.

Carte de compilation pour les langages d'action. Un travail sur l'efficacité temporelle et spatiale des requêtes et transformations liées à la représentation d'actions a été entrepris. Le travail réalisé s'est intéressé à plusieurs langages pour la représentation d'actions non-déterministes (le cas déterministe présentant peu de défis) : PDDL, dans une version non-déterministe et propositionnelle, une version non-déterministe de STRIPS conditionnel, les théories d'actions en NNF, et DL-PPA. Nous avons montré un paysage très riche en termes de concision relative des langages et de leur capacité à traiter les requêtes au cœur de la planification : déterminer si un état est successeur d'un autre, si une action est applicable dans un état, si une succession d'action assure la satisfaction d'une formule logique. Les résultats obtenus ont été publiés à la conférence internationale ICAPS'21.

Compilation pour la représentation de politiques d'actions pour la planification contingente. On s'intéresse à des domaines dans lesquels un agent a à sa disposition un ensemble d'actions pour réaliser un but, mais où l'observabilité est partielle : l'agent ne connaît pas exactement l'état de l'environnement à un instant donné, il a seulement des croyances sur cet état, résultant de croyances initiales et d'observations (imparfaites) apportées par ces actions. Il s'agit alors de calculer des politiques, qui sont des objets donnant l'action optimale à effectuer dans tout état de croyance atteignable ou, de façon équivalente, pour toute séquence d'observations effectuées. Nous avons étudié la représentation de politiques par des programmes à base de connaissances, dont nous avons montré qu'ils fournissent une représentation possiblement exponentiellement plus succincte que toute représentation standard, la contrepartie étant une difficulté algorithmique (modérément) plus importante pour retrouver les actions à effectuer au moment d'exécuter la politique. Ces résultats ont fait l'objet d'un article paru dans la revue internationale *Artificial Intelligence*.

Efficacité spatiale des circuits arithmétiques. Les circuits arithmétiques (AC) sont des circuits sur les nombres réels avec des variables d'entrée dans 0/1 et dont les portes calculent la somme ou le produit de leurs entrées. L'existence d'algorithmes efficaces pour de nombreuses opérations utiles au raisonnement probabiliste sur ces représentations requiert d'imposer des restrictions structurelles au circuit considéré. Dans ce travail, nous avons étudié l'efficacité spatiale de plusieurs classes de circuits arithmétiques obtenues en imposant différentes restrictions. Nous avons produit une carte de concision inconditionnelle pour les classes de circuits arithmétiques monotones - c'est-à-dire ceux dont les étiquettes constantes sont des réels non négatifs – respectant les restrictions considérées. Nous avons étendu une petite partie de cette carte aux classes de circuits arithmétiques positifs. Ceux-ci sont connus pour être généralement exponentiellement plus succincts que leurs homologues monotones, mais nous avons montré que pour les circuits dits déterministes, il n'y a pas de différence d'efficacité spatiale. Nous avons aussi mis en évidence des bornes inférieures exponentielles pour les circuits arithmétiques positifs structurellement décomposables. Les résultats obtenus ont été publiés à la conférence internationale KR'21.

Contraintes pseudo-booléennes et DNNF. Les contraintes pseudo-booléennes (PB) sont des inégalités ou équations linéaires sur les littéraux de variables booléennes régulièrement utilisées pour représenter des problèmes en IA. Nous avons prouvé l'existence de contraintes PB dont les représentations DNNF sont de taille exponentielle. Ce résultat renforce les études sur la séparation entre les systèmes de contraintes PB et les formes compilées usuelles. Une publication décrivant ce travail est parue dans les actes de la conférence internationale SAT'20.

2. Compilation approchée

Approximation de représentations par des d-DNNFs. Dans de nombreuses applications de la compilation de connaissances, la connaissance disponible constitue un modèle approché de la réalité. Ainsi, les modèles probabilistes - par exemple des réseaux bayésiens - construits par apprentissage sont nécessairement imparfaits. Dans ce contexte, on peut envisager de compiler des approximations des modèles pour obtenir des formes approchées plus compactes. Une forme approchée reste une approximation de la réalité et, moyennant la possibilité de contrôler l'erreur d'approximation, de précision comparable à celle de la forme exacte. Nous nous sommes concentrés sur les représentations d-DNNF, lesquelles sont d'intérêt pour les modèles probabilistes puisqu'elles permettent le comptage (potentiellement pondéré) de modèles, auquel l'inférence probabiliste se réduit. Nous avons raffiné la notion d'approximation jusqu'alors considérée dans la littérature, que nous appelons approximation faible et qui, parce qu'elle autorise l'approximation des fonctions de peu de modèles par la fonction nulle, est inadaptée à nos besoins. Nous avons défini une approximation forte qui s'affranchit de ce problème et offre des garanties vis-à-vis du comptage de modèles. Nous avons ensuite prouvé les limites des approximations faible et forte pour la compilation en d-DNNF. Nous avons d'abord exhibé des classes de fonctions dont les d-DNNF exactes et les d-DNNF de faibles approximations sont nécessairement de taille exponentielle. Puis nous avons mis en évidence des classes de fonctions, dont il existe des approximations faibles triviales et sans intérêt (typiquement par la fonction nulle) mais dont les fortes approximations n'ont que des d-DNNF de tailles exponentielles. Une publication décrivant ce travail est parue dans les actes de la conférence internationale IJCAI-PRICAI'20.

Compilation approchée. Nous avons étudié dans quelle mesure le pré-processeur B+E pour le comptage de modèles pouvait être utile à la compilation approchée de formules propositionnelles. Par compilation approchée, il faut entendre ici compilation restreinte à un sous-alphabet du vocabulaire propositionnel de la formule considérée. Il est attendu que la forme compilée approchée à calculer constitue une conséquence logique (*upper approximation*) de la formule de départ qui doit être logiquement équivalente à la projection de cette formule d'entrée sur le sous-alphabet considéré. Nous avons montré que le pré-processeur B+E peut être utilisé en amont d'un compilateur exact ou approché en préservant la condition énoncée, à condition de contraindre la bipartition des variables qu'il calcule à inclure toutes les variables d'intérêt dans l'ensemble des variables d'entrée. Nous avons réalisé des expérimentations pour évaluer les bénéfices pouvant être obtenus en utilisant B+E en amont des compilateurs exacts c2d et d4, disponibles en ligne. Grâce à ces expérimentations, nous avons montré qu'en pratique, les temps de compilation et les tailles des formes compilées pouvaient être réduits (parfois drastiquement) en utilisant B+E. Le travail réalisé a donné lieu à une publication dans la revue internationale *Artificial Intelligence* en 2020. Le calcul de bipartition α , de son côté, été amélioré récemment de façon très significative en pratique (publication à JELIA'23).

3. Compilation et systèmes de preuve

Réfutations et DNNF. L'étude de la complexité de preuves a pour objectif l'obtention de bornes inférieures sur des réfutations de formules propositionnelles dans des systèmes de preuves. Nous avons étudié les réfutations par résolution régulière pour des formules de Tseitin. Nos travaux ont abouti à une caractérisation des formules de Tseitin non satisfiables,

dont la réfutation par résolution régulière ne nécessite qu'un nombre polynomial de clauses. Notre contribution est la preuve que les plus petites réfutations de ce type pour les formules de Tseitin sont de longueur polynomiale si et seulement si la «*treewidth*» (ou largeur d'arbre) des graphes qui en décrivent la structure est en $O(\log(n))$, où n désigne le nombre de nœuds du graphe. L'originalité de ces travaux réside dans le lien qui y est fait entre la longueur de la plus petite réfutation régulière par résolution de formules de Tseitin non satisfiables et la taille des représentations DNNF de formules de Tseitin satisfiables travaillant sur les mêmes graphes. Nous montrons que si ces représentations DNNF sont de taille au moins exponentielle alors il est en est de même pour la longueur des réfutations étudiées, puis nous prouvons que les DNNF des formules de Tseitin sont effectivement exponentielles en la *treewidth* du graphe de la formule. Une publication décrivant ce travail a été acceptée à la conférence internationale SAT'21.

Définition et étude de systèmes de preuve pour les formules booléennes quantifiées (QBF).

Nous avons proposé et étudié des systèmes de preuve pour les formules booléennes quantifiées (QBF) opérant sur des diagrammes de décision binaire ordonnés (OBDD). Nous avons obtenu des séparations exponentielles des longueurs des preuves en comparaison aux systèmes de preuve standard à base de formules sous forme normale conjonctive (CNF), de type QU-Resolution ou IR-Calc. Nous avons également développé une technique de borne inférieure pour les systèmes de preuve QBF, basée sur l'extraction de stratégies, qui permet de généraliser les bornes connues, issues de la théorie de la complexité de communication. Cette technique nous permet de dériver de meilleures bornes inférieures pour les systèmes de preuve QBF, qui sont indépendantes de l'ordre des variables utilisé pour définir les diagrammes de décision et qui restent valides même quand le système de preuve est équipé d'un oracle NP. Une publication décrivant ce travail a été acceptée à la conférence internationale SAT'21.

4. Compilation de connaissances et IA de confiance

Compilation certifiée. L'utilisation croissante de la compilation de connaissances dans divers cadres (configuration de produits assemblés, provenance dans les bases de données, méthodes formelles, IA explicable, etc.) rend impérieux le développement d'outils permettant de vérifier les formes compilées produites. Aucun outil n'existe à ce jour pour réaliser de telles tâches de vérification efficacement (i.e., en temps polynomial) et vraisemblablement, un tel outil n'existera jamais. En effet, quand les contraintes sont exprimées en CNF, le test d'équivalence entre une représentation en d-SDNNF et une représentation en CNF est coNP-complet. Il importe donc de définir de nouveaux langages de compilation, qui bien sûr devront offrir les propriétés attendues des formes compilées (le test de cohérence, le conditionnement, etc.) et garantir en plus la CNF-certifiabilité. Nous nous sommes attachés à la réalisation de cet objectif et, à cet effet, introduit de nouveaux langages de compilation dans la famille DNNF, appelés langages des DNNF décorées. La décoration ajoutée joue le rôle d'un certificat permettant une vérification en temps polynomial. Nous avons montré comment transformer un compilateur DNNF top-down « classique » en compilateur top-down produisant des formes compilées décorées. Une publication décrivant ce travail a été acceptée à la conférence internationale AAAI'21.

Enumération d'impliquants premiers. Nous avons étudié le problème Enum-IP d'énumération des impliquants premiers de fonctions booléennes représentées par des circuits dec-DNNF. En particulier, nous nous sommes intéressés à l'énumération des impliquants premiers représentant des raisons suffisantes du classement d'instances, visant à expliquer les prédictions obtenues par des modèles d'apprentissage automatique. Nous avons montré que l'énumération des impliquants premiers correspondant aux raisons suffisantes d'une instance n'est vraisemblablement pas dans OutputP , c'est-à-dire réalisable en temps polynomial dans la taille de l'entrée plus la taille de la sortie, et cela même si le classeur considéré en entrée est une fonction booléenne représentée par un arbre de décision. Cela contraste avec la plupart des requêtes d'IA explicative, qui sont traitables pour les arbres de décision. Une publication décrivant ce travail a été acceptée à la conférence internationale IJCAI-PRICAI'22.

Traductions entre modèles d'apprentissage. Nous avons étudié l'efficacité spatiale de divers modèles d'apprentissage automatique, et plus précisément l'existence de traductions en temps polynomial ou en espace polynomial entre divers langages de représentation de classeurs. Les langages considérés incluent les arbres de décision, les forêts aléatoires, plusieurs types d'arbres boostés, les réseaux de neurones binaires, les perceptrons booléens multicouches et diverses représentations logiques de classeurs binaires. Nous avons obtenu une carte complète indiquant pour chaque couple de langages C, C' s'il existe ou non une traduction temps polynomial / espace polynomial de C vers C' . Nous avons montré également comment tirer parti de la carte résultante à des fins d'IA explicative. Une publication décrivant ce travail a été acceptée à la conférence internationale IJCAI'23.

C.5 EXPLOITATION DES RESULTATS

Les résultats obtenus ouvrent un très large ensemble de perspectives, à la fois théoriques et pratiques. Leur diffusion dans des supports de premier plan et la formation des deux doctorants recrutés dans le cadre du projet et de divers stagiaires contribueront à promouvoir la thématique dans les années à venir.

C.6 DISCUSSION

Les objectifs initiaux du projet ont été largement réalisés. Les principales modifications apportées ont concerné, du point de vue théorique, l'abandon de l'étude de la compilabilité paramétrée au profit de l'étude des liens entre formes compilées et systèmes de preuve, et du point de vue pratique, la focalisation sur les questions d'IA de confiance (sujet d'importance majeure depuis quelques années) en lieu et place des autres champs applicatifs qui avaient été ciblés au départ.

C.7 CONCLUSIONS

Le projet PING/ACK a permis d'étendre la portée scientifique de la compilation de connaissances, à la fois du côté théorique et du côté pratique, en identifiant de nouvelles questions de recherche et de nouveaux champs applicatifs où elle présente un intérêt certain. Le principal souci rencontré fut la difficulté de se réunir pour collaborer, à cause de la pandémie COVID-19. Pour autant, le projet a donné lieu à une production scientifique notable.

C.8 REFERENCES

A. Darwiche. Decomposable Negation Normal Form. *Journal of the ACM*, 48(4):608–647, 2001.

A. Darwiche and P. Marquis. A Knowledge Compilation Map. *Journal of Artificial Intelligence Research*, 17:229–264, 2002.

D LISTE DES LIVRABLES

Deux rapports ont été produits ; il s’agit de synthèses de travaux existants sur deux thèmes abordés dans PING/ACK :

- *KC maps of preference languages*
- *Parameterized algorithms and lower bounds for classical problems and well-established languages*

Un site web a été mis en place dès le début du projet. Le workshop final qui avait été prévu au départ n’a pas pu avoir lieu mais il a été remplacé par l’édition d’un numéro spécial de la revue internationale « *Annals of Mathematics and Artificial Intelligence* » dédié à la compilation de connaissances et dont un des éditeurs invités est le porteur du projet PING/ACK. Ce numéro spécial est en voie de finalisation. Les autres livrables prévus au départ ont été abandonnés.

E IMPACT DU PROJET

E.1 INDICATEURS D’IMPACT

Nombre de publications et de communications

		Publications multipartenaires	Publications monopartenaires
International	Revue à comité de lecture	1 revue	6 revues
	Ouvrages ou chapitres d’ouvrage		
	Communications (conférence)		17 conférences 1 atelier
France	Revue à comité de lecture		
	Ouvrages ou chapitres d’ouvrage		
	Communications (conférence)		4 conférences
Actions de diffusion	Articles vulgarisation		
	Conférences vulgarisation		
	Autres		2 thèses

E.2 LISTE DES PUBLICATIONS ET COMMUNICATIONS

Thèses

A. de Colnet

Hard Functions in Knowledge Compilation: from Lower Bounds to Applications

Thèse de doctorat d’université, université d’Artois, 2022.

S. Scheck

Knowledge compilation for nondeterministic action languages

Thèse de doctorat d'université, université de Caen Normandie, 2022.

Revue internationale

J.-M. Lagniez, E. Lonca, and P. Marquis.
Definability for Model Counting
Artificial Intelligence 281: 103229 (2020).

B. Zanuttini, J. Lang, A. Saffidine, and F. Schwarzentruher.
Knowledge-based programs as succinct policies for partially observable domains.
Artificial Intelligence 288: 103365 (2020).

S. Mengel.
No Efficient Disjunction or Conjunction of Switch-Lists
J. Satisf. Boolean Model. Comput. 13(1): 1-4 (2022).

A. de Colnet and S. Mengel.
Characterizing Tseitin-Formulas with Short Regular Resolution Refutations.
J. Artif. Intell. Res. 76: 265-286 (2023).

E. Grandjean, Th. Grente, and V. Terrier.
Inductive definitions in logic versus programs of real-time cellular automata
Theoretical Computer Science (2023, accepté sous révision mineure).

H. Fargier, S. Mengel, and J. Mengin.
An Extended Knowledge Compilation Map for Conditional Preference Statements-based and Generalized Additive Utilities-based Languages
Annals of Mathematics and Artificial Intelligence (2023, accepté sous révision mineure).

A. Niveau, H. Palacios, S. Scheck, and B. Zanuttini.
A Knowledge Compilation Perspective on Queries and Operations for Nondeterministic Planning
Annals of Mathematics and Artificial Intelligence (2023, accepté sous révision mineure).

Conférences internationales

E. Grandjean and Th. Grente.
Descriptive complexity for minimal time of cellular automata.
Proc. of LICS'19, 1-13.

A. de Colnet and S. Mengel.
Lower Bounds for Approximate Knowledge Compilation
Proc. of IJCAI-PRICAI'20, 1834-1840.

A. de Colnet.
A lower bound on DNNF encodings of pseudo-Boolean constraints
Proc. of SAT'20, 312-321.

- J. Clément and A. Genitrini.
Binary Decision Diagrams: from Tree Compaction to Sampling.
Proc. of LATIN'20, 571-583.
- H. Fargier and J. Mengin.
A knowledge compilation map for conditional preference statements-based languages
Proc. of AAMAS'21, 492-500.
- F. Capelli, J.-M. Lagniez, and P. Marquis.
Certifying Top-Down Decision-DNNF Compilers
Proc. of AAI'21, 6244-6253.
- S. Scheck, A. Niveau, and B. Zanuttini.
Knowledge Compilation for Nondeterministic Action Languages
Proc. of ICAPS'21, 308-316.
- A. de Colnet and S. Mengel.
Characterizing Tseitin-Formulas with Short Regular Resolution Refutations
Proc. of SAT'21, 116-133.
- A. de Colnet and S. Mengel.
A Compilation of Succinctness Results for Arithmetic Circuits
Proc. of KR'21, 205-215.
- S. Mengel and F. Slivovsky.
Proof Complexity of Symbolic QBF Reasoning
Proc. of SAT'21, 399-416.
- S. Gamblin, A. Niveau, and M. Bouzid.
A Symbolic Representation for Probabilistic Dynamic Epistemic Logic.
Proc. of AAMAS 2022, 445-453.
- A. de Colnet and S. Mengel.
Lower Bounds on Intermediate Results in Bottom-Up Knowledge Compilation
Proc. of AAI'22, 5564-5572.
- A. de Colnet and P. Marquis.
On the Complexity of Enumerating Prime Implicants from Decision-DNNF Circuits
Proc. of IJCAI-ECAI'22, 2583-2590.
- A. de Colnet and P. Marquis.
On Translations between ML Models for XAI Purposes
Proc. of IJCAI'23, 3158-3166.
- S. Mengel.
Bounds on BDD-Based Bucket Elimination

Proc. of SAT'23, 16:1-16:11.

J.-M. Lagniez and P. Marquis.

Boosting Definability Bipartition Computation using SAT Witnesses

Proc. of JELIA'23, to appear.

J. Clément and A. Genitrini.

An iterative approach for counting reduced ordered binary decision diagrams

Proc. of MFCS'23, to appear.

Ateliers internationaux

Th. Grente and E. Grandjean.

Conjunctive Grammars, Cellular Automata and Logic

Proc. of 27th IFIP WG 1.5 International Workshop on Cellular Automata and Discrete Complex Systems (AUTOMATA'21) 8:1-8:19.

Conférences nationales

S. Scheck, A. Niveau, et B. Zanuttini.

Knowledge Compilation for Action Languages

Actes des 15es Journées Francophones sur la Planification, la Décision et l'Apprentissage pour la conduite de systèmes (JFPDA'20).

H. Fargier et J. Mengin.

An incomplete knowledge compilation map for conditional preference statements-based languages

Actes des 14es Journées d'Intelligence Artificielle Fondamentale (JIAF'20).

S. Scheck, A. Niveau et B. Zanuttini.

Explicit Representations of Persistency for Propositional Action Theories.

Actes des 16es Journées sur la Planification, la Décision et l'Apprentissage pour la conduite de systèmes (JFPDA'21).

S. Scheck, A. Niveau et B. Zanuttini.

A KC Map for Variants of Nondeterministic PDDL

Actes des 16es Journées d'Intelligence Artificielle Fondamentale (JIAF'22).

E.3 LISTE DES ELEMENTS DE VALORISATION

Le développement du projet PING/ACK a conduit à l'émergence d'autres projets et d'actions de diffusion. Parmi celles-ci figure l'action KOCOON (*KnOwledge COmpilatiOn Network*) qui s'est inscrite dans le cadre du dispositif d'accompagnement de la recherche régionale « SACRe : Soutien à l'Animation de Collectifs de Recherche » de la Région Hauts-de-France.

KOCOON avait pour objectif principal de renforcer les coopérations existantes et de développer de nouvelles collaborations pour créer un réseau de recherche international pérenne sur la thématique de la compilation de connaissances. Le levier employé a été

l'organisation d'un séminaire de recherche international de quatre jours en 2019, à Arras. Il s'est tenu dans les locaux de la maison de la recherche de l'université d'Artois, du 16 au 19 décembre 2019. Le séminaire a regroupé 42 participants, issus de divers pays d'Europe (Allemagne, Autriche, Belgique, France, Portugal, République Tchèque, Royaume-Uni) et au-delà (Japon, USA). Le programme scientifique a été très riche et a inclus des présentations sur des sujets variés relevant de la compilation de connaissances et de ses applications et outils.

Par ailleurs, les nombreux échanges scientifiques réalisés avec Adnan Darwiche (UCLA) nous ont conforté dans l'idée de construire un laboratoire international associé, porté par le CRIL (université d'Artois / CNRS) et UCLA, et centré sur la compilation de connaissances. Ce type de structure (maintenant appelée « *International Research Project* ») est porté et financé par le CNRS pour une durée de cinq ans. Le projet de laboratoire international associé appelé MAKC (« *Modern Approaches to Knowledge Compilation* ») a reçu un accueil favorable du CNRS et a été créé le 1^{er} janvier 2020 pour cinq ans.

Enfin, le développement des recherches sur la compilation de connaissances impulsé dans PING/ACK ont poussé son porteur à proposer un projet de chaire ANR d'enseignement et de recherche. Ce projet, baptisé EXPEKCTATION pour « EXPLainable artificial intelligence: a KnowlEdge CompilaTion FoundATIOn » concerne le développement d'approches d'IA explicable pour un apprentissage automatique interprétable et robuste, utilisant des méthodes de raisonnement automatisé par contraintes, en particulier la compilation de connaissances. Il s'agit de définir des techniques de prétraitement capables d'associer un prédicteur de boîte noire à une boîte blanche de substitution, qui puissent être utilisée pour fournir diverses formes d'explication et pour répondre à des requêtes de vérification sur la boîte noire correspondante. L'objectif est d'obtenir des systèmes d'IA auxquels l'utilisateur peut faire confiance. Le projet EXPEKCTATION a été retenu par l'ANR. Il a débuté en septembre 2020 et se poursuit encore aujourd'hui.

E.4 BILAN ET SUIVI DES PERSONNELS RECRUTES EN CDD (HORS STAGIAIRES)

Identification				Avant le recrutement sur le projet			Recrutement sur le projet				Après le projet				
Nom et prénom	Sexe H/F	Adresse email (1)	Date des dernières nouvelles	Dernier diplôme obtenu au moment du recrutement	Lieu d'études (France, UE, hors UE)	Expérience prof. Antérieure, y compris post-docs (ans)	Partenaire ayant embauché la personne	Poste dans le projet (2)	Durée missions (mois) (3)	Date de fin de mission sur le projet	Devenir professionnel (4)	Type d'employeur (5)	Type d'emploi (6)	Lien au projet ANR (7)	Valorisation expérience (8)
DE COLNET Alexis	H	decolnet@cril.fr	août 2023	Diplôme d'ingénieur + master	France et Singapour	-	CRIL	doctorant	38	31/10/2022	post-doctorant étranger	université	chercheur	non	oui
SCHECK Sergej	H	Sergej.scheck@unicaen.fr	août 2023	Master	Allemagne	-	GREYC	doctorant	36	23/12/2022	recherche d'emploi (en Allemagne)				