numéro 4

1024

BULLETIN
de la société informatique
de France

octobre 2014

COMITÉ DE RÉDACTION

Contact : 1024@societe-informatique-de-france.fr

SOCIÉTÉ INFORMATIQUE DE FRANCE
Institut Henri Poincaré, 11 rue Pierre et Marie Curie, 75231 Paris Cedex 05

Prix public : 32 € (adhérents SIF : −30%)

Directeur de la publication : Colin de la Higuera ISSN : 2270-1419

Couverture : d'après une maquette réalisée par Lollygraph.com.

SOMMAIRE DU Nº 4

Le mot du président

Colin de la Higuera [1]

La recherche en informatique

Commençons par les bonnes nouvelles. Il y a quelques jours, le CNRS a décerné sa médaille d'or : « Alain Fuchs, président du CNRS, vient de l'annoncer : l'informaticien Gérard Berry est le lauréat 2014 de la plus prestigieuse récompense scientifique française. »

La SIF a tenu à se joindre à ceux qui ont félicité Gérard Berry pour son prix. Nous ajoutons qu'en plus de ses travaux, pour lesquels il a été justement récompensé, nous remercions Gérard pour le temps qu'il consacre à de nombreuses opérations que la SIF soutient. Il est membre assidu de notre Conseil scientifique et a accompagné différentes délégations de la SIF auprès de ministères et organismes pour présenter en particulier notre point de vue (partagé) concernant l'enseignement de la discipline.

La recherche, que Gérard représente si bien, est cependant en difficulté. De nombreux chercheurs, laboratoires, départements, organismes ont décidé de soutenir le mouvement *Sciences en marche* qui a voulu alerter, ce mois d'octobre, les pouvoirs publics. Comme d'autres sociétés savantes, la SIF s'est interrogée sur son rôle vis-à-vis d'un tel mouvement. Le Conseil d'administration de la SIF a décidé de soutenir *Sciences en marche*, les questions posées concernant la recherche en général s'appliquant tout à fait à l'informatique en particulier [2].

1. Président de la Société informatique de France, Professeur à l'Université de Nantes, Courriel : cdlh@univ-nantes.fr.

2. http://www.societe-informatique-de-france.fr/2014/09/la-sif-soutient-sciences-en-marche/

D'autres problèmes concernent la recherche en informatique. Son absence dans Horizon 2020 semble montrer que, contrairement à d'autres disciplines scientifiques, les efforts visant à la promouvoir n'ont pas porté leurs fruits. Si l'informatique « est partout », elle ne figure pas quelque part en particulier. La question de la représentation de la discipline au niveau européen est posée. Avec d'autres, en Europe, nous cherchons à y apporter une réponse.

Nos relations avec les autres sciences

L'informatique a également à construire sa place, entre et avec les autres sociétés savantes. Les rencontres avec les sociétés de mathématiques, en particulier, sont fréquentes. Nous traitons fructueusement des dossiers ensemble. Par exemple, un travail commun avec la SMF, la SMAI, la SFDF et *Femmes et Mathématiques* doit mener à la publication commune d'une brochure produite par l'ONISEP sur « les métiers des Mathématiques et de l'Informatique ». Organisé avec la SMAI et INRIA, le Forum des lauréats de prix en informatique et mathématiques appliquées verra, en décembre, des exposés des récipiendaires de différents prix en Mathématiques appliquées et Informatique. Une autre occasion de montrer les coopérations entre informatique et mathématiques.

C'est également pour le compte de la coopération que nous pouvons parler de la journée annuelle « recherche » de la SIF, dont le thème était, cette année, le lien entre formation et innovation dans le domaine informatique. Les excellentes relations avec l'association Pasc@line se sont traduites par une collaboration effective : l'organisation d'une journée commune [3]. Les échanges furent tout à fait passionnants et on notera une *montée en puissance* avec les participations (et exposés denses) de Guy Mamou-Mani, Président du Syntec Numérique et de notre *premier* député : Émeric Bréhier, auteur d'un rapport sur le doctorat.

Médiation scientifique

Au moment d'écrire ces lignes, comme de nombreux membres de la SIF, je suis fatigué mais heureux par le travail accompli ces derniers jours pour parler d'informatique pendant la Fête de la science. Toutes celles et tous ceux qui chaque année prennent un peu de temps pour expliquer aux jeunes (et aux moins jeunes !) ce qu'est l'informatique en apprennent beaucoup sur la discipline, sur la façon d'en parler.

En effet, pour que l'informatique soit pleinement reconnue comme science et technique il est essentiel que nous ne nous limitions pas à en faire, à chercher, à enseigner. Il faut absolument que les efforts en direction du public soient renforcés. Plus nous serons nombreux à le faire, meilleur sera le résultat.

3. http://www.societe-informatique-de-france.fr/recherche/les-journees-recherche/journee-recherche-de-la-sif-2014/

Un premier public que nous devons cibler est le public informatique : que nous ayons une culture commune, que nous connaissions notre histoire, que nous sachions pourquoi untel a été jugé mériter tel prix. C'est dans ce contexte que 1024 est un outil précieux. En quelques numéros, les échos que nous avons est que cette mission se remplit : qu'il soit lu par un étudiant en informatique, par un technicien, par un ingénieur, un enseignant-chercheur, un chercheur ou au sens large un informaticien, c'est tout à fait motivant pour l'équipe de 1024 d'entendre dire que le but fixé est atteint.

Un second public est celui beaucoup plus large qui est « confronté » à l'informatique. On lui dit que « c'est la faute de l'informatique ! », que ses enfants vont devoir l'étudier, que les algorithmes vont bientôt tout contrôler, que ça a disparu et été remplacé par le numérique ou le codage, que… Ce public est souvent curieux ; il a envie de comprendre, alors que personne ne lui a jamais expliqué. Le rôle de `binaire` est naturellement essentiel, mais là encore, les efforts vers le grand public doivent être beaucoup plus importants.

Un effet de bord des actions de 1024 et `binaire` est que notre communauté est en train de *se créer* des écrivains de l'informatique. Chaque semaine, de nouveaux articles, écrits par de nouveaux auteurs, nous sont proposés. Dans certains cas, il convient de retravailler le matériel : écrire pour le grand public n'est pas aussi simple que l'on pourrait le croire ! L'intérêt de cet effort collectif est que nous pouvons espérer passer d'un mode où les thèmes à traiter dépendent d'abord des propositions des auteurs, à un mode dans lequel, devant une question importante, nous saurons vers qui nous tourner pour demander un article.

Qu'on ne s'y méprenne pas ! Il ne s'agira pas de remplacer la sympathique improvisation actuelle par un traitement trop systématique. Mais nous devons aller vers cela et j'encourage les lecteurs de ce numéro de 1024 à devenir les auteurs du prochain.

L'école

L'informatique va-t-elle bientôt s'enseigner à l'école, au collège, au lycée ? À l'heure d'écrire ces lignes, les annonces ministérielles n'ont pas encore été faites. Je ne peux donc que vous livrer ici l'analyse qu'a faite la SIF des événements, rencontres, déclarations, discussions de ces derniers mois. Cela permettra au lecteur, à l'adhérent, de comprendre les positions que nous prenons en tant que participants à ce débat.

En juin dernier, le CSP (Conseil supérieur des programmes) a remis son rapport sur le « Socle commun de connaissances, de compétences et de culture ». On peut notamment y lire (Domaine 1) : *« L'élève sait que les équipements informatiques utilisent une information codée et il est initié au fonctionnement, au processus et aux règles des langages informatiques ; il est capable de réaliser de petites applications*

utilisant des algorithmes simples. » Ainsi que : *« Ce domaine (Domaine 4) comprend un vaste pan de la culture regroupant les mathématiques, l'informatique, les sciences de la vie et de la Terre, la physique, la chimie, la technologie. (...) En abordant les sciences dès le début de l'école primaire et en les pratiquant de façon active jusqu'à la fin du collège, l'élève se familiarise avec la démarche scientifique. Il est initié à la démarche d'investigation grâce à des pédagogies adaptées ou en étant confronté à des problèmes ouverts. »*

Sur la base de ce texte, la SIF a décidé de réagir positivement, et a invité différents organismes et associations à signer un texte commun [4]. Les signataires représentaient l'AFDEL, l'EPI, INRIA, l'INS2I-CNRS, Pasc@line, la PEEP, le Syntec Numérique, ainsi que la SIF.

En juillet, suite à ce communiqué, mais aussi à différentes annonces, différents membres des Conseils d'administration et scientifique de la SIF sont intervenus dans les médias, pour montrer leur adhésion aux débuts de propositions faites tout en rappelant à quel point la formation des enseignants était un enjeu prioritaire [5].

Le rapport du Conseil national du numérique [6], publié en octobre, est venu dire avec des arguments et des chiffres parfois nouveaux des choses que de nombreux informaticiens avaient envie d'entendre. Des propositions précises ont été formulées : ainsi, le CNNum n'hésite pas à dire que la discipline informatique doit être enseignée, et qu'elle doit l'être par des informaticiens, n'hésitant pas à demander à la fois un CAPES et une Agrégation en informatique ! Les débats suivant cette publication ont été vifs sur la toile et dans les médias.

Pour les informaticiens, il me semble essentiel que nous soyons plus particulièrement attentifs au discours de celles et ceux qui, au sein du CNNum, ne sont pas des informaticiens mais ont participé aux travaux du groupe ayant préparé le rapport. Si les participants à ce groupe se sont retrouvés sur un texte commun si fort et qui, par bien des aspects, va beaucoup plus loin que ce que la SIF pouvait espérer, c'est aussi parce que les réflexions ont convergé. Notre devoir d'informaticiens ne peut pas consister à n'y prendre que ce qui nous intéresse et ignorer ou combattre le reste. Il me semble indispensable que nous étudiions également ce qui ne fait toujours pas l'unanimité.

Dans une déclaration ultérieure à la publication du rapport, Daniel Kaplan, délégué général de la Fondation internet nouvelle génération (FING), résume [7] le chemin parcouru : *« . . . soit le Capes est une mauvaise manière de recruter tous les profs (ça se discute, sûrement !) et il faut le supprimer, soit c'est la manière standard et*

4. http://www.societe-informatique-de-france.fr/2014/06/communique-rapport-csp/

5. La récente *Newsletter* de la SIF contient les liens vers les principales interventions :http://www.societe-informatique-de-france.fr/newsletter/

6. http://www.cnnumerique.fr/education-2/

7. http://www.internetactu.net/2014/10/13/jules-ferry-3-0-recit-dune-convergence/

reconnue et l'on se demande bien pourquoi seuls les profs d'informatique y échapperaient ? »

Il est bien entendu impossible pour nous de ne pas être d'accord avec Daniel Kaplan là-dessus. On pourrait se contenter de ne garder que cet extrait, mais il est bien plus intéressant d'examiner le reste.

Ailleurs, le même Kaplan écrit : *« D'où l'idée autour de laquelle nous avons convergé, entre ceux qui, dès le départ, militaient en faveur de l'enseignement de l'informatique et ceux qui n'y croyaient pas : faire de cet enseignement le"cheval de Troie" par lequel élèves et enseignants explorent de nouvelles formes de travail et d'apprentissage. Nous considérons en effet qu'il serait absurde et même néfaste d'enseigner l'informatique au tableau noir (voire au tableau blanc interactif), que son enseignement passe nécessairement par une organisation en projets et par un travail collectif. Et où trouver des "projets" qui ont du sens ? Dans les autres disciplines ! »*

Les premières réactions des informaticiens tiennent — admettons-le — du grincement de dents : si, au Royaume-Uni, la position de départ à l'introduction du *computer science* à l'école a consisté à la considérer comme une science à égalité avec les autres, voilà qu'en France on pourrait plutôt lui demander un statut spécial, un statut qui lui donne un rôle pionnier, mais également un rôle menant à une différence pédagogique, à priver ses enseignants de leur liberté pédagogique ?

Or si l'on regarde les réactions négatives (de notre bord) au propos de Daniel Kaplan, on arrive en fait à s'en réjouir ! *Comment ça, au tableau noir ? Monsieur Kaplan devrait voir ce qui se faisait lorsque l'informatique était enseignée, ce qui se fait aujourd'hui dans les Écoles et les Universités, ou tout simplement en ISN ! L'approche par projets a toujours été privilégiée !* Quelle belle information. Cela signifie qu'au détail près que nous ne savons pas toujours assez bien montrer nos méthodes pédagogiques, nous sommes tous d'accord !

Il faut donc, je crois, répondre collectivement **oui** à toute proposition d'organisation de l'enseignement en France basée sur le rapport du CNNum, un **oui** global sur l'ensemble des propositions, un **oui** pour dire que nous serons, informaticiens, partants pour élaborer ce projet pour l'école, un **oui** pour dire que la SIF agira pour convaincre les signataires du communiqué commun de juin (ainsi que d'autres qui nous ont dit qu'ils l'auraient signé s'ils y avaient été invités) que l'adhésion à un projet en continuité avec les propositions du CNNum…

Un **oui** dans sa globalité, c'est-à-dire qui ne prend pas dans ce rapport ce qui va dans son sens et rejette le reste et permettre ainsi une large union, bien au-delà de l'informatique et en faveur du « grand plan numérique pour l'école » voulu par le Président de la République.

Conclusion

Avec un peu plus de place, je vous en dirais un peu plus sur les échanges du printemps pédagogique dernier sur les MOOCs ou sur le prochain congrès qui s'annonce passionnant : « Femmes et Informatique », et qui sera l'occasion de nous réunir, d'échanger et de discuter, à Orléans, en février prochain.

Mais je vous laisse surtout entre les mains du 4e numéro de 1024 que l'équipe éditoriale dirigée par Éric Sopena a préparé pour vous. À moins que ce ne soit l'inverse.

Actualité de la SIF

Informations collectées par Sylvie Alayrangues

Une rubrique pour vous tenir informé de la vie de la SIF et vous inviter à y participer. Vous trouverez notamment ici des annonces de manifestations organisées ou soutenues par la SIF, un point sur les dossiers en cours, des appels à participation...

La SIF vous parle d'elle

Lors des réunions du Conseil scientifique d'avril et octobre dernier, deux postes clés de ce conseil ont été renouvelés. Christine Froidevaux succède à Jean-Marc Petit en tant que secrétaire et Gilles Dowek prend la suite de Serge Abiteboul à la présidence. C'est l'occasion d'adresser un grand merci à Jean-Marc et Serge pour tout le travail accompli lors de ce mandat et pour leur implication qui va bien entendu se poursuivre puisqu'ils continuent à œuvrer au sein du Conseil scientifique.

Comme chaque année, le Conseil d'administration sera partiellement renouvelé lors de l'Assemblée générale de la SIF qui se tiendra en février 2015 à Orléans à l'occasion de de notre congrès annuel. Si vous souhaitez vous impliquer dans les nombreux dossiers qui nous occupent, n'hésitez pas à présenter votre candidature jusqu'au 31 octobre 2014. Tous les adhérents ont dû recevoir un formulaire de candidature dans leur boîte courriel.

En juillet, tous les adhérents ont également reçu une lettre pour faire un point sur l'utilisation de leurs cotisations [1].

1. societe-informatique-de-france.fr/wp-content/uploads/2014/02/lettre_finances.pdf

Pour améliorer la circulation de l'information, la SIF s'est dotée récemment d'une newsletter, la « Lettre d'information de la SIF », qui parvient par courriel à ses adhérents et peut se retrouver ensuite sur son site [2].

La SIF, bien sûr, continue d'alimenter ses comptes sur les réseaux sociaux :
- LinkedIn, `www.linkedin.com/groups?gid=4757149`,
- Facebook, `facebook.com/groups/sif.informatique/`,
- Twitter, `twitter.com/SocInfoFr`.

La SIF vous propose

Du mardi 3 février au jeudi 5 février 2015 : Congrès de la SIF « Femmes et Informatique »

Le congrès se déroulera à Orléans. Laissons la parole aux organisateurs :

« Venez assister ou nous faire part de vos expériences, nous nous retrouverons pour échanger sur : les prochaines générations (« de la crèche au doctorat »), l'enseignement supérieur et la recherche, l'entreprise, la situation à l'étranger...

Les orateurs invités confirmés sont :
- Isabelle Collet (Université de Genève), chercheuse en sciences de l'éducation (informaticienne scientifique de formation initiale), auteure de *L'informatique a-t-elle un sexe ?*,
- Pierre-Henri Gouyon (MNHN), généticien, qui travaille sur Inné, Acquis / Sexe, Genre. Il devrait nous en parler plus particulièrement dans le cas de l'informatique,
- Pierre Mounier-Kuhn (CNRS & Paris-Sorbonne University), historien des sciences, dont les derniers travaux portent sur la féminisation dans l'informatique (histoire récente à nos jours)

La traditionnelle Assemblée générale se déroulera le 3 février en fin d'après-midi et une rencontre avec les différentes communautés scientifiques (GDR, Association thématique...) sera organisée avant l'AG.

Le prix de thèse Gilles Kahn sera remis le 4 février.

Nous comptons sur vous ! »

Les co-présidents du comité scientifique,

Jérôme Durand-Lose, LIFO, Orléans
jerome.durand-lose@univ-orleans.fr

Florence Sèdes, IRIT, Toulouse
florence.sedes@irit.fr

2. `societe-informatique-de-france.fr/newsletter/`

Toutes les informations seront bientôt en ligne sur :
societe-informatique-de-france.fr/congres-2015/

La SIF soutient

Pixees

Derrière le doux nom de Pixees s'affiche l'espace de médiation scientifique du réseau de médiation scientifique INRIA et de ses partenaires, dont la SIF. Il rassemble les ressources (documentaires et humaines) en médiation scientifique des sciences du numérique. C'est avant tout un espace collaboratif auquel contribuent (co-création de contenu, réalisation d'intervention, réflexion commune...) de nombreux acteurs qui forment une communauté active sur ces sujets.

Bref, si vous avez besoin d'un coup de main pour réaliser des activités de médiation scientifique ou si vous avez envie de partager vos expériences, allez-y [3] !

Des ateliers d'informatique débranchée

Dans le cadre de la semaine européenne du code, la SIF a participé avec INRIA et MagicMakers à un atelier de découverte du codage et des algorithmes en famille [4].

L'espace Mendès France, le centre de médiation de Culture scientifique technique et industrielle de Poitiers propose depuis cette année des ateliers ouverts à partir de 8 ans (sans limite d'âge haute) pour initier à l'informatique. Les ateliers comprennent une partie « débranchée » pour découvrir des concepts en s'amusant (algorithmique, cryptographie...), une partie programmation utilisant Scratch et des échanges entre les participants et des chercheurs en informatique. Ces ateliers sont animés par des acteurs du Lieu multiple (pôle de création numérique de l'espace Mendès France) et par des enseignants-chercheurs de l'Université de Poitiers (adhérents de la SIF) [5].

Le prix Bernard Novelli lancé par le magazine Tangente

Pour la deuxième année la SIF est partenaire et membre du jury du Prix Bernard Novelli [6] proposé par le magazine Tangente. Ce prix est dédié à la mémoire du créateur de jeux Bernard Novelli et récompense des lycéens (en particulier les élèves en spécialité Informatique et sciences du numérique) auteurs de projets informatiques autour du jeu. La remise du prix aura lieu le mercredi 19 novembre au Palais du Luxembourg.

---

3. pixees.fr
4. societe-informatique-de-france.fr/2014/09/semaine-du-code-codage-et-algorithme-en-famille/
5. emf.fr/20366/programmer-pour-ne-pas-pas-etre-programme/
6. prixnovelli.fr

Le concours Castor informatique

L'édition 2014 du concours Castor informatique [7] se prépare. Elle aura lieu du mercredi 12 novembre à 7h au mercredi 19 novembre à 20h, heure de Paris. Si vous êtes enseignant de collège ou lycée, n'hésitez pas à proposer le concours à vos élèves.

La SIF prend position et agit

En enseignement

À l'initiative de la SIF, huit organismes et associations (l'AFDEL, le CNRS, l'EPI, INRIA, Pasc@line, la PEEP, la SIF, le Syntec Numérique) ont signé en juin un communiqué commun [8] pour saluer le rapport du Conseil supérieur des programmes, remis le 10 juin, qui octroie à l'informatique une réelle place dans le projet de socle commun de connaissances, de compétences et de culture. L'ensemble des signataires est également prêt à mettre la main à la pâte pour que les conclusions du conseil supérieur des programmes puissent être mises en œuvre dans de bonnes conditions.

La SIF s'est associée au communiqué [9] de l'EPI (association Enseignement public & informatique) du 6 septembre 2014. Celui-ci prône deux approches complémentaires dans l'éducation : l'utilisation de l'informatique dans l'ensemble des disciplines et l'enseignement de l'informatique en tant que discipline à part entière. Il met en exergue quelques points clés de l'enseignement de l'informatique au primaire, au collège et au lycée et pose une nouvelle fois la question de la formation des enseignants.

Le printemps pédagogique 2014 de la SIF s'est tenu sur deux jours à Paris (au tout début de l'été !), les 23 et 24 juin dernier. Pendant deux jours, les MOOCs étaient au cœur du débat. Vous pouvez retrouver les thèmes abordés ainsi qu'une partie des supports des orateurs sur la page dédiée [10]. Vous pouvez également suivre les échanges auxquels elles ont donné lieu sur Twitter (mot-dièse #sifmooc).

L'an dernier, la SIF s'était impliquée dans un collectif créé par la CNIL pour faire de l'éducation au numérique une grande cause nationale. C'est finalement l'engagement associatif qui a été choisi comme thème de grande cause nationale en 2014. Une nouvelle candidature est envisagée pour 2015, avec une place plus importante réservée à la formation à l'informatique dans la formation au numérique.

---

7. castor-informatique.fr
8. societe-informatique-de-france.fr/2014/06/communique-rapport-csp/
9. societe-informatique-de-france.fr/2014/09/communique-epi/
10. societe-informatique-de-france.fr/enseignement/j-pedago/j-pedago-2014-programme/

En recherche

En tant que société savante représentant la science informatique, la SIF partage les inquiétudes des différents acteurs de la recherche scientifique quant à l'avenir de cette recherche en France. Elle a donc apporté son soutien au mouvement « Sciences en marche [11] » qui a conduit pendant plusieurs semaines des chercheurs à sillonner la France, à pied ou à vélo pour converger sur Paris le 17 octobre. Cette manifestation visant à alerter sur la dégradation des conditions de travail des chercheurs (précarisation des métiers de la recherche, diminution chronique des financements...) a regroupé plusieurs milliers de personnes.

Le dynamisme de la recherche passe aussi par le transfert de la recherche fondamentale vers l'innovation. La journée recherche de la SIF de septembre dernier s'est intéressée aux liens entre les formations scientifiques et techniques (niveau bac+5 et plus) et l'innovation en informatique portée par les entreprises du secteur. Cette journée, organisée en partenariat avec Pasc@line, a rassemblé une centaine de participants venus de divers horizons (responsables de formations, industriels, conseillères d'orientations...). Les interventions des deux conférenciers invités, Émeric Bréhier, député, auteur d'un rapport sur le doctorat et Guy Mamou-Mani, président du Syntec Numérique, ainsi que les débats qu'elles ont suscités ont été particulièrement appréciés. La qualité et la diversité des intervenants des tables rondes : acteurs de l'innovation dans les principaux organismes de recherche publique, ou dans des entreprises privées, responsables de master ou d'école d'ingénieur, ont aussi largement contribué à enrichir la discussion.

Vous pouvez retrouver les thèmes abordés et les supports des intervenants sur le site de la SIF [12], et les vidéos filmées lors de l'événement sur la « chaîne SIF » [13].

En médiation scientifique

Le blog `binaire` [14] est en lice pour la 5^{e} édition des « Golden blog awards ». La remise des prix aura lieu le 12 novembre. Vous pouvez voter pour lui [15] !

Un numéro spécial de 1024 sur la médiation scientifique, faisant suite au congrès 2014, sortira prochainement. Il s'agit, comme au congrès, de donner la parole, ou plutôt la plume, à celles et ceux qui agissent et partagent autour d'eux des grains de science informatique. Vous y retrouverez des témoignages d'expériences vécues, des pistes de méthodologie, des analyses... Vous reconnaitrez une partie des intervenants du congrès... et en découvrirez d'autres.

---

11. `sciencesenmarche.org`
12. `societe-informatique-de-france.fr/recherche/les-journees-recherche/journee-recherche-de-la-sif-2014/`
13. `videolectures.net/sif/`
14. `binaire.blog.lemonde.fr/`
15. `www.golden-blog-awards.fr/blogs/binaire.html`

Au cœur des données

Entretien avec Serge Abiteboul (*réalisé par Valérie Schafer*)

Des bases de données au « data déluge » actuel, Serge Abiteboul se passionne depuis plus de trente ans pour un domaine qui ne cesse d'évoluer et de nous confronter à de nouveaux défis, techniques, mais aussi sociétaux. Il a reçu en 2013 le prix Milner de la Royal Society et vient d'achever le projet Webdam[1] *sur la gestion des données, pour lequel il avait été lauréat en 2008 d'une bourse ERC. Membre de l'Académie des sciences et du Conseil national du numérique, président du Conseil scientifique de la SIF, co-fondateur du blog* BINAIRE[2]*, Serge Abiteboul, chercheur à Inria, revient sur sa carrière, ses recherches et éclaire pour nous quelques-uns des enjeux que pose aujourd'hui le numérique.*

Valérie Schafer[3]

Valérie Schafer : *Vous avez démarré votre carrière scientifique au milieu des années 1970, en faisant des études à Telecom Paris et au Département d'informatique du Technion de Haïfa (Israël). Vous obtenez un diplôme de Telecom Paris en 1978, faites votre service militaire à l'ESEAT*[4] *de Rennes, puis vous partez pour la Californie*

1. http://webdam.inria.fr
2. http://binaire.blog.lemonde.fr
3. Valérie Schafer, CNRS, ISCC. Entretien co-publié avec la revue Technique et Science Informatique (TSI).
4. Maintenant, École des transmissions (de l'armée de terre).

et vous réalisez un PhD sous la direction de Seymour Ginsburg[5]*. Revenons pour commencer sur ces années si vous le voulez bien. Qu'est-ce qui vous pousse à vous tourner vers l'informatique et à partir aux États-Unis ?*

Serge Abiteboul : Les gens parlent souvent de vocations précoces... Ce n'est pas le cas pour moi. Je ne savais rien de l'informatique quand j'ai commencé mes études. J'ai fait des classes préparatoires scientifiques alors que j'aurais préféré des études littéraires. Puis je suis rentré à Télécom, tout bêtement parce que c'était la meilleure école que j'avais obtenue. Lors de ma troisième année je suis parti en Israël, dans le cadre d'un échange d'étudiants. J'ai choisi l'informatique parce que cela me permettait de passer un an en Israël. Et c'est là-bas que j'ai vraiment commencé à comprendre ce que c'était. À l'époque, il faut reconnaître que ce qu'on nous enseignait à Télécom n'était pas ragoutant. C'était l'« informatique bidouille » ; on apprenait à écrire des programmes en bricolant ; on nous laissait dans une ignorance crasse des fondements théoriques du domaine. La théorie des langages, l'algorithmique, la complexité, c'est en Israël que j'ai découvert tout ça. Je suis ensuite revenu en France à l'ESEAT de Rennes pour mon service militaire. Après, je n'avais pas envie de devenir ingénieur ; j'avais plutôt envie de voyager. Alors, j'ai obtenu un poste d'assistant d'enseignement à Los Angeles et je suis parti faire un PhD. C'était la vie d'étudiant en Californie… et j'y ai découvert la recherche avec Seymour Ginsburg.

V.S. : *Parlez-nous un peu de Seymour Ginsburg si vous le voulez bien.*

S.A. : Seymour Ginsburg avait une vraie densité scientifique. C'était une personnalité pas ordinaire, un personnage étonnant par des côtés parfois ridicules, pas très sociable, mais quand il expliquait une preuve… Tout à coup il devenait passionnant ! Ses travaux traitaient de théorie des langages. Je ne me suis pas engagé dans cette voie, et ce pour une raison simple : la Mecque mondiale de la théorie des langages, c'était l'école de Maurice Nivat en France ! Pour Ginsburg, Nivat était l'un des meilleurs informaticiens au monde. Je ne connaissais pas Nivat personnellement à l'époque, nous nous sommes rencontrés depuis. Toujours est-il qu'il aurait été ridicule d'aller aux États-Unis pour travailler sur un sujet où l'excellence était en France. Seymour Ginsburg voulait regarder du côté des bases de données, un domaine qui démarrait à peine. Alors pour quelqu'un de fainéant comme moi, une quinzaine de papiers à lire, face à des centaines de publications dans la théorie des langages ! Je n'ai pas hésité. Et puis, il y avait l'attrait de la nouveauté. Certains choisissent la difficulté, aiment s'attaquer à des problèmes auxquels d'autres se sont confrontés pendant des années sans succès, moi je préfère la facilité et la nouveauté, définir

5. Pionnier de la théorie des langages formels et de la théorie des bases de données, voir notamment l'hommage qui lui a été rendu en 2004 et qui retrace son parcours dans *SIGMOD Record*, vol. 34, n° 1, March 2005. http://www.sigmod.org/publications/sigmod-record/0503/p5.special.vianu.pdf.

mes propres problèmes. De plus, le domaine des bases de données était à la frontière entre informatique théorique et appliquée, ce qui m'attirait aussi. J'aime pousser les limites de la formalisation, investiguer de nouvelles frontières.

« For me, Seymour was at first the professor with a strange hat and immutable clothes going down the corridor of the department, a bit abrupt but always so helpful to the young ones. Seymour Ginsburg, the famous professor, a pioneer of formal languages, a knight of computer science theory, always found seemingly unlimited time for his PhD students. I remember my visits to his place on Saturdays to go over my thesis together, after the invariable spaghetti lunch. He would check every single definition, result, every single proof, sentence, every single comma. [...]

After many years, when I look back, I see that no one influenced my work more than Seymour. I realize when I talk to my PhD students, that I am repeating his lessons. When I write a paper, I try to follow his very precise guidelines. When I get started on some new problem, I wonder : what questions would have Seymour asked ? »

Serge Abiteboul, in
« In memory of Seymour Ginsburg 1928-2004 »,
SIGMOD Record, vol. 34, n° 1, March 2005. `http://goo.gl/F7FMoJ`

V.S. : *Sur quoi travaillait-on concrètement à l'époque quand on faisait de la recherche sur les bases de données ?*

S.A. : Les bases de données à l'époque, c'étaient les bases de données relationnelles, les tableaux à deux dimensions, des logiciels comme System R d'IBM ou Oracle. Ma thèse a porté sur les propriétés algébriques des bases de données relationnelles. Très rasoir. Il n'y avait aucune motivation pratique. J'ai appris des tas de choses en rédigeant ma thèse. Mais j'ai surtout appris que je ne me satisfaisais pas de tels travaux purement mathématiques. J'avais besoin d'accroches sur des aspects plus pratiques, des applications. Je n'affirme ici aucun dogme, aucune hiérarchie de valeur. Il n'y a pas une seule façon de faire la recherche en informatique. Il y a des recherches très théoriques et d'autres plus expérimentales. Pour moi, le premier moteur de recherche de Google et son PageRank par Sergeï Brin et Larry Page, c'est ni plus ni moins de la recherche que, par exemple, la preuve de l'indécidabilité de l'implication (finie) pour les dépendances fonctionnelles et d'inclusion par Ashok Chandra et Moshe Vardi. Mon goût personnel, c'est d'essayer de concilier des aspects à la fois applicatifs et de recherche fondamentale ; c'est tout à fait possible en bases de

données, pas dans tous les domaines ; c'est aussi pour ça que je me suis plu en bases de données. Et puis, c'est un domaine qui bouge, qui vit. C'est important de bien choisir son champ, de ne pas se laisser enfermer dans un domaine en train de mourir. Toute la problématique des données, de l'information, des connaissances, ne cesse de poser de nouveaux enjeux, d'évoluer ; il y a eu un cheminement du domaine et ma recherche évolue avec.

V.S. : *Vous rentrez ensuite en France ?*

S.A. : Je suis entré à Inria Rocquencourt en 1982 dans l'équipe de François Bancilhon, qui est passé assez vite après dans l'industrie, et qui est resté un ami. C'est l'actuel PDG de Datapublica. J'avais l'idée de passer un an ou deux en France et de retourner aux États-Unis…

V.S. : *...et vous êtes toujours chez Inria 30 ans après...*

S.A. : C'est un super cadre de travail pour faire de la recherche en informatique. Ce n'est peut-être pas le confort de Stanford ou Oxford, mais en France c'est ce qui se fait de mieux. Et puis, j'ai bougé de Rocquencourt, à Saclay, puis à Cachan. (rire) Plus sérieusement, Inria offre un très bon support d'aide à la recherche. Par exemple, quand vous montez une entreprise, Inria sait vraiment aider.

V.S. : *Ce que vous avez fait avec Xyleme SA, devenue Xyleme Inc.*

S.A. : Oui, avec des amis, notamment Sophie Cluet, nous avons lancé Xyleme. Je m'y suis investi à partir de 2000, pendant cinq ans environ, avant même la création de la startup. Déjà un an et demi avant, nous étions sur la recherche qui a conduit à l'entreprise et sur le montage de celle-ci, son business plan, etc. Mais il faut souligner que pendant ces années où nos activités étaient axées sur le développement d'un logiciel à transférer, je n'ai pas vu de diminution de la qualité scientifique de l'équipe. Si l'on prend des critères comme le nombre de publications, de doctorants, etc., critères qui valent certes ce qu'ils valent, mais constituent tout de même des indications, il n'y a pas eu de baisse de productivité de recherche. Le projet a attiré d'excellents étudiants, motivés par cet aspect appliqué et industriel. Le logiciel portait sur la gestion de documents semi-structurés. Parmi les clients importants que nous avons eus chez Xyleme, il y avait par exemple *Le Monde*, qui a mis toutes ses archives sur notre système. Mais dès que j'ai pu, je suis retourné à la recherche. Je reste fondamentalement un chercheur.

V.S. : *Vous aimez aussi transmettre ? Vous avez des activités d'enseignement ?*

S.A. : Oui, en effet. Actuellement, j'enseigne la gestion de données sur le web [6] au master MPRI [7] et les bases de données relationnelles en licence à Cachan.

V.S. : *Et vous avez été un an au Collège de France en 2011-2012* [8].

S.A. : Oui. Ça a été une expérience extraordinaire avec un public très différent. Cette année au Collège de France m'a amené à réfléchir sur des aspects plus sociétaux, qui maintenant me passionnent ; une partie de mon temps aujourd'hui est dédiée à des questions moins techniques. Je suis notamment au Conseil national du numérique. À l'intérieur du conseil, le boulot est vraiment différent. J'ai fait par exemple partie d'un groupe sur l'inclusion numérique [9]. Nous avons rencontré des tas de gens. Je me doutais bien des problèmes d'exclusion générés par le numérique, mais ils étaient restés abstraits. De rencontrer des gens qui sont confrontés au quotidien à ces problèmes-là, cela change votre perspective. Et puis après, il faut regarder ce qui se fait, imaginer des solutions. C'est passionnant. Les chercheurs devraient regarder davantage autour d'eux ce qui se passe dans la vraie vie.

V.S. : *Sur ce point les pays anglo-saxons n'ont-ils pas pris un peu d'avance sur nous ? S'occupe-t-on suffisamment en France d'éthique des TIC ?*

S.A. : La France ne me semble pas en retard. Elle est même très en avance sur certains aspects comme la protection des données avec la Cnil par exemple. L'Europe en général me paraît plus que les États-Unis sensible à ces enjeux. En France, nous avons un système d'associations très dynamiques avec, par exemple, la FING [10] ou encore la Quadrature du Net [11]. Les associations font énormément sur le terrain. Plus récemment, en informatique, en France, un comité piloté par Max Dauchet, la Cerna [12], a été mis en place pour travailler sur les problèmes d'éthique. On aimerait que les chercheurs en informatique s'impliquent plus dans la société. Évidemment, c'est une généralité. Certains le font déjà énormément.

---

6. Voir notamment Serge Abiteboul, Ioana Manolescu, Philippe Rigaux, Marie-Christine Rousset et Pierre Senellart, *Web Data Management*, Cambridge University Press, 2011.

7. Master parisien de recherche en informatique.

8. La leçon inaugurale du 8 mars 2012 de Serge Abiteboul, intitulée « Sciences des données : de la logique du premier ordre à la toile » est disponible à l'adresse : `http://www.college-de-france.fr/site/serge-abiteboul/inaugural-lecture-2012-03-08-18h00.htm` Voir également sa *Royal Society Milner Award Lecture* « From data and information to knowledge : the Web of tomorrow », intervention à la *Royal Society* de Londres du 12 novembre 2013 : `http://royalsociety.org/events/2013/web-of-tomorrow/`.

9. Voir le rapport « Citoyens d'une société numérique, Accès, littératie, médiations, pouvoir d'agir : pour une nouvelle politique d'inclusion ». `http://www.cnnumerique.fr/inclusion/`

10. Fondation Internet Nouvelle Génération. `http://fing.org/`

11. `https://www.laquadrature.net/fr`

12. Commission de réflexion sur l'Éthique de la Recherche en sciences et technologies du Numérique (CERNA). `http://www.inria.fr/actualite/mediacenter/creation-de-la-cerna`

V.S. : *En parlant des États-Unis, j'ai été frappée de voir que vous aviez été consultant pour la NASA. Les liens entre le monde de la recherche en informatique et le monde militaire semblent tellement plus forts outre-Atlantique...*

S.A. : Oui, dans ma carrière aux États-Unis, j'ai été payé par la NASA mais j'ai aussi travaillé sur des contrats avec ARPA et plusieurs autres organismes publics américains. Pendant les six ans que j'ai passés aux États-Unis, j'ai été régulièrement payé par les militaires, mais toujours pour faire la recherche que je voulais, par exemple sur les modèles de données semi-structurés (les ancêtres de XML) ou des modèles logiques du e-commerce. C'est une vraie différence entre la France et les États-Unis. J'en ajouterai une deuxième : aux États-Unis, les industriels venaient souvent voir ce que je faisais dans mon labo, y compris d'ailleurs des industriels français, alors qu'en France, les mêmes, s'ils viennent me voir, c'est qu'ils cherchent une subvention de l'ANR ou de la CE. Évidemment, mon point de vue est biaisé. Mes copains industriels vous diront peut-être que les enseignants-chercheurs américains sont plus faciles à approcher que les français. Il faudrait travailler pour changer les mentalités, fluidifier les liens entre les labos académiques et les équipes de R&D des entreprises.

V.S. : *C'est aux États-Unis que vous découvrez Internet ?*

S.A. : Effectivement quand j'ai commencé ma thèse à Los Angeles, il y avait Internet, on s'envoyait des emails. Quand je suis revenu en France, on m'a dit qu'on ne pouvait envoyer des emails qu'à l'intérieur d'Inria... La préhistoire quoi ! Enfin, grâce à des copains au Bâtiment 8, j'ai pu garder le contact par mèl avec les copains aux États-Unis. Quant au web, je l'ai découvert à sa naissance. J'ai tout de suite pensé que cela pouvait tout changer pour les bases de données, les rendre disponibles de partout. On pouvait sortir une base de données du deuxième sous-sol où elle était cantonnée pour la mettre à disposition dans le monde entier ! François Bancilhon avait créé une start-up qui avait développé un système de gestion de bases de données objet, O_2 ; avec un étudiant, on a fait O_2Web pour accéder par le web à une base de données O_2. Cette notion d'information mondialisée explosait le paradigme de la gestion de données. À l'époque, le web restait confidentiel. Il apparaissait comme le deuxième alinéa d'une sous-rubrique dans le rapport prospectif d'Inria. Puis, c'est allé très vite. Quelques années après la naissance du web, tout le monde s'est mis à publier sur le web et on a vu arriver les premiers moteurs de recherche, puis Google... J'étais dans la même équipe de recherche de Stanford que les « Google Guys », Lawrence Page et Sergeï Brin. Ils nous ont présenté leurs idées. Elles n'étaient pas si révolutionnaires au premier abord. Ils reprenaient un algorithme d'IBM. Mais ils montraient qu'on pouvait faire ça sur le web. Ils nous ont fait part de leur idée de la première architecture et des algorithmes. Nous étions sceptiques face à la taille de l'index nécessaire. Quelques mois après, ils me faisaient leur première démo. J'ai testé leur système en posant la requête « Rodin » sur Infoseek et Google. Il n'y avait

pas photo, leur moteur de recherche était bien meilleur [13]. Qui se rappelle encore d'Infoseek, un des moteurs de recherche du web les plus utilisés à l'époque ?

V.S. : *On parle aujourd'hui de tsunami de données, de déluge informationnel (Data Deluge). Ce ne sont pas des termes très positifs. Faut-il craindre la révolution annoncée ?*

S.A. : Évidemment, il y a la masse, la déferlante. Nous générons de plus en plus d'information. Plus de gens génèrent des données. De plus en plus d'objets connectés génèrent des données. Ce qu'on produit comme informations et données double tous les 18 mois. Ce doublement est-il soutenable ? On n'a pas le choix ! Cela pose des questions : Que faire de toutes ces données ? Comment les gérer ? Prenez les données personnelles, mon sujet de recherche actuel. Cela devient impossible aujourd'hui de gérer ses données perso. Par exemple, avant on pouvait retrouver une information sur son PC en faisant une recherche avec *Google desktop*. Maintenant, Google a arrêté Google desktop, parce que nos données ne sont plus dans le desktop ; elles sont sur dix systèmes différents, dans les nuages. Vous ne pouvez plus faire de recherche globale sur vos données personnelles. Le problème n'est pas technique. La question, c'est que vous ne contrôlez plus vos données, qu'elles sont ailleurs. Comment faire coopérer ces données que vous avez sur dix ou cent systèmes différents, comment gérer par exemple les contrôles d'accès ? On est en plein dans la gestion de données distribuées. Mais on est aussi dans des problèmes de société. Comment faire pour que Prism ne sache pas tout de moi ? Comment faire disparaître une photographie que je souhaite ne plus voir apparaître sur le web ? Il faut des lois, il faut de meilleures pratiques, il faut des utilisateurs avec une vraie littératie numérique, il faut aussi de nouvelles solutions techniques. C'est drôle, parfois j'entends presque la même question dans un séminaire de recherche et dans une réunion du Conseil national du numérique. La recherche en gestion de données découvre sans cesse de nouveaux challenges.

V.S. : *Puisque nous nous rencontrons en ce tout début d'année 2014, je ne peux résister à la tentation de vous demander si l'homme de science que vous êtes a pris des « bonnes résolutions » professionnelles pour la nouvelle année... Et si l'on devait souhaiter une bonne année à la science informatique, que pourrait-on lui souhaiter ? Quels vœux lui adresser ?*

S.A. : Mes bonnes résolutions ? Travailler moins (rire). Pour ce qui est de la recherche, c'est un peu tôt pour répondre. J'ai fini le mois dernier le projet Webdam [14]

13. Voir notamment Sergueï Brin et Lawrence Page. The anatomy of a large-scale hypertextual web search engine. *Proceedings of the 7th International Conference on World Wide Web*, Amsterdam, Elsevier, Computer Networks and ISDN Systems, vol. 30, n° 1-7, 1998, p. 107-117, doi : 10.1016/S0169-7552(98)00110-X.

14. `http://webdam.inria.fr`

de l'*European Research Council*. Quant à la science informatique, je souhaite le décollement véritable de l'enseignement de l'informatique en France [15, 16], qu'on sorte de l'ornière ; cela ne peut venir que d'une décision politique. Il faut la volonté politique de former des écoliers, des collégiens, des lycéens, des citoyens qui sachent tous programmer, des littéraires qui sachent ce qu'est un moteur de recherche. Il faut former toute la population à l'informatique. C'est indispensable. Pas juste pour former une génération d'informaticiens, mais pour avoir aussi des scientifiques, des ingénieurs plus performants, plus innovants. Avions, voitures, appareils photo, tous ces objets incluent maintenant des tas d'informatique. Mais un artiste aussi a besoin aujourd'hui de maîtriser l'informatique. Un médecin, un plombier, un cultivateur. Tout le monde. L'informatique doit entrer dans l'école au même titre que les mathématiques, la physique, la chimie ; l'informatique doit faire partie du socle fondamental. Chacun doit apprendre à maîtriser l'informatique, pour ne pas être juste un utilisateur de logiciels, ne pas devenir l'esclave de logiciels développés par d'autres. On a l'impression de radoter quand on dit ça tellement ça a été répété. Mais je crois que c'est vraiment essentiel.

15. Rapport de l'académie des sciences, *L'enseignement de l'informatique en France, Il est urgent de ne plus attendre*, mai 2013, www.academie-sciences.fr/activite/rapport/rads_0513.pdf

16. Voir l'article de Serge Abiteboul, Gilles Dowek et Colin de la Higuera du 24 janvier 2013 « Recherche profs d'informatique désespérément », dans *01 Business*, disponible sur : http://pro.01net.com/editorial/585095/recherche-profs-dinformatique-desesperement/

Bio-algorithmique des ARN : petite promenade aux interfaces

Yann Ponty [1]

Introduction

La bioinformatique est un champ disciplinaire consacré à l'analyse des données biologiques par des méthodes informatiques. S'agissant d'une discipline transverse, elle concerne de nombreux aspects de la recherche en informatique, allant des bases de données au calcul haute performance, en passant par l'algorithmique combinatoire. La bioinformatique aide principalement la biologie à exploiter, et parfois à établir, des relations causales entre des phénomènes apparaissant en biologie. De façon plus large, elle vient s'inscrire dans une démarche de modélisation, traditionnellement centrale à la biologie. Elle y permet alors non seulement de compléter l'éventail des méthodes expérimentales, autorisant une validation/invalidation des modèles proposés, mais offre aussi la possibilité d'inférer des modèles compatibles avec les données. Dans ce cas en particulier, elle requiert une confrontation à des grands volumes de données, parfois si énormes que des complexités temps/mémoire surlinéaires sont totalement exclues. Elle soulève donc de nombreuses questions spécifiques, inspirant des développements informatiques *amont*, motivant par exemple des études poussées de complexité algorithmique, ou de problèmes combinatoires dans des fonctions d'évaluation complexes.

La pratique quotidienne de la bioinformatique n'est cependant pas sans difficultés. L'évolution rapide des techniques, des volumes de données disponibles et des

1. Yann Ponty est Chargé de Recherche CNRS au laboratoire LIX de l'École polytechnique à Palaiseau, Yann.Ponty@lix.polytechnique.fr

problématiques biologiques rend parfois difficile la pérennisation des approches algorithmiques. Il n'est d'ailleurs pas rare qu'un traitement *ad hoc* hyper-spécialisé soit préféré à un algorithme élégant à fort potentiel de généralisation, et offrant de meilleures garanties de performances, mais arrivé quelques mois trop tard, ou publié dans des revues plus théoriques que ses compétiteurs... Par ailleurs, l'appétence modérée des biologistes pour les approches trop formelles rend parfois difficile la dissémination des méthodes basées sur une algorithmique non-triviale. Une étude en 2012 de Fawcett et Higginson [11] observe par exemple une corrélation négative entre la densité d'équations et la visibilité d'une étude [2], initiant un débat passionné outre-atlantique [16, 12] sur la nécessité d'introduire un meilleur bagage formel dans les formations en biologie. Cette corrélation n'implique certes pas une causalité, mais elle illustre *a minima* une difficulté récurrente de communication aux interfaces. La dissémination de résultats ayant une dimension informatique, dans des journaux du domaine d'application, s'apparente alors parfois à un exercice de schizophrénie douce, mais très certainement passionnant et enrichissant !

Principe de parcimonie et optimisation combinatoire

En dépit de ces difficultés, de nombreux problèmes algorithmiques bien définis, à l'origine d'outils à l'impact démontré sur la biologie, jalonnent la jeune histoire de la bioinformatique. Les algorithmes, et plus particulièrement les algorithmes d'optimisation combinatoire, y jouent un rôle prépondérant. Des problèmes d'optimisation combinatoire complexes apparaissent en effet très naturellement dans tous les contextes où plusieurs explications existent pour une observation empirique. Le principe de parcimonie, aussi parfois appelé *rasoir d'Ockham*, amène à privilégier, toutes choses étant égales par ailleurs, l'explication la plus simple (ou la plus probable) pour un phénomène observé. L'ensemble des explications admissibles pour l'observé peut être alors assimilé à un espace de recherche, sur lequel on cherchera à optimiser le nombre d'événements atomiques. Ce principe se généralise au principe du maximum de vraisemblance en associant à chaque explication une probabilité, ou vraisemblance, qu'on cherchera à maximiser. Les deux notions coïncident évidemment quand tous les événements sont indépendants et équiprobables, mais ce paradigme permet aussi de modéliser des phénomènes de dépendances, potentiellement complexes, entre les événements.

Un exemple emblématique est celui de l'alignement de séquences, que l'on présente traditionnellement comme la mise en correspondance des positions de deux séquences, de façon à mettre en évidence leurs similarités et différences. Un point de vue parcimonieux sur ce problème consiste à considérer les deux séquences comme homologues, c'est-à-dire des versions contemporaines d'une même séquence ancestrale. On suppose alors que les deux séquences ont évolué indépendamment, chacune

2. Environ 28% de citations en moins, en moyenne, pour chaque équation supplémentaire par page !

subissant une suite d'opérations atomiques (insertion/délétion/mutation d'un caractère). Imaginons maintenant que la séquence ancestrale soit indisponible, et que nous ne disposions que des deux séquences produites par l'évolution. Toutes choses étant égales par ailleurs, un principe de parcimonie nous amènera à favoriser une suite d'opérations faisant apparaître un nombre minimal de modifications. Ce nombre correspond alors exactement à la distance de Levenshtein, efficacement calculable par programmation dynamique dès les années 60 [22]. En modifiant très légèrement les équations de programmation dynamique utilisées pour son calcul, de façon à maximiser la vraisemblance, on obtient l'algorithme de Smith-Waterman [33], qui permet de détecter des similarités locales. Ce dernier fut jadis l'un des algorithmes les plus utilisés de la bioinformatique (cité plus de 7 000 fois), au point de faire l'objet, à son heure de gloire, d'implémentations matérielles dédiées sous la forme de serveurs spécialisés.

Ce paradigme est aussi omniprésent dans les travaux ayant trait à la phylogénie, champs disciplinaire s'attachant à comprendre les mécanismes de l'évolution en inférant des relations d'ancestralité entre espèces, gènes et autres entités biologiques. Ici, l'un des challenges est la reconstruction d'une phylogénie, c'est-à-dire l'inférence d'un arbre de la vie cohérent avec les données contemporaines, tout en induisant un nombre minimal d'événements évolutifs. Par exemple, si l'on souhaite baser la reconstruction sur la séquence associée à un gène (ARN, protéine...) présent dans une famille d'organismes, on pourra choisir une phylogénie induisant un nombre minimal de mutations, insertions et délétions. Des principes similaires gouvernent les algorithmes de prédiction des réarrangements génomiques, l'inférence des réseaux biologiques, la reconstruction des séquences ancestrales, les études d'association génotype/phénotype...

Algorithmique des ARN

Les travaux effectués depuis les années 1970 sur la structure secondaire de l'ARN constituent un exemple d'interaction pluridisciplinaire particulièrement fructueuse, à l'interface entre mathématiques discrètes, biologie, physique, chimie et informatique.

L'ARN, une molécule polyvalente à l'origine de la vie ?

Les Acides RiboNucléiques (ARN) sont des molécules composées de nucléotides Adénine, Cytosine, Guanine et Uracile, assimilables à des séquences sur un alphabet $\{\mathsf{A},\mathsf{C},\mathsf{G},\mathsf{U}\}$. Ils sont transcrits comme des copies de portions de l'ADN, les Thymines de ce dernier y étant remplacées par l'Uracile ($\mathsf{T}\rightarrow\mathsf{U}$). Les molécules d'ARN sont de tailles très variables, et sont encodées par des séquences de tailles variant de 20 nucléotides (nts) à 3 000 nts environ dans la cellule, et pouvant même atteindre jusqu'à 30 000 nts pour les génomes entiers, constitués d'ARN, de certains

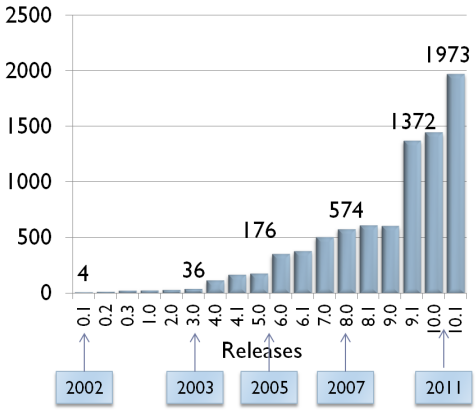


FIGURE 1. Evolution du nombre de familles d'ARN répertoriées au sein de la base de donnée RFAM [17]. Cette croissance est loin d'être ralentie, et la version 11.0 de la base, disponible depuis août 2012, s'est élargie à 2208 familles d'ARN.

virus comme le coronavirus SARS-CoV. Cette diversité de taille est la conséquence directe de la grande variété de rôles joués par l'ARN au sein de la cellule, parmi lesquels :

— Médiateur de l'information génétique : les ARN messagers encodent les protéines, synthétisées comme des chaînes d'acides aminés, chaque acide aminé étant encodé par un triplet de nucléotides. De façon assez remarquable, un unique ARN messager peut parfois encoder simultanément plusieurs protéines fonctionnelles. Plusieurs mécanismes (épissage alternatif, décalage de phase, translecture...) contribuent à ce non-déterminisme, influencé à la fois par des motifs présents dans la séquence, la structure, ou encore les concentrations relatives des différents acteurs du monde cellulaire ;

— Partie-prenante de la machinerie traductionnelle : outre leur rôle de *code source*, les ARN sont aussi acteurs de la traduction des ARN en protéines. Ils participent ainsi au ribosome, un énorme complexe moléculaire composé de nombreuses protéines et d'ARN ribosomaux, qui décodera les ARN messagers en protéines. Des ARN de transfert fournissent aussi la matière première au ribosome, en lui apportant les acides aminés, véritables *briques de base* qui, assemblés linéairement, formeront les protéines ;

— Acteur de la régulation : non seulement les ARN rendent possible la synthèse des protéines, mais ils en gouvernent aussi l'expression quantitative. Par exemple, des petits ARN simple-brin sont au cœur de l'interférence à ARN,

un mécanisme de régulation au cours duquel ils se fixent sur des ARN messagers et y empêchent la fixation du ribosome, inhibant *in fine* la synthèse des protéines encodées.

Cette liste est loin d'être exhaustive, et la base de données RFAM [17], qui recense et organise en familles fonctionnelles les ARN identifiés dans la littérature, a connu une croissance explosive depuis sa création en 2002, comme l'illustre la Figure 1. De plus, l'initiative ENCODE [10], qui a analysé systématiquement 1% du génome humain, a révélé en 2007 qu'une grande majorité (environ 93%) de celui-ci est transcrit en ARN. En extrapolant sur l'ensemble du génome, et en excluant les 30% de gènes codant pour des protéines, dont 2% directement pour les acides aminés, il subsiste une foule d'ARN synthétisés, la plupart sans fonction connue à l'heure actuelle. Si certaines de ces séquences sont sans doute le produit de mécanismes stochastiques assimilables à un *bruit de fond*, il est difficile d'imaginer l'avantage sélectif procuré par une synthèse coûteuse d'ARN à une si grande échelle. Il est donc raisonnable d'anticiper la découverte, dans les décennies qui viennent, de nouvelles fonctions, modes d'action et familles pour l'ARN, à l'instar des 197 nouvelles familles de longs ARN non-codants récemment identifiées [3] dans des régions auparavant ignorées car ne codant pas pour des gènes identifiés.

La diversité fonctionnelle de l'ARN est telle que celui-ci constitue actuellement le support biochimique le plus probable à l'origine de la vie ! En effet, les mécanismes du vivant reposent sur une dichotomie entre gènes qui encodent les éléments fonctionnels et enzymes actrices responsables de la réplication du matériel génétique. Chez les organismes contemporains, les gènes sont codés par l'ADN, et les protéines assument une grande partie des rôles enzymatiques. Or, les capacités enzymatiques de l'ADN sont trop limitées pour permettre une réplication efficace, et les protéines se prêtent mal à une auto-réplication fidèle. Cette situation soulève donc une question cruciale :

> Comment *démarrer* un système (la vie) dont chacun des deux acteurs principaux (ADN et protéines) semble avoir besoin de l'autre pour exister et se reproduire ?

L'ARN résout cet apparent *paradoxe de l'œuf et de la poule* en cumulant capacité de stockage, les génomes de certains virus étant par exemple entièrement constitués d'ARN, et fonctions enzymatiques autorisant l'auto-réplication. Un *monde prébiotique à ARN* constitue donc actuellement l'hypothèse la plus séduisante pour l'origine de la vie, et fait l'objet des travaux de toute une communauté mêlant biologistes moléculaires et biochimistes [4].

Comprendre la structure de l'ARN pour en comprendre la fonction

À la différence des ADN, les ARN sont synthétisés sous la forme d'une copie simple, aussi appelée simple-brin, et n'adoptent donc pas nécessairement la structure en double-hélice qui caractérise l'ADN. Au contraire, l'ARN adopte des formes complexes en se repliant sur lui-même. Il est ainsi soumis à des fluctuations à échelle

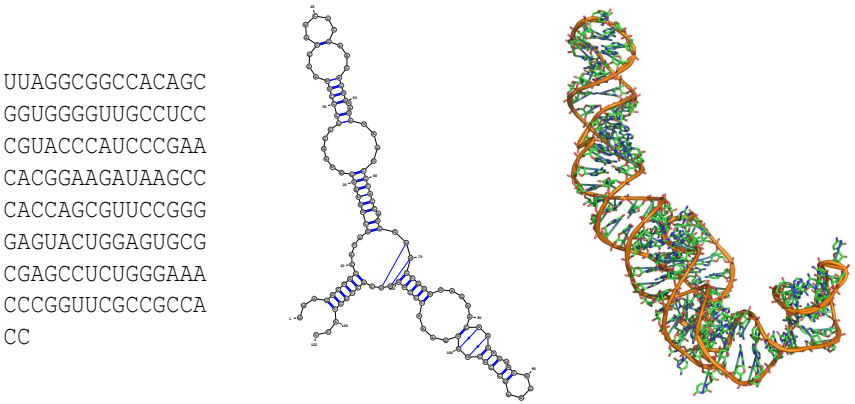


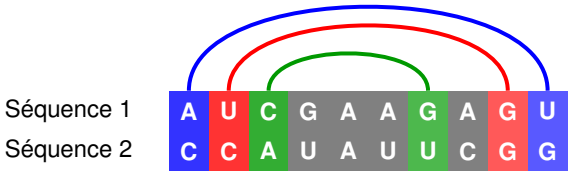
FIGURE 2. Trois principaux niveaux de représentation pour un ARN ribosomal (Id. PDB : 1K73 :B) correspondant aux étapes de structuration dans l'hypothèse d'un repliement hiérarchique.

nanométrique, et se retrouve stabilisé dans des conformations par l'appariement de certains de ces nucléotides via la formation de liaisons hydrogènes. Bien que presque toutes les paires de nucléotides soient susceptibles de former un appariement, les plus stabilisatrices sont les paires dites canoniques, parmi lesquelles on distingue les paires Watson-Crick (G–C ou A–U) et Wobble (G–U). Il résulte de ce processus une (parfois plusieurs) forme tridimensionnelle complexe essentielle, chez la plupart des ARN, pour la réalisation d'une fonction spécifique.

Il est généralement admis que le processus du repliement des ARN s'effectue de façon hiérarchique [34] comme l'illustre la Figure 2. Initialement, l'ARN se replie sur lui-même, établissant un ensemble d'appariements canoniques sans croisement, donnant lieu à une structure arborescente appelée la structure secondaire. Cette dernière peut être représentée de nombreuses façons équivalentes. Dans un deuxième temps, la forme tri-dimensionnelle adoptée par l'ARN rend possible la formation de motifs plus faiblement stabilisateurs. Parmi ceux-ci, on trouve des liaisons non-canoniques et des motifs topologiques complexes appelés pseudo-nœuds, constitués de couples d'appariements en situation de croisement quand dessinés sur le demi-plan supérieur. La prédiction de la forme adoptée par un ARN à l'issu de son repliement commence donc souvent par la prédiction d'une (ou plusieurs) structure(s) secondaire(s) candidate(s) pour celui-ci. Cette prédiction initiale est ensuite complétée par la modélisation des éléments supplémentaires, et enfin par l'agencement

tri-dimensionnel de celle-ci.

Chez les ARN non-codants, qui ne sont pas traduits en protéines, la structure secondaire constitue souvent le principal déterminant de la fonction. La structure est en effet tout autant, voire parfois plus, préservée par l'évolution que la séquence précise des nucléotides. Comme le montre l'*exemple jouet* ci-dessous, deux séquences très différentes peuvent servir de support à une même structure secondaire, à travers des substitutions locales préservant les appariements :



La structure secondaire est donc au cœur de nombreuses approches basées sur la conservation d'une structure commune parmi une famille de molécules homologues. Par exemple, elle aide à rechercher de nouvelles occurrences d'un ARN dans un ou plusieurs génomes, ou encore à prédire le repliement commun à un ensemble de séquences d'ARN. La prédiction de la structure secondaire fonctionnelle d'un ARN constitue donc une première étape indispensable pour comprendre son (ou ses) mode(s) d'action(s), et le replacer dans le contexte plus large des systèmes biologiques.

Représentations et propriétés combinatoires de la structure secondaire d'ARN

Du fait de la contrainte de non-croisement, la structure secondaire est analogue à un arbre, une propriété fort sympathique d'un point de vue algorithmique. Par ailleurs, des contraintes à la fois géométriques et biochimiques interdisent l'appariement canonique de positions consécutives dans un ARN, matérialisant une difficulté des biopolymères à effectuer des *demi-tours complets* dans leur trajectoire tridimensionnelle. Il en résulte donc une contrainte de distance minimale $\theta \geq 1$ entre des positions appariées.

La structure arborescente sous-jacente à la structure secondaire autorise de nombreuses représentations et caractérisations équivalentes (voir Figure 3), fournissant une interface naturelle avec de nombreuses communautés de recherche en informatique et en mathématiques discrètes :

(1) un mot de Motzkin, c'est-à-dire une expression bien parenthésée sur l'alphabet $\{(,),\bullet\}$, dans laquelle on interdit le motif $()$;

(2) un graphe planaire extérieur biconnexe de degré maximal $\Delta \leq 3$;

(3) un arbre unaire-binaire, dont les pères des feuilles gauches sont unaires ;

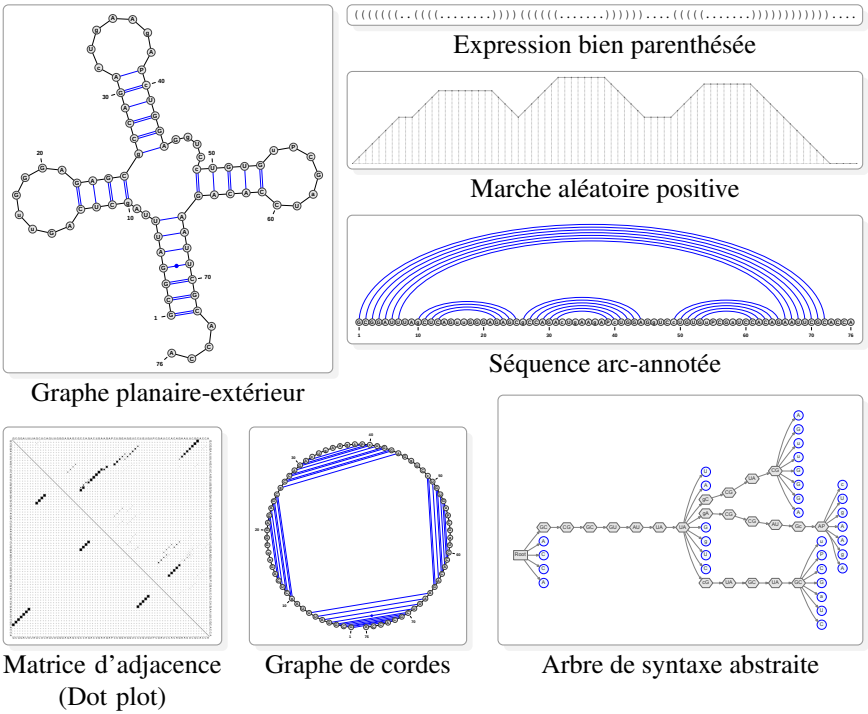


FIGURE 3. Représentations alternatives (équivalentes) de la structure secondaire d'un ARN de transfert.

(4) un arbre de syntaxe abstraite associé à la grammaire hors-contexte

$$S \to (\, S^{\overline{\varepsilon}} \,)\, S \mid \bullet S \mid \varepsilon \qquad\qquad S^{\overline{\varepsilon}} \to (\, S^{\overline{\varepsilon}} \,)\, S \mid \bullet S\,;$$

(5) un ensemble stable de sommets (non-adjacents) d'un graphe de cordes, défini comme le graphe d'intersection d'un ensemble de cordes dessinées sur un cercle ;

(6) une marche (aléatoire) positive à pas dans l'ensemble

$$\{\mathtt{U} : (+1,+1),\ \mathtt{D} : (+1,-1),\ \mathtt{H} : (+1,0)\},$$

commençant et finissant à l'ordonnée 0, et excluant le motif *pyramidal* `UD`...

Ces représentations diverses permettent la formation d'intuitions algorithmiques complémentaires, ainsi que l'utilisation d'approches classiques dans différentes disciplines, afin de mieux comprendre et analyser ces objets. Par exemple, à l'interface

des mathématiques discrètes et de la biologie, l'arsenal de la combinatoire énumérative/analytique peut être utilisé pour prouver que le nombre de structures secondaires compatibles avec un ARN est, en moyenne, exponentiel sur la taille de celui-ci [3].

Prédiction du repliement par maximisation du nombre d'appariements

Une première question algorithmique se pose naturellement : Comment prédire la structure secondaire la plus stable pour une séquence d'ARN donnée ? À l'équilibre thermodynamique, celle-ci est en effet la plus probablement observée, et il est donc raisonnable d'imaginer que l'évolution ait favorisé celle-ci comme conformation fonctionnelle. Le problème d'optimisation combinatoire associé est appelé prédiction de la structure secondaire par minimisation de l'énergie libre, notion sur laquelle nous reviendrons dans la section suivante. Il exclut *a priori* toute approche gloutonne, et l'explosion du nombre de structures secondaires compatibles avec une séquence d'ARN interdit une énumération exhaustive des structures secondaires candidates.

À la fin des années 70, un premier algorithme est proposé par Nussinov *et al* [29] dans un modèle où le nombre d'appariements canoniques est considéré comme indicateur de la stabilité. La minimisation de l'énergie libre y revient donc à maximiser le nombre de paires de bases. Le problème informatique associé peut alors être résolu efficacement par une technique appelée programmation dynamique : on commence par calculer efficacement le nombre maximal d'appariements pour l'ARN par le biais d'une récurrence, puis on reconstruit une structure secondaire possédant de tels appariements en retraçant les étapes de la récurrence ayant permis d'obtenir le score maximal.

Plus précisément, l'algorithme de Nussinov part de l'observation que toute structure secondaire, compatible avec une séquence d'ARN donnée, possède (au moins) une des propriétés suivantes :

— les premier et dernier nucléotides sont appariés, sous réserve de compatibilité des nucléotides présents à ces positions ;
— le premier (resp. dernier) nucléotide est non-apparié ;
— la structure secondaire se partitionne en deux sous-structures indépendantes, c'est-à-dire ne partageant pas d'appariement.

Il est alors possible d'étendre ce raisonnement à toute sous-séquence définie sur un intervalle de la séquence d'ARN. On en déduit une récurrence permettant de calculer, en temps/mémoire $\Theta(n^3)/\Theta(n^2)$, le nombre maximal de paires de bases réalisables sur chaque intervalle d'une séquence d'ARN composée de n nucléotides. Une fois ces valeurs calculées, il est possible de retracer les étapes du calcul pour retrouver l'ensemble de paires de bases ayant contribué à maximiser le nombre d'appariements, et d'en déduire ainsi une structure secondaire optimalement stable.

3. Plus précisément, le nombre de structures secondaires compatibles avec un ARN composé de n nucléotides admet un équivalent asymptotique de l'ordre de $1.87^n/n\sqrt{n}$ [41].

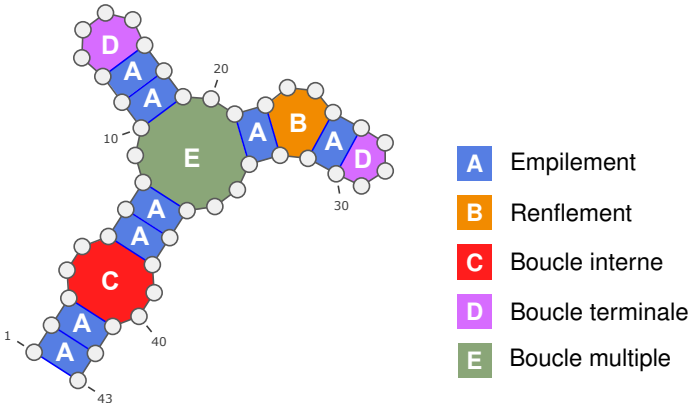


FIGURE 4. Principaux éléments de la décomposition en k-boucles d'une structure secondaire d'ARN. Un appariement *virtuel* peut être ajouté entre les premières et dernières positions afin de prendre en compte d'éventuels nucléotides non-appariés sur la face extérieure.

De façon intéressante, l'algorithme proposé par Nussinov est extrêmement similaire à celui proposé indépendamment par Gavril [15] pour la recherche d'un ensemble stable de cardinalité maximale dans les graphes de cordes. On y retrouve aussi une spécialisation de l'algorithme CYK [20] permettant l'analyse syntaxique à partir de grammaires algébriques, dans sa version pondérée populaire en traitement automatique du langage. Ce lien entre bioinformatique des ARN et traitement du langage naturel est aussi à l'origine d'avancées récentes.

Au-delà du discret : modèle de Turner et MFold

Dans le modèle de stabilité considéré par l'algorithme de Nussinov, seules les paires de bases contribuent à la stabilité d'un repliement. Or la réalité biochimique des ARN est bien plus complexe, et l'ARN est stabilisé par de nombreux motifs structurels. Ceux-ci interviennent dans le calcul de l'énergie libre, définie de façon relative, une structure étant d'autant plus stable que son énergie est négative.

En particulier, le modèle de Turner décompose la structure secondaire en occurrences de motifs structure/séquence, associés à des contributions énergétiques, comme l'illustre la Figure 4. Ces dernières sont typiquement mesurées empiriquement, au travers d'ingrates et exhaustives expériences systématiques de spectrophotométrie [36]. On obtient alors l'énergie libre d'une structure secondaire en additionnant les contributions des motifs y apparaissant.

Un fossé menace alors de se creuser entre les biochimistes, qui utilisent en pratique le modèle de Turner de façon heuristique et largement manuelle, et les proto-bioinformaticiens (informaticiens et mathématiciens de la première heure) de l'ARN, qui privilégient une optimisation exacte dans un modèle désormais considéré comme naïf. Heureusement, en 1981, Zuker et Stiegler [42] réconcilient le domaine, en remarquant que le modèle de Turner peut être complètement pris en compte via une modification subtile de l'algorithme de Nussinov.

Plus précisément, ces auteurs proposent une décomposition canonique des structures secondaires en un ensemble couvrant de k-boucles, illustrée par la Figure 4. Ils remarquent alors que les contributions énergétiques primitives du modèle de Turner peuvent être associées aux occurrences des k-boucles dans la décomposition. Ils en déduisent alors un ensemble de récurrences permettant de calculer l'énergie libre minimale sur chaque intervalle, puis un algorithme de programmation dynamique permettant la reconstruction de la structure d'énergie libre minimale dans le modèle de Turner. En imposant des restrictions réalistes sur les tailles de certains motifs, le calcul des récurrences peut être effectué en temps/mémoire $\Theta(n^3)/\Theta(n^2)$. On obtient alors l'algorithme `MFold` [42], un des plus grands succès de la bioinformatique, dont les cinq principales publications totalisent plus de 15 000 citations [4], témoignant d'un usage quasi-quotidien du logiciel par les biologistes et bioinformaticiens de l'ARN.

Structures sous-optimales et paradigme thermodynamique

Malgré la grande popularité de `MFold`, largement méritée de part sa rapidité et la fiabilité générale de ses prédictions — allant de 60% [14] à 73% [25] de sensibilité en moyenne selon les bases de séquences considérées —, la méthode rencontre des difficultés pour prédire certaines classes d'ARN :

— Certaines de ces difficultés sont de nature intrinsèque, et liées à un modèle d'énergie établi expérimentalement, mettant en jeu des valeurs extrapolées à partir de plus petits motifs ou de températures/concentrations ioniques différentes des conditions biologiques ;

— Par ailleurs, l'ARN se replie *in vivo* en interaction avec son environnement, constitué de protéines, d'ARN et de petites molécules, qui peuvent *guider* son repliement vers des structures *a priori* peu stables. Certaines catégories d'ARN voient même certains de leur nucléotides modifiés par des enzymes, perturbant ainsi leurs propriétés physico-chimiques, ce qui peut déstabiliser certaines structures ;

— Enfin, la structure secondaire, en ignorant les couples d'appariements en situation de croisement (pseudo-nœuds), néglige des contributions potentiellement stabilisatrices.

4. Source : Google Scholar, avril 2014.

Tous ces facteurs sont susceptibles de renverser l'ordre relatif de stabilité, perçu par l'algorithme, pour deux structures secondaires compatibles avec une séquence.

Une première approche pour contourner ce problème consiste à considérer les structures sous-optimales. Zuker propose en 1989 un premier algorithme [39] pour la génération des structures *P*-optimales, l'ensemble des structures stables faisant apparaître certains appariements jugés crédibles. Cette définition possède comme mérite manifeste d'éviter une explosion du nombre de structures candidates. Elle interdit cependant une étude systématique de l'ensemble des sous-optimaux, et ne permet pas de déterminer, par exemple, si la structure d'énergie minimale est isolée ou, au contraire, si elle est concurrencée par d'autres structures. Des travaux de l'école Viennoise de biochimie théorique viendront, une décennie plus tard, compléter cette première définition en proposant un algorithme de génération exhaustive des structures secondaires sous-optimales à différence d'énergie [38].

Une alternative très prisée actuellement propose un changement de paradigme complet, basé sur un concept central de la physique statistique, l'équilibre de Boltzmann. En effet, en focalisant sur la structure d'énergie minimale, on oublie que le phénomène du repliement est un phénomène intrinsèquement stochastique. La stabilité d'un repliement est certes importante, mais la probabilité d'observer un ARN dans une topologie donnée dépend aussi du nombre de conformations de stabilités comparables, présentes au sein des structures possibles pour l'ARN. En postulant un équilibre thermodynamique de Boltzmann, on suppose que l'ARN est resté suffisamment longtemps dans des conditions stables, et que la probabilité d'observer une structure secondaire $S \in \mathscr{S}_\omega$ pour un ARN ω est donnée par

$$\mathbb{P}(S \mid \omega) = \frac{e^{\frac{-E_\omega(S)}{kT}}}{\mathscr{Z}_\omega} \qquad \{\text{Distribution de Boltzmann}\}$$

$$\text{où} \quad \mathscr{Z}_\omega = \sum_{S \in \mathscr{S}_\omega} e^{\frac{-E_\omega(S)}{kT}}, \qquad \{\text{Fonction de partition de } \omega\}$$

où k est la constante de Boltzmann et T la température en Kelvins. L'équilibre de Boltzmann peut aussi être interprété comme la distribution stationnaire d'un processus Markovien modélisant la dynamique du repliement.

En pondérant exponentiellement les structures selon leur énergie, cette distribution opère un compromis subtil entre la stabilité des molécules et l'explosion combinatoire du nombre de conformations sous-optimales. La structure d'énergie minimale reste certes la plus probable, mais sa probabilité décroît exponentiellement sur la taille des séquences considérées. Il en résulte une diversité de paysages thermodynamiques bien plus riche que ne l'aurait laissé présager le simple examen de la structure d'énergie minimale. Une telle distribution permet alors de donner un sens mathématique rigoureux aux questions suivantes :

— Le processus du repliement est-il *bien défini* ? [40]

— Combien y a-t-il de conformations *principales* dans l'ensemble ? [9]
— Quels sont les appariements récurrents dans l'ensemble de Boltzmann ? [26]
— Fournissent-ils une indication de qualité pour les prédictions ? [25]
— Quelles mutations sont les plus susceptibles de perturber la fonction ? [37]
— Quelle distribution de distance à une structure de référence ? [13]

Le verrou algorithmique principal pour répondre à ces questions concerne le calcul de la fonction de partition, *a priori* complexe car défini ci-dessus comme la somme d'un nombre exponentiel de termes. Or, John McCaskill remarque en 1990 [26] que, sous certaines hypothèses techniques, un schéma de programmation dynamique pour la minimisation d'énergie peut être transformé *syntaxiquement*, à travers un simple changement d'algèbre, en un algorithme d'efficacité comparable pour calculer la fonction de partition. Des réponses algorithmiques aux questions posées ci-dessus sont alors obtenues par des modifications de ce schéma, ou à travers des méthodes d'échantillonnage exact elles-aussi dérivées du calcul de la fonction de partition.

L'évolution comme alliée : approches comparatives

Malgré les gains substantiels, à la fois en qualité des prédictions et en quantification de la confiance des prédictions [25], induits par ce changement de paradigme, la recherche d'un repliement hautement précis pour un nouvel ARN restait un problème ardu. Heureusement, à l'heure où les approches énergétiques atteignaient leurs limites, des *approches comparatives*, l'exploitation des propriétés évolutives de l'ARN, allaient permettre un nouveau bond en avant. En effet, l'hypothèse d'une relation forte entre structure et fonction motive les études structurelles mais, en renversant le point de vue, permet aussi de postuler que des ARN *homologues*, ayant une fonction commune, partagent très probablement une structure commune.

On peut donc déjà exploiter cette remarque pour améliorer notre capacité de prédiction, dès lors que des ARN homologues sont identifiés. On se pose alors le problème de la prédiction d'une structure secondaire commune à deux ou plusieurs séquences. Si celles-ci sont déjà alignées, alors l'exécution d'un algorithme de type `MFold`, légèrement étendu pour incorporer les contributions simultanées des séquences, permet la prédiction d'une structure consensus pour une famille d'ARN préalablement alignés [18]. Cependant, cette situation est assez rare, et l'absence de contraintes fonctionnelles fortes s'exprimant sur la séquence entraîne souvent une évolution rapide de celle-ci. La prédiction d'un alignement est alors délicate tant que la structure commune reste inconnue, or la prédiction de celle-ci est justement notre objectif premier !

Pour sortir de ce paradoxe de l'œuf et de la poule, David Sankoff propose en 1985 [31] un algorithme pour le repliement/alignement simultané d'une famille de

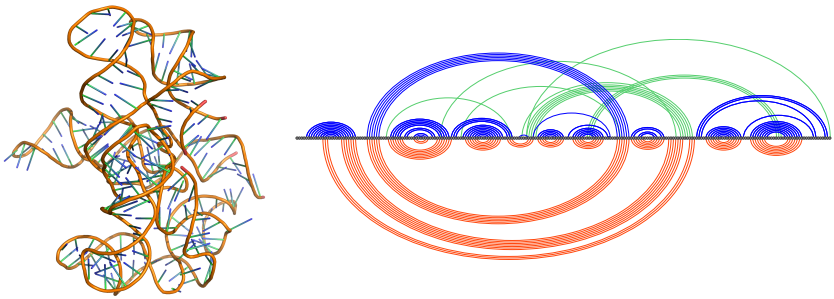


FIGURE 5. Pour ce ribozyme du groupe I (PDB [5] ID : 1Y0Q :A, gauche), les performances d'une prédiction par minimisation d'énergie sont affectées par l'omission des appariements en situation de croisement. Un logiciel de type `MFold` prédira donc des appariements (bas) substantiellement différents, à la fois de ceux lisibles dans la structure 3D (haut), et de leur sous-ensemble sans croisement de cardinalité maximale (haut, trait gras).

séquences optimisant la somme des scores d'alignement et des énergies de repliement. De complexité exponentielle pour un nombre non-borné de séquences, il se simplifie en un algorithme en $\Theta(n^6)/\Theta(n^4)$ pour deux séquences. Malgré sa complexité élevée, cette catégorie d'approches permet d'obtenir un tel gain de qualité sur les prédictions réalisées qu'elle a aujourd'hui la faveur des bioinformaticiens de l'ARN [14]. Cette popularité a d'ailleurs motivé de nombreux compromis algorithmiques afin de rendre réaliste l'exécution de cette approche à grande échelle [35].

Cette approche comparative est aussi utilisée afin de rechercher des homologues, dans les génomes, à des ARN issus d'une famille déja identifiée. Une fois un alignement obtenu, il est alors possible de construire une structure secondaire consensus pour la famille. En conjonction avec l'alignement multiple, cette dernière permet alors la construction d'un modèle de covariation [28], c'est-à-dire une grammaire stochastique grâce à laquelle de nouvelles séquences peuvent être rapatriées des bases de données publiques. Ces nouvelles séquences viennent alors enrichir l'alignement, amorçant un cercle vertueux qui a rendu possible la caractérisation de nombreuses familles d'ARN, et constituent donc l'un des piliers de la base de données RFAM [17].

Enrichir les espaces de conformations

Parallèlement à ces développements, une direction de recherche orthogonale s'est développée, visant à explorer des espaces de recherche plus riches. L'hypothèse d'un repliement hiérarchique, initialement sans croisement, n'est en effet pas une règle absolue, et de nombreux ARN contiennent des nœuds, pseudo-nœuds et autres motifs

topologiques complexes (voir Figure 5). Ces motifs, qui sont représentés par des ensembles d'appariements en croisement, sont parfois indispensables pour qu'un ARN remplisse sa fonction. De plus, ils influencent fortement la modélisation 3D en limitant ses degrés de liberté, et leur prédiction constitue donc une étape préliminaire essentielle vers des prédictions de structure à résolution atomique.

Cependant, autoriser des appariements en croisement soulève des problèmes algorithmiques importants. Dans le modèle de Nussinov maximisant le nombre de paires de bases, ou leur poids total, le problème de prédiction d'une structure minimisant l'énergie avec pseudo-nœuds peut être résolu en temps polynomial comme une instance particulière de la recherche, dans un graphe, d'un ensemble stable (pondéré) maximal [8]. Cependant, le problème devient NP-difficile [24] dès qu'un modèle d'énergie, analogue à celui de Turner, est pris en considération, et le reste même dans un modèle se limitant aux empilements d'appariements [23]. Enfin, les modèles d'énergie les plus expressifs semblent même exclure (à moins que $\mathsf{P} = \mathsf{NP}$) tout schéma d'approximation polynomial de ratio constant [32].

Cette situation, en apparence désespérée, a suscité des travaux où l'espace de recherche se prête *par construction* à la programmation dynamique, le tout en couvrant la plus grande proportion possible des structures rencontrées dans la littérature. Akutsu [1] propose ainsi un algorithme exact en $\Theta(n^4)$ permettant de prédire des pseudo-nœuds simples. S'ensuivent alors de nombreux travaux cherchant à optimiser le compromis expressivité/efficacité. En particulier, des contributions élégantes, issues de la physique théorique, proposent de considérer le genre topologique, un paramètre associé à la carte obtenue en dessinant tous les appariements de la structure secondaire dans le demi-plan supérieur [6]. On constate alors que les structures connues ont une valeur faible pour le genre, motivant la conception d'algorithmes pour le repliement en genre limité. Le problème est cependant loin d'être résolu, et fait encore l'objet de développements supplémentaires, tant sur l'algorithmique qu'au niveau des modèles d'énergie/conformations sous-jacents, dans un dialogue fructueux à l'interface de la biochimie et de l'optimisation combinatoire.

Des travaux très similaires permettent aussi de prédire les interactions ARN-ARN. En effet, celles-ci se modélisent assez naturellement comme la conjonction de deux repliements *internes* à chacun des ARN, et d'un ensemble d'appariements *interactions*, reliant des positions laissées libres dans les deux structures. Malheureusement, le problème de la prédiction simultanée des repliements et interactions devient alors NP-difficile, même dans sa version purement combinatoire, où l'on cherche à maximiser le nombre d'appariements, où chacune des deux structures *internes* est une simple structure secondaire, et où les appariements en interaction sont eux-mêmes sans croisement [2]. De même que pour la prédiction des pseudo-nœuds, la communauté recherche actuellement une stratégie de contournement pour ce résultat négatif, explorant des approches heuristiques [7] et des résolutions exactes sur des espaces de recherche restreints [27].

Une autre direction pour l'extension des espaces de conformations consiste à considérer, non seulement les traditionnels appariements G-C, A-U et G-U, mais aussi les appariement non-canoniques [21]. Bien que ces derniers soient plus rares, et *a priori* moins contributifs à la stabilité d'une conformation, ils s'organisent en modules bien identifiés, récurrents, conservés par l'évolution, et pour la plupart fonctionnels. De plus, leur prédiction fournirait un réseau de contraintes plus riches, permettant d'éliminer des degrés de liberté dans la recherche d'une conformation 3D. Un modèle de conformation étendant la notion de k-boucle a en effet été proposé en 2008 par Major et Parisien [30], permettant alors de reconstruire la conformation 3D d'ARN connus à des résolutions record. Ces travaux prometteurs peinent cependant à être étendus et systématisés, malgré l'existence d'un cadre algorithmique unifié [19]. Ce retard tient principalement au fait que l'absence de données thermodynamiques associées, en conjonction avec une population de modèles 3D déterminés expérimentalement bien plus modeste pour l'ARN que pour les protéines, rend délicate, voire irréaliste à l'heure actuelle, la mise en œuvre d'approches basées sur des potentiels statistiques.

Conclusion et discussion

Pour terminer ce petit tour d'horizon (non exhaustif) des problématiques et avancées algorithmiques en bioinformatique des ARN, je voudrais rappeler la constante confrontation interdisciplinaire qui jalonne l'histoire de la bioinformatique des ARN, et conclure sur l'importance des contributions, passées et à venir, de l'algorithmique au sein d'une démarche de modélisation.

Comme ont pu l'illustrer les nombreuses références qui émaillent ce texte, la bioinformatique des ARN a fait l'objet de nombreuses contributions aux interfaces entre l'informatique et plusieurs domaines scientifiques. Des biochimistes expérimentaux y ont collaboré avec des bioinformaticiens pour produire un modèle d'énergie entièrement paramétré, le modèle de Turner, à un niveau de granularité permettant une algorithmique de prédiction efficace. Des physiciens y ont contribué par des concepts permettant une classification topologique des structures déterminée expérimentalement, plus tard transformés en algorithmes efficaces pour l'optimisation dans des espaces de conformations expressifs. Des spécialistes de l'évolution ont étudié la relation structure/séquence chez l'ARN comme un modèle de la théorie de l'évolution neutre, résultant un peu par hasard en les premiers outils et algorithmes pour la conception d'ARN synthétiques (Design d'ARN). La biologie moléculaire y suscite et utilise des outils bioinformatiques pour comprendre de nouveaux mécanismes de régulation impliquant l'ARN, parmi lesquels l'interférence par ARN, à l'origine de nombreuses promesses et objet du prix Nobel de physiologie et médecine attribué à Craig et Mello en 2006.

Toutes ces avancées ont été rendues possibles par de nombreuses contributions, issues de champs très variés de l'informatique. Historiquement, l'utilisation de langages formels a permis la recherche de nouveaux ARN dans les génomes, ou encore la modélisation d'espaces de repliements expressifs. Des concepts issus du traitement automatique du langage ont fourni un cadre probabiliste, alternatif à l'approche thermodynamique, dont s'inspirent encore toute une classe d'approches en vogue pour l'étude du repliement. La théorie algorithmique des graphes a fourni un langage pour une reformulation claire, mathématiquement bien définie, de nombreux problèmes. La combinatoire a rendu possible une analyse en moyenne des structures d'ARN aléatoires et des algorithmes œuvrant sur celles-ci, et a fourni un cadre privilégié pour la conception des algorithmes de programmation dynamique omniprésents en bioinformatique. L'optimisation combinatoire, et plus particulièrement la programmation mathématique, est de plus en plus fréquemment utilisée pour contourner les résultats de difficulté algorithmique apparaissant, entre autres, dans la prédiction de structure/appariements, l'étude de la cinétique ou le design d'ARN fonctionnels. Enfin, des techniques de fouille de données, reposant sur des structures d'indexation succinctes issues de l'algorithmique du texte, jouissent d'une popularité croissante pour la détection de nouveaux gènes ARN à travers plusieurs génomes. La recherche en bioinformatique des ARN est donc une recherche profondément pluridisciplinaire, et d'impact démontré sur les pratiques des biologistes, y compris à un niveau fondamental.

En effet, éminemment empiriques, les sciences du vivant reposent de façon cruciale sur la notion de modèle, permettant d'exprimer en des termes testables et réfutables une explication avancée pour un phénomène d'intérêt. Des algorithmes exacts y jouent un rôle essentiel dans la validation/invalidation des modèles comme explication de la réalité. En effet, un modèle n'est qu'une construction logique tant que ses conséquences, les propriétés dérivées sous l'hypothèse de celui-ci, n'ont pas été confrontées à une réalité observable. Si les conséquences du modèle sont compatibles avec l'observation empirique, alors la force d'explication du modèle, et notre foi en celui-ci, s'en trouvent renforcées. En cas de contradiction apparente, le modèle doit être remis en question, raffiné, ou tout simplement abandonné.

C'est ici qu'une algorithmique exacte ou, à défaut, offrant des garanties théoriques, devient indispensable : Quelle conclusion tirer d'une incompatibilité entre la réalité observable et les conséquences, estimées bien approximativement et sans garanties théoriques, d'un modèle ? S'agit-il juste d'un artefact lié à un calcul incorrect ? En l'absence d'algorithmes bien conçus, et calculables en temps raisonnable, les modèles formulés en biologie deviennent donc bien difficiles à valider et réfuter. Malheureusement, certains modèles discrets s'avèrent extrêmement difficiles à analyser, en particulier quand ceux-ci sont associés à des problèmes algorithmiques difficilement résolubles au sens de la théorie de la complexité. Les choix de modélisation devraient donc être guidés, non seulement par une connaissance intime

des phénomènes étudiés et des données disponibles, mais aussi par une culture algorithmique suffisamment large pour guider la recherche d'un *modèle compromis*, conjuguant expressivité suffisante et espoir de résolution exacte. Une telle recherche justifie, à elle seule, la poursuite et le renforcement d'un dialogue pluridisciplinaire, d'ores et déjà fructueux, à l'interface entre l'informatique et les sciences du vivant.

Remerciements

L'auteur tient à remercier chaleureusement Christine Froidevaux et Alain Denise pour leurs relectures attentives de versions préliminaires de ce document.

Références

[1] T. Akutsu. Dynamic programming algorithms for RNA secondary structure prediction with pseudoknots. *Discrete Appl. Math.*, 104(1-3) :45–62, 2000.

[2] C. Alkan, E. Karakoc, J. H. Nadeau, S C. Sahinalp, and K. Zhang. RNA-RNA interaction prediction and antisense RNA target search. *J Comput Biol*, 13(2) :267–282, March 2006.

[3] P. P. Amaral, M. B. Clark, D. K. Gascoigne, M. E. Dinger, and J. S. Mattick. lncRNAdb : a reference database for long noncoding RNAs. *Nucleic Acids Res*, 39(Database issue) :D146–D151, January 2011.

[4] J. Atkins, R. Gesteland, and T. Cech. *RNA Worlds : From Life's Origins to Diversity in Gene Regulation*. Cold Spring Harbor Perspectives in Biology. Cold Spring Harbor Laboratory Press, 2011.

[5] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne. The protein data bank. *Nucleic Acids Res*, 28(1) :235–242, January 2000.

[6] M. Bon, G. Vernizzi, H. Orland, and A. Zee. Topological classification of RNA structures. *J Mol Biol*, 379(4) :900–911, June 2008.

[7] A. Busch, A. S. Richter, and R. Backofen. IntaRNA : efficient prediction of bacterial sRNA targets incorporating target site accessibility and seed regions. *Bioinformatics*, 24(24) :2849–2856, 2008.

[8] R. B. Cary and G. D. Stormo. Graph-theoretic approach to RNA modeling using comparative data. *Proceedings International Conference on Intelligent Systems for Molecular Biology*, 3 :75–80, 1995.

[9] Y. Ding, C. Y. Chan, and C. E. Lawrence. RNA secondary structure prediction by centroids in a boltzmann weighted ensemble. *RNA*, 11(8) :1157–1166, August 2005.

[10] ENCODE Project Consortium. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature*, 447(7146) :799–816, June 2007.

[11] T. W. Fawcett and A. D. Higginson. Heavy use of equations impedes communication among biologists. *Proceedings of the National Academy of Sciences*, 109(29) :11735–11739, 2012.

[12] T. W. Fawcett and A. D. Higginson. Reply to chitnis and smith, fernandes, gibbons, and kane : Communicating theory effectively requires more explanation, not fewer equations. *Proceedings of the National Academy of Sciences*, 109(45) :E3058–E3059, 2012.

[13] E. Freyhult, V. Moulton, and P. Clote. Boltzmann probability of RNA structural neighbors and riboswitch detection. *Bioinformatics*, 23(16) :2054–2062, August 2007.

[14] P. P. Gardner and R. Giegerich. A comprehensive comparison of comparative RNA structure prediction approaches. *BMC Bioinformatics*, 5 :140, September 2004.

[15] F. Gavril. Algorithms for a maximum clique and a maximum independent set of a circle graph. *Networks*, 3(3) :261–273, 1973.

[16] J. Gibbons. Do not throw equations out with the theory bathwater. *Proceedings of the National Academy of Sciences*, 109(45) :E3054, 2012.

[17] S. Griffiths-Jones, A. Bateman, M. Marshall, A. Khanna, and S. R. Eddy. Rfam : an RNA family database. *Nucleic Acids Res*, 31(1) :439–441, January 2003.

[18] I. L. Hofacker, M. Fekete, and P. F. Stadler. Secondary structure prediction for aligned RNA sequences. *J Mol Biol*, 319(5) :1059–1066, June 2002.

[19] C. Höner zu Siederdissen, S. H. Bernhart, P. F. Stadler, and I. L. Hofacker. A folding algorithm for extended RNA secondary structures. *Bioinformatics*, 27(13) :i129–i136, July 2011.

[20] T. Kasami. An efficient recognition and syntax-analysis algorithm for context-free languages. Scientific report AFCRL-65-758, Air Force Cambridge Research Lab, Bedford, MA, 1965.

[21] N. B. Leontis and E. Westhof. Geometric nomenclature and classification of RNA base pairs. *RNA*, 7(4) :499–512, April 2001.

[22] V. I Levenshtein. Binary codes capable of correcting deletions, insertions, and reversals. *Soviet Physics Doklady*, 10 :707–10, 1966.

[23] R. B. Lyngsø. Complexity of pseudoknot prediction in simple models. In *Proceedings of ICALP*, pages 919–931, 2004.

[24] R. B. Lyngsø and C. N. Pedersen. RNA pseudoknot prediction in energy-based models. *J Comput Biol*, 7(3-4) :409–427, 2000.

[25] D. H. Mathews, J. Sabina, M. Zuker, and D. H. Turner. Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J Mol Biol*, 288(5) :911–940, May 1999.

[26] J. S. McCaskill. The equilibrium partition function and base pair binding probabilities for RNA secondary structure. *Biopolymers*, 29(6-7) :1105–1119, 1990.

[27] U. Mückstein, H. Tafer, J. Hackermüller, S. H. Bernhart, P. F. Stadler, and I. L. Hofacker. Thermodynamics of RNA-RNA binding. *Bioinformatics*, 22(10) :1177–1182, May 2006.

[28] E. P. Nawrocki, D. L. Kolbe, and S. R. Eddy. Infernal 1.0 : inference of RNA alignments. *Bioinformatics*, 25(10) :1335–1337, May 2009.

[29] R. Nussinov, G. Pieczenik, J. Griggs, and D. Kleitman. Algorithms for loop matchings. *SIAM Journal on Applied Mathematics*, 35(1) :68–82, 1978.

[30] M. Parisien and F. Major. The MC-Fold and MC-Sym pipeline infers RNA structure from sequence data. *Nature*, 452(7183) :51–55, March 2008.

[31] D. Sankoff. Simultaneous solution of the RNA folding, alignment and protosequence problems. *SIAM Journal on Applied Mathematics*, 45(5) :810–825, October 1985.

[32] S. Sheikh, R. Backofen, and Y. Ponty. Impact of the energy model on the complexity of RNA folding with pseudoknots. In J. Krkkinen and J. Stoye, editors, *Combinatorial Pattern Matching*, volume 7354 of *Lecture Notes in Computer Science*, pages 321–333. Springer Berlin Heidelberg, 2012.

[33] T. Smith and M. Waterman. Identification of common molecular subsequences. *J Mol Biol*, 147(1) :195–197, 1981.

[34] I J Tinoco, O. C. Uhlenbeck, and M. D. Levine. Estimation of secondary structure in ribonucleic acids. *Nature*, 230(5293) :362–367, April 1971.

[35] H. Touzet and O. Perriquet. CARNAC : folding families of related RNAs. *Nucleic Acids Res*, 32(Web Server issue) :W142–W145, July 2004.

[36] D. H. Turner and D. H. Mathews. NNDB : the nearest neighbor parameter database for predicting stability of nucleic acid secondary structure. *Nucleic Acids Res*, 38(Database issue) :D280–D282, January 2010.

[37] J. Waldisphl and Y. Ponty. An unbiased adaptive sampling algorithm for the exploration of RNA mutational landscapes under evolutionary pressure. *J Comput Biol*, 18(11) :1465–1479, November 2011.

[38] S. Wuchty, W. Fontana, I. L. Hofacker, and P. Schuster. Complete suboptimal folding of RNA and the stability of secondary structures. *Biopolymers*, 49(2) :145–165, February 1999.

[39] M. Zuker. On finding all suboptimal foldings of an RNA molecule. *Science*, 244(4900) :48–52, 1989.

[40] M. Zuker and A. B. Jacobson. Using reliability information to annotate RNA secondary structures. *RNA*, 4(6) :669–679, 1998.

[41] M. Zuker and D. Sankoff. RNA secondary structures and their prediction. *Bull Math Biol*, 46(4) :591–621, 1984.

[42] M. Zuker and P. Stiegler. Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. *Nucleic Acids Res*, 9(1) :133–148, January 1981.

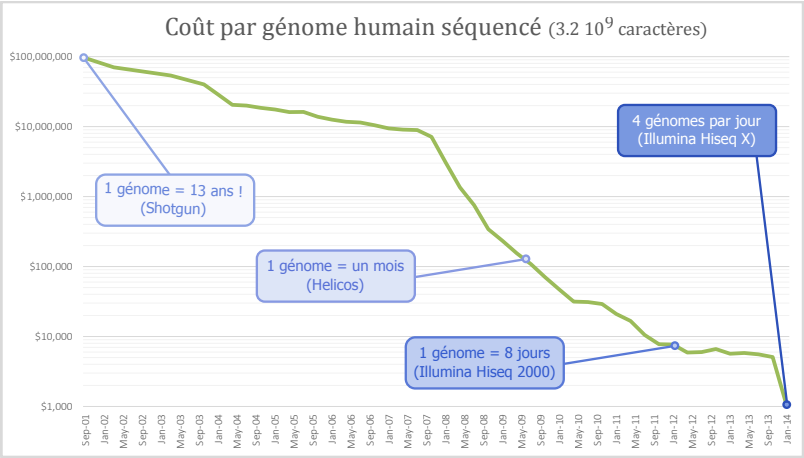


FIGURE 6. Évolution (échelle logarithmique) du coût associé au (re)séquençage d'un génome humain entier (Source : NHGRI [67]).

Complément 1

SÉQUENÇAGE HAUT-DÉBIT ET ALGORITHMIQUE DU TEXTE

Des algorithmes efficaces sont aussi indispensables afin de faire face au déluge de données produites, à un débit sans précédent, par des techniques expérimentales en constante évolution. Les pratiques quotidiennes de pans entiers de la recherche en biologie ont ainsi été révolutionnés par l'arrivée, la pérennisation puis la démocratisation des techniques de séquençage haut-débit.

Au début des années 2000, le séquençage shotgun, initié par Sanger et perfectionné dans le cadre du séquençage du génome humain, permettait de produire des fragments d'un ADN à un débit relativement faible et pour un coût important. La dernière décennie a vu se succéder des techniques de séquençage haut-débit de plus en plus rapides et complexes pour des coûts de plus en plus faibles. En particulier, comme illustré par la Figure 6, le constructeur Illumina aurait récemment franchi une barre symbolique en annonçant une machine capable de séquencer un génome individuel (couverture moyenne 30×) en une journée pour un coût total inférieur à 1 000 $, prix incluant les consommables et le travail des techniciens [5].

Un tel coût, dérisoire au regard du niveau de vie des pays développés, permet désormais d'envisager un recours généralisé au séquençage, et ouvre la porte à des

5. Ce chiffre exclut cependant l'investissement initial, qui s'élève tout de même à 10 000 000$!

applications au potentiel révolutionnaire, telle une médecine personnalisée selon le contenu précis de notre génome ou de notre flore intestinale. Cette situation soulève aussi des problématiques de stockage des données et de traitement algorithmique des données produites, atteignant par exemple de l'ordre de 300 Go par reséquençage de génome humain dans le contexte du diagnostic médical. Elle soulève aussi de très concrètes et pressantes questions d'ordre éthique : Qui disposera des données des patients ? Sous quelle forme ? Comment empêcher des séquençages *pirates* ? Plusieurs thèmes de l'informatique (cryptographie distribuée, confidentialité différentielle...) pourront fournir des réponses techniques partielles, mais ne sauront se substituer à un débat démocratique, ainsi qu'à une volonté politique forte pour en appliquer les conclusions.

Reséquençage de génomes connus. Quelque révolutionnaires que soient les nouvelles technologies de séquençage, elles ne produisent jamais que des petits fragments de notre ADN. Deux approches différentes peuvent être utilisées pour les transformer en séquences génomiques, en fonction de la disponibilité ou non d'un génome de référence, appartenant à un individu suffisamment proche, au sens de l'évolution, de l'individu séquencé.

Si un génome de référence est disponible, alors il est possible d'effectuer un reséquençage, c'est-à-dire une projection des fragments sur celui-ci, par exemple afin d'identifier ou reconnaître des mutations induisant un phénotype. Cette technique est aussi largement utilisée dans d'autre types d'expériences, parmi lesquelles les expériences RNA-Seq, où l'on cherche à quantifier la production d'ARN messagers à partir du nombre de fragments chevauchant les régions génomiques leur étant associées. Une telle entreprise soulève typiquement d'importants problèmes de renormalisation, certaines régions génomiques étant plus *couvertes* que d'autres par le séquençage, pour des raisons intrinsèques ou relevant d'un biais expérimental. En conséquence, on utilise souvent ces expériences de façon *différentielle*, en comparant les profils de couverture obtenus dans différences conditions, afin par exemple d'identifier l'effet d'un changement environnemental sur l'expression des gènes, ou identifier des comportements pathologiques. Afin de faire face à l'explosion des données produites, des structures de données succinctes et efficaces (Filtres de Blum [47]), ainsi que des techniques d'encodage et de compression dédiées (Transformée de Burrows/Wheeler [61]) sont indispensables.

Assemblage de novo. En l'absence de génome de référence, on souhaite alors effectuer un assemblage de novo, c'est-à-dire reconstruire une grande séquence d'ADN compatible avec un ensemble de fragments. L'approche dominante passe par la reconstruction d'un grand graphe dirigé, appelé graphe de De Bruijn, dont la construction est illustrée par la Figure 7. La reconstruction de la séquence d'ADN initiale s'apparente alors à la recherche d'un chemin eulérien, un chemin passant par chacune des arêtes du graphe. Un tel chemin n'est pas forcément unique, par exemple

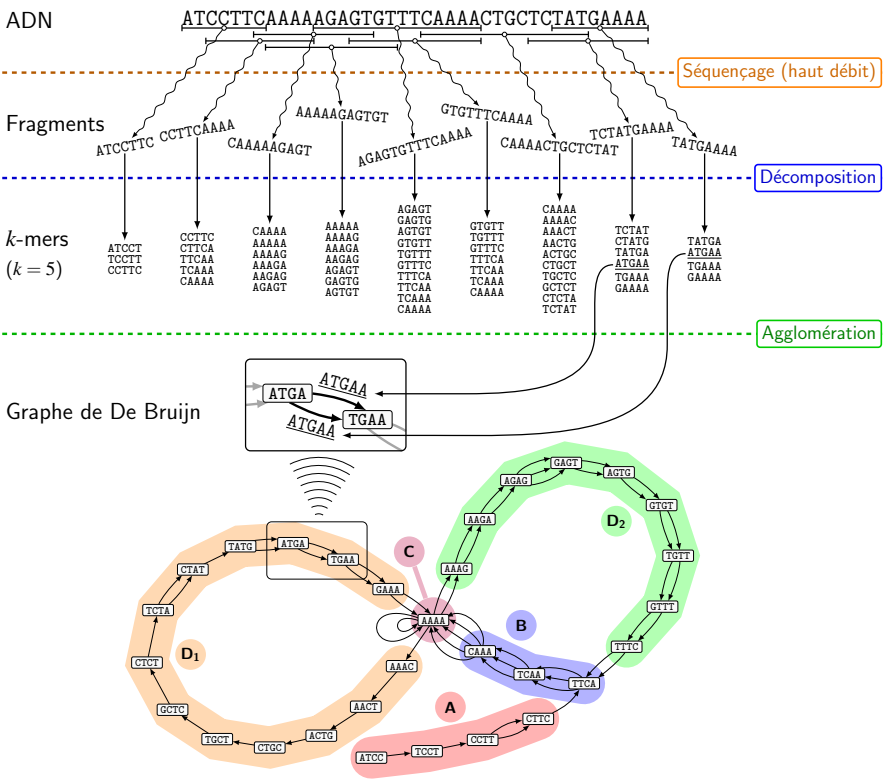


FIGURE 7. Graphe de De Bruijn pour l'assemblage. Les fragments obtenus par séquençage de l'ADN sont décomposés en k-mers, des petits fragments de taille k homogènes. On reconstruit alors le graphe de De Bruijn, dont les nœuds sont les $(k-1)$-mers, et les arêtes relient deux $(k-1)$-mers apparaissant dans un k-mer séquencé. La reconstruction de la séquence initiale s'apparente alors à la recherche d'un chemin eulérien, passant par toutes les arêtes du graphe.

si des motifs répétés apparaissent dans la séquence initiale, telle la composante **C** de la Figure 7. Ici, les données disponibles ne permettent pas de distinguer la séquence initiale (correspondant au chemin **A.B.C.C.D$_2$.B.C.D$_1$.C**) parmi l'ensemble des séquences alternatives (dont le chemin **A.B.C.D$_1$.C.C.D$_2$.B.C**). Certaines portions peuvent aussi n'être recouvertes par aucun fragment. Dans de telles situations, des expériences complémentaires visant la production de plus grands fragments, ou encore l'utilisation d'un modèle de maximisation de vraisemblance, permettront de

départager les séquences candidates.

Au-delà de la génomique. Le champ d'application de ces techniques ne se limite plus désormais au séquençage des génomes et à l'analyse des niveaux de transcription. Elles pénètrent aussi la biologie moléculaire et structurale, où elles apparaissent comme une étape-clé au sein de méthodes permettant, à haut-débit :

— une caractérisation de la structure de la chromatine, l'architecture 3D de l'ADN, à travers une détermination de ses contacts inter et intra-chromosomiques (méthodes 5C, Hi-C), qui permet de comprendre son influence sur les mécanismes cellulaires, ou encore de déterminer l'architecture 3D des génomes fortement structurés tel celui de la levure [48] ;

— une détermination des interactions ARN-Protéines (Clip-SEQ) [63], permettant de mieux comprendre l'articulation de ces deux acteurs du monde cellulaire, et leur insertion dans les réseaux de régulation ;

— une estimation des profils d'accessibilité dans les ARN structurés (`FragSeq` [64] ou `SHAPE-Seq` [44]), aidant à la prédiction de structure.

Ces données ne fournissent souvent qu'une information incomplète, assimilable à une *projection* du phénomène d'intérêt. Elles sont donc souvent incorporées dans des algorithmes d'optimisation multiobjectif, et la conception d'algorithmes de prédiction et d'annotation tirant efficacement partie de ces données est aujourd'hui un des challenges algorithmiques majeurs en bioinformatique. En particulier, les nombreux biais systématiques apparaissant dans les données haut-débit soulèvent des problèmes, particulièrement ardus, d'inférence de paramètres cachés couplés à une maximisation de la vraisemblance, souvent résolus en pratique par des heuristiques propres aux méthodes Bayésiennes.

Complément 2

PETIT DÉTOUR PAR LA COMBINATOIRE

Remontons maintenant en 1978, à l'origine des travaux théoriques sur la structure secondaire de l'ARN, et effectuons comme Waterman et Smith [66] un petit exercice de combinatoire énumérative/analytique afin d'énumérer les structures secondaires d'ARN de taille n. On rappelle que celles-ci sont engendrées par la grammaire hors-contexte

$$S \to (\, S^{\overline{\varepsilon}} \,)\, S \mid \bullet S \mid \varepsilon \qquad\qquad S^{\overline{\varepsilon}} \to (\, S^{\overline{\varepsilon}} \,)\, S \mid \bullet S\,;$$

Considérons alors les séries génératrices

$$S(z)=\sum_{n\geq 0} s_n z^n \qquad \text{et} \qquad S^{\overline{\varepsilon}}(z)=\sum_{n\geq 0} s^{\overline{\varepsilon}}_n z^n,$$

où s_n (resp. $s^{\overline{\varepsilon}}_n$) est le nombre de structures secondaires (resp. non vide), et z est une variable formelle. On peut alors remarquer que la grammaire ci-dessus est non-ambiguë, et qu'elle peut donc être traduite *automatiquement* en le système d'équations fonctionnelles

$$S(z)\to zS^{\overline{\varepsilon}}(z)zS(z)+zS(z)+1 \qquad \Rightarrow \qquad S^{\overline{\varepsilon}}(z)=zS^{\overline{\varepsilon}}(z)zS(z)+zS(z)\ ;$$

En résolvant ce système, on trouve deux solutions conjuguées, dont une seulement admet un développement à coefficient positifs :

$$S(z)=\frac{1-z+z^2-\sqrt{1-2z-z^2+z^4}}{2z^2}. \tag{1}$$

Cette série génératrice est, par construction, algébrique et admet une singularité dominante, unique et de type *racine*, en $z=\rho:=\frac{3-\sqrt{5}}{2}$. Une analyse de singularité *automatique* (on trouvera plus de détails dans la *bible* de Flajolet/Sedgewick [49]) nous donne alors un équivalent asymptotique précis pour le nombre de structures secondaires de taille n

$$s_n:=[z^n]S(z)=\kappa\cdot\frac{(1/\rho)^n}{n\sqrt{n}}(1+\mathscr{O}(1/n)),\ \text{où } 1/\rho\approx 2.618\ldots$$

où κ est une constante.

On peut aussi, de façon similaire, s'intéresser au nombre moyen de structures compatibles avec une séquence. On remarque alors que l'ensemble des couples structure/séquence compatibles, c'est-à-dire telles que les positions appariées dans la structure correspondent à des paires de nucléotides canoniques, est engendré par la grammaire

$$T\to \begin{array}{l}({}_{\mathsf{A}}T^{\overline{\varepsilon}})_{\mathsf{U}}T\\({}_{\mathsf{U}}T^{\overline{\varepsilon}})_{\mathsf{A}}T\\({}_{\mathsf{G}}T^{\overline{\varepsilon}})_{\mathsf{C}}T\\({}_{\mathsf{C}}T^{\overline{\varepsilon}})_{\mathsf{G}}T\\({}_{\mathsf{U}}T^{\overline{\varepsilon}})_{\mathsf{G}}T\\({}_{\mathsf{G}}T^{\overline{\varepsilon}})_{\mathsf{U}}T\end{array}\Bigg|\begin{array}{l}\bullet_{\mathsf{A}}T\\\bullet_{\mathsf{U}}T\\\bullet_{\mathsf{G}}T\\\bullet_{\mathsf{C}}T\end{array}\Bigg|\,\varepsilon \qquad S^{\overline{\varepsilon}}\to \begin{array}{l}({}_{\mathsf{A}}T^{\overline{\varepsilon}})_{\mathsf{U}}T\\({}_{\mathsf{U}}T^{\overline{\varepsilon}})_{\mathsf{A}}T\\({}_{\mathsf{G}}T^{\overline{\varepsilon}})_{\mathsf{C}}T\\({}_{\mathsf{C}}T^{\overline{\varepsilon}})_{\mathsf{G}}T\\({}_{\mathsf{U}}T^{\overline{\varepsilon}})_{\mathsf{G}}T\\({}_{\mathsf{G}}T^{\overline{\varepsilon}})_{\mathsf{U}}T\end{array}\Bigg|\begin{array}{l}\bullet_{\mathsf{A}}T\\\bullet_{\mathsf{U}}T\\\bullet_{\mathsf{G}}T\\\bullet_{\mathsf{C}}T\end{array}.$$

On la transforme alors en système d'équation, qu'on résout pour obtenir la série génératrice

$$T(z):=\sum_{n\geq 0} t_n z^n=\frac{1-4z+6z^2-\sqrt{1-8z+4z^2-48z^3+36z^4}}{12z^2},$$

où les t_n comptent désormais les couples structures/séquences compatibles de taille n, asymptotiquement équivalents à

$$t_n := [z^n]\,T(z) = \kappa' \cdot \frac{(1/\rho')^{-n}}{n\sqrt{n}}(1+\mathcal{O}(1/n)),\ \text{où } 1/\rho' \approx 8.164\ldots$$

où κ' est une constante. On en conclut donc qu'un ARN de taille n, tiré uniformément parmi les 4^n séquences possibles sur $\{\mathsf{A},\mathsf{C},\mathsf{G},\mathsf{U}\}$, est en moyenne compatible avec $\Theta(2.04^n/n\sqrt{n})$ structures secondaires. Cette croissance exponentielle chute à $\Theta(1.87^n/n\sqrt{n})$ en imposant une distance minimale, souvent utilisée dans la littérature, de $\theta = 3$ entre les nucléotides appariés [41].

Complément 3

C'ÉTAIT DEMAIN… RÉSULTATS RÉCENTS ET PROBLÈMES OUVERTS

Outre les questions algorithmiques laissées ouvertes sur la prédiction du repliement par les approches *ab initio* et comparatives, plusieurs questions bien définies animent actuellement la communauté, et seront très probablement l'objet de nombreuses contributions futures à l'interface informatique/sciences du vivant.

Approches génériques. On l'imagine aisément à la lecture des équations de programmation dynamique de la littérature (s'étalant parfois sur plusieurs pages), l'implémentation des algorithmes évoqués ici s'avère parfois concrètement problématique. Leur conception elle-même est complexe, avec des problèmes de bornes sur les sommes, bornes prudemment omises dans ce document. Comme l'écrivait au début des années 2000 un relecteur anonyme, fréquemment cité par Robert Giegerich : « *Le développement d'une récurrence de programmation dynamique réussie est une question d'expérience, de talent et de chance* ! »[6] Plusieurs formalismes déclaratifs ont ainsi été proposés pour abstraire un schéma de programmation dynamique de son équation, de façon à limiter au maximum les erreurs liées à son implémentation.

Au cours des années 1990, Fabrice Lefebvre propose d'utiliser des grammaires attribuées comme formalisme unificateur [56]. Ces travaux sont ensuite poursuivis et étendus dans les années 2000 par le groupe de Robert Giegerich, qui introduit le formalisme de la programmation dynamique algébrique [52, 62, 53], couplé avec des compilateurs efficaces vers des langages *bas-niveau*. L'utilisation, même par des informaticiens, de tels formalismes pour exprimer leurs algorithmes paraissait appartenir à la science-fiction il y a une petite dizaine d'années. Pourtant, nombreux sont

6. *The development of successful dynamic programming recurrences is a matter of experience, talent and luck.* Anonymous Reviewer.

ceux désormais, parmi lesquels des bioinformaticiens de formation, qui utilisent ce type de formalisme pour tester des nouvelles hypothèses/algorithmes à faible coût.

Optimisations de la programmation dynamique. Le problème du repliement de l'ARN est représentatif d'une grande classe de problèmes se prêtant à la programmation dynamique sur des structures arborescentes. À ce titre, ils ont été rencontrés dans d'autres communautés, où des améliorations ont été apportées à leur complexité. Par exemple, le calcul de la fonction de partition et les probabilités de paires de bases à pu être amélioré à $\Theta(n^{2.38})$ [69] par utilisation de la méthode de Valiant [65]. Cette dernière consiste à ramener le calcul de la matrice de programmation dynamique au produit d'une famille de matrices, permettant alors de profiter des dernières optimisations dans le domaine. L'astuce des *quatre-russes* [43] permet aussi de gagner un facteur logarithmique sur la complexité du repliement [50], mais les constantes importantes apparaissant dans son implémentation limitent sa portée en pratique.

De façon plus spectaculaire, plusieurs auteurs ont utilisé un principe de calcul clairsemé (*sparsification*) appliqué à la programmation dynamique prédictive pour les ARN. L'idée est de calculer, lors des étapes initiales de la récurrence, non seulement un score optimal mais aussi une liste de candidats (indices des boucles) non-trivialement dominés, et méritant une exploration lors des étapes ultérieures. Le calcul est alors toujours exact mais un gain en temps n'est pas toujours garanti. Sous certaines hypothèses (distribution des paires de bases en loi de puissance), la cardinalité de ces listes est bornée par une constante. Il en résulte alors un gain algorithmique linéaire, et le repliement est alors résoluble en temps $\Theta(n^2)$ [45]. Une caractérisation des classes de séquences garantissant un tel gain reste cependant ouverte.

Design d'ARN. Le design d'ARN consiste à concevoir une séquence se repliant, au sens de la minimisation d'énergie, en une structure secondaire donnée en entrée. Il est en ce sens le problème inverse du repliement, et plusieurs travaux y font référence sous le nom de *repliement inverse* de l'ARN. De façon assez surprenante, et malgré deux décennies de recherches sur le sujet [55], la complexité théorique du problème n'est toujours pas établie, une situation totalement exceptionnelle en bioinformatique.

Mimant en un certain sens l'évolution, plusieurs auteurs ont proposé des approches basées sur la recherche locale [46], l'algorithmique génétique [58] ou encore une résolution exponentielle exacte basée sur de la programmation par contraintes [51]. Cependant, le comportement de ce type d'approches est très peu prévisible, et celles-ci s'avèrent difficilement adaptables (en temps raisonnable) à des ensembles de contraintes plus expressives, nécessaires dans les contextes biologiques. Une étude plus approfondie de ce problème est donc à la fois motivée par son intrigante (absence de) structure exploitable algorithmiquement, et par des applications potentielles fascinantes en biologie synthétique [60].

Considérations cinétiques. En étudiant l'ARN par une approche probabiliste de type *fonction de partition* ou, *a fortiori*, en prédisant le repliement par minimisation d'énergie, on travaille implicitement sous l'hypothèse que le processus du repliement a atteint un régime stationnaire, l'équilibre de Boltzmann. Or, le temps de mélange de la chaîne de Markov associée est loin d'être indépendant des propriétés de la séquence d'ARN considérée et peut, dans certains cas, dépasser la durée de vie de l'ARN dans la cellule. On peut donc rencontrer des pièges cinétiques, des minimas locaux de l'énergie libre, dont un ARN aura bien du mal à sortir avant sa dégradation finale par les enzymes. L'ARN peut alors n'être jamais observable dans un état potentiellement extrêmement stable, et l'hypothèse d'un équilibre de Boltzmann donne une vision faussée de sa fonction.

Malheureusement, on rencontre rapidement plusieurs difficultés quand on cherche à étudier les propriétés cinétiques fines d'un ARN. D'une part, il existe un nombre exponentiel de pièges cinétiques potentiels parmi les conformations compatibles avec une séquence [57]. De plus, le calcul de la barrière d'énergie, c'est-à-dire la quantité minimale d'énergie devant être fournie pour passer d'une conformation à une autre, est NP-complet même dans un modèle de Nussinov, et pour des conformations sans pseudo-nœuds [59].

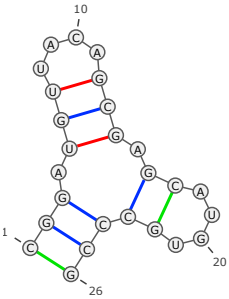
Les approches existantes reposent donc sur une simulation [68], ou sur des heuristiques exécutées sur un sous-ensemble de structures sous-optimales stables [54]. Cependant cette situation, où l'on recherche les propriétés dynamiques de processus stochastiques associés à de très grands espaces d'états, semble faire écho à des problématiques rencontrées en vérification. Il est donc permis d'espérer des contributions issues de ce domaine dans un avenir proche.

Complément 4

Modèles d'énergie

Plusieurs modèles d'énergie libre ont été considérés pour la prédiction de structure par minimisation d'énergie. Ceux-ci diffèrent principalement par le choix des motifs structurels correspondant à des contributions énergétiques élémentaires. L'énergie libre d'une structure est alors, en général, approchée par une simple somme des contributions de ses motifs.

En illustration, voici les différentes énergies libres associées à une structure secondaire dans plusieurs modèles d'énergie issus de la littérature :



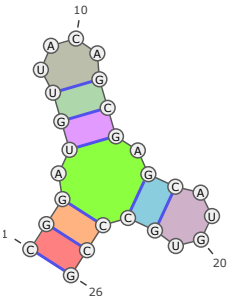
— #Paires de bases

$$E(S) = -8 \text{ kcal.mol}^{-1}$$

— #Liaisons hydrogènes

$$E(S) = 4\Delta_G\left(\begin{smallmatrix}G\\C\end{smallmatrix}\right) + 2\Delta_G\left(\begin{smallmatrix}C\\G\end{smallmatrix}\right) + 2\Delta_G\left(\begin{smallmatrix}U\\G\end{smallmatrix}\right)$$
$$= -(4\times 3 + 2\times 2 + 2\times 1) = -18 \text{ kcal.mol}^{-1}$$

— Empilements (Turner 2004)



$$E(S) = \Delta_G\left(\begin{smallmatrix}C&G\\G&C\end{smallmatrix}\right) + \Delta_G\left(\begin{smallmatrix}G&G\\C&C\end{smallmatrix}\right) + \Delta_G\left(\begin{smallmatrix}U&G\\G&C\end{smallmatrix}\right)$$
$$+ \Delta_G\left(\begin{smallmatrix}U&G\\G&C\end{smallmatrix}\right) + \Delta_G\left(\begin{smallmatrix}U&G\\G&C\end{smallmatrix}\right)$$
$$= -(2.4 + 3.3 + 1.4 + 2.5 + 3.4) = -13 \text{ kcal.mol}^{-1}$$

— Plus proche voisin (Turner 2004)

$$E(S) = \Delta_G\left(\begin{smallmatrix}C&G\\G&C\end{smallmatrix}\right) + \Delta_G\left(\begin{smallmatrix}G&G\\C&C\end{smallmatrix}\right) + \Delta_G\left(\begin{smallmatrix}\cdot&\cdot\\\cdot&\cdot\end{smallmatrix}\right) + \Delta_G\left(\begin{smallmatrix}U&G\\G&C\end{smallmatrix}\right)$$
$$+ \Delta_G\left(\begin{smallmatrix}U&G\\G&C\end{smallmatrix}\right) + \Delta_G\left(\begin{smallmatrix}\cdot&\cdot\\\cdot&\cdot\end{smallmatrix}\right) + \Delta_G\left(\begin{smallmatrix}U&G\\G&C\end{smallmatrix}\right) + \Delta_G\left(\begin{smallmatrix}\cdot&\cdot\\\cdot&\cdot\end{smallmatrix}\right)$$
$$= -2.4 - 3.3 + 3.6 - 1.4 - 2.5 + 5.4 - 3.4 + 4.1$$
$$= 0.1 \text{ kcal.mol}^{-1}$$

Remarquons que la variabilité des énergies estimées par ces différents modèles n'est pas nécessairement problématique, l'énergie libre étant définie de façon relative. En revanche, l'énergie positive de la structure dans le modèle *plus proche voisin* indique que la structure *ouverte*, sans aucun appariement et ayant une énergie nulle par convention, lui sera préférée par un algorithme de prédiction.

Références supplémentaires

[43] V. L. Arlazarov, E. A. Dinic, M. A. Kronrod, and I. A. Faradev, *On economical construction of the transitive closure of a directed graph*, Soviet Mathematics—Doklady **11** (1970), no. 5, 1209–1210.

[44] S. Aviran, C. Trapnell, J. B. Lucks, S. A. Mortimer, S. Luo, G. P. Schroth, J. A. Doudna, A. P. Arkin, and L. Pachter, *Modeling and automation of sequencing-based characterization of RNA structure.*, Proc Natl Acad Sci U S A **108** (2011), no. 27, 11069–11074 (eng).

[45] R. Backofen, D. Tsur, S. Zakov, and M. Ziv-Ukelson, *Sparse RNA folding : Time and space efficient algorithms*, Journal of Discrete Algorithms **9** (2010), no. 1, 12–31.

[46] A. Busch and R. Backofen, *INFO-RNA–a fast approach to inverse RNA folding.*, Bioinformatics **22** (2006), no. 15, 1823–1831 (eng).

[47] R. Chikhi and G. Rizk, *Space-efficient and exact de bruijn graph representation based on a bloom filter*, Algorithms for Molecular Biology **8** (2013), no. 1, 22.

[48] Z. Duan, M. Andronescu, K. Schutz, C. Lee, J. Shendure, S. Fields, W. S. Noble, and C., *A genome-wide 3C-method for characterizing the three-dimensional architectures of genomes.*, Methods **58** (2012), no. 3, 277–288 (eng).

[49] P. Flajolet and R. Sedgewick, *Analytic combinatorics*, Cambridge University Press, 2009.

[50] Y. Frid and D. Gusfield, *A simple, practical and complete O(n3/log n)-time algorithm for RNA folding using the four-russians speedup.*, Algorithms Mol Biol **5** (2010), 13 (eng).

[51] J. A. Garcia-Martin, P. Cote, and I. Dotu, *RNAiFOLD : A constraint programming algorithm for RNA inverse folding and molecular design*, J Bioinform Comput Biol **11** (2013), no. 02, 1350001, PMID : 23600819.

[52] R. Giegerich and C. Meyer, *Algebraic dynamic programming*, AMAST, 2002, pp. 349–364.

[53] R. Giegerich and H. Touzet, *Modeling dynamic programming problems over sequences and trees with inverse coupled rewrite systems*, Algorithms **7** (2014), no. 1, 62–144.

[54] I. L. Hofacker, C. Flamm, C. Heine, M. T. Wolfinger, G. Scheuermann, and P. F. Stadler, *Barmap : RNA folding on dynamic energy landscapes.*, RNA **16** (2010), no. 7, 1308–1316 (eng).

[55] I. L. Hofacker, W. Fontana, P. Stadler, L. Bonhoeffer, M. Tacker, and P. Schuster, *Fast folding and comparison of RNA secondary structures*, Monatshefte für Chemie / Chemical Monthly **125** (1994), no. 2, 167–188 (English).

[56] F. Lefebvre, *An optimized parsing algorithm well suited to RNA folding*, ISMB, 1995, pp. 222–230.

[57] W. A. Lorenz and P. Clote, *Computing the partition function for kinetically trapped RNA secondary structures.*, PLoS One **6** (2011), no. 1, e16178 (eng).

[58] R. B. Lyngsø, J. W J. Anderson, E. Sizikova, A. Badugu, T. Hyland, and J. Hein, *Frnakenstein : multiple target inverse RNA folding.*, BMC Bioinformatics **13** (2012), 260 (eng).

[59] J. Manuch, C. Thachuk, L. Stacho, and A. Condon, *NP-completeness of the energy barrier problem without pseudoknots and temporary arcs*, Natural Computing **10** (2011), no. 1, 391–405.

[60] G. Rodrigo, T. E. Landrain, and A. Jaramillo, *De novo automated design of small RNA circuits for engineering synthetic riboregulation in living cells.*, Proc Natl Acad Sci U S A **109** (2012), no. 38, 15271–15276 (eng).

[61] M. Salson, T. Lecroq, M. Léonard, and L. Mouchard, *A four-stage algorithm for updating a burrows-wheeler transform*, Theoretical Computer Science **410** (2009), no. 43, 4350–4359.

[62] G. Sauthoff, S. Janssen, and R. Giegerich, *Bellman's GAP : a declarative language for dynamic programming*, PPDP, 2011, pp. 29–40.

[63] J. Ule, K. Jensen, A. Mele, and R. B. Darnell, *CLIP : a method for identifying protein-RNA interaction sites in living cells.*, Methods **37** (2005), no. 4, 376–386 (eng).

[64] J. G. Underwood, A. V. Uzilov, S. Katzman, C. S. Onodera, J. E. Mainzer, D. H. Mathews, T. M. Lowe, S. R. Salama, and D. Haussler, *FragSeq : transcriptome-wideRNA structure probing using high-throughput sequencing.*, Nat Methods **7** (2010), no. 12, 995–1001 (eng).

[65] L. G. Valiant, *General context-free recognition in less than cubic time*, J. Comput. Syst. Sci. **10** (1975), no. 2, 308–314.

[66] M. S. Waterman and T. F. Smith, *Secondary structure of single stranded nucleic acids*, Math Biosci **42** (1978), 257–266.

[67] K A Wetterstrand, *DNA sequencing cost : Data from the NHGRI genome sequencing program (GSP)*, February 2014.

[68] A. Xayaphoummine, T. Bucher, F. Thalmann, and H. Isambert, *Prediction and statistics of pseudoknots in RNA structures using exactly clustered stochastic simulations.*, Proc Natl Acad Sci U S A **100** (2003), no. 26, 15310–15315 (eng).

[69] S. Zakov, D. Tsur, and M. Ziv-Ukelson, *Reducing the worst case running times of a family of RNA and CFG problems, using valiant's approach.*, Algorithms Mol Biol **6** (2011), no. 1, 20 (eng).

[70] M. Zuker and D. Sankoff, *RNA secondary structures and their prediction*, Bull Math Biol **46** (1984), no. 4, 591–621.

L'algorithme de test de planarité de R. E. Tarjan

Robert Cori [1]

La planarité ?

Savoir si un graphe peut être dessiné sans croisement dans le plan est un problème très simple à exposer : il s'agit, étant donnés un ensemble $X = \{x_1, x_2, \ldots, x_n\}$ de n *sommets* et un ensemble $E = \{e_1, e_2, \ldots, e_m\}$ de m *arêtes*, qui sont des paires d'éléments de X, de vérifier si on peut représenter les sommets par des points sur un plan et chaque arête $e_k = \{x_i, x_j\}$ par une courbe joignant x_i à x_j, de façon telle que les courbes ne se coupent pas. Depuis Euler on sait que ceci est impossible si $m > 3n-6$ et depuis Kuratowski qu'une telle chose est possible si et seulement si le graphe ne contient pas par réduction (en un sens un peu compliqué que l'on ne développera pas ici) l'un des deux graphes suivants :

— le graphe K_5 contenant toutes les 10 arêtes formées avec $\{x_1, x_2, x_3, x_4, x_5\}$,

— le graphe $K_{3,3}$ contenant toutes les 9 arêtes $\{x_i, x_j\}$ formées avec $\{x_1, x_2, x_3, x_4, x_5, x_6\}$ telles que i est pair et j impair.

Chercher un algorithme efficace pour résoudre ce problème était une question ouverte à la fin des années 60 — il s'agissait alors de réaliser des circuits électriques ou des composants électroniques de grande taille (chaque couche d'un circuit imprimé doit correspondre à un ensemble de connexions électriques réalisées sans croisements).

1. Professeur émérite à l'université de Bordeaux, membre du LaBRI, UMR 5800 CNRS. http://www.labri.fr/perso/cori/

L'application du théorème de Kuratowski donnant un algorithme nécessitant à priori un nombre d'opérations en n^7, plusieurs chercheurs, comme Auslander et Parter [2], Lempel, Even et Cederbaum [8] et Demoucron, Malgrange et Pertuiset [3], se sont attelés à proposer des algorithmes. La complexité de ces algorithmes n'était pas évaluée rigoureusement, ce n'était pas à l'ordre du jour alors. L'efficacité se mesurait en fonction du temps mis par l'ordinateur pour résoudre le problème sur des graphes ayant de l'ordre de 10, 100 ou 1000 sommets.

À cette époque, à Paris (Institut de Programmation), il était beaucoup question dans ce cadre d'un *Contrat Graphes* sous la responsabilité de Jean Berstel, Claude Girault et Jean-François Perrot, pour lequel Alain Jacques et Claude Lenormand avaient implanté ces algorithmes en proposant des structures de données élaborées.

Quelques années plus tard la thèse de R. E. Tarjan, soutenue à Stanford en 1971, a révolutionné le domaine des algorithmes pour les graphes car elle proposait de nouvelles idées et une présentation précise des algorithmes qui permettait l'évaluation de leur complexité. Le résultat principal de cette thèse était un algorithme en temps dépendant linéairement du nombre de sommets qui testait si un graphe est planaire. Je me propose de présenter les grandes lignes de cet algorithme en insistant sur deux points que je considère comme les plus importants permettant d'assurer cette complexité : l'algorithme de parcours en profondeur et les « piles de bi-piles ».

Une version très simplifiée du problème

On suppose dans un premier temps que les sommets $x_1, x_2, \ldots x_n$ sont des points placés sur un cercle, dans cet ordre, que n est pair ($n = 2m$), qu'il y a m arêtes en plus de celles du cercle — des cordes donc —, et enfin que chacun des x_i apparaît dans une et une seule de ces paires. On pose la question de savoir si on peut relier les points par des cordes à l'intérieur du cercle sans que celles-ci se coupent. En terme de graphes on suppose que le graphe est cubique (chaque sommet a trois voisins) et que l'on en connaît un cycle Hamiltonien (un cycle passant une seule fois par tous les sommets) ; on demande alors s'il est représentable par un graphe planaire pour lequel le cycle Hamiltonien détermine une face de longueur n.

On pourrait tester que les cordes ne se coupent pas en les traitant deux à deux, ce qui donnerait un algorithme en $O(n^2)$. On propose ici un algorithme en $O(n)$ qui fait intervenir une pile d'entiers. Cet algorithme sera modifié dans la section suivante pour traiter un problème plus complexe.

On suppose donnés les éléments de E sous forme d'un tableau de longueur n, formé en mettant les paires les unes à la suite des autres, triées dans l'ordre croissant de leur plus petit élément, et en ordonnant chaque paire de façon décroissante.

Par exemple (voir Figure 1), si l'ensemble des paires est :

$$\{1,6\}\ \{2,12\}\ \{3,5\}\ \{4,8\}\ \{7,11\}\ \{9,10\}$$

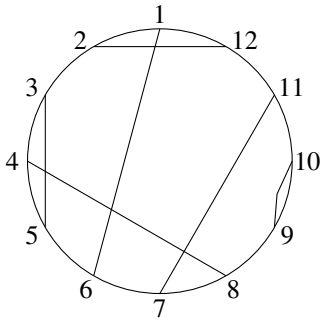


FIGURE 1. Un ensemble de cordes qui se coupent

le tableau représentant ces paires sera :

$$t = 6,1,12,2,5,3,8,4,11,7,10,9$$

On peut obtenir ce tableau en temps linéaire à partir des cordes données dans un ordre quelconque par une procédure simple, laissée à la sagacité du lecteur.

On note que l'exemple donné plus haut n'est pas réalisable sans croisement car, par exemple, les cordes $\{1,6\}$ et $\{2,12\}$ se coupent et la corde $\{4,8\}$ coupe trois autres cordes.

L'Algorithme 1 utilise une pile contenant des entiers. Ceux-ci doivent sortir de la pile dans l'ordre croissant de 1 à n.

À chaque étape, pour $i = 1,\ldots,n$, on effectue une itération qui supprime le haut de la pile s'il est égal au dernier élément sorti de la pile augmenté de 1 ; ensuite, on compare le nombre t_i avec le nombre qui se trouve au sommet de la pile, que l'on notera a dans la suite. Si la pile est vide, ou si $t_i < a$, on empile t_i ; sinon on peut montrer qu'il y a alors un croisement entre la corde dont une extrémité est a et celle dont une extrémité est t_i. En effet, on constate d'abord que la règle d'empilement implique la croissance des entiers dans la pile depuis le haut jusqu'au fond de la pile. Ainsi les t_k pour les valeurs de k paires font une entrée dans la pile suivie immédiatement d'une sortie. En posant $a = t_j$ on constate que les deux cordes considérées sont $c_1 = \{t_{j+1}, t_j\}$ et $c_2 = \{t_{i+1}, t_i\}$. On a ainsi $t_{j+1} < t_{i+1}$ car la corde c_1 apparaît avant c_2 dans t ; de plus $t_{i+1} < t_j$, car lorsque t_k se présente, avec k pair, tous les nombres qui se trouvent dans la pile sont supérieurs à t_k. De ce fait :

$$t_{j+1} < t_{i+1} < t_j < t_i$$

ce qui implique que c_1 et c_2 se croisent.

Données : Un ensemble de cordes représenté par un tableau t les contenant sous la forme décrite plus haut, P est une pile d'entiers supposée contenir initialement un entier plus grand que n de façon à éviter les accès à la pile vide.
Résultat : Vrai si aucune corde n'en coupe une autre.

```
dernierSorti ← 0;
pour i = 1,2,...,n faire
    tant que (Top(P) = dernierSorti + 1) faire
        Depiler(P) ;
        dernierSorti ← dernierSorti + 1
    fin
    si (non vide(P) et Top(P) < t[i]) alors
        return false
    sinon
        Empiler(P,t[i])
fin
return true
```

Algorithme 1 : Vérifier qu'un ensemble de cordes ne se coupent pas

On peut remarquer que la suite des nombres dans le tableau t est une permutation des entiers $1,2,\ldots,n$ et que l'algorithme teste si cette permutation peut être triée en utilisant une pile. En fait, on peut se contenter de vérifier que le tableau obtenu en ne retenant qu'un entier sur deux de t, ceux en position impaire, est bien triable par une pile. On peut aussi montrer que le nombre de permutations triables par une pile est égal au nombre d'ensembles de cordes sans croisement. Il se trouve que c'est la suite des *nombres de Catalan* $(1,2,5,14,42,132,\cdots)$, bien connue en combinatoire énumérative.

Une version moins simple

On considère maintenant une version du problème dont l'énoncé paraît semblable au précédent mais dont la solution est nettement moins simple. Elle fait intervenir un concept important introduit par R. E. Tarjan pour traiter le cas des graphes planaires quelconques. La donnée est la même que pour le cas précédent mais on autorise toutefois les cordes à être dessinées à l'intérieur ou à l'extérieur du cercle (voir l'exemple donné en Figure 2). Avec l'interprétation précédente, cela revient à tester si le tableau t peut être trié à l'aide de *deux* piles, l'une triant les cordes mises à l'intérieur l'autre celles à l'extérieur.

Une façon de faire consiste à construire un *graphe des croisements* dont les sommets correspondent aux cordes et tel que deux sommets sont reliés par une arête si

les cordes correspondantes se coupent lorsqu'elles sont mises du même coté, c'est-à-dire si elles sont de la forme $\{x_i, x_j\}$ et $\{x_k, x_\ell\}$ avec $i < k < j < \ell$. Tester si on peut représenter les cordes sans croisement revient alors à tester si les sommets du graphe ainsi formé sont coloriables en deux couleurs de façon telle que deux sommets reliés par une arête soient toujours de couleurs différentes. Ceci donne un algorithme dont le nombre d'opérations est linéaire en le nombre d'arêtes du graphe des croisements et qui semble donc efficace. Toutefois la construction de ce graphe nécessite l'examen de toutes les paires de cordes ce qui est de l'ordre de n^2. R. E. Tarjan a mis au point dans sa thèse une structure de données — qu'il a ensuite reprise de façon plus claire avec P. Rosenstiehl dans [11] — pour obtenir un algorithme en temps linéaire permettant de tester la bi-coloriabilité du graphe des croisements (voir Algorithme 2).

Cette structure fait intervenir une pile dont les cellules sont des couples de piles d'entiers : une pile gauche et une pile droite. Les opérations définies sur les cellules sont les suivantes :

— À partir d'un entier on peut **créer** une cellule dont la pile gauche contient cet entier et la pile droite est vide.
— On peut **échanger** la pile gauche et la pile droite d'une cellule.
— On peut **concaténer** deux cellules en mettant l'une à la suite de l'autre les deux piles gauches et les deux piles droites.
— On peut **accéder** à la pile gauche $cell.G$ ou à la pile droite $cell.D$ de la cellule $cell$ et y effectuer les opérations classiques sur les piles d'entiers.

Sur les piles de cellules on dispose des opérations courantes sur les piles :

— **Tester** si la pile est **vide**, **Empiler** une cellule en haut de la pile, **Dépiler** la cellule du haut de la pile.
— **Accéder** à la cellule en haut d'une pile P de deux façons différentes, soit par $Top(P)$ pour obtenir la cellule entière, soit par $TopG(P)$ (ou $TopD(P)$) pour obtenir l'entier qui se trouve en haut de la pile gauche (ou de la pile droite) de cette cellule.

On peut considérer ici la vérification de la réalisabilité comme la possibilité de trier la permutation représentée par le tableau t à l'aide de deux piles en parallèle : à chaque étape de l'algorithme on doit choisir de dépiler l'une des deux piles ou d'empiler l'élément suivant de la permutation sur l'une ou l'autre des piles.

En fait, il s'agit de représenter les possibilités de choix pour l'algorithme d'utiliser l'une ou l'autre des piles à sa disposition par cette structure de « pile de bi-piles » qui remet à plus tard le choix de la pile sur laquelle on empile un élément. Le choix est effectué lorsque l'on trouve une incompatibilité entre deux cordes qui se coupent.

L'Algorithme 2 permet de tester si une permutation quelconque peut être triée par deux piles, dans le cas où chaque élément ne peut être inséré que dans une seule des

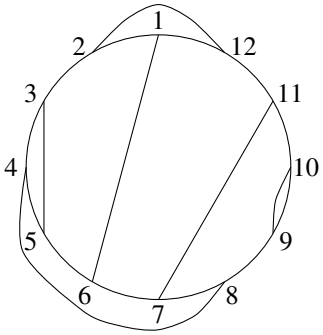


FIGURE 2. Un ensemble de cordes qui peut être représenté en interne et externe.

deux piles, au choix. Le nombre de ces permutations a été évalué de façon précise très récemment par Michael Albert et Mireille Bousquet-Mélou [1].

L'algorithme, dans le cas général

L'algorithme de planarité présenté dans la thèse de R. E. Tarjan — puis sous une forme simplifiée dans [6] —, utilise les piles de bi-piles dans sa dernière phase. Dans la première phase il s'agit de construire un arbre recouvrant [2] obtenu par une recherche en profondeur. Un tel arbre est également appelé *arbre de Trémaux*. En effet, ce bel algorithmique récursif a été pour la première fois proposé par Charles Trémaux (1859–1882), polytechnicien de la promotion 1876 et ingénieur des télégraphes, répondant à une question d'Édouard Lucas (1842–1891), professeur de Mathématiques en classe de spéciales au Lycée Saint-Louis, connu pour ses travaux en théorie des nombres et son ouvrage *Récréations Mathématiques* [9].

Il s'agissait de donner un algorithme précis permettant de sortir d'un labyrinthe, ce que l'on peut traduire par trouver un chemin qui visite toutes les arêtes d'un graphe. Édouard Lucas cite dans son ouvrage (voir page 47) une solution à son problème proposée par Charles Trémaux. Cet ouvrage est disponible sur le site Gallica de la Bibliothèque Nationale de France [3].

Plusieurs chercheurs, dont R. E. Tarjan, ont retrouvé l'algorithme proposé par Trémaux. Le mérite de R. E. Tarjan est d'avoir utilisé la numérotation des sommets obtenue dans l'ordre de ce parcours pour répondre à d'autres questions algorithmiques sur les graphes.

2. Un arbre recouvrant est un arbre contenant tous les sommets du graphe et obtenu en sélectionnant un sous-ensemble de ses arêtes.

3. http://gallica.bnf.fr/ark:/12148/bpt6k3943s

Données : Un ensemble de cordes représentées par un tableau t les contenant sous la forme décrite plus haut, P est une pile de cellules supposée contenir initialement une cellule avec deux piles contenant chacune un entier plus grand que n de façon à éviter les accès à la pile vide.
Résultat : Vrai si on peut donner une représentation planaire des cordes de manière interne ou externe au cercle, Faux sinon.

```
doitSortir ← 1;
pour i = 1,2,...,n faire
    tant que (TopG(P) = doitSortir ou TopD(P) = doitSortir) faire
        Depiler(Top(P).G) ou Depiler(Top(P).D) suivant le cas;
        si (estVide(Top(P).G) et estVide(Top(P).D) alors
            Depiler(P);
        doitSortir ← doitSortir + 1
    fin
    cell1 ← Top(P);
    cell2 ← CreerCellule(t[i]);
    tant que (t[i] > Top(cell1.G) ou t[i] > Top(cell1.D)) faire
        si t[i] < Top(cell1.D) alors
            Echanger(cell1) ;
        si t[i] > Top(cell1.G) alors
            return false
        sinon
            cell2 ← Concatener(cell2, cell1);
            Depiler(P);
            cell1 ← Top(P);
        fin
    fin
    Empiler(P, cell2);
fin
return true
```

Algorithme 2 : Vérifier qu'un ensemble de cordes peut être réalisé en interne et en externe

Son idée fondamentale a été d'associer à chaque sommet x_i une valeur qu'il a appelée « point le plus bas » (*lowpoint*) : elle désigne le numéro du sommet de plus faible numéro dans l'arbre de Trémaux qui peut être atteint par une seule arête en dehors de cet arbre depuis un sommet situé dans le sous-arbre ayant pour racine x_i. Ce nouveau paramètre a fait merveille dans d'autres algorithmes sur les graphes comme la recherche des composantes biconnexes ou triconnexes, le test de forte connexité pour les graphes orientés et d'autres applications.

Une deuxième phase consiste à décomposer le graphe en un ensemble de chemins, il y en a autant que d'arêtes en dehors de l'arbre. Ces chemins jouent ensuite

un rôle similaire aux cordes que nous avons considérées dans les problèmes étudiés plus haut. Il s'agit alors de savoir pour chaque chemin s'il doit être plongé à l'intérieur ou à l'extérieur d'un cycle principal de la construction. Cette décision doit être prise en fonction d'incompatibilités qui ont un statut très voisin du graphe construit sur les cordes. L'utilisation d'une pile de bi-piles est alors cruciale pour obtenir un algorithme qui donne un résultat en temps linéaire.

Il serait bien long de donner la totalité de l'algorithme dans ces lignes. J'ai choisi de mettre en lumière deux points précis qui me paraissent être les points clefs de l'algorithme. Les lecteurs souhaitant connaître les détails complets peuvent consulter l'article original de Hopcroft et Tarjan [6] ou une tentative de présentation plus intuitive par William Kocay [7].

La fonction « point le plus bas » de R. E. Tarjan

L'algorithme de parcours en profondeur (*DFS*) ou de Trémaux consiste, lorsque l'on atteint un sommet au cours du parcours, à trouver un sommet voisin non encore visité ; si tous les voisins ont déjà été visités il faut alors retourner en arrière par l'arête qui a permis d'atteindre ce sommet pour la première fois et recommencer l'opération. À chaque fois que l'on atteint un nouveau sommet on le numérote par l'entier le plus petit non encore utilisé. L'ensemble des arêtes qui permettent d'atteindre un nouveau sommet forment alors un arbre et les arêtes en dehors sont appelées *arêtes de retour*. En orientant les arêtes de l'arbre du sommet de numéro le plus petit vers celui de numéro le plus grand et les arêtes de retour du sommet de numéro le plus grand vers le sommet de numéro le plus petit on obtient ce que Tarjan appelle un *palmier*.

La fonction *lwpt* (point le plus bas) associe au sommet x_i le numéro le plus petit du sommet qui peut être atteint par un chemin orienté composé d'une suite d'arêtes de l'arbre (dont les numéros vont donc en croissant) et d'un seul arc de retour (qui permet de faire décroître le numéro). La détermination des valeurs de *lwpt* peut s'obtenir en ajoutant quelques instructions à l'algorithme de parcours *DFS* qui calcule l'arbre recouvrant et qui numérote aussi les sommets du graphe dans l'ordre du parcours (voir Algorithme 3).

L'entier k, les tableaux d'entiers $pere, num, lwpt$ sont des variables globales, *num* est un tableau initialisé à 0 pour tous les sommets du graphe, qui est ensuite modifié lorsqu'on atteint le sommet pour la première fois. Un sommet x_0 du graphe est distingué comme la racine de l'arbre recouvrant à construire, la fonction est appelée par $LWPT(x_0)$. À la suite du calcul, les *num* et *lwpt* sont obtenus pour tous les sommets. L'arbre est fourni par la fonction *pere*.

Données : Un graphe donné par les listes des voisins de ses sommets.
Résultat : Détermination de l'arbre recouvrant par la fonction *pere*, numérotation des sommets dans l'ordre du parcours en profondeur, et de la fonction $lowpt$.

```
k ← 1;
pour Tout i = 1...n faire
    num[x_i] ← 0
fin

DFSLWPT(x) :
lwpt[x] ← num[x];
pour Tout voisin y de x faire
    si num[y] ≠ 0 alors
        lwpt[x] ← Min(num[y], lwpt[x])
    sinon
        pere[y] ← x;
        num[y] ← k;
        k ← k + 1;
        DFSLWPT(y);
        lwpt[x] ← Min(lwpt[y], lwpt[x])
    fin
fin
```

Algorithme 3 : La fonction récursive $LWPT(x)$

Cette fonction présente plusieurs intérêts en dehors de l'algorithme de planarité. Par exemple, elle permet de savoir si un sommet x d'un graphe est un point d'articulation, c'est-à-dire si la suppression de x et des arêtes qui lui sont incidentes déconnecte le graphe. En effet, il suffit de vérifier que $lwpt(y) = x$ où y est l'un des fils de x dans un arbre de Trémaux. Sur l'exemple donné en Figure 3, un graphe est dessiné avec un arbre de Trémaux en traits épais, les sommets sont numérotés dans l'ordre de parcours de l'arbre, et les valeurs de $lwpt$ sont données par :

i	1	2	3	4	5	6	7	8	9	10	11	12	13	14
$lwpt[i]$	1	1	1	3	3	4	4	6	7	1	1	2	12	12

On remarque que 12 est un point d'articulation : en effet, son fils 13 vérifie $lwpt[13] = 12$. Il en est de même de 3 car $lwpt[4] = 3$.

Pour ce qui est de l'algorithme de planarité, on peut considérer que cette fonction permet de remplacer chaque chemin composé d'une suite d'arêtes de l'arbre et d'une seule arête de retour par une corde sur un cercle formé par un cycle du graphe ; on se ramène ainsi à un problème voisin de ceux qui ont fait l'objet des deux premières sections de ce texte.

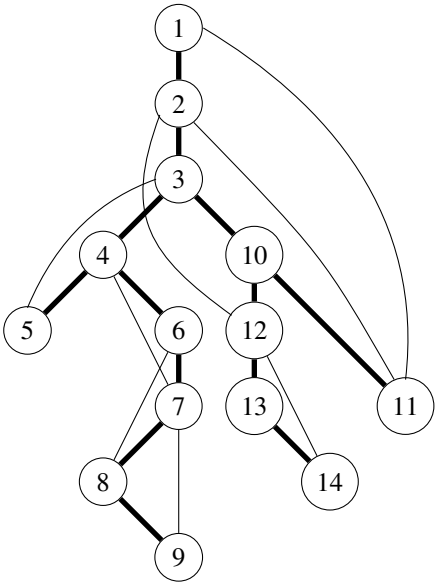


FIGURE 3. Un arbre de Trémaux, les arêtes de retour sont en trait fin

Histoire de l'algorithme

Cet algorithme a été proposé par R. E. Tarjan dans sa thèse de Ph.D. à l'Université de Stanford, sous la direction de Robert W. Floyd et Donald Knuth. Cette thèse fait 133 pages de texte et contient le texte du programme qui implante l'algorithme. Il est écrit en Algol W et fait moins de 1000 lignes de code et commentaires. Le programme avait été testé sur des graphes planaires allant jusqu'à 900 sommets et 2500 arêtes. Il prenait une dizaine de secondes sur un IBM 360/67 pour vérifier la planarité de tels graphes.

Après sa thèse, R. E. Tarjan a été nommé professeur à Cornell où il a repris l'algorithme de sa thèse avec John Hopcroft, professeur à Cornell. Ils en ont donné une version plus courte et plus lisible (voir [6]). Par la suite, de nombreux chercheurs ont implanté cet algorithme en l'améliorant, c'est le cas en particulier de Pierre Rosenstiehl et ses élèves Hubert de Fraysseix et Patrice Ossona de Mendez.

Les travaux de Hopcroft et Tarjan sur les algorithmes et les structures de données ont été couronnés par le prix Turing en 1986, la plus haute reconnaissance pour des chercheurs en informatique. Les travaux de R. E. Tarjan, qui est actuellement professeur à l'Université de Princeton, ont aussi été reconnus par de nombreux autres prix et distinctions.

En ce qui me concerne, j'ai connu cet algorithme au début de l'année 1973. Je venais d'être nommé à Bordeaux et j'ai été destinataire d'un exemplaire de la thèse de R. E. Tarjan grâce à Jean Vuillemin qui avait fréquenté pendant plusieurs années l'Université de Stanford dans une équipe de chercheurs gravitant autour de Donald Knuth. J'avais tout de suite proposé comme sujet de projet, à un groupe d'étudiants de Maîtrise d'Informatique, la réalisation de cet algorithme en Algol sur les machines IBM du Centre de calcul de l'Université. Un jeune garçon brillant avait réussi à conduire ce groupe qui a mené à terme ce projet. Il s'agit de Dominique Gouyou-Beauchamps, qui s'est illustré par la suite en Combinatoire et Algorithmique à Bordeaux puis à Orsay.

Références

[1] M. Albert, M. Bousquet-Mélou. Permutations sortable by two stacks in parallel and quarter plane walks. **arXiv** :1312.4487, 35p. (2013).

[2] L. Auslander, S. Parter. On Imbedding Graphs in the Sphere. *J. Math. Mechanics* **10**, 517–523 (1961).

[3] G. Demoucron, Y. Malgrange, R. Pertuiset. Graphes Planaires : reconnaissance et construction de représentations. *Revue Française de Rech. Op.* **8**, 33–47 (1964).

[4] H. de Fraysseix, P. Ossona de Mendez. Trémaux trees and planarity. *Eur. J. Combinatorics* **33**, 279–293 (2012).

[5] H. de Fraysseix, P. Rosenstiehl. A characterization of planar graphs by Trémaux orders. *Combinatorica* **5**, 127–135 (1985).

[6] J. Hopcroft, R. E. Tarjan. Efficient planarity testing. *J. Assoc. Comp. Mach.* 21, 10–22 (1974).

[7] W. Kocay. The Hopcroft-Tarjan planarity algorithm (1993). `http://www.combinatorialmath.ca/G&G/articles/planarity.pdf`

[8] A. Lempel, S. Even, I. Cederbaum. An algorithm for planarity testing of graphs. In *Theory of Graphs, International Symposium, Rome*, Gordon and Breach, New York (1967).

[9] E. Lucas, *Récréations Mathématiques*, 2e édition, Librairie Albert Blanchard Paris (1891). Nouveau tirage 1992.

[10] R. E. Tarjan. *An efficient planarity algorithm*. Ph.D. Thesis, Stanford University (1972).

[11] P. Rosenstiehl, R. E. Tarjan. Gauss codes, planar Hamiltonian graphs and stack-sortable permutations. *J. Algorithms* **5**, 375–390 (1984).

Le *Bitcoin*, première crypto-monnaie

Jean-Paul Delahaye [1]

Conçu grâce aux progrès de la cryptographie et rendu possible par les réseaux pair à pair (P2P), le Bitcoin *fonctionne depuis 2009. Sa capitalisation dépasse aujourd'hui 5 milliards d'euros.*

Les États ont depuis toujours souhaité émettre et contrôler la monnaie. Les monnaies locales, privées et complémentaires sont peu utilisées et surveillées ; de plus elles dépendent toujours d'une autorité centrale. La cryptographie moderne qui propose des protocoles sûrs et réalisant des opérations qu'on pensait impossibles (comme les signatures digitales), et l'installation dans le monde entier des infrastructures permettant le fonctionnement quasiment sans failles d'un système de réseaux mondiaux, changent la donne. Une nouvelle forme de monnaie (le terme est discuté mais nous l'accepterons) semble possible : purement numérique, gérée collectivement sans intervention d'aucune autorité centrale, permettant des transferts sûrs, instantanés et d'un coût minime d'un point du globe à un autre, ce sont les *monnaies cryptographiques* ou *crypto-monnaies*. Après quelques essais inaboutis, la première version robuste de ce nouveau type de monnaie, le *Bitcoin* (dont la capitalisation représente aujourd'hui 92% de la capitalisation de toutes les monnaies cryptographiques) est né en 2009. Il suscite la passion, aussi bien de ses détracteurs que de ceux qui veulent y croire et qui ont entrepris de construire un secteur économique lié à son existence. Cela pourrait à terme changer profondément l'économie et même nos sociétés.

1. Université de Lille 1, Sciences et Technologies, Laboratoire d'Informatique Fondamentale de Lille, UMR 8022 CNRS, Bât M3-ext, 59655 Villeneuve d'Ascq Cedex. E-mail : delahaye@lifl.fr.

N'est-il pas remarquable que le total des devises sorties d'un protocole cryptographique réussisse l'exploit de valoir aujourd'hui l'équivalent de plusieurs milliards d'euros (plus de 5 milliards d'euros le 27 août 2014, voir `http://coinmarketcap.com/`).

Le *Bitcoin* est le résultat d'un subtil assemblage de protocoles cryptographiques élaborés fin 2008 et mis en œuvre le 3 janvier 2009 par un chercheur — ou peut-être plusieurs ? — caché sous le nom de Satoshi Nakamoto. Comprendre la logique de cette monnaie numérique et tenter de savoir si on peut s'y fier sera notre but. On va voir que le sujet est à la fois passionnant — du fait de l'originalité, du mystère et du succès et de cette construction mathématique et informatique — important — car il s'agit incontestablement d'un nouveau type de monnaie pouvant jouer un rôle central — et particulièrement délicat — personne aujourd'hui ne sait vraiment ce que ce montage numérique va devenir. Les avis les plus extrêmes s'expriment à son sujet [2].

Les monnaies électroniques ne sont pas une nouveauté et, d'une certaine façon, toutes les monnaies sont déjà électroniques. Il y a longtemps en effet que la totalité des opérations bancaires ne sont plus que des jeux d'écritures opérés dans les mémoires des ordinateurs. C'est important de le noter car cela signifie que l'on sait faire des systèmes informatiques robustes manipulant l'argent même quand il s'agit de dizaines de milliards d'euros. Certes les pannes, les « bug », les virus, les pirates informatiques existent, mais on réussit assez bien à s'en protéger : l'informatisation du stockage et du transfert massif d'argent n'est pas la porte ouverte à de colossales catastrophes financières. Si on s'en donne les moyens — ce qui est le cas car quand il s'agit d'argent, c'est sérieux –, on y arrive très bien. Les crises financières comme celle de 2009 n'ont pas pour origine le dysfonctionnement ou la fraude informatique, mais des erreurs commises par des humains qui se trompent dans leurs analyses économiques et financières ou sont trop voraces, voire malhonnêtes.

Autorité centrale

Aujourd'hui toute monnaie repose sur une autorité centrale : une banque avec, derrière, un État ou un ensemble d'États associés. C'est le cas aussi de tous les systèmes de pseudo-monnaies électroniques privées, dénommées monnaies complémentaires ou alternatives. Elles permettent des paiements par internet, le commerce au sein d'un jeu sur le réseau (le *dollar Linden* de *Second Life*) ou la fidélisation des clients (les *miles* des compagnies aériennes, les *points* que votre supérette inscrit sur votre compte à chaque passage aux caisses).

La caractéristique principale des *Bitcoins* est qu'à l'inverse, ils ont été conçus comme devant s'autoréguler sans autorité centrale. Le bon fonctionnement des échanges est garanti par une organisation générale que tout le monde peut examiner

2. voir le *Complément 7* à la fin de l'article.

car tout y est public : les protocoles de bases, les algorithmes cryptographiques utilisés, les programmes les rendant opérationnels et — c'est plus étonnant — les données des comptes.

À tout instant, chacun peut savoir combien il y a de *Bitcoins* sur chaque compte existant et participer à la vérification des nouvelles transactions. Cette publicité totale des outils et du contenu des comptes n'empêche pas l'anonymat puisque les propriétaires des comptes ne sont pas tenus de se déclarer. C'est presque un paradoxe : tout mouvement de *Bitcoins* est public et pourtant le système produit un certain anonymat des détenteurs (dont il faut préciser qu'il est parfois imparfait).

Porte-monnaie virtuel

Posséder des *Bitcoins*, c'est connaître une suite de chiffres et de lettres qui constituent un compte. Une personne peut bien sûr posséder plusieurs comptes. Chaque compte comporte le montant en *Bitcoins* de l'argent qu'il contient, une clef publique qu'on peut laisser circuler (c'est le numéro de compte), et une clef privée qui doit absolument rester secrète car quiconque en dispose peut dépenser l'argent du compte.

Voici des numéros de compte :

1FxkfJQLJTXpW6QmxGT6oF43ZH959ns8Cq
13cia2KGVASavNmRs4niK5RSRfwkB1uLAu
1A6dpTWvoLWmTgwezLmyQti8oDUcLtjTKX

Voici aussi un exemple de clef secrète [3] :

E9 87 3D 79 C6 D8 7D C0 FB 6A 57 78 63 33 89 F4 45 32 13 30 3D A6
1F 20 BD 67 FC 23 3A A3 32 62

Tout support est bon pour conserver la suite de symboles définissant votre compte, y compris un papier, une clef USB ou votre mémoire si vous le souhaitez et êtes capable de retenir les séries de chiffres et de lettres des numéros. Vous pouvez gérer vos comptes à l'aide d'un logiciel appelé « wallet » ou « porte-monnaie ». Vous pouvez aussi confier la gestion de vos comptes à un site internet de confiance, mais alors vous renoncerez à l'anonymat puisque le site en question aura besoin de savoir qui vous êtes. De plus — et c'est arrivé — les gestionnaires du site peuvent s'emparer de votre argent et disparaître avec ou se le faire voler. Le plus prudent sera donc peut-être de gérer soi-même ses comptes sur son ordinateur ou son smartphone. Les logiciels permettant cela, et plus généralement permettant le contrôle des *Bitcoins*, sont souvent développés dans le cadre de projets « open source » (licence MIT) : les programmes ne sont pas secrets et ceux qui le veulent contrôlent ce qu'ils font ou contribuent à leur amélioration. Aujourd'hui les logiciels permettant de gérer sur sa machine des comptes *Bitcoins* sont disponibles gratuitement (le

3. Le *Complément 1* explique ce que sont les *clefs privées* et *clefs publiques*, introduites en cryptographie il y a une quarantaine d'années.

plus souvent) pour tous les types de machines. Pour une liste de porte-monnaie, voir https://bitcoin.org/fr/choisir-votre-porte-monnaie.

Pour avoir des *Bitcoins* sur un compte, il faut qu'un détenteur de *Bitcoins* vous en ait donnés — par exemple en échange d'un bien. Autre possibilité : passer par une plateforme de change qui accepte de convertir des devises classiques (euros, dollars, etc.) en *Bitcoins* (il faut donner son identité).

Il existe aussi des machines distributeurs de *Bitcoins* (ATM-Bitcoin : Automated Teller Machine Bitcoin, voir http://en.wikipedia.org/wiki/Bitcoin_ATM). Vous entrez un ou plusieurs billets (en euros par exemple) dans la machine ; la machine envoie la somme convertie en *Bitcoins* (moins quelques frais) sur le compte *Bitcoin* que vous lui avez indiqué (par exemple en présentant le QRcode du numéro de compte à un lecteur optique). Quelques minutes plus tard vous pouvez vérifier que la somme a été ajoutée au compte en question. Une telle machine (que j'ai testée) se trouve à la *Maison du Bitcoin*, 35 rue du Caire à Paris.

Dernière possibilité pour avoir des *Bitcoins* : vous les gagnez en participant aux opérations de contrôle collectif de la monnaie appelé « minage ». On verra plus loin comment.

La gestion d'un compte doit être menée très soigneusement. Si vous effacez les clefs par mégarde, alors son contenu sera définitivement perdu, exactement comme quand vous lancez par-dessus bord une pièce de monnaie au milieu de l'océan. De nombreux *Bitcoins* ont ainsi déjà été perdus par des utilisateurs imprudents ou négligents. Il n'est pas impossible non plus qu'on vous vole vos *Bitcoins* en allant fouiller votre ordinateur et en y dénichant la clef secrète d'un de vos comptes : celui qui la connait peut dépenser le contenu du compte. Cela peut se produire à l'occasion de l'intrusion d'un hacker accédant aux données de votre machine par l'intermédiaire du réseau. Pour éviter cela, certains porte-monnaie contenant d'importantes sommes en *Bitcoins* sont gardés « au froid », c'est-à-dire sur des ordinateurs non connectés au réseau, ou même que l'on éteint. Sur ces questions consulter https://bitcoin.org/fr/securiser-porte-monnaie.

L'argent, c'est la mémoire

La cohérence des comptes — et donc la solidité de la monnaie *Bitcoin* — se fonde sur un principe général qui est l'application moderne de la théorie « Money is memory » de Narayana Kocherlakota (voir http://www.minneapolisfed.org/research/sr/sr218.pdf).

Ce principe général s'exprime ici sous la forme suivante :

— toutes les transactions faites depuis le début des *Bitcoins* le 3 janvier 2009 sont publiques, et donc, à chaque instant, la somme totale des *Bitcoins* émis est connue de tous, ainsi que le contenu de chaque compte qui en détient ;

— seul celui qui connaît la clef secrète d'un compte peut dépenser son contenu en envoyant tout ou une partie de celui-ci à un autre compte, cela publiquement sur le réseau à la vue de tous, ce qui permet à tous de connaître à chaque instant le contenu de chaque compte ;

— tous ceux qui le souhaitent participent au calcul général de la répartition des *Bitcoins* créés, cela à l'aide de logiciels — open-source et gratuits —, dont la correction et le comportement sont contrôlables par tous.

La cryptographie mathématique à clef publique n'est pas utilisée ici pour cacher de l'information, mais seulement pour signer les transactions, autrement dit, pour que personne ne puisse dépenser à votre place les *Bitcoins* qu'il y a sur vos comptes [4].

Une transaction est irréversible (sauf en cas d'accord explicite des deux contractants pour faire une transaction inverse). En effet, une fois que vous avez dépensé l'argent d'un compte, personne n'a l'autorité pour demander à celui qui a reçu l'argent de le rendre. Lui seul connaît la clef privée de son compte qui est nécessaire pour faire une dépense de l'argent du compte. C'est là une grande différence avec les monnaies numériques à autorité de contrôle centralisée où, assez fréquemment, des transactions sont annulées, parfois plusieurs jours après qu'elles ont été opérées, ce qui donne lieu à toutes sortes d'escroqueries. L'absence d'autorité centrale et l'anonymat des comptes ont cette conséquence dont il faut être conscient : l'échange est rapide et sans frais, mais une fois effectué, il sera impossible d'agir sur celui qui possède le compte ayant reçu vos *Bitcoins*... même s'il ne vous livre pas l'achat que vous pensiez régler.

Un système simplifié pour comprendre

Pour bien saisir l'idée des *Bitcoins* nous allons décrire un système simplifié des *Bitcoins*. Nous énumérerons ses défauts avant de voir comment il a été perfectionné par Satoshi Nakamoto pour aboutir au véritable protocole.

Le *système simplifié des Bitcoins* consiste en un *fichier de comptes* que tous les utilisateurs — dont la liste est fixée à l'avance et ne peut pas évoluer — calculent chacun de leur côté et mettent à jour en permanance, sur une feuille de papier, ou à l'aide de leur ordinateur. Ce *fichier de comptes* tenu à jour par tous les utilisateurs contient toutes les opérations de transfert (transactions) d'un compte vers un autre et permet donc de savoir quelles sommes sont présentes sur les comptes. Ce *fichier de comptes* pourrait ne conserver à chaque instant que l'information du solde de chaque compte, mais s'il contient l'historique des toutes les transactions, c'est plus précis et cela permet aussi le calcul du solde de chaque compte.

Les seules transactions possibles sont du type « le compte A verse la somme S au compte B ». Seul le détenteur du compte A peut enclencher une telle transaction. À chacune d'elles, tous les utilisateurs sont consultés, et donnent leur accord, après

4. Dans le *Complément 2* on explique le protocole d'une transaction.

avoir contrôlé en utilisant leur *fichier de comptes* que celui qui dépense l'argent, A, le possède bien. Une fois l'accord unanime obtenu, la transaction a lieu, ce qui signifie que chacun met son *fichier de comptes* à jour : le solde du compte A est diminué de la somme S, le solde du compte B est augmenté d'autant.

Ce système simplifié fonctionnerait très bien pour gérer une caisse entre une dizaine d'amis qui décideraient par exemple aussi que l'unité de compte de leur système vaut un euro. S'ils sont honnêtes et attentifs, ces amis seront toujours unanimes pour dire à chaque instant quelle somme se trouve sur chaque compte. Ils seront donc toujours d'accord pour valider les demandes honnêtes de dépense d'un compte vers un autre.

L'argent des comptes dans ce système simplifié serait purement virtuel : ce serait la mémoire que le *fichier de comptes* commun en a. Cette caisse permettrait par exemple aux dix amis de vivre ensemble dans un appartement en contribuant inégalement aux dépenses communes (faites avec de vrais euros), que le *fichier de comptes* rééquilibrerait. Quand Jean dépense 100 euros (véritables) pour les courses de l'appartement, ses 9 amis lui versent chacun 10 unités sur son compte. Au démarage des comptes, il n'y aurait pas besoin de faire le moindre versement, chacun se voyant attribuer par exemple 500 unités. Si les dix amis souhaitent mettre fin au système, ils rééquilibrent les comptes en faisant de vrais échanges entre eux. Une fois l'équilibre atteint, ils oublient la caisse et le *fichier de comptes*.

Transposer cela sur le réseau et à une échelle plus grande est difficile. Les échanges électroniques ne sont ni parfaits ni instantanés. Certaines parties d'un réseau sont parfois temporairement déconnectées du reste du réseau. De plus, tous les utilisateurs de *Bitcoins* ne souhaitent pas participer à la vérification continue des transactions et au re-calcul permanant du solde des comptes, car cela demande une puissance informatique non négligeable et beaucoup de mémoire. Faire l'hypothèse que personne ne voudra jamais tricher (par exemple en se retirant après avoir vidé son compte) est un peu naïf. Il est aussi très ennuyeux que la liste des utilisateurs du modèle simplifié soit fixée au départ et ne puisse pas évoluer. Il faut donc perfectionner le modèle simplifié pour l'adapter et lui donner plus de souplesse et de robustesse.

Insistons sur le fait que le *système simplifié des Bitcoins* réalise le plus simplement possible l'idée que « l'argent c'est la mémoire ». Admettre qu'il fonctionne parfaitement pour gérer une caisse entre une dizaine d'amis est le premier pas pour saisir précisément la nature du *Bitcoin*, et pourquoi cela fonctionne et ne constitue en rien une escroquerie. Les insuffisances du système simplifié ont contraint Satoshi Nakamoto à proposer un système plus compliqué, organisé autour d'une série de dispositifs cryptographiques, mais l'idée économique est celle de la caisse des dix amis, gérée par un *fichier de comptes* que chacun suit, opération après opération, en déplaçant des unités monétaires virtuelles.

Pannes, tricheurs, nouveaux arrivants

Bien des avantages résultent des perfectionnements de Nakamoto. En effet le système des *Bitcoins* mis en place en janvier 2009 possède les propriétés suivantes.

— Des nouveaux utilisateurs (comptes) peuvent s'introduire à chaque instant ou se retirer.

— Les utilisateurs ne sont pas tous contraints de suivre une à une les opérations faites d'un compte à un autre.

— Les opérations peuvent être plus complexes que le seul versement d'une somme S du compte A vers le compte B.

— Un change flottant de l'unité de compte (le *Bitcoin*) permet à sa valeur d'évoluer : aucune valeur n'est attribuée au départ au *Bitcoin* ; celle-ci s'établit progressivement, puis une fois adoptée, évolue en fonction de l'offre et de la demande.

— Les pannes de certaines machines du réseau, la coupure et même l'isolement de certaines parties du réseau, de longs délais de transmission entre nœuds du réseau, le désaccord de certains nœuds, les tentatives de tricheries — par exemple les doubles dépenses — n'auront pas d'effet sur le fonctionnement général du système.

Personne ne peut tricher, à la seule condition qu'un nombre minimum de participants acceptent de suivre le *fichier de comptes*.

Les améliorations faisant passer du *modèle simplifié* au *modèle réel du Bitcoin* se fondent sur une série de protocoles particuliers qui font la nouveauté du système des *Bitcoins* et qui aboutissent à un montage subtil et complexe — sinon il aurait été inventé bien avant ! — mais qui rend la monnaie *Bitcoin* résistante à toutes sortes de dysfonctionnements en même temps qu'à toutes sortes de tentatives de manipulation de la monnaie. La participation au contrôle ne sera menée que par les nœuds du réseau qui le souhaitent. Pour éviter que trop peu de nœuds du réseau participent au travail de contrôle un système de rémunération est prévu. Ce délicat agencement a étonné les spécialistes et prouve que l'auteur anonyme qui a conçu les *Bitcoins* est, très probablement, un cryptologue averti ou un groupe incluant au moins un cryptologue averti.

Il ne faut jamais oublier que cette monnaie ne tient que par *la cohérence et l'accord général et unanime de ceux qui y participent et s'entendent sur le contenu de chaque compte que rien ne matérialise, et qu'aucune autorité ne garantit*. Il faut donc que la construction logicielle et cryptographique assure par elle-même que personne ne peut augmenter le total des *Bitcoins* détenus, ni modifier des comptes sans que tout le monde le découvre dans un délai très court. Il n'y a pas de police, la conception même de la monnaie doit empêcher seule la fraude et les dysfonctionnements.

Cela semble impossible et c'est pourquoi la construction conçue par Satoshi Nakamoto est parfois qualifiée de géniale. Personne avant lui n'avait imaginé un système robuste réalisant cette gestion infalsifiable d'un *fichier de comptes*. Le scepticisme sur la robustesse de la nouvelle monnaie, assez fort au départ, tend à s'atténuer. Le fait que la monnaie ait résisté plus de cinq ans malgré les attaques qu'elle a eu à subir est une preuve par l'expérience que le protocole tient. C'est l'une des explications de la valeur actuelle du *Bitcoin*.

Une page toutes les 10 minutes

L'idée pour l'amélioration du modèle simplifié consiste à gérer dans le cadre d'un réseau pair à pair (P2P) un *fichier de comptes* numérique (dont le nom technique est *Blockchain*) qui est complété progressivement par ajout de nouvelles *pages de transactions* (nommées *block*) toutes les 10 minutes environ. Cette *Blockchain* est le *fichier de comptes* du modèle simplifié. Il ne sera pas modifié à chaque opération, mais seulement toutes les 10 minutes.

Chaque ajout d'une page est validé par ceux qui participent à la gestion et à la surveillance décentralisée des comptes et qui détiennent chacun une copie de la *Blockchain*. Pour inciter à participer à la vérification, un concours se déroule en permanence. Une sorte de tirage au sort désigne toutes les dix minutes environ celui des participants qui ajoute la nouvelle page à la *Blockchain*, et qui est rémunéré pour cela par 25 *Bitcoins* créés ex nihilo. Lorsque la nouvelle page est ajoutée à la *Blockchain*, cela valide les transactions qui y apparaissent. La création de *Bitcoins* pour rémunérer le gagnant déterminé toutes les dix minutes est la seule création de *Bitcoins* possible. Tous les *Bitcoins* existants sont apparus de cette façon.

Lors d'une transaction en ma faveur, mon ordinateur connecté au réseau consulte la *Blockchain* qui est un fichier commun partagé par tous les nœuds vérificateurs du réseau P2P du *Bitcoin*. En consultant la *Blockchain*, mon ordinateur contrôle que le compte qui m'envoie des *Bitcoins* ne les a pas déjà dépensés.

Cependant, à cause de la possibilité d'une double dépense simultanée, une transaction n'est considérée comme valide que si elle apparaît dans la *Blockchain*, et donc pour être assuré de l'irréversibilité (par exemple avant d'envoyer le livre qu'on vient de vous acheter en vous faisant parvenir un paiement en *Bitcoins*), il faut attendre dix minutes et voir la transaction sur la nouvelle page du cahier. Les problèmes sont assez rares, et pour de petites transactions, on n'attend donc même pas les 10 minutes. La *Blockchain* peut être explorée à partir de `http://blockchain.info/fr`.

Dédoublement de la *Blockchain*

La possibilité d'ennuis sur le réseau et les délais de transmission des messages ont pour conséquence que parfois deux ajouts de pages à la *Blockchain* se feront

quasi-simultanément dans deux parties éloignées du réseau, créant temporairement un dédoublement de la *Blockchain*. Les deux versions peuvent alors contenir une dernière page sensiblement différente, ce qui rend alors possible une double dépense. L'événement est rare, mais comme il est possible et inévitable à cause de l'imperfection des communications, un procédé de remise en ordre du système est prévu. Les deux *Blockchains* continueront chacune de leur côté à se voir ajouter des pages toutes les 10 minutes environ. Le temps nécessaire à l'ajout est lié au processus de tirage au sort qui désigne le gagnant des 25 *Bitcoins* et donc les ajouts de pages aux deux *Blockchains* malencontreusement créées ne se feront pas à la même vitesse exactement. La *Blockchain* la plus longue (donc celle qui a été complétée de plusieurs nouvelles pages le plus rapidement) est considérée comme la bonne. Cette règle traduite dans les programmes conduit à l'élimination de l'autre *Blockchain* et donc à la reconstitution d'un état cohérent du système où ne persiste qu'une seule *Blockchain* et où d'éventuelles doubles dépenses sont impossibles.

Ces ennuis temporaires, rares mais inévitables, dans la gestion de la *Blockchain* ont pour conséquence au final que pour être certain qu'une transaction est définitivement valide — c'est important dans le cas de grosses sommes –, il faut non pas attendre dix minutes, mais plusieurs fois 10 minutes. On considère qu'une heure produit une garantie parfaite.

Une ruée vers l'or numérique

La désignation des gagnants des 25 *Bitcoins* toutes les dix minutes se fait par un processus cryptographique qui en assure la parfaite honnêteté et surtout une totale imprévisibilité et « infalsifiabilité » (il est impossible de manipuler le choix du gagnant que personne absolument ne peut connaître à l'avance). Ce tirage au sort se fait selon un procédé qui vous donne d'autant plus de chances de gagner que vous disposez de plus de puissance de calcul. Plus vous acceptez de consacrer des ressources de calcul à tenter de gagner, plus vous augmentez vos chances de gagner [5]. Le travail fait par vos machines pour tenter de gagner porte le nom de *minage*, par analogie au travail dans une mine qui conduit ceux qui ont de la chance à trouver de l'or.

Aujourd'hui participer à ces tirages au sort, et donc participer au contrôle général des comptes, est assez tentant puisque 25 *Bitcoins* s'échangent contre plus de 9000 euros (27 août 2014). Du coup, les *mineurs de Bitcoins*, comme ils se nomment, se sont multipliés ce qui renforce le système de contrôle général des comptes. Les *mineurs de Bitcoins* ont progressivement perfectionné leurs outils avec l'espoir d'augmenter leurs chances de gagner. Dans un premier temps, les mineurs ont programmé des cartes graphiques pour faire, le plus rapidement possible, les calculs demandés par le minage (les cartes graphiques disposent d'une puissance de calcul importante

5. Voir le *Complément 4* sur les *preuves de travail* qui explique la méthode cryptographique qui réalise cela.

qu'on peut détourner). Les cartes graphiques sont maintenant insuffisantes pour avoir de bonnes chances de gagner car au fur et à mesure que plus de mineurs se sont mis à jouer, il est devenu plus difficile de gagner. Le système de Nakamoto est conçu pour qu'il y ait un gagnant toutes les dix minutes environ et il s'ajuste automatiquement pour que ce temps moyen ne diminue pas. Des entreprises de hardware se sont donc mis à fabriquer des cartes et des machines spécialisées dont le seul but est de miner les *Bitcoins*. La puissance — et donc la consommation électrique ! — consacrée au minage s'est considérablement accrue depuis un an. Le phénomène ressemble un peu à une ruée vers l'or, sauf qu'ici tout se déroule dans le monde des réseaux et des ordinateurs en faisant circuler des bits d'information et rougir des microprocesseurs dédiés.

La puissance de calcul nécessaire pour gagner fait qu'il devient impossible même pour un acteur très puissant — et on sait qu'il y en a ! — de s'emparer de tous les gains. L'analyse générale du protocole des *Bitcoins*, effectuée dès 2008 par Nakamoto, montre précisément que si un acteur pouvait disposer de la moitié de la puissance consacrée au minage, alors il serait en mesure de perturber gravement le fonctionnement de la monnaie *Bitcoin*. L'accroissement des efforts faits pour miner des *Bitcoins* rend de plus en plus difficile de réunir ces 50%, et indirectement renforce donc la monnaie *Bitcoin*. Le système conçu par Nakamoto se consolide au fur et à mesure que des gens s'y intéressent : plus le cours du *Bitcoin* monte, plus il devient intéressant de miner des *Bitcoins*, plus nombreux sont ceux qui minent, plus le *Bitcoin* devient robuste y compris aux attaques d'acteurs très puissants, et donc, plus son cours a des chances de monter.

Vingt mille fois le plus puissant des ordinateurs

La puissance globale consacrée aujourd'hui au minage de *Bitcoins* est de 2 250 000 pétaflops (le 27 août 2014, voir `http://bitcoinwatch.com/` ; 1 pétaflops = 10^{15} opérations en virgule flottante). C'est plus de 20 000 fois la puissance du plus puissant ordinateur du monde (le « Tianhe-2 » détenu par la Chine qui espère atteindre en 2015 une puissance de 100 pétaflops) et c'est largement plus de cent fois la puissance cumulée des 500 ordinateurs les plus puissants.

C'est considérable ! Ce qu'on peut voir comme un énorme gâchis empirera si le *Bitcoin* s'impose et que son cours (qui bien sûr détermine l'argent que les mineurs sont prêts à investir) progresse. Les défenseurs du *Bitcoin* argumentent en disant qu'une monnaie basée sur l'or est aussi absurde que le *Bitcoin*. En effet on n'utilise pas l'or qui reste dans des coffres, qu'il faut surveiller. Même pour les monnaies usuelles comme l'euro ou le dollar, une quantité très importante de ressources est consacrée à leur création (conception et impression des billets) à leur transport, à leur surveillance (dans des coffres aussi), à la recherche des faussaires, etc.

Même si le minage des *Bitcoins* est utile à la consolidation de la monnaie *Bitcoin*, la chose apparaît parfois absurde car ces calculs menés pour augmenter les chances de gagner 25 *Bitcoins* n'ont aucune utilité directe. C'est sans doute pourquoi l'informaticien Sunny King a conçu un procédé pouvant se substituer à celui aujourd'hui utilisé pour miner les *Bitcoins*. Avec sa méthode, les calculs faits par les candidats produisent un tirage au sort équitable et infalsifiable, mais, en même temps, ils font découvrir des chaînes de nombres premiers intéressant les mathématiciens. King a mis en œuvre son idée en créant en juillet 2013 une nouvelle crypto-monnaie *Primecoin* concurrente des *Bitcoins* et qui a déjà conduit à découvrir des chaînes de nombres premiers record. La capitalisation de tous les *Primecoins* est, aujourd'hui (27 août 2014), d'environ un million d'euros. Le *Primecoin* arrive en position 26 par importance dans le classement (par capitalisation décroissante) des crypto-monnaies. Peut-être qu'on finira par imaginer encore mieux, et concevoir un protocole de minage qui — par exemple — aide aux calculs nécessaires pour déterminer le repliement des protéines, ce qui serait cette fois utile à la recherche médicale.

Aujourd'hui, en utilisant seul son ordinateur pour miner des *Bitcoins*, on n'a aucune chance de gagner des *Bitcoins*. Cette situation a conduit à la création de « coopératives de mineurs » (« *mining pool* »). Les mineurs associés décident de partager les gains qu'ils feront en proportion de la puissance de calcul qu'ils consacrent à miner. Ces regroupements assurent donc de gagner régulièrement, car la coopérative (si elle est puissante) remportera assez fréquemment les 25 *Bitcoins* qu'elle redistribuera à ses membres. Toutefois ne vous faites pas d'illusion : en rejoignant une coopérative, si vous n'offrez que la puissance de votre ordinateur personnel, la part qui vous reviendra sera minuscule et ne paiera sans doute même pas l'électricité que vous dépenserez pour faire tourner votre machine. Vous trouverez une liste de coopératives de minage en `http://www.bitcoin.fr/pages/Mining-pools`.

Vingt-et-un millions de *Bitcoins* en tout

Les protocoles de Nakamoto (qui sont fixés quasi définitivement et traduits dans les programmes utilisés pour la gestion décentralisée de la monnaie *Bitcoin*) prévoient que tous les quatre ans, la somme distribuée aux gagnants du minage est divisée par deux. Au départ, elle était de 50 *Bitcoins*, le 22 novembre 2012, elle est passée à 25 *Bitcoins*, et elle passera à 12,5 *Bitcoins* dans 2 ans. Du fait qu'un *Bitcoin* ne peut pas être divisé en unités plus petites que le cent millionième de *Bitcoin*, le gain toutes les dix minutes finira par arriver à 0. Un calcul montre que le processus d'émission de nouveaux *Bitcoins* aura cessé en 2140 et qu'il y aura alors un total de 21 millions de *Bitcoins*. À partir de cette date, aucun nouveau *Bitcoin* ne sera plus jamais créé.

De manière à éviter que tous les mineurs — essentiels au bon fonctionnement du protocole — désertent et que la construction et la validation continue de la *Blockchain* cesse, Nakamoto a prévu qu'à chaque transaction, on donne une commission à celui qui ajoutera la page contenant la transaction. L'intérêt de miner sera donc préservé, même au-delà de 2140. Donner une telle commission n'est pas obligatoire et aujourd'hui même si vous ne laissez rien, vos transactions sont quand même validées et passent dans la *Blockchain*. Après 2140, il deviendra souhaitable de laisser un petit quelque chose à chaque transaction... on a le temps d'y penser. En fait dans certaines circonstances pour décourager les trop petites transactions qui encombreraient le système (et pourraient même le faire s'écrouler), laisser une commission — très minime — est déjà obligatoire : voir `http://bitcoinfees.com/` ou `https://en.bitcoin.it/wiki/Transaction_fees`.

L'impossible devenu réalité... et valeur

Le système mis en fonctionnement il y a bientôt cinq ans tient solidement. Dans un premier temps, le cours du *Bitcoin* était dérisoire. Il valait environ 10 euros en janvier 2013. Il a atteint 200 euros le 9 avril 2013. Il a ensuite chuté, puis en décembre 2013 le *Bitcoin* est monté jusqu'à près de 900 euros (voir `http://www.bitcoin.fr/pages/Cours-du-bitcoin`). Depuis il est revenu à 390 euros (le 27 août 2014). Certains ont fait d'excellentes affaires, soit en en achetant quand le *Bitcoin* ne valait rien, soit en minant les *Bitcoins* quand c'était facile. L'instabilité du cours fait qu'acheter des *Bitcoins* est un pari.

La page `http://www.bitcoin.fr/pages/Cours-du-bitcoin` indique clairement : « Bitcoin *est une expérience inédite. N'y investissez que ce que vous pouvez vous permettre de perdre.* »

Au fur et à mesure que son usage se répandra et que des commerçants accepteront d'être payé en *Bitcoins*, on peut espérer que le cours se calmera. Tout a été dit sur le *Bitcoin* et les avis sont partagés sur son devenir à long terme [6]. Il semble bien cependant que quelque chose d'important se soit produit avec la naissance de cette monnaie que quelques années d'existence et une valorisation du total des *Bitcoins* qui se compte en milliards d'euros ont installé pour longtemps dans le monde réel. Une question cependant doit être posée : pourquoi est-ce que cela ne s'est pas produit plus tôt ?

La réponse est assez simple : avant 2009, il était impossible d'envisager une telle monnaie qui doit son existence aux progrès récents dans plusieurs domaines et à leur association dans la construction de Nakamoto :

> (1) Il fallait un réseau mondial fiable ; sans lui rien ne serait possible ; le *Bitcoin* cesserait d'exister immédiatement en cas d'arrêt du réseau (il reprendrait à sa remise en marche).

6. Voir le *Complément 7*.

(2) Rien de possible non plus sans d'importantes puissances de calcul et de mémorisation informatique : la *Blockchain* fait 21 giga-octets (le 27 août 2014). Ce n'est que récemment — grâce à la loi de Moore — qu'elles sont devenues suffisantes pour que la tenue et la vérification des comptes — même en considérant toutes les transactions depuis la création de la monnaie — soient possibles simultanément par des milliers d'acteurs différents et indépendants se contrôlant donc les uns les autres.

(3) Le génie d'un (ou plusieurs ?) informaticien qui, en s'appuyant sur une discipline ayant formidablement progressé depuis trente ans — la cryptographie mathématique et informatique — a produit un protocole étonnamment subtil et robuste que personne n'imaginait possible, et qui a réussi à le faire fonctionner.

(4) Essentielle aussi est la communauté des passionnés qui s'occupe des programmes libres et des réseaux P2P rendant l'utilisation pratique des *Bitcoins* possible gratuitement par tous et évitant qu'un groupe, une banque ou un État ne s'empare de ce qui est parfois présenté comme une monnaie commune, universelle et démocratique.

Ceux qui, à propos du *Bitcoin*, parlent de pyramide de Ponzi ou de construction sur du vide susceptible de s'écrouler du jour au lendemain n'ont probablement rien compris à cette nouveauté remarquable, produit des mathématiques, des avancées techniques et de l'ingéniosité de Nakamoto. Ils n'ont peut-être rien compris non plus aux monnaies actuelles qui reposent toutes sur la confiance (depuis l'abandon général de la convertibilité en or) et donc sont — autant que le *Bitcoin* — de la création de valeur à partir de rien.

Plutôt que de continuer à faire reposer l'indispensable monnaie sur des institutions centralisées qui se sont cent fois révélées défaillantes, il serait peut-être plus rationnel de s'appuyer sur une organisation et un fonctionnement général de la monnaie fondés sur des protocoles cryptographiques reconnus, sur des logiciels que tout le monde peut examiner, et sur une transparence des comptes les rendant infalsifiables. Cela mérite au moins d'être discuté. Le *Bitcoin* pose une multitude de problèmes techniques, réglementaires, législatifs, fiscaux, policiers, et politiques. L'avenir du *Bitcoin* dépend de la façon dont ils seront traités et résolus. Sous la forme qu'il a aujourd'hui ou sous une autre forme (une autre crypto-monnaie), il est très probable qu'il continuera d'exister, et que ceux qu'il gêne et qui voudraient s'en débarrasser échoueront : on ne désinventera pas le *Bitcoin*.

Pour en savoir plus...

[1] James Angel and Douglas M. McCabe. *The Ethics of Payments : Paper, Plastic, or Bitcoin ?*, 2014.
http://papers.ssrn.com/sol3/Papers.cfm?abstract_id=2379233

[2] Jan Bergstra and Peter Weijland. *Bitcoin : a money-like informational commodity*. arXiv preprint arXiv :1402.4778, 2014.
http://arxiv.org/pdf/1402.4778.pdf

[3] Félix Brezo, Pablo G. Bringas. *Issues and Risks Associated with Cryptocurrencies such as Bitcoin*. SOTICS 2012 : The Second International Conference on Social Eco-Informatics.
http://www.thinkmind.org/index.php?view=article&articleid=sotics_2012_1_40_30101

[4] Nicolas Courtois, Marek Grajek, Rahul Naik. *The Unreasonable Fundamental Incertitudes Behind Bitcoin Mining*, 2013.
http://arxiv.org/pdf/1310.7935v2.pdf

[5] Jean-Paul Delahaye. *Le* Bitcoin*, la cryptogmonnaie*. Pour la science, décembre 2013, pp. 86-91.
http://www.lifl.fr/~delahaye/pls/2013/241.pdf

[6] Jean-Paul Delahaye. *Les preuves de travail*. Pour la science, avril 2014, pp. 86-91.
http://www.lifl.fr/~delahaye/pls/2014/245.pdf

[7] Danielle Drainville. *An Analysis of the Bitcoin Electronic Cash System*. University of Waterloo 12-2012.
https://math.uwaterloo.ca/combinatorics-and-optimization/sites/ca.combinatorics-and-optimization/files/uploads/files/Drainville,%20Danielle.pdf

[8] Ittay Eyal, Emin Gün Sirer. *Majority is not Enough.* arXiv preprint arXiv :1311.0243, 2013.
http://arxiv.org/pdf/1311.0243v5.pdf

[9] FBI Directorate of Intelligence Cyber Intelligence. *Bitcoin Virtual Currency : Unique Features Present Distinct Challenges for Deterring Illicit Activity*, 2012.
http://www.wired.com/images_blogs/threatlevel/2012/05/Bitcoin-FBI.pdf

[10] Stephen Goldfeder et al. *Securing Bitcoin wallets via threshold signatures*, 2014.
http://www.cs.princeton.edu/~stevenag/bitcoin_threshold_signatures.pdf

[11] Philippe Herlin. *La révolution du Bitcoin et des monnaies complémentaires*. Editions Eyrolles, 2013.

[12] Joshua Kroll, Ian Davey, Edward Felten. *The Economics of Bitcoin Mining, or Bitcoins in the Presence of Adversaries*. The Twelfth Workshop on the Economics of Information Security, 2013.
http://www.weis2013.econinfosec.org/papers/KrollDaveyFeltenWEIS2013.pdf

[13] William Luther. *Bitcoin is Memory*. Social Science Network, 2013.
http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2275730

[14] Satoshi Nakamoto. *Bitcoin : A Peer-to-Peer Electronic Cash System*, 2009.
http://bitcoin.org/bitcoin.pdf

[15] Michael Nielsen. *How the bitcoin protocol actually works*, 2013.
http://www.michaelnielsen.org/ddi/how-the-bitcoin-protocol-actually-works/

[16] Mark Pinklinton. *Bitcoin and Complexity Theory : Some Methodological Implications*. Social Science Network, 2013.
http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2340007

[17] Fergal Reid, Martin Harrigan. *An analysis of Anonymity in the Bitcoin System*. Security and Privacy in Social Networks, 2013, pp 197-223.
http://arxiv.org/pdf/1107.4524.pdf

[18] Tetsuya Saito. *Bitcoin a Search Theoretic Approach*. Research Institute of Economic Science College of Economics, Nihon University, 2013.
http://www.eco.nihon-u.ac.jp/center/economic/publication/pdf/13-01.pdf

[19] Bitcoin Foundation (consulté en août 2014).
https://bitcoinfoundation.org

[20] Sites des développeurs principaux du Bitcoin (consulté en août 2014).
https://bitcoin.org/en/ouhttps://bitcoin.org/fr/

[21] Wikipedia, Bitcoin (consulté en août 2014).
http://en.wikipedia.org/wiki/Bitcoin ou http://fr.wikipedia.org/wiki/Bitcoin

[22] Bitcoin wiki (consulté en août 2014).
https://en.bitcoin.it/wiki/Main_Page

[23] La Blockchain (consulté en août 2014).
http://blockchain.info/

Complément 1

La signature par cryptographie à double clef

Un protocole de signature à double clef est la donnée de deux fonctions f et g permettant de signer les messages et d'interpréter les signatures. Ces fonctions sont connues de tous.

Supposons par exemple qu'Alice dispose de deux clefs A_{pri} (clef privée) et A_{pub} (clef publique). Ce sont des suites de symboles ou, ce qui revient au même, des nombres entiers.

La clef A_{pub} est transmise à tout le monde, mais la clef A_{pri} n'est connue que par Alice. Si le protocole de signature à double clef est bon, il est impossible en pratique de déduire A_{pri} à partir de A_{pub}.

Les deux fonctions f et g servent à signer un message et à lire la signature.
Exemple. Soit M un message à signer. Alice applique f aux données A_{pri} et M

$$f(A_{pri}, M) = M',$$

ce sera le message signé par Alice.

Toute personne ayant en main M' et connaissant la clef publique d'Alice vérifiera sans peine que c'est bien Alice qui a signé le message. Pour cela, elle appliquera la fonction de lecture g aux données A_{pub} et M', ce qui produira M :

$$g(A_{pub}, M') = M.$$

Le fait qu'il soit nécessaire d'appliquer la clef publique d'Alice à M' pour prendre connaissance du message empêche toute falsification du message signé. Il est parfois

commode pour Alice de transmettre à la fois M et M', M' servant seulement à contrôler pour ceux qui le veulent que Alice a bien signé M avec sa clef privée.

Il existe de nombreuses façons de choisir les fonctions f et g. Celle qui sert pour la monnaie *Bitcoin* est basée sur la cryptographie à courbes elliptiques, dite ECDSA (*Elliptic Curve Digital Signature Algorithm*). La courbe employée est **secp256k1**. Une autre solution aurait pu être le RSA (Rivest-Shamir-Adleman), plus connu mais demandant des clefs plus longues.

Et si le protocole de signature du Bitcoin *était cassé ?*

Si le protocole de signature venait à être cassé et que quelqu'un disposant d'une clef publique A_{pub} sache calculer facilement la clef privée associée A_{pri}, alors cette personne serait en mesure de dépenser le contenu de tous les comptes. À condition de le faire progressivement pour ne pas se faire repérer, cette personne serait donc très riche ! Les détenteurs des comptes ne s'en apercevraient que lorsque, consultant la somme liée à leur compte (et donc allant lire la *Blockchain*), ils découvriraient que leur compte est vide.

Notons bien que ce vidage des comptes (par quelqu'un sachant calculer A_{pri} à partir de A_{pub}) peut s'opérer même si le porte-monnaie est sur un disque dur déconnecté du réseau et éteint, ou même si la clef privée du porte-monnaie n'est écrite que sur une feuille de papier, ou même si elle est perdue.

On considère cependant que personne ne disposera jamais du moyen pratique de calculer A_{pri} à partir de A_{pub}, ou que si cela se produit, la faiblesse du protocole de signature aura été repérée avant qu'elle soit devenue réellement utilisable, et donc, que la communauté des développeurs *Bitcoin* aura opéré à temps le changement de cette partie du protocole *Bitcoin*. La possibilité d'adapter et donc de corriger des faiblesses qu'on repèrerait dans le protocole *Bitcoin* est prévue dans le protocole *Bitcoin*. Mais cette éventuelle réparation d'une défaillance du protocole de signature n'est envisageable que si on s'en aperçoit...

On peut même imaginer que si des vidages de comptes étaient opérés en trop grand nombre, la communauté *Bitcoin* arrêterait toutes les transactions et tenterait de réparer le protocole et de rétablir les comptes dans un état antérieur à la détection de la fraude. De telles situations obligeant à une intervention des équipes de développement du *Bitcoin* se sont déjà produites deux fois. Une première fois, en août 2010, à la suite de la découverte d'un bug qui avait permis la création de 184 millions de *Bitcoins* frauduleux qui furent immédiatement annulés. Une seconde fois en mars 2013, à la suite d'une duplication de la *Blockchain* (voir `http://en.wikipedia.org/wiki/History_of_Bitcoin`). Au total, même dans les cas extrêmes de mise en cause du protocole *Bitcoin*, il est donc possible d'en limiter les conséquences et donc d'éviter l'écroulement total du système. Notons que lors de ces rares événements, le principe de décentralisation du *Bitcoin* cesse temporairement d'être respecté, et qu'une équipe restreinte de développeurs reprend la main et

opère des choix qui ne sont pas inscrits d'avance dans le protocole. Ces choix sont faits publiquement et à la suite d'accords entre les développeurs, mais — temporairement et pour sauver le *Bitcoin* de catastrophes graves — il s'agit quand même d'écarts inquiétants à la doctrine de base.

Complément 2

LE PROTOCOLE D'UNE TRANSACTION

Lorsqu'Alice veut faire un paiement en *Bitcoins* à Bernard (par exemple en échange d'un livre), leurs ordinateurs vont opérer une série d'opérations. Ces opérations sont gérées automatiquement par le logiciel qu'ils ont installé sur leur ordinateur. Les communications sur le réseau *Bitcoin* se font directement. Ce *réseau* pair à pair ou P2P (peer-to-peer) est le cœur du système. L'existence de tels réseaux est essentielle pour la monnaie *Bitcoin* qui n'est gérée par aucun nœud central qui contrôlerait l'ensemble des communications. La transaction qui résulte des échanges entre Alice et Bernard est publique (tous les ordinateurs présents sur le réseau y auront accès) et permet la mise à jour par tous de la *Blockchain* qui indique combien de *Bitcoins* sont déposés dans chaque compte existant.

(1) Alice souhaite envoyer N *Bitcoins* à Bernard.

(2) Bernard communique sa clef publique B_{pub} (c'est-à-dire son numéro de compte) à Alice.

(3) Alice constitue un message M de transaction contenant la clé publique de Bernard B_{pub} et la somme N à transférer : $M = B_{pub}N$.

(4) Alice signe la transaction M avec sa clé privée, c'est-à-dire calcule une suite de symboles $M' = f(A_{pri}, M)$ qui, avec sa clef publique, redonne M :

$$g(A_{pub}, f(A_{pri}, M)) = g(A_{pub}, M') = M.$$

Tout le monde peut donc contrôler que c'est Alice qui a signé, mais personne ne peut signer à sa place.

(5) Alice diffuse la transaction signée sur le réseau afin qu'elle soit vue par tout le monde.

Le protocole réel est légèrement plus compliqué (il contrôle qu'Alice dispose bien de la somme N dans son porte-monnaie).

En regardant cette transaction depuis l'extérieur, tout le monde voit qu'Alice a donné son accord pour transférer N *Bitcoins* à Bernard.

Ne disposant pas de la clef privée d'Alice, personne d'autre qu'elle ne peut envoyer une telle transaction sur le réseau. Son envoi est donc la preuve qu'Alice était d'accord pour le transfert. Tout le monde considérera donc le transfert comme valide.

Complément 3

LE HACHAGE ET LES PREUVES DE TRAVAIL

Une fonction de hachage est une fonction h qui à toute suite de symboles S (par exemple des chiffres) associe une autre suite de symboles (plus courte) $h(S) = R$ et surtout qui est telle qu'il est impossible en pratique pour une valeur possible R de la fonction h de trouver un S tel que $h(S) = R$. Si h est une bonne fonction de hachage les valeurs $h(S)$ produites par quelqu'un qui essaie diverses valeurs pour S, sont aussi imprévisibles que si elles étaient tirées au hasard avec une roue de loterie.

Disposant d'une telle fonction h on peut définir un *travail* qu'il sera impossible de faire rapidement :

Travail de niveau k : Trouver S tel que $h(S)$ commence par k fois le symbole '0'.

Plus k est grand, plus il faut essayer de nombreux S avant de trouver un S convenable. En moyenne, ceux qui prétendent avoir trouvé un tel S ont fourni un travail de calcul qui est d'autant plus important que k est grand.

C'est un peu comme si on demandait à quelqu'un de lancer deux dés jusqu'à obtenir un double 6 (il faudrait en moyenne qu'il les lance 36 fois pour réussir).

On vérifiera facilement que les S prétendument trouvés sont bons, en en demandant la communication, et en calculant $h(S)$ qui doit être un résultat avec k '0' en tête.

L'idée de ces « preuves de travail » a été proposée en cryptographie dans le but par exemple de lutter contre le courrier électronique indésirable (spam) : si chaque ordinateur qui veut accéder à ma boîte de messages doit prouver qu'il a effectué un certain travail dépendant de ma boîte (par exemple trouver un S tel que $h(S)$ commence par 10 fois '0' pour une fonction h qui m'est propre), il devient impossible à celui qui le voudrait d'envoyer de milliers de spam, car la preuve de travail nécessaire à chaque envoi devient trop lourde au total. Cette barrière à l'entrée d'une boîte à lettres électronique n'est pas ennuyeuse pour celui qui ne veut envoyer que quelques messages, car la preuve de travail à fournir est raisonnable si le nombre d'envois est petit. À ma connaissance, cette méthode n'est pas encore utilisée pour lutter contre le spam.

La technique de la *preuve de travail* est au cœur du système des *Bitcoins*. C'est elle qui est utilisée pour le tirage au sort de celui qui ajoute une page à la *Blockchain* et remporte, toutes les dix minutes, 25 nouveaux *Bitcoins*. Tous ceux qui participent se lancent dans la recherche du S, le premier qui en trouve un est le gagnant. Un paramètre contextuel dans la définition de h, connu seulement au moment où un nouveau tirage est lancé, empêche les participants de commencer à chercher le S en avance.

Ajustables en faisant varier l'entier k, les *preuves de travail* exigées pour emporter les 25 *Bitcoins* créés toutes les 10 minutes sont devenues de plus en plus difficiles au cours des mois. Depuis que des puces spécialisées ont été conçues pour calculer très vite les $h(S)$ la difficulté du travail demandé a été augmentée, cela de façon à ce que le temps moyen entre deux gains reste toujours de 10 minutes environ. Cette augmentation de la difficulté des *preuves de travail* est conforme au protocole fixé par Nakamoto, le concepteur des *Bitcoins*.

Pour un article général sur les preuves de travail, voir Jean-Paul Delahaye, *Les preuves de travail*, Pour la science, avril 2014, pp. 86-91, `http://www.lifl.fr/~delahaye/pls/2014/245.pdf`.

Complément 4

Les cours, la capitalisation, les « bulles »

La première page de la *Blockchain Bitcoin* a été publiée le 3 janvier 2009.

Les *Bitcoins* sont émis à un rythme régulier :

— 50 *Bitcoins* nouveaux ont été créés toutes les 10 minutes, jusqu'au 22 novembre 2012.

— Depuis, 25 nouveaux *Bitcoins* sont créés toutes les 10 minutes.

— Le nombre de *Bitcoins* émis sera divisé par 2 tous les quatre ans.

— Le total des *Bitcoins* émis ne dépassera jamais 21 millions.

Aujourd'hui (26 août 2014, 9h20) le nombre de *Bitcoins* émis est 13 188 050. Ils valent exactement 5 002 201 euros (pour une mise à jour, aller en `http://bitcoinwatch.com/`).

Le 9 février 2011 : Le *Bitcoin* a atteint la parité avec le dollar.

Le 11 avril 2013 : Effondrement de la valeur du *Bitcoin* qui passe de 266 $ à 105 $ et descend même brièvement à 60 $.

En décembre 2013, il atteint 1147 $ (900 euros). Depuis, il est redescendu et vaut 504 $ (27 août 2014).

La taille de la *Blockchain* est 21 gigaoctets (27 août 2014).

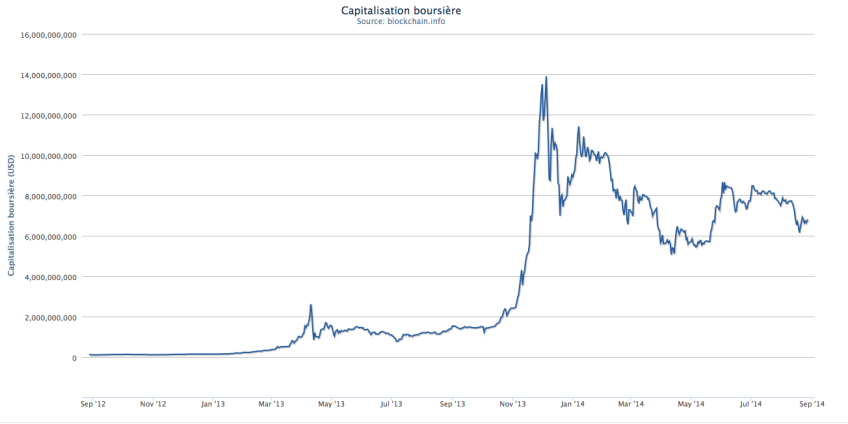


FIGURE 1. Capitalisation boursière jusqu'au 27 août 2014

D'autres informations et graphiques sur la monnaie *Bitcoin* sont mis à jour en continu en :

`https://blockchain.info/fr/charts` (voir Figure 1)
`http://bitcoinwatch.com/`
`http://www.quandl.com/markets/bitcoin`
`http://www.bitcoin.fr/pages/Cours-du-bitcoin`
`http://www.coindesk.com/price/`
`http://coinmarketcap.com/` (Capitalisation de toutes les crypto-monnaies)

Les cinq « bulles » du Bitcoin.

Les montées brusques du cours du *Bitcoin*, suivies de baisses, elles-aussi violentes, ne sont pas un phénomène nouveau comme le croient certains observateurs des récentes variations. En examinant ce cours depuis 2010 (voir Figure 2), on observe que c'est en réalité la cinquième « bulle » qui gonfle et éclate (partiellement).

Ces « bulles » doivent-elles vraiment être appelées des « bulles » puisqu'à chaque fois le cours finit par se stabiliser au-dessus de sa valeur d'avant la « bulle » ?

En 2010, on passe de moins d'un dixième de dollars à plus d'un dollar.
En 2011, on passe d'un dollar à 5 dollars.
En 2012, on passe de 5 dollars à 10 dollars.
Au premier semestre 2013 on passe de 10 dollars à 100 dollars.
Au second semestre 2013, on passe de 100 dollars à plus de 400 dollars.

Parler de « bulles qui éclatent » est étrange : si vous êtes pris dans le gonflement et détenez des *Bitcoins* alors, en vendant après l'éclatement, vous réalisez une excellente affaire. Personne ne regrettera d'avoir acheté à 10 si, quelques mois après, il

FIGURE 2. Les cinq « bulles » du *Bitcoin*

dispose de 100, et cela même s'il a manqué de vendre à 200 ! Il semblerait plus sage de parler de « bouffées de volatilité », car à part des pics temporaires (sans doute excessifs) qui font à l'arrivée quand même monter le cours, il n'y a jamais eu à proprement parler d'éclatement de bulle ! Bien sûr, cela peut venir, et le cours du *Bitcoin* pourrait retomber à moins d'un dollar comme l'annoncent certains.

Complément 5

LA NAISSANCE DU *Bitcoin*

Le *Bitcoin* a été défini en 2008 par un personnage qui se présente sous le nom de Satoshi Nakamoto et qui a écrit avoir travaillé deux ans à la conception de sa monnaie. Nakamoto garde l'anonymat. Il se peut qu'il s'agisse en fait d'un groupe de plusieurs personnes, mais vue la façon dont le protocole qui gère la nouvelle monnaie numérique a été fixé, il est certain que Nakamoto possède des connaissances de très bon niveau en cryptographie. Une sorte de grande traque se déroule sur internet pour identifier qui est le génial inventeur. On analyse la façon dont il s'est exprimé en anglais, on fait des listes de personnes pouvant avoir les compétences requises... et on spécule.

Récemment des arguments assez forts semblent désigner Nick Szabo de l'Université George Washington à Washington D.C. aux États-Unis. Voir :

http://techcrunch.com/2013/12/05/who-is-the-real-satoshi-nakamoto-one-researcher-may-have-found-the-answer/

https://likeinamirror.wordpress.com/2013/12/01/satoshi-nakamoto-is-probably-nick-szabo/

http://www.coindesk.com/linguistic-researchers-name-nick-szabo-author-bitcoin-whitepaper/

L'anonymat des utilisateurs des *Bitcoins* est assuré par le fait que seuls les numéros de compte et les contenus des comptes sont nécessaires au maintien de la cohérence de la *Blockchain*. La clef privée d'un compte assure son propriétaire que lui seul pourra dépenser l'argent qui s'y trouve et donc, en théorie, l'anonymat des détenteurs de comptes est assuré. La réalité est plus complexe et l'anonymat n'est pas absolu. D'une part, on peut suivre le déplacement des *Bitcoins* d'un compte à l'autre et de cette façon en déduire certaines informations sur le propriétaire unique d'une série de comptes visiblement gérés par une seule personne. De plus, au moment de transformer des *Bitcoins* en Euros ou en une devise classique l'anonymat n'est plus possible.

Des chercheurs, dont Sergio Lerner, en étudiant la *Blockchain* concluent que Nakamoto possède probablement l'équivalent de 10% des *Bitcoins* émis à ce jour. Il est en effet à peu près certain qu'au lancement de la monnaie, il fut le seul à « miner les *Bitcoins* » pour se constituer un pécule personnel et qu'il a regroupé ce pécule sur quelques comptes en nombre assez limités qu'on réussit plus ou moins à suivre. L'invention des *Bitcoins* aurait donc permis à Nakamoto de se constituer une belle fortune de l'ordre de 500 millions d'Euros (somme à réévaluer en fonction du cours). Il lui sera sans doute difficile de les remettre sur le marché sans dévoiler son identité (et sans faire baisser les cours)... à moins qu'il procède lentement et qu'il imagine et mette en œuvre des techniques de brouillage faisant perdre sa trace, comme il s'en développe. Voir par exemple :

http://app.bitlaundry.com/

http://bittumble.com/

https://en.bitcoin.it/wiki/Category:Mixing_Services

https://en.bitcoin.it/wiki/Mixing_service

Complément 6

LE PLUS ET LE MOINS DU *Bitcoin*

***Nouveautés, forces et qualités des* Bitcoins**

— La monnaie *Bitcoins* est basée sur un réseau pair à pair et des logiciels libres et gratuits. Elle est indépendante de toute banque, n'est soumise à aucune autorité centralisée et offre une transparence sans équivalent dans toute l'histoire des monnaies.

— Les transactions de *Bitcoins* sont rapides et irréversibles. Personne ne peut agir sur les *Bitcoins* de vos comptes sans votre consentement (sauf s'il en connaît la clef privée).

— Il n'y a pas de frais de transaction ou de gestion, ou ceux-ci sont minimes (électricité, réseau, commissions infinitésimales ou déterminées par l'utilisateur).

— Le nombre de *Bitcoins* est rigoureusement fixé et ne dépassera jamais 21 millions. Avec les *Bitcoins*, vous échappez au risque qu'un acteur dominant (une banque centrale) décide de faire fonctionner la planche à billets et par ce moyen indirect capte une part de votre argent par l'inflation créée.

— Anonymat : le réseau fonctionne à partir de comptes. Posséder un compte, c'est connaître la clef privée qui lui est associée. L'identité des utilisateurs n'est utile à aucun moment. (L'anonymat n'est cependant pas total : voir le Complément précédent.)

— Un *Bitcoin* peut être divisé en fractions de *Bitcoin* jusqu'au 1/100 000 000.

— Le *Bitcoin* (à cause du nombre maximum de *Bitcoins* en circulation) est sans doute intrinsèquement déflationniste : il est amené à prendre petit à petit de la valeur. Non seulement vos économies ne sont pas rongées par l'inflation, mais elles s'apprécient.... à moins que vous ne perdiez tout parce que le protocole *Bitcoin* s'effondre !

— Le *Bitcoin* a été conçu d'une manière telle que l'intérêt de ceux qui s'en occupent est qu'il fonctionne bien. Du fait de sa conception, plus il prend de la valeur plus les contrôles auxquels il est soumis sont nombreux.

— Les protocoles et programmes permettant la gestion des transactions peuvent évoluer, mais cela ne peut se faire que lentement, par une sorte de vote de la communauté, et donc en respectant l'intérêt commun.

Doutes, fragilités et risques des **Bitcoins**

— L'anonymat de Nakamoto « l'inventeur » des *Bitcoins* et l'argent qu'il a gagné facilement au départ de la monnaie créent un sentiment désagréable et font craindre une combine. Certains ont même imaginé que le *Bitcoin* est une création de la NSA.
— Aujourd'hui le *Bitcoin* est économiquement très petit à côté des autres monnaies internationales auxquelles il ne peut donc pas prétendre se substituer : il y a quelques milliards de dollars en *Bitcoins*, alors que la devise américaine par exemple a été émise à hauteur de 1 200 milliards (uniquement en billets).
— La monnaie *Bitcoin* repose sur des protocoles cryptographiques dont la robustesse n'est pas prouvée mathématiquement. Il faut faire confiance à la science cryptographique d'aujourd'hui et à son état de l'art.
— Le système de gestion des *Bitcoins* repose sur un ensemble de protocoles qui ont été rendus opérationnels par des programmes. Des erreurs peuvent s'y trouver.
— Le *Bitcoin* reste assez compliqué à comprendre et suscite donc la méfiance du plus grand nombre (qui ne saisit peut-être pas mieux la façon dont fonctionne les monnaies classiques !).
— Relativement peu de sites et peu de commerçants acceptent les *Bitcoins* aujourd'hui.
— Le *Bitcoin* (comme l'argent liquide) favorise le blanchiment d'argent sale, facilite les trafics en tout genre, et permet la fraude fiscale.
— Le *Bitcoin* semble intrinsèquement déflationniste ce que certains considèrent comme négatif car cela constitue un frein à la circulation de l'argent, et surtout, les cours du *Bitcoin* sont très volatils du fait des incertitudes qui l'entourent.
— Le *Bitcoin* pourrait être l'objet d'interdictions ou de contrôles stricts imposés par des États voulant protéger leurs propres monnaies. Il n'est pas impossible que le *Bitcoin* soit victime d'attaques menées par des agences comme la NSA qui tenteraient de briser toute confiance en lui, pour maintenir les monopoles monétaires actuels.
— L'anonymat y est imparfait (du fait que toutes les transactions sont publiques).
— Le succès des *Bitcoins* a inspiré toutes sortes de Nakamoto et des dizaines de nouvelles crypto-monnaies directement copiées sur lui ont vu le jour. Bien qu'aujourd'hui le total des autres crypto-monnaies représente moins de 10% du *Bitcoin*, certaines crypto-monnaies un peu différentes et mieux conçues pourraient capter l'intérêt et faire se déplacer l'argent misé aujourd'hui sur les *Bitcoins*.

— L'évolution possible des protocoles et programmes — prévue mais au fonctionnement délicat — conduit à la mise en place d'une forme d'administration centralisée constituée par l'ensemble des nœuds les plus puissants du réseau collectif de contrôle. Cela pourrait conduire à terme à faire ressembler le *Bitcoin* aux monnaies usuelles dont Nakamoto voulait se démarquer [7].

Complément 7

QUELQUES AVIS TRANCHÉS

Contre

`http://www.gaullistelibre.com/2013/05/bitcoin-ou-la-folie-de-la-monnaie.html`

• 14 mai 2013, Laurent Pinsole (Blogueur, militant dans le mouvement « Debout la République ») :

Bien sûr, les plus libéraux des libéraux s'enthousiasmeront de cette expérimentation du marché, sans la contrainte étatique. Mais sur le fond, ces monnaies virtuelles posent des problèmes qui justifient largement leur interdiction. Tout d'abord, leur conception, garantissant parfois un anonymat complet, peut permettre le recyclage d'argent sale. Ensuite, il s'agit d'un flux monétaire sur lequel l'État a moins de maîtrise, ce qui peut faciliter la désertion fiscale des entreprises qui les utilisent.

Mais surtout, la monnaie est un pilier des sociétés modernes, à la fois unité nécessaire à l'échange, mais aussi unité de compte qui permet l'épargne ou le prêt. En cela, il est difficile de ne pas comprendre qu'il s'agit d'un service public, qui relève éminemment de l'État pour garantir la valeur de cette monnaie. Confier cela au marché aboutit précisément aux errements constatés avec Bitcoin, à savoir une variation totalement erratique de sa valeur, qui peut faire du mal aux citoyens.

En outre, l'exemple de Bitcoin amène à se poser la question de qui profite de la création de ces monnaies virtuelles. En effet, les dix millions d'unités créées ne sont pas gagées sur quoi que ce soit. Du coup, il faut reconnaître que quelqu'un profite de la création ex nihilo de ces monnaies et l'absence de régulation en ce domaine pose également problème. Bref, rien ne justifie que de la monnaie soit créée de la sorte par des entreprises privées. C'est l'État qui doit gérer la monnaie.

7. Sur ce point voir : Joshua Kroll, Ian Davey, Edward Felten, *The Economics of Bitcoin Mining*, ou *Bitcoins in the Presence of Adversaries*, The Twelfth Workshop on the Economics of Information Security, 2013.

L'expérience Bitcoin montre à nouveau tous les dangers du « laisser-faire » en matière monétaire. Non seulement, les marchés seuls sont incapables d'assurer la stabilité nécessaire à la monnaie, mais les bénéfices de la création de la monnaie doivent rester dans le giron public.

---

http://www.atlantico.fr/decryptage/et-folle-envolee-bitcoin-annoncait-en-realite-disparition-prochaine-benoist-rousseau-903206.html

• 20 novembre 2013, Benoist Rousseau (Ingénieur, Informaticien et historien économiste) :

Cette monnaie virtuelle est aujourd'hui trop instable, ces phases de hausses extrêmes font aussi place à des krachs réguliers, en avril 2013 le Bitcoin a ainsi perdu 60% de sa valeur en moins de 5 jours. Qui accepterait aujourd'hui d'être payé en Bitcoins avec de telles fluctuations ? Le Bitcoin est actuellement en train de devenir un simple instrument spéculatif qui n'est pas sans rappeler le début d'une formation d'une bulle que l'on retrouve dans de nombreux krachs boursiers.

[...] L'envolée récente des cours au mois de novembre 2013 est une réponse en miroir à l'effondrement du Bitcoin du mois d'octobre 2013. On parlait alors de la fin du Bitcoin... Du fait du faible nombre de Bitcoins en circulation, il est très aisé pour une structure organisée disposant de capitaux importants de faire fluctuer le cours comme elle le souhaite. C'est une monnaie « casino » actuellement.

[...] Du fait de la faiblesse du volume des transactions, le Bitcoin doit « appartenir » en fait à quelques personnes fortement capitalisées qui peuvent « jouer » avec lui. Cette monnaie « libre » est sûrement l'une des plus manipulables et des moins libres paradoxalement... Quelques organisations douteuses peuvent se comporter comme les banques centrales sur cet actif et faire la pluie et le beau temps, le prix à payer pour l'utopie d'une dérégulation totale... Ce ne sont plus les États qui peuvent garantir un jeu basé sur des règles établies et connues de tous afin d'éviter les ententes, les délits d'initiés... mais les plus forts qui agissent comme ils l'entendent sur les cours du Bitcoin, sans limite, sans règle. Faut-il préférer les mafias ou les banques centrales ? Quant à savoir quand cette bulle éclatera, personne ne le sait, c'est le principe des bulles, tout le monde la voit mais personne ne sait quand. L'envolée récente peut être vue comme une tentative de sauvetage du Bitcoin après son effondrement d'octobre. Dans un marché si étroit, les variations sont toujours excessives, à la hausse comme à la baisse.

---

http://finance.blog.lemonde.fr/2013/11/20/le-bitcoin-est-une-monnaie-de-casino-qui-profite-a-des-manipulateurs-incontroles/

• 4 décembre 2013, Georges Ugeux (Banquier d'Affaires) :

Le Bitcoin est une « monnaie » de casino qui profite à des manipulateurs incontrôlés.

Je n'ai aucune opinion sur le souhait de ceux et celles qui s'adonnent au jeu, et utilisent la monnaie des casinos. Après tout, le jeu est aussi vieux que le monde. Mon problème est quand les fondateurs de Bitcoin prétendent que c'est une monnaie légitime. Cette semaine a été une brillante démonstration de l'incompétence des politiciens et de la stupidité moutonnière des marchés. Depuis les tulipes hollandaises, nous savons que les modes sont un phénomène régulier dans les marchés. Le tout, c'est de s'en sortir à temps.

Lorsque le château de cartes s'effondrera, qui seront les victimes ? Ceux et celles qui, inconscients du danger, se retrouveront avec des jetons de 900 dollars et découvriront qu'ils ne valent plus rien. C'est là que se trouve la légitimité d'une intervention sérieuse et musclée du législateur. La « Fondation Bitcoin » comporte plus de 100 membres permanents et le même nombre de membres annuels. Ils standardisent, promeuvent et protègent le Bitcoin. Inutile de dire qu'ils n'assument aucune responsabilité quelconque sur la valeur ou les actifs. La « fondation » fait du marketing, et ses membres n'apportent de crédibilité que par leur nombre.

`http://www.banque-france.fr/fileadmin/user_upload/banque_de_france/publications/Focus-10-stabilite-financiere.pdf`

• 5 décembre 2013, La Banque de France (Focus n°10) :

Les bitcoins : une monnaie non régulée qui n'offre aucune garantie.

[...] Une conception qui alimente la spéculation.

[...] Des plates-formes internet proposent, sans aucune garantie de prix ni de liquidité, l'achat/vente de bitcoins contre des devises ayant cours légal.

[...] Par son caractère anonyme, le bitcoin favorise le contournement des règles relatives à la lutte contre le blanchiment des capitaux et le financement du terrorisme.

[...] Même si le bitcoin ne remplit pas à ce jour les conditions pour devenir un support d'investissement crédible et poser ainsi un risque significatif pour la stabilité financière, il représente un risque financier certain pour les acteurs qui le détiennent.

[...] N'offrant aucune garantie de sécurité, de convertibilité et de valeur, le bitcoin présente peu ou pas d'intérêt pour une utilisation par les acteurs économiques, au-delà des aspects marketing et publicitaire, tout en les exposant à des risques importants.

[...] En limitant la quantité maximale de bitcoins pouvant être créée et en faisant fluctuer le rythme de création au cours du temps, les concepteurs ont « organisé » la pénurie de cette monnaie virtuelle et lui ont ainsi conféré son caractère hautement spéculatif.

---

http://wallstreetpit.com/101867-greenspan-i-guess-bitcoin-is-a-bubble/
http://www.businessinsider.com/alan-greenspan-bitcoin-comment-reaction-2013-12

• 5 décembre 2013, Alan Greenspan (Économiste, ancien président de la Réserve Fédérale, la Banque centrale des États-Unis) :

I guess Bitcoin is a bubble.

You really have to stretch your imagination to infer what the intrinsic value of Bitcoin is. I haven't been able to do it. But if you ask me, "Is this a bubble in Bitcoin ?" "Yeah, it's a bubble".

---

http://www.washingtonpost.com/blogs/the-switch/wp/2013/12/10/this-finance-expert-thinks-bitcoin-will-fall-99-percent-by-june/

• 10 décembre 2013, Timothy B. Lee (Journaliste économique) :

In recent weeks, some Bitcoin critics have been rethinking their initial Bitcoin skepticism. But others are as convinced as ever that the cryptocurrency is doomed. One of the harshest critics is Mark Williams, who teaches finance at the Boston University School of Management. He predicts that in the first half of 2014, bitcoins will lose almost 99 percent of their value, falling below $10.

---

https://al3x.net/2013/12/18/bitcoin.html https://al3x.net/

• 18 décembre 2013, Alex Payne (Informaticien, écrivain) :

Bitcoin, Magical Thinking, and Political Ideology.

[...] Most charitably, Bitcoin is regarded as a flawed but nonetheless worthwhile experiment, one that has unfortunately attracted outsized attention and investment before correcting any number of glaring security issues.

To those less kind, Bitcoin has become synonymous with everything wrong with Silicon Valley : a marriage of dubious technology and questionable economics wrapped up in a crypto-libertarian political agenda that smacks of nerds-do-it-better paternalism. With its influx of finance mercenaries, the Bitcoin community is a grim illustration of greed running roughshod over meaningful progress.

Far from a "breakthrough", Bitcoin is viewed by many technologists as an intellectual sinkhole. A person's sincere interest in Bitcoin is evidence that they are disconnected from the financial problems most people face while lacking a fundamental understanding of the role and function of central banking. The only thing "profound" about Bitcoin is its community's near-total obliviousness to reality.

[...] The push toward Bitcoin comes largely from the libertarian portion of the technology community who believe that regulation stands in the way of both progress

and profit. Unfortunately, this alarmingly magical thinking has little basis in economic reality. The gradual dismantling of much of the US and international financial regulatory safety net is now regarded as a major catalyst for the Great Recession. The "financial or political constraints" many of the underbanked find themselves in are the result of unchecked predatory capitalism, not a symptom of a terminal lack of software.

Silicon Valley has a seemingly endless capacity to mistake social and political problems for technological ones, and Bitcoin is just the latest example of this selective blindness

---

http://www.atlantico.fr/decryptage/que-bitcoin-revele-enfer-terrestre-que-serait-paradis-libertarien-eric-verhaeghe-929581.html?page=0,1

• 19 décembre 2013, Eric Verhaeghe (Écrivain, journaliste, ancien élève de l'ENA) :

Ce que le Bitcoin nous révèle de l'enfer terrestre que serait le paradis libertarien.

Il est bien probable que les péripéties que les bitcoins traversent donnent le signal prématuré d'une mort doctrinale pour le libertarisme, et d'un scepticisme généralisé pour les monnaies privées.

Premier problème : le bitcoin est accusé de couvrir les pires activités mafieuses. Alors que le secret bancaire est de plus en plus fragilisé, et que tout flux financier doit, de façon grandissante, être « blanchi », le caractère totalement privé du bitcoin attire les convoitises. Quel système mieux adapté que la relation « peer to peer », loin des règles contraignantes de la puissance publique, pour recycler de l'argent sale ?

Deuxième problème : le bitcoin est une monnaie extrêmement spéculative qui enrichit quelques détenteurs futés. Selon certaines sources, la moitié de la masse de bitcoin serait détenue par moins de mille particuliers. Un tiers du stock serait détenu par moins de cinquante personnes. Cette extrême concentration souligne le premier intérêt du système : enrichir ses créateurs, et rien de plus.

Car, troisième problème : le bitcoin est extrêmement volatil. En quelques jours, il peut perdre une part importante de sa valeur. C'est le cas en ce moment : ce mercredi, le bitcoin a perdu 50% de sa valeur, après des annonces inquiétantes en Chine.

Cet épisode marquera les esprits. Un monde sans État et sans pouvoir public est toujours possible. Mais c'est un monde opaque, inégalitaire, et fondamentalement instable. Or la stabilité et la confiance sont des conditions nécessaires à la prospérité.

---

http://krugman.blogs.nytimes.com/2013/12/28/bitcoin-is-evil/?_r=0

• 28 décembre, Paul Krugman (Économiste, Prix Nobel d'économie 2008) :

Bitcoin is evil. [...] I have had and am continuing to have a dialogue with smart technologists who are very high on BitCoin — but when I try to get them to explain to me why BitCoin is a reliable store of value, they always seem to come back with explanations about how it's a terrific medium of exchange. Even if I buy this (which I don't, entirely), it doesn't solve my problem. And I haven't been able to get my correspondents to recognize that these are different questions.

`http://www.livemint.com/Opinion/fEXo8r2OcAIZHWOMbZmmfO/Investors-beware-of-Bitcoin.html`

• 13 janvier 2014, Avinash Persaud (Directeur d'une société spécialisée dans les produits financiers) :

[...] Bitcoin is a cryptographer's wet dream rather than a useful monetary system.

[...] Bitcoins have been stolen and there are allegations of covert creation of bitcoins and cornering of supply by a few larger computer networks.

[...] Bitcoins will be worthless within a couple years.

[...] Because of its anonymity, bitcoins are attractive in the laundering of illicit activities.

[...] Sooner rather than later, holders will find that there are a diminishing number of greater fools left to buy bitcoins from them, and its price will collapse. Don't be tempted.

Pour

`http://bitcoin.org/bitcoin.pdf`

• 2008, Satoshi Nakamoto (personnage anonyme concepteur du Bitcoin, voir *Complément 6*) :

A purely peer-to-peer version of electronic cash would allow online payments to be sent directly from one party to another without going through a financial institution. Digital signatures provide part of the solution, but the main benefits are lost if a trusted third party is still required to prevent double-spending. We propose a solution to the double-spending problem using a peer-to-peer network. The network timestamps transactions by hashing them into an ongoing chain of hash-based proof-of-work, forming a record that cannot be changed without redoing the proof-of-work. The longest chain not only serves as proof of the sequence of events witnessed, but proof that it came from the largest pool of CPU power. As long as a majority of CPU power is controlled by nodes that are not cooperating to attack the network, they'll

generate the longest chain and outpace attackers. The network itself requires minimal structure. Messages are broadcast on a best effort basis, and nodes can leave and rejoin the network at will, accepting the longest proof-of-work chain as proof of what happened while they were gone.

---

http://www.paristechreview.com/2012/01/20/bitcoin-devise-complementaire-universelle/

• 20 janvier 2012, Pierre Noizat (Cofondateur de Paymium (Bitcoin Central) et spécialiste du Bitcoin) :

Tout comme l'or, les bitcoins peuvent être assimilés à des obligations sans échéance. Mais à la différence de l'or, les bitcoins peuvent être divisés indéfiniment et n'impliquent aucun frais de stockage. D'après les estimations de GFMS, fin 2010, le stock d'or extrait se chiffre à hauteur de 166 600 tonnes, ce qui, au prix moyen de l'année 2010, représente 6500 milliards de dollars dont environ 2400 milliards constituent des réserves privées ou officielles, sous forme de pièces ou de lingots. Le stock total moins les 30 000 tonnes correspondant aux réserves officielles mondiales en août 2011, nous donne une estimation de 1230 milliards de dollars pour le marché de l'or, en tant que réserve de valeur. Si l'on devait calculer un taux de change du bitcoin avec le dollar sur la base de ces chiffres, on obtiendrait un taux de change de 600$ pour 1 BTC, si les bitcoins représentaient 1% du marché privé de l'or comme instrument de couverture.

Dans le même esprit, si l'économie bitcoin devait croître à hauteur de 5% du PIB des États-Unis (c'est-à-dire 725 milliards de dollars) et en supposant une vélocité monétaire du bitcoin égale à 50, équivalente au dollar pour les petits montants, un bitcoin représenterait l'équivalent de 700$. Cela équivaudrait à une valeur projetée de 15 milliards de dollars pour le réseau bitcoin. C'est un ordre de grandeur cohérent avec la capitalisation boursière de Visa, Inc. (55 milliards de dollars) ou de MasterCard (39 milliards de dollars). Acheter des bitcoins aujourd'hui, c'est acheter des actions pour un nouveau réseau mondial de transactions électroniques. À 10$ en août 2011, soit une valorisation du réseau bitcoin établie à 210 millions de dollars, les bitcoins sont clairement sous-évalués même en admettant que d'autres devises universelles pourraient entrer en lice.

[...] Du fait qu'ils sont échangés électroniquement, les bitcoins, contrairement à l'or, sont infiniment divisibles et permettent une grande vélocité monétaire. Ainsi, une forte déflation des prix ne ferait que limiter bitcoin à un rôle de réserve de valeur, plus pratique que l'or. En fait, la déflation des prix n'aurait de conséquences économiques néfastes que si les bitcoins étaient la devise exclusive d'une aire géographique donnée. Cela n'est absolument pas le cas puisqu'en tant que devise complémentaire, les bitcoins coexistent avec la devise locale sponsorisée par l'État, sans chercher à la remplacer. Les prix dans les commerces de proximité continueront à

être exprimés en devises locales. Dans une transaction électronique en ligne, le prix exprimé en devise universelle peut facilement être ajusté en temps réel par rapport à un taux de change variable. C'est uniquement pour les transactions qui ne sont pas en ligne que la stabilité des prix est une exigence pour toute devise universelle. En bref, la déflation des prix augmente l'attractivité des bitcoins en tant que valeur refuge et n'affecte que marginalement son application en tant que moyen d'échange.

Conclusions. Dans une économie mondialisée, la naissance d'une ou de plusieurs devises universelles est inéluctable dès lors qu'elle est technologiquement faisable et économiquement souhaitable.

Bitcoin, en tant que première devise du genre, a largement ouvert la voie à de nouvelles applications. En particulier, bitcoin peut fortement améliorer l'efficacité des transferts d'argent là où il y en a le plus besoin, notamment pour l'aide au développement, longtemps considérée (selon les mots de l'économiste Peter Bauer) comme « une excellente façon de transférer l'argent des pauvres des pays riches aux riches des pays pauvres ». Bitcoin peut profiter de la généralisation des téléphones mobiles dans les pays en voie de développement pour permettre de transférer de l'argent directement, sans passer par des intermédiaires, qu'ils soient bureaucratiques ou bancaires. L'institution ou l'organisme non-gouvernemental responsable du transfert assignera simplement des adresses bitcoin aux destinataires et les marchants locaux pourront alors réaliser des transferts d'argent et des paiements en bitcoins.

Cette technologie permet à la fois un nouveau type de transaction sur le réseau et une nouvelle devise universelle.

Par analogie, il est intéressant de noter que la gouvernance du World Wide Web est régie par une organisation à but non-lucratif — le W3C — composée de 300 membres parmi les plus grandes entreprises du secteur des hautes technologies. Clairement, tout appui de la part d'un gouvernement à l'un des membres du W3C peut être contrebalancé par les autres si cela ne convient pas à l'intérêt général. Si ce principe parvient avec succès à réguler un domaine où la technologie sert de nouvelles méthodes de production et de partage du savoir, il est permis d'espérer qu'une organisation similaire peut également superviser les caractéristiques techniques du protocole bitcoin. Cela permettrait à bitcoin de préserver son intégrité et son potentiel d'innovation face aux aléas des mesures macro-économiques.

Philippe Herlin, *La révolution du Bitcoin et des monnaies complémentaires*, Éditions Eyrolles, 2013.

• Mai 2013, Philippe Herlin (Docteur en économie du CNAM, chercheur en finance) :

Soyons clairs, le bitcoin constitue une remise en cause frontale de leur capacité à tout régenter dans le domaine monétaire et financier. Les États, pour la plupart, creusent leur déficit, alourdissent leur dette, demandent à leur banque centrale de

racheter cette dette (c'est la « planche à billets »), ce qui dévalue la valeur de la monnaie. Et ils s'autorisent désormais à ponctionner directement les comptes bancaires et à instaurer un contrôle des mouvements de capitaux, comme on l'a vu à Chypre. Le bitcoin permet d'échapper à tout cela, d'avoir un compte inviolable, une monnaie solide, et d'effectuer facilement des transactions. Face aux États qui jouent aux apprentis sorciers avec leur monnaie, on peut considérer que les citoyens ont le droit à l'autodéfense.

[...] La révolution monétaire, cela consiste tout simplement à ne plus être prisonnier d'une seule monnaie ni du système bancaire, et à retrouver sa liberté, à choisir les systèmes auxquels on accorde sa confiance. Les monnaies officielles resteront encore longtemps prédominantes, mais elles seront de moins en moins exclusives. Elles seront mises en concurrence, comparées, et il y a peu de chance que ça tourne en leur faveur. De la même façon, suite à l'affaire de Chypre et à la ponction opérée sur les comptes bancaires, la volonté de sortir du circuit bancaire ne va faire que croître.

Les alternatives existent, nous l'avons vu, avec les monnaies complémentaires en fort développement depuis les années 1990, la possible remonétisation de l'or, et surtout le bitcoin, universel, électronique, indépendant de toute puissance étatique ou financière, offrant à chacun un compte inviolable et des transactions parfaitement sécurisées.

Les grands groupes internationaux des télécoms s'intéressent également de près à cette révolution et, déjà, des systèmes bancaires alternatifs complets existent dans plusieurs régions du monde. Les grands acteurs de l'électronique grand public ne restent pas inactifs : Google avoue s'intéresser au bitcoin, Amazon a créé sa propre monnaie avec l'Amazon Coin, et d'autres initiatives ne devraient pas manquer d'intervenir.

Quelle sera l'ampleur et la rapidité de cette révolution ? Difficile à dire, mais plus les dettes publiques et privées augmentent à travers le monde, plus les planches à billets des banques centrales tournent, plus le phénomène sera rapide. Les États feront tout pour garder le contrôle de leur monnaie, y compris en en créant une nouvelle si l'ancienne explose en vol, mais tous ses détenteurs auront été ruinés au passage, au contraire de ceux qui possèdent des bitcoins, de l'or, des monnaies complémentaires indexées sur des matières premières, et des actifs réels en général (immobilier, œuvres d'art, terrains agricoles). À bon entendeur...

Face à des États qui s'endettent toujours plus, qui manipulent leur monnaie et n'hésitent pas — comme à Chypre — à ponctionner les comptes et à restreindre les mouvements de capitaux, les citoyens ont à leur disposition de nouveaux moyens d'autodéfense.

http://www.latribune.fr/journal/edition-du-2008/opinions/780725/le-bitcoin-une-exigence-democratique-.html

• 19 août 2013, La tribune :

Le Bitcoin, une exigence démocratique.

[...] Tout bitcoin ou fraction de bitcoin composant le montant d'une transaction trouve son origine dans une transaction précédente, dite transaction de génération, qui crée 25 bitcoins. Ces transactions interviennent toutes les dix minutes et les bitcoins générés sont distribués par un algorithme complexe aux nœuds du réseau qui vérifient et relaient les transactions. Parce que tout le monde peut participer ainsi librement au réseau bitcoin, cette technologie change notre rapport de négociation avec les banques, jusqu'à présent entièrement dominé par le banquier pour tous nos échanges économiques. De même que, grâce à l'email, nous utilisons les services de la Poste lorsqu'ils nous sont vraiment utiles, nous pourrons désormais décider de passer par la banque seulement quand elle aura une valeur ajoutée réelle.

Le bitcoin, plus pratique et plus sécurisé.

Dans la plupart de nos transactions effectuées dans un contexte connu et pour des montants limités (commerce de proximité, amis, famille, etc), la technologie bitcoin sera préférée pour des raisons pratiques et politiques. Sur un plan pratique en effet, les commissions de transactions sur le réseau bitcoin sont fixées librement par le payeur qui peut les mettre à zéro. Et contrairement à un virement bancaire ou un paiement par carte bancaire, une transaction bitcoin est diffusée en quelques secondes sur le réseau. Au plan politique, le citoyen préférera un système de paiement comme bitcoin qui lui laisse maîtriser la divulgation des données de transaction plutôt que la capture systématique de ces données par des multinationales.

---

http://www.washingtonpost.com/blogs/the-switch/wp/2013/11/05/when-will-the-people-who-called-bitcoin-a-bubble-admit-they-were-wrong/

• 5 novembre 2013, Timothy B. Lee (Journaliste économique) :

When will the people who called Bitcoin a bubble admit they were wrong ?

[...] The word "bubble" comes up a lot in discussion of Bitcoin. Bubble talk last peaked in April of this year, when the value of one Bitcoin soared from less than $100 at the start of the month to an all-time high of $266 on April 10. And in the next few weeks, it appeared that the skeptics had been vindicated. Bitcoin prices tumbled, falling to a low of about $50 before stabilizing back around $100. Bitcoin critics argued that the Bitcoin-buying public had fallen prey to irrational exuberance, and that one unit of the crypto-currency couldn't possibly be worth $266. But then something interesting happened. After a few months of relative stability, Bitcoin prices began rising sharply again. Since the start of October, Bitcoin prices have doubled from 125*to*250*. That's just shy of the all-time high of $266.*

That poses a problem for those who called the high prices of last April a bubble. When people describe a market move as a "bubble", they're generally referring to more than a rapid price increase that's followed by a decline — that's just garden-variety volatility. Rather, the term "bubble" refers to appreciation that is driven by hype rather than a sober consideration of market fundamentals. When bubbles pop, they tend to stay popped, as the market comes to its senses and realizes that the earlier, higher price was irrational. So if a price returns to its earlier highs a few months later, that suggests that the original appreciation might not have been so irrational after all.

`http://www.contrepoints.org/2013/12/09/149249-bitcoin-la-monnaie-qui-derange`

• 9 décembre 2013, Jérémy Berthet (Ingénieur informatique) :

La nature déflationniste et décentralisée du système Bitcoin le rend complètement étranger aux mécanismes d'une pyramide de Ponzi. On ne peut pas en dire autant de nos monnaies d'État, réputées sûres et régulées, mais contrôlées par une entité centrale au sommet très accommodante avec les premiers arrivants que sont les banques et qui ne paient pas le contrecoup de l'inflation, à la différence des citoyens en bas de l'échelle.

Tout comme l'échange « peer to peer » de fichiers devrait amener l'industrie culturelle à revoir son modèle, le Bitcoin, pour autant qu'on le laisse vivre, devrait conduire les acteurs de la finance à s'adapter à un « nouveau » paradigme monétaire plus sûr et plus fiable que l'actuel. Internet est une technologie dont l'évolution permet de réduire drastiquement le nombre d'intermédiaires commerciaux. Si Bitcoin n'y parvient pas pour la monnaie, suite à une réaction de peur des États, vous pouvez être certain que d'autres y arriveront plus tard, ce n'est qu'une question de temps.

`http://www.digitaltrends.com/opinion/everyone-planet-invest-bitcoin/`

• 10 décembre 2013, Andrew Couts (Journaliste spécialisé dans les nouvelles technologies) :

At its heart, Bitcoin is a protocol in the same way the Web (HTTP) and the Internet (TCP/IP) are protocols. Whereas HTTP dictates how Internet-connected computers "talkèè to each other, the Bitcoin protocol allows for the encrypted, anonymous transfer of funds (i.e. a unit of value) across the Internet.

To take the Internet comparison further, just as the Internet allows for the free flow of information, Bitcoin allows for the free flow of "money" in ways impossible before its invention. Use a legacy banking system, PayPal, or anything besides cash in an envelope, and you're looking at fees, delays, and headaches. Bitcoin solves all

of this, offering the possibility for instant fund transfers to anywhere in the world, zero or extremely low transaction fees (even when transferring millions of dollars worth of Bitcoin), and the ability to send someone Bitcoin (or fractions of a Bitcoin) simply by knowing their Bitcoin address — no other account information, names, or routing numbers needed.

On top of all these benefits, the Bitcoin protocol is based upon extremely strong encryption and a public ledger system (called the "blockchain") that uses distributed transaction confirmations to ensure the integrity of every single Bitcoin transaction. (Told you it would get nerdy...)

Bitcoin Believers will tell you that no one really knows Bitcoin's true potential, just as the Vint Cerf and Tim Berners-Lee never envisioned what the Internet and Web would become when they created the respective protocols on which these world-changing technologies are based.

---

http://www.lefigaro.fr/secteur/high-tech/2013/12/11/01007-20131211ARTFIG00469-pourquoi-le-bitcoin-fait-grincer-des-dents.php

• 11 décembre 2013, Pierre Noizat (Cofondateur de Paymium (Bitcoin Central) et spécialiste du Bitcoin) :

Si le monde veut rentrer dans l'économie numérique, il faut s'adapter aux changements technologiques plutôt que de crier au loup. [...] Comme toute nouvelle technologie, le bitcoin reste réservé aux personnes averties, conclut Pierre Noizat. Mais une fois que les banques auront compris qu'il s'agit d'un vrai business, les barrières à l'entrée seront levées et la monnaie plus accessible pour les particuliers.

Dans cinq ans, le bitcoin sera adopté par tous.

---

http://cdixon.org/2013/12/12/coinbase/

• 12 décembre 2013, Chris Dixon (Entrepreneur et investisseur internet américain) :

Bitcoin is the first plausible proposal for an economic protocol for the Internet.

This matters for two reasons :

1) It fixes serious problems with existing payment systems that depend on centralized services to verify the validity of transactions. These services are both expensive (roughly a 2.5% tax on all transactions) and prone to failure (Internet payment fraud is rampant).

2) More importantly, Bitcoin is a platform upon which new technologies can be developed. Developers have created some early applications, and speculated about future applications. Some potential applications include : a) micropayments as a replacement for banner ads or subscription fees, b) machine-to-machine payments to reduce spam and denial-of-service attacks, c) a way to offer low-cost financial

services to people who, because of financial or political constraints, don't have them today.

But to proliferate widely, Bitcoin needs a killer app the same way HTTP had web browsers and SMTP had email clients. That's why today I'm excited to announce that Andreessen Horowitz is leading a $25M financing of Coinbase, a service that provides an accessible interface to the Bitcoin protocol. Consumers can use Coinbase to convert to and from other currencies and to pay for goods and services. Merchants can use Coinbase to accept payments and convert currencies. Developers can build new services using Coinbase's API.

Coinbase has grown extremely fast and is now the most widely used Bitcoin service in the US. The founders of Coinbase, Brian Armstrong and Fred Ehrsam, have worked closely with banks and regulators to ensure that the service is safe and compliant. We think Coinbase can significantly accelerate Bitcoin's proliferation, and as that happens the Internet will enter a new phase of invention and opportunity.

[Une réponse en `https://al3x.net/2013/12/18/bitcoin.html`]

`http://motherboard.vice.com/blog/bitcoin-becomes-a-real-job-and-wall-street-is-hiring`

• 2 janvier 2014, Alec Liu (Écrivain spécialisé dans les nouvelles technologies) :

Bitcoin Becomes a Real Job and Wall Street Is Hiring

[...] Yesterday, Wall Street analysts Wedbush added their endorsement (`http://www.businessinsider.com/winners-and-losers-in-bitcoin-2014-1`), noting that they "believe Bitcoin and its associated technology represent a potential disruption to our covered companies", and adding that "Bitcoin's potential lies beyond the 'coin' as the underlying blockchain protocol can be used to replace traditional intermediaries by acting as an exchange mechanism for a multitude of transactions".

Wall Street was also the location of Bitcoin's New Year's Eve party in New York the other day, when Steve Stockman, a Republican representative from Texas, promised to sponsor a pro-Bitcoin bill. [...]. BOND New York, a real estate brokerage firm just announced that it would start accepting Bitcoin as payment for real estate transactions. The litany of pro-Bitcoin news has helped push the price over $800 again on Mt. Gox. Bitcoin, it appears, is as resilient as ever.

The influx of more established financial players will inevitably help decrease Bitcoin's price volatility, which currently prevents it from being a more useful currency.

`http://www.lemonde.fr/idees/article/2014/01/06/pourquoi-les-economistes-devraient-etre-interesses-par-le-bitcoin_4343607_3232.html`

• 6 janvier 2014, Nicolas Houy (Économiste, chercheur CNRS) :

Pourquoi les économistes devraient être intéressé par le Bitcoin

[...] Les économistes ont vite fait de mettre le bitcoin dans une catégorie dont ils ont l'habitude de traiter et de le trouver au mieux inutile, le plus souvent néfaste, en tout cas, à éviter.

Mais le bitcoin n'est rien de tout cela. C'est un protocole. Comme le sont le HTTP (le protocole derrière l'Internet que vous utilisez tous les jours) ou le SMTP (un des protocoles derrière les e-mails). Ainsi, le bitcoin est un langage, un moyen de communication entre ordinateurs.

Grâce à ce moyen de communication, on peut échanger de l'argent entre deux points du globe sans coût. C'est déjà pas mal. Mais le bitcoin ne doit pas être résumé à cela, au risque de passer à côté du sujet. Comme tout protocole, chacun peut s'emparer de ce langage et en faire ce qu'il veut.

[...] le bitcoin ne doit pas être vu seulement comme une nouvelle version d'un objet économique déjà existant (vous entendrez certainement monnaie ou or 2.0). C'est un langage offrant des horizons infinis à quiconque décide de s'en emparer. Et les économistes, qui passent tant de temps à étudier les situations dans lesquelles les contraintes physiques empêchent la contractualisation, pourraient être bientôt les plus intéressés à en imaginer les usages futurs.

D'autres citations sélectionnées, « Pour » et « Contre », se trouvent en :

`http://www.lifl.fr/~delahaye/Bitcoin/Bitcoin.pdf`.

L'expérience des « 58 lycées »

Jacques Baudé [1]

Il est des moments, dans ce long fleuve pas forcément tranquille qu'est le déploiement de « l'informatique pédagogique » [2] au sein du système éducatif français, qui ont laissé des traces plus ou moins précises dans les mémoires. J'ai retracé à grands traits, dans le numéro 2 de 1024 [3], l'histoire de l'option informatique des lycées dans les années 80 et 90. Dans cet article, je remonte un peu plus loin dans le temps pour évoquer l'expérience des 58 lycées, moins connue car plus ancienne, mais qui a marqué durablement la suite des événements.

J'ai participé à cette expérience depuis son tout début au lycée de La Celle Saint-Cloud doté, dès la fin de la décennie 60, d'un ordinateur 10 010 de la CII [4]. Bien qu'ayant vécu cette expérience de près, je n'ai pas voulu me fier seulement à ma mémoire. Cet article s'appuie sur l'étude de l'INRP « Dix ans d'informatique dans l'enseignement secondaire (1970-1980) » [5]. Les historiens devront obligatoirement s'y reporter.

1. Président d'honneur de l'association Enseignement Public et Informatique (EPI).
2. Expression créée par l'EPI : http://www.epi.asso.fr/revue/48/b48p003.htm
3. http://www.societe-informatique-de-france.fr/wp-content/uploads/2014/02/1024-2-baude.pdf
4. *Introduction de l'informatique dans l'enseignement secondaire*, étude effectuée d'octobre 1971 à juin 1972 par MM. Pitié et Scherer : http://ensmp.net/pdf/1971/etude_pitie-scherer.pdf
5. *Dix ans d'informatique dans l'enseignement secondaire (1970-1980)*, collection Recherches pédagogiques, *Recherches pédagogiques* n° 113, INRP, 4e trimestre 1981, 182 pages. Archivé sur LARA, URI : http://hdl.handle.net/2332/1250.

Je me réfère également à de nombreux documents, tous en ligne sur le site de l'association Enseignement Public et Informatique, et notamment : « Pour une histoire de l'informatique dans l'enseignement français, premiers jalons » [6] *par Émilien Pélisset.*

Le séminaire de Sèvres

Même si un certain nombre d'expériences ont eu lieu dans les années 60 (voir note n° 6) il est couramment admis que l'introduction de l'informatique dans l'enseignement général français trouve son inspiration dans le séminaire du Centre d'études et de recherches pour l'innovation dans l'enseignement (CERI de l'OCDE, créé en 1968) au Centre national d'études pédagogiques de Sèvres en mars 1970 [7].

La lecture des conclusions et recommandations du séminaire est intéressante. On y retrouve les grands thèmes qui vont marquer les décennies à venir, certains étant encore en débat (discipline informatique ou pas), d'autres toujours pas mis en œuvre de façon durable 45 ans après (la formation des enseignants).

Ainsi, page 33, discipline ou pas : « *Envisagé comme enseignement ayant son propre statut, ou intégré dans une autre disciple d'enseignement, l'informatique...* ».

Page 34, l'enseignement de l'informatique nécessite des machines : « *Dans la mesure où le contenu de cet enseignement prévoit une initiation à la programmation conduisant à la rédaction de programmes par les élèves, il est nécessaire, au moins pour des raisons pédagogiques, que ces derniers voient le résultat de leur travail et prennent ainsi conscience de ce qu'un ordinateur est en mesure de faire...* ».

Page 35, à propos de la formation des enseignants : « *Il est évident qu'une action importante doit être entreprise dans le domaine de la formation des enseignants à l'informatique* » et, page 37 : « *Si l'on ne veut pas se contenter d'une formation au rabais, avec comme conséquence une dégradation presque immédiate de cet enseignement de l'informatique au niveau des élèves, ce seront des ressources importantes qui devront être consacrées à cet aspect du problème tant en argent qu'en hommes* ».

Tout a été dit, ou presque, au cours de ce séminaire international d'une semaine, du lundi 9 mars au samedi 14 mars 1970.

---

6. *Pour une histoire de l'informatique dans l'enseignement français, premiers jalons*, Émilien Pélisset : `http://www.epi.asso.fr/revue/histo/h85ep.htm`.

7. Actes du Séminaire pour « *l'enseignement de l'informatique à l'école secondaire* » organisé par l'OCDE (OCDE/Centre pour la recherche et l'innovation dans l'enseignement) avec la collaboration de la Direction de la coopération du MEN. Centre international d'études pédagogiques de Sèvres, 9 au 14 mars 1970. Les conclusions (pages 33 à 40) : `http://www.epi.asso.fr/revue/histo/h70ocde.htm`.

La politique volontariste du ministère de l'Éducation nationale

Rapidement, à la suite de ce séminaire, le MEN français va mettre en œuvre une politique ambitieuse comportant une formation des enseignants (avant même qu'il y ait des machines dans les établissements, c'est à noter !), une réflexion pédagogique collective dans le cadre de groupes disciplinaires au sein du Service de l'informatique éducative (SIE-INRDP), l'équipement (entre 1973 et 1976) de 58 lycées-collèges en mini-ordinateurs Mitra15 (CII) et T1600 (Télémécanique) [8].

Une conséquence importante de ce séminaire fut également la création, au ministère de l'Éducation nationale, d'une « mission informatique » dont le responsable avisé sera Wladimir Mercouroff.

L'appellation trompeuse « d'expérience des 58 lycées » fut donnée après que collèges et lycées aient été séparés (loi de 1975 sur les C.E.S.).

La formation des enseignants

Dès l'année scolaire 1970-1971, 80 enseignants (sélectionnés parmi 1024 candidats, ça ne s'invente pas !) sont répartis dans 3 centres de formation, 40 chez IBM, 20 à la CII et 20 chez Honeywell-Bull). Tout un ensemble de problèmes pédagogiques et organisationnels vont se poser. Les stagiaires ressentant la nécessité d'une structure spécifique, l'association Enseignement Public et Informatique (EPI) fut créée et officiellement déclarée le 1er février 1971.

Le positionnement des stagiaires est clair : c'est à eux, membres de l'enseignement public, de décider comment ils utiliseront l'informatique. Et apparemment, plusieurs voies sont possibles.

Ainsi, dans l'éditorial de premier bulletin de l'EPI (décembre 1971) [9], on peut lire :

« *...Notre attitude doit rester accueillante à l'égard des diverses expériences pédagogiques. L'introduction de l'informatique se présente actuellement, semble-t-il, sous trois aspects :*

> *— comme l'enseignement d'une matière nouvelle (sections et établissements spécialisés, cours pour volontaires, enseignement intensif de quelques jours pour tous les élèves d'un même niveau) ;*
>
> *— comme l'enseignement d'une méthode de pensée à l'intérieur des matières existantes, chaque professeur retrouvant dans sa discipline les notions fondamentales de modèle, d'algorithme, d'information ;*

8. Liste des 58 lycées : `http://www.epi.asso.fr/revue/histo/h70-58lycees.htm`

9. Éditorial du premier Bulletin de l'EPI (décembre 1970) : `http://www.epi.asso.fr/revue/01/b01p001.htm`

— *comme l'utilisation d'un moyen nouveau, comparable à ce que fut le premier livre imprimé, aidant le professeur dans la partie répétitive de son travail.*

Si la politique du ministère consiste à favoriser la deuxième solution, et malgré l'intérêt d'un tel choix qui facilite les rapports entre les différentes spécialités, nous pensons que nous devons nous intéresser à ces trois types d'expériences. L'avenir nous dira quelle est la voie la plus sûre. Mais peut-être s'agit-il de trois aspects complémentaires qui devraient coexister. En tout cas, nous ne pouvons ni ne devons choisir dès maintenant... »

Les années suivantes (de 1971 à 1976) les formations « lourdes » se déroulent dans des centres universitaires à Grenoble, Nancy, Toulouse, Paris, Rennes. Au total, 528 enseignants de différentes catégories et disciplines furent formés.

Avec des variations selon les centres de stage, on trouvait sensiblement les mêmes éléments :

— notions générales d'informatique (information, algorithmes, structure de l'ordinateur...),

— initiation à la programmation, qui avait tendance à dominer compte tenu de la rareté, voire de l'absence, de logiciels,

— les applications de l'informatique dans les différentes activités humaines et notamment une réflexion sur les applications pédagogiques possibles (je rappelle qu'à l'époque, il n'existait pratiquement pas en France de logiciels utilisables dans les disciplines générales, en dehors de ceux qui seront progressivement conçus et réalisés par les stagiaires. Dans ma discipline, deux des premiers logiciels de simulation LINKO et NUTRIT datent de 1974).

Une formation « légère » fut également mise en place par le Centre National de Télé-Enseignement (C.N.T.E.) de Vanves sous forme d'un cours par correspondance complété par quatre jours de stage sur machine. Elle concerna environ 5000 enseignants. Enfin, nombreux furent les autodidactes autour des machines.

Alors que la circulaire 70-232 du 21 mai 1970 (BOEN n° 22 du 28 mai) envisage un enseignement de l'informatique « *soit à l'occasion des cours traditionnels, soit dans des cours spéciaux pour volontaires* », le comité pédagogique animé par Wladimir Mercouroff, chargé de mission à l'informatique, donne la place prépondérante à l'outil pédagogique dans les différentes disciplines. C'est ce qui va orienter l'expérience des 58 lycées.

Les groupes disciplinaires coordonnés par l'INRDP

La Section Informatique et Enseignement (SIE), créée dès 1971 à l'INRDP, a regroupé un certain nombre d'enseignants formés dans des groupes disciplinaires (j'ai appartenu durant la décennie 70 au groupe « Informatique et Sciences naturelles »).

Ces enseignants recevaient des décharges de service pour leur donner le temps nécessaire à leur recherche et à la conception-réalisation de logiciels. Tout était à inventer et je peux témoigner, pour ce qui me concerne, que cette période pionnière qui comprenait à la fois l'enseignement de l'informatique, la recherche pédagogique, la conception de logiciels testés sur mes élèves de terminale, la formation des collègues... est la plus passionnante de toute mon existence professionnelle.

Le lycée de La Celle Saint-Cloud

Le lycée-collège de la Celle Saint-Cloud est un peu à part dans la mesure où une expérience d'enseignement optionnel de l'informatique, pilotée par l'INRDP, a été engagée à partir de l'année 1969-1970 grâce à un ordinateur 10 010 prêté par la CII (plusieurs parents travaillaient dans cette entreprise) (voir note n°4). Cette expérience se poursuivra avec le Mitra15 (reçu en février 1973) et le langage LES.

Il fut le premier à recevoir un Mitra 15. Cette attribution première (février 1973) n'a pas été qu'un avantage ! Nous avons essuyé les plâtres à la fois par les plantages du système LSE (vite réparés par Yves Noyelle de SupElec) mais surtout par les multiples pannes hard et une maintenance CII dépassée. Il fallait avoir la foi bien chevillée au corps pour survivre aux travaux pratiques (de Biologie, pour ce qui me concerne) interrompus en pleine séance ou reportés à la dernière minute. Nous avions appris à toujours prévoir une solution de remplacement pour nos élèves. Heureusement qu'à l'époque, l'ordinateur exerçait un tel attrait qu'il faisait oublier ses faiblesses. Comme disait Emilien Pélisset, ex-président de l'EPI, « on amenait les élèves à l'ordinateur comme la vache au taureau ».

Les matériels : Mitra 15 - T1600

Les équipements s'échelonneront de 1973 à 1976. Un des premiers établissements équipé fut celui de La Celle Saint-Cloud qui reçu en février 1973 le Mitra15 ayant fait un séjour à SupElec pour la mise au point du LSE. Il prenait la succession d'un 10 010 de la CII (voir encart).

Mitra15 (CII puis SEMS) et T1600 (Télémécanique) étaient des mini-ordinateurs plutôt imposants (2 mètres de haut, 200 kg) accompagnés de 9 terminaux et, au début, d'un lecteur de ruban perforé sur télétype. Il fallait plus d'une heure pour charger le

système LSE par cette voie ! Le lecteur de ruban fut remplacé par la suite par un disque souple (*floppy disk*).

Le disque dur, au départ de 256 Ko (eh oui !), se révéla très vite insuffisant pour stocker les logiciels et les travaux des élèves.

Le LSE [10]

Il résulte d'un contrat passé par la Délégation à l'informatique avec l'École Supérieure d'Électricité pour la définition d'un langage adapté aux besoins de l'enseignement secondaire.

Le rapport de définition de LSE fut remis au MEN (Mission à l'informatique) en octobre 1971. Il restait à réaliser le système de programmation et le système d'exploitation en temps partagé sur les deux machines retenues, avec comme date limite la rentrée scolaire 1972-1973.

Le système LSE (Langage Symbolique pour l'Enseignement) de syntaxe française fonctionnait en temps partagé, c'est-à-dire que chaque utilisateur, en alternance, disposait de la totalité du système. Chaque utilisateur pouvait dialoguer avec les différentes ressources du système par le télétype pilote et par sa console de visualisation. « *Le plus étonnant c'est que ça marche !* ». L'exclamation n'est pas de moi mais d'un des concepteurs du système qui venait régulièrement réparer les plantages.

L'EPI fit activement la promotion de ce système qui assurait une compatibilité et une portabilité bien supérieures à ce qu'aurait permis alors toute autre combinaison. Si le premier manuel LSE fut un numéro spécial du *Bulletin de liaison de l'INRP* (avril 1973), l'EPI produisit par la suite plusieurs co-éditions successives avec le CRDP.

Comme a tenu à le souligner Jacques Hebenstreit dans sa préface au « *Parler LSE* » de Michel Canal (Ed. Eyrolles) : « *Il est erroné de dire que LSE soit inspiré par BASIC. En 1970, BASIC (conçu en 1964) était un langage simple mais très rudimentaire, quasiment inexistant en France. En vérité, LSE a deux sources dont la première, injustement oubliée, est le langage PAF (interactif et à syntaxe française), conçu et réalisé par la SEA (Société d'Électronique et d'Automatisme) pour les ordinateurs CAB 500, et la seconde n'est autre que le langage ALGOL, né, lui aussi, vers les années 60.* »

Toujours est-il que ce langage de syntaxe française, gérant les chaînes de caractères dès sa création, a permis pendant des années les échanges au sein d'une communauté d'enseignants soucieux d'enseigner la programmation (clubs) et surtout de créer des logiciels pédagogiques pour les différentes disciplines. Ce ne fut pas le moindre de ses mérites. Comme disait un collègue, il aurait fallu qu'il nous revienne du Japon pour qu'il s'impose durablement en France.

10. « *La Saga du LSE et de sa famille (LSD/LSG/LST)* » par Yves Noyelle : `http://www.epi.asso.fr/revue/54/b54p216.htm`

Les logiciels et le LSE (ainsi que les sources) étaient diffusés gratuitement dans le système éducatif. L'esprit du « libre » s'enracine profondément dans la démarche de l'informatique pédagogique française...

Les logiciels

Dès les premières installations de mini-ordinateurs et du langage LSE, des enseignants commencèrent à concevoir et écrire des logiciels, soit isolément, soit dans des équipes locales, soit dans le cadre des groupes disciplinaires de l'INRP. Certains de ces logiciels n'eurent que des usages locaux, d'autres furent estampillés « INRP » et diffusés comme tels.

Parmi les 500 professeurs « formés lourds », beaucoup enseignent leur discipline d'origine dans des lycées équipés (ce fut mon cas). Ils disposent de décharges de service (toujours insuffisantes et trop tardives, mais néanmoins elles existaient !) pour accompagner leurs collègues, animer des clubs, concevoir des logiciels, les tester avec leurs élèves. Dans les années 80, beaucoup deviendront des formateurs, notamment pour l'opération Informatique Pour Tous (qui fera l'objet d'un prochain article), d'autres s'engageront dans la voie de l'option informatique (voir 1024 n° 2)...

Cette expérience fondatrice d'introduction de l'informatique dans l'enseignement secondaire a vu la conception, la réalisation (allant jusqu'à la programmation en LSE), l'échange gratuit de logiciels développés par les enseignants eux-mêmes en liaison étroite avec le SIE-INRP. Environ 800 logiciels ont été produits, toutes disciplines confondues, de 1972 à 1979. Plus de la moitié n'a été utilisée que localement, mais environ 400 ont été répertoriés et diffusés dans les 58 lycées (et au-delà par la suite) soit sous forme de rubans perforés soit sous forme de floppy plus tard. Chaque logiciel était accompagné d'une fiche pédagogique [11] comprenant le livret du maître, le livret de l'élève, le descriptif du programme et le listing en LSE.

Je n'ai pas la place ici de traiter complètement le sujet et renvoie à la troisième partie (pages 55 à 140) de l'étude de l'INRP (note n° 5) qui traite de la banque de logiciels, de l'analyse des logiciels et de leur utilisation.

Je dirai seulement que, contrairement à l'idée qu'on peut se faire ou qu'on a pu se faire (surtout à l'époque de l'enseignement « assisté » par ordinateur), les approches pédagogiques sont variées et souvent très novatrices. Certes, il y a les logiciels d'enseignement « tutoriel » et les exercices d'entraînement et de contrôle des connaissances, mais il y a aussi les logiciels de simulation, de traitement de données et les jeux pédagogiques.

11. Première page de la fiche accompagnant le logiciel de simulation expérimentale MENDEL : `http://www.epi.asso.fr/blocnote/Mendel.jpg`

Après adaptation aux micro-ordinateurs, puis au nano-réseau, une partie de ces logiciels se retrouvera dans les opérations « 10 000 micros » et « 100 000 micros », puis dans les « valises » du plan Informatique Pour Tous (IPT) (note n° 6) et même au-delà.

Les clubs

Dans leurs établissements, les enseignants formés sont très souvent à l'origine de « clubs informatiques », où l'on organise un enseignement selon des modalités variées où domine la programmation. Ces clubs concernent aussi bien les collégiens que les lycéens.

Ainsi, au concours de programmes [12] ouvert aux jeunes d'âge scolaire, organisé en 1981 par l'AFCET, avec la participation active de l'EPI, sur les 113 dossiers retenus 65 avaient comme point d'appui l'un des 58 lycées-collèges (16 dossiers provenaient de collégiens). La plupart de ces travaux n'auraient pas été menés à bien sans l'ambiance intellectuelle des clubs informatiques qui fonctionnaient depuis plusieurs années dans les établissements équipés.

On voit la complémentarité des approches se dessiner peu à peu, évidemment à échelle encore très réduite : l'informatique « outil » dans les différentes disciplines et l'informatique « objet » pratiquée dans les clubs informatiques.

L'accompagnement

Un bulletin de liaison INRP « *L'informatique dans l'enseignement secondaire* » a été diffusé dans les établissements équipés. Dans ces numéros, des comptes-rendus de réunions de travail et de séminaires, mais surtout, des comptes-rendus d'expériences et de pratiques dans les différentes disciplines.

Un numéro spécial (avril 1973) fut un manuel d'utilisation du LSE, complété et amélioré par un nouveau numéro (janvier 1975) rédigé par François Chédeville, professeur de lettres au lycée de La Celle Saint-Cloud, madame Faure de l'INRDP et Christian Lafond, responsable du SIE.

Le numéro spécial de décembre 1976 fait le point sur l'expérience des « 58 lycées ». On y retrouve « les deux frères ennemis », Jacques Arsac et Jacques Hebenstreit, l'un plaidant pour l'informatique « objet », l'autre pour l'informatique « outil ». Le débat ne date pas d'hier. Il sera — provisoirement — tranché en 1981 [13] !

À ce bulletin de liaison s'ajoutait le bulletin trimestriel de l'EPI.

---

12. À propos du concours AFCET « L'informatique à l'école » : `http://www.epi.asso.fr/revue/22/b22p059.htm`. Condition de participation au concours : « La réalisation d'un programme représentant un phénomène simple ou complexe relevant des disciplines enseignées dans le secondaire général et technique ».

13. *L'option informatique des lycées dans les années 80 et 90*. Première partie, « *La naissance d'une option* » : `http://www.epi.asso.fr/revue/histo/h10oi_jb1.htm`

La volonté des enseignants appartenant aux groupes disciplinaires a été de faire connaître les résultats de leurs recherches et notamment leur production logicielle à leurs collègues. C'est ainsi, par exemple, que le groupe « Informatique et sciences naturelles » a été la cheville ouvrière du thème « Biologie et informatique » du Congrès et journées nationales de l'APBG de juillet 1979 à Toulouse. Un dossier spécifique était paru, en deux parties, dès 1978 dans les numéros 3 et 4 du Bulletin de cette association, avec une double introduction de Wladimir Mercouroff et de moi-même [14].

Naturellement, des démarches analogues eurent lieu en direction d'autres associations de spécialistes. Il était important que les travaux de quelques un.e.s soient connus du plus grand nombre et, pour cela, la presse associative est un excellent vecteur.

L'évaluation

L'année 1975-1976 a marqué un tournant, l'arrêt de l'expérience (cf. ci-dessous), mais il n'avait pas échappé au Directeur des lycées que quelque chose d'important s'était passé dans les 58 lycées et collèges. Il commanda à l'INRP une nouvelle mission, celle d'évaluation (lettre du 9 août 1976 au responsable de la Section Informatique et Enseignement, Christian Lafond). Cette évaluation aboutit, en 1981, au rapport déjà cité plusieurs fois (note n° 5).

« *Cette étude est le résultat d'un travail collectif effectué par le groupe auquel l'INRP a confié le soin d'évaluer "l'expérience d'introduction de l'informatique dans l'enseignement secondaire". Au moment de la rédaction, ce groupe était constitué des personnes suivantes : Georges Baron, Michèle Bounay, Philippe Dautrey, Josée Guelfucci, Danie Hebert, Pierre Muller, Monique Schwob, Philippe Tourtelier. Il a bénéficié des contributions extérieures ou des conseils de Mmes Bonin, Dassonville et Torcq et de M. J.Baudé (Secrétaire général de L'EPI, ex-membre du groupe Informatique et Sciences naturelles).* » (page 3)

Il s'agit d'un essai de bilan global de nature à servir aux opérations à venir. La première partie de ce rapport dresse le cadre matériel et administratif dans lequel s'est déroulée l'expérience et fournit un certain nombre de données quantitatives sur le fonctionnement des centres informatiques des lycées. La sensibilisation à l'informatique, premier objectif que s'était fixé le Ministère, est l'objet de la deuxième partie où les auteurs examinent les contenus qu'elle a essayé de transmettre. Dans la troisième partie, il est traité longuement des logiciels pédagogiques, envisagés successivement sous l'angle de leur contenu et de leur utilisation. La quatrième partie, enfin, tente de dégager les différents apports de l'informatique aux élèves et aux enseignants. On trouve dans la conclusion à la fois un constat des acquis, les limites

14. Bulletin pédagogique de l'APBG, n° 3, 1978 ; 2me partie dans le Bulletin n° 4, 1978, http://www.epi.asso.fr/tic/78apbg.htm

d'une expérience et un ensemble de propositions pour la poursuite d'une généralisation déjà amorcée.

Parmi les acquis de l'expérience :

— Il faut dire tout d'abord que les résultats obtenus — pour une expérience qui ne concernait que 58 établissements — sont loin d'être négligeables. Outre l'expérience acquise qui se révèlera importante pour les années à venir, plus de 45 000 élèves et plus de 1 000 enseignants ont chaque année utilisé l'ordinateur (à une époque où il s'agissait d'un objet rare !). Toutes utilisations confondues, les centres informatiques ont fonctionné en moyenne 32 heures par semaine.

— La formation des enseignants ayant été posée comme préalable, l'utilisation de l'ordinateur a permis de nouvelles pratiques pédagogiques (simulation, exploitation de banques de données...) suscitant des motivations nouvelles de la part des élèves. Elle a permis la constitution d'équipes compétentes capables d'innover et notamment de participer à la conception de logiciels disciplinaires en liaison avec la recherche pédagogique. La prise de conscience de la nécessité d'une recherche pédagogique de qualité à l'écoute du « terrain » s'est imposée.

— L'introduction de l'informatique a pu s'effectuer à travers toutes les disciplines. On peut parler ici de pluridisciplinarité effective et même d'interdisciplinarité pour un certain nombre de démarches. Dans nombre d'établissements, la cohabitation entre l'EAO [15] et les clubs informatiques fut bénéfique pour les élèves et les enseignants.

— Les décharges de service pour les enseignants impliqués dans la recherche et la conception de logiciels (répondant aux besoins réels, et non supposés, des enseignants), la disponibilité de logiciels gratuits (diffusés avec le code source) furent également des facteurs déterminants.

Mais l'expérience a connu *quelques difficultés*, certaines conjoncturelles et en partie résolues (matériels, par exemple), d'autres étant toujours d'actualité :

— Les insuffisances du matériel (pannes, lenteur des entrées-sorties, incompatibilité des disques souples, engorgement de la salle informatique...). Une certitude s'impose, les matériels doivent être en nombre suffisant et leur maintenance assurée.

— Les insuffisances administratives (planning de la salle informatique, horaires des élèves, classes devant être dédoublées, décharges de service insuffisantes ou parvenant très en retard dans les établissements, manque de moyens au niveau du pilotage national...).

Difficultés insuffisantes pour décourager des « pionniers » mais de nature à compromettre une généralisation. Ce qui se vérifiera d'ailleurs pas la suite.

15. Enseignement Assisté par Ordinateur.

Le coup d'arrêt de l'année 1975-1976

En septembre 1975, le Ministre de l'Éducation nationale supprime brutalement la Mission à l'informatique. Les difficultés économiques entraînent un plan d'austérité général et un budget de rigueur pour l'Éducation nationale en 1976. La Direction générale de la programmation et de la coordination (D.G.P.C.), qui a confié la gestion de l'expérience à la Direction des Lycées, supprime les formations et décide l'arrêt des équipements (janvier-février 1976).

On voit apparaître les premières « dents de scie » d'une politique de déploiement de l'informatique dans le système éducatif et on commence à se rendre compte que ce déploiement sera plus chaotique que linéaire. Les changements de ministres de l'Éducation nationale (l'EPI en a connu vingt depuis sa fondation) ne rajoutant rien.

Toutefois, le volume des décharges de service et une partie des postes attribués antérieurement sont conservés ; quelques améliorations techniques facilitent l'usage des matériels en place (notamment lecteurs/enregistreurs de disquettes 8 pouces). Une nouvelle série de cours du C.N.T.E., assortie d'un stage de trois jours, est diffusée entre 1976 et 1979 mais elle est payante. Elle touchera moins d'un millier de collègues. Il n'y aura aucune formation approfondie de 1976 à 1981.

Les suites de l'expérience : le plan « 10 000 micro-ordinateurs »...

« *L'informatisation de la société* », rapport de Simon Nora et Alain Minc, est remis en janvier 1978 au Président de la République qui, en novembre, demande au gouvernement d'élaborer un nouveau plan informatique pour accroître l'efficacité, la compétitivité du système économique.

Une décision de décembre 1978 relance un projet d'ampleur et de nature sensiblement différents : il s'agit de l'opération « 10 000 micro-ordinateurs ». En effet, les micro-ordinateurs sont apparus vers 1975, faisant prévoir un rapport nouveau du public à l'informatique. C'est aussi l'occasion d'un encouragement des pouvoirs publics aux entreprises françaises qui semblent alors capables en ce domaine de tenir une place dans la concurrence mondiale. Mais le projet concerne encore essentiellement les lycées et les collèges. Pourtant, en 1980, commence à se développer un plan d'équipement des Écoles normales et de formation des personnels, en même temps que se réunit régulièrement une « cellule de réflexion » sur l'introduction de l'informatique dans les écoles, à l'initiative de la Direction des Écoles. Mais toujours pas de reprise des formations « lourdes » et, cette fois, les matériels précèdent les formations quand il y en a.

Mais ceci est une autre histoire et nous sommes déjà sortis de l'expérience des « 58 lycées ».

JACQUES BAUDÉ, juin 2014, jacquesbaude@free.fr

Je découvre l'informatique et la programmation avec Martin

Entretien avec Martin Quinson [1]

Dans le monde numérique qui est le nôtre, l'informatique est un incontournable élément de culture générale qui n'est pas encore partagé par tous. Il s'agit donc de promouvoir les actions d'information, d'initiation et d'éducation à cette discipline.

Nous nous entretenons ici avec Martin Quinson, enseignant-chercheur en informatique qui œuvre dans toutes ces dimensions de médiation scientifique et enseignement de l'informatique.

Philippe Marquet [2]

Martin, tu es enseignant-chercheur en informatique et enseignes cette discipline à des élèves ingénieurs de TELECOM Nancy. Tu cherches aussi à partager tes connaissances avec d'autres publics.
Chacun devrait apprendre à programmer ? Nous devrions tous devenir programmeur ?

Martin Quinson : Non, bien sûr. Il me semble clair que nous aurons demain encore besoin de plus de boulangers, de plombiers et de chirurgiens-dentistes que de programmeurs capables de réécrire un module noyau avant le petit-déjeuner. Mais étant donnée la percolation de l'informatique dans toute notre société, il me semble tout aussi clair que ceux qui n'auront même pas des bases de programmation auront un sérieux désavantage par rapport aux autres.

1. Université de Lorraine, `http://www.loria.fr/~quinson/`.
2. Université Lille 1. Vice-président de la Société informatique de France.

Si on n'a jamais essayé de programmer, il est difficile de ne pas développer de pensée magique sur ce qu'un ordinateur peut ou ne peut pas faire. On ne peut plus se permettre d'avoir une majorité de citoyens ignares et des décideurs mal informés en informatique après Snowden. Ils risquent soit d'avoir des comportements irresponsables, soit de développer des sentiments de rejet vis-à-vis du numérique pourtant indispensable dans nos sociétés. Il est donc primordial pour la nation que les individus dépassent les usages numériques et aient des bases de programmation.

À titre personnel, si on sait programmer un peu, on peut automatiser des tâches de la vie de tous les jours. Mon voisin qui est ingénieur à EDF sait écrire des macros Excel et des tout petits programmes en Visual Basic. Cela lui simplifie souvent la vie dans son travail bien qu'il ne se sente pas du tout « programmeur ». Pour moi, écrire des règles de filtrage des mails, c'est aussi programmer pour se simplifier la vie professionnelle même quand on n'est pas programmeur. L'une des clés du succès de Minecraft [3] (et Quake avant lui) me semble tenir à la simplicité d'écrire des extensions. De nombreux individus créatifs sans être « programmeurs » se sont emparés de ce support. Je crois que nous sommes ici face à un problème de la poule et de l'œuf : il n'y a finalement pas tellement de situations que l'on peut adapter facilement en programmant, mais c'est aussi parce que trop peu de monde sait programmer et que le besoin n'est pas là. Les gens utilisent assez peu des choses comme AppleScript car ils n'ont aucune idée de ce que programmer permet. Mais chaque fois que la taille de la communauté atteint un seuil critique, on voit des choses incroyables émerger. J'ai hâte de voir un tel effet boule de neige sur toute la société civile quand suffisamment de gens sauront programmer et réclameront la possibilité d'adapter leur environnement numérique à leurs besoins personnels !

Ce qui est sûr, c'est que le monde où 1% de la population est parfaitement versé dans l'art de la programmation et 99% parfaitement ignare en matière de fonctionnement du numérique est derrière nous. La démarcation ne sera pas si claire à l'avenir. J'irais même plus loin en affirmant qu'apprendre à programmer est utile même si par extraordinaire, on fait ensuite un métier sans aucun ordinateur. Programmer enseigne à rationaliser sa pensée afin de la rendre compréhensible par l'ordinateur. Un peu de rationalité, c'est très utile dans la vie. La magie des ordinateurs, c'est que programmer permet de marier très naturellement cette rationalité avec la créativité, puisque les nouveaux médias permettent de s'exprimer au travers de créations multimédia originales, que l'on peut réaliser chez soi, très facilement.

Cette envie de partager tes connaissances de la programmation, de l'informatique, avec tous les publics, comment cela a-t-il commencé ?

MT : Comme beaucoup d'enseignants-chercheurs, j'ai été recruté sur mon pedigree de recherche, l'enseignement n'étant évalué que quantitativement. J'ai eu la chance

3. Minecraft est un jeu vidéo dont le joueur, qui évolue dans un monde qu'il peut modifier à volonté, est libre de définir l'objectif.

d'avoir des cours de niveau BAC+5 à mon arrivée. Mais j'ai très vite réalisé que je prenais plus de plaisir à enseigner les bases de la programmation aux débutants qu'à enseigner des notions plus avancées à des étudiants parfois mal assurés en programmation. Depuis plusieurs années maintenant, je réalise la quasi totalité de mon service en première année d'ingénieurs. Il faut faire preuve d'une certaine imagination pédagogique pour enseigner rapidement la programmation à des débutants absolus. Au fil des ans, nous avons réalisé avec Gérald Oster une plate-forme pédagogique complète pour donner les bases de l'algorithmique le plus efficacement possible aux élèves entrant à l'école. La PLM (Programmer's Learning Machine) est maintenant devenue un projet de recherche soutenu par Inria, qui finance un ingénieur pour deux ans afin d'en faire une plate-forme expérimentale pour l'enseignement de l'informatique. Nous souhaitons par exemple mesurer l'efficacité de différentes approches et exercices, détecter les barrières didactiques de notre discipline ou encore apporter une assistance automatisée aux apprenants qui font des erreurs classiques.

Pour diffuser la connaissance de l'informatique à un public encore plus large, j'ai participé à la fête de la science pour la première fois en 2011 avec Jean-Christophe Bach. Nous avons réalisé pour l'occasion une petite série d'activités débranchées pour parler d'algorithmique sans ordinateur. Ce fut un tel succès que nous avons repris et augmenté ces activités (avec l'aide d'Inria, aussi), qui ont été jouées des dizaines de fois depuis, par nous et par d'autres. Je compte améliorer et étendre ce projet (qui s'appelle Sciences Manuelles du Numérique), mais le temps me manque pour l'instant.

Tu es également soucieux de diffuser largement ce matériel que tu produis ?

Oui, cette large diffusion de la connaissance me tient à cœur, et c'est pour cela que je participe au grand mouvement du logiciel libre depuis 15 ans. En parallèle de ma thèse, j'ai animé le groupe francophone de traduction de la distribution Debian. Nous avons œuvré pour que les logiciels libres deviennent vraiment accessibles à tous, même ceux qui ne parlent pas anglais. Encore maintenant, je suis un fervent défenseur des logiciels libres au sens large. En particulier, tous les projets dont je parle ici sont placés sous licences libres (GPL ou CC-BY-SA). Mon objectif est toujours de donner en bien commun afin que d'autres puissent bénéficier et améliorer ce que je fais mais sans pouvoir empêcher les suivants de profiter eux aussi de ces projets. Je donne dans l'objectif de bénéficier de l'aide de la communauté, et cela fonctionne plutôt bien : la PLM est par exemple traduite en brésilien et italien alors que je ne parle pas ces langues, et quelques leçons proposées ont été écrites par des volontaires que je ne connais pas forcément.

Pour en revenir à l'initiation à l'informatique, il reste quand même quelques étapes à franchir à un enfant ou au grand public qui a découvert l'informatique par ces activités débranchées pour aborder la programmation sur ta plateforme PLM ?

MT : En effet, il y a un an ou deux, j'étais impliqué dans deux projets d'initiation à l'informatique : les SMN, Sciences Manuelles du Numérique, pour faire découvrir l'algorithmique à la fête de la science, et la PLM pour apprendre à programmer à de futurs spécialistes. C'était très intéressant (même si un peu chronophage), mais j'avais du mal à boucler la boucle. Il est difficile de se lancer dans la PLM après seulement deux heures de SMN, et c'est en tout cas impossible pour des enfants (les miens ont entre 8, 9 et 11 ans). En furetant un peu sur le web, j'ai découvert les Coding Goûters [4]. Cette superbe initiative invite les enfants et leurs parents à s'initier à la programmation créative, c'est-à-dire à s'approprier juste ce qu'il faut de programmation pour pouvoir s'exprimer librement dans ses propres créations sur ordinateur.

J'aime beaucoup cette idée car cela réconcilie informatique et créativité, sans le côté austère d'un cours où un « professeur » donne des exercices à la chaîne. J'ai aussi découvert par la suite qu'inviter les parents à écouter les activités de leurs enfants était une technique classique quand on cherche à réduire en douceur le niveau d'illettrisme d'une population. C'est une bonne idée d'adapter cela à l'illettrisme numérique ;). Nous avons donc organisé quelques coding goûters dans l'école d'ingénieurs où j'enseigne (TELECOM Nancy). L'accueil du public a été enthousiaste : nous avions imprimé plusieurs centaines de flyers à destination des écoles voisines afin de remplir les 35 places du premier après-midi. Toutes sont parties en deux heures après un simple mail aux employés de l'université de Lorraine. Maintenant, ce sont nos élèves qui nous pressent d'organiser d'autres coding goûters dans l'école. Nous avons décidément beaucoup de chance.

Ces coding goûters permettent une découverte de la programmation. Comment prolonger cette première initiation ?

Pour aller plus loin qu'une simple après-midi de découverte, nous avons noué un partenariat avec la MJC centre social Nomade de Vandœuvre-lès-Nancy afin d'organiser une activité de programmation créative hebdomadaire pour les 8-12 ans. Là encore, Inria soutient le projet (nommé CodCodCoding) en finançant un doctorant qui fera son monitorat à ce public un peu particulier, assisté d'un animateur de la MJC qui apporte sa connaissance de notre public. Notre objectif cette première année est d'explorer pour découvrir ce que ce genre d'activité et de public implique. Nos découvertes sont consignées dans un blog tenu par le doctorant. À terme, nous souhaitons essaimer le plus largement possible, au sein du réseau Inria et au-delà. Nous avons organisé des visio-conférences de « formation » pour inciter les collègues de partout en France à se lancer. J'ai aussi participé à un séminaire de formation « La main à la pâte » à destination de formateurs d'enseignants, et un autre est prévu en mai. Je suis prêt à faire des cours de rattrapage pour ceux qui ont raté les premières séances : Il suffit de me contacter si vous voulez vous lancer.

4. http://codinggouter.org/

C'est qu'il naît environ 800 000 enfants par an chaque année. À raison de 35 enfants par samedi après-midi, il faudrait donc plusieurs dizaines de milliers de semaines par an pour que chacun puisse bénéficier d'une simple initiation de type coding goûter. En attendant que la République ne prenne ses responsabilités et forme les enfants comme il se doit, il est urgent de faire essaimer les initiatives et de diffuser les bonnes pratiques entre ces initiatives locales. C'est pour cela que nous avons fondé le collectif `jecode.org` l'an passé avec Bastien Guerry et David Roche. Nous étions inspirés à la fois du site `code.org` qui a beaucoup fait outre-Atlantique pour réveiller les consciences sur l'urgence de l'enseignement de l'informatique, mais aussi du site `codeclub.org.uk` qui a énormément fait pour organiser le milieu associatif britannique. Code.org a fait une série de vidéos mettant en scène des stars qui expliquent l'importance du « code » dans le monde d'aujourd'hui (c'est de là que vient tout le buzz actuel autour de ce mot). CodeClub fournit des ressources clé en main, et a accompagné la création de plusieurs milliers de club d'informatique à travers le Royaume-Uni.

Plus modestement, nous avons attiré de nombreux associatifs au sein du collectif `jecode.org`, où nous discutons de nos techniques et retours d'expériences sur la programmation créative pour les enfants. Si le sujet vous intéresse, n'hésitez pas à nous rejoindre. Nous n'avons pas encore les moyens de réaliser une vidéo pour insister sur l'importance de la programmation, mais cela viendra peut-être `:)`. En attendant, nous avons récolté les témoignages personnels de programmeurs afin de montrer la diversité des motivations à savoir programmer, même un peu. Nous recensons également les initiatives locales pour permettre aux différents acteurs de se retrouver sur le territoire.

Je rêve d'un scénario à l'anglaise, où le premier ministre a dit un beau jour que tous les enfants devaient être correctement formés à l'informatique au travers du « code » et de la programmation. Les deux ans qui ont suivi, c'est le milieu associatif qui a assuré cette initiation, et ensuite les enseignants ont pris le relai une fois formés. Cette prise en charge par les enseignants de l'État me semble indispensable pour ne pas aggraver les inégalités territoriales. Aucune association ne peut prendre en charge 800 000 enfants par classe d'âge, c'est-à-dire plusieurs millions d'enfants en continu. Cela demande des dizaines de milliers d'intervenants correctement formés, et des moyens dont seul l'État dispose.

La bonne nouvelle, c'est que ce scénario à l'anglaise n'est pas impossible chez nous, où un ministre de passage a dit que le code devait faire son entrée à l'école primaire sur le temps périscolaire. Bon, il est reparti avant même la rentrée, mais les choses progressent, je garde espoir. Si le grand plan numérique annoncé ne sert pas qu'à financer des tablettes qui feraient des gamins de parfaits petits consommateurs, cette fois est peut-être la bonne.

En savoir plus...

– www.loria.fr/~quinson/JLM/ PLM – Programmer's Learning Machine.
– www.loria.fr/~quinson/SMN/ SMN – Sciences manuelles du numérique.
– iww.inria.fr/codcodcoding/ Blog de CodCodCoding, l'activité de programmation créative de la MJC centre social Nomade à Vandœuvre-lès-Nancy.
– jecode.org rassemble les initiatives encourageant l'apprentissage de la programmation pour les enfants.

ENSEIGNEMENT

L'édition scolaire au temps de l'informatique

Jean-Pierre Archambault [1]

L'informatique a plongé le monde de l'édition scolaire, et pas que lui, dans une période de turbulences fortes. La banalisation des logiciels de réalisation de contenus (traitement de texte, logicicels de publication...) et d'Internet, qui permet un travail collectif à grande échelle, a favorisé les productions autonomes des enseignants et leur diffusion. On constate une transférabilité des méthodes du logiciel libre et de ses réponses en termes de droit d'auteur à d'autres ressources immatérielles (les licences Creative Commons *sont issues de la GNU-GPL), pédagogiques en particulier. Dans* Le Monde Diplomatique*, en décembre 2002, quand il évoque les risques de privatisation du génome humain, John Sulston, prix Nobel de médecine, dit que «* les données de base doivent être accessibles à tous, pour que chacun puisse les interpréter, les modifier et les transmettre, à l'instar du modèle de l'open source pour les logiciels *».*

L'économie de l'information, longtemps limitée à celle de ses moyens de diffusion, ne peut plus se confondre avec l'économie du support puisque que les biens informationnels ne sont plus liés rigidement à un support donné. Les coûts marginaux pour produire et diffuser un exemplaire supplémentaire sont négligeables... Les choses changent au temps de l'informatique.

Des manuels scolaires libres

Des manuels scolaires libres, papier et/ou numériques : il en existe depuis quelques années déjà. Les plus connus et emblématiques sont ceux de l'association

1. Président de l'EPI, `http://www.epi.asso.fr/`.

Sésamath [1]. Dans le champ émergeant de la production pédagogique numérique libre, l'association Sésamath, pionnière en la matière dès le début des années 2000, est synonyme d'excellence. Elle compte une centaine de professeurs de mathématiques, de collège principalement. Plusieurs centaines de contributeurs proposent régulièrement des améliorations des logiciels développés et des documents pédagogiques les accompagnant, mis gratuitement et librement en ligne (licences de type *Creative Commons*). Sésamath ne soutient que des projets collaboratifs dont elle favorise et encourage les synergies. Elle instaure des partenariats avec des éditeurs privés et publics (CRDP). Avec l'éditeur Génération 5, elle représente de l'ordre de 25% du manuel de mathématiques en collège. Le modèle économique est simple : d'un côté donc des ressources accessibles gratuitement et librement en ligne, de l'autre l'édition de livrets, de manuels, de produits pour ordinateur et tablette à partir de ces ressources mises sur le web. La rémunération se fait donc avec des produits matériels dérivés. Classique et compatible avec le libre.

On citera également, toujours sous licences de type *Creative Commons*, lelivrescolaire.fr [2]. Et chez Eyrolles, le manuel Informatique et sciences du numérique Spécialité ISN en Terminale S [3]. Au CRDP de Paris, *Introduction à la science informatique pour les enseignants de la discipline en lycée* [4]. Ces manuels ont rencontré leur public.

Dans ce texte, nous parlons des manuels scolaires libres, du manuel scolaire papier, ne serait-ce que parce qu'il a encore un avenir, des ressources pédagogiques éditorialisées, numérisées ou non, des rapports entre droit d'auteur et pédagogie, et du présent et de l'avenir de l'édition scolaire [5].

Le manuel scolaire, les ressources éditorialisées et les autres ressources pédagogiques

Les ressources pédagogiques utilisées dans les cours ont toujours joué un rôle important dans l'enseignement, correspondant à des démarches et des conceptions d'apprentissage. On peut les classer en trois catégories. Il y a les ressources produites par les enseignants eux-mêmes, aujourd'hui dans le contexte radicalement nouveau issu de l'omniprésence de l'ordinateur et d'Internet, à des échelles dont les ordres de grandeur sont sensiblement différents de ceux de l'ère du pré-numérique. Il y a aussi les ressources que se procurent les enseignants. Éditées à des fins explicitement pédagogiques ou matériaux bruts non conçus initialement pour des usages scolaires, mais y trouvant naturellement leur place, comme une œuvre musicale ou une reproduction de tableau. Leurs modèles économiques et leurs modalités de droit d'auteur sont distincts. Le manuel scolaire est une ressource éditorialisée parmi d'autres, relevant d'un secteur économique avec une activité dont il doit vivre. Les ressources brutes, elles, appellent une « exception pédagogique ».

L'exception pédagogique

L'exception pédagogique, c'est-à-dire l'exonération des droits d'auteurs sur les œuvres utilisées dans le cadre des activités d'enseignement et de recherche, et des bibliothèques, concerne potentiellement des productions qui n'ont pas été réalisées à des fins éducatives. Les enseignants en utilisent, dans toutes les disciplines, mais particulièrement dans certaines d'entre d'elles comme l'histoire-géographie, les sciences économiques et sociales ou la musique : récitation d'un poème, lecture à haute voix d'un ouvrage, consultation d'un site Web... Ces utilisations en classe ne sont pas assimilables à l'usage privé. Elles sont soumises au monopole de l'auteur dans le cadre du principe de respect absolu de la propriété intellectuelle. Cela peut devenir mission impossible, tellement la contrainte et la complexité des droits se font fortes. Ainsi pour les photographies : droits du photographe, de l'agence, droit à l'image des personnes qui apparaissent sur la photo ou droit des propriétaires dont on aperçoit les bâtiments... Difficile d'imaginer les enseignants n'exerçant leur métier qu'avec le concours de leur avocat !

L'arrivée de l'informatique a fait entrer le monde de l'édition scolaire, pas que lui, dans une période de turbulences. Nous y reviendrons. Des illustrations des tensions provoquées avec de « surprenantes » entraves à l'activité pédagogique figurant dans des accords passés en 2006 entre le Ministère de l'Éducation nationale et les syndicats d'éditeurs. Yves Hulot, professeur d'éducation musicale à l'IUFM de Versailles (Cergy-Pontoise), en pointait quelques-unes [6]. La combinaison d'une résolution limitée à 400×400 pixels et d'une définition de 72 dpi de la représentation numérique d'une œuvre avec une utilisation portant sur l'œuvre intégrale empêche le recours à des photos de détails des œuvres. Par exemple : « *Le tableau* Les noces de Cana *de Véronèse comporte une intéressante* viola da braccio *au premier plan. Avec une telle limite de résolution sur cet immense tableau, il est impossible de projeter et de zoomer convenablement sur ce détail. Qui peut croire justifiées de telles limitations ?* » La durée des extraits d'une œuvre musicale crée également des obstacles de nature pédagogique. « *Si je travaille en classe de troisième sur la compagnie des Ballets russes, il me semble évident que pour apprécier l'importance de la révolution qu'elle apporta, une analyse d'au moins deux ballets s'impose, en l'occurrence* L'après-midi d'un faune *et* Le sacre du Printemps*, ce qui permet d'aborder deux musiciens capitaux, Debussy et Stravinsky. Le premier dure environ dix minutes, le second environ trente. Habitués que sont nos élèves à côtoyer majoritairement le genre chanson, il me semble utile de les confronter à d'autres durées et d'autres langages musicaux ou chorégraphiques. Mais je crains qu'à trop limiter l'activité pédagogique des enseignants on finisse par empêcher ceux qu'ils éduquent de réellement avoir les clés d'accès à la culture !* »

Les licences *Creative Commons* contribuent, en tout cas sont un puissant levier, à développer un domaine public élargi de la connaissance. Et la GNU GPL et le CeCILL qui permettent aux élèves et aux enseignants de retrouver, dans la légalité, leurs environnements de travail sans frais supplémentaires, ce qui est un facteur d'égalité et de démocratisation. Mais la question de l'exception pédagogique dans sa globalité, une vraie question, reste posée avec une acuité accrue de par le numérique : quelque part, le *copyright* est antinomique avec la logique et la puissance d'Internet.

L'activité d'enseignement est désintéressée et toute la société en bénéficie. L'éducation n'est pas un coût mais le plus nécessaire (et le plus noble) des investissements. L'exception pédagogique a donc une forte légitimité sociétale.

Des résistances

Si les modèles de Sésamath et des autres ouvrages libres éditorialisés fonctionnent bien, ils n'en sont pour autant pas devenus la norme. Depuis la signature de l'accord-cadre par le MEN et l'AFUL [2] en octobre 1998, le logiciel libre a fait sa place dans le système éducatif, devenant au fil des ans une composante à part entière de l'informatique scolaire. Ce ne fut cependant pas un long fleuve tranquille. Mais le cheminement est plus lent pour les ressources pédagogiques éditorialisées, en premier lieu les manuels scolaires. Il faut dire que les enjeux économiques et financiers sont très forts avec le public (captif) des 12 millions d'élèves. Il y a des résistances, il y a toujours des résistances au changement [7].

Pourtant, les choses doivent nécessairement évoluer. Le monde de la production matérielle et celui de la production immatérielle, leurs réponses en termes de propriété intellectuelle (brevets, droit d'auteur...) ne sauraient se confondre. D'où des turbulences qui ne concernent pas que le manuel scolaire.

L'économie de l'information et des biens immatériels

L'économie de l'information s'est longtemps limitée à une économie de ses moyens de diffusion, c'est-à-dire à une économie des médias. Elle ne peut désormais plus se confondre avec l'économie du support, puisque les biens informationnels ne sont plus liés rigidement à un support donné. L'essentiel des dépenses était constitué par les coûts de production, de reproduction matérielle et de distribution dans les divers circuits de vente. Aujourd'hui, les techniques de traitement de l'information, la numérisation et la mise en réseau des ordinateurs réduisent les coûts de duplication et de diffusion jusqu'à les rendre à peu près nuls. Dans ces conditions, la valeur économique de l'information ne peut plus se construire à partir de l'économie des supports physiques servant à la distribution, mais à partir de services permettant son appropriation, sauf s'il y a des produits dérivés, comme un manuel papier. L'éditeur

2. Association Francophone des Utilisateurs de Logiciels Libres, `https://aful.org/`.

est placé dans une situation, pour les biens numérisés, où les coûts fixes de production, dépensés avant la première commercialisation, sont importants, et les coûts marginaux pour produire et diffuser un exemplaire supplémentaire négligeables.

Il faut dissocier le texte du support. Un livre classique est composé de deux choses : un texte et un support. Le texte est une suite de symboles et d'illustrations qui peuvent exister sur différents supports. Le support papier est un ensemble de feuilles reliées et couvertes d'encre. De l'invention de l'écriture à celle de l'ordinateur, le texte et son support étaient indissociables, ce qui nous a donné la mauvaise habitude d'utiliser le mot « livre » pour désigner l'un et l'autre. Mais texte et support sont des objets effectivement très différents du point de vue économique. Le support est un bien rival : chaque exemplaire supplémentaire coûte. Le texte, en revanche, est non rival. Une fois écrit, il peut être dupliqué à l'infini à un coût quasi nul ; il n'y a donc aucun obstacle à laisser tout le monde en profiter gratuitement. La vente à l'exemplaire de biens non rivaux est absurde : les consommateurs n'ont aucun intérêt à payer chaque exemplaire, les créateurs n'ont aucun intérêt à priver ceux qui n'ont pas les moyens ou l'envie de payer leurs créations. C'est pour cela que les créateurs de biens non rivaux réfléchissent à d'autres moyens de financer leur création : se faire payer une fois et laisser le fruit du travail en libre accès à tous. Si la licence globale pour la musique — payer une redevance de 100 € par an pour pouvoir accéder à toute la musique au lieu d'acheter 20 € chaque CD — peine à se mettre en place, essentiellement parce qu'elle contrarie les intérêts d'un tout petit nombre de musiciens très populaires et de leurs maisons de disques, il nous semble que, dans le cas des manuels scolaires, déjà payés par l'impôt, ne pas mettre en place une telle licence globale et continuer à payer le texte à l'exemplaire (le texte et non le support papier répétons-le) est une mauvaise utilisation de l'argent public. La question du modèle économique de l'exemplaire papier est posée [8].

Et, d'une manière générale, de nouvelles sources de valeur sont en train d'apparaître. Le modèle économique de mise en valeur de l'information déplace son centre de gravité des vecteurs physiques vers des services annexes ou joints dont elle induit la consommation ou qui permettent sa consommation dans de bonnes conditions (services d'adaptation d'un logiciel au contexte d'une entreprise, de facilitation de l'accès à des ressources numérisées, commerce induit sur des produits dérivés, etc.). On retrouve la logique du libre, à savoir la rémunération par le service [9].

La gratuité ne doit pas brouiller le débat

En effet, elle n'est pas le problème. Les produits du travail humain ont un coût, le problème étant de savoir qui paye, quoi et comment. La production d'un logiciel, qu'il soit propriétaire ou libre, nécessite une activité humaine. Elle peut s'inscrire dans un cadre de loisir personnel ou associatif, écrire un programme étant un hobby comme il en existe tant. Elle n'appelle alors pas une rémunération, la motivation

des hackers (développeurs de logiciels dans des communautés) pouvant résider dans la quête d'une reconnaissance par les pairs. En revanche, si la réalisation se place dans un contexte professionnel, elle est un travail qui, toute peine méritant salaire, signifie nécessairement rémunération. Le logiciel ainsi produit ne saurait être gratuit, car il lui correspond des coûts. Et l'on sait, répétons-le, que dans l'économie des biens immatériels, contrairement à l'économie des biens matériels, les coûts fixes sont importants tandis que les coûts marginaux (coûts de production et diffusion d'un exemplaire supplémentaire) sont peu élevés. Dupliquer un cédérom de plus ou fabriquer un deuxième avion, ce n'est pas la même chose. Télécharger un fichier ne coûte rien ou presque (sauf l'accès au réseau).

Mais, même quand un logiciel n'est pas gratuit, il le devient lorsqu'il a été payé ! Enfin, quand il est sous licence libre. Qu'est-ce à dire ? Prenons le cas d'un client qui commande la réalisation d'un logiciel à une société et la lui paye intégralement, dans une relation de sous-traitance. Un travail a été fait et il est rémunéré. Si, en plus, le client a conservé ses droits de propriété sur le logiciel et décide de le mettre à disposition des autres sous une licence libre, ledit logiciel est alors librement et gratuitement accessible pour tout un chacun. Exit les licences par poste de travail. Bien commun, un logiciel libre est à la disposition de tous. À charge de revanche, mais c'est l'intérêt bien compris des uns et des autres de procéder ainsi, même s'il est vrai qu'il n'est pas toujours évident d'enclencher pareil cycle vertueux...

Le manuel scolaire

La problématique du manuel scolaire s'inscrit donc dans un contexte nouveau fait des turbulences dans lesquelles l'informatique et le numérique ont précipité le monde de l'édition en général, scolaire en particulier.

Le manuel des élèves est la figure centrale, l'« enfant chéri » du modèle français de l'édition scolaire, essentiellement privée. Ce modèle a acquis ses lettres de noblesse et fait ses preuves depuis deux siècles. L'histoire de l'édition scolaire en France est l'histoire des relations entre trois acteurs majeurs, l'État, les éditeurs et les auteurs — des enseignants et des inspecteurs — et de leurs rapports de force mouvants, dans lesquels il arrive que la technique intervienne. Ainsi, à partir de 1811, un auteur ne peut plus soumettre directement à l'État un manuscrit sous prétexte, parmi d'autres raisons, que son examen est long et difficile. Une proposition de manuel doit être obligatoirement communiquée sous forme d'imprimé. L'éditeur devient, de fait, incontournable.

Depuis les années 70, les manuels scolaires se sont profondément transformés [10]. La conséquence en sera qu'ils vivront l'arrivée du numérique dans une situation de fragilité. Quelques rappels, indispensables car ils permettent de concevoir la place du livre papier dans le nouveau contexte issu du numérique.

Les évolutions du manuel scolaire

Quelles évolutions ? L'iconographie occupe jusqu'à 50% de la surface de par les progrès techniques, la quadrichromie. La structure linéaire est abandonnée. L'ouvrage permet des lectures plurielles et des usages multiples, qui préfigurent ceux de l'hypermédia. C'est à l'enseignant de définir un itinéraire parmi des éléments disjoints regroupés en rubriques (activités préparatoires, dossiers, iconographies, exercices) plutôt qu'en chapitres. La structure récurrente de double-page accueille des éléments éclatés. Il n'y a plus de cours en tant que tel, avec des exceptions en langues vivantes (des méthodes), en mathématiques (des démonstrations) et en sciences expérimentales (des expérimentations). On a tendance à se perdre malgré des règles précises (couleurs, plages fonctionnelles), et il faut un mode d'emploi. Le livre devient davantage un outil parmi d'autres à la disposition de l'enseignant qu'un ouvrage de référence pour l'élève. Le résultat combiné de cette complexité croissante des manuels, de la concurrence des méthodes actives, de l'arrivée de nouveaux publics scolaires et de l'hétérogénéité des élèves et des classes qui s'ensuit et de la souplesse de la photocopie est sans appel : on constate depuis les années 80 une tendance à la perte de vitesse de l'utilisation du manuel traditionnel, même si l'attachement symbolique demeure. L'enfant chéri de l'édition scolaire connaît une forme de crise...

Le numérique arrive dans ce contexte. Rude concurrent ! En effet, le manuel scolaire, ce livre d'une centaine de pages, qu'on ne lit plus d'une façon linéaire mais dans lequel on navigue avec des index et des renvois, ce livre ne saurait rivaliser avec Internet et le multimédia, leurs hyperliens, leurs millions de pages et leurs outils de recherche automatisée. Non seulement le livre papier n'est désormais plus « l'enfant unique » mais, de surcroît donc, il aborde « fragilisé » la nouvelle phase introduite par le numérique. Un partage des rôles se dessine. À chacun ses points forts : au livre papier les exposés de connaissances disciplinaires validées, l'écrit, la lecture (sur Internet on passe moins de temps à lire une page), la maniabilité, le temps de l'assimilation ; au numérique, l'accès facilité à une masse de documents grâce aux portails et aux moteurs de recherche, la simulation, les phénomènes dynamiques, les figures en géométrie...

Une autre rupture

Le numérique transforme radicalement le paysage éditorial installé. Nous avons vu la dissociation du texte et du support et ses conséquences mais il y a aussi les relations entre les trois acteurs traditionnels. Certes, les enseignants ont toujours réalisé des documents à l'intention de leurs élèves, en préparant leurs cours. Jusqu'à l'arrivée de l'ordinateur et d'Internet, une élaboration collaborative avec des collègues et la visibilité des ressources produites ne pouvaient aller au-delà d'un cercle restreint

et rapproché. Modifier un document écrit à la main était et demeure une opération lourde, qui plus est quand il circule et que chacun y met sa griffe. Dans les années 70 encore, les photocopieuses étaient rarissimes, les machines à alcool fastidieuses à utiliser.

Des échanges sur une plus grande échelle supposaient de mettre en forme des notes manuscrites, et la machine à écrire manquait de souplesse, ne tolérant pas vraiment les fautes de frappe. Le manuel scolaire était alors la seule perspective pour une diffusion élargie, l'éditeur le passage obligé pour l'enseignant voulant se faire connaître et publier, et on lui accordait d'autant plus facilement des droits sur la fabrication des ouvrages que l'on ne pouvait pas le faire soi-même ! Mais, aujourd'hui, les conditions de la production des ressources pédagogiques, numériques pour une part en augmentation régulière, ont donc radicalement changé, du fait de la banalisation des outils informatiques de réalisation des contenus (du traitement de texte aux logiciels de publication) et d'Internet qui favorise à la fois les productions individuelles et le développement du travail collectif des enseignants, dans leur établissement ou disséminés sur un vaste territoire, à la manière des programmeurs qui écrivent des logiciels libres. La profusion des ressources éducatives que l'on peut consulter sur Internet est là pour en témoigner. Les sites personnels, associatifs, institutionnels se multiplient ainsi que les communautés d'enseignants auteurs-utilisateurs.

Il y a une transférabilité de l'approche du libre, des logiciels à la réalisation des autres biens immatériels. Internet permet aux auteurs de toucher un vaste public potentiel qui peut aisément reproduire leurs documents, les utiliser, les modifier... Utilisant à plein les potentialités d'interaction du Web, les communautés d'enseignants auteurs-utilisateurs fonctionnent comme les communautés de développeurs de logiciels libres. Leurs membres ont une vision et une identité communes. Organisées pour fédérer les contributions volontaires, dans une espèce de synthèse de « la cathédrale et du bazar », ces communautés répondent à des besoins non ou mal couverts et doivent compter en leur sein suffisamment de professionnels ayant des compétences en informatique. Les maîtres mots de la division du travail y sont « déléguez » et « distribuez ». La démarche du libre investit le monde de la production pédagogique. En nombre, les enseignants auteurs-utilisateurs sont devenus des acteurs autonomes à part entière de la production de ressources pédagogiques et de l'édition scolaire. De nouveaux « rapports de force » se sont installés et ils sont eux aussi cause des turbulences qui agitent le secteur [11]. Cela étant, s'il doit redéfinir son rôle, l'éditeur n'en a pas moins des raisons de se montrer optimiste !

« Éditeurs, ayez confiance, le web est une formidable opportunité »

Sur le blog du CNS (Canal numérique des savoirs), Laurent Catach (dictionnaires

Le Robert) propose à ses collègues des voies de sortie des turbulences, leur demandant de se montrer optimistes : « *Éditeurs, ayez confiance, le web est une formidable opportunité, comme sans doute il n'en arrive que tous les quelques siècles...* ». Il s'emploie à les rassurer : « *Il est en effet logique et inévitable que plus la quantité d'informations augmente plus on a besoin de la hiérarchiser, de la filtrer, de la commenter, de l'animer et de la fédérer. Comment par exemple feront les élèves pour se repérer et trouver une juste information dans les 15 millions de livres numérisés de Google ?* ». Il voit « *se profiler un véritable eldorado éditorial !* ». Il met les points sur les i : « *Nous avons à notre disposition un matériau informationnel extraordinaire (toute la richesse du web) à mettre en forme et à mettre en scène. Et l'information et sa mise en scène, n'est-ce pas là très précisément notre métier ?... La question n'est donc pas de savoir si les éditeurs scolaires ont un rôle à jouer sur le web : c'est une évidence. Et c'est même leur responsabilité vis-à-vis des jeunes générations de ne pas laisser les élèves se débrouiller tous seuls avec Internet. La seule et unique question est de savoir comment ils seront rémunérés.* » [12].

Une redéfinition du rôle de l'éditeur

N'étant plus le passage obligé pour qui veut publier, l'éditeur peut s'appuyer, dans la redéfinition partielle de son rôle, sur ses fonctions traditionnelles de sélection, de co-auteur (mise en forme de documents notamment multimédia), de regroupement dans des collections, de facilitation de l'accès (thématique par exemple) à la pléthore de documents disponibles pour les enseignants qui n'ont pas nécessairement le temps nécessaire à leur recherche... qui subsistent. Il doit partager avec les enseignants la « certification de la qualité », coopérer avec eux autour de la fonction de prescription qui reste fondamentale pour un bien d'expérience dont l'utilisateur ne peut savoir a priori s'il lui convient ou non, indépendamment du fait qu'il est coûteux de rechercher une information, surtout lorsque l'on ne sait pas précisément ce que l'on cherche. Il partagera la prescription avec les forums de discussion, les critiques des enseignants, les leaders d'opinion. L'éditeur peut favoriser la constitution d'espaces publics de production, de plates-formes interopérables à base de standards ouverts [13]. Rémunération par le service aussi pour les manuels scolaires papier.

Une édition de produits libres

Le cercle vertueux évoqué ci-avant vaut pour les manuels papier (dont le texte est libre) et pour les ressources pédagogiques numérisées, livres électroniques compris. On peut penser qu'il s'agit d'une tendance profonde sur le long terme : l'expérience montre que l'on ne peut pas s'opposer sur la durée à la logique intrinsèque d'un outil. Or nous avons vu que le numérique et l'informatique permettent la reproduction et la diffusion à tous à coûts (marginaux) quasi nuls, et que, quelque part, le *copyright*

est antinomique avec la logique et la puissance d'Internet. « Drôle d'idée » que de vouloir verrouiller un outil pareil de diffusion de la connaissance. Vaine également...

On assiste cependant à une offensive du « propriétaire », qui n'a jamais abdiqué, avec les tablettes et les livres électroniques : interdiction de télécharger un logiciel de son choix mais produit par un concurrent, difficulté voire impossibilité de transférer un fichier à un collègue, des verrous partout, non-interopérabilité et incompatibilité érigées en « valeurs suprêmes », standards ouverts inconnus au bataillon... On a déjà eu ça avec les tableaux numériques interactifs, un enseignant ne pouvant pas réutiliser ses réalisations s'il change d'école et si le TNI n'est pas le même. Comment espérer le développement d'un marché de masse dans ces conditions ? Et dire que certains se plaignent de l'isolement de l'enseignant dans sa classe. On ne peut pas dire que la coopération et la mutualisation soient facilitées avec ces contextes informatiques verrouillés !

Notes

[1] http://www.sesamath.net/
http://manuel.sesamath.net/?ticket=53595581bd8e8e277c17f579241ca2a580ffe7508933b

[2] http://lelivrescolaire.fr/
http://lelivrescolaire.fr/cgu

[3] Gilles Dowek, Jean-Pierre Archambault, Emmanuel Baccelli, Hugues Bersini, Claudio Cimelli, Albert Cohen, Christine Eisenbeis, Guillaume Le Blanc, Thierry Viéville et Benjamin Wack. (Préface de Gérard Berry), *Informatique et sciences du numérique - Édition spéciale Python*, août 2013, Eyrolles, collection Noire, 342 pages.
http://www.editions-eyrolles.com/Livre/9782212136760/informatique-et-sciences-du-numerique-edition-speciale-python

[4] *Introduction à la science informatique*, Manuel collectif, sous la direction de Gilles Dowek, préface de Gérard Berry, édité par le CRDP de Paris, juillet 2011, 369 pages.
http://crdp.ac-paris.fr/Introduction-a-la-science,27388
http://science-info-lycee.fr/actualites/incourtournable-a-lire/

[5] Ce texte correspond à une intervention lors de la journée organisée le 3 juin 2014 par le consortium couperin.org « Livre électronique et open access ». http://jle-couperin.sciencesconf.org/
http://jle-couperin.sciencesconf.org/resource/page/id/1

[6] http://www.epi.asso.fr/revue/articles/a0904b.htm
http://www.education.gouv.fr/bo/2007/5/MENJ0700078X.htm

[7] Déjà, Confucius mettait en garde : « Lorsque tu fais quelque chose, sache que tu auras contre toi ceux qui voulaient faire la même chose, ceux qui voulaient faire le contraire et l'immense majorité de ceux qui ne voulaient rien faire. »

[8] Gilles Dowek et Jean-Pierre Archambault, *Pourquoi pas une licence globale pour les manuels scolaires ?*, Rue89, 30 novembre 2012.
http://www.rue89.com/rue89-culture/2012/11/30/pourquoi-pas-une-licence-globale-pour-les-manuels-scolaires-237472
http://www.epi.asso.fr/revue/articles/a1301c.htm

[9] Et rappelons que dans les années 80, le Minitel, terminal qui avait un coût non négligeable de l'ordre de 5000 F (800 €), était fourni gratuitement par la Direction Générale des Télécommunications. Et il a engendré une économie de services télématiques florissante et beaucoup de communications.

[10] Alain Choppin (INRP), *Les manuels scolaires : histoire et actualité*, Hachette 1992.

[11] Jean-Pierre Archambault, *Les turbulences de l'édition scolaire*, colloque SIF « Les institutions éducatives face au numérique », organisé par la Maison des Sciences de l'Homme de Paris-Nord, Paris les 12 et 13 décembre 2005.
http://lamaisondesenseignants.com/download/document/sifArchambault.pdf

[12] Jean-Pierre Archambault, *Numérique, droit d'auteur et pédagogie*, Terminal n° 102, Automne-Hiver 2008-2009, édition l'Harmattan, pp. 143-155. EpiNet n° 114, avril 2009 (note 2).
http://www.epi.asso.fr/revue/articles/a0904b.htm

[13] Jean-Pierre Archambault, *Un spectre hante le monde de l'édition*, Médialog n° 63 de septembre 2007, pp. 42-45.
http://medialog.ac-creteil.fr/ARCHIVE63/jpa63.pdf

Leslie Lamport : l'importance de la pensée mathématique pour l'informatique

Stephan Merz [1], Anca Muscholl [2]

En mars dernier, l'ACM a annoncé que le prix Turing 2013 était attribué à Leslie Lamport pour ses « contributions fondamentales à la théorie et à la pratique des systèmes répartis et paralléles, notamment l'invention de concepts comme la causalité et les horloges logiques, de sûreté et de vivacité, les machines à états répliqués et la consistance séquentielle » [3].

L'original en anglais de cette citation, ainsi que plusieurs développements, se trouvent sur les sites :

`http://www.acm.org/news/featured/awards/turing-award-2013`

et

`http://amturing.acm.org/award_winners/lamport_1205376.cfm`

L'article le plus cité de Leslie Lamport, qui est également un des plus cités de l'histoire de l'informatique, s'intitule « Time, Clocks, and the Ordering of Events in a Distributed System *», et pose les fondements du calcul réparti. On le trouve à :*

`http://amturing.acm.org/p558-lamport.pdf`

1. Directeur de recherche au Centre de recherche Inria Nancy - Grand Est, membre du LORIA.
2. Professeure à l'université de Bordeaux, membre du LaBRI et du Conseil scientifique de la SIF.
3. Voir également l'article paru dans le numéro 3 de 1024, `http://www.societe-informatique-de-france.fr/wp-content/uploads/2014/05/1024-3-lamport.pdf`.

La rubrique REGARDS de ce numéro présente les opinions de Lamport sur l'importance de la pensée mathématique pour l'informatique, qui motivent ses travaux actuels sur la spécification formelle et la vérification de programmes parallèles.

Enfin, n'oublions pas son rôle essentiel dans le développement du logiciel LaTeX *dont la communauté scientifique ne saurait se passer.*

Marie-Claude Gaudel

En janvier 2013 Leslie Lamport [4] est invité à donner son opinion sur les pratiques courantes en matière de programmation au site grand public `www.wired.com`. Son article *Pourquoi nous devrions construire des logiciels comme nous construisons des maisons* [5] suggère que les programmeurs devraient consacrer un peu de leur temps à décrire, ne serait-ce qu'en langue courante, ce qui est attendu d'un morceau de code, et ceci avant de le produire. « Réfléchir ne garantit pas que nous ne commettrons pas d'erreurs. Ne pas réfléchir garantit que nous en allons commettre. » Cet article suscite l'incompréhension de beaucoup de lecteurs qui jugent notamment que les contraintes de l'industrie moderne du logiciel ne laissent pas de temps pour rédiger des spécifications qui deviendront d'ailleurs très vite obsolètes à cause des évolutions incessantes des besoins.

Lamport n'en est pas à son coup d'essai : dès 1977, inspiré par les travaux de Floyd et de Hoare, il avait réfléchi sur la perception de la programmation dans un monde moderne qui en est de plus en plus dépendant. Dans une présentation à l'Université de Kiel (Allemagne) en 2003, il compare les programmes modernes à des systèmes biologiques, dont la complexité les rend difficilement compréhensibles. Il y prône une approche rigoureuse selon laquelle un programme est un objet mathématique complexe mais qui reste à la portée d'une analyse logique. D'après Lamport, la logique doit intervenir dans la conception de programmes et le raisonemment sur leurs propriétés, et elle ne doit pas être réservée aux seuls spécialistes, mais ses concepts fondamentaux doivent être compris par les développeurs et ingénieurs : « La pensée mathématique est quelque chose que les ingénieurs devraient apprendre à l'université plutôt que d'y être confrontés seulement en milieu professionnel. Ils peuvent apprendre au travail les spécificités d'un langage de programmation, mais la façon de penser doit être enseignée par l'université. De nos jours, les étudiants américains peuvent apprendre C++, Java et les détails d'Unix. Mais très peu d'universités américaines offrent des cours qui abordent la conception des systèmes réels sous un angle mathématique et logique. Les universités sont appelées à faire plus d'efforts dans cette direction. »

4. `http://www.lamport.org/`
5. `http://www.wired.com/2013/01/code-bugs-programming-why-we-need-specs/`

Les contributions principales de Lamport concernent la conception d'algorithmes répartis, c'est-à-dire exécutés par des nœuds indépendants dans un réseau d'ordinateurs. Dans ce domaine, selon le dicton bien connu de Lamport, « la panne d'un ordinateur dont vous ignoriez même l'existence peut faire capoter le calcul ». Les apports de Lamport ont été profonds et ont marqué le domaine, que ce soit par la formalisation de notions clés comme celle de tolérance aux pannes (problème des généraux byzantins) ou par la conception d'algorithmes. Par exemple, l'algorithme *Paxos* qui vise à garantir un consensus parmi des nœuds répartis est à la base de nombreux services Internet et Cloud modernes, comme la gestion de documents modifiables par des utilisateurs distants ou les services d'achats en ligne.

Énoncer précisément les fonctionnalités attendues d'un algorithme réparti et décrire sans ambiguïté son opération, ainsi que les hypothèses (mode de communication, pannes...) sur lesquelles repose son fonctionnement est fondamental pour concevoir des systèmes fiables. Aussi existe-t-il un sous-domaine de l'informatique qui conçoit des langages et formalismes pour la spécification d'algorithmes et de systèmes informatiques et pour raisonner sur leurs propriétés. Fidèle à sa formation initiale de mathématicien, Lamport se montre sceptique à l'égard de formalismes ad-hoc et prône l'utilisation, autant que possible, des notations et concepts mathématiques habituels dont il considère qu'ils s'appliquent parfaitement aux systèmes informatiques. De ces idées est né le langage logique de spécification TLA^+, dont le but est de rendre la conception et le raisonnement formel autour des systèmes concurrents complexes plus abordable.

Quelle conclusion tirer sur les contributions de Leslie Lamport, qui a marqué l'informatique moderne par son approche originale, à la fois scientifique et pourvue d'un sens prononcé pour les problèmes réels ? Un des messages clés de l'activité de Lamport est que l'informatique, et en particulier la programmation, doit se nourrir d'une conception rigoureuse, mathématique, qui font de l'informatique la science que nous aimons, développons et enseignons : « Lorsque des gens qui ne pensent pas de façon logique conçoivent de grands systèmes, ces systèmes deviennent incompréhensibles. Et alors nous commençons à regarder ces systèmes comme des systèmes biologiques. Les systèmes biologiques étant trop complexes, il semble très naturel de penser aussi que les systèmes informatiques sont trop complexes pour être maîtrisés. Nous ne devons pas accepter cette situation — « nous » signifiant les informaticiens en première ligne, mais aussi les utilisateurs de l'informatique. »

Gérard Berry : un informaticien médaille d'or du CNRS 2014

Serge Abiteboul [1], Laurent Fribourg [2] et Jean Goubault-Larrecq [3]

C'est un chercheur en informatique qui vient de recevoir la médaille d'or du CNRS, la plus haute distinction scientifique française toutes disciplines confondues. Les informaticiens sont rares à avoir été ainsi honorés : ce n'est que la seconde fois, après Jacques Stern [4] en 2006.

Gérard Berry, traqueur de bugs

Gérard Berry [5] est un pionnier dans un nombre considérable de domaines informatiques : le lambda-calcul, la programmation temps réel, la conception de circuits intégrés synchrones, la vérification de programmes et circuits, l'orchestration de services Web. Il a été l'un des premiers informaticiens académiciens des sciences, le premier professeur d'informatique au Collège de France.

Parmi ses grandes inventions, essayons d'en expliquer une, le langage Esterel [6].

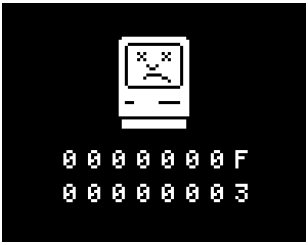
Nous sommes entourés de systèmes d'une incroyable complexité : téléphones, moteurs de recherche, avions, centrales nucléaires. Ils fonctionnent tous avec du matériel informatique (des circuits) et du logiciel informatique (des programmes). Mais alors que le plantage d'un téléphone ou même d'un moteur de recherche est

1. Directeur de Recherche Inria et professeur à l'ENS Cachan.
2. Directeur de Recherche CNRS à l'ENS Cachan.
3. Professeur à l'ENS Cachan.
4. http://www.di.ens.fr/~stern/
5. http://www.college-de-france.fr/site/gerard-berry/#course
6. http://fr.wikipedia.org/wiki/Esterel_%28langage%29

anodin, il en est tout différemment des avions, des centrales nucléaires ou encore des pacemakers. Pour ces derniers, le bug peut provoquer un désastre. Pour ne donner qu'un exemple, c'est un bug qui est à l'origine de la destruction d'Ariane 5, de l'Agence spatiale européenne, quarante secondes seulement après son décollage, le 4 juin 1996. Or un bug, c'est souvent une seule ligne de code erronée sur des millions qui composent un programme. Ça vient vite ! Et ça peut faire mal.

Gérard Berry en cours au Collège de France

Comment éviter les bugs ? On peut bien sûr tester davantage les programmes. Cela permet de trouver beaucoup d'erreurs, mais combien d'autres passeront à travers les mailles du filet ? Une autre solution, c'est d'intervenir en amont dans le processus de création de programme, par exemple en fournissant aux informaticiens de meilleurs outils de conception, de meilleurs langages de programmation. C'est l'approche que prône Gérard Berry.



Écran indiquant un code erreur sur les premières versions de Macintosh
`fr.wikipedia.org/wiki/Macintosh`

Les langages de programmation standards sont mal adaptés aux situations rencontrées dans des systèmes aussi complexes que des avions. Il faut tenir compte à la fois du matériel et du logiciel, du fait que de nombreuses tâches s'exécutent en parallèle, que parfois la même tâche est exécutée plusieurs fois pour se protéger d'une panne d'un composant. Surtout, il faut utiliser des modèles qui tiennent compte du temps, des délais de réponse, des mécanismes de synchronisation. Le nouveau concept de langage synchrone a permis de répondre à cette situation. Ce concept a été découvert et promu par Gérard Berry et ses collègues au travers notamment du langage Esterel. Ce langage ainsi que d'autres langages synchrones développés en France, comme Lustre et Signal, ont eu un impact majeur dans le monde entier.

Mais comment un langage peut-il aider à résoudre des problèmes aussi complexes ? D'abord, parce qu'il permet de décrire les algorithmes que l'on veut implémenter sous une forme compacte et proche de l'intention du programmeur. Ensuite, parce que ce langage est accompagné de toute une chaîne d'outils de compilation et de vérification automatique, qui garantit que le produit final est correct.

Les langages synchrones ont constitué une avancée scientifique majeure. Gérard Berry est allé plus loin encore, et les a emmenés dans l'aventure industrielle en cofondant la société Esterel Technologies [7]. Ces langages sont aujourd'hui incontournables dans des domaines comme l'aérospatial ou l'énergie.

Sa notoriété, Gérard l'a aussi mise au service du partage de la culture scientifique pour permettre à chacune et chacun de comprendre « pourquoi et comment le monde devient numérique ». Il passe un temps considérable, souvent avec des jeunes, à expliquer les fondements et les principes de la science informatique. Un défi !

L'enseignement de l'informatique est l'un de ses grands combats. Il a dirigé avec Gilles Dowek l'écriture d'un rapport important sur la question [8]. Un autre défi ! Alors qu'il s'agit d'éducation, de science et de technique, l'État se focalise souvent sur le haut débit et l'achat de matériel. Un pas en avant et au moins un en arrière. Mais il en faut bien plus pour entamer l'enthousiasme de Gérard Berry.

Les mardis de la science
`www.lairedu.fr/science-conscience-chez-les-shadoks/`

Gérard Berry est un inventeur. Au-delà d'Esterel, c'est un découvreur des modèles stables du lambda-calcul, un inventeur de machine abstraite chimique, un concepteur de langages d'orchestration d'objets communicants, comme HipHop. Gérard Berry s'investit avec enthousiasme dans ses nombreuses fonctions des plus académiques, comme professeur au Collège de France, aux plus mystérieuses, comme régent de déformatique du Collège de Pataphysique.

Et quand vous passerez devant une centrale nucléaire, admirez le fait que même s'il utilise des logiciels et matériels bien plus compliqués que votre téléphone, son système informatique ne « plante » pas, contrairement à celui de votre téléphone. Quand vous volerez, peut-être au-dessus de l'Atlantique, réjouissez-vous que le fonctionnement de votre avion soit plus fiable que celui de votre tablette. Et puis, de loin en loin, pensez que tout cela est possible parce que des chercheurs en informatique comme Gérard Berry ont mis toute leur

7. `http://www.esterel-technologies.com/`
8. `http://www.academie-sciences.fr/activite/rapport/rads_0513.pdf`

créativité, toute leur intelligence pour développer cette science et cette technique qui garantissent la fiabilité des systèmes informatiques.

Pour aller plus loin

[1] « Science et conscience chez les Shadoks ! », vidéo, http://www.youtube.com/watch?v=g2Ct67WJjhM

[2] « L'enseignement de l'informatique en France. Il est urgent de ne plus attendre », rapport de l'Académie des sciences, 2013, http://www.academie-sciences.fr/activite/rapport/rads_0513.pdf

[3] « L'informatique du temps et des événements », cours au Collège de France (2012-2013), http://www.college-de-france.fr/site/gerard-berry/index.htm#%7Cm=course%7Cq=/site/gerard-berry/course-2012-2013.htm

[4] « Penser, modéliser et maîtriser le calcul », cours au Collège de France (2009-2010), http://www.college-de-france.fr/site/gerard-berry/index.htm#%7Cm=course%7Cq=/site/gerard-berry/course-2009-2010.htm

[5] « Pourquoi et comment le monde devient numérique », cours au Collège de France (2007-2008), http://www.college-de-france.fr/site/gerard-berry/index.htm#%7Cm=course%7Cq=/site/gerard-berry/course-2007-2008.htm

[6] *Entretien avec Gérard Berry*, Valérie Schafer, paru dans le numéro 2 de 1024, janvier 2014, http://www.societe-informatique-de-france.fr/wp-content/uploads/2014/02/1024-2-berry.pdf.

[7] Une lecture qui tente d'aller plus loin dans les explications : « Mais pourquoi une médaille d'or de la recherche à un informaticien ? », Thierry Viéville, http://www.scilogs.fr/intelligence-mecanique/mais-pourquoi-une-medaille-dor-de-la-recherche-un-informaticien/

Post-scriptum. Une citation de Gérard Berry, pataphysicien : « L'informatique, c'est la science de l'information, la déformatique, c'est le contraire. »

Cet article est également paru sur le blog BINAIRE [9].

9. http://binaire.blog.lemonde.fr/

Le bon algorithme de jeu ?

Jean-Paul Delahaye [1]

La rubrique récréation informatique propose une petite énigme algorithmique ou à propos d'un thème de mathématiques discrètes susceptible d'intéresser un lecteur de 1024. La solution est donnée dans le numéro suivant.

Rappel et solution du problème précédent

Vous êtes face à une machine qui, lorsque vous lui donnez un entier n, vous répond par $P(n)$, où P est un polynôme à coefficients entiers positifs dont vous ignorez le degré et les coefficients. Vous voulez connaître le polynôme que la machine utilise (donc son degré et ses coefficients). Chaque calcul de $P(n)$ est payant et vous coûte 100 euros. Combien allez-vous dépenser en vous y prenant au mieux ?

SOLUTION. La bonne réponse m'a été envoyée par plusieurs lecteurs qui sont dans l'ordre d'arrivée des messages : Julien Bernard, Denis Monnerat, Eric Wegrzynowski et Arnaud Fréville. Merci à eux de leur participation.

Il fallait répondre 200 euros. En effet, deux évaluations de P suffisent pour connaître P.

On commence par demander $P(1)$.

— Si $P(1) = 0$, c'est que tous les coefficients sont nuls.

— Si $P(1) = 1$, c'est qu'il n'y a qu'un seul coefficient non nul qui vaut 1. Le polynôme est donc de la forme $P(X) = X^k$. Pour connaître k, on demande $P(2) = a$, qui bien sûr permet (en prenant le logarithme en base 2 de a) de connaître k, et donc P.

1. Université de Lille 1, Sciences et Technologies, Laboratoire d'Informatique Fondamentale de Lille, UMR 8022 CNRS, Bât M3-ext, 59655 Villeneuve d'Ascq Cedex. E-mail : delahaye@lifl.fr.

— Si $P(1) = a > 1$, on demande $P(a+1) = b$. L'entier $a+1$ est un entier strictement supérieur à chaque coefficient du polynôme et

$$b = P(a+1) = a_0 + a_1(a+1)^1 + a_2(a+1)^2 + ... + a_k(a+1)^k.$$

On calcule alors la décomposition en base de numération $(a+1)$ de b (il faut quelques divisions euclidiennes). On trouve bien sûr que b s'écrit

$$a_k \ldots a_2 a_1 a_0.$$

On a donc les coefficients du polynôme, et bien sûr son degré.

Nouveau problème : le bon algorithme de jeu ?

Cinquante boîtes sont placées en ligne sur une table devant vous :

```
1 2 3 4 5 ... 48 49 50
```

Elles sont numérotées de 1 à 50. Dans chacune a été enfermée une somme d'argent dont le montant est inscrit bien clairement sur le couvercle. Vous pouvez vérifier que tout est exact. Vous allez jouer contre un adversaire qui pourra utiliser un ordinateur pour s'aider s'il le souhaite, alors que vous devrez vous contenter de calculs mentaux. La règle est simple : à tour de rôle, vous prendrez une boîte à l'une des deux extrémités de la ligne des boîtes encore présentes, et elle vous sera temporairement acquise. Quand la ligne de boîtes sera épuisée, celui des deux joueurs qui aura acquis la plus forte somme sera déclaré vainqueur et emportera alors le total du contenu de toutes les boîtes. Pour éviter les parties nulles, le total de l'argent contenu dans les boîtes est impair et chaque boîte contient un nombre entier d'euros.

Puisque votre adversaire a l'avantage de pouvoir s'aider d'un ordinateur pour faire ses choix, on vous laisse commencer. Pouvez-vous gagner de manière certaine ? Si oui, quel est l'algorithme de jeu qui permet cela ? Sinon, pourquoi ?

Envoyez vos réponses à delahaye@lifl.fr. Le nom des premiers lecteurs à me donner la bonne réponse (et à la justifier) seront mentionnés dans le prochain numéro de 1024.

1024

BULLETIN

de la société informatique de France

Institut Henri Poincaré,
11 rue Pierre et Marie Curie,
75231 Paris Cedex