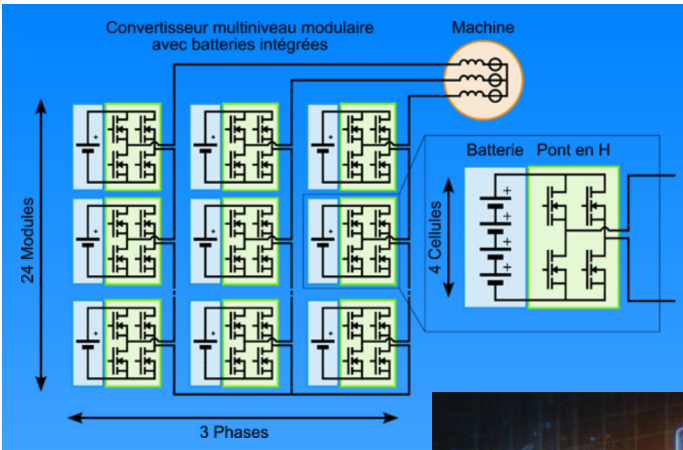
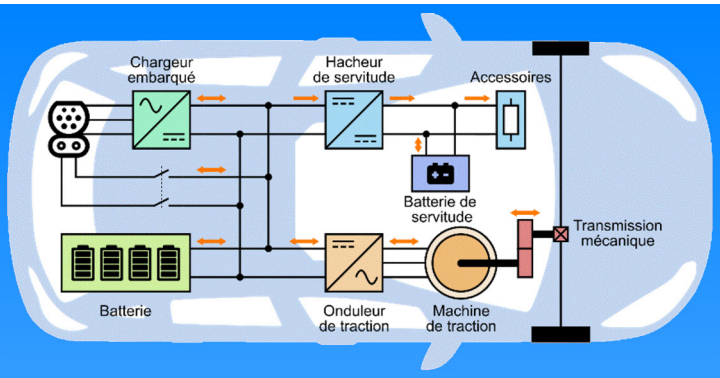
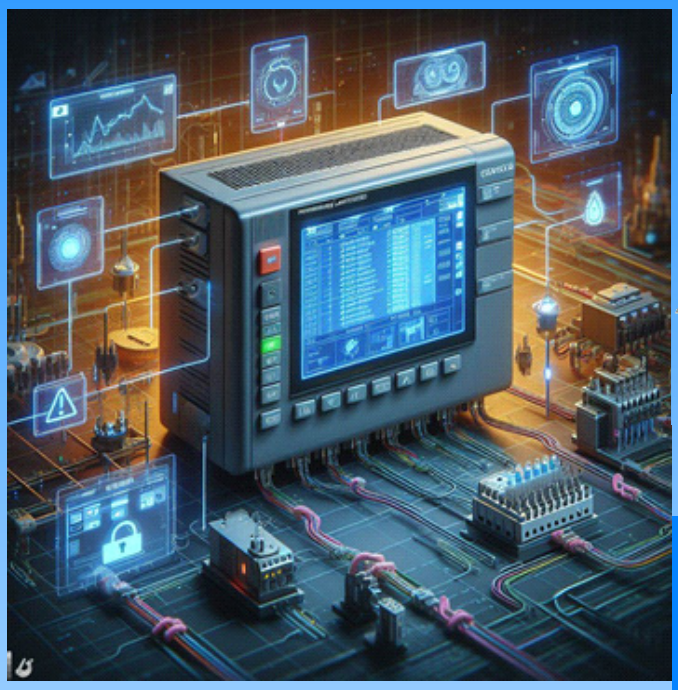
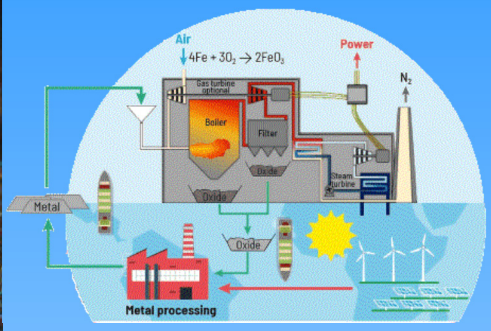
29ème année

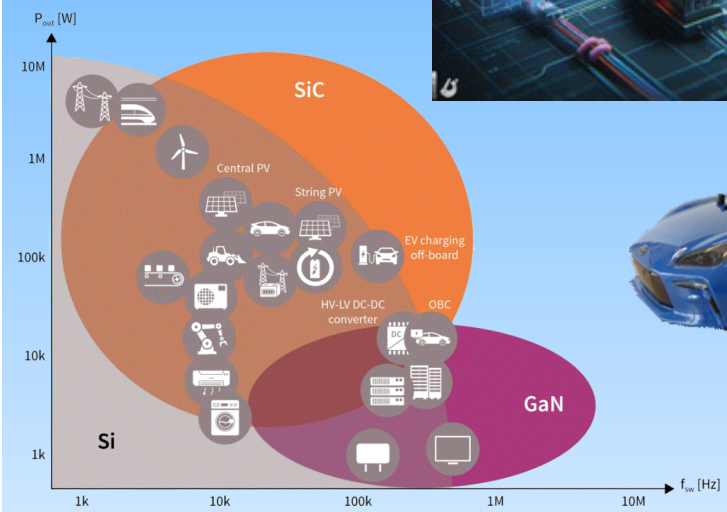
La Revue 3EI











Cybersécurité des systèmes industriels - Électronique de puissance

Publication trimestrielle du Cercle Thématique 13.01 de la SEE

ENSEIGNER L'ÉLECTROTECHNIQUE ET L'ÉLECTRONIQUE INDUSTRIELLE

Société de l'Électricité, de l'Électronique
et des Technologies de l'Information
et de la Communication

N°111 – Janvier 2024

La Revue 3E.I
Publication trimestrielle de la SEE

Hébergé par :
Culture Sciences de l'Ingénieur

SOCIETE de l'ELECTRICITE, de l'ELECTRONIQUE et des TECHNOLOGIES de l'INFORMATION et de la COMMUNICATION.

17, rue de l'Amiral Hamelin, 75116 PARIS
Tél : 01 56 90 37 17
www.see.asso.fr

SEE, association reconnue d'utilité publique par le décret du 7 décembre 1886
Siret 785 393 232 00042, APE 9412 Z, n° d'identification FR 44 785 393 232

4 avenue des Sciences, 91190 Gif sur Yvette
tel : 01 81 87 55 22
https://eduscol.education.fr/sti/si-ens-paris-saclay

La Revue 3E.I

3EI : Enseigner l'Électrotechnique et l'Électronique Industrielle

La Revue 3EI, *Édition SEE,*
17 rue de l'Amiral Hamelin
75116 PARIS

Directeur de la publication
François GERIN
Président de la SEE

Rédacteur en Chef
Franck LE GALL

Adresser les propositions d'article à :
revue3ei@gmail.com

Communication :
Mme. Mélisande DE LASSENCE
Communication1@see.asso.fr
01 56 90 37 17

Dépôt Légal : 1er trimestre 2024

ISSN 1252-770X

Sommaire du n° 111

Editorial

La Revue 3EI *est de retour !*

Après un an de réflexions sur l'avenir de notre revue, la voici de nouveau entre vos mains. Gardant l'esprit de partage qui anime le comité de rédaction depuis sa création en 1995, et après analyse des réponses à la consultation que nous avons lancée, nous avons souhaité la renouveler pour qu'elle soit :

- ***Gratuite** : Partagez-là sans limite avec vos collègues, vos étudiants…*
- ***Numérique** : Disponible sur les sites de la SEE et de Culture Science de l'Ingénieur, vous pourrez aussi y trouver des ressources numériques qui compléteront le contenu des articles.*
- ***Participative** : C'est votre revue. Vous pouvez proposer des thèmes, des articles, des informations. La liste de diffusion (https://groupes.renater.fr/sympa/info/revue3ei) est créée pour cela. Inscrivez-vous !*
- ***Ouverte** sur tous les domaines du Génie Electrique, de l'Informatique Industrielle et de la Physique Appliquée.*
- ***Un lien** entre les enseignants des différentes formations du supérieur du domaine GEII (IUT, BTS, Ecoles d'Ingénieurs, Université).*

Nous ouvrons cette nouvelle version de La Revue 3EI qui restera trimestrielle avec deux dossiers dédiés respectivement à la cybersécurité des systèmes industriels et à l'électronique de puissance. Ces dossiers ouverts dans ce numéro 111 s'enrichiront au fil du temps d'autres articles disponibles sur le site Culture Science de L'Ingénieur. Vous pouvez dès maintenant y contribuer en proposant à l'adresse revue3ei@gmail.com vos propres textes sur vos retours d'expériences dans vos enseignements ou votre pratique professionnelle.
La section hors thème de la revue est là aussi pour accueillir vos articles sur des sujets qui ne sont pas couverts par les dossiers déjà ouverts.

« Dossier : Cybersécurité des systèmes industriels »

La cybersécurité des systèmes industriels qui se trouve à l'intersection entre la cybersécurité des systèmes informatiques et l'informatique industrielle est confrontée à une augmentation continue du nombre des attaques malveillantes.
La prise de conscience de ces menaces induit la nécessité d'une meilleure formation des acteurs à la cybersécurité, d'où notamment l'évolution du BTS Systèmes Numériques en BTS Cybersécurité, Informatique et Réseaux, Électronique (CIEL) et l'introduction de la cybersécurité dans le programme national des BUT GEII et R&T.
Accompagnant ce mouvement, ce dossier vise à proposer aux enseignants d'informatique industrielle des ressources théoriques, des exemples de travaux pratiques et des témoignages d'industriels, organisés en 3 domaines :

- *Cybersécurité des systèmes automatisés industriels.*
- *Cybersécurité des objets connectés.*
- *Cybersécurité des systèmes embarqués (essentiellement automobiles).*

En complément, une dernière partie présentera une introduction à la cybersécurité au niveau matériel des systèmes informatiques, problématique commune à tous les domaines de l'informatique.

« Dossier : Innovations en cours en électronique de puissance »

L'électronique de puissance est une branche relativement récente du génie électrique (début des années 1960). A l'heure actuelle la dynamique d'évolution de l'électronique de puissance reste importante, en partie grâce aux nouveaux semiconducteurs et aux progrès des matériaux diélectriques et magnétiques. Par ailleurs, les impératifs de développement durable imposent à présent de considérer, lors de la phase d'étude et de conception d'un convertisseur électronique de puissance, des critères de minimisation des impacts environnementaux sur cycle de vie, ce qui nécessite une approche système, une meilleure compréhension des mécanismes de défaillance et de dégradation, mais également de considérer la recyclabilité et la réparabilité des dispositifs.
L'objectif de ce dossier « électronique de puissance » est de fournir aux lecteurs les éléments de compréhension des évolutions en cours et les enjeux technologiques et sociétaux de cette discipline grâce à quelques exemples illustratifs d'applications innovantes.

« Hors Thème »

La rubrique hors thème regroupe des articles qui vont de l'informatique embarquée à l'énergie en passant par l'électronique.
Les quatre premiers articles dédiés à l'informatique embarquée décrivent la mise en œuvre des voitures et du simulateur utilisés dans le cadre de la Course Voitures Autonomes Paris Saclay (CoVAPSy).
L'article de Rémi Al Ajroudi et ses collègues directement exploitable pour un projet avec des étudiants traite d'une application de contrôle de la température à l'aide d'un module Peltier.
Nous refermons ce numéro avec la publication de Driss Laraqui qui nous fait découvrir la nouvelle filière énergétique des combustibles métalliques renouvelables dans le contexte actuel de la décarbonation des activités humaines.

Introduction à la cybersécurité des systèmes industriels

La Revue 3E.I
Culture Sciences de l'Ingénieur

Édité le 15/02/2024

école normale supérieure paris—saclay

Maxime SECHEHAYE[1], Anthony JUTON[2]

[1] *Etudiant en M2 à l'ENS Paris-Saclay – DER Nikola Tesla*

[2] *Professeur agrégé à l'ENS Paris-Saclay – DER Nikola Tesla*

Cette ressource fait partie du N°111 de La Revue 3EI de janvier 2024.

Cette ressource introduit le dossier sur la cybersécurité en définissant ses objectifs, les principes fondamentaux de la cybersécurité ainsi que ses enjeux.

D'après le dictionnaire *Le Robert*, la cybersécurité désigne « l'ensemble des moyens utilisés pour assurer la sécurité des systèmes et des données informatiques d'un État, d'une entreprise, etc. ». Aborder la cybersécurité dans son ensemble est donc une tâche colossale qui nécessite des connaissances et des compétences dans de très nombreux domaines : économie, diplomatie, droit, sociologie, renseignement, etc. Se former en cybersécurité, ce n'est donc pas uniquement apprendre de bonnes pratiques en informatique, le champ d'application est bien plus vaste.

Cette ressource définit donc les limites de ce dossier qui ne se veut pas exhaustif, puis précise le public visé, avant de présenter quelques attaques récentes liées à des problématiques de cybersécurité pour commencer à sensibiliser le lecteur.

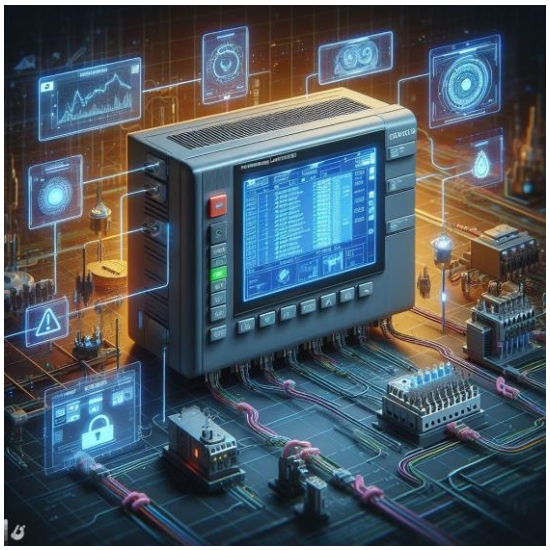


Figure 1 : Cybersécurité des systèmes automatisés industriels, vue par Microsoft Designer / DALL-E3

1 - Les champs de la cybersécurité couverts par ce dossier

Comme beaucoup des disciplines informatiques, la cybersécurité des systèmes informatiques est un sujet largement traité avec des ressources de qualité accessibles librement et destinées à des informaticiens.

La cybersécurité des systèmes industriels est à l'intersection entre la cybersécurité des systèmes informatiques et le champ disciplinaire nommé en France « informatique industrielle » qui regroupe l'automatisme industriel et l'informatique embarquée. Pour le premier, l'« industrie 4.0 » et pour le second, l'« internet des objets », sont deux expressions (aux contours peu précis) très utilisées qui traduisent la tendance à la « connexion de tout, toujours et partout ». Ceci élargit la surface vulnérable des systèmes industriels et amène à une augmentation du nombre des attaques. Ces systèmes, automatisés, embarqués ou IoT, ont de plus une durée de vie importante allant souvent au-delà de la période couverte par les mises à jour de sécurité (quand elles sont faites...).

Cela implique une meilleure formation des acteurs, notamment ceux issus de l'informatique industrielle, à la cybersécurité, d'où notamment l'évolution du BTS Systèmes Numériques en BTS Cybersécurité, Informatique et réseaux, Électronique (CIEL) et l'introduction de la cybersécurité dans le programme national des BUT GEII et R&T (pour lequel existe désormais un parcours cybersécurité).

Accompagnant ce mouvement, ce dossier vise à proposer aux enseignants d'informatique industrielle des ressources théoriques, des exemples de travaux pratiques et des témoignages d'industriels, organisés en trois domaines, pas complètement indépendants, après une partie introductive :

- Cybersécurité des systèmes automatisés industriels,
- Cybersécurité des objets connectés,
- Cybersécurité des systèmes embarqués (essentiellement automobiles).

Pour chacun de ces ensembles de systèmes, le dossier s'intéresse essentiellement aux vulnérabilités et protections au niveau réseau, en restant dans le cadre des compétences attendues d'un technicien ou ingénieur en GEII.

En complément, une dernière partie présentera une introduction à la cybersécurité au niveau matériel des systèmes informatiques, problématique commune à tous les domaines de l'informatique.

2 - À qui s'adresse ce dossier ?

Ce dossier s'adresse à la fois aux enseignants, aux techniciens et aux ingénieurs en ingénierie électrique ou en informatique embarquée.

Par sa grande complexité, abordée dans le paragraphe introductif, la cybersécurité est un domaine d'experts. Ce dossier n'a pas vocation à former des experts en cybersécurité. Il se propose au contraire de sensibiliser les acteurs travaillant dans des domaines concernés par la cybersécurité pour qu'ils puissent prendre en compte ces problématiques lors de la conception de systèmes.

En effet, la prise en compte des problématiques de cybersécurité dès la phase de conception d'un système avec l'application des bonnes pratiques et des solutions existantes est le plus souvent suffisante pour se protéger des attaques classiques réalisées par des individus malveillants isolés.

3 - Exemples d'attaques liées à des problématiques de cybersécurité

Cette partie présente trois exemples, issus des fiches d'incidents [1] publiées périodiquement par le Clusif, association française de promotion de la cybersécurité.

3.1 - Systèmes automatisés – Prise de contrôle du système de production d'une aciérie

Figure 2 : fiche 32 [1] : Prise de contrôle du système de production d'une aciérie

Cette fiche a été choisie car représentative des risques en cybersécurité des systèmes industriels et des conséquences importantes. L'intrusion a eu lieu par une campagne de mails frauduleux (*phishing*) et a permis de s'introduire sur le réseau de bureautique. De ce réseau bureautique, les hackers ont eu accès au réseau industriel pour prendre les commandes des systèmes de production et désactiver les mises en sécurité d'un haut fourneau jusqu'à provoquer de lourds dégâts.

La fiche rédigée par le Clusif à propos de cet incident propose les contre-mesures suivantes :

- Sensibilisation des agents aux méthodes d'attaque par *spear phishing* ;
- Restriction des droits accordés aux profils d'agent sur le réseau et les systèmes, de façon à détecter, voire empêcher toute action suspecte (prise de contrôle de systèmes, de terminaux…) ;
- Cloisonnement des réseaux de bureautique exposés aux attaques et aux intrusions, et des réseaux de contrôle des systèmes de production ;
- Mise en place de mécanismes de sûreté indépendants du système de conduite.

Parmi les fiches Clusif, un nombre important relate des attaques sur des systèmes industriels (exemple : empoisonnement de l'eau dans une usine de production d'eau, fiche 19) ou énergétiques (exemple : Black Energy, fiche 4 – coupure de l'électricité en Ukraine). La majeure partie pourrait être évitée par une bonne application des règles de cybersécurité abordées dans la partie 1 du dossier, consacrée aux systèmes automatisés industriels.

Quelques attaques sont le résultat d'un affrontement entre puissances étrangères, notamment l'attaque Stuxnet (fiche 36) qui a permis en 2010 aux services secrets israéliens de saboter les centrifugeuses iraniennes enrichissant l'uranium.

L'attaque Stuxnet en 2010 a révélé la vulnérabilité des systèmes automatisés et permis de prendre conscience des risques encourus et de la nécessité de mettre en œuvre une politique de cybersécurité pour les systèmes industriels également. Les nombreuses attaques qui ont suivi, comme celle présentée ici, mettent en évidence la lente formation des automaticiens à la cybersécurité et la longue marche vers la sécurisation de tous les équipements, pour certains en fonctionnement depuis longtemps.

3.2 - Objets connectés – Attaque sur une pompe à insuline

Quelques fiches s'intéressent aux objets connectés, notamment la fiche 41 qui présente la prise de contrôle à distance d'une pompe à insuline.

Figure 3 : Fiche 41 [1] : Attaque sur une pompe à insuline

Après l'analyse de la documentation constructeur (manuel d'utilisation, analyse des brevets, numéro de série de l'appareil…), un chercheur est parvenu à intercepter les communications échangées entre les capteurs et sa pompe à insuline et établir la liste des codes de commande utiles de l'équipement.

Le chercheur a alors imaginé plusieurs scénarios d'attaque : rejeu (l'entité malveillante intercepte puis réitère une transmission de données valide) de valeurs transmises à la pompe par les sondes, envoi de commandes forgées directement à la pompe (accès physique requis pour connaître le numéro de série nécessaire à l'envoi).

La fiche à propos de cet incident propose les contre-mesures suivantes :

- Forcer l'authentification mutuelle des sondes et pompes à insuline ;
- Chiffrer les signaux échangés ;
- En conclusion : intégrer la sécurité dans la phase de conception de ces objets.

Cette attaque met en valeur le manque de mesures de sécurité lors de la conception d'objets connectés. Les objets, parfois d'un coût peu élevé, sont conçus par des électroniciens qui valident le bon fonctionnement sans toujours connaître les attaques que leur dispositif risque d'affronter, parfois dans plusieurs années.

La fiche 22 du Clusif montre comment un adolescent a pu reproduire une télécommande d'aiguillage de Tramway. Martin Hron, chercheur en sécurité chez Avast, a publié pour sa part un article présentant la possibilité d'attaquer une machine à café connectée [3]. Les attaques peuvent porter uniquement sur la machine (dérèglement dangereux de la machine pouvant mener à sa destruction si une rançon n'est pas payée) mais aussi sur le réseau domestique (la machine à café sert de porte d'entrée au réseau).

La prise de contrôle d'objets connectés peu sécurisés (car peu chers et peu dangereux) permet à des hackers de lancer depuis ces milliers d'objets contrôlés des attaques DDoS (Distributed Denial of Service) en saturant un serveur de requêtes. Par exemple, le 21 octobre 2016, le malware Mirai, après avoir infecté des dizaines de millions d'objets connectés (notamment des caméras de surveillance de bébés, ce qui a participé à son succès médiatique) a rendu indisponible le gestionnaire de noms de domaine américain Dyn, ce qui a mis hors connexion les sites de clients importants comme Twitter, Spotify, et PayPal.

Une fois les objets connectés dans les mains de particuliers, non enregistrés et non visés par les attaques, il est très difficile de faire procéder à leur mise à jour, ce qui contribue à expliquer que Mirai refait parfois parler de lui.

La partie 2 du dossier, consacrée aux objets connectés, présente les dispositifs de cybersécurité sur les principaux réseaux IoT et une application pratique Bluetooth vulnérable et sa sécurisation.

3.3 - Systèmes embarqués – Prise de contrôle d'un véhicule automobile

En 2015, les chercheurs américains Chris Valasek et Charlie Miller ont révélé des failles de sécurité dans des applications embarquées dans un véhicule Jeep [2]. Ces failles concernaient un réseau Wifi disponible en option à l'intérieur du véhicule mais pouvaient aussi être exploitées directement par internet car les véhicules concernés étaient connectés au réseau cellulaire de l'entreprise Sprint.

Il fut alors possible de réaliser des commandes normalement gérées via le tableau de bord : augmentation du volume sonore de la radio, activation de la ventilation, etc. Ce sont de nombreuses commandes qui peuvent surprendre le conducteur et donc mener à des comportements dangereux sur la route.

Figure 4 : Fiche 23 [1] Prise de contrôle d'un véhicule automobile

La vulnérabilité la plus dangereuse était la possibilité de modifier le firmware du contrôleur V850 qui, a priori, ne pouvait que lire des informations sur le bus CAN mais ne pouvait pas y envoyer des commandes. Une fois le firmware modifié, les chercheurs ont réussi à envoyer des commandes à distance pour par exemple bloquer le système de freinage ou faire changer le véhicule de direction.

La fiche à propos de cet incident propose les contre-mesures suivantes :

- Utilisation d'un algorithme assurant une génération de clé non prédictible ;
- Mise en place d'un mécanisme empêchant la mise à jour du Firmware du contrôleur V850 par un code non signé ;
- Filtrage des communications entre le contrôleur V850 et le bus CAN

Cette attaque a fait l'effet d'un électrochoc dans l'industrie automobile en 2015 et a abouti au rappel de plus d'un million de véhicules. Ici, l'attaque est d'une complexité très élevée et a pris plusieurs années pour être conçue par les deux chercheurs. Elle souligne toutefois qu'avec l'augmentation de la connectivité, la cybersécurité est devenue un axe de travail important pour l'industrie automobile.

Deux autres fiches (exemples : 26, 27) montrent que les voitures modernes avec des logiciels de plus en plus complexes et des connectivités (wifi, 4G/5G, Bluetooth) importantes, notamment des interactions avec des applications smartphone, ont une surface vulnérable aux attaques plus importantes et deviennent des cibles de choix pour les hackers. Comme dans l'industrie, les mises à jour régulières et le cloisonnement du réseau multimédia et du réseau de terrain sont les premiers éléments mis en avant.

Enfin, la fiche 39 présente un exemple de leurre d'un récepteur GPS, menace prise très au sérieux, en particulier par les équipes travaillant sur les véhicules autonomes.

La partie 3 de ce dossier présente des vulnérabilités dans l'automobile et les travaux actuels des industriels pour renforcer la cybersécurité dans l'automobile.

4 - Plan du dossier

Après avoir défini la cybersécurité et le cadre de ce dossier, ces quelques exemples ont permis de souligner les problématiques de cybersécurité pour les systèmes industriels, dont une partie concernent des aspects réseaux : authentification, confidentialité, intégrité, non répudiation, disponibilité. C'est l'objet de la ressource « Fondamentaux de la sécurité réseau » [6], ce qui amène à introduire le plan prévu pour ce dossier :

Introduction à la cybersécurité des systèmes industriels

Fondamentaux de la sécurité réseau [6]

Cybersécurité des systèmes automatisés industriels

Cybersécurité des systèmes automatisés industriels [7]

Mise en œuvre du protocole sécurisé OPC-UA (à paraitre)

La Cybersécurité chez Eiffage Energie Systèmes [8]

Cybersécurité des objets connectés

Sécurité du protocole Bluetooth Low Energy (à paraitre)

Création d'une application Bluetooth Low Energy sur STM32WB (à paraitre)

Analyse de la sécurité d'une application Bluetooth Low Energy (à paraitre)

Sécurité du protocole LoraWan (à paraitre)

Sécurité du protocole Zigbee (à paraitre)

Wattsense - Siemens, une entreprise pour une GTB sécurisée [9]

Cybersécurité des systèmes embarqués (automobile)

(à paraitre)

Cybersécurité matérielle / les attaques par canaux auxiliaires

Mise en œuvre de la carte ChipWhisperer (à paraitre)

Références

[1]: *Fiches incidents cyber SI industriels – Fiche 22*, Clusif, 2022
https://clusif.fr/publications/fiches-incidents-cyber-si-industriels/
[2]: *Hackers Remotely Kill a Jeep on the Highway – With Me in It*, Andy Greenberg, WIRED, 2015
https://www.wired.com/2015/07/hackers-remotely-kill-jeep-highway/
[3]: *The Fresh Smell of ransomed coffee*, Martin Hron, Avast, 2020

https://decoded.avast.io/martinhron/the-fresh-smell-of-ransomed-coffee/

[4]: Essonne : un centre hospitalier visé par une cyberattaque, une rançon de 10 millions de dollars exigée, Le Figaro, 2022

https://www.lefigaro.fr/secteur/high-tech/essonne-un-centre-hospitalier-vise-par-une-cyberattaque-une-rancon-de-10-millions-de-dollars-exigee-20220822

[5]: 10 choses à savoir sur les attaques DDoS massives contre Dyn, Le Monde informatique, 25 Octobre 2016, https://www.lemondeinformatique.fr/actualites/lire-10-choses-a-savoir-sur-les-attaques-ddos-massives-contre-dyn-66325.html

[6]: Fondamentaux de la sécurité réseau, M. Sechehaye, A. Juton, février 2024, https://eduscol.education.fr/sti/si-ens-paris-saclay/ressources_pedagogiques/fondamentaux-dela-securite-reseau

[7]: Cybersécurité des systèmes automatisés industriels, A. Juton, février 2024, https://eduscol.education.fr/sti/si-ens-paris-saclay/ressources_pedagogiques/cybersecurite-des-systemes-automatises-industriels

[8]: La Cybersécurité chez Eiffage Energie Systèmes, J. Zindy, F. Le Gall, février 2024, https://eduscol.education.fr/sti/si-ens-paris-saclay/ressources_pedagogiques/cybersecurite-chez-eiffage-energie-systemes

[9]: Wattsense - Siemens, une entreprise pour une GTB sécurisée, M. Zenadi, M. Sauvergeat, février 2024, https://eduscol.education.fr/sti/si-ens-paris-saclay/ressources_pedagogiques/wattsense-siemens-une-entreprise-pour-une-gbt-securisee

Ressource publiée sur Culture Sciences de l'Ingénieur : https://eduscol.education.fr/sti/si-ens-paris-saclay

Fondamentaux de la sécurité réseau

La Revue 3EI
Culture Sciences de l'Ingénieur

Édité le 16/02/2024

école normale supérieure paris-saclay

Maxime SECHEHAYE[1] - Anthony JUTON[2] - Magali SAUVERGEAT[3]

[1] *ENS Paris-Saclay – DER Nikola Tesla*
[2] *Professeur agrégé à l'ENS Paris-Saclay – DER Nikola Tesla*
[3] *Enseignante BTS CIEL, Arpajon*

Cette ressource fait partie du N°111 de La Revue 3EI de janvier 2024.

Si les attaques les plus nombreuses viennent de l'intérieur après une arrivée par mail (virus ou cheval de Troie téléchargé par un membre du personnel), nombre d'attaques parmi les plus spectaculaires exploitent les vulnérabilités des communications réseau. Avant d'aborder les différents champs d'application de la cybersécurité des systèmes industriels, cette ressource rappelle les principes fondamentaux de la sécurité réseau : confidentialité, intégrité, disponibilité, authentification et non-répudiation. Elle détaille ensuite le protocole TLS, très populaire, qui illustre trois de ces principes. La ressource se prolonge par une liste de vidéos d'illustration ou d'approfondissement, en français et en anglais, pouvant servir de support pour des séquences de classe inversée ou un co-enseignement anglais-GEII.

1 - Principes fondamentaux

Cette première partie présente les principes fondamentaux (confidentialité, intégrité, disponibilité, authentification et non-répudiation) de la cybersécurité qui, s'ils sont correctement implémentés, garantissent la protection via le réseau, excepté en cas d'exploitation de failles non documentées, ce qui est hors du champ de ce dossier.

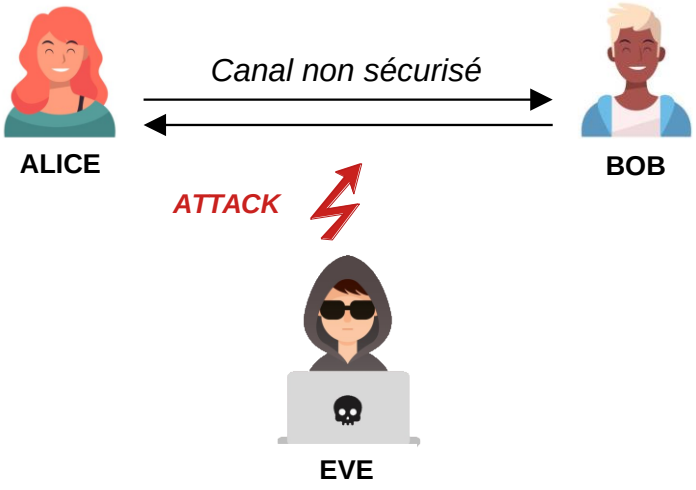


Figure 1 : Schéma simplifié d'une attaque

On imagine qu'Alice et Bob veulent échanger et que Eve souhaite nuire à leur relation. Alice et Bob peuvent être par exemple :

- Des personnes,
- Un terminal de paiement et un serveur de banque,
- Un automate industriel et sa supervision,
- Des calculateurs embarqués automobiles,
- Un smartphone et une serrure connectée.

Les communications sans fil peuvent être interceptées facilement et les communications sur Internet utilisent des chemins non connus et passent par des routeurs potentiellement écoutés. Le canal est donc souvent non sécurisé.

La **confidentialité** est l'impossibilité pour Eve de comprendre les messages circulant sur le canal, l'**intégrité** est la non-modification par Eve des messages (ou la détection d'une modification), l'**authentification** assure Bob que le message vient bien d'Alice et inversement, la **non répudiation** empêche Alice de nier avoir envoyé tel ou tel message et la **disponibilité** est la possibilité pour Bob et Alice de continuer à échanger.

1.1 - Confidentialité

La confidentialité permet de s'assurer que seules les personnes autorisées peuvent avoir accès à l'information transmise. Elle permet au concepteur d'un système d'information d'empêcher tout attaquant de récolter des informations sensibles en faisant de l'écoute clandestine (on parle d'*eavesdropping* en anglais).

La technique la plus courante pour garantir la confidentialité des échanges est l'utilisation de clés pour le **chiffrement** des données. Le chiffrement de données consiste en leur transformation en d'autres données, pas forcément de même longueur, grâce à un algorithme de chiffrement qui va utiliser une **clé** comme paramètre. Le terme clé est assez explicite : L'émetteur doit avoir une clé pour enfermer le message dans le coffre qui permet son transport et le destinataire doit aussi avoir une clé (la même ou une clé « associée ») pour ouvrir le coffre et lire le message.

Les algorithmes de chiffrement sont longs à développer et à vérifier. Les fuites survenant inévitablement, on ne peut envisager de redévelopper un algorithme de chiffrement à chaque fuite. C'est pourquoi le **Principe de Kerckhoffs** explique que l'algorithme doit être public et seules les clés doivent être secrètes. Celles-ci peuvent alors être renouvelées régulièrement.

On imagine ainsi qu'Alice souhaite envoyer un message à Bob. Ce message circulant sur un canal de communication accessible (ondes ou Internet par exemple), il peut être lu mais ne doit pas pouvoir être compris par Eve.

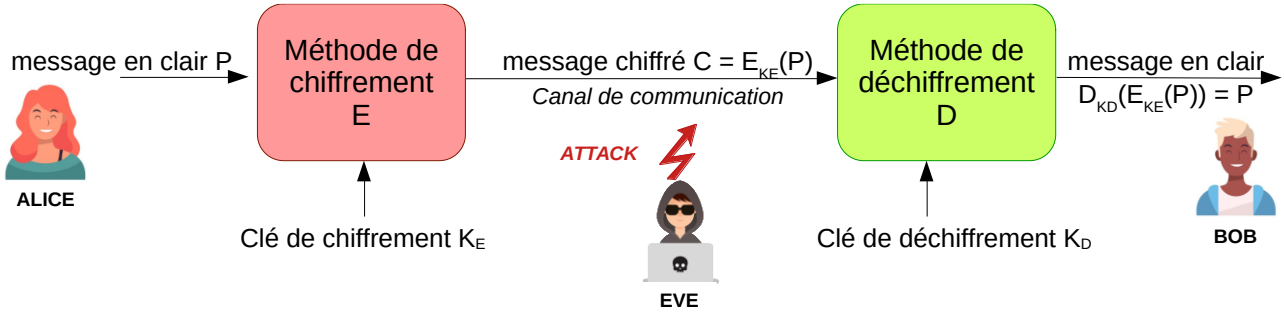


Figure 2 : Schéma de principe du chiffrement

Le développement des méthodes de chiffrement est la cryptographie, la recherche de méthode de déchiffrement sans connaître la clé est la cryptanalyse qui peut exploiter :

→ le texte chiffré seul,

→ un texte en clair connu (voir comme illustration le film *Imitation Game*),

→ un texte en clair choisi.

Chiffrement symétrique

Le chiffrement symétrique est le plus simple : l'émetteur et le récepteur **partagent le même secret**, la clé de chiffrement.

On peut imaginer par exemple deux individus qui discutent dans une langue qu'ils ont créée. La clé de chiffrement dans ce cas est le dictionnaire de cette langue. A priori, toute personne écoutant leur conversation et ne connaissant pas le dictionnaire de cette langue ne peut rien comprendre.

Un exemple classique de chiffrement symétrique est le **chiffrement par décalage** où l'on décale les lettres de l'alphabet d'un même nombre qui sera la clé de chiffrement. Par exemple, si l'on choisit un décalage de 5, `BONJOUR` sera chiffré en `GTSOTZW`. Pour déchiffrer le message, le récepteur doit seulement connaître la valeur du décalage et décaler les lettres dans le sens inverse. D'autres méthodes simples (OU exclusif avec un motif par exemple) consistent également à remplacer un caractère par un autre. On parle de **chiffrement par substitution**.

Il est aussi possible de modifier l'ordre des lettres, on parle alors de **chiffrement par permutation**.

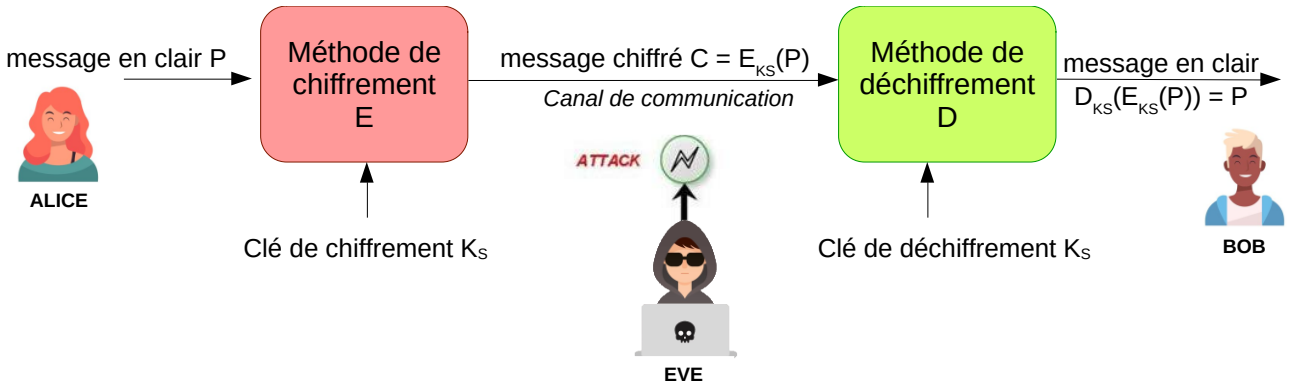


Figure 3 : Schéma de principe du chiffrement par clé symétrique

Un algorithme performant utilise à la fois permutation et substitution, pour éviter les méthodes de cryptanalyse statistique (pour du texte, on cherche le motif le plus courant et on lui associe la lettre la plus fréquente). Il utilise des clés suffisamment longues pour éviter les recherches de clés par force brute (tests successifs de toutes les clés possibles). L'algorithme de chiffrement symétrique le plus couramment utilisé (pour le wifi WPA2 ou pour HTTPS notamment) est l'algorithme **AES (Advanced Encryption Standard), avec des clés symétriques de 256 bits (soit 10^{77} clés possibles).**

> Les algorithmes de chiffrement symétriques sont plus rapides que les algorithmes de chiffrement asymétriques. Leur sécurité repose cependant sur le transfert de la clé secrète entre les deux appareils.

Chiffrement asymétrique

Le chiffrement asymétrique est plus complexe mais évite d'avoir à distribuer à l'avance une clé secrète aux participants : deux clés sont utilisées dans le processus de chiffrement : une **clé**

publique, connue de tous, et une **clé privée**, gardée secrète le plus souvent par le récepteur des données.

> La confidentialité des échanges est assurée lorsqu'on chiffre les données avec la clé publique : dans ce cas, seul le possesseur de la clé privée pourra déchiffrer le message.

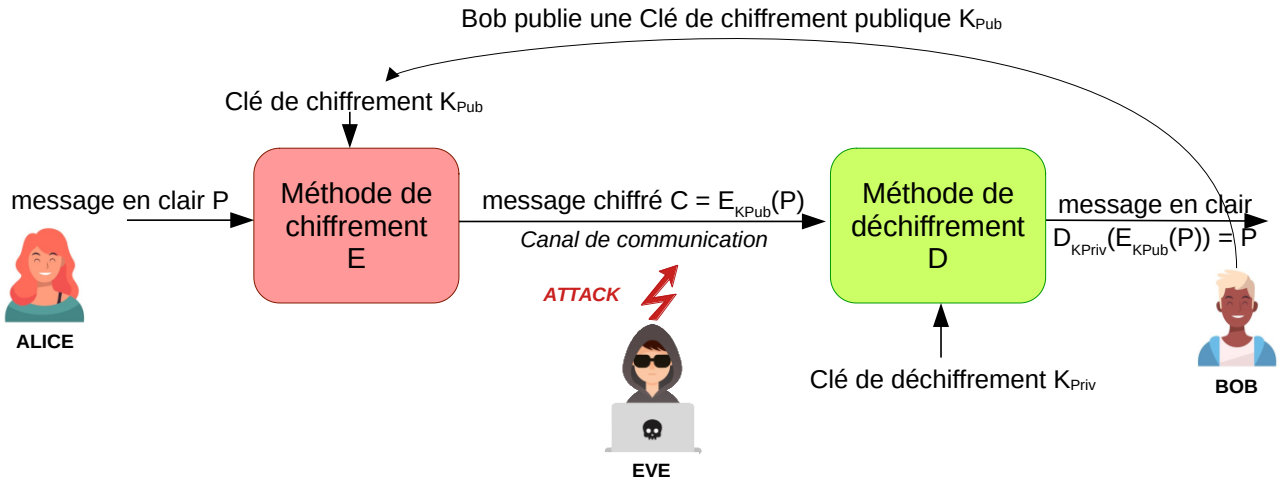


Figure 4 : Schéma de principe du chiffrement par clés asymétriques

Une analogie courante pour décrire le chiffrement asymétrique est celle de l'envoi d'un message de Bob vers Alice, en utilisant un coffre :

1° Alice dispose de cadenas identiques, s'ouvrant avec une clé et se refermant sans la clé (comme un cadenas classique). Elle envoie les cadenas ouverts à toute personne souhaitant lui envoyer un message secret, dont Bob. Elle garde la clé avec elle

2° Bob dispose d'un coffre et met son message à l'intérieur. Il prend le cadenas reçu d'Alice et le referme pour verrouiller le coffre.

3° Alice reçoie le coffre fermé avec son cadenas. Elle est la seule à pouvoir ouvrir ce cadenas.

L'algorithme de chiffrement asymétrique le plus souvent utilisé est l'algorithme **RSA**. Celui-ci est détaillé dans la ressource « Déchiffrez c'est gagné » [8], du même dossier.

Le protocole d'échange de clés **Diffie-Hellman** utilise aussi un algorithme asymétrique pour générer et partager de manière sécurisée une clé symétrique commune aux deux participants.

> Les algorithmes asymétriques sont les plus coûteux en ressources pour les calculs et en temps. Ils sont toutefois plus sécurisés que les algorithmes symétriques car aucune donnée secrète ne doit être partagée entre les deux protagonistes. On les utilise le plus souvent comme un moyen d'établir et de communiquer une clé symétrique à chaque nouvelle session (avec la méthode Diffie-Hellman par exemple).

On note que les clés publiques et privées de RSA sont réversibles, ce qui servira dans la partie suivante sur l'intégrité :

$D_{KPriv}(E_{KPub}(P)) = P$ mais aussi $D_{KPub}(E_{KPriv}(P)) = P$

1.2 - Intégrité

L'intégrité des données échangées désigne le fait qu'elles n'ont pas été modifiées durant le cycle de vie du message. Lors du téléchargement d'un logiciel par exemple, celui-ci n'est pas secret, il serait lourd de chiffrer l'ensemble du logiciel. Pour garantir l'intégrité du logiciel, l'éditeur met à disposition un code issu de l'ensemble des données de ce logiciel passé par une fonction de hachage cryptographique (**HMAC *Hash Message Authentication Code***).

Fonctions de hachage

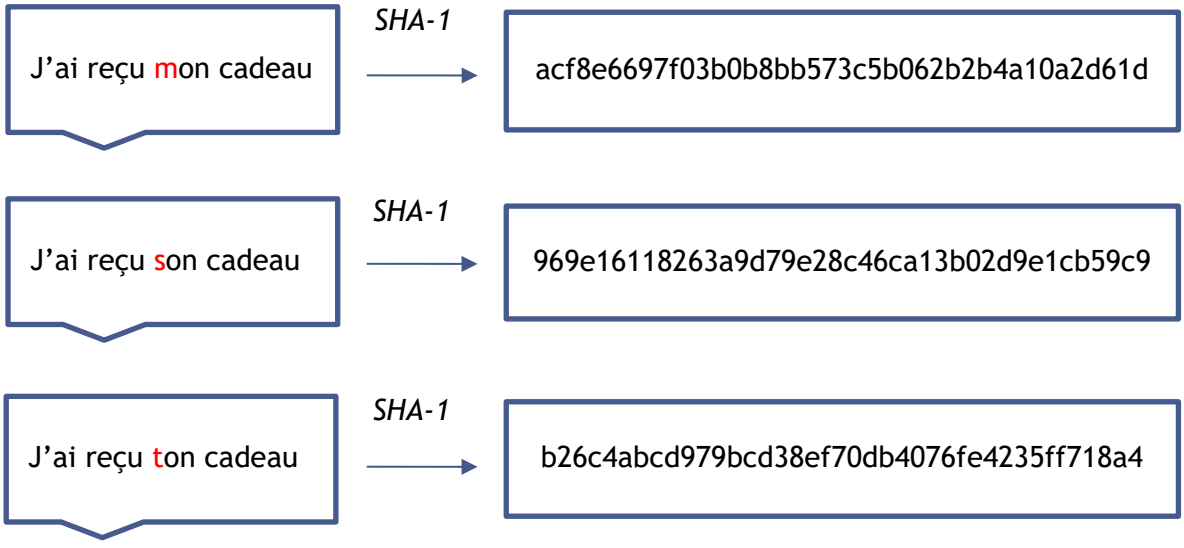
Les fonctions de hachage sont très souvent rencontrées en cybersécurité. Une fonction de hachage est une fonction qui, pour une donnée en entrée de longueur variable (qui peut être très longue, comme une vidéo ou un logiciel), retourne une valeur de longueur fixe (quelques octets), nommée condensat (ou en anglais *digest*).

Les codes de détection d'erreur CRC (Cyclic Redundancy Check) utilisés pour vérifier l'intégrité aux perturbations électromagnétiques des trames Ethernet ou CAN sont des fonctions de hachage, mais pas de cybersécurité.

Les fonctions de hachage classiques ne sont pas conçues pour répondre à toutes les problématiques de cybersécurité. C'est pourquoi on utilise des fonctions particulières, appelées **fonctions de hachages cryptographiques**. Les plus utilisées sont SHA-1 (*Secure Hash Algorithm*) et SHA-2, qui comprend notamment SHA-256.

Ces fonctions répondent aux exigences suivantes :

- Déterminisme : la même valeur de hachage (condensat) sera systématiquement générée pour un même message ;
- Rapidité de calcul ;
- **Non-réversibilité** : on ne peut pas reconstruire un message en ne connaissant que son condensat ;
- Résistance aux **collisions** : on ne peut pas trouver facilement un second message produisant le même condensat ;
- **Effet d'avalanche** : une légère modification du message entraîne une importante modification du condensat (voir exemple ci-dessous avec la fonction *SHA-1*).



Outre pour la signature numérique présentée ci-dessous, les fonctions de hachage sont utilisées pour le stockage des mots de passe. Les mots de passe ne sont normalement pas stockés en clair dans une application, pour éviter leur divulgation en cas de fuite. L'application se contente de stocker le condensat du mot de passe (mot de passe souvent concaténé avec une chaîne de caractère nommée « sel » pour améliorer la robustesse du stockage du condensat).

Signature numérique

On utilise ici la réversibilité des clés asymétriques. Si Alice chiffre un document avec sa clé privée K_{PRIVA}, qu'elle seule connaît, toute personne ayant la clé publique d'Alice peut déchiffrer le message et être sûr qu'il provient d'Alice.

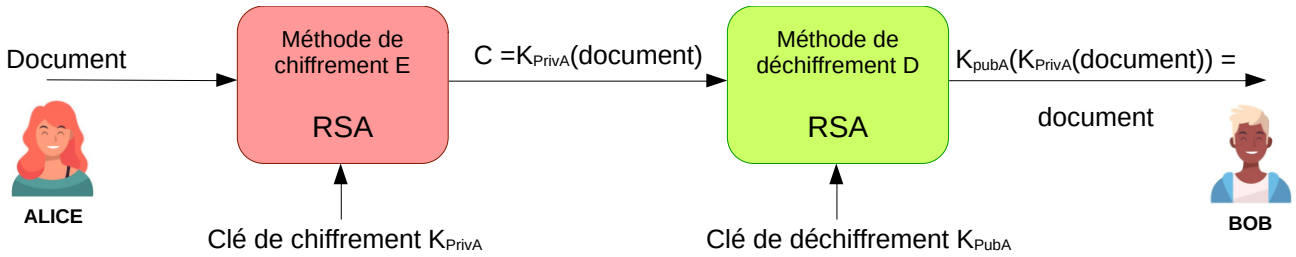


Figure 5 : Schéma de principe de la signature numérique

Cependant, chiffrer un gros fichier avec une clé asymétrique est gourmand en temps et en énergie. Il est plus économe de le hacher et de ne chiffrer avec la clé privée que le condensat.

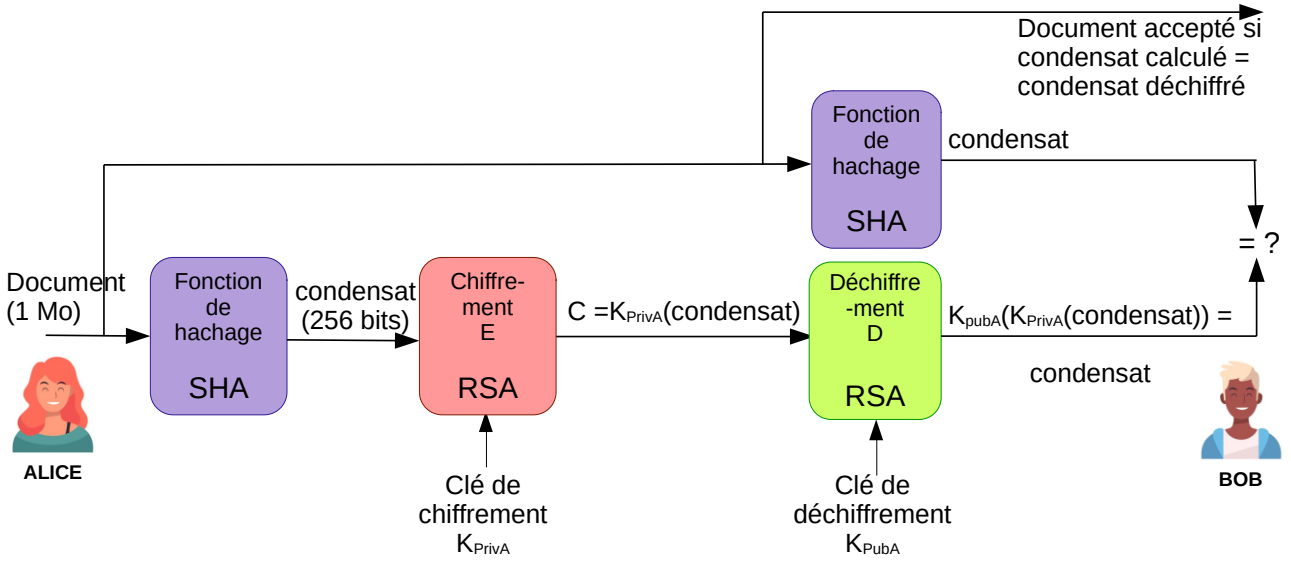


Figure 6 : Schéma de principe de la signature numérique avec fonction de hachage

Bob a ainsi l'assurance que le document reçu n'a pas été modifié entre son émission par Alice et sa réception.

1.3 - Non-répudiation

La non-répudiation assure qu'une action réalisée par une entité ne peut être niée ou ignorée. Elle apporte donc des preuves irréfutables qui pourront être utilisées a posteriori.

L'utilisation de **signatures numériques** peut assurer la non-répudiation. En chiffrant les messages avec sa clé privée (voir Figure 6), l'entité ne peut pas nier avoir signé le message car elle est la seule à posséder cette clé privée. De plus, certains mécanismes de signature numérique utilisent l'horodatage dans les données générant la signature numérique. Ainsi, même si la clé privée est volée par la suite, on peut s'assurer que la signature est bien valide car générée avant le vol.

1.4 - Authentification

L'authentification est le processus de vérification de l'identité d'une entité (utilisateur, système, appareil, etc.). Elle permet de s'assurer que, dans la communication, chacun est réellement qui il prétend être.

Les mécanismes d'authentification sont nombreux. On peut séparer d'une part les authentifications d'utilisateurs pré-enregistrés et les authentifications par **certificats**.

La première famille regroupe l'authentification par nom d'utilisateur et mot de passe ou l'authentification biométrique (empreinte digitale, reconnaissance faciale, etc.). Elle demande à l'un des partenaires de l'échange de connaître les identifiants de l'autre partenaire. Ce ne peut être une solution pour des échanges entre deux protagonistes ne se connaissant pas.

> Pour renforcer l'authentification, il est fréquent de mettre en place une authentification à plusieurs facteurs (deux voire trois). L'authentification à deux facteurs peut par exemple demander en plus d'un mot de passe la saisie d'une valeur envoyée par SMS.

Pour la seconde famille, on s'intéresse maintenant à deux acteurs ne se connaissant pas à l'avance, pour illustrer l'intérêt des certificats dans l'authentification.

La simple déclaration n'est évidemment pas suffisante, Eve pouvant déclarer être Alice.

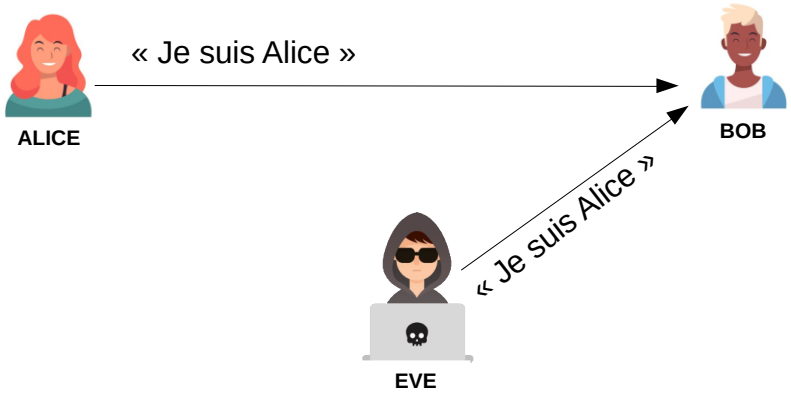


Figure 7 : Usurpation d'identité

Une déclaration chiffrée (ce qui demanderait d'avoir une clé symétrique) peut être rejouée par Eve pour se faire passer pour Alice.

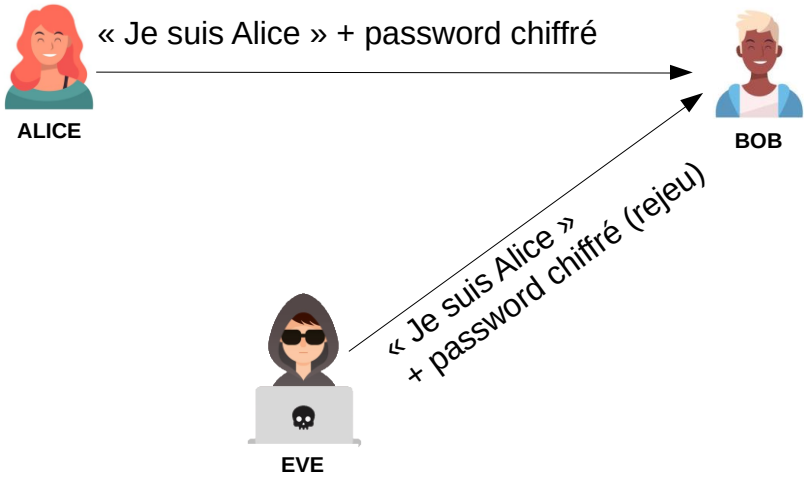


Figure 8 : Usurpation d'identité chiffrée par rejeu

Pour éviter le rejeu, Bob envoie un nombre R à usage unique. Comme les acteurs n'ont pas partagé au préalable une clé secrète, un mécanisme à clés asymétriques est utilisé. K_{PUBA} et K_{PRIVA} sont les clés publiques et privées d'Alice. Alice renvoie R chiffré avec sa clé privée et diffuse sa clé publique. On joue sur la réversibilité des clés publiques et privées. En connaissant la clé publique d'Alice, Bob (et toute autre personne voyant passer le message) peut déchiffrer le message $K_{PrivA}(R)$ et donc être sûr qu'il provient d'Alice.

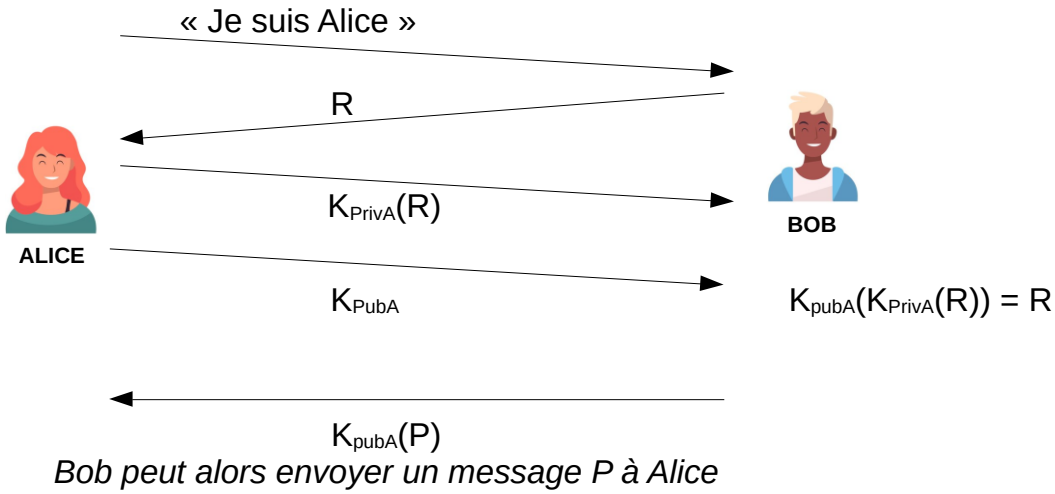


Figure 9 : Authentification par signature d'un nombre R

Reste à être sûr que K_{pubA} est bien la clé publique d'Alice. En effet, il est possible d'imaginer qu'Eve s'interpose entre Bob et Alice et intercepte les communications de l'un comme de l'autre. On appelle cette attaque « Man in the Middle ». Eve se fait passer pour Alice auprès de Bob et pour Bob auprès d'Alice. Elle peut alors lire et/ou modifier le message P envoyé par Bob à Alice.

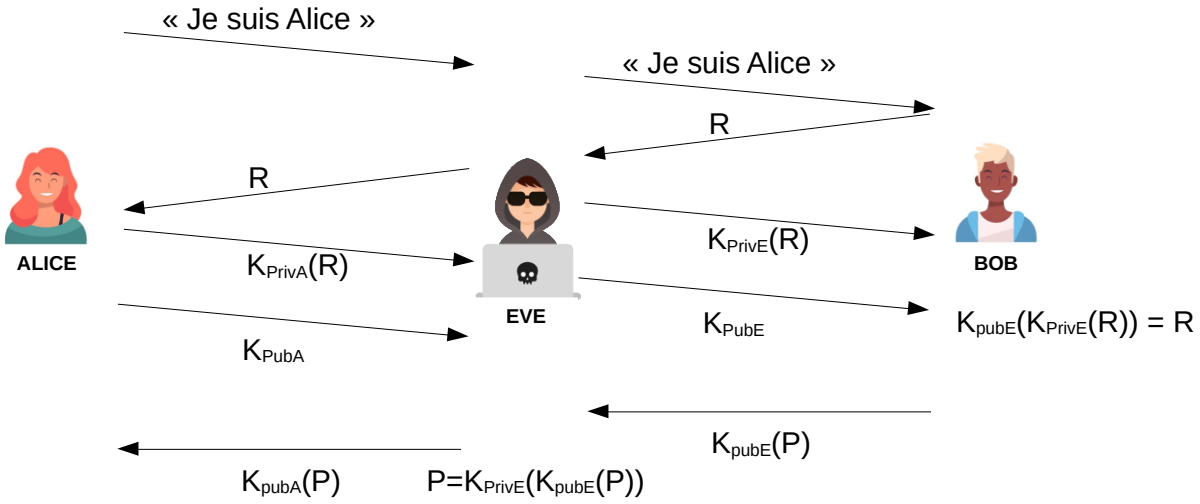


Figure 10 : Principe de l'attaque Woman in the middle

Il est donc nécessaire de certifier les clés publiques, ce que permettent les certificats. X509 est le protocole de gestion de certificats le plus populaire. Un tiers de confiance nommé autorité de certification signe numériquement (c'est-à-dire chiffre avec sa clé privée un condensat) le certificat contenant notamment la clé publique et le nom du protagoniste. La clé publique de cette autorité de certification est connue (les navigateurs web ont par exemple les clés publiques des principales autorités de certification) ou alors elle est diffusée elle-même avec un certificat d'une autorité de certification connue.

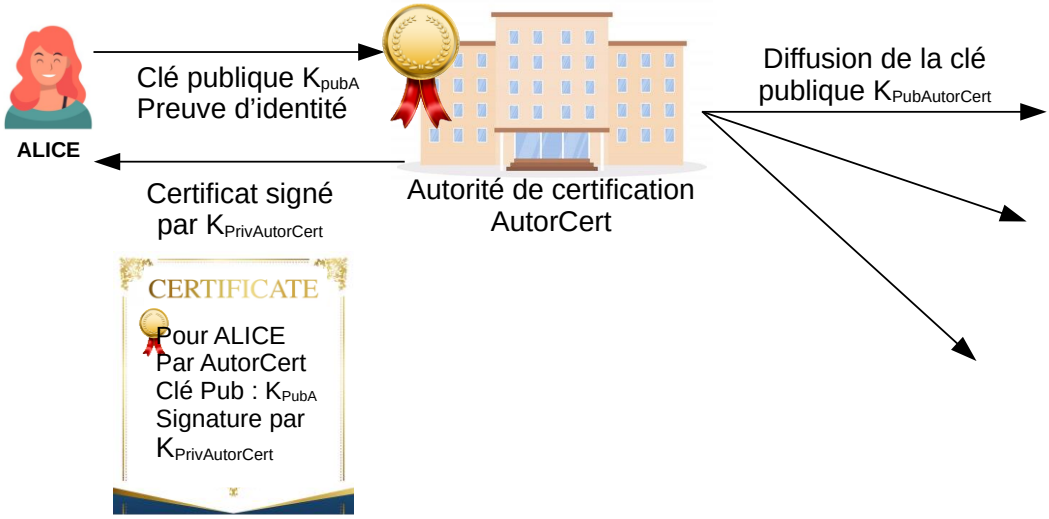


Figure 11 : Obtention d'un certificat auprès d'une autorité de certification

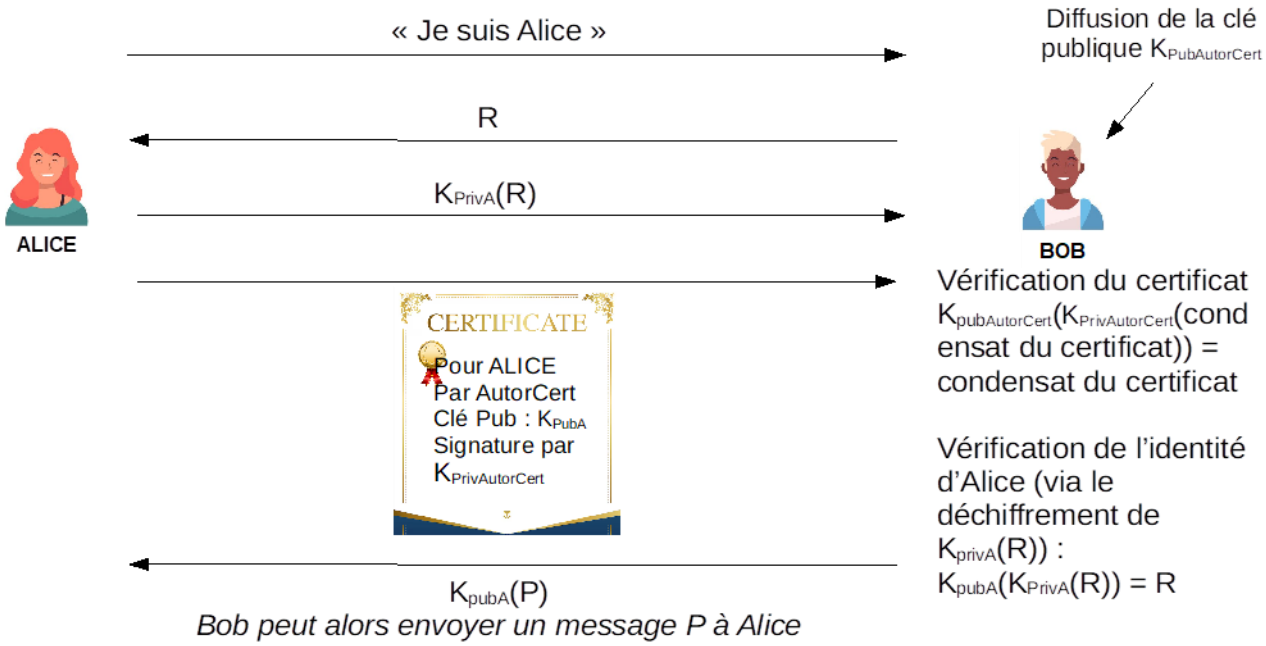


Figure 12 : Authentification avec utilisation d'un certificat

On résout ainsi le risque d'usurpation d'identité par Eve. On retrouve ces concepts en vidéo sur la chaîne d'Hervé Discours [7].

Figure 13 : Copie d'écran de la vidéo d'Hervé Discours sur les clés et certificats

1.5 - Disponibilité

La disponibilité désigne la capacité d'un service, d'un système à être accessible par un utilisateur autorisé quand il le souhaite.

> La disponibilité d'un système d'information est l'un des piliers de son bon fonctionnement. En effet, une attaque rendant impossible l'accès à un service en ligne fourni par une société nuit grandement à l'image de cette dernière.

Les attaques par déni de service distribué (DDoS, Distributed Denial of Service) sont particulièrement difficiles à bloquer car elles n'utilisent que des requêtes légales, issues d'appareils détournés, pour saturer un serveur. Les dernières attaques DDoS atteignent des débits de plusieurs $Tb.s^{-1}$ (2,3 $Tb.s^{-1}$ pour l'attaque d'AWS en février 2020)

Les requêtes étant légales, les mécanismes permettant de garantir la disponibilité sont complexes mais très étudiées au vu des enjeux financiers ou organisationnels liés à la coupure de service réseau :

- **Redondance** des systèmes ;
- Répartition de la charge de travail sur plusieurs serveurs pour éviter la surcharge ;
- Filtrage en amont des requêtes avec des pare-feux avancés.

2 - Exemple du protocole TLS

Le protocole TLS (*Transport Layer Security*, sécurité de la couche de transport) est un protocole cryptographique permettant de sécuriser les communications sur un réseau informatique (ici confidentialité via une clé symétrique AES-128, intégrité via une fonction de hachage SHA-256 et authentification via un certificat X509 et des échanges par clés asymétriques Diffie Hellman), au-dessus de la couche transport TCP et en-dessous de la couche application. Son utilisation la plus connue est le protocole HTTPS qui consiste en une version sécurisée du protocole HTTP. Il est aussi utilisé pour des protocoles industriels (OPC UA, IEC61850) et dans l'automobile (IEC15118).

> On utilise parfois le terme « SSL/TLS ». SSL désigne l'ancêtre du protocole TLS et toutes ses versions sont obsolètes depuis 2015. Il convient donc d'utiliser uniquement le terme TLS dans la conception d'un système d'information et de ne pas utiliser abusivement le terme « SSL/TLS ».

Cette partie explique les principes de fonctionnement de ce protocole afin d'illustrer l'implémentation des principes fondamentaux exposés précédemment. La version 1.3 étant un peu plus simple que la 1.2, elle sert de support à cette explication. Les principes sont les mêmes, avec un peu plus d'échanges pour la version 1.2, encore utilisée.

2.1 - Négociation de la connexion sécurisée

Dans un premier temps, le client demande au serveur de s'authentifier, les deux se mettent d'accord sur les paramètres de la connexion sécurisée et, via des clés asymétriques, échangent les clés symétriques de session.

Premier contact par le client - *Client Hello*

Dans ce premier message, le client communique entre autres au serveur les versions de TLS qu'il prend en charge, l'ensemble des algorithmes cryptographiques qu'il peut utiliser par ordre de préférence et un nombre aléatoire qui servira pour générer les clés de cette session. Pour reprendre une session interrompue, il est possible de fournir un identifiant de session, *Session ID*.

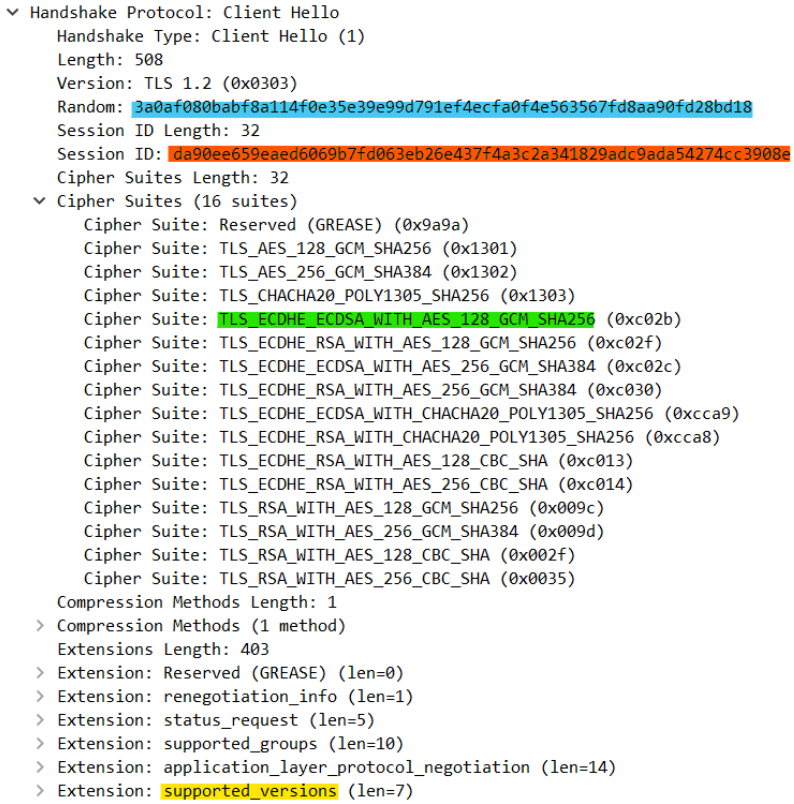


Figure 14 : Extrait d'un ClientHello sur Wireshark

Il est intéressant d'analyser les types d'algorithmes cryptographiques utilisés :

TLS_ECDHE_ECDSA_WITH_AES_128_GCM_SHA256

Elliptic Curve Diffie-Hellman Ephemeral

Algorithme asymétrique qui permet la génération et l'échange d'une clé symétrique utilisée pour l'établissement des clés de session

Elliptic Curve Digital Signature Algorithm

Algorithme asymétrique qui permet l'authentification, la non-répudiation et l'intégrité.

AES 128 bits Galois/Counter Mode

Algorithme de chiffrement symétrique qui permet la confidentialité des communications

Secure Hash Algorithm 256 bits

Fonction de hachage qui permet de générer un condensat pour vérifier l'intégrité des données échangées

Réponse du serveur - *Server Hello*

Le serveur répond alors au client en sélectionnant pour chaque paramètre une valeur parmi celles proposées par le client et qui convient à ses capacités.

Le serveur envoie également un nombre aléatoire qui servira pour générer les clés de cette session

Envoi et vérification du certificat du serveur

Le serveur fournit ensuite au client un certificat électronique, aussi appelé certificat numérique. Il va permettre au client de faire confiance à l'entité avec laquelle il communique. Il s'agit donc d'une étape essentielle pour garantir la sécurité de la connexion qui est sur le point d'être établie.

On peut considérer ce certificat comme la carte d'identité du serveur. On observe sur la Figure 15 le certificat numérique du site de l'ENS Paris-Saclay (accessible en cliquant sur le cadenas dans la barre d'adresse).

Figure 15 : lecture du certificat numérique du site de l'ENS Paris-Saclay sur un navigateur Google Chrome

Le certificat numérique est émis par une autorité de certification. Sur la Figure 15, le certificat a été fourni par l'organisation GEANT. Cette organisation fournit, moyennant rétribution, le certificat pour une période donnée, cette période est d'un an pour celui de la Figure 15.

> Il existe de nombreuses autorités de certifications qui opèrent à différentes échelles. Cela permet de ne compromettre que les serveurs situés dans une zone précise si une autorité de certification n'est plus sûre. Ainsi, les autorités de certification intermédiaires doivent aussi présenter un certificat numérique et le client doit vérifier tous les certificats numériques jusqu'à remonter à une autorité de certification de confiance.

Le client va donc devoir :

- Vérifier l'intégrité du message avec le condensat (condensant nommé « Empreinte SHA » sur la Figure 15) ;
- Vérifier l'authentification grâce à la signature du certificat par la clé privée de l'autorité de certification ;
- S'assurer qu'il peut faire confiance à l'autorité de certification en remontant la chaîne des certificats numériques, c'est-à-dire en vérifiant les certificats des autorités de certification jusqu'à arriver sur une autorité de certification de référence, connue du logiciel (le navigateur pour HTTPS par exemple) (voir Figure 16).

Figure 16 : Chaîne de vérification des certificats pour le site de l'ENS Paris-Saclay

Calcul du secret partagé

L'algorithme de Diffie Hellman, un peu différent de RSA, permet d'établir une clé commune à partir d'éléments de clés publiques envoyés par les 2 acteurs. Chacun peut alors déchiffrer à l'aide de sa clé privé les échanges.

Le client connaît désormais via le certificat les éléments de clés publiques du serveur et a transmis les siens lors du premier échange. Indépendamment, le client et le serveur vont donc pouvoir calculer à partir de ces éléments de clés publiques un nouveau secret commun, le *master secret*.

Calcul des clés de session

Le client peut alors communiquer de manière chiffrée avec le serveur, par le *master secret*. Il peut ainsi envoyer des clés symétriques AES au serveur, de manière chiffrée. 4 clés sont partagées : 1 pour les données client → serveur, 1 pour la vérification de l'intégrité client → serveur et la même chose dans le sens serveur → client.

Les différentes données utilisées pour établir la sécurité de l'application (valeurs aléatoires, *pre-master secret*, *master secret*, etc.) ne seront pas réutilisées lors de futures sessions entre les mêmes client et serveur. Cela participe au respect d'un principe appelé « confidentialité persistante » (*Forward Secrecy* en anglais) qui garantit que la découverte d'un secret privé ne compromet pas la confidentialité des échanges passées.

2.2 - Établissement de la connexion sécurisée

Une fois les clés de session établies, le client et le serveur vont chacun envoyer un message permettant de vérifier qu'ils possèdent bien les mêmes clés et signalant que les échanges futurs seront chiffrés.

Les échanges sont à présent sécurisés.

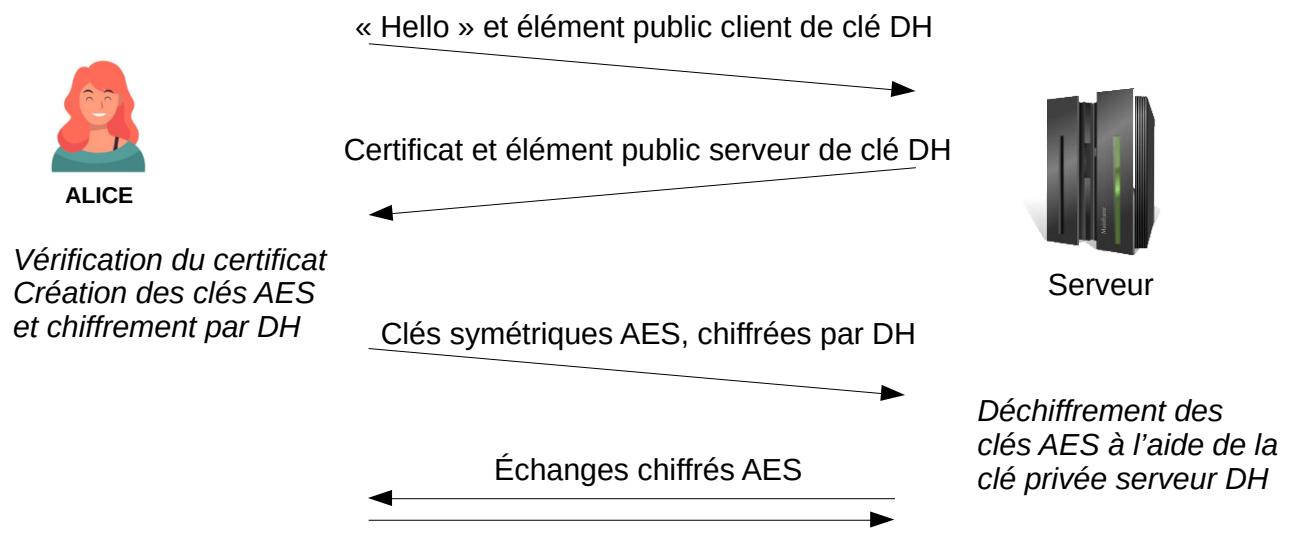


Figure 17 : Echanges TLS 1.3 menant à la mise en place d'une communication sécurisée

Les acquisitions ci-dessous présentent une acquisition wireshark de la mise en place d'un échange TLS 1.2 et TLS 1.3. Seuls les messages TLS sont affichés, les acquittements TCP notamment n'apparaissent pas, pour plus de lisibilité. Application Data signifie que les échanges chiffrés sont commencés.

Source	Destination	Protocol	Length	Info
192.168.1.41	129.175.212.146	TLSv1.2	744	Client Hello
129.175.212.146	192.168.1.41	TLSv1.2	1434	Server Hello
129.175.212.146	192.168.1.41	TLSv1.2	1430	Certificate
129.175.212.146	192.168.1.41	TLSv1.2	354	Server Key Exchange, Server Hello Done
192.168.1.41	129.175.212.146	TLSv1.2	192	Client Key Exchange, Change Cipher Spec, Encrypted Handshake Message
129.175.212.146	192.168.1.41	TLSv1.2	117	Change Cipher Spec, Encrypted Handshake Message
192.168.1.41	129.175.212.146	TLSv1.2	1930	Application Data
129.175.212.146	192.168.1.41	TLSv1.2	685	Application Data

192.168.1.41	138.231.176.51	TLSv1.3	721	Client Hello
138.231.176.51	192.168.1.41	TLSv1.3	4150	Server Hello, Change Cipher Spec, Application Data
138.231.176.51	192.168.1.41	TLSv1.3	2739	Application Data, Application Data, Application Data
192.168.1.41	138.231.176.51	TLSv1.3	134	Change Cipher Spec, Application Data
192.168.1.41	138.231.176.51	TLSv1.3	679	Application Data
138.231.176.51	192.168.1.41	TLSv1.3	325	Application Data

Figure 18: Comparaison entre l'établissement d'une connexion TLS 1.2 et celui d'une connexion TLS 1.3

3 - Conclusion

Cette ressource a expliqué les principes fondamentaux de la sécurité réseau que l'on retrouvera dans les autres ressources de ce dossier avant de voir comment ces principes sont mis en œuvre par le protocole TLS.

La ressource associée « Informatique débranchée : Déchiffrez c'est gagné » [8] donne un exemple de mise en pratique des algorithmes de chiffrement par les étudiants.

Les principes présentés ici sont issus de l'informatique et appliqués en automatisme industriel (OPC UA, IEC 61850, …) ou dans l'automobile (IEC 15118). Dans les objets connectés, aux capacités de calcul et à la consommation plus réduites, il est important d'appliquer les mêmes principes, avec des algorithmes peu gourmands. Cela passe notamment par l'intégration hardware de blocs de chiffrement, ce qui fige les algorithmes utilisés.

Par ailleurs, la puissance des ordinateurs augmentant, la taille des clés croît également pour résister aux attaques par force brute, passant par exemple de 1024 à 2048 bits pour RSA. Cela pose la question de la mise à jour logicielle, voire matérielle, des systèmes industriels à la longue durée de vie.

Enfin, l'arrivée des premiers ordinateurs quantiques, avec des capacités estimées très importantes pour craquer les clés asymétriques, amène dès aujourd'hui à réfléchir à de nouveaux algorithmes de chiffrement, plus robustes face à ces nouvelles machines. Que restera-t-il alors de la sécurité réseau des systèmes industriels installés en 2024 ?...

4 - Annexe : Quelques références supplémentaires

Cette annexe propose une sélection de vidéos YouTube, en français ou en anglais, permettant d'illustrer ou d'approfondir les différentes notions expliquées dans cette ressource.

Cette sélection restreinte parmi les nombreuses vidéos sur le sujet permet aussi de découvrir quelques chaînes YouTube dédiées à la sécurité réseau. Le lecteur intéressé par la pédagogie ou le niveau technique d'une des vidéos pourra explorer les autres vidéos de la chaîne.

Ces vidéos peuvent être utilisées notamment pour réaliser des séquences pédagogiques sous forme de classe inversée ou comme supports pour le co-enseignement en anglais en BTS CIEL.

4.1 - Exemples d'attaques liées à des problématiques de cybersécurité

La chaine **Cyber Vox** regroupe des vidéos, en français et en anglais, sur des attaques historiques telles que Wannacry, notPetya, stuxnet solarWinds : https://www.youtube.com/@CyberVox

4.2 - Principes fondamentaux : Vidéos de vulgarisation

4.2.1 - Confidentialité

Chiffrement symétrique / Chiffrement asymétrique

Cette vidéo de vulgarisation sur le chiffrement symétrique et asymétrique, issue de la chaîne en anglais Code.org, s'appuie sur des animations très pédagogiques :

The Internet : Encryption & Public Keys : https://www.youtube.com/watch?v=ZghMPWGXexs

La chaine Exo7Math apporte les notions mathématiques pour comprendre le protocole RSA du point de vue algorithmique :

Cryptographie - partie 5 : arithmétique pour RSA : https://youtu.be/M7vOxKVLsVY

Cryptographie - partie 6 : chiffrement RSA : https://www.youtube.com/watch?v=Xlal_d4zyfo

4.2.2 - Intégrité

Fonctions de hachage

Cette vidéo, issue de la chaîne **Bande de Codeurs**, détaille les fonctions de hachage et leurs différentes utilisations :

Comprendre les fonctions de HACHAGE : https://www.youtube.com/watch?v=OHXfKCH0b6s

4.2.3. - Disponibilité

Cette vidéo issue de la chaîne **@IBM technology** aborde la résistance aux ransomware :

Protecting Yourself from Ransomware : https://www.youtube.com/watch?v=eizn9TC68E8

4.2.4 - Authentification

Cette vidéo de la chaîne **kubucation,** en anglais, approfondit les notions de certificat et d'autorité de certification appliquées à HTTPS :

How does HTTPS work? What's a CA? What's a self-signed Certificate? :
https://www.youtube.com/watch?v=T4Df5_cojAs

4.2.5 - Non-répudiation

La chaîne **LeDroitpourMoi** propose une vidéo s'intéressant au côté juridique de la signature électronique :

Signature électronique : comment ça marche ? https://www.youtube.com/watch?v=GhTZUbp9M-8

4.3 - Sujets divers

La chaîne L'informateur aborde en français tous les sujets de la sécurité réseau, notamment les réseaux VPN.

Sécurité 13 : IPSec et comment fonctionne un VPN :
https://www.youtube.com/watch?v=V9bTy0gbXIQ&list=PLOapGKeH_KhFBC39ltMDhkEx1aI3hlwSK

La chaîne **@IBM technology** présente, en anglais, le concept Zero Trust :
Why Implement Zero Trust : https://www.youtube.com/watch?v=lT11tGaEC3s

Références

[1]: *Fiches incidents cyber SI industriels – Fiche 22*, Clusif, 2022
https://clusif.fr/publications/fiches-incidents-cyber-si-industriels/

[2]: *Hackers Remotely Kill a Jeep on the Highway – With Me in It*, Andy Greenberg, WIRED, 2015
https://www.wired.com/2015/07/hackers-remotely-kill-jeep-highway/

[3]: *The Fresh Smell of ransomed coffee*, Martin Hron, Avast, 2020
https://decoded.avast.io/martinhron/the-fresh-smell-of-ransomed-coffee/

[4]: *Essonne : un centre hospitalier visé par une cyberattaque, une rançon de 10 millions de dollars exigée*, Le Figaro, 2022
https://www.lefigaro.fr/secteur/high-tech/essonne-un-centre-hospitalier-vise-par-une-cyberattaque-une-rancon-de-10-millions-de-dollars-exigee-20220822

[5]: Images issues de Freepik : https://fr.freepik.com/vecteurs-libre/avatars-gens-heureux_7085154.htm et https://fr.freepik.com/vecteurs-libre/jeune-pirate-anonyme-au-design-plat_2753360.htm et https://fr.freepik.com/vecteurs-libre/ensemble-batiments-ville_8270967.htm

[6]: Computer Networks, Andrew Tanenbaum, Nick Feamster, David Wetherall, Pearson Education Limited

[7]: Cybersécurité, Clés publique/privée, Certificat X509, PKI, Chiffrement RSA, fichier.pem OPC-UA, Hervé Discours, https://www.youtube.com/watch?v=58FUQzWxs3Y

[8]: Informatique débranchée : Déchiffrez c'est gagné, A. Juton, février 2024, https://eduscol.education.fr/sti/si-ens-paris-saclay/ressources_pedagogiques/informatique-debranchee-dechiffrez-cest-gagne

[9]: Computer Networking, a top-down approach, Jim Kurose, Keith Ross, Pearson; 8th edition (September 13, 2020), http://gaia.cs.umass.edu/kurose_ross

Ressource publiée sur Culture Sciences de l'Ingénieur : https://eduscol.education.fr/sti/si-ens-paris-saclay

Informatique débranchée : Déchiffrez c'est gagné

La Revue 3EI
Culture Sciences de l'Ingénieur
école normale supérieure paris-saclay

Anthony JUTON[1]

Édité le 13/02/2024

[1] *Professeur agrégé à l'ENS Paris-Saclay – DER Nikola Tesla*

Cette ressource fait partie du N°111 de La Revue 3EI de janvier 2024.

La ressource « Fondamentaux de la sécurité réseau » [2] présente notamment les principes de chiffrement symétrique et asymétrique. Pour illustrer le cours, parfois un peu théorique, est née cette activité ludique d'informatique débranchée. L'objectif de l'activité est de sensibiliser les étudiants aux mécanismes de communication sur canal non sécurisés en pratiquant un peu de chiffrement par clés symétriques et asymétriques, dans l'esprit du protocole TLS.

L'activité est présentée comme elle a eu lieu en décembre 2023, avec 20 étudiants de Master, pendant le cours de réseau. Des adaptations ou une assistance de l'enseignant sont à prévoir en fonction du niveau en mathématiques des étudiants. Une version 2 est en réflexion pour faire intervenir l'authentification par certificat.

1 - Le contexte et les règles du jeu

Trois équipes de 6/7 étudiants séparées en 2 chiffreurs, 2 déchiffreurs et 2 hackeurs (répartition variable suivant les équipes).

Les chiffreurs doivent faire deviner un mot de 8 lettres reçu dans une enveloppe aux déchiffreurs. Le seul moyen de communication, non sécurisé, est un tableau commun pour tous. Tout le monde joue en même temps. Les calculatrices sont autorisées.

Le premier mot déchiffré vaut 3 points, le deuxième vaut 2 points et le dernier vaut 1 point, qu'il soit déchiffré par l'équipe associée à ce mot ou par le hackeur d'une autre équipe (l'équipe du hackeur obtient alors le ou les points).

Le protocole proposé est inspiré de TLS (le déroulement précis est décrit en partie 4) :

1. Le chiffreur calcule et diffuse une clé publique RSA via le tableau.
2. Le déchiffreur reçoit la clé publique et l'utilise pour chiffrer une clé symétrique créée pour l'occasion (type AES, mais plus simple) et l'envoyer via le tableau.
3. Le chiffreur utilise alors la clé symétrique pour chiffrer le mot à faire deviner et envoie le mot chiffré à ses partenaires via le tableau.
4. Le déchiffreur déchiffre le mot et annonce la réponse.

Pendant ce temps, les hackeurs tentent de craquer les clés privées adverses pour obtenir la clé symétrique, déchiffrer le mot de l'équipe adverse avant elle et obtenir le point.

2 - Le protocole à clé symétrique

TLS utilise le protocole à clé symétrique AES-128 ou AES-256. AES combine des substitutions et des permutations. Pour faire simple et rapide à chiffrer, on utilise uniquement des substitutions, avec un Ou exclusif utilisant une clé symétrique 20 bits.

Les caractères, uniquement des lettres majuscules, sont codés sur 5 bits.

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26

Exemple : coder GEII avec la clé 0xabcde puis déchiffrer le message chiffré avec la même clé.

```
Message GEII      : 00111001010100101001
XOR clé 0xabcde   : 10101011110011011110
Message chiffré   : 10010010100111110111


Message chiffré   : 10010010100111110111
XOR clé 0xabcde   : 10101011110011011110
Message déchiffré : 00111001010100101001
                      G    E    I    I
```

Pour le jeu, on utilise une clé de 20 bits et un algorithme de chiffrement Ou exclusif. Le mot à chiffrer faisant 8 caractères, on concatène la clé avec elle-même pour chiffrer 40 bits.

3 - Le protocole à clés asymétriques RSA

Pour transmettre les clés symétriques, on utilise l'algorithme de chiffrement à clé publique RSA, également utilisé par TLS 1.2 (mais plus par TLS 1.3).

Algorithme à clé publique RSA (du nom de ses inventeurs Rivest, Shamir, Adleman)

- Choisir deux grands nombres premiers p et q (p et q sont normalement des nombres sur 1024 voire 2048 bits).
- Calculer n = p * q et z = (p-1)*(q-1)
- Choisir un nombre d (d<n) tel que d et z n'aient pas de facteur commun
- Trouver e tel que e * d = 1 mod z
- Former des blocs de k bits tels que 2^k<n

- Calculer pour chaque bloc $C=P^e \bmod n$
- Déchiffrer en calculant $P = C^d \bmod n$

{e,n} est la clé publique

{d,n} est la clé privée

C'est réversible : le contraire ({d,n} publique et {e,n} privée) marche aussi

Exemple, p = 3, q = 11, d=7. Chiffrer « ENS ».

Les caractères, uniquement des lettres majuscules, sont codés sur 5 bits.

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26

On trouve n = 33, z = 20

On choisit d = 7 (7<33 et 7 et 20 n'ont pas de facteur commun)

On trouve e = 3 (3*7 = 21 = 1 mod 20)

On choisit k = 5 ($2^5 = 32 < 33$)

Chaque caractère est un bloc. P = {E,N,S} = {5,14,19}.

On le chiffre avec la clé publique {e,n} = {3,33}

- E → $55^3 \bmod 33 = 26$
- N → $1414^3 \bmod 33 = 5$
- S → $1919^3 \bmod 33 = 28$

Le message transmis sur le canal est donc C = {26,5,28}

On déchiffre avec la clé privée {7, 33}

- $26^7 \bmod 33 = 5$→ E
- $5^7 \bmod 33 = 14$→ N
- $28^7 \bmod 33 = 19$→ S

Le message déchiffré est bien P = {E;N;S}

Pour le jeu, on limite à p<50 et q<50 et on impose k=5 (donc n > 32)

Le site Dcode [3] permet de calculer des clés publiques et privées,de chiffrer des messages et de craquer des petites clés. Il permet également de voir le temps nécessaire à un PC pour craquer la clé privée à partir de la clé publique. Il faut de très (très) grands nombres pour obtenir un temps significatif.

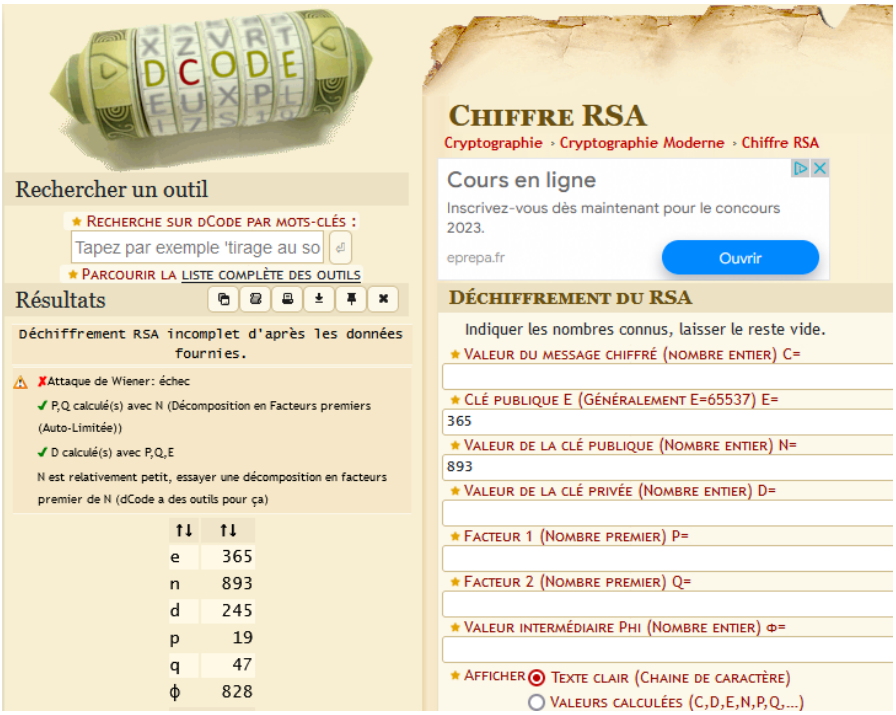


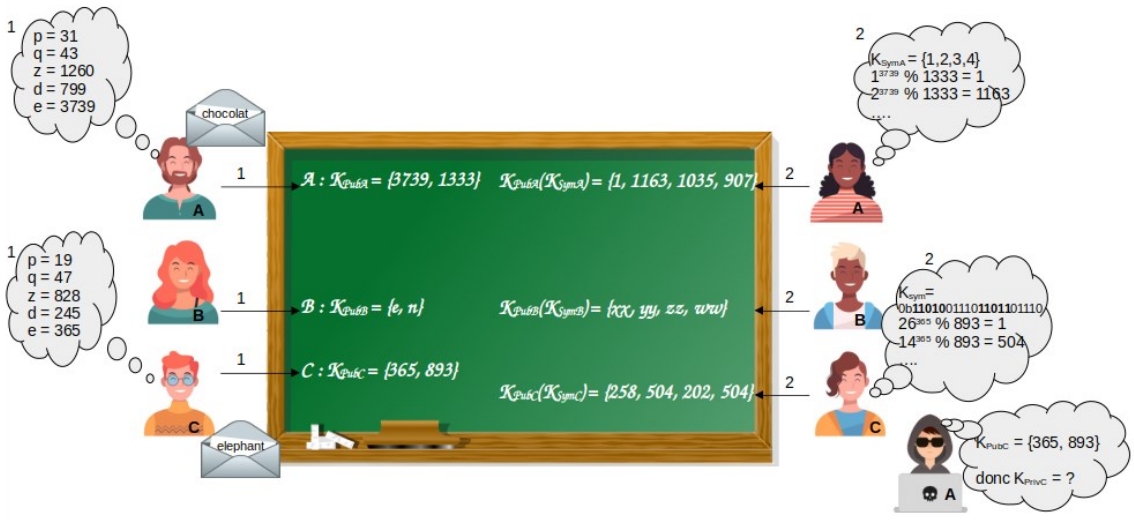
Figure 1 : Copie d'écran du site Dcode > RSA [3]

4 - Déroulement du jeu

Les règles sont expliquées, les étudiants sont répartis entre les chiffreurs d'un côté de la salle avec leur enveloppe et les autres de l'autre côté de la salle.

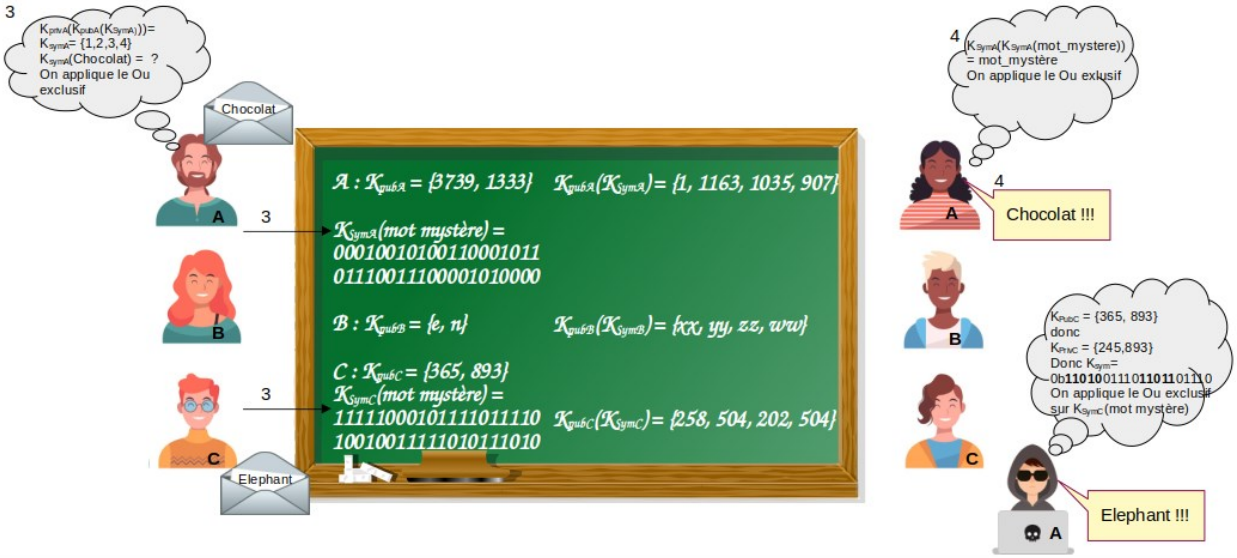
Le tableau est à la vue de tous.

Les chiffreurs doivent trouver un jeu de clés RSA. L'enseignant aide un peu si besoin. Une fois la clé publique obtenue, le chiffreur écrit celle-ci au tableau (1). Les déchiffreurs de son équipe codent alors, avec cette clé publique, la clé symétrique qu'ils ont choisie et écrivent la clé chiffrée au tableau (2). Pendant ce temps les hackeurs des autres équipes cherchent la clé privée pour pouvoir déchiffrer la clé symétrique.



Une fois la clé symétrique chiffrée affichée sur le tableau, le chiffreur la déchiffre avec sa clé privée et l'utilise pour chiffrer le mot mystère puis écrit le mot chiffré au tableau (3). Le déchiffreur utilise la clé symétrique pour déchiffrer le mot mystère et annonce le résultat (4).

Pendant ce temps les hackeurs tentent de trouver les clés privées des chiffrements RSA des autres équipes. S'ils vont suffisamment vite, ils peuvent déchiffrer le message contenant la clé symétrique et ainsi déchiffrer les mots mystère des autres équipes et obtenir leur point.



5 - Conclusion

Les exercices sur les clés symétriques et asymétriques faits pendant le cours, le jeu a duré une petite heure, explication comprise. L'émulation a bien fonctionné, la perspective de gagner quelques chocolats donnant un argument pour tenter de craquer le code de l'équipe adverse. L'équipe A a trouvé le mot A et craqué la clé de l'équipe B, sans avoir le temps de deviner le mot B avant l'équipe B. L'équipe C s'est trompé dans le transfert de sa clé publique et a fini par trouver le mot C avec l'aide de tous. L'enseignant a aidé les uns et les autres pour trouver les clés, chiffrer ou déchiffrer.

Le jeu a permis de mettre en évidence la possibilité de communiquer de manière confidentielle à travers un canal non sécurisé, sans avoir échangé des clés à l'avance. Le fait de s'écarter des ordinateurs a amené les étudiants à prendre la mesure des calculs nécessaires pour le chiffrement symétrique, le chiffrement asymétrique et le craquage d'une clé.

La limite à 50 pour les nombres premiers est sans doute un peu élevée. 40 serait plus raisonnable. Il faudrait peut-être donner une méthode pour trouver d et e.

L'an prochain, un essai d'introduction des certificats avec autorité de confiance est prévu.

Références

[1]: Computer Networks, Andrew Tanenbaum, Nick Feamster, David Wetherall, Pearson Education Limited

[2]: Fondamentaux de la sécurité réseau, M. Sechehaye, A. Juton, février 2024, https://eduscol.education.fr/sti/si-ens-paris-saclay/ressources_pedagogiques/fondamentaux-dela-securite-reseau

[3]: Dcode, Outil pour déchiffrer/chiffrer avec RSA https://www.dcode.fr/chiffre-rsa

[4]: Cybersécurité, Clés publique/privée, Certificat X509, PKI, Chiffrement RSA, fichier .pem OPC-UA, Hervé Discours, https://www.youtube.com/watch?v=58FUQzWxs3Y

Ressource publiée sur Culture Sciences de l'Ingénieur : https://eduscol.education.fr/sti/si-ens-paris-saclay

Cybersécurité des systèmes automatisés industriels

La Revue 3E.I
Culture Sciences de l'Ingénieur

Édité le 15/02/2024

école normale supérieure paris—saclay

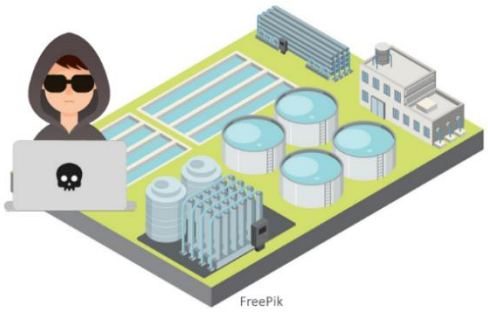
Anthony JUTON[1]

[1] *Professeur agrégé à l'ENS Paris-Saclay – DER Nikola Tesla*

Cette ressource fait partie du N°111 de La Revue 3EI de janvier 2024.

Cet article est une mise à jour l'article publié en juillet 2018 dans la revue 3EI [7]. Il introduit la partie du dossier cybersécurité des systèmes industriels consacrée aux systèmes automatisés. L'automatisme industriel étant essentiellement enseigné en BUT GEII, c'est le public visé par cet article, qui s'efforce de fournir des exemples et des applications pratiques.

Les risques liés à la cybersécurité pour les industries et les services sont réels comme le montre le blocage d'une usine Renault et d'hôpitaux anglais par le ransomware WannaCry en mai 2017 ([3] – Fiche 1). Consciente du risque cyber sur l'industrie et des implications pour le fonctionnement et la sécurité de l'Etat, la loi de programmation militaire de 2013 impose un renforcement de la sécurité informatique à des entreprises privées ou publiques considérées vitales pour la France, regroupées sous le terme Opérateurs d'Importance Vitale (OIV). On trouve notamment parmi les OIV des usines de traitement des eaux, des centrales de production d'énergie, des aéroports, des usines pharmaceutiques.



Ce renforcement concerne l'ensemble des équipements informatiques, ce qui comprend les systèmes gérés habituellement par les services informatiques (IT information technology) normalement sensibilisés à la cybersécurité mais aussi ceux gérés par les services automatisme (OT operational technology), qui doivent se former à ce nouveau risque.

La sécurisation d'une installation industrielle est donc le fruit d'une collaboration entre informaticiens et automaticiens. Cela passe par une implication des informaticiens dans la production et par une formation des automaticiens aux bases de la cybersécurité.

Prenant acte de ce contexte, la licence professionnelle SARII (Systèmes Automatisés Réseaux et Informatique Industrielle) de l'IUT de Cachan (aujourd'hui dissoute dans le BUT3 parcours AII) a créé en 2018 un module de cybersécurité des systèmes industriels pour compléter la formation en automatisme, réseaux et supervision de ses techniciens. Cet article repose essentiellement sur la démarche et les contenus de ce module prévu pour 4h de cours/TD et 12h de TP. L'ensemble s'appuie essentiellement sur les supports proposés par l'Agence Nationale de la Sécurité des

Systèmes d'Information (ANSSI) [1] et par le groupe de travail cybersécurité des systèmes industriels du Club de la Sécurité de l'Information Français (CLUSIF) [2], [3] et [4]. La consultation de ces 4 ressources permettra au lecteur intéressé d'approfondir le sujet.

Dans un premier temps l'article rappelle le risque cyber pour l'industrie, avant d'aborder la démarche proposée par l'ANSSI pour sécuriser un site. L'analyse de quelques incidents récents souligne les bonnes pratiques pour les éviter et l'article termine par une étude de cas d'usine pharmaceutique, support d'une exploitation possible en TD et TP.

1 - L'industrie est soumise à un risque cyber

1.1 - Les types d'attaque

Une attaque peut être ciblée contre l'entreprise (exemple de Stuxnet visant les usines d'enrichissement de l'uranium iraniennes [3] fiche 36 ou de BlackEnergy visant les postes électriques ukrainiens [3] fiche 4) ou non (exemple de WannaCry attaquant tous les systèmes Windows XP ou 7 non mis à jour [3] fiche 1).

L'attaque nécessite une intrusion dans le système (ou dans beaucoup de systèmes extérieurs pour les attaques par déni de service) et un mécanisme de sabotage.

1.1.1 – Solutions pour permettre l'intrusion dans le système

- Un **spyware** ou logiciel espion est un programme qui enregistre les frappes au clavier, webcam, microphone pour récupérer des informations (login et mot de passe notamment). Il peut s'installer lors d'une installation d'un logiciel depuis un site web malveillant, par l'introduction d'un média amovible infecté ou lors de l'ouverture d'un document contenant des macros.
- Le **phishing** ou hameçonnage est un mail utilisant un aspect officiel pour demander la saisie de données personnelles. Plus il est personnalisé (logo de l'entreprise, utilisation de détails concernant la cible), plus il est efficace.
- Un **ver** est un programme qui se reproduit sur plusieurs ordinateurs en utilisant le réseau informatique.

1.1.2 – Mécanisme permettant le sabotage ou la neutralisation du système industriel

- Un **cheval de Troie** est un programme qui permet de prendre à distance le contrôle de l'ordinateur cible. Si un PC de supervision ou de programmation des automates est infecté, le pirate peut modifier dangereusement le comportement du système.
- Un **ransonware** ou cryptovirus est un programme qui chiffre les fichiers et qui demande une rançon pour les déchiffrer. Une fois les fichiers chiffrés, le système est neutralisé.
- Un **virus** est un programme qui s'attache à un autre pour modifier son fonctionnement.
- Le **déni de service** est une attaque qui rend impossible l'utilisation d'un service, notamment via l'utilisation de botnet, réseaux de robots informatiques (souvent installés sur des systèmes informatiques peu protégés) qui vont ensemble saturer un serveur de requêtes.

1.2 - Les opérateurs d'importance vitale

La loi de programmation militaire de 2013 précise les obligations pour 12 secteurs d'opérateurs vitaux pour l'intégrité du territoire de ses habitants ou son économie :

- Activités civiles de l'Etat
- Activités judiciaires
- Activités militaires de l'Etat
- Alimentation
- Communications électroniques, audiovisuel et information
- Energie
- Espace et recherche
- Finances
- Gestion de l'eau
- Industrie
- Santé
- Transports.

Dans ces secteurs, plus de 200 services publics ou entreprises privées dont les activités sont indispensables au bon fonctionnement et à la survie de la Nation sont classés opérateurs d'importance vitale (OIV). La liste est confidentielle, on y trouve des acteurs industriels dans le traitement de l'eau, la production d'électricité, la fabrication de médicaments, la gestion technique des aéroports... La loi impose à ces OIV le renforcement de la sécurité des systèmes d'information critiques qu'ils exploitent, nommés « systèmes d'information d'importance vitale » (SIIV).

Les entreprises qui ne sont pas OIV sont également encouragées à prendre des mesures de cybersécurité, ne serait-ce que pour assurer leur survie économique en cas d'attaque.

1.3 - Les spécificités des systèmes industriels

Les systèmes informatiques industriels sont très proches des systèmes informatiques de gestion (utilisation de réseaux Ethernet/TCP/IP, utilisation de PC et serveurs pour la supervision, utilisation de bases de données SQL…) mais ont des spécificités qui les rendent vulnérables et difficiles d'accès aux informaticiens :

- Certains systèmes informatiques industriels (centrales nucléaires, usines en 5x8, aéroports...) doivent être disponibles sans interruptions, rendant difficiles les mises à jour, les tests de vulnérabilité, etc...
- Certains systèmes informatiques industriels mettent en jeu la vie des personnes (centrales nucléaires, machines médicales, usines de production de médicaments…) et pour cela reçoivent des habilitations/qualifications qui ne sont plus valables en cas de mise à jour majeure d'un équipement.
- Les équipements de contrôle-commande ont une durée de vie très longue (on trouve encore en fonctionnement dans les usines de très fiables automates Siemens S5 des années 80) et forment un parc souvent hétérogène (chaque machine peut avoir un modèle d'automate ou pire, une marque d'automate différent). Cela rend le suivi des vulnérabilités et des mises à jour plus fastidieux. Ces automates sont souvent inconnus des informaticiens en charge de la cybersécurité.
- Les productions alimentaires et pharmaceutiques notamment doivent garantir la traçabilité de leur production. Cela rend nécessaire les connexions entre les machines de terrain

(automates, superviseurs, ... regroupés sous le sigle OT Operational Technology) et les machines de l'administration (suivi de la qualité, traçabilité, ... regroupés sous le sigle IT Information Technology).

- Les réseaux de terrain traditionnels (profibus, CANopen, DeviceNet, Modbus RTU...) et certains protocoles TCP/IP largement utilisés en automatisme (Modbus TCP, BACnet/IP, …) ne sont pas sécurisés et pas sécurisables. Ces protocoles sont souvent inconnus des informaticiens en charge de la cybersécurité.
- La maîtrise exigée pour certaines tâches (par exemple la régulation de l'humidité dans des bâtiments d'architecture originale, la mécanique de précision ou l'intégration de robots industriels à des machines automatisées…) demande l'intervention de sous-traitants spécialisés (mécanique, automatisme, robotique, supervision…). A cela s'ajoute la volonté de réduire la masse salariale des entreprises, pratique très présente dans l'automobile. Il est bien sûr plus difficile de maîtriser l'intégrité et la formation en cybersécurité des sous-traitants que celles de ses salariés.

1.4 - Vulnérabilité des systèmes industriels

Pour souligner les vulnérabilités d'un système informatique industriel, voici trois cas représentatifs :

1.4.1 – Système non connecté à Internet

Le système informatique industriel basique comprend typiquement un ou plusieurs automates supervisés par un PC de supervision via un réseau Ethernet, pas forcément connecté à Internet. Un serveur de base de données pour l'archivage peut aussi être présent localement. L'automate contrôle divers équipements via des bus de terrain standard ou basés sur Ethernet.

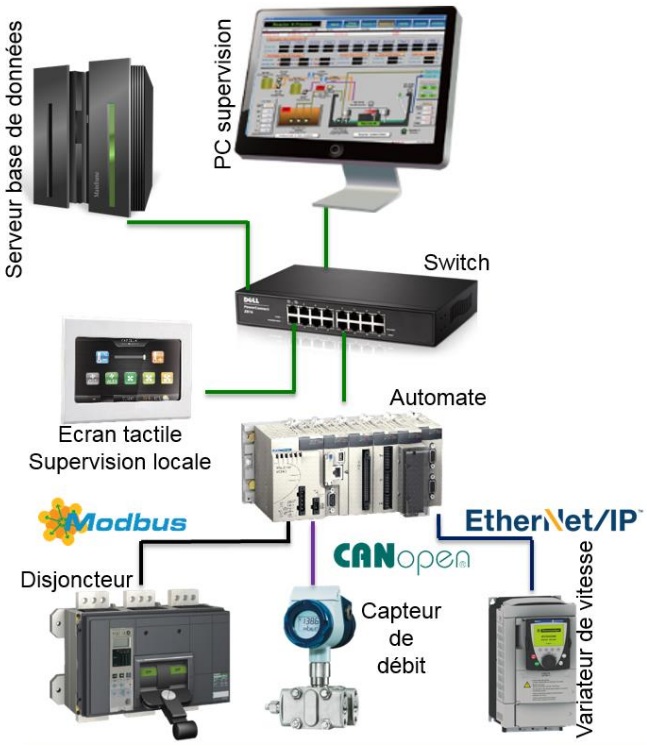


Figure 1 : Architecture type d'un système informatique industriel non connecté à Internet

Les problèmes peuvent survenir de :

- La modification du programme automate via un accès direct à l'automate ou la modification des consignes de supervision par du personnel non autorisé (sous-traitant par exemple),

- L'introduction d'un virus par une clé USB,
- L'introduction d'un virus par un PC maintenance (d'un sous-traitant ou d'un salarié) se connectant au réseau local, pour une mise à jour par exemple.

Le virus peut alors simplement neutraliser le PC de supervision (en chiffrant ses données comme wannacry [3]) ou plus rarement, car beaucoup plus complexe, neutraliser un équipement industriel (automate comme le malware Triton, [3] fiche 7 ou superviseur comme le malware Havex [3] fiche 5) ou plus complexe encore, modifier le programme automate ou les consignes envoyées par le PC de supervision (un seul exemple, Stuxnet [3]).

1.4.2 – Système connecté à Internet

Le réseau de l'atelier est connecté au réseau de l'administration via un routeur. Le réseau de l'administration est lui-même connecté à Internet via un routeur.

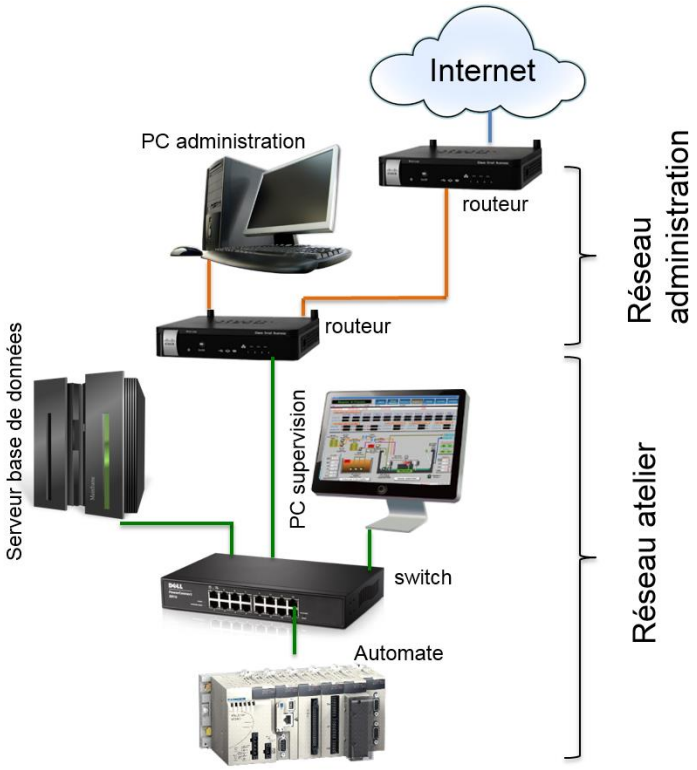


Figure 2 : Architecture type d'un système informatique industriel connecté à Internet

La connexion à Internet, outre une possibilité supplémentaire d'introduction de virus, amène le risque suivant : une personne malveillante peut s'introduire sur le réseau atelier via un accès ouvert (pour la télégestion par exemple) en dérobant des identifiant et mot de passe de connexion ou, plus complexe, via un cheval de Troie installé par un virus sur un PC de bureau par exemple. Cette personne peut alors neutraliser des équipements, les espionner ou en prendre le contrôle à distance.

Tout accès à distance à une installation fait courir le risque d'une prise de contrôle par une personne malveillante. (Exemple de l'attaque d'une station d'épuration, [3] fiche 16, ou [3] fiche 21, parmi beaucoup d'autres)

1.4.3 – Système informatique industriel distribué

Un système distribué désigne un système dont les organes de contrôle-commande (automates, variateurs, modules d'entrées/sorties déportés) ne sont pas localisés dans le même local. Les

grands sites de stockage d'hydrocarbure, les réseaux ferroviaires, les tunnels routiers et les bâtiments en général sont des systèmes distribués.

N'est représentée sur le schéma ci-dessous que la partie contrôle-commande. La supervision et la connexion éventuelle à Internet sont identiques aux architectures présentées ci-dessus.

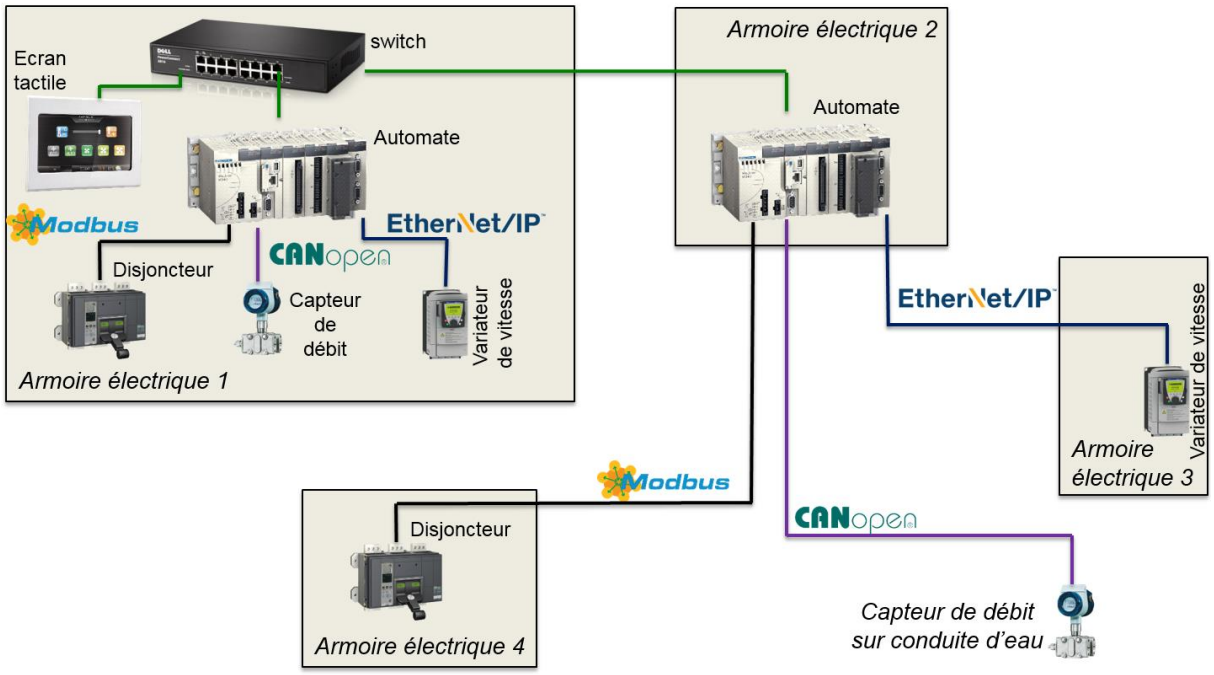


Figure 3 : Architecture type d'un système informatique industriel distribué (partie contrôle/commande uniquement)

Outre les risques précédemment cités, les réseaux de terrain standards, très répandus, très connus dans le monde industriels et très documentés, sur liaison RS485 (Modbus, Profibus, BACnet MSTP), sur bus CAN (CANopen, DeviceNet) ou spécifiques (LON, DALI, KNX…) ne sont pas sécurisés et pas sécurisables simplement (un standard sécurisé de KNX existe mais n'est pas compatible avec les équipements déjà installés). Le passage de ces bus dans des zones publiques ou faciles d'accès fait courir le risque d'une intrusion sur le réseau et de l'envoi sur ce réseau d'informations de capteurs fausses ou de commandes d'actionneurs dangereuses (exemple d'envoi de commande de déversement d'eaux usées en utilisant le réseau industriel local [3] fiche 18 ou exemple de la prise de contrôle d'éoliennes via un accès physique au réseau local [3] fiche 8)

Les réseaux de terrain modernes sur Ethernet (ProfiNet, EtherNet/IP, Ethercat) sont plus récents et leurs standards plus complexes tendent à prendre mieux en compte le risque cyber (authentification de la machine se connectant par exemple). Le protocole OPC-UA, sur Ethernet TCP/IP, utilisé pour la communication entre le superviseur et l'automate, ou entre automates, est même chiffré, ce qui lui donne une popularité certaine actuellement. La ressource [5] traite spécifiquement d'OPC-UA.

2 - Les mesures à appliquer

Une fois les risques présentés, le cours présente alors une version simplifiée de la méthode proposée par l'ANSSI [1] pour la sécurisation des systèmes informatiques industriels et la mise en place d'une politique de sécurité des systèmes d'information (PSSI).

« L'objectif de la Sécurité des Systèmes Informatiques (SSI) est d'étudier les vulnérabilités des systèmes (matériel, logiciel, procédures, aspects humains) afin de déployer des mesures pour les limiter et permettre d'assurer la continuité des fonctions métier à un niveau acceptable. ». Il s'agit

d'assurer la disponibilité, l'intégrité, la confidentialité et la traçabilité du système informatique industriel.

La cybersécurité doit être envisagée par les automaticiens comme la sûreté de fonctionnement des machines :

- On identifie les risques puis la probabilité et les conséquences du risque,
- On met en place des mesures proportionnées au risque (coût et contraintes sur les utilisateurs), sous peine de les voir rejetées.
- On ne peut raisonner en termes de retour sur investissement.

La mise en place de la sécurisation du système informatique industriel doit impliquer les automaticiens, bons connaisseurs de leurs équipements industriels et les responsables cybersécurité de l'informatique générale, formés en cybersécurité mais souvent hermétiques à l'automatisme.

La méthode suit 7 étapes, détaillées ensuite :

1. Sensibilisation des personnels
2. Cartographie des installations et analyse de risque
3. Prévention : concept de la défense en profondeur
4. Surveillance des installations et détection des incidents
5. Traitement des incidents, chaîne d'alerte
6. Veille sur les menaces et les vulnérabilités
7. Plans de reprise et de continuité d'activité

La méthode ne fait appel qu'à des compétences d'automaticien moderne (les compétences en réseaux sont évidemment importantes). Elle montre que la cybersécurité est surtout une question d'organisation et de temps alloué à cette organisation, d'où l'obligation d'un PSSI pour les OIV.

2.1 - Sensibilisation des personnels

La majorité des incidents est liée à l'imprudence des personnels de l'entreprise :

- Utilisation de clé USB,
- Logiciel « cheval de Troie » ou virus installés en ouvrant un fichier ou en installant un logiciel de provenance douteuse,
- Divulgation de ses identifiant/mot de passe en réponse à un mail de phishing,
- Mot de passe écrit sur un post-it ou stocké en clair sur une machine,
- Identifiant/mot de passe identique pour tous les techniciens (y compris ceux qui quittent l'entreprise…),
- Machines non protégées ou avec les identifiant/mot de passe par défaut, avec des mises à jour logicielles non effectuées.

Plus de 10 fiches « incident » du Clusif [3] ont pour origine un vol de mot de passe par une campagne de phishing, plusieurs autres ont un lien aussi avec des manquements à la sécurité des employés (non révocation des accès d'un employé licencié, installation de logiciel infecté...)

La sensibilisation de tous les personnels aux règles d'« hygiène informatique » contribue à réduire fortement les vulnérabilités et les opportunités d'attaques. La sensibilisation doit être régulière car les risques évoluent en permanence et les mauvaises pratiques reviennent.

2.2 - Cartographie des installations et analyse de risque

Comme pour la sécurité des machines, la seconde étape est un audit de l'installation :

- Quels sont les objectifs métier (production, distribution, protection des biens et des personnes…) et les services assurés ?
- Quels sont les impacts en cas d'interruption de service ? En cas de modification du comportement du système ?
- Quelles sont les fonctions indispensables à l'atteinte des objectifs, et en particulier :
  - leurs niveaux d'implication et de criticité dans la réalisation des services,
  - les systèmes qui les portent,
  - si ces systèmes sont centralisés, distribués, accessibles à distance, etc.

Cela amène donc à un inventaire des installations matérielles (référence de l'équipement, version, protections, accès), de l'architecture réseau et des communications entre les équipements internes et externes. Cela amène également à séparer les équipements critiques devant être très protégés des autres. Les plus critiques, systèmes d'information d'importance vitale (SIIV), doivent être déclarés à l'ANSSI.

L'ANSSI propose une méthode d'analyse du risque nommée EBIOS [1].

2.3 - Prévention : concept de la défense en profondeur

On entre ici dans la partie technique de la cybersécurité : la défense en profondeur consiste à protéger les installations en les entourant de plusieurs barrières de protection autonomes et successives, de sorte d'assurer la protection même en cas de compromission d'un équipement. Ces barrières peuvent être technologiques ou liées à des procédures organisationnelles ou humaines.

Première règle : tout est interdit par défaut. On autorise juste les accès nécessaires aux personnes concernées, on n'ouvre que les connexions UDP/TCP utiles, on n'installe que les logiciels indispensables.

Voici les protections à mettre en place :

- **Protection physique** : c'est la protection la plus simple, les équipements de contrôle-commande doivent être dans des armoires fermées à clé, le local de supervision ne doit être accessible qu'aux personnes autorisées. Siemens propose par exemple des verrous bloquant l'accès aux ports Ethernet des équipements (automates, switch...).

Figure 4 : Verrou Siemens pour prise RJ45 : IE RJ 45 Port Lock

- **Pare-feu** : sur les routeurs et sur les Pcs et automates, on bloque les ports UDP et/ou TCP non utilisés pour empêcher les requêtes indésirables d'arriver jusqu'à la machine. Les pare-feux avancés permettent de faire de l'inspection de paquets en profondeur (le pare-feu vérifie que le contenu d'un paquet arrivant sur le port 502, dédié à Modbus, est bien une requête Modbus par exemple).
- **Cloisonnement des réseaux**, notamment pour séparer le réseau industriel (OT) du réseau de l'administration (IT) : Les VLANs et les pare-feux des routeurs permettent de filtrer les échanges entre un sous-réseau et un autre. On veillera notamment à n'autoriser que les requêtes indispensables à entrer dans le sous-réseau atelier. Il est possible d'installer des **diodes réseau** (les informations ne passent physiquement que dans une direction, ce qui interdit les communications TCP) ou passerelles unidirectionnelles (un peu plus avancées, elles acceptent les établissements de connexion TCP et les acquittements, pour permettre un flux d'information TCP dans une seule direction).
  Il est possible également de mettre une **passerelle de rupture protocolaire**. Celle-ci, en passant le message d'un protocole de communication à l'autre, permet d'éviter l'exploitation de failles dans un des protocoles.
- **Protection antivirale**, les PCs (supervision, programmation) doivent avoir un antivirus à jour. Une procédure explicite de mise à jour des antivirus doit exister.
- **Durcissement des configurations** :

  Pour un PC de supervision :

  - Ne garder que les logiciels indispensables à la supervision (pas de logiciel de programmation des automates, pas de navigateur web, pas de logiciels de bureautique…) ;
  - Bloquer les médias amovibles (clés USB) sur les ports usb ;
  - Mettre un mot de passe sur le Bios pour notamment empêcher un démarrage sur un autre support que le disque de la machine ;
  - Supprimer ou désactiver les fonctions non utilisées mais activées par défaut.
  - Mettre à jour le système d'exploitation et le logiciel de supervision. Il peut être nécessaire pour cela d'avoir une installation miroir réduite pour tester les mises à jour avant leur mise en production ;
  - Distinguer clairement les profils (OS et supervision) utilisateur et administrateur. Le PC de supervision est sur un profil utilisateur (pas de droit pour installer des logiciels) et chacun dispose sur la supervision d'un profil adapté à ses besoins. Chaque personne a des identifiant/mot de passe uniques ;
  - Choisir un logiciel de supervision offrant les meilleures caractéristiques pour répondre aux exigences de sécurité et mettre en place ces fonctionnalités : mécanismes d'authentification, ségrégation des droits (la personne chargée de la maintenance peut acquitter les alarmes mais ne peut modifier les consignes du système par exemple) ;
  - Les logiciels de programmation des automates sont sur des PCs éteints et stockés sous clé ;

  Sur les automates :

  - Changer les configurations par défaut (mot de passe par exemple),
  - Mettre à jour régulièrement le firmware de l'automate (disponible sur le site du fabricant),
  - Supprimer ou désactiver les fonctions non utilisées mais activées par défaut (serveur web, utile pour la configuration mais pas pour l'utilisation, serveur FTP...).

Dans ce cadre de défense en profondeur, des procédures accompagnent ces défenses techniques, notamment concernant les interventions des sous-traitants qui doivent être planifiées précisément (mots de passe, accès, utilisation de ses propres outils ou non, échanges de matériels, qualifications…). L'ANSSI publie un guide de l'externalisation pour accompagner les entreprises dans la mise en place des procédures d'intervention des sous-traitants.

2.4 - Surveillance des installations et détection des incidents

Les équipements réseaux proposent des journaux et pour les plus avancés des alarmes permettant d'indiquer un trafic anormal. Surveiller le réseau en lisant ces journaux système et en configurant ces alarmes mais aussi en formant le personnel à détecter et signaler des comportements suspects de leur machine n'empêchera pas un incident mais permettra de le détecter au plus tôt et d'en limiter autant que possible les effets.

Plus un incident sera détecté tôt, plus il sera possible de mettre en place des mesures pour en réduire et confiner les effets comme par exemple :

- Isoler physiquement les installations en cas d'attaque virale pour limiter les risques de propagation (on déconnecte du réseau les machines),
- Arrêter une installation avant sa dégradation si des données de configuration ne sont plus intègres.

2.5 - Traitement des incidents, chaîne d'alerte

Le dispositif de détection des incidents est associé à une organisation et des procédures pour traiter les incidents :

- Que faire lors de la détection d'un incident ?
- Qui alerter ?
- Quelles sont les premières mesures à appliquer ?

La gestion des incidents doit également intégrer une phase d'analyse post incident qui permettra d'améliorer l'efficacité des mesures déployées initialement.

2.6 - Veille sur les menaces et les vulnérabilités

La sécurité informatique est une action continue nécessitant des efforts permanents (on revient à l'importance du PSSI). La ou les personnes en charge de la cybersécurité du système industriel doivent mettre en place une organisation pour :

- Se tenir informés de l'évolution des menaces, des vulnérabilités, sur le site Internet de l'ANSSI ([1] et [6]) et sur celui des équipementiers qui doivent indiquer les vulnérabilités et les mises à jour disponibles de préférence via des envois d'alerte sécurité.
- Mettre à jour régulièrement les micrologiciels (firmwares) des automates et autres équipements (variateurs, écrans tactiles…) et les systèmes d'exploitation et les applications des PCs de supervision et autres serveurs de bases de données. Comme indiqué précédemment, cela peut nécessiter d'avoir une installation miroir réduite pour les tests de ces mises à jour avant leur mise en production.

L'entreprise doit donc accepter un coût et une période de maintenance pour ces mises à jour et les tests associés. La mise à jour doit parfois passer par la migration d'un OS non maintenu à sa version suivante. Précisément, Windows XP n'étant plus maintenu, il ne devrait plus être utilisé sur

des systèmes industriels critiques. Wannacry a mis en évidence la vulnérabilité de Windows XP et le coût potentiel de sa conservation. (Le coût de Wannacry qui exploitait une faille connue de XP a été estimé entre 1 et 4 Milliards de dollars par différentes agences nationales de sécurité informatique).

Pour les systèmes qualifiés avec des versions d'un firmware et d'un système d'exploitation, il est nécessaire lors de la conception du projet, de prendre en compte la mise à jour des firmwares des automates et des logiciels de supervision et systèmes d'exploitation et d'intégrer des mécanismes de requalification des équipements si besoin. Si ce n'est pas possible (système ancien), le système doit être isolé du réseau avec uniquement des communications précises, surveillées et unidirectionnelles vers le réseau.

2.7 - Les plans de reprise et de continuité

L'objectif du plan de reprise est de se préparer à faire face à des événements exceptionnels pour lesquels toutes les mesures précédentes auraient échoué afin de minimiser les impacts et permettre de redémarrer l'activité le plus rapidement possible.

Il est important pour cela de disposer d'une sauvegarde de chaque automate, des équipements réseau et du PC de supervision, des codes sources et des données et de prévoir des modes de fonctionnement dégradé (le système continue la production, mais moins vite ou avec plus d'interventions manuelles). Les sauvegardes doivent être stockées sur des supports amovibles ou sur des machines éteintes ou déconnectées du réseau (sauvegardes froides).

Pour des systèmes critiques, on prévoira un approvisionnement (automates, PC) pour limiter la durée de l'arrêt de la machine.

3 - Analyse d'incidents

La méthode décrite, le cours reprend à partir des fiches du CLUSIF [3] avec 4 incidents de cybersécurité et invite les étudiants à chercher quelle vulnérabilité a été exploitée et quelle étape de la méthode aurait pu empêcher une telle attaque :

- Fiche 16 : Attaque d'une station d'épuration des eaux
- Fiche 19 : Empoisonnement de l'eau potable
- Fiche 32 : Prise de contrôle du système de production d'une aciérie
- Fiche 4 : Coupure générale d'électricité - BlackEnergy

Ces 4 études de cas montrent qu'à chaque fois, ce sont 2 failles successives qui ont permis l'attaque. La mise en place de la méthode de sécurisation des installations, assez accessible, aurait permis de les éviter.

Les fiches étant bien détaillées, le lecteur pourra s'y référer pour organiser une activité similaire, et y puiser d'autres exemples.

4 - Étude de cas pratique

Le cours de cybersécurité termine par une étude de cas fictive de sécurisation d'un site OIV associé à sa mise en œuvre en Travaux Pratiques sur 12h.

Le contexte : on considère une usine pharmaceutique française disposant d'un atelier de fabrication et d'un atelier d'emballage. L'usine produisant de l'insuline (nécessaire aux personnes

diabétiques), elle est classée Opérateur d'Importance Vitale (OIV). De plus, la réglementation pharmaceutique exige une traçabilité importante de la qualité de la production. Le siège de l'entreprise situé au Danemark héberge la base de données comprenant les autorisations d'accès et doit pouvoir récupérer les données de production de l'usine française.

La supervision des deux ateliers a lieu dans un local de supervision situé dans l'atelier. La supervision, outre l'affichage des informations sur le système (mesures, alarmes… utilisées par les services maintenance et qualité), permet au chef d'atelier de passer certaines machines en mode manuel et de modifier les cadences, notamment pour adapter le débit de la production à celui de l'emballage.

Pour travailler en salle de TP, les IP publiques des routeurs site sont remplacées par des IP du sous-réseau 192.168.2.0/24, qui joue le rôle de réseau « public » de la salle de TPs. Une connexion sécurisée VPN relie le site Danois et le site Français.

L'installation est donc la suivante lors de l'arrivée des étudiants « sur site » :

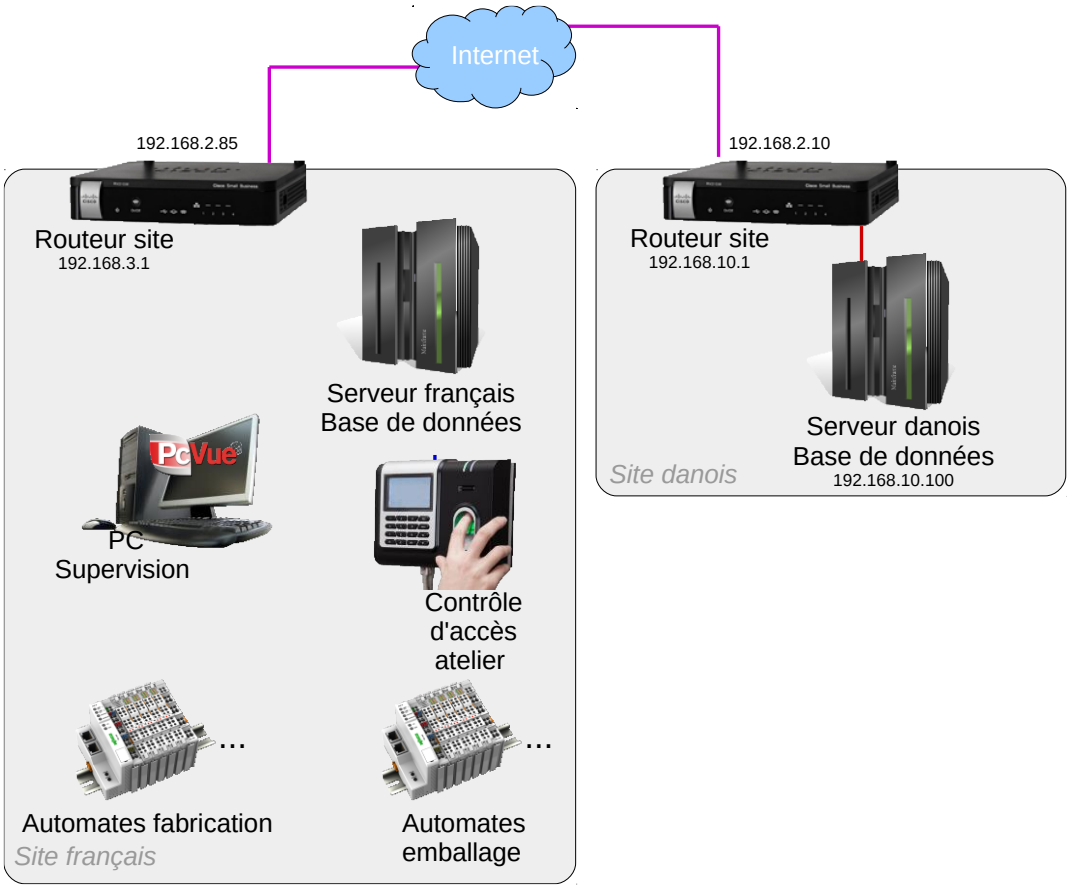


Figure 5 : Equipements de l'entreprise de production avant sécurisation du site

Tout d'abord, les étudiants mettent en place la connexion sécurisée VPN site-à-site entre les 2 routeurs (Cisco RV215W).

Ensuite, les étudiants mettent en place une supervision simplifiée : une entrée Tout ou Rien de l'automate de fabrication remonte à la supervision en modbus TCP. Celle-ci contrôle également une sortie Tout ou Rien de ce même automate.

Les étudiants demandent ensuite au superviseur PCVue le stockage des valeurs de l'entrée et de la sortie de l'automate dans le serveur de base de données SQLServer (requête SQL passant sur TCP/IP).

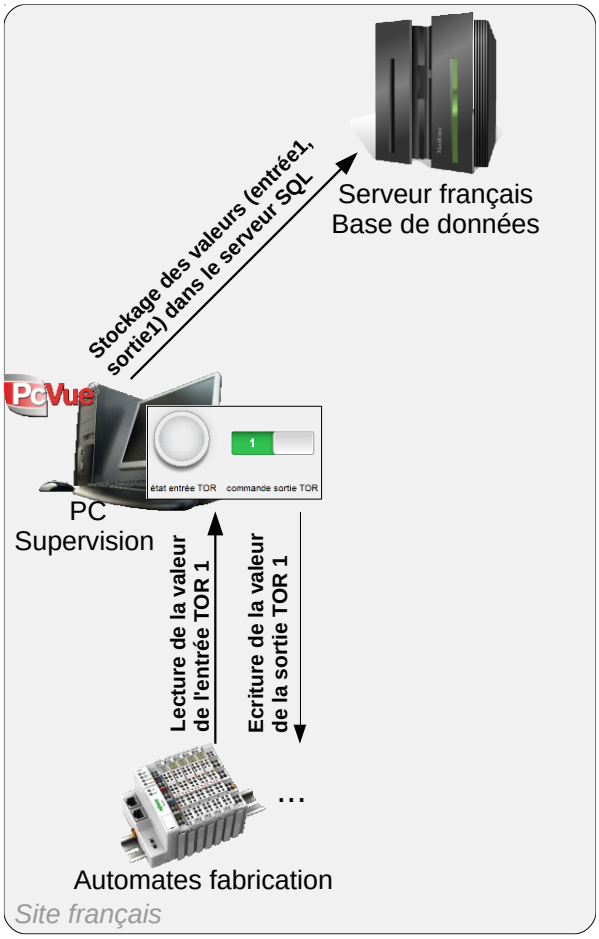


Figure 6 : Echanges entre les équipements du site français

On reprend alors à travers ce TP les différentes étapes de la méthode pour sécuriser un site :

4.1 - Sensibilisation des personnels

Les étudiants doivent expliquer rapidement la mise en place d'une politique de sensibilisation des personnels à la sécurité du site et à la cybersécurité, en insistant notamment sur le phishing, les clés USB, les fichiers douteux, le choix et le non stockage des mots de passe et l'alerte à donner en cas de détection d'un comportement suspect de machine.

4.2 - Cartographie des installations et analyse du risque

Le site étant d'importance vitale et un contrôle à distance non nécessaire (un chef de production est toujours présent sur le site), l'analyse du risque amène à choisir de bloquer tous les accès entrants dans l'atelier.

Les données de traçabilité doivent tout de même être disponibles pour les Danois. Elles seront partagées via un serveur SQL situé à l'extérieur de l'atelier, dans une DMZ. Une DMZ (zone démilitarisée, en référence à la zone « neutre » servant notamment pour l'échange de prisonniers entre les 2 Corées) est un sous-réseau du site accessible depuis l'extérieur et depuis les sous-réseaux sensibles du site (ici, l'atelier). Il n'est pas possible d'accéder au réseau atelier depuis la DMZ. Les équipements des sous-réseaux sensibles y déposent les données que les équipements extérieurs viendront y chercher, sans avoir besoin d'entrer dans le sous-réseau sensible.

Les données pourraient aussi, en plus, être stockées localement sur le PC de supervision (ou sur un serveur SQL dans l'atelier) si l'on voulait être sûr de les conserver en cas d'attaque. Pour tenir en 12h, ce dernier point n'est pas retenu.

Dans les ateliers, l'isolation des réseaux amène à un VLAN et sous-réseau par atelier (192.168.4.0/24 pour la fabrication et 192.168.5.0/24 pour l'emballage) et un VLAN et sous-réseau (192.168.6.0/24) pour le contrôle d'accès du bâtiment. Ainsi, l'infection d'une machine aura plus de mal à se propager d'un atelier à un autre, les zones de diffusion étant segmentées. Le PC de supervision doit communiquer avec les deux VLANs des ateliers. Pour cela, soit sa carte de réseau accepte les VLANs tagués et est rattachée aux 2 VLANs ateliers, avec 2 IP, soit on met 2 cartes réseaux dans le PC de supervision chacune attachée à un VLAN avec une IP dans le sous-réseau associé.

La cartographie des installations et l'analyse du risque amènent les étudiants à proposer l'architecture réseau suivante :

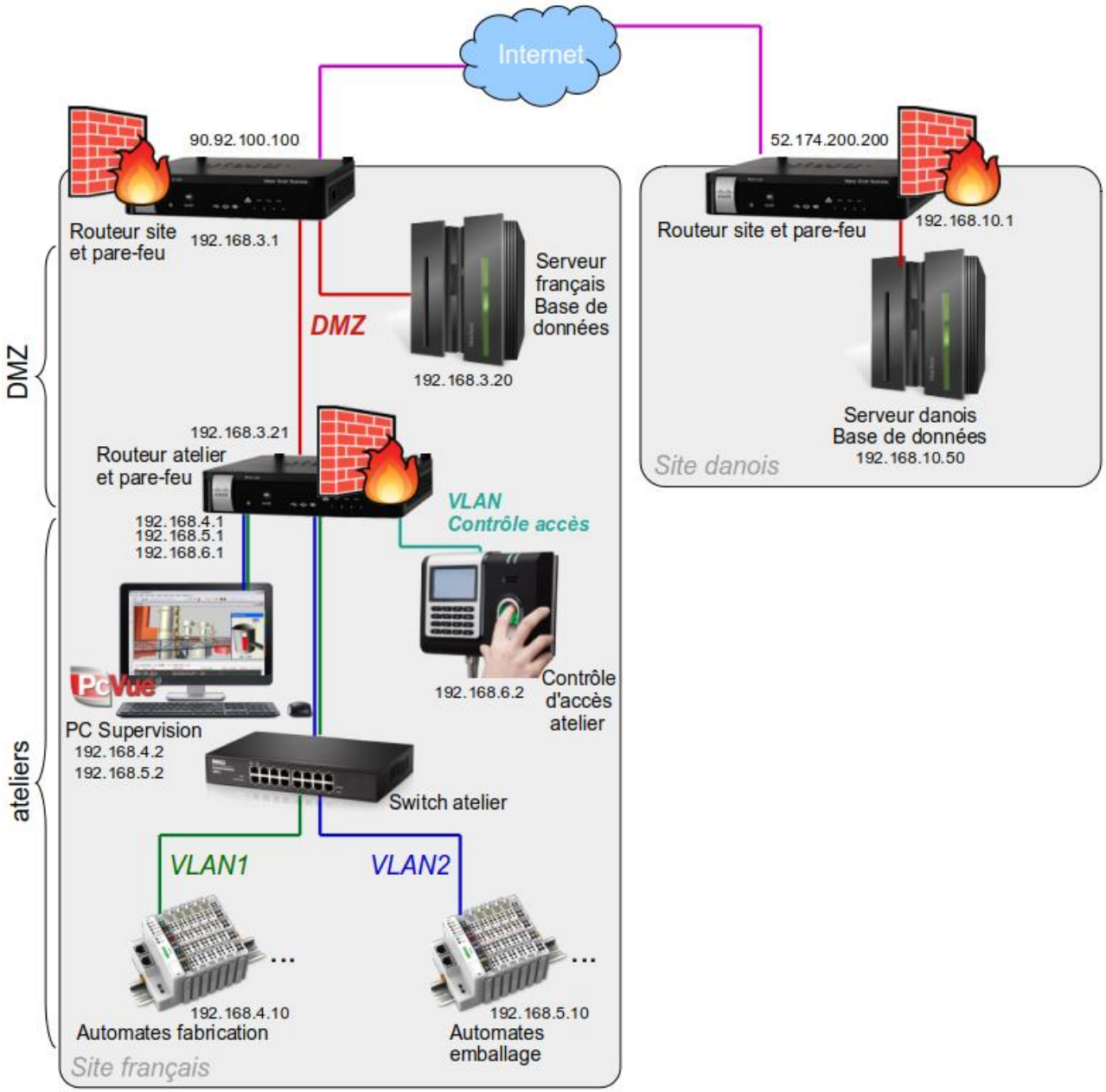


Figure 7 : Architecture réseau de l'entreprise de production de médicaments

4.3 - Prévention : défense en profondeur

Les étudiants câblent leur réseau et abordent alors les différents aspects de la défense en profondeur de l'installation.

4.3.1 – protection physique des équipements

Les étudiants doivent indiquer les mesures de protection physique minimales : armoires électriques et réseaux sous clés, accès à l'atelier et au local de supervision soumis à un contrôle d'accès strict, suppression des accès des employés ayant quitté l'entreprise.

4.3.2 – Cloisonnement des réseaux et durcissement de leur configuration

Les étudiants configurent les pare-feux :

- Le routeur atelier refuse toutes les requêtes entrantes (fonctionnement en passerelle unidirectionnelle) et n'accepte que les requêtes sortantes du PC de supervision vers le port 1433 du serveur SQL.
- Le routeur site français refuse également les requêtes entrantes. Il est configuré en VPN site-à-site avec le routeur site danois, ce qui permet un accès au serveur de base de données français par le serveur danois.

Sur les serveurs de base de données (des PCs équipées de SQL Server Express), SQL Management Studio Express sert de client SQL permettant de lire dans une base de données locale ou distante.

4.3.3 – Durcissement du PC de supervision

Sur le PC de supervision, sur lequel les étudiants sont administrateurs :

- Vérification de l'activation du firewall du PC,
- Établissement d'un identifiant/mot de passe pour chaque utilisateur (le chef d'atelier et le technicien de maintenance) avec des droits limités (pas de possibilité d'installation de logiciels et pas de possibilité de modifier la configuration réseau),
- Établissement également d'une ségrégation des droits sous PCVue. Le chef d'atelier a les droits pour modifier la variable de sortie de l'automate. Le technicien de maintenance peut juste observer les variables,
- Blocage des médias amovibles usb : blocage de la détection des périphériques de stockage USB, désactivation du pilote de gestion des périphériques de stockage USB, désactivation de l'exécution automatique, en suivant la procédure proposée sur le site de l'ANSSI.
- Indication du fait qu'il faudrait supprimer les logiciels autres que le superviseur PCVue du PC, en particulier les logiciels de programmation des automates, très dangereux (ils permettraient à une personne prenant le contrôle du PC de supervision de modifier les programmes de production des médicaments), et les logiciels de bureautique, très attaqués.
- Le PC de supervision étant client des automates serveurs et client SQL, son pare-feu doit être configuré pour ne pas accepter les requêtes entrantes. C'est la configuration par défaut du pare-feu windows. On peut vérifier qu'elle est correcte et que le pare-feu est bien activé au lancement du PC (Menu Sécurité Windows des paramètres de la machine)

4.3.4 – Durcissement des automates

Les étudiants récupèrent le firmware à jour chez le fabricant et le chargent dans les automates.

Ils désactivent le serveur web embarqué de l'automate. Les serveurs web embarqués sont très utilisés pour la configuration des équipements industriels mais sont vulnérables car utilisant des technologies web standard (parfois anciennes, les automates ayant une longue durée de vie).

4.4 - Surveillance des installations et détection des

- Les étudiants utilisent le port mirroring du routeur site et Wireshark (avec des filtres bien choisis) pour surveiller les échanges réseaux. Ils peuvent vérifier que les trames extérieures au site sont bien chiffrées (protocole ESP) et détecter les incidents ;

	Time	Source	Destination	Protocol	Length	Info
379	17.166214	192.168.2.10	192.168.2.85	ESP	310	ESP (SPI=0x51d0f1a7)
380	17.173467	192.168.10.100	192.168.3.102	TDS	246	SQL batch
381	17.174008	192.168.3.102	192.168.10.100	TDS	142	Response
382	17.175583	192.168.2.85	192.168.2.10	ESP	214	ESP (SPI=0x11bfd739)

```
Frame 379: 310 bytes on wire (2480 bits), 310 bytes captured (2480 bits) on interface 0
Ethernet II, Src: Cisco_aa:c7:24 (10:bd:18:aa:c7:24), Dst: Cisco_ab:c6:99 (10:bd:18:ab:c6:99)
Internet Protocol Version 4, Src: 192.168.2.10, Dst: 192.168.2.85
Encapsulating Security Payload
```

```
Frame 380: 246 bytes on wire (1968 bits), 246 bytes captured (1968 bits) on interface 0
Ethernet II, Src: Cisco_ab:c6:98 (10:bd:18:ab:c6:98), Dst: Dell_04:14:53 (d8:9e:f3:04:14:53)
Internet Protocol Version 4, Src: 192.168.10.100, Dst: 192.168.3.102
Transmission Control Protocol, Src Port: 4854, Dst Port: 1433, Seq: 20248, Ack: 7069, Len: 192
Tabular Data Stream
```

Figure 8 : Surveillance des échanges au niveau du routeur site

- Les étudiants ouvrent les journaux du routeur pour connaître l'historique des connexions VPN ;
- Les étudiants ouvrent les journaux du serveur SQL (accessibles par SQL Management Studio) pour connaître l'historique des connexions au serveur SQL.

4.5 - Traitement des incidents, chaîne d'alerte

Les étudiants proposent une procédure en cas de virus sur le serveur SQL : mise en place d'un archivage (archivage obligatoire pour la traçabilité de la fabrication de médicaments) local par exemple en attendant la remise en état du serveur SQL.

4.6 - Veille sur les menaces et les vulnérabilités

Les étudiants proposent un plan de veille sur les mises à jour de sécurité du fabricant des automates, de Windows et de PCVue.

4.7 - Les plans de reprise et de continuité d'activité

Les étudiants sauvegardent les programmes automates, l'application de supervision et les configurations des 2 routeurs sur un dossier qu'ils expliquent devoir garder sous clé sur une machine ou un support non connecté au réseau.

5 - Étude de cas simulée

Période propice aux discussions sur l'importance vitale de l'industrie pharmaceutique, mai 2020 a amené à faire le TP cybersécurité à distance. Cisco Packet Tracer permet de travailler sur la configuration des équipements réseaux (VPN, pare-feu, VLAN, DMZ), sans toutefois apporter l'ensemble des savoir-faire d'un véritable TP : programmation du superviseur, ségrégation des droits, configuration des restrictions sur les PCs, mise à jour des automates, mise en place des requêtes SQL…

Le fichier pkt « professeur » est donné à titre indicatif en pièce jointe à cette ressource. Toutes les fonctionnalités n'ont pas été testées.

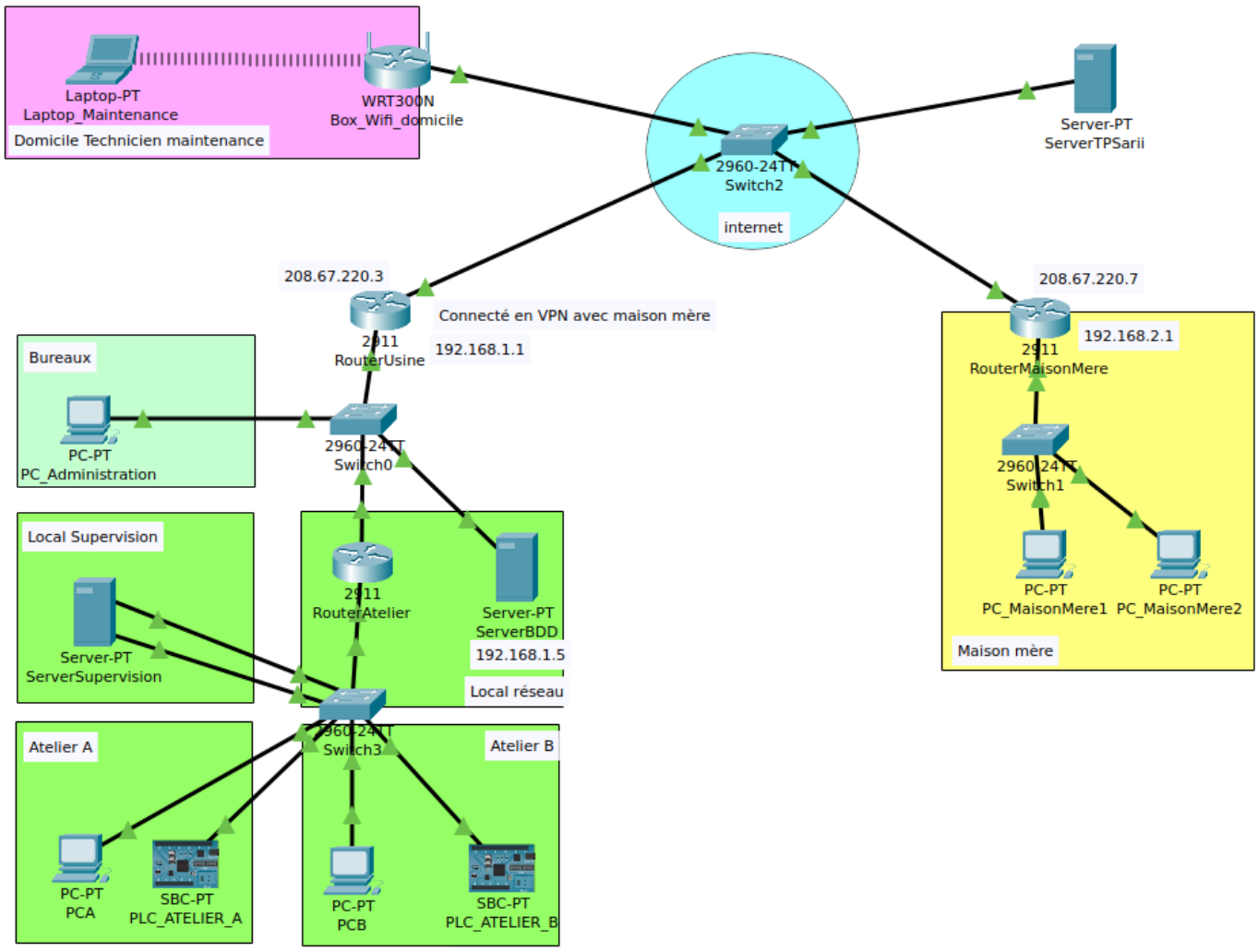


Figure 9 : copie d'écran de l'installation pharmaceutique sous cisco packet tracer

6 - Conclusion

Le cours et le TP cybersécurité des systèmes automatisés industriels permettent aux étudiants d'être sensibilisés au risque cyber, de connaître les principales mesures à mettre en place pour sécuriser un site. Cette séquence pédagogique leur fait prendre conscience que ces mesures, à leur portée, permettent de contrer la plupart des attaques ciblées ou non contre un réseau informatique industriel.

Ce cours/TP est à compléter par la présentation et l'exploitation du protocole sécurisé OPC-UA, pris en compte par les automates industriels modernes (Siemens S7-1500, Schneider M251 par exemple) et objet d'une seconde ressource de ce dossier [5].

Références :

[1]: ANSSI : Publications sur la cybersécurité des systèmes industriels
https://cyber.gouv.fr/publications/la-cybersecurite-des-systemes-industriels

[2]: Guide Cybersécurité des systèmes industriels, Clusif, 2021
https://clusif.fr/publications/guide-cybersecurite-des-systemes-industriels-2021/

[3]: Fiches incidents cyber si industriels, Clusif, 2022
https://clusif.fr/publications/fiches-incidents-cyber-si-industriels/

[4]: Panorama des référentiels - Cybersécurité des systèmes industriels
https://clusif.fr/publications/panorama-des-referentiels-2eme-edition-2/

[5]: OPC UA, un protocole sécurisé pour l'automatisme industriel (à paraître)

[6]: ANSSI – CERT-FR : Centre Gouvernemental de veille, d'alerte et de réponse aux attaques informatiques. https://www.cert.ssi.gouv.fr/

[7]: La Revue 3EI – N°93 - juillet 2018 : Cyber-sécurité et réseau électrique,
https://eduscol.education.fr/sti/si-ens-paris-saclay/ressources_pedagogiques/3ei-n93-juillet2018-cybersecurite-et-reseau-electrique

La Cybersécurité chez Eiffage Energie Systèmes

La Revue 3E.I
Culture Sciences de l'Ingénieur

Édité le 01/02/2024

école normale supérieure paris—saclay

Jocelyn ZINDY[1] – Franck LE GALL[2]

[1] *Eiffage Energie Systèmes,* [2] *ISEN Brest*

Cette ressource fait partie du N°111 de La Revue 3EI de janvier 2024.

Cet article est issu d'un entretien avec M. Zindy, directeur du développement commercial Cybersécurité - Industries du Futur de Eiffage Energie Systèmes. Il présente d'une part le cadre général de la Cybersécurité dans le domaine de l'OT (Operational Technology) en précisant les spécificités de l'OT vis-à-vis de l'IT et d'autre part comment un grand groupe comme Eiffage intègre celle-ci dans ses offres.

1 - Introduction

Les cyberattaques sont devenues monnaie courante dans nos sociétés modernes. Elles touchent les réseaux internet, mobiles, Wifi et les systèmes qui leurs sont connectés. On pense en premier lieu aux ordinateurs et à tout l'univers de l'IT (Information Technology) mais cette menace est aussi bien réelle pour le monde industriel. Les grandes entreprises du domaine sont particulièrement attentives à ces évolutions et proposent des solutions pour y remédier.

Le groupe Eiffage exerce dans de nombreux domaines : construction, infrastructures, concessions et énergie. Il s'agit du troisième groupe de constructions et de concessions français, derrière Vinci et Bouygues, et du quatrième groupe européen. En 2023, il comprend plus de 70 000 collaborateurs. C'est dans la branche « Energie systèmes » qui compte environ 29 000 personnes que travaillent une trentaine d'experts en cybersécurité.

Dans le monde de la cybersécurité, Eiffage Energie Systèmes n'est ni un « pure player », ni spécialisé dans la cybersécurité IT, il reste proche de son cœur de métier en offrant un volet cybersécurité dans le domaine de l'OT (Operational Technology). En tant qu'intégrateur multi technologies, il propose à ses clients de répondre aux clauses de cybersécurité qui sont maintenant présentes dans les appels d'offre.

Il existe de nombreuses entreprises qui proposent des solutions de cybersécurité. On peut citer Atos ou Sopra Steria qui bien que très majoritairement orientées vers l'IT (environ 95%) réalisent aussi des prestations orientées OT (environ 5%). D'autres acteurs de plus petite taille (quelques dizaines de personnes), sont des « pure players » c'est-à-dire des petites entreprises spécialisées dans la cybersécurité OT.

Figure 1 : La cybersécurité dans l'Operational Technology

2 - OT versus IT

Qu'est que l'OT ? Nous sommes habitués à entendre parler de l'IT qui est exploité au sein des systèmes d'information d'entreprise alors que l'OT est mise en œuvre pour la gestion et la prise en charge des systèmes d'Information Industriels ou techniques. Alors que l'IT concerne principalement les réseaux, les logiciels, la téléphonie, c'est à dire la gestion de l'entreprise, l'OT est plus proche des process et prend en compte les automates, les capteurs …. Le marché de la cybersécurité se réparti actuellement à hauteur d'environ 80 % d'IT et 20 % d'OT.

La frontière n'est pas étanche entre ces deux domaines d'autant plus que l'évolution des usages fait que la demande de remontée d'informations, de mesures est croissante. La nécessité de connecter des outils industriels qui étaient souvent déconnectés du réseau (filaire ou Wifi) se fait de plus en plus présente ouvrant ainsi mécaniquement une porte d'accès pour des malveillants.

Le besoin de continuité numérique et d'analyse de données a, par exemple, amené les hôpitaux à développer leurs réseaux en connectant de plus en plus de systèmes sur internet et en offrant des accès Wifi. Ces réseaux mal sécurisés ont fait l'objet de nombreuses attaques ces dernières années avec des demandes de rançons sous la menace de divulgation de dossiers confidentiels ou une paralysie des services et des soins critiques comme, par exemple les salles d'opération. Dans cet exemple, le vol de dossiers confidentiels peut être rangé dans le domaine de l'IT alors que la prise de contrôle d'une salle d'opération le sera dans celui de l'OT.

Autre exemple moins médiatisé mais tout aussi problématique. Celui d'une station traitement des eaux dont les opérateurs ont perdu le contrôle de vanne sous l'effet d'une cyberattaque, menaçant ainsi de contaminer des circuits d'eaux potables. On comprend bien que de telles actions, sans être nécessairement d'une grande complexité, peuvent entraîner des conséquences délétères sur la santé des personnes et l'intégrité des matériels.

Alors, comment distinguer l'IT de l'OT ? Comme nous l'avons écrit, la frontière est poreuse mais l'OT comporte tout de même deux contraintes spécifiques :

- Les contraintes liées aux process. Nombreux process industriels nécessitent un fonctionnement en continu avec des exigences élevés en temps de réponse. Il est alors plus difficile d'intervenir et d'implémenter une solution cyber, et de la maintenir à jour …
- L'hétérogénéité des équipements : les systèmes industriels comportent de nombreux systèmes, capteurs, ordinateurs, automates … dont les caractéristiques sont très

hétérogènes. Les intervenants doivent posséder une solide culture industrielle pour y implémenter des solutions cyber.

Eiffage Energie Systèmes propose ses services de cybersécurité (majoritairement OT) dans l'industrie, les infrastructures et réseaux, le tertiaire et les villes et collectivités. Ils proposent de sécuriser les composantes de systèmes complexes (Serveur, postes utilisateurs, réseaux, automates, logiciels …) (figure 2).

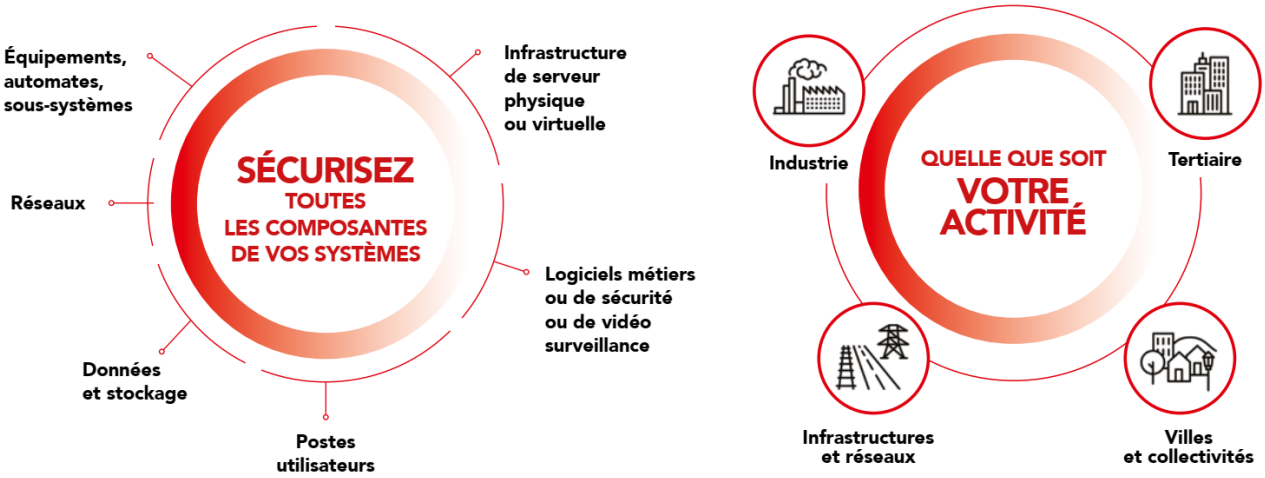


Figure 2 : Les domaines de la cybersécurité OT chez Eiffage

3 - Une prise de conscience d'un besoin à grande échelle

Ces menaces sont largement prises au sérieux par les autorités qui ont chargé l'ANSSI (Agence Nationale de la Sécurité des Systèmes d'Information) d'un rôle de communication et de préconisation auprès des opérateurs et des industriels.

On peut lire sur son site que « L'ANSSI, autorité nationale de cybersécurité, propose au premier ministre les mesures destinées à répondre aux crises affectant la sécurité des systèmes d'information des autorités publiques et des opérateurs régulés. Elle coordonne l'action gouvernementale et anime l'écosystème national. »

L'ANSSI a donc pour interlocuteurs des entreprises qui sont classées en fonction de leur importance lors de crises de cybersécurité. Ce sont des :

- Opérateurs d'Importance Vitale (OIV) : environ 200 organisations ayant des activités indispensables à la survie de la nation ou dangereuses pour la population [Wikipédia]. C'est le cas, par exemple, des gestionnaires d'installations nucléaires ou d'alimentation en eau, des acteurs du secteur militaire, ou des organismes opérant dans le domaine de la santé. Ces OIV sont tenus de développer des stratégies de cybersécurité pour protéger leurs activités.
- Opérateurs de Services Essentiels (OSE) : environ 1000 « opérateurs, publics ou privés, offrant des services essentiels au fonctionnement de la société ou de l'économie et dont la continuité pourrait être gravement affectée par des incidents touchant les réseaux et systèmes d'information nécessaires à la fourniture desdits services » (Article 5 de la loi du 26 février 2018.) Pour être considéré comme un OSE, l'entreprise doit répondre à trois critères :
    - Fournir au moins un service essentiel à la continuité d'activités économiques ou sociétales critiques ;

- Être tributaire des réseaux et des systèmes d'information pour la fourniture de ce service ;
- Être confronté à un risque de suspension de ce service en cas de cyber-incident.

Afin que ces OIV et OSE atteignent leurs objectifs l'ANSI édite en particulier :

- Des recommandations ;
- Un référentiel PIMSEC pour l'intégrateur (PAX pour le Neuf, PAMS pour la maintenance) ;
- Une méthode de gestion des risques EBIOS.

Les entreprises comme Eiffage Energie Systèmes qui proposent leurs services peuvent obtenir des qualifications pour les audits, la détection d'incident ou encore l'intervention après incident.

En complément des actions de l'ANSSI on peut noter que certaines collectivités territoriales sont pro actives dans ce domaine. Ainsi, la région Grand-Est propose une aide aux entreprises en prenant en charge 50% du coût d'un diagnostic de cybersécurité.

4 - Quelques « Success Stories »

Les projets réalisés par Eiffage Energie Systèmes sont très variés et basés sur trois types de prestations :

- **Audit, diagnostic et conseils :** il s'agit d'inventorier, d'évaluer les risques, de proposer un plan d'action pour atteindre un niveau de sécurité homologué.
- **Maintien en conditions de sécurité :** la veille de vulnérabilité et détection/qualification de nouvelles menaces, la formation de sensibilisation, l'amélioration continue, la simulation d'attaques et la réponse sur incident permettent de maintenir le niveau de sécurité à un niveau acceptable.
- **Sécurisation des systèmes :** l'intégration de solutions matérielles et logicielles (détection et protection), la remédiation et l'évaluation des risques résiduels permet de protéger les systèmes des risques de cyberattaque.

Á titre d'exemple de sécurisation, on peut citer :

- La sécurisation du contrôle-commande d'une ventilation nucléaire qui a nécessité la prise en compte de 3 000 entrées sorties physiques, 400 synoptiques.
- La sécurisation du contrôle-commande d'unité de contrôle des opérations sur le banc d'essai des accélérateurs à poudre situé à Kourou en Guyane française.
- La rénovation du système de vidéo-surveillance d'un site militaire par l'étude des risques de sécurité (cyberattaques, intrusions, piratages…) à l'aide de la méthode EBIOS.
- La rénovation du système de contrôle-commande d'une centrale électrique munie de groupes thermiques Diesel.
- La sécurisation de l'application TETRA de communication opérationnelle interne aux agents de la RATP.

Figure 3 : Visuel cybersécurité de Eiffage Energie Systèmes

La méthode suivie lors de ces projets commence par du « bon sens » : par exemple pour la sécurisation d'un automate, il faut renforcer les mots de passe, fermer les ports inutilisés, utiliser un VPN, segmenter l'architecture, ajouter des sondes …. Au-delà de ces premières mesures, il s'agit principalement d'un travail d'intégration basé sur la Norme de sécurisation de l'OT (ISO 62443).

Pour certains projets, il est nécessaire de réaliser des développements sur mesure. Ainsi dans le cas d'un système de supervision de contrôle commande avec des contraintes temps réel trop élevées, il a été nécessaire de développer un logiciel ad hoc.

5 - Et demain …

Chez Eiffage Energie Systèmes, on estime que le secteur de la cybersécurité OT est en pleine croissance et qu'il va falloir de nombreuses années pour mettre au niveau les systèmes industriels.

En effet, les menaces cyber ne font que croitre et les autorités souhaitent rehausser les niveaux de protection. Dans ce cadre, la directive NIS-2 (Directive Network and Information System Security 2) prévue pour le mois octobre 2024, aura pour conséquence de multiplier par 30 à 40 le nombre d'entreprises qui seront obligées de se cyber-sécuriser.

Ainsi, alors qu'aujourd'hui, Eiffage Energie Systèmes dispose de compétences Cyber dans quelques entités opérationnelles, il paraît fort probable qu'il faudra à terme des équipes cyber réparties sur tout le territoire au plus près des utilisateurs. Le potentiel d'emploi dans ce domaine est donc prometteur.

Ressource publiée sur Culture Sciences de l'Ingénieur : https://eduscol.education.fr/sti/si-ens-paris-saclay

Wattsense - Siemens, une entreprise pour une GTB sécurisée

La Revue 3EI
Culture Sciences de l'Ingénieur

Édité le 15/02/2024

école normale supérieure paris-saclay

Mohamed ZENADI[1] - Magali SAUVERGEAT[2]

[1] *Responsable Technique – Wattsense*
[2] *Enseignante BTS CIEL Arpajon*

Cette ressource fait partie du N°111 de La Revue 3EI de janvier 2024.

Cette ressource issue d'un entretien avec Mohamed Zenadi de la société Wattsense, présente la prise en compte de la cybersécurité pour les objets connectés par l'entreprise Wattsense. On y retrouve les concepts de cybersécurité réseaux décrits dans la ressource « Fondamentaux de la sécurité réseau » [5].

1 - L'historique de Wattsense

Wattsense est une start-up lyonnaise, elle est située à Dardilly, à proximité de Lyon, elle a démarré son activité en 2017. En octobre 2021, elle est acquise par Siemens et devient une unité autonome au sein de la division Siemens Smart Infrastructure. Les solutions développées par Wattsense permettent aux entreprises implantées dans l'Union européenne de se conformer à la directive sur la performance énergétique des bâtiments (DPEB).[1] [2]

Logo Wattsense - Siemens

Cette directive prévoit l'installation de systèmes d'automatisation et de contrôle des bâtiments dans le secteur tertiaire afin d'améliorer leur efficacité énergétique et de réduire leurs émissions de CO2.

2 - L'équipement : Wattsense Tower

Wattsense Tower est conçue pour connecter tous les types d'équipements de tous les bâtiments : capteurs IoT LoRa, compteurs, matériels de chauffage, de climatisation ou de traitement d'air, systèmes de gestion technique du bâtiment (GTB). Cet équipement s'interface également avec des équipements communicants et des consoles de supervisions via les protocoles M-Bus, KNX, Modbus, BACnet. [3]

Wattsense Tower se connecte automatiquement au cloud Wattsense via la 3G/4G, uniquement avec un port sortant (réduction de la surface vulnérable). Les données sont transférées régulièrement sur le cloud Wattsense, via un protocole sécurisé utilisant le protocole TLS qui assure authentification, confidentialité et intégrité. [4] Wattsense est autorité de certification pour délivrer les certificats TLS aux équipements.

Pour garantir la continuité de la sécurité, Wattsense Tower peut assurer la mise à jour de son firmware automatiquement via le réseau (mise à jour OTA Over the Air). Les firmwares téléchargés sont authentifiés via des certificats et leur intégrité est vérifiée.

Les données, stockées sur des bases de données dans des clouds privés virtuels (VPC) avec des pare-feux dédiés, peuvent être consultées via des pages Web sécurisées (HTTPS), permettant ainsi au client de configurer un ou plusieurs tableaux de bord (DashBoard). Le client a la liberté de sélectionner les données à afficher, de générer des graphiques, et même de définir des seuils d'alerte selon ses préférences.

Les connecteurs Webhook et MQTT offrent également la possibilité aux clients de recevoir le flux de données en temps-réel dans leurs infrastructures cloud d'entreprise, leur permettant ainsi de gérer ces flux. Parmi les plateformes possibles, on peut citer IoT Hub, IoT Core, Node-RED, et bien d'autres encore.

Le client peut également consulter à distance ces données via une API REST pour peupler une base de données interne, ou réaliser des applications internes de visualisation.

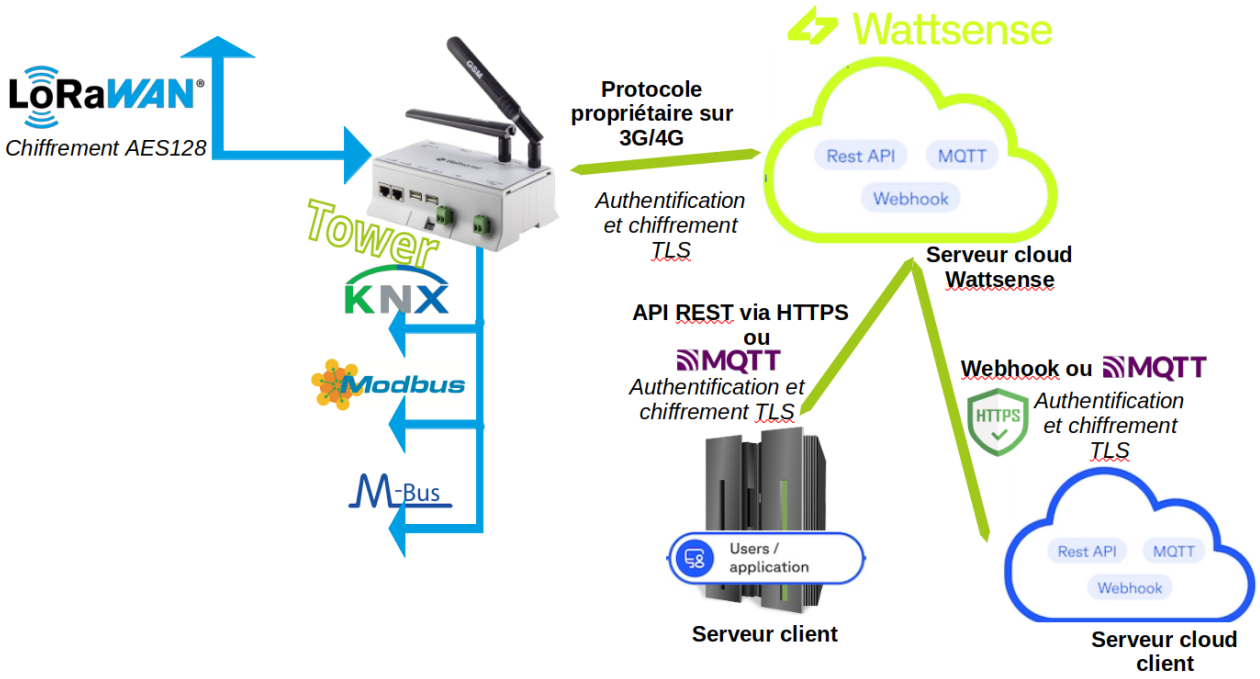


Schéma de l'architecture : Wattsense Tower

3 - Entretien avec Mohamed Zenadi, Wattsense

3EI : Pourquoi les trames LoRa sont déchiffrées dans l'équipement Tower et non sur le Cloud comme le prévoit l'architecture LoRaWan ?

Wattsense : Wattsense Tower s'intègre dans les architectures de supervision de nos clients avec par exemple le protocole Mbus. Afin de permettre l'intégration des équipements LoRaWan au sein de cette infrastructure locale, le décodage des trames doit être effectué au niveau de l'équipement Tower. Les données sont ensuite envoyées sur le cloud Wattsense en 3G/4G via le protocole sécurisé TLS.

3EI : Comment sont stockées les données sur le cloud Wattsense ?

Wattsense : Les données récoltées sont une première fois sauvegardées sous forme de données brutes puis sur une base de données MongoDB Time Series et redondées sur un deuxième cloud distant.

3EI : L'équipement Tower supporte le protocole LoRaWan v1.0, pourquoi le protocole LoRaWan v1.1 n'est-il pas supporté ?

Wattsense : La version 1.1 sera supportée dans la prochaine version de Tower. Cependant nos clients ne la demandent pas, car sur le marché de la GTB, il y a très peu d'équipements qui communiquent en LoRaWan v1.1.

3EI : Quels sont les fabricants d'équipements conseillés pour s'interfacer avec Wattsense Tower ? Exemple : (ADEUNIS, ATIM, ENLESS ...)

Wattsense : Wattsense Tower supporte 64 fabricants, il m'est difficile de privilégier un fabriquant plutôt qu'un autre. Je comprends que vous citiez ADEUNIS, ATIM, ENLESS car ce sont des fabricants français qui produisent des équipements de qualité, mais nous ne pouvons pas faire de recommandations pour un fabriquant plutôt qu'un autre.

3EI : MQTT et Webhook , quels sont les utilisations pour les entreprises ?

Wattsense : Lorsque le client souhaite récupérer le flux de données sans réaliser de développement interne, les webhooks, par leur facilité d'intégration, sont une solution performante qui permet au client d'automatiser des tâches et de déclencher des actions en temps réel.

MQTT est plus adapté aux communications bidirectionnelles, à la messagerie asynchrone et aux systèmes de messagerie plus complexes. Par rapport à MQTT, les Webhooks sont généralement plus simples à mettre en œuvre et sont souvent utilisés pour des communications unidirectionnelles en temps réel.

L'API REST est disponible pour les entreprises souhaitant réaliser un développement spécifique et récupérer les données archivées en interrogeant en différé le cloud Wattsense.

Références

[1] : Article *lyon-entreprises*
https://www.lyon-entreprises.com/actualites/article/internet-des-objets-la-start-up-lyonnaise-wattsense-rachetee-par-le-geant-allemand-siemens

[2] : DPEB : Performance énergétique des bâtiments
https://www.europarl.europa.eu/news/fr/press-room/20230206IPR72112/performance-energetique-des-batiments-neutralite-climatique-d-ici-2050

[3] : Le site Web de Wattsense
https://www.wattsense.com/

[4] : Wattsense : Sécurité IoT
https://www.wattsense.com/fr-fr/resources/securite-iot/

[5]: Fondamentaux de la sécurité réseau, M. Sechehaye, A. Juton, février 2024, https://eduscol.education.fr/sti/si-ens-paris-saclay/ressources_pedagogiques/fondamentaux-dela-securite-reseau

Introduction au dossier « électronique de puissance »

La Revue 3E.I
Culture Sciences de l'Ingénieur
Édité le 22/01/2024
école normale supérieure paris—saclay

François COSTA[1,2]

[1] *Université Paris-Saclay, ENS Paris-Saclay, CNRS, SATIE UMR8029, 91190 Gif-sur-Yvette, France*
[2] *Université Paris Est Créteil UPEC, 94000 Créteil, France*

Cette introduction fait partie du N°111 de La Revue 3EI de janvier 2024.

L'électronique de puissance est une branche relativement récente du génie électrique : les premiers convertisseurs statiques à semiconducteurs de puissance sont en effet apparus industriellement au début des années 1960 avec la mise sur le marché du thyristor, comme l'illustrent les quelques exemples ci-dessous.

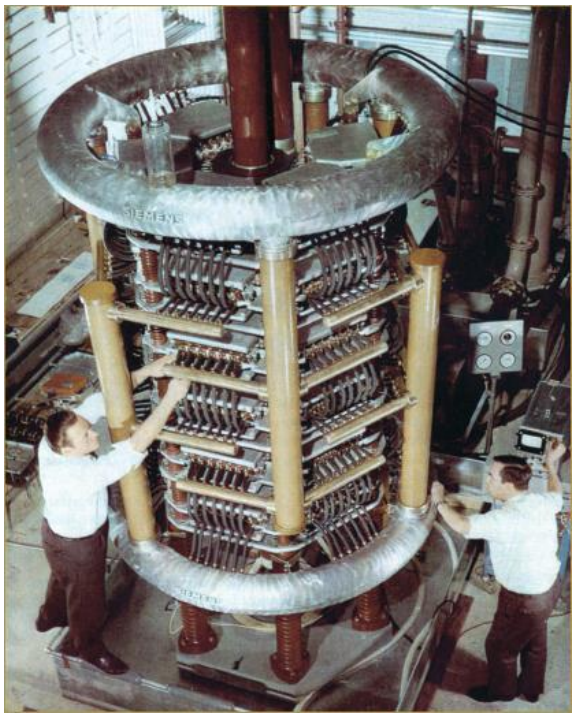


Figure 1a : 100 kV, 1 kA prototype de valve thyristor (192 thyristors 1,6 kV en série) (photo : Siemens Corporate Archives, 1967)

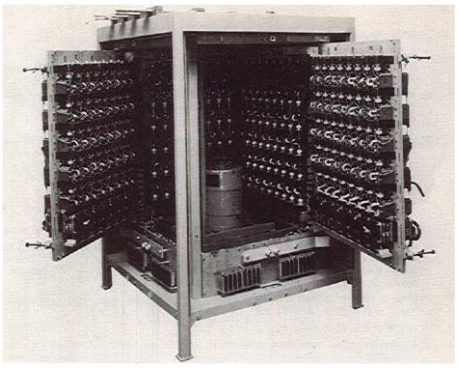


Figure 1b : locomotive BB16685 4,1 MW, bloc de puissance à thyristors, 2x1900 A 1100 V, marche et récupération, doc Alstom

Cette discipline a connu depuis 60 ans un développement extraordinaire en termes de concepts scientifiques, de technologies et de domaines d'application. Elle a permis de développer un usage à haute efficacité de l'énergie électrique grâce à l'excellent rendement de conversion et à la souplesse de contrôle qu'elle autorise : limitée initialement à la traction électrique ferroviaire (figure 1b) et à quelques usages industriels lourds, elle envahit aujourd'hui tous les secteurs et tous les niveaux de puissance : gestion des réseaux, transports terrestres aériens et maritimes, traitement de l'information, domotique, éclairage, etc.

Ainsi, le convertisseur électronique de puissance n'est pas seulement un assemblage de composants qu'on étudie comme tel, mais sa structure, ses composants et sa loi de commande sont le résultat d'un formalisme qui prend en considération l'environnement, la fonction à réaliser, les technologies disponibles et les performances souhaitées. De cette méthodologie résulte le concept de fonctions de conversion qui, assemblées entre elles permettent l'élaboration de systèmes électriques complexes, par exemple pour piloter un réseau électrique ou assurer avec efficacité l'interfaçage de panneaux photovoltaïques.

L'électronique de puissance est aussi une technologie invisible mais dont l'impact sociétal est très fort grâce à l'amélioration de l'efficacité énergétique sans égal dans des secteurs très impactant (transports, bâtiment, industrie). On peut citer les quelques chiffres[1] ci-dessous, correspondant à des valeurs moyennes mondiales sur 25 ans (1990-2015) :

- **Secteur automobile** : l'allumage et le contrôle électronique ont réduit la consommation des moteurs thermiques de plus de 10%, soit environ 4800 TWh et évité la production de 10 MdT CO_2, (pour référence : la consommation électrique annuelle en France est d'environ de 430 TWh, l'émission de CO_2 est d'environ 0,33 MdT en 2018)

- **Secteur industriel** : l'usage de la variation de vitesse électrique dans l'industrie a réduit la consommation en énergie primaire carbonée d'environ 42 000 TWh (soit environ 1 700 TWh par an à mettre en regard des 160 000 TWh de besoin mondial annuel en énergie primaire)

- **Secteur du bâtiment** : l'usage de sources d'éclairage basse consommation a réduit la consommation d'environ 9100 TWh.

C'est donc une technologie clé qui qui joue un rôle majeur pour accroître la soutenabilité des activités humaines en favorisant la progression rapide de l'électricité dans le mix énergétique mondial. Ainsi, les progrès technologiques successifs ont permis des avancées très importantes en termes d'efficacité énergétique, de miniaturisation et de réduction des coûts.

L'électronique de puissance est une discipline enseignée depuis les années 60 : initialement dans les lycées (ouverture des BTS en 1962) et les écoles d'ingénieurs (ENSEEIHT Toulouse, ENSIEG Grenoble pour les premières) puis dans les universités. Elle constitue un domaine scientifique et technologique à part entière qui agrège de nombreux principes physiques et techniques pour concevoir, commander et réaliser des convertisseurs statiques performants généralement inclus dans des systèmes : l'automobile électrique, qui fait l'objet de l'article « Apport des convertisseurs multiniveaux modulaires aux véhicules électriques » [1] présenté dans ce dossier, est un exemple d'une telle approche qui met en évidence la problématique d'architecture de conversion de puissance, de lois de commandes et de technologies des semiconducteurs de puissance.

Actuellement, face à l'électrification rapide et massive des usages sociétaux (transport, habitat, industrie) on assiste à une demande accrue de formation dans tous les domaines du génie électrique et en particulier en électronique de puissance.

Un enjeu technologique fort réside encore et toujours dans l'intégration de puissance : les dispositifs doivent être peu volumineux (ou peu lourds), peu polluants électromagnétiquement, à haut rendement et d'une fiabilité bien contrôlée. La figure ci-dessous donne quelques valeurs de

[1] webinar de J. Baliga, 26/08/2015, Engineering a sustainable society with power semiconductor devices, North Carolina State University, Raleigh, https://www.youtube.com/watch?v=TC2C9NCrva4

densité volumique de puissance atteintes dans différents domaines applicatifs pour des dispositifs industriels et pour des dispositifs de recherche.

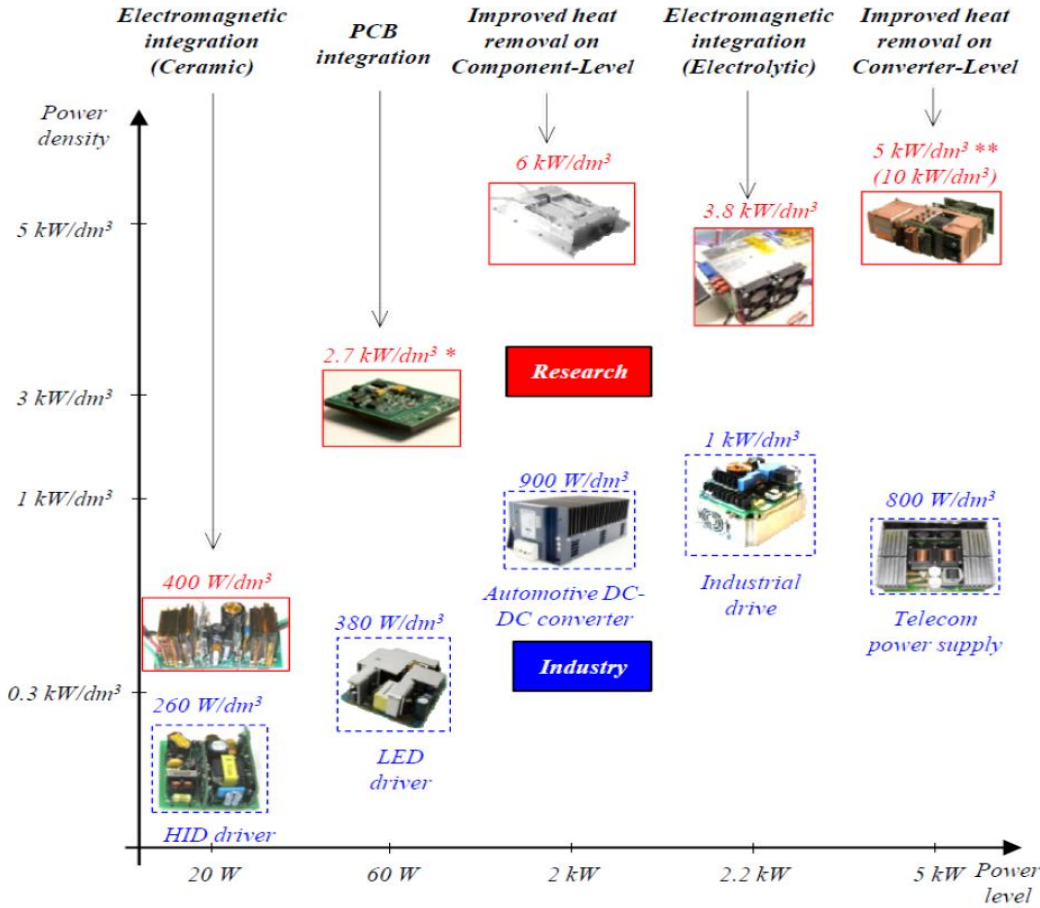


Figure 2 : densité volumique de puissance en fonction de la puissance d'un convertisseur, comparaison des technologies disponibles (doc. ECPE Tutorial "Power Electronics Packaging" Delft Septembre 2014).

Pour illustrer cette évolution, l'article du dossier «Caractérisation Thermoélectrique et Thermomécanique d'Assemblages PCB Intégrants des Puces de Puissance » [2] aborde ce thème de l'accroissement de la densité massique/volumique de puissance en s'intéressant à une nouvelle technique de prise des contact sur les semiconducteurs enfouis dans le circuit imprimé par des mousses métalliques qui remplacent les traditionnels fil de bonding, mutualisant ainsi les fonctions d'amenée de courant et de drainage thermique tout en réduisant fortement les inductances parasites sources de perturbations électromagnétiques.

Le potentiel d'évolution du domaine reste donc très important alors que se profile la révolution des semiconducteurs à large bande interdite (encore appelés grands gap) : carbure de silicium (SiC) pour les hautes tensions et nitrure de gallium (GaN) pour les basses et moyennes tensions ouvrent la voie des hautes fréquences (> MHz) et des hautes températures (> 150°C) !

La figure suivante présente une fresque historique des semiconducteurs de puissance dans laquelle on peut observer la diversification croissante des composants notamment liée leur facilité de commande et aux domaines d'usages de plus en plus larges (calibres de tension et courant, vitesse de commutation).

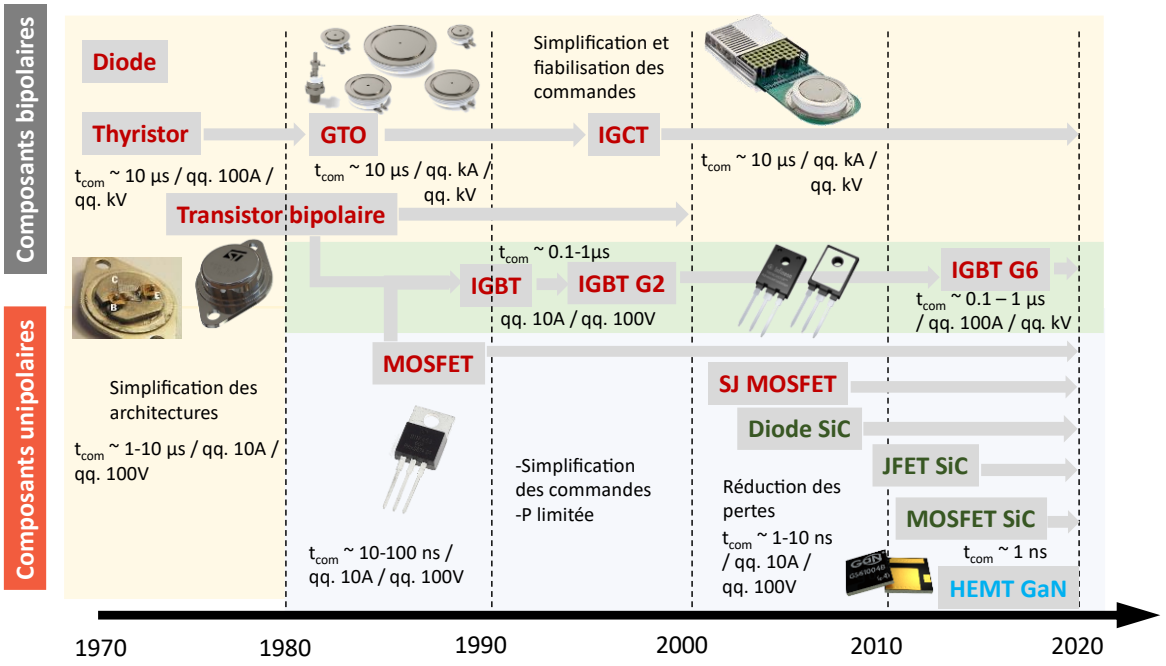


***Figure 3** : fresque historique des semiconducteurs de puissance, d'après Hong LIN, YOLE, PCIM Asia 2015.*

L'article « Technologie des transistors au nitrure de gallium » [3] présente ces nouveaux semiconducteurs GaN et montre leurs potentialités mais aussi leurs défauts de jeunesse et les nouvelles perspectives qu'ils offrent en termes d'accroissement de densité massique de puissance mais aussi de contraintes, notamment en compatibilité électromagnétique (CEM).

L'électronique de puissance a permis une gestion toujours plus efficace et fiable de l'énergie électrique dans des applications de plus en plus larges : électromobilités, industrie, smart grids, liaisons HVDC etc… L'évolution de ses performances est liée aux matériaux semiconducteurs (Si, SiC, GaN, diamant…), à l'inventivité en termes d'architectures, aux procédés de fabrication et au packaging avec des limitations dues au drainage thermique et à la CEM. La figure suivante permet de situer les technologies mises en œuvre selon les applications dans un repère fréquence de commutation/puissance apparente commutée.

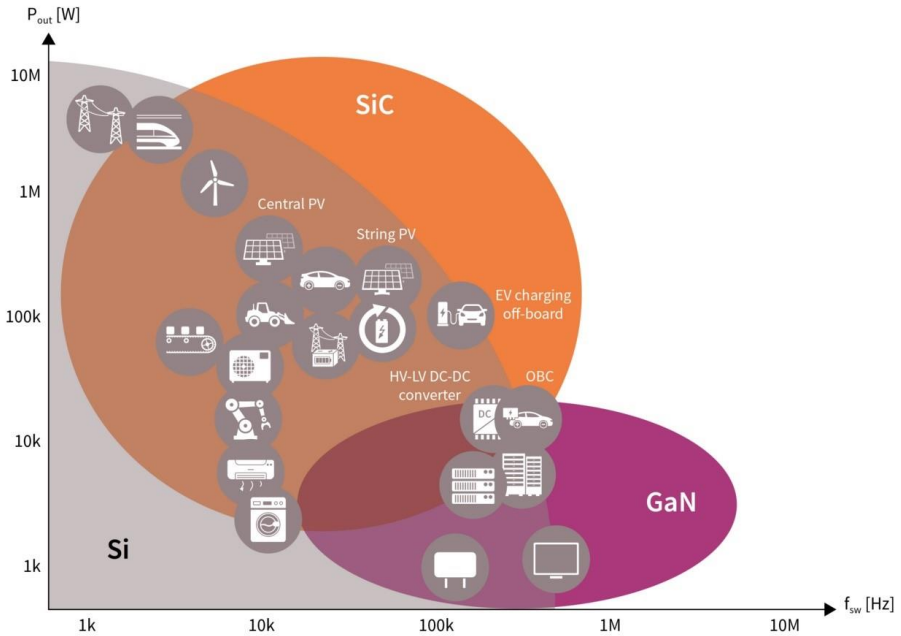


***Figure 4** : domaines applicatifs selon les technologies de semiconducteurs :*
https://www.infineon.com/cms/en/product/technology/wide-bandgap-semiconductors-sic-gan/

Conclusion

La dynamique d'évolution de l'électronique de puissance reste importante, en partie grâce aux nouveaux semiconducteurs et aux progrès des matériaux diélectriques et magnétiques. Cependant, les impératifs de développement durable imposent à présent de considérer, lors de la phase d'étude et de conception d'un convertisseur électronique de puissance, des critères de minimisation des impacts environnementaux sur cycle de vie, ce qui nécessite une approche système, une meilleure compréhension des mécanismes de défaillance et de dégradation, mais également de considérer la recyclabilité et la réparabilité des dispositifs.

L'objectif de ce dossier « électronique de puissance » est de fournir aux lecteurs les éléments de compréhension des évolutions en cours et les enjeux technologiques et sociétaux de cette discipline grâce à quelques exemples illustratifs d'applications innovantes.

Références

[1] Gaël Pongnot et al. « Apport des convertisseurs multiniveaux modulaires aux véhicules électriques », https://eduscol.education.fr/sti/si-ens-paris-saclay/ressources_pedagogiques/apport-des-convertisseurs-multiniveaux-modulaires-aux-ve

[2] Mounira Bouarroudj et al., « Caractérisation Thermoélectrique et Thermomécanique d'Assemblages PCB Intégrants des Puces de Puissance », https://eduscol.education.fr/sti/si-ens-paris-saclay/ressources_pedagogiques/caracterisation-thermoelectrique-thermomecanique-assemblages-pcb-puces-puissance

[3] Matthieu Landel, « Technologie des transistors au nitrure de gallium », https://eduscol.education.fr/sti/si-ens-paris-saclay/ressources_pedagogiques/technologie-des-transistors-au-nitrure-de-gallium

Apport des convertisseurs multiniveaux modulaires aux véhicules électriques

La Revue 3E.I
Culture Sciences de l'Ingénieur

Gaël PONGNOT[1] - Anatole DESREVEAUX[1,3] - Clément MAYET[2]
Denis LABROUSSE[1,3] - Francis ROY[4] - Thomas PEUCHANT[5]

Édité le 22/01/2024

école normale supérieure paris—saclay

[1] *Université Paris-Saclay, ENS Paris-Saclay, CNRS, SATIE, F-91190, Gif-sur-Yvette, France*
[2] *Univ. Lille, Arts et Métiers Institute of Technology, Centrale Lille, Junia, ULR 2697 – L2EP, Lille, F-59000, France*
[3] *Le Cnam, Paris, F-75003, France, HESAM Université*
[4] *Stellantis*
[5] *Saft*

Cet article fait partie du N°111 de La Revue 3EI de janvier 2024. Les travaux présentés ont été réalisés dans le cadre du projet IBIS, un projet d'investissement d'avenir de l'ADEME.

Cet article expose le fonctionnement d'un véhicule électrique à batterie actuel, en particulier sa chaine de traction, et s'intéresse à une structure innovante fusionnant les batteries avec l'étage de conversion DC-AC pour créer un convertisseur multiniveau modulaire. Le principe de fonctionnement de ce convertisseur est détaillé, avec les enjeux associés à sa mise en œuvre. Un certain nombre d'atouts potentiels face aux véhicules actuels sont présentés, et une modélisation des pertes de la chaine de traction est proposée. Ce travail se base sur les recherches effectuées dans le cadre du projet IBIS, regroupant des industriels et des chercheurs.

1 - Développement des véhicules électriques

Les véhicules électriques (à batterie et hybrides rechargeables) représentent 21% des ventes en 2022 en Europe, part en constante augmentation. Cet élan est porté par l'essor de ces véhicules en Allemagne, au Royaume-Uni et en France notamment qui représentent une part importante du marché européen. Sur le marché français, 26% des véhicules vendus en 2023 étaient électriques, plus d'un quart du marché, dont 16.8% pour les véhicules électriques à batterie [1]. Le premier marché mondial restant la Chine (Figure 1), avec 14 millions de véhicules électriques en circulation cette même année, contre 8 millions en Europe [2].

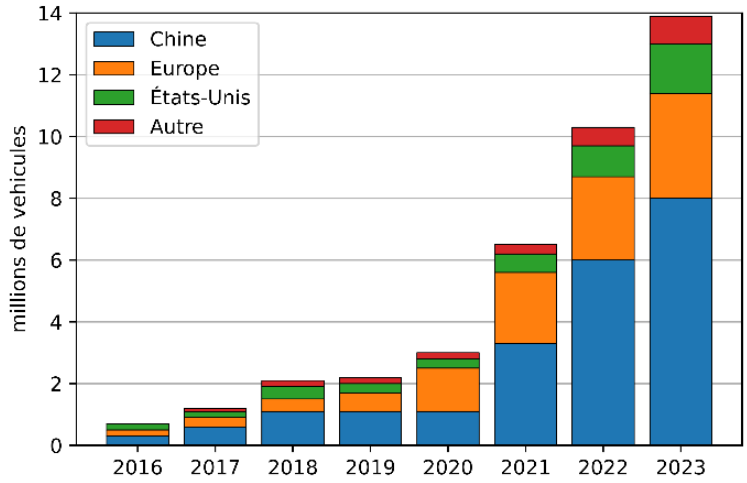


Figure 1 : Ventes annuelles de véhicules électriques dans le monde

Cependant une pénétration profonde du marché exige que les véhicules électriques (VE), et en particulier les véhicules électriques à batterie (VEB), aient une autonomie et un coût acceptables. Cela est permis, principalement, par l'augmentation de la densité énergétique des batteries grâce aux technologies Li-ion et à la maitrise croissante de leurs procédés de fabrication [3]. Un autre axe développé est l'amélioration de l'efficacité énergétique de l'ensemble de la chaine cinématique afin de réduire toujours plus la consommation des véhicules, l'effort étant porté par la recherche sur l'amélioration des chaines de traction électrique [4].

De plus, l'augmentation de la part des énergies intermittentes sur le réseau électrique conduit au nécessaire accroissement des capacités de stockage d'énergie [5]. Dans cette perspective, la création de centres de stockage stationnaire d'énergie dans des batteries commence à être mise en œuvre [6], [7]. En complément, les VEB pourraient alors être mis à contribution dans de futurs réseaux électriques communicants (Vehicule to Grid, V2G) [8]. Ce dernier point nécessite la réversibilité en puissance du chargeur embarqué, ce qui n'est pas toujours le cas.

Dans la suite de cet article, nous allons mettre de côté la recherche sur les technologies de batterie pour nous concentrer sur l'utilisation de l'énergie embarquée et les architectures de VEB associés à ces usages. En particulier, nous regarderons le cas d'un véhicule électrique conventionnel et celui d'un véhicule qui exploiterait une nouvelle architecture modulaire. Nous nous intéresserons à ses atouts en termes de performance, mais également de polyvalence et de résilience.

2 - Architecture des véhicules électriques à batterie

Un véhicule électrique doit répondre à deux principaux cas d'utilisation : le conducteur doit pouvoir le conduire et le recharger. La Figure 2 décompose la conduite en trois actions : accélérer, ralentir et diriger le véhicule, ceci est commun à tout type de véhicule. La recharge peut également être décomposée en trois modes : monophasée, triphasée ou continue, tous les véhicules ne sont pas compatibles avec tous les modes de recharge [9].

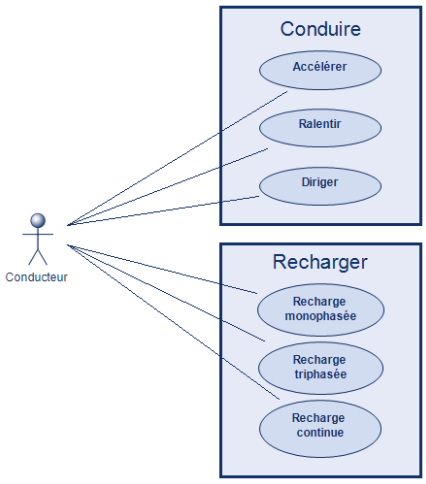


Figure 2 : Diagramme des cas d'utilisation d'un véhicule électrique

Dans un véhicule électrique à batterie conventionnel, on retrouve les éléments présentés à la Figure 3. L'énergie stockée dans la batterie est distribuée aux différents organes du véhicule par le biais d'un bus de tension continue. La traction et la récupération d'énergie au freinage sont assurées par l'ensemble : onduleur de traction, machine de traction et transmission mécanique.

Le réseau de bord (calculateurs, interface avec l'utilisateur, système multimédia, climatisation, etc…) est alimenté grâce à une batterie de servitude comme dans le cas d'un véhicule thermique. Cette batterie est rechargée, non pas avec un alternateur relié au moteur comme sur un véhicule thermique, mais via un convertisseur DC-DC qui fait le lien entre bus haute tension (batterie

principale) et bus basse tension (servitude). Enfin la recharge s'effectue soit à travers un chargeur embarqué (convertisseur AC-DC), soit en se connectant directement au bus haute tension (recharge rapide en courant continu).

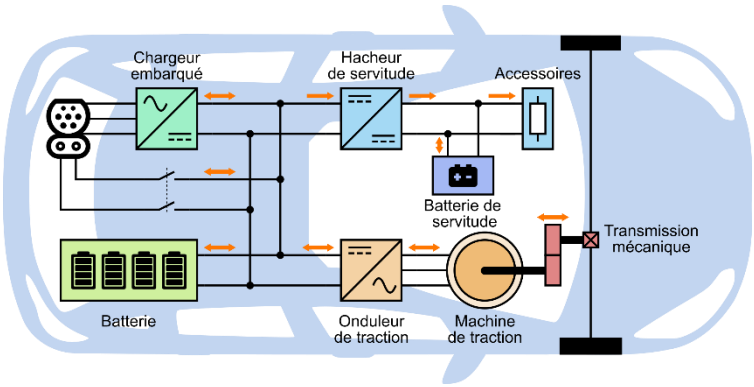
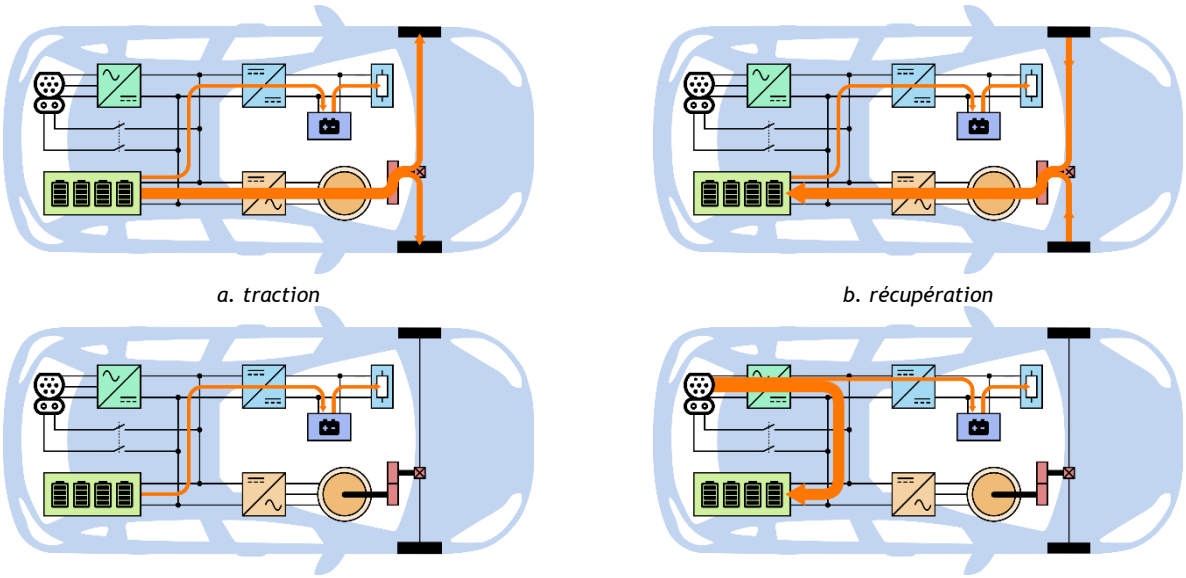


Figure 3 : Schéma synoptique d'un véhicule électrique conventionnel

2.1 - Transferts de puissance et mode d'utilisation

Le recours à un bus de tension permet de connecter des sources de puissance entre elles, ici on compte : un élément de stockage (batterie), un consommateur final (réseau de bord) et deux interfaces de transfert (réseau électrique et machine de traction). Selon les cas d'utilisation, les transferts de puissances différent, Figure 4.



a. traction *b. récupération*

c. statique *d. recharge*

Figure 4 : Transferts de puissance dans différents modes d'utilisation

En traction, la machine électrique fournit de la puissance aux roues pour mettre en mouvement le véhicule, Figure 4a. Lorsque le conducteur ralenti, les véhicules électriques ont la particularité de pouvoir récupérer une partie de l'énergie cinétique pour la stocker dans la batterie, Figure 4b. Cette récupération est possible, car toute la chaine énergétique (onduleur, machine, transmission) est réversible en puissance. Les freins mécaniques sont alors moins sollicités.

Le cas statique se superpose à ces deux cas de mouvement. Aucun transfert d'énergie n'est réalisé avec les roues, le seul transfert présent est dirigé vers le réseau de bord, nécessitant une puissance bien plus faible que la traction, mais de manière permanente, Figure 4c.

Enfin, en recharge AC, la majorité du flux de puissance traverse le chargeur embarqué pour atteindre la batterie, Figure 4d. En recharge DC, ce convertisseur ne perçoit plus le flux de

puissance qui arrive directement de la prise vers la batterie. Le recours au V2G est essentiellement envisagé dans le cas alternatif et nécessite que le chargeur embarqué soit réversible pour inverser le flux de puissance par rapport à la recharge, la batterie du véhicule fournit alors de l'énergie au réseau.

2.2 - Structure de la chaine de traction

L'utilisation essentielle d'un véhicule est la conduite. Dans ce cadre, l'essentiel du flux de puissance transite par la chaine de traction qui transforme l'énergie stockée sous forme chimique dans les batteries en énergie mécanique en sortie de machine en passant par le vecteur énergétique *électricité*. Les machines utilisées dans les véhicules étant pour la plupart triphasées [10], un onduleur de tension est nécessaire pour convertir la tension continue du bus batterie en tension triphasée alternative de fréquence et d'amplitude variable pour alimenter la machine électrique.

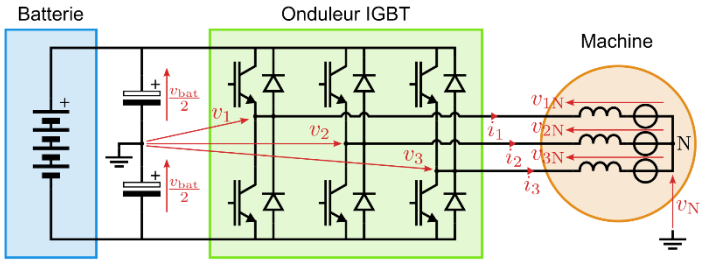


Figure 5 : Schéma électrique d'une chaine de traction conventionnelle

Mise en équation

Chaque bras d'onduleur dans une architecture triphasée classique comporte deux interrupteurs de puissance pilotés qui forment une cellule de commutation (Figure 5) : à tout instant, ces deux interrupteurs sont dans des états complémentaires, dépendant de la fonction de modulation $f_{\mathrm{m}\,i}$ binaire (pouvant prendre les valeurs 0 ou 1). Lorsque $f_{\mathrm{m}\,i} = 1$ le transistor du haut du bras i est passant et celui du bas est bloqué, conduisant ainsi à écrire :

$$v_i = V_{\mathrm{bat}}\left(f_{\mathrm{m}\,i} - \frac{1}{2}\right) \tag{1}$$

Les tensions perçues par la machine dépendent du potentiel du point neutre, qui s'exprime, dans le cas d'une machine équilibrée à neutre non relié :

$$v_N = \frac{v_1 + v_2 + v_3}{3} = V_{bat}\left(\frac{f_{m1} + f_{m2} + f_{m3}}{3} - \frac{1}{2}\right) \tag{2}$$

Ainsi on peut écrire la relation entre les tensions perçues par les enroulements de la machine et les ordres de commande des bras d'onduleur.

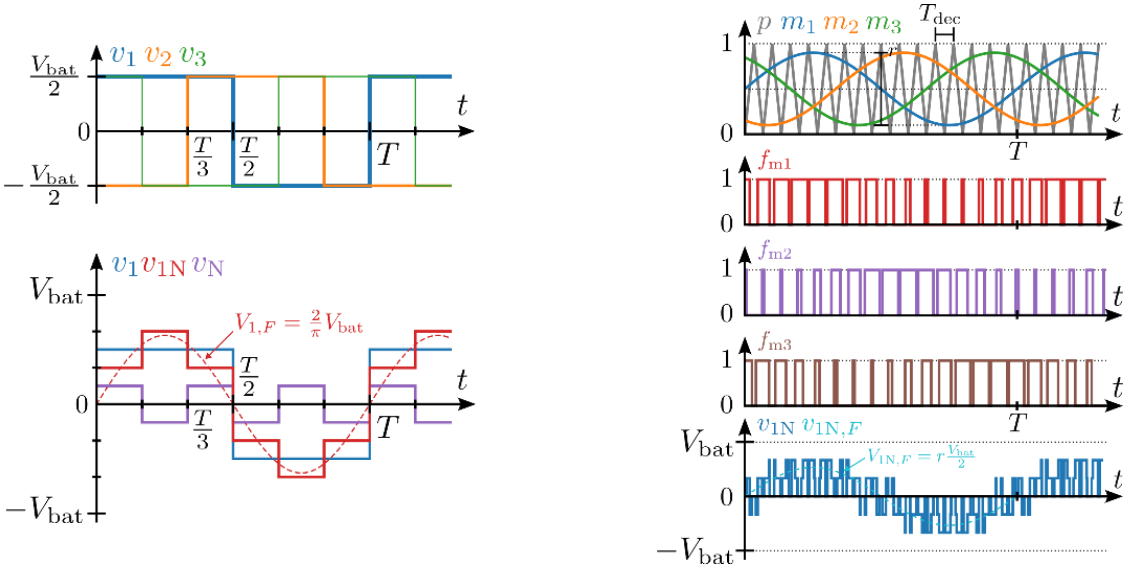
$$\begin{pmatrix} v_{1N} \\ v_{2N} \\ v_{3N} \end{pmatrix} = \frac{V_{bat}}{3}\begin{pmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{pmatrix}\begin{pmatrix} f_{m1} \\ f_{m2} \\ f_{m3} \end{pmatrix} \tag{3}$$

Les paragraphes suivants présentent trois stratégies de commande courantes pour ce type de structure : la commande pleine onde, la commande par modulation de largeur d'impulsion (MLI) et la commande vectorielle.

Commande pleine onde

Dans le mode de fonctionnement, les interrupteurs sont alternativement bloqués ou passants pendant la moitié de la période T. Les tensions v_i sont alors des créneaux de rapport cyclique 50%

et d'amplitude $\frac{V_{bat}}{2}$, déphasés deux à deux de $\frac{2\pi}{3}$. De par l'équation (3), les harmoniques multiples de 3 ne sont pas présents dans les tensions v_{iN}, ce qui amène aux formes d'onde présentées par la Figure 6a. Le fondamental des tensions v_{iN} a pour valeur $V_{1,F} = \frac{2}{\pi} V_{bat}$ [11].



a. pleine onde *b. MLI intersective*

Figure 6 : Allure des tensions pour des commandes scalaires

La tension produite a un spectre très riche en basse fréquence. Celui-ci contient les harmoniques de $F = 1/T$ non-multiples de 3, ce qui a des effets néfastes sur le rendement de la machine. En outre, la fréquence est bien réglable, mais l'amplitude de la tension de sortie ne l'est pas, elle est proportionnelle à la tension du bus continu. C'est pourquoi cette solution n'est pas retenue dans le cadre des véhicules, car elle nécessiterait l'ajout d'un convertisseur DC-DC en amont.

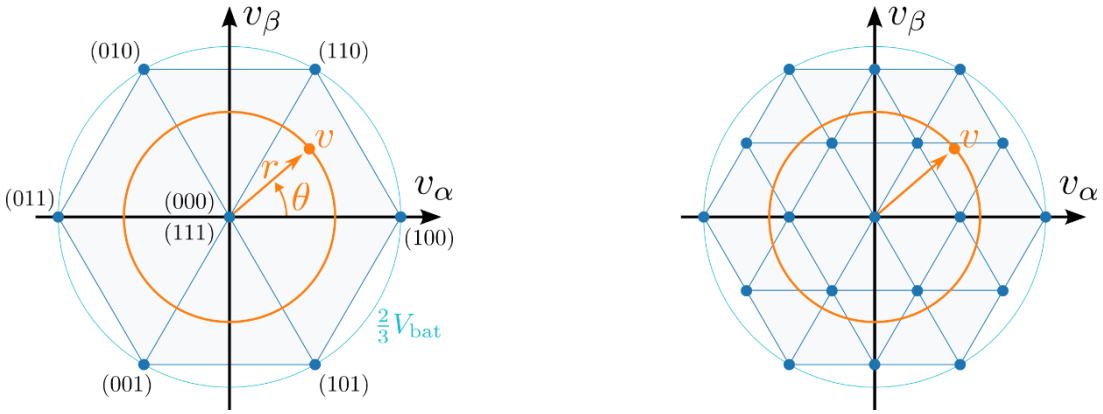
Commandes par modulation de largeur d'impulsion

La modulation de largeur d'impulsion (MLI) d'une cellule de commutation peut être obtenue par comparaison d'un signal triangulaire (porteuse $\boldsymbol{p}$) qui détermine la fréquence de découpage avec le signal de modulation $\boldsymbol{m_i}$, on parle alors de MLI intersective [12]. La Figure 6b présente le cas d'une MLI centrée (porteuse triangulaire symétrique).

Le spectre de la tension de sortie est riche en contenu haute fréquence, autour de $F_{dec} = 1/T_{dec}$ et ses multiples, qui sont plus facilement filtrables. L'amplitude du fondamental de cette tension est donnée par $V_{1,F} = r\frac{V_{bat}}{2}$ avec $r \in [0,1]$. L'amplitude et la fréquence sont alors réglable, avec une amplitude maximale de $V_{1,F}^{max} = \frac{V_{bat}}{2}$.

Commandes vectorielles

Les commandes précédentes traitent chaque phase indépendamment, elles sont alors qualifiées de scalaires. Les commandes vectorielles utilisent la transformation de Clarke (4) ou de Concordia pour représenter les tensions dans le plan complexe α, β. Les 8 configurations possibles de l'onduleur triphasé délimitent alors 6 secteurs, comme le montre la Figure 7a, deux configurations conduisant à la tension nulle. La transition d'un point à un voisin ne nécessite la commutation que d'un seul bras.



a. Onduleur 2 niveaux *b. Onduleur 3 niveaux*

Figure 7 : Principe de la commande vectorielle

Pour synthétiser la tension complexe souhaitée, une pondération des trois points délimitant le secteur est calculée pour faire correspondre leur barycentre avec le point désiré. Ces pondérations correspondent aux fractions de période de commutation à consacrer à chaque état [13].

$$x_{\alpha\beta o} = \frac{2}{3}\begin{pmatrix} 1 & -\frac{1}{2} & -\frac{1}{2} \\ 0 & \frac{\sqrt{3}}{2} & -\frac{\sqrt{3}}{2} \\ \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \end{pmatrix} x_{123} \tag{4}$$

Le régime linéaire de la commande vectorielle se limite au cercle inscrit à l'hexagone, Figure 7a. Dès lors, l'amplitude de tension maximale possible avec cette commande est de $V_{1,F}^{\max} = \frac{V_{\text{bat}}}{\sqrt{3}}$.

Dans le cadre d'onduleur multiniveau, la constellation de points s'enrichit et le nombre de secteurs augmente, Figure 7b. Si bien que lorsque le nombre de niveaux est suffisant, il est possible de s'affranchir du découpage pour se placer sur le point le plus proche de l'objectif avec une marge d'erreur acceptable. Cette commande est appelée par vecteur le plus proche (NVC).

Perspective technologique

La plupart des onduleurs de tractions sont aujourd'hui basés sur des transistors IGBT silicium. Or les transistors à base de silicium (Si) font aujourd'hui face à leurs limites physico-chimiques. Des technologies émergentes apparaissent, regroupées sous le nom de matériaux à large bande interdite (WBG : *Wide-Band Gap*) [14]. La bande interdite est une caractéristique quantique des semi-conducteurs, son augmentation repousse les limites physiques actuelles, en particulier en termes de résistance à l'état passant, Figure 8.

Notamment, des transistors à base de carbure de silicium (SiC) et de nitrure de gallium (GaN) sont aujourd'hui sur le marché. Les transistors SiC sont cependant dans un état de développement plus avancé que les GaN. L'utilisation de tels transistors octroie une amélioration des rendements des convertisseurs [15].

La technologie est encore peu répandue, en particulier à cause d'un coût élevé. Malgré cela, les transistors SiC sont déjà dans les véhicules. Tesla a fait le choix du SiC dès 2017 pour son Model 3, cela avec une augmentation espérée de l'autonomie du véhicule de 10% [16].

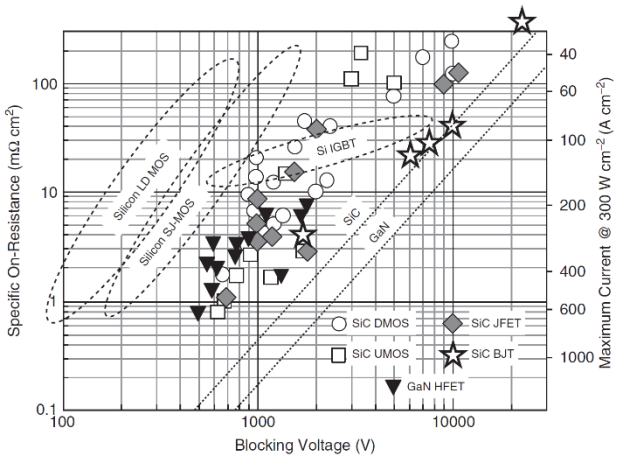


Figure 8 : Limites théoriques des résistances à l'état passant en fonction de la tension de claquage, issue de [14]

En parallèle, certains constructeurs envisagent d'augmenter la tension de batterie, pour passer d'une tension aujourd'hui autour de 400 V à 800 V. Cela afin d'améliorer le rendement de la chaine, mais surtout d'accélérer la recharge rapide des véhicules.

3 - Convertisseur multiniveau : fusion des batteries et de l'onduleur

Face à ces dynamiques orientées sur les technologies nouvelles et l'augmentation de la tension, une autre stratégie est envisageable : concevoir un convertisseur multiniveau modulaire sur la base de transistors silicium basse tension et de faible coût.

3.1 - Le projet IBIS

Les convertisseurs multiniveaux modulaires (MMC) sont exploités aujourd'hui sur les réseaux pour les liaisons haute tension en courant continu (France-Espagne notamment). L'idée d'utiliser cette technologie dans les véhicules électriques a émergé à la fin du XXe siècle [17], mais elle a trouvé son réel intérêt en l'associant à la technologie des cellules Li-ion.

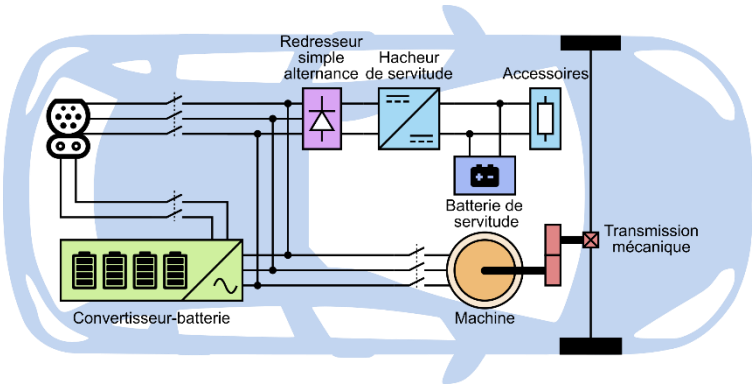


Figure 9 : Schéma synoptique d'un véhicule électrique IBIS

Dans les années 2010, les recherches et les idées développées autour de ces structures dans les laboratoires ont rencontré les acteurs industriels qui ont trouvé un intérêt potentiel dans cette nouvelle approche. C'est ainsi que le projet IBIS (*Intelligent Battery-Inverter System*) est né en associant deux grands acteurs industriels (Stellantis et Saft), quatre laboratoires de recherche (Satie, GeePs, LEPMI et Institut Lafayette) et deux PME (E2CAD et Sherpa Engineering). Ce projet, en partie financé par l'ADEME, permet un transfert technologique des laboratoires vers l'industrie de par l'étude et la réalisation de prototypes, jusqu'à une potentielle commercialisation dans les années à venir [18], [19].

La particularité de la structure, détaillée juste après, est de fusionner le convertisseur DC-AC avec la batterie, créant ainsi un convertisseur-batterie (IBIS) qui distribue une tension triphasée pour le véhicule. Ce système peut alors être directement connecté à la machine de traction ou au réseau électrique selon le cas d'utilisation. L'emploi de ce même système est envisagé par Saft pour le stockage stationnaire d'énergie pour le réseau. Celui-ci pourrait utiliser les batteries du véhicule en seconde vie.

3.2 - Principe de fonctionnement

La structure étudiée est un convertisseur multiniveau modulaire de type « onduleur à ponts en H cascadés », avec la particularité que les sources de tension isolées sont ici des batteries Li-ion de faible tension. Le schéma électrique présenté à la Figure 10 illustre le cas du prototype actuellement en fonctionnement.

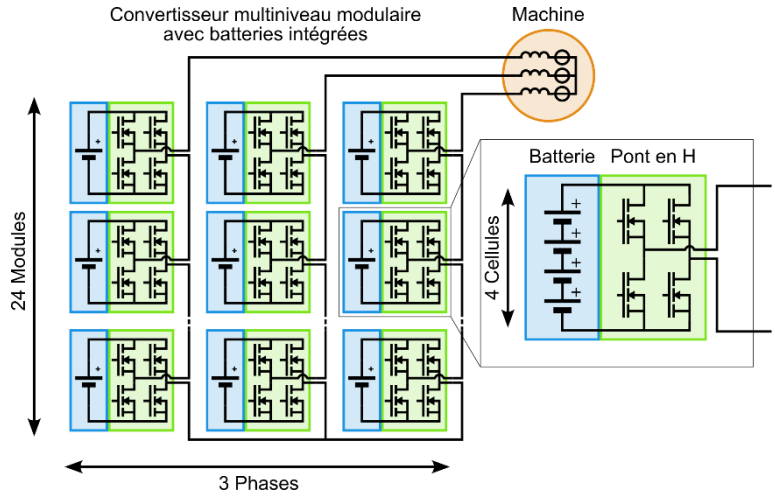


Figure 10 : Schéma électrique d'une chaine de traction à convertisseur multiniveau modulaire

Chaque module peut fournir trois niveaux de tension différents en fonction de l'état du pont en H. En notant $N = 4$ le nombre de cellules de batterie dans un module, $M = 24$ le nombre de modules dans une phase, $V_{\text{cell}} = 3{,}3\ \text{V}$ la tension d'une cellule de batterie et $u_{p,m} \in \{-1{,}0{,}1\}$ l'ordre de commande du module, avec $m \in \{1 \dots M\}$ de la phase $p \in \{1{,}2{,}3\}$, on exprime alors la tension de sortie d'un module :

$$v_{p,m} = u_{p,m}\, N\, V_{\text{cell}} \tag{5}$$

Les modules d'une phase sont reliés en série, ce qui conduit à :

$$v_p = N V_{\text{cell}} \sum_{m=1}^{M} u_{p,m} \tag{6}$$

Comme précédemment, il est possible de réaliser une commande pleine onde. Pour cela, tous les modules d'une même phase sont dans le même état ($\forall m, u_{p,m} = u_{p,1}$). Les créneaux auront ici une amplitude plus élevée qu'une structure classique comportant une batterie ayant le même nombre de cellules, mais en série ($S = N\ M$). En effet, la structure présentée est capable de produire une tension $\pm V_{\text{bat}} = \pm N\ M\ V_{\text{cell}}$, contrairement à un onduleur classique limité aux tensions $\pm \frac{V_{\text{bat}}}{2}$.

Pour exploiter les différents niveaux du convertisseur, plusieurs stratégies existent [20] : les commandes MLI, la commande par niveau le plus proche (NLC) et la commande par vecteur le plus proche (NVC). Dans le cadre du projet, la commande NLC a été retenue pour son faible nombre de commutations par période et sa simplicité de mise en œuvre.

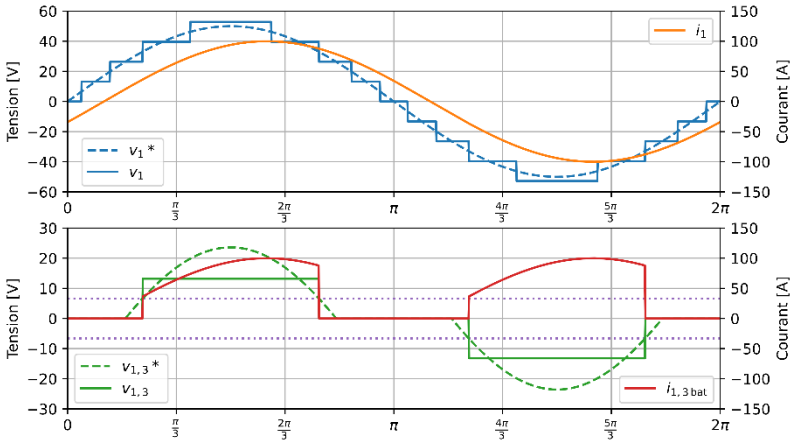


Figure 11 : Exemple de mise en œuvre de la commande NLC

La Figure 11 présente une mise en œuvre de la commande NLC pour une tension de consigne v_1^* d'amplitude 50 V et un courant sinus i_1 de 100 A d'amplitude. L'analyse de la tension de sortie du 3e module sollicité $v_{1,3}$ illustre bien les 3 états possibles, on dénombre alors 4 commutations par période électrique. Ces transitions sont déterminées en comparant la tension de référence du module $v_{1,3}^*$ avec $\pm NV_{\text{cell}}/2$. Le courant traversant les cellules $i_{p,m\,\text{bat}}$ est quant à lui fluctuant : nul, égal ou opposé au courant de phase selon l'état du module.

Onduleur triphasé classique			Architecture IBIS		
Nombre de cellules	$S = 120$ $P = 2$	$V_{\text{bat}} = S\,V_{\text{cell}}$ $= 396$ V	Nombre de cellules	$M = 20$ $N = 4$	$V_{\text{bat}} = MN\,V_{\text{cell}}$ $= 264$ V
Commande pleine onde	$\frac{2}{\pi}V_{\text{bat}}$	252 V	Commande pleine onde	$\frac{4}{\pi}V_{\text{bat}}$	336 V
Commande MLI	$\frac{V_{\text{bat}}}{2}$	198 V	Commande NLC	V_{bat}	264 V
Commande vectorielle	$\frac{V_{\text{bat}}}{\sqrt{3}}$	227 V	Commande NVC	$\frac{2\,V_{\text{bat}}}{\sqrt{3}}$	305 V

Tableau 1 : Comparaison des tensions fondamentales simples maximales

La comparaison d'IBIS avec un onduleur triphasé classique associé à une batterie de capacité similaire est présentée au Tableau 1. Le pack batterie classique est une architecture série-parallèle avec S et P les nombres de cellules en série et en parallèle. Le nombre total de cellules embarquées est $SP = 3MN = 240$. Dans cette configuration, la structure permet d'accéder à des niveaux de tension plus élevés et donc de diminuer le courant de phase à puissance équivalente. Cela est dû principalement à l'utilisation de ponts en H, qui permettent une tension de sortie des modules bidirectionnelle. En contrepartie, le nombre de cellules en série sollicitées pour produire une tension intermédiaire est plus faible, ce qui conduit à l'augmentation du courant traversant ces cellules. Ces différences de niveaux de tension ont une incidence sur le point de fonctionnement couple-vitesse de la machine électrique et par conséquent cela nécessite une adaptation de la chaine de traction.

3.3 - Degrés de liberté et communications

Là où les configurations possibles pour un onduleur classique sont au nombre de 8, elles sont de 3^{3M} soit environ 10^{30} dans le cas $M = 24$. Cela conduit à un nombre impressionnant de configurations microscopiques possibles permettant d'obtenir la même configuration macroscopique. Que faire de

ces possibilités ? On peut commencer par gérer indépendamment chaque phase et imposer une règle simple : tous les ordres de commandes $u_{p,m}$ d'une même phase sont de même signe.

Les degrés de liberté peuvent ensuite être utilisés pour équilibrer les états de charges des modules entre eux et réguler la température des modules. Ce nombre impressionnant de configurations, associé à la nécessité d'une intelligence centralisée, requiert des communications à haut débit qui sont un enjeu majeur de la mise en œuvre d'une telle structure.

3.4 - Équilibrage des cellules

L'équilibrage des cellules est réalisé en sollicitant en priorité les modules les plus chargés en phase de traction, et les moins chargés en phase de régénération [21]. Cela passe par une gestion centralisée des états de charge des modules par le calculateur central. Ce dernier trie les modules selon cette valeur et détermine les modules à activer en fonction de la tension de consigne et du signe de la puissance.

Deux stratégies sont alors possibles, s'intéresser à la puissance moyenne ou la puissance instantanée. Cette dernière permet de mieux équilibrer les cellules, mais au prix d'une augmentation du nombre de commutations. Une fois la stratégie déterminée, la quantité d'informations à transmettre est ainsi réduite et le débit requis est diminué.

3.5 - Recharge

Une des particularités de ce système est la suppression du chargeur embarqué. De fait, avec un nombre de cellules de batterie adapté, il est possible pour IBIS de se connecter directement à un réseau électrique triphasé domestique (230 V en tension simple). C'est cet aspect qui intéresse particulièrement les acteurs du stockage d'énergie pour le réseau. La recharge monophasée ou en courant continu nécessite une légère reconfiguration, mais reste possible : les trois « phases » sont alors connectées en série.

4 - Estimation des pertes

Afin de comparer la pertinence de la structure IBIS pour une utilisation automobile, la comparaison aux onduleurs deux niveaux IGBT et SiC est cruciale. Le critère pertinent est alors la consommation sur cycle, qui semble être à l'avantage d'IBIS dans l'article de Chang [15], bien que la commande utilisée soit une MLI. Le calcul des pertes par commutation dans le cas NLC est plus complexe et nécessite la connaissance des instants de commutation. Le calcul analytique étant trop complexe, la méthode présentée passe par le développement d'un simulateur numérique permettant de prendre en compte les différentes dynamiques électriques [22].

4.1 - Modélisations

La modélisation du système et sa commande sont organisées à l'aide du formalisme REM : Représentation Énergétique Macroscopique [23], Figure 12. La chaine de conversion d'énergie est alors décomposée en blocs représentant les fonctions énergétiques des différents éléments physiques selon les principes d'interaction (action et réaction) et de causalité (intégrale).

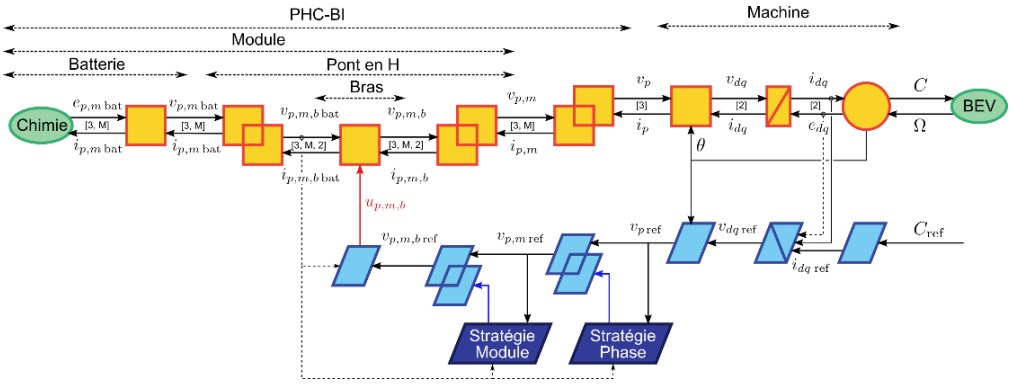


Figure 12 : Représentation énergétique macroscopique de la chaine de traction

Batterie

Les batteries au Lithium sont des systèmes électrochimiques complexes encore difficilement modélisés. On représente généralement le comportement dynamique des batteries à l'aide de réseau RC série, avec une résistance statique. Ce modèle permet de décrire suffisamment fidèlement le comportement des cellules.

$$v_{p,m\,bat} = e_{p,m\,bat} - R_{bat} i_{p,m\,bat} \tag{7}$$

$$e_{p,m\,bat} = N V_{cell} \tag{8}$$

Des études ont montré que la prise en compte des comportements dynamiques conduit à une amélioration du rendement dans le cas des ponts en H cascadés (PHC) grâce à la composante basse fréquence du courant [24]. Pour simplifier la simulation, nous considérons ici un modèle statique (7), car les dynamiques chimiques sont lentes devant les dynamiques électriques. Cela conduit à une surestimation des pertes estimée à environ 20% [24].

Électronique de puissance

Les ponts en H sont constitués de deux bras à base de MOSFET Si basse tension. Ces quatre transistors induisent des pertes par conduction et des pertes par commutation. Un transistor MOSFET à l'état passant se comporte comme une résistance notée $\mathrm{R_{mos}}$. Cette résistance dépend fortement de la température, ce qui complexifie le calcul des pertes par conduction en introduisant un couplage thermique. Ici, la température de fonctionnement est considérée comme constante et égale à celle d'un régime permanent cible : 80°C.

$$v_{p,m,b} = u_{p,m,b}\, v_{p,m,b\,\mathrm{bat}} - R_{\mathrm{mos}}\, i_{p,m,b} \tag{9}$$

La prise en compte des pertes par commutation est complexe dans le cas des MOSFET, contrairement aux IGBT les constructeurs ne fournissent pas de caractéristique des énergies dissipées. Des modèles analytiques existent, mais impliquent un grand nombre de paramètres [25]. Cependant, la commande NLC implique une forte réduction du nombre de commutations par période électrique. Les pertes induites deviennent alors négligeables face aux pertes par conduction. Elles ne seront pas prises en compte dans le modèle.

Machine

La machine considérée est une machine synchrone à aimants permanents et à pôles saillants. Des simulations par éléments finis (EF), non détaillées ici, fournissent les informations caractéristiques de la machine ainsi que la loi de commande. Le comportement de la machine est alors exprimé dans Park par les équations suivantes. Les effets dynamiques sont calculés à partir des inductances L_d et L_q déterminées par les simulations EF statiques, en fonction des courants.

$$\begin{cases} L_{\mathrm{d}}(i_{\mathrm{d}}, i_{\mathrm{q}})\, i_{\mathrm{d}} = \int (v_{\mathrm{d}} - e_{\mathrm{d}} - R_{\mathrm{Cu}}(\omega)\, i_{\mathrm{d}})\, dt \\ L_{\mathrm{q}}(i_{\mathrm{d}}, i_{\mathrm{q}})\, i_{\mathrm{q}} = \int (v_{\mathrm{q}} - e_{\mathrm{q}} - R_{\mathrm{Cu}}(\omega)\, i_{\mathrm{q}})\, dt \end{cases} \quad (10)$$

Les pertes cuivre sont modélisées par une résistance en série avec les enroulements $R_{\mathrm{Cu}}(\omega)$. Cette résistance présente une composante continue, représentative des pertes en courant continu, et une composante quadratique avec la vitesse ω, caractéristique de l'effet de peau dans les conducteurs. Elle intervient dans l'équation des inductances, représentées sur la Figure 12 par l'élément d'accumulation d'énergie.

Les simulations EF fournissent les flux magnétiques Φ_{d} et Φ_{q} en fonction des courants i_{d} et i_{q}. Ces flux conduisent aux expressions des forces électromotrices (fem) e_{d} et e_{q}, et du couple électromagnétique C_{em} :

$$\begin{cases} e_{\mathrm{d}} = -\omega \Phi_{\mathrm{q}}(i_{\mathrm{d}}, i_{\mathrm{q}}) \\ e_{\mathrm{q}} = +\omega \Phi_{\mathrm{d}}(i_{\mathrm{d}}, i_{\mathrm{q}}) \end{cases} \quad (11)$$

$$C_{\mathrm{em}} = n_{\mathrm{p}}\left(\Phi_{\mathrm{d}}(i_{\mathrm{d}}, i_{\mathrm{q}})\, i_{\mathrm{q}} - \Phi_{\mathrm{q}}(i_{\mathrm{d}}, i_{\mathrm{q}})\, i_{\mathrm{d}}\right) \quad (12)$$

$$\omega = n_p\, \Omega \quad (13)$$

La prise en compte des pertes fer s'effectue via un couple résistant C_{Fe} calculé pour un point de fonctionnement couple-vitesse. Cet élément est particulièrement dépendant à la loi de commande.

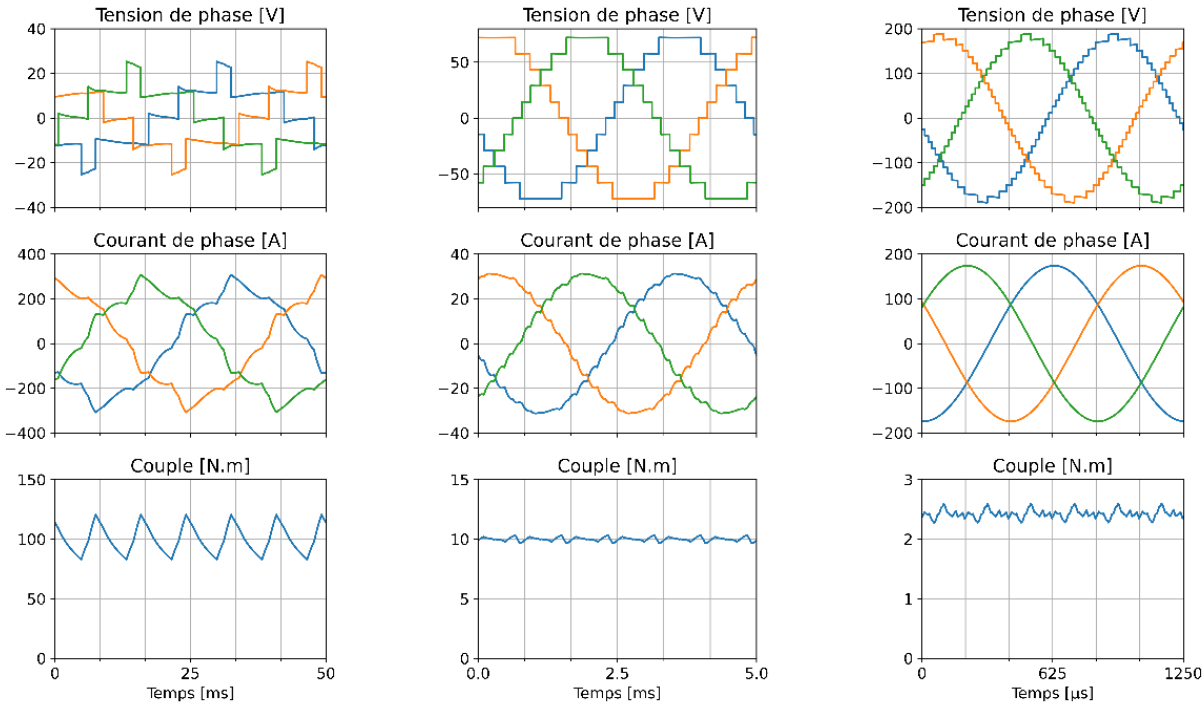
4.2 - Simulations

Les modèles présentés sont simulés à l'aide de Matlab/Simulink sur différents points de fonctionnement mécaniques couple-vitesse. La loi de commande de la machine a été déterminée, grâce à des simulations par élément finis, afin de minimiser l'amplitude du courant (MTPA : *Maximum Torque per Ampere*). Elle fournit les courants de références $i_{\mathrm{d\,ref}}$ et $i_{\mathrm{q\,ref}}$ en fonction du couple de référence C_{ref} et de la vitesse mesurée Ω.

L'inversion de l'élément d'accumulation représentant les inductances de la machine est réalisée à l'aide d'un correcteur PI pour asservir les courants. Enfin la suite de la chaine de commande est réalisée par inversion de modèle en utilisant la commande NLC. La Figure 13 présente les formes d'onde obtenues pour trois points de puissance identiques (3 kW), mais avec une répartition couple-vitesse différente [22].

La Figure 13a illustre un point de fonctionnement à basse vitesse et fort couple. Dans cette configuration, le nombre de modules utilisés est réduit : ici deux modules pour cinq niveaux de tension. Lorsque la vitesse augmente, la tension augmente également, conséquence de la fem, et donc le nombre de niveaux de tension s'accroît, Figure 13b et c. Cela induit une amélioration du contenu spectral du courant, et donc une réduction de l'ondulation de couple.

La puissance apparente est plus élevée sur la Figure 13c. Cela est dû à la stratégie de commande de la machine qui impose un défluxage, injection de courant i_{d} en opposition aux aimants, à haute vitesse. L'important courant et le faible nombre de niveaux rendent visible l'effet des pertes par conduction sur la Figure 13a. Le fort courant qui traverse les résistances séries des batteries et des transistors déforme les paliers de tensions qui ne sont plus constants comme espérés. Cet effet est moins marqué sur la Figure 13b.



a. 300 tr/min et 100 N.m *b. 3000 tr/min et 10 N.m* *c. 12000 tr/min et 2,5 N.m*

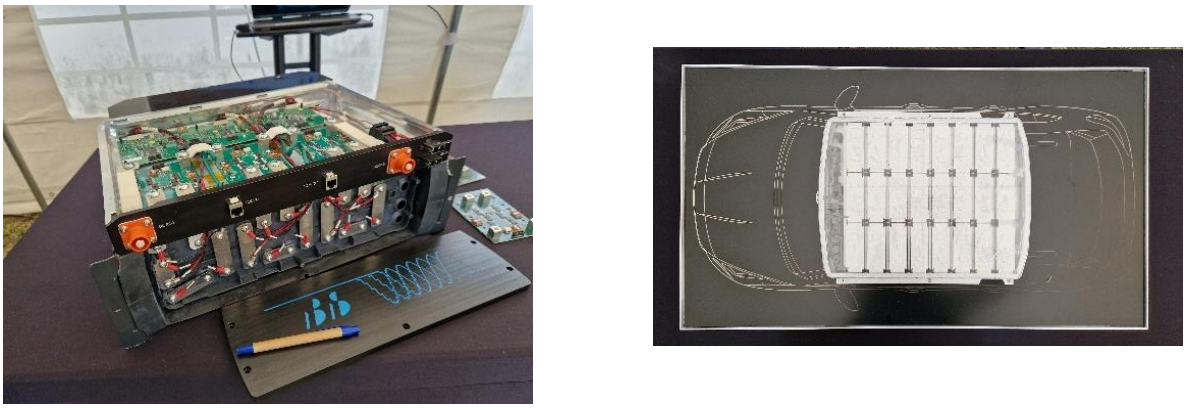
Figure 13 : Formes d'onde simulées

5 - Perspectives de développement

5.1 - Le futur d'IBIS

Un démonstrateur statique a été réalisé et a permis de démontrer la faisabilité technique du système. Ce système modulaire, dont on peut voir un module sur la Figure 14a, est installé au laboratoire GeePs. Des essais de connexion au réseau (recharge) et de traction/récupération avec une machine de traction réelle (conduite) y sont réalisés.

Un démonstrateur du véhicule est en cours de conception et devrait être opérationnel à l'été 2024. Son objectif sera de démontrer la faisabilité de l'intégration d'un tel système dans un véhicule et d'étudier son comportement en conditions réelles. C'est une étape supplémentaire vers la commercialisation d'un premier véhicule à l'horizon 2030 [19].



a. phase 1, démonstrateur statique *b. phase 2, démonstrateur véhicule*

Figure 14 : Prototypes expérimentaux

5.2 - Résilience, maintenance et réutilisation

La force de la modularité du système est dans sa résilience. En cas de défaut d'une cellule de la batterie, le véhicule est immobilisé et le pack batterie doit être entièrement remplacé. Avec IBIS, le module contenant la cellule défectueuse est isolé (bypass) et le véhicule peut continuer à fonctionner (en mode légèrement dégradé) jusqu'à un passage au garage.

La maintenance est également rationalisée puisqu'il suffit alors de remplacer uniquement les modules défectueux ou en fin de vie. Dans une application véhicule, les cellules sont considérées en fin de vie lorsqu'elles ont perdu 20% de leur capacité initiale, la densité d'énergie stockée devient alors trop faible pour cette utilisation. Cependant la capacité restante demeure suffisante pour des applications où l'encombrement est un enjeu secondaire : le stockage sur le réseau. Ainsi, les mêmes modules pourront être réutilisés (seconde vie) dans ce type d'installation.

Figure 15 : Vue d'artiste d'un parc de stockage d'énergie connecté au réseau électrique, issue de [26]

5.3 - Hybridation

La structure en pont en H cascadé n'impose pas la nature des sources de tension continue. On peut donc imaginer mélanger des cellules de batterie de natures différentes (forte densité de puissance ou d'énergie) ou introduire d'autre technologies de stockage comme des supercondensateurs. Une hybridation avec une pile à combustible est également envisageable. Sur des installations stationnaires, des panneaux photovoltaïques peuvent être intégrés comme source d'alimentation de certains modules [27].

Conclusion

Les véhicules électriques sont en plein développement à l'échelle mondiale, mais font face dans leurs améliorations aux limites imposées par les systèmes existants : batterie, électronique de puissance, machine, etc. Certains constructeurs s'orientent vers les transistors WBG dans l'espoir d'améliorer les rendements, mais ceux-ci restent coûteux. Une autre voie est d'utiliser des convertisseurs multiniveaux modulaires en fusionnant la batterie et l'étage de conversion DC-AC. Cette voie, explorée par Stellantis dans le cadre du projet IBIS, semble prometteuse en termes de performance, mais elle comporte également des atouts en matière de soutenabilité.

Remerciements

Cette étude a été effectuée dans le cadre du projet IBIS, lequel est financé par l'Agence de l'Environnement et de la Maîtrise de l'Energie (ADEME) grâce au programme d'Investissement d'Avenir (PIA).

Références

[1] A. Aloisi, « Plus d'un quart des véhicules vendus en 2023 en France étaient électriques ! », Electroniques. Consulté le: 8 janvier 2024. [En ligne]. Disponible sur: https://www.electroniques.biz/auto-train/plus-dun-quart-des-vehicules-vendus-en-2023-en-france-etaient-electriques/

[2] IEA, « Global EV Outlook 2023 : Trends in electric light-duty vehicles », IEA. Consulté le: 8 janvier 2024. [En ligne]. Disponible sur: https://www.iea.org/energy-system/transport/electric-vehicles

[3] V. Sauvant-Moynot, F. Orsini, et A. Juton, « État de l'art et perspectives des batteries de voitures électriques », *La Revue 3EI*, nº 99, janvier 2020.

[4] X. Rain, « Les solutions actuelles de motorisations pour véhicules électriques », *La Revue 3EI*, nº 99, janvier 2020.

[5] RTE, « Futurs énergétiques 2050 _ rapport complet.pdf », févr. 2022.

[6] TotalEnergies, « Stockage par batteries : nos projets et réalisations », TotalEnergies. Consulté le: 8 janvier 2024. [En ligne]. Disponible sur: https://totalenergies.com/fr/projets/electricite/stockage-batteries-nos-projets-realisations

[7] B. Crépin, « Stockage d'énergie par batteries : « On est au début d'une nouvelle ère » », Techniques de l'Ingénieur, Énergie. Consulté le: 8 janvier 2024. [En ligne]. Disponible sur: https://www.techniques-ingenieur.fr/actualite/articles/stockage-denergie-par-batteries-on-est-au-debut-dune-nouvelle-ere-111535/

[8] ENEDIS, « Pilotage de la recharge de véhicules électriques », ENEDIS, déc. 2020. Consulté le: 8 janvier 2024. [En ligne]. Disponible sur: https://www.enedis.fr/sites/default/files/documents/pdf/enedis-rapport-pilotage-de-la-recharge-de-vehicules-electriques.pdf

[9] C. Saber et N. Rouhana, « Chargeurs de batteries de véhicule électrique », *La Revue 3EI*, nº 99, janvier 2020.

[10] A. Juton *et al.*, *Technologies des voitures électriques: Motorisations, batteries, hydrogène, interactions réseau*, Illustrated édition. Malakoff: Dunod, 2021.

[11] J.-P. Louis et C. Bergmann, « Commande numérique des machines - Systèmes triphasés : régime permanent », *Tech. Ing. Convers. Lénergie Électr.*, nov. 1996, doi: 10.51257/a-v1-d3642.

[12] H. Foch, F. Forest, et T. Meynard, « Onduleurs de tension - Structures. Principes. Applications », *Tech. Ing. Convers. Lénergie Électr.*, p. 21, 1998.

[13] P.-É. Vidal, B. Trajin, et F. Rotella, « Stratégie et technique pour le pilotage en modulation des convertisseurs statiques », *Tech. Ing. Électronique*, nº E3969, déc. 2019, doi: 10.51257/a-v1-e3969.

[14] T. Kimoto et J. A. Cooper, *Fundamentals of Silicon Carbide Technology: Growth, Characterization, Devices and Applications*, vol. 9781118313527. in Fundamentals of Silicon Carbide Technology: Growth, Characterization, Devices and Applications, vol. 9781118313527. 2014, p. 538. doi: 10.1002/9781118313534.

[15] F. Chang, O. Ilina, M. Lienkamp, et L. Voss, « Improving the Overall Efficiency of Automotive Inverters Using a Multilevel Converter Composed of Low Voltage Si mosfets », *IEEE Trans. Power Electron.*, vol. 34, nº 4, p. 3586-3602, avr. 2019, doi: 10.1109/TPEL.2018.2854756.

[16] C. Morris, « Silicon carbide is becoming the material of choice for EV power electronics », Charged EVs. Consulté le: 4 septembre 2023. [En ligne]. Disponible sur: https://chargedevs.com/newswire/silicon-carbide-is-becoming-the-material-of-choice-for-ev-power-electronics/

[17] L. M. Tolbert, F. Z. Peng, et T. G. Habetler, « Multilevel converters for large electric drives », *IEEE Trans. Ind. Appl.*, vol. 35, nº 1, p. 36-44, janv. 1999, doi: 10.1109/28.740843.

[18] ENS Paris-Saclay, « Projet IBIS, une coopération exemplaire pour des batteries révolutionnaires | ENS-PARIS-SACLAY », ENS Paris-Saclay. Consulté le: 10 janvier 2024. [En ligne]. Disponible sur: https://ens-paris-saclay.fr/actualite/projet-ibis-une-cooperation-exemplaire-pour-des-batteries-revolutionnaires

[19] J. Leblanc, « Batterie IBIS. Stellantis travaille sur des voitures électriques plus simples et plus durables », L'argus. Consulté le: 28 août 2023. [En ligne]. Disponible sur: https://www.largus.fr/actualite-automobile/batterie-ibis-stellantis-travaille-sur-des-voitures-electriques-plus-simples-et-plus-durables-30028369.html

[20] L. G. Franquelo, J. Rodriguez, J. I. Leon, S. Kouro, R. Portillo, et M. A. M. Prats, « The age of multilevel converters arrives », *IEEE Ind. Electron. Mag.*, vol. 2, nº 2, p. 28-39, juin 2008, doi: 10.1109/MIE.2008.923519.

[21] C. Mayet, D. Labrousse, A. Dittrick, B. Revol, R. Bkekri, et F. Roy, « Simulation and Control of a New Integrated Battery System for Automotive Applications », in *PCIM Europe digital days 2021*, mai 2021, p. 1-6.

[22] G. Pongnot, C. Mayet, et D. Labrousse, « Répartition des pertes dans une chaîne de traction utilisant un onduleur à ponts en H cascadés avec batteries intégrées », in *Symposium de Génie Électrique*, Lille, France, juill. 2023. Consulté le: 18 octobre 2023. [En ligne]. Disponible sur: https://hal.science/hal-04158082

[23] W. Lhomme, P. Delarue, A. Bouscayrol, et P. Barrade, « La REM, formalisme multiphysique de commande de systèmes énergétiques », *Tech. Ing. Convers. Lénergie Électr.*, nº D3066, nov. 2014, doi: 10.51257/a-v1-d3066.

[24] O. Theliander, A. Kersten, M. Kuder, W. Han, E. A. Grunditz, et T. Thiringer, « Battery Modeling and Parameter Extraction for Drive Cycle Loss Evaluation of a Modular Battery System for Vehicles Based on a Cascaded H-Bridge Multilevel Inverter », *IEEE Trans. Ind. Appl.*, vol. 56, nº 6, p. 6968-6977, nov. 2020, doi: 10.1109/TIA.2020.3026662.

[25] D. Christen et J. Biela, « Analytical Switching Loss Modeling Based on Datasheet Parameters for mosfets in a Half-Bridge », *IEEE Trans. Power Electron.*, vol. 34, nº 4, p. 3700-3710, avr. 2019, doi: 10.1109/TPEL.2018.2851068.

[26] J.-C. Lefebvre et V. Gillot, « IBIS : Stellantis et Saft dévoilent une 'batterie intelligente' et plus efficace pour les véhicules électriques et le stockage stationnaire », Stellantis. Consulté le: 8 janvier 2024. [En ligne]. Disponible sur: https://www.stellantis.com/fr/actualite/communiques-de-presse/2023/july/ibis-stellantis-et-saft-devoilent-une-batterie-intelligente-et-plus-efficace-pour-les-vehicules-electriques-et-le-stockage-stationnaire

[27] C. Sirico *et al.*, « PV Module-Level CHB Inverter with Integrated Battery Energy Storage System », *Energies*, vol. 12, nº 23, Art. nº 23, janv. 2019, doi: 10.3390/en12234601.

Caractérisation Thermoélectrique et Thermomécanique d'Assemblages PCB Intégrants des Puces de Puissance

La Revue 3E.I
Culture Sciences de l'Ingénieur

Édité le 29/01/2024

école normale supérieure paris—saclay

Mounira BOUARROUDJ[1,2] – Mickaël PETIT[1] – Said BENSEBAA[1,4]
Stéphane LEFEBVRE[1] – Nicolas SCHMITT[2, 3]

[1] *SATIE, ENS Paris-Saclay, CNRS, UCP, Cnam, 91190, Gif-sur-Yvette*
[2] *Université Paris Est Créteil UPEC, 94000 Créteil, France*
[3] *LMPS, ENS Paris-Saclay, 91190, Gif-sur-Yvette, France*
[4] *Synchrotron SOLEIL, Gif-sur-Yvette, France*

Cet article fait partie du N°111 de La Revue 3EI de janvier 2024.

Un procédé d'intégration PCB (printed ciruit board) basé sur l'utilisation d'une mousse métallique pressée pour assurer le contact électrique des puces de puissance est présenté. Ce procédé permet de réduire l'inductance parasite qui est liée aux connectiques (fils de bonding), il permet également de réduire les contraintes mécaniques dans la puce comparée à un assemblage avec prise de contact face arrière par brasure. Le principal avantage de ce procédé est la simplicité de sa réalisation et son faible coût.

Cet article propose une description du procédé de fabrication suivie d'une partie de caractérisation électrique et thermique des différents contacts et matériaux utilisés dans l'assemblage. À l'issue de cette étude un choix du matériau approprié est effectué (Mousse Nickel et/ou cuivre), enfin des résultats de tenue au cyclage passif sont présentés.

1 - Introduction

Les modules de puissance standards utilisant les fils de bonding, figure 1, pour la prise de courant présentent certaines limitations, à savoir : i) l'inductance parasite (de l'ordre de 5 à 20 nH) qui génère des surtensions lors des commutations, particulièrement pour les composants à grand gap (GaN), ii) une limitation thermique, où la dissipation de la chaleur se fait uniquement par la face arrière, iii) en terme de fiabilité, ces fils présentent une zone de fragilité limitant ainsi la durée de vie de ces modules.

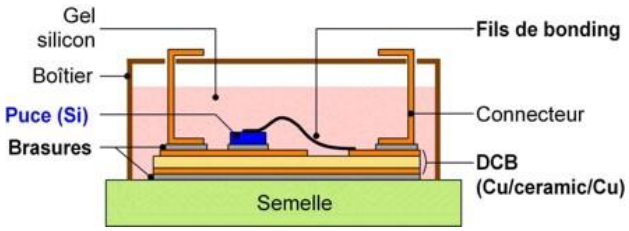


Figure 1 : Assemblage montrant les connectiques (fil de bonding en face avant et brasure en face arrière) de la puce dans un assemblage de puissance standard.

L'enfouissement des puces au cœur du PCB semble être une alternative intéressante en vue de remédier aux contraintes énoncées précédemment. Dans ce cas, l'amenée de courant se fait généralement par des Vias en cuivre réalisés par électro déposition et la puce est brasée sur le substrat PCB [1][2]. L'utilisation de la mousse métallique pour assurer le contact supérieur de la puce a été vérifiée et peut être une alternative simple et pas chère [3]. Dans cet article, nous

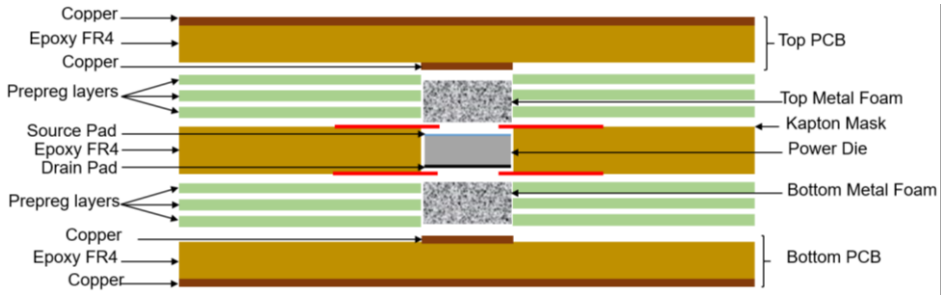
proposons un assemblage pour les puces enfouis au cœur du PCB avec une prise de contact supérieur et inférieur assurée par des mousses métalliques pressées sans l'utilisation de brasure. Nous présentons des travaux de caractérisations électriques et thermiques de l'assemblage ainsi que des résultats de tests de vieillissement thermiques passifs selon les normes PCB [MIL-STD-883E].

2 - Procédé d'intégration

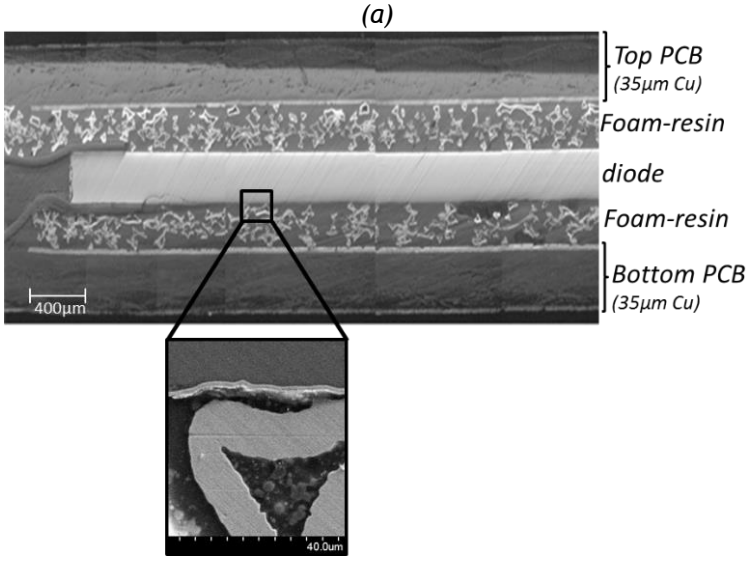
L'assemblage proposé comprend une diode enfouie entre deux mousses métalliques au cœur du PCB. Ces mousses assurent le contact entre les faces avant et arrière de la puce. Le procédé peut être décrit en deux étapes. La première étape, figure 2.a, consiste à effectuer un empilement de bas en haut des couches suivantes :

- PCB inferieur (Bottom PCB),
- Mousse métallique (Contact inferieur), entourée par des couches de Prepreg,
- Puce,
- Mousse métallique (Contact supérieur) entourée par des couches de Prepreg,
- PCB supérieur (Top PCB).

La deuxième étape consiste à laminer cet empilement de matériaux sous des conditions de pression et de température bien spécifiées (des conditions fournis par le fabricant des couches de Prepreg, voir figure 3). Cette étape est effectuée dans une presse (LPKF MultiPress S).



(a)



(b)

Figure 2 : a) Schémas du procédé d'intégration proposé, b) Coupe microscopique d'une diode intégrée au cœur du PCB, avec zoom sur la zone de contact, observation au MEB (Microscope Electronique à Balayage).

La mousse utilisée sur la face avant ne doit pas court-circuiter la métallisation de la puce et l'anneau de garde. Pour cela un masque isolant en Polyimide est ajouté, il permet de spécifier la surface de contact entre mousse et puce, il peut également d'être utilisé pour isoler la grille et la source dans le cas des transistors. Pour des raisons de symétrie, le masque est également ajouté sur la face arrière de l'assemblage. La figure 2.b présente le schéma de l'assemblage ainsi qu'une micro-section de la zone de contact entre la mousse et puce (observation au MEB).

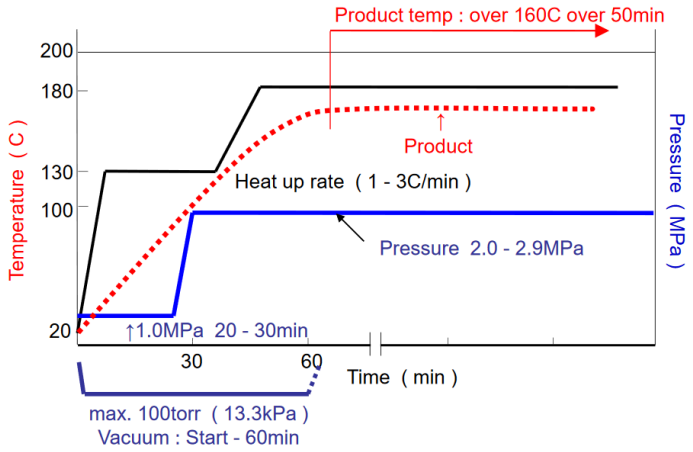


Figure 3 : Recommandations générales sur les conditions de pression et de température appliqués au niveau de la machine de pressage [4].

3 - Caractérisation électrique et thermique

Après laminage, la résine contenue dans le Prepreg pénètre à l'intérieur de la mousse, générant ainsi un nouveau matériau composite (mousse-résine). Les propriétés physiques de ce mélange doivent être déterminées si l'on souhaite estimer la robustesse et les contraintes induites dans cet assemblage. La qualité des contacts entre les différentes couches doit être également évaluée. La démarche et les résultats de ces travaux de caractérisations sont détaillés dans cette section.

3.1 - Caractérisation thermique

Dans un premier temps, nous avons cherché à caractériser les propriétés thermiques du matériau composite obtenu après laminage (constitué d'un mélange de mousse pressée et de la résine). Dans un second temps, nous avons estimé les résistances thermiques de contact.

3.1.1 - Protocole de mesure

Le banc d'essais utilisé est illustré sur la figure 4 [5], où l'échantillon à caractériser est placé entre deux plaques de cuivre (cuivre supérieur et inférieur) à l'intérieur desquelles une sonde de température PT100 est insérée. Cet ensemble est monté sur un cylindre à base de duralumin ayant une résistance thermique connue $R_{th\_duralumin}$ (préalablement calibrée). Ce cylindre comporte également deux sondes de température PT100 pour la mesure des températures aux extrémités ($T_{Top-dur}$ *et* $T_{bottom-dur}$). Cela permet ainsi de mesurer le flux de chaleur qui le traverse. Le dispositif est considéré thermiquement isolé. Sous cette hypothèse, le flux de puissance, P, mesuré dans le cylindre est égal à celui qui traverse l'échantillon. Il peut être exprimé par l'équation (1).

$$P = \frac{T_{Top-dur} - T_{bottom-dur}}{R_{th\_duralium}} \qquad (1)$$

Connaissant la puissance qui traverse l'échantillon ainsi que la température entre ses bornes (en haut et en bas des plaques de cuivre), on pourra alors calculer la résistance thermique de l'échantillon avec l'équation (2).

$$R_{th\_measured} = \frac{T_{top-copper} - T_{bottom-copper}}{P} \qquad (2)$$

Pour assurer un meilleur transfert de chaleur entre les plaques de cuivre et l'échantillon, des interface thermique (TIM) sont insérées. Par conséquent, et afin d'obtenir uniquement la résistance de l'échantillon, les résistances thermiques de ces TIM ont d'abord été caractérisées puis soustraites de la résistance mesurée, voir équation (3).

$$R_{th\_sample} = R_{th\_measured} - 2.R_{th\_TIM} \quad (3)$$

Les résistances thermiques des plaques de cuivre sont négligées (0,0015 et 0,0025 K/W) devant celles de l'échantillon à caractériser. Les résultats sont enregistrés directement sur MATLAB Simulink©, avec un contrôle instantané de la puissance et de la température dans le dispositif.

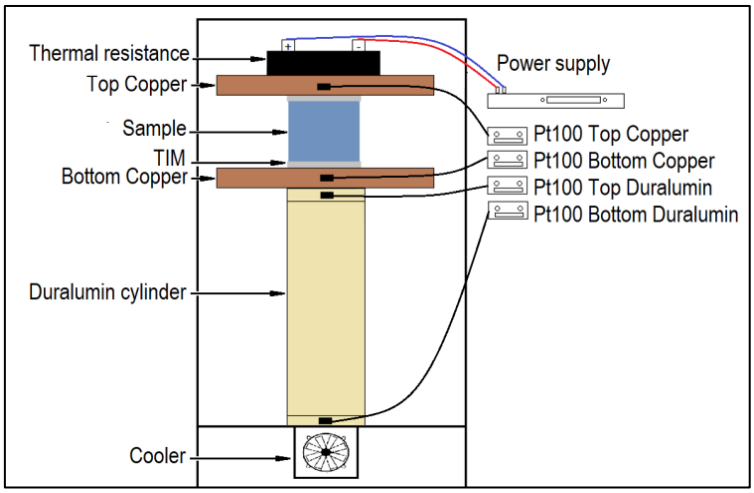


Figure 4 : Schémas du dispositif de mesure de la résistance thermique.

3.1.2 - Caractérisation thermique du mélange mousse-résine

La conductivité thermique du mélange mousse-résine dépend de plusieurs paramètres structurels, tels que la porosité, la densité relative et la taille des pores de la mousse. Afin d'évaluer l'effet de la proportion de la mousse dans le mélange mousse-résine sur la conductivité thermique, des échantillons sont réalisés avec différentes épaisseurs. L'épaisseur finale de l'échantillon dépend de la pression appliquée lors du pressage, plus elle augmente plus la densité de la mousse dans le mélange augmente. Avant pressage, la densité moyenne initiale de la mousse (solide / air) est connue, ce qui permet de déduire le volume de mousse (V_{foam}) et le volume d'air (V_{air}). Lors du pressage, la résine remplace l'air, donc le volume de résine qui pénètre dans la mousse est égal au volume d'air. Lorsqu'une pression est appliquée sur l'échantillon, le volume de mousse (V_{foam}) est toujours le même, mais le volume de résine diminue (V_{resin}). Ainsi, le rapport entre le volume de mousse (V_{foam}) et le volume total (V_{foam} + V_{resin}) du mélange mousse-résine augmente. Cette fraction volumique (X), est donnée par l'équation (4).

$$X = \frac{V_{foam}}{V_{foam} + V_{resin}} \quad (4)$$

La conductivité thermique λ mesurée est ainsi exprimée en fonction de la fraction volumique X. Les résultats obtenus pour le mélange mousse Cuivre-résine sont représentés sur la figure 5.

La conductivité thermique du mélange augmente avec l'augmentation de la fraction volumique, ce qui est logique, car la conductivité thermique du mélange augmente lorsque la densité du cuivre augmente. Les résultats obtenus sont comparés avec le modèle de conductivité thermique "Bhattachaya" d'une mousse à pores ouverts [6], équation (5).

$$\lambda_{eff} = M \cdot \left(zK_f + (1-z) \cdot K_s\right) + \frac{1-M}{\left(\frac{z}{K_f} + \frac{1-z}{K_s}\right)} \quad (5)$$

Où K_s et K_f représentent la conductivité thermique du métal et du fluide respectivement, z la proportion du métal dans le mélange et M est à déterminer expérimentalement.

Néanmoins, on observe que pour une forte proportion de cuivre (X_{cu}=0.7), la conductivité thermique mesurée est très faible (4.5 W/mK) par rapport à celle du cuivre solide (400 W/mK). Cela est probablement dû à la résistance thermique des contacts entre les différentes couches de l'assemblage.

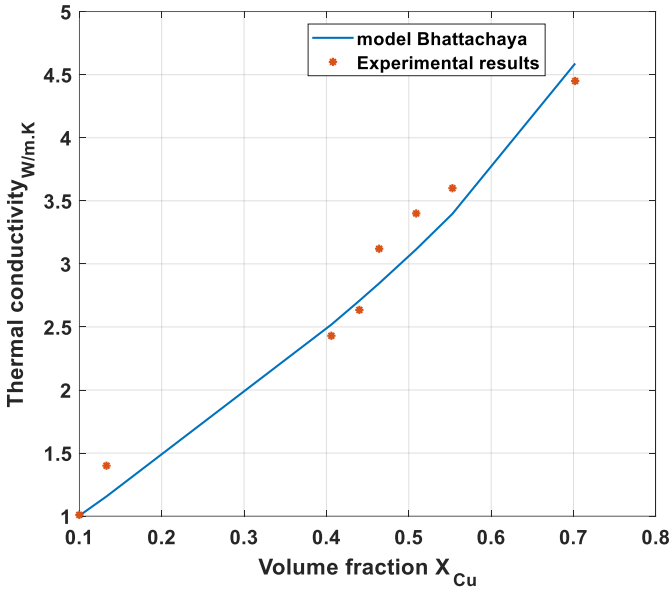


Figure 5 : Conductivité thermique λ du mélange mousse Cuivre-résine en fonction de la fraction volumique (X_{cu}).

3.1.3 - Caractérisation thermique du contact Mousse-Puce

La deuxième partie de la caractérisation thermique consiste à étudier le contact thermique entre la puce et la mousse. Pour cela, la résistance thermique $R_{th(j\text{-}case)}$ (jonction-refroidisseur) de l'assemblage est mesurée. Pour effectuer cette mesure, on polarise la puce, cette dernière dissipe une puissance (*P*). La température du refroidisseur est mesurée avec une sonde de température PT100 et celle de jonction est obtenue en utilisant un paramètre électrique thermosensible (TSEP), basé sur la tension directe V_f de la diode à l'état passant. En effet, il existe une dépendance linéaire entre la tension directe et la température de jonction au niveau de la puce (environ -2mV/°C pour les composants à base de Silicium). Ainsi, pour obtenir l'image de la température de jonction, un faible courant est injecté (2 mA) dans la puce et la mesure de V_f fournit l'image de la température (figure 6).

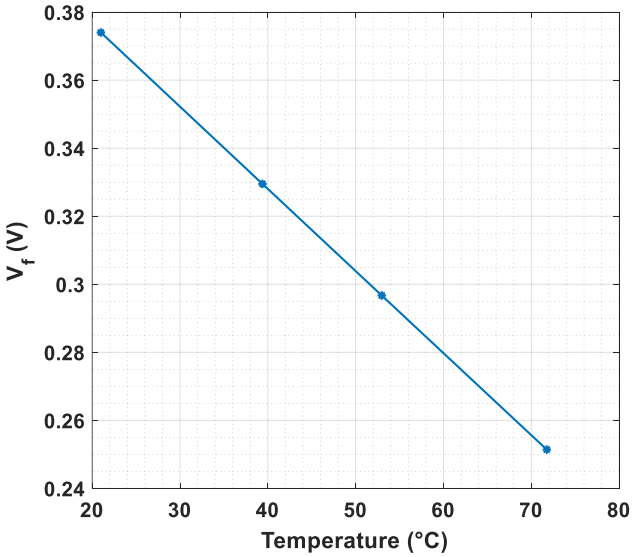
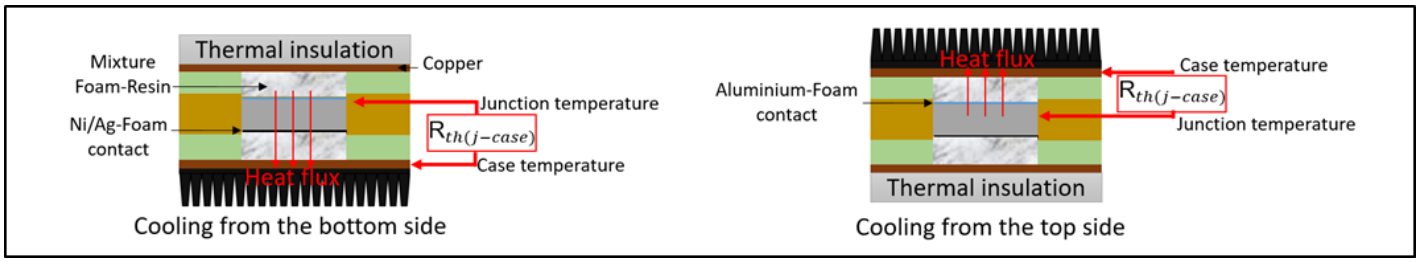


Figure 6 : Résultats de caracterisation de la tension de seuil V_f (evolution de V_f en fonction de la temperature T).

Une calibration préalable de l'évolution de la tension V_f en fonction de la température est réalisée en utilisant un conditionneur thermique. Les résultats de caractérisation sont donnés par l'équation (6).

$$\frac{\Delta V_f}{\Delta T} = -2{,}412\ mV.K^{-1} \qquad (6)$$

La diode possède deux terminaisons différentes, à savoir une métallisation en aluminium sur la face supérieure et une terminaison nickel sur la face inférieure. Le refroidissement par face avant ou arrière peut donc conduire à des résultats différents. Nous avons cherché à caractériser le transfert thermique par les deux faces. Le dispositif de mesure est présenté dans la figure 7. La diode est polarisée en direct et dissipe une puissance contrôlée avec MATLAB Simulink©. Dans un premier temps, le refroidissement se fait par la face avant afin de caractériser le contact supérieur (Al/mousse). Dans un second temps il se fait par la face inférieure pour caractériser le contact Ni/mousse.



(a) *(b)*

Figure 7 : Dispositif de mesure de Rth (j-case), (a) Refroidissement par face arrière, (b) refroidissement par face avant

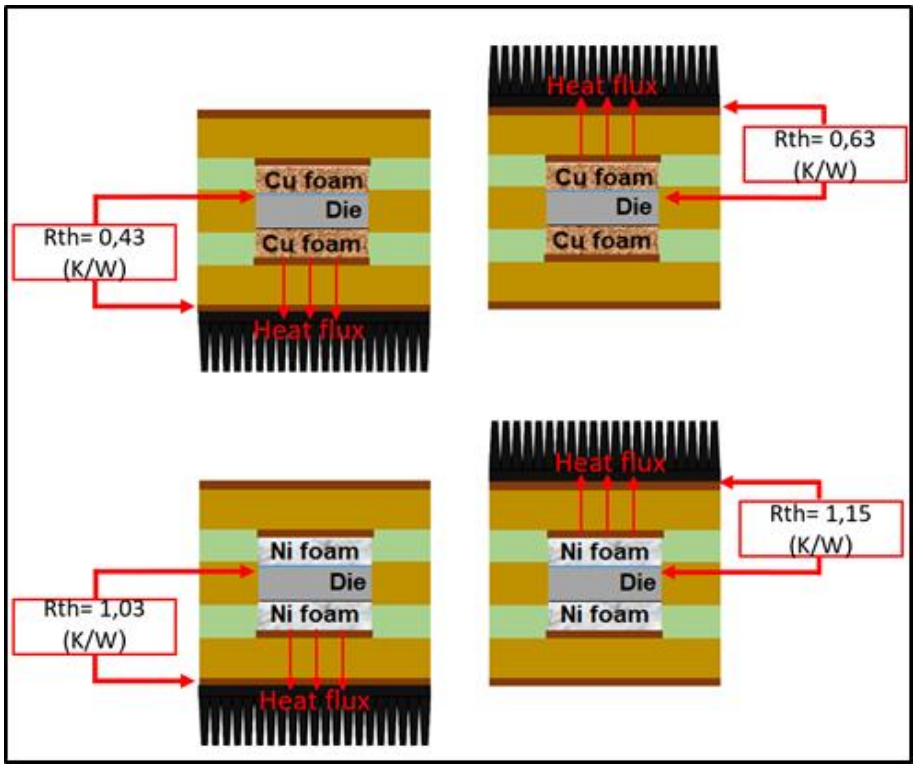


Figure 8 : Résultats de mesure de la résistance thermique $R_{th(j\text{-}case)}$

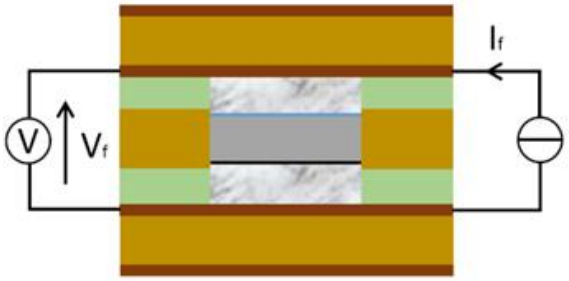
Comme indiqué sur la (figure 8), la résistance thermique mesurée $R_{th(j\text{-}case)}$ est plus faible lorsqu'on refroidit par la face arrière et cela quelle que soit la nature de la mousse (cuivre ou nickel). Ce résultat peut être lié à la nature des contacts entre mousse et puce (Ni/Ag face arrière, Al face avant) et/ou à la présence d'anneaux de garde sur la face avant de la puce, ce qui réduit la surface de transfert de chaleur entre la puce et la mousse. De plus, l'utilisation de la mousse de cuivre

permet de réduire la résistance thermique par un facteur de 2 par rapport à la mousse nickel. Ainsi, d'après ces résultats, pour maximiser le transfert de chaleur, il est préférable de refroidir par la face arrière et d'utiliser une mousse pressée à base de cuivre.

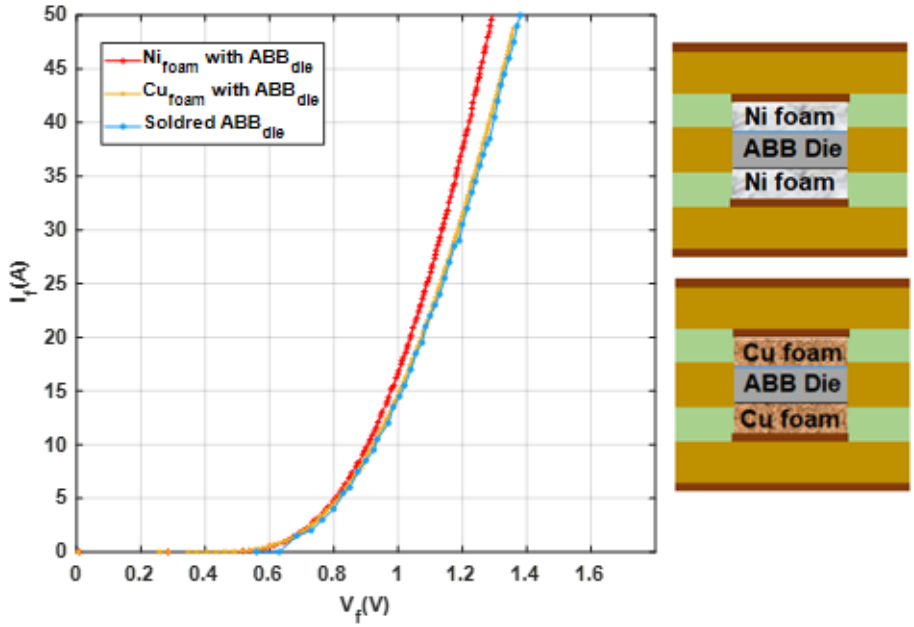
3.2 - Caractérisation électrique

Les caractéristiques statiques directes des diodes intégrées ont été tracées à l'aide d'un traceur de courbe Tektronix 371 (mesure 4 fils), (figure 9.a). La résistance électrique à l'état passant de l'assemblage représente la pente de la courbe $I_F = f(V_F)$ pour un courant donné (dans notre cas 40 A, voir figure 9.b).

Dans un premier temps, des assemblages utilisant le même type de mousse sur les deux faces ont été caractérisés (soit Nickel/Nickel ou Cuivre/Cuivre). Les résultats obtenus sont également comparés aux assemblages utilisant une mousse en Nickel sur la face avant et une brasure sur la face arrière [5]. Ces derniers montrent une résistance électrique équivalente à celle des assemblages utilisant une mousse en cuivre sur les deux faces.



(a)



(b)

Figure 9 : (a) Circuit de mesure 4 fils, (b) caractéristiques statiques directs $I_F = f(V_F)$ des assemblages utilisant uniquement de la mousse (Cuivre ou Nickel) et ceux utilisant mousse Nickel face avant et brasure face arrière.

Les résultats obtenus sont résumés dans le Tableau 1, ils montrent que les assemblages qui utilisent la mousse en Nickel sur les deux faces possèdent une résistance plus faible comparée à celle obtenue dans le cas des assemblages qui utilisent la mousse en Cuivre. Ce résultat est contradictoire puisque le cuivre est moins résistif que le nickel. On présume alors que le problème n'est pas lié au type de mouse, mais probablement à la nature des contacts aux interfaces entre mousse et puce.

Tableau 1 : Résistances directes moyennes des assemblages (calculées pour un courant de 40 A)

	Echantillons avec mousse Nickel des deux faces	Echantillons avec mousse Cuivre des deux faces	Echantillons avec mousse Nickel et brasure
Résistance (mΩ)	7,4	9,09	9

Dans un second temps, la résistance électrique de contact entre la mousse métallique et les deux faces de la puce est étudiée (métallisation Aluminium sur la face avant et métallisation Nickel sur la face arrière). Pour cela, on a réalisé des assemblages où la puce a été remplacée par un métal ayant les mêmes dimensions que la puce : i) Métal en Al qui correspond à la métallisation avant de la puce et ii) un métal en Al nickelé, ce qui correspond à la métallisation arrière de la puce. La couche de revêtement doit être fine, afin d'avoir approximativement la même résistance électrique pour de l'aluminium et de l'aluminium nickelé (afin d'être représentatif de la terminaison arrière d'une puce).

Dans ce cas, la résistance mesurée représente la somme des résistances du métal, des mousses et des différents contacts entre mousse et métal, tel qu'indiqué dans la figure 10. Les courbes I=f(V) obtenues pour les différents assemblages réalisés sont présentées sur la figure 11.

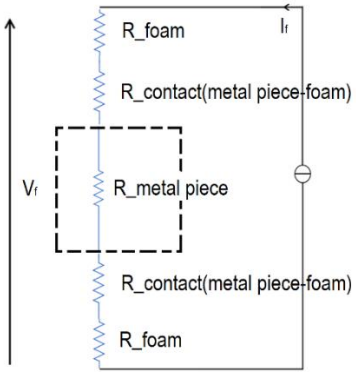


Figure 10 : Différentes résistances dans l'assemblage.

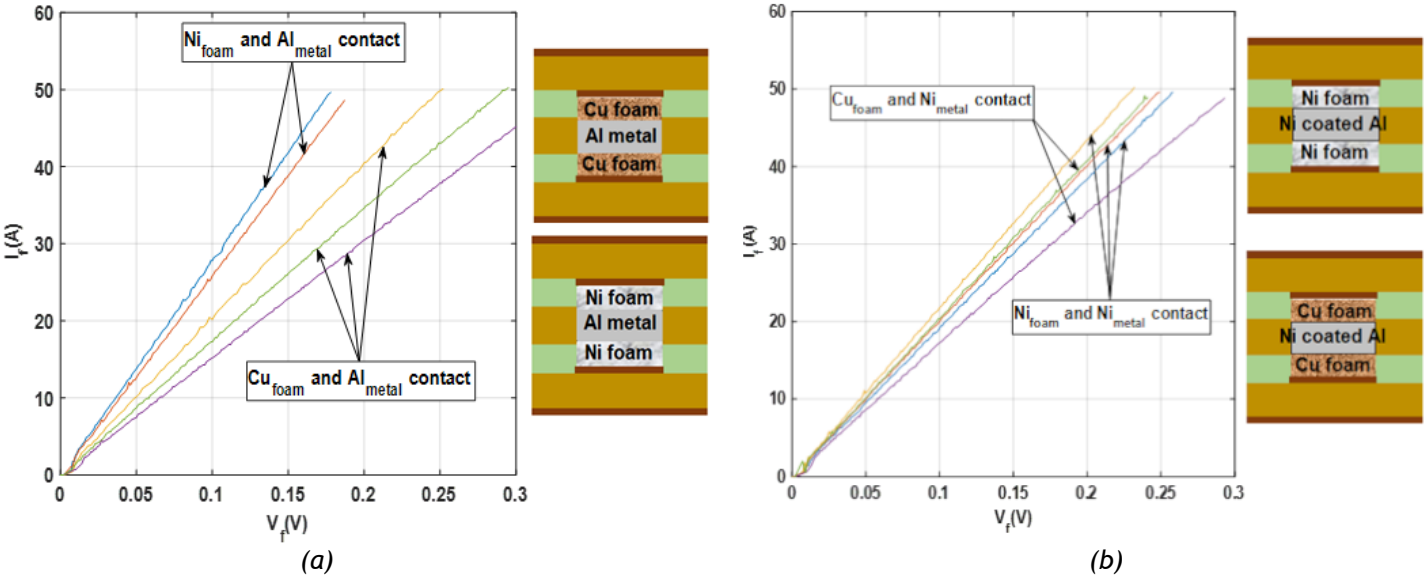


Figure 11 : (a) Caractéristique statique direct des contacts Cu_{foam}-$Al_{métal}$ et Ni_{foam}-$Al_{métal}$, (b) Caractéristique statique direct des contacts Ni_{foam}-$Ni_{métal}$ et Cu_{foam}-$Ni_{métal}$.

Les résultats sont résumés dans le Tableau 2, ils indiquent que le contact $Cu_{foam}/Al_{métal}$ possède une résistance électrique plus élevée, environ 1,5 fois, que le contact $Ni_{foam}/Al_{métal}$. De ce fait, il est préférable d'utiliser la mousse Nickel sur la face avant de la puce (terminaison Al). La résistance électrique du contact lors de l'utilisation de mousse en cuivre ou nickel avec un métal d'aluminium

nickelée ($Ni_{métal}$) est similaire, alors au niveau de la face arrière de la puce (terminaison Ni/Ag) on peut utiliser indifféremment une mousse en cuivre ou en nickel.

Tableau 2 : Résistance obtenue pour les différents assemblages (calculée à un courant de 40 A)

Package		Résistance (mΩ/cm²)
Face avant de la puce (métal Al)	Contact Ni - Al	2,18
	Contact Cu - Al	2,84 to 3,55
Face arrière de la puce (métal Ni)	Contact Ni - Ni	2,72
	Contact Cu - Ni	2,67 to 3,25

3.3 - Caractérisation mécanique

Le résumé des résultats précédents nous amène à la conclusion que la mousse de cuivre sur la face inférieure permet une meilleure évacuation de la chaleur et que la mousse de nickel sur le contact supérieur permet de réduire la résistance de contact électrique.

Dans cette partie, nous détaillerons la caractérisation thermomécanique. Même si le mélange Mousse-Résine présente un comportement élasto-plastique, cette étude s'est limitée à l'identification des propriétés thermoélastiques du mélange.

3.3.1 - Coefficient d'expansion thermique du mélange mousse-résine

Le coefficient de dilatation thermique (CTE) α_i dans la direction X_i a été mesuré à l'aide d'un analyseur thermomécanique NETZSCH TMA 402 F1 Hyperion (voir figure 12). Le protocole de mesure consiste à imposer à l'échantillon caractérisé, une variation de température $\Delta T = T\text{-}T_0$, (T variant entre 200 et -50°C et T_0=25°C) et en mesurant la variation de la longueur dans la direction donnée, $\Delta l_i = l_i - l_{0i}$. L'équation (7) donne les relations correspondantes :

$$\alpha_i(\mathrm{T}) = \frac{\varepsilon_i^{th}}{\Delta T} \quad \text{avec} \quad \varepsilon_i^{th} = \frac{\Delta l_i}{l_{oi}} \qquad (7)$$

Le long de l'axe Z, le CTE est d'environ 5×10^{-5} K^{-1}, entre -50°C et 100°C. Il augmente ensuite, d'abord lentement jusqu'à 70 ppm/°K à 150°C et enfin brusquement entre 150°C et 200°C pour atteindre 250 ppm/°K à 200°C et dépasser 250 ppm/°K, une fois que la température de transition vitreuse (Tg) de l'époxy est atteinte. Pour les axes *X* et *Y*, le CET est d'environ 46 ppm/°K. Ainsi, la température maximale de fonctionnement de l'assemblage ne doit pas dépasser la *Tg* de la résine époxy.

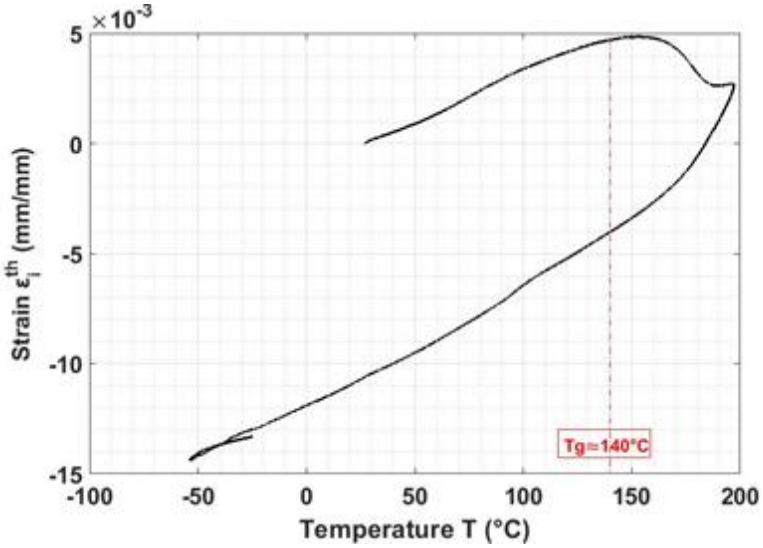


Figure 12 : Essai de dilatométrie pour la mesure du CTE le long de l'axe z en fonction de la température pour le contact de nickel pressé.

3.3.2 - Propriétés élastiques du mélange mousse-résine

Les propriétés élastiques dans les trois directions (*X, Y, Z*) et pour plusieurs températures ont été caractérisées à l'aide du même analyseur thermomécanique sur des échantillons de dimension 6x6 mm et d'une épaisseur maximale l_{0i} = 0,79 mm. Des essais de chargement-relaxation en compression ont été réalisés en appliquant une charge F inférieure à 3 N pour s'assurer que l'échantillon reste élastique. La contrainte normale est calculée par $\sigma_{ii} = F/S_i$et la déformation normale associée $\varepsilon_{ii} = \Delta l_i / l_{oi}$ à partir de la mesure du changement de longueur de l'échantillon $\Delta l_i = l_i - l_{oi}$. On déduit ainsi, à partir des deux mesures le module d'Young E_i dans la direction X_i, (équation. 8).

$$E_i(\mathrm{T}) = \frac{\sigma_{ii}}{\varepsilon_{ii}} \qquad (8)$$

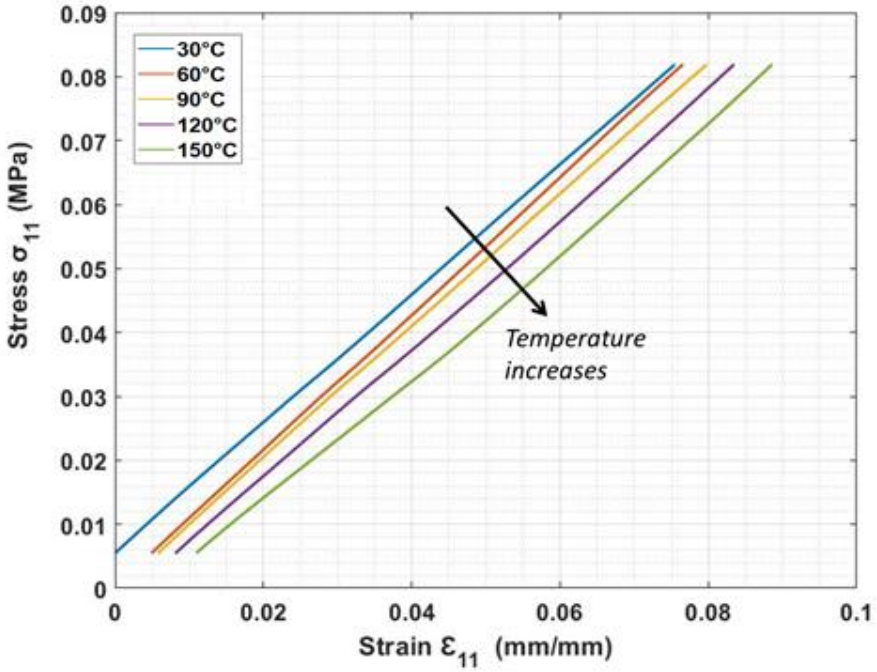


Figure 13 : Diagramme Contraintes-déformations du mélange mousse-résine à différentes températures, dans la direction Z.

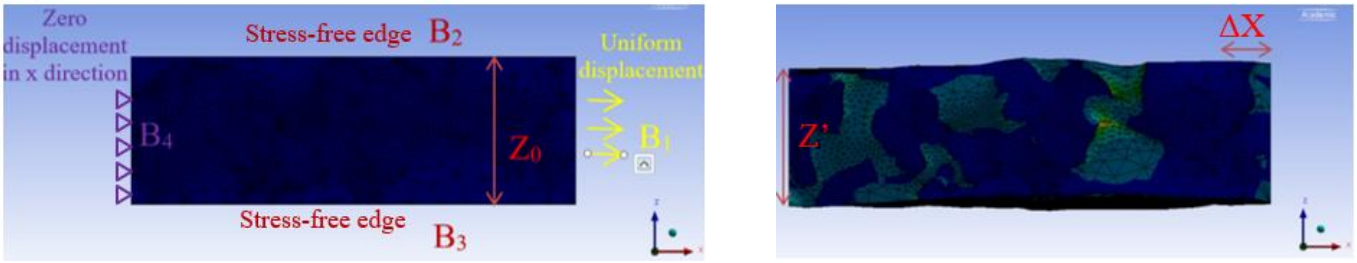
Les essais ont été réalisés à plusieurs températures inférieures à 150°C. Pour chaque température, un essai de chargement/déchargement par compression a été appliqué. Comme le montre la figure 13, la température n'affecte pas le module d'Young dans la plage de température considérée.

En raison des faibles valeurs de charge qui peuvent être appliquées par l'analyseur thermomécanique, nous avons rencontré des problèmes liés à la faible rigidité des contacts entre l'échantillon et les plaques et, par conséquent, les valeurs du module d'Young sont trop faibles par rapport à ce que nous aurions dû attendre d'un tel mélange. En effet, avec un mélange constitué de résine (module d'Young de 3.5 GPa) et de cuivre (131 GPa) nous devrions trouver une valeur entre ces deux valeurs. Par conséquent, nous avons utilisé une méthode de caractérisation numérique en utilisant un modèle géométrique 3D de la microstructure obtenu à partir de l'analyse tomographique et nous avons calculé le module d'Young.

Les relations contraintes-déformations effectives sont données par l'équation (9), où εij et σij sont respectivement la déformation élastique moyenne et la contrainte Gauchy moyenne, et E_i, υ_{ij} et G_i, les paramètres élastiques de tensor.

$$\begin{bmatrix}\bar{\varepsilon}_{xx}\\ \bar{\varepsilon}_{yy}\\ \bar{\varepsilon}_{zz}\end{bmatrix}=\begin{bmatrix}\frac{1}{E_x} & \frac{-\upsilon_{xy}}{E_x} & \frac{-\upsilon_{xz}}{E_x}\\ \frac{-\upsilon_{yx}}{E_y} & \frac{1}{E_y} & \frac{-\upsilon_{yz}}{E_y}\\ \frac{-\upsilon_{zx}}{E_z} & \frac{-\upsilon_{zy}}{E_z} & \frac{1}{E_z}\end{bmatrix}\cdot\begin{bmatrix}\bar{\sigma}_{xx}\\ \bar{\sigma}_{yy}\\ \bar{\sigma}_{zz}\end{bmatrix} et \begin{bmatrix}\bar{\varepsilon}_{xy}\\ \bar{\varepsilon}_{yz}\\ \bar{\varepsilon}_{zx}\end{bmatrix}=\begin{bmatrix}\frac{1}{G_{xy}} & 0 & 0\\ 0 & \frac{1}{G_{yz}} & 0\\ 0 & 0 & \frac{1}{G_{zx}}\end{bmatrix}\cdot\begin{bmatrix}\bar{\sigma}_{xy}\\ \bar{\sigma}_{yz}\\ \bar{\sigma}_{zx}\end{bmatrix} \quad (9)$$

Le module d'Young E_X et les coefficients de Poisson (υ_{xy} et υ_{xz}) ont été identifiés par la simulation d'un essai de traction effectué dans la direction de l'axe X. La figure 14.a montre les conditions aux limites appliquées. Une charge de déplacement uniforme est appliquée sur le bord B1 (dans la direction X). Le déplacement du bord B4 dans la direction X n'est pas autorisé, tandis que les déplacements sur les bords B2 et B3 sont considérés comme étant sans contrainte.



(a) *(b)*

Figure 14 : (a) Conditions aux limites pour l'estimation du module d'Young selon l'axe X, (b) état final après un essai de traction.

Le module d'Young E_X correspond à la pente de la partie linéaire de la courbe $\bar{\sigma}_{xx} - \varepsilon_{xx}$, équation (10). Le coefficient υ_{xz}représente le rapport entre la déformation ε_{zx}mesurée dans la direction z (c'est-à-dire perpendiculaire à la contrainte σ_{xx}) et la déformation mesurée dans la direction x, ε_{xx}, équation (11). De même, le coefficient υ_{xy} est le rapport entre la contrainte ε_{yx}et la contrainte mesurée dans la direction x ε_{xx}, équation (12). Les autres paramètres Ey, Ez et υ_{yz}ont été calculés en modifiant les conditions aux limites et en appliquant des déplacements dans les directions y et z.

$$E_x=\frac{\bar{\sigma}_{xx}}{\bar{\varepsilon}_{xx}} \quad with \quad \bar{\varepsilon}_{xx}=\frac{\Delta\bar{l}_x}{l_{x0}}=\frac{\bar{l}_x-l_{x0}}{l_{x0}} \quad (10)$$

$$\upsilon_{xz}=\frac{\bar{\varepsilon}_{xz}}{\bar{\varepsilon}_{xx}} \quad \text{with} \quad \bar{\varepsilon}_{xz}=\frac{\Delta\bar{l}_z}{l_{z0}}=\frac{\bar{l}_z-l_{z0}}{l_{z0}} \quad (11)$$

$$\upsilon_{xy}=\frac{\bar{\varepsilon}_{xy}}{\bar{\varepsilon}_{xx}} \quad \text{with} \quad \bar{\varepsilon}_{xy}=\frac{\Delta\bar{l}_y}{l_{y0}}=\frac{\bar{l}_y-l_{y0}}{l_{y0}} \quad (12)$$

Les résultats obtenus sont synthétisés dans le tableau 3.

Tableau 3 : Estimation numérique des paramètres mécaniques effectifs du mélange mousse(Ni)-résine.

	Young modulus (GPa)	Poisson ratio	CTE (ppm.K^{-1})
Mélange mousse- Resin (Numerical value)	37,6 (z-axis)	0,2 (XY)	47,9 (z-axis)
	37,9 (x-axis)	0,22 (YZ)	35,8 (x-axis)
	44,8 (y-axis)	0,17 (ZX)	39,9 (y-axis)
Nickel (Datasheet)	204	0,31	12,7
Resin (Datasheet)	3,36	0,4	61

4 - Conclusions

Dans cet article, nous avons montré la faisabilité d'une prise de contact électrique sur une puce enfouie au cœur du PCB, en utilisant uniquement des mousses métalliques pressées, ce qui a permis

de s'affranchir des fils de bonding et de la brasure. Des travaux de caractérisations thermiques et électriques ont été réalisés et montrent des performances électriques et thermiques prometteuses. Selon les résultats, l'utilisation d'une mousse cuivre pour le contact arrière assure de bonnes performances thermiques et une mousse nickel sur la face avant permet de réduire les pertes électriques.

Références :

[1] T. Huesgen, "Printed circuit board embedded power semiconductors: A technology review," *Power Electron. Devices Components*, vol. 3, no. July, p. 100017, 2022, doi: 10.1016/j.pedc.2022.100017.

[2] C. Chen, "A Review of SiC Power Module Packaging: Layout, Material System and Integration," *CPSS Trans. Power Electron. Appl.*, vol. 2, no. 3, pp. 170–186, 2017, doi: 10.24295/cpsstpea.2017.00017.

[3] Y. Pascal, A. Abdedaim, D. Labrousse, M. Petit, S. Lefebvre, et F. Costa, « Using Laminated Metal Foam as the Top-Side Contact of a PCB-Embedded Power Die », IEEE Electron Device Lett., vol. 38, no 10, p. 1453-1456, oct. 2017, doi: 10.1109/LED.2017.2748223.

[4] P. G. Panasonic Corporation, "Pressur Process." https://industrial.panasonic.com/content/data/EM/PDF/processguideline_R-1766GH.pdf

[5] S. Zhang, "Intégration dans un substrat PCB de composants à semi-conducteur grand gap pour le développement d ' un convertisseur d ' électronique de puissance à forte densité To cite this version : HAL Id : tel-02275807 Wide Bandgap Semiconductor Components Integrat," 2018.

[6] A. Bhattacharya, V. V. Calmidi, and R. L. Mahajan, "Thermophysical properties of high porosity metal foams," *Int. J. Heat Mass Transf.*, vol. 45, no. 5, pp. 1017–1031, 2002, doi: 10.1016/S0017-9310(01)00220-4.

Technologie des transistors au nitrure de gallium

La Revue 3E.I
Culture Sciences de l'Ingénieur

Édité le 01/02/2024

école normale supérieure paris-saclay

Matthieu LANDEL[1]

[1] *Ingénieur à l'IFPEN, docteur de l'Université Paris-Saclay*

Cet article fait partie du N°111 de La Revue 3EI de janvier 2024.

Cet article vise à donner au lecteur les éléments nécessaires pour appréhender les enjeux de la révolution qui s'opère actuellement en électronique de puissance avec l'arrivée de cette nouvelle technologie de composants, tant en termes de structures de conversion qu'en termes d'enjeux sociétaux [1] [2], d'applications et de performances mais aussi de contraintes nouvelles notamment en compatibilité électromagnétique.

Les convertisseurs électroniques de puissance reposent sur des composants à semi-conducteur, et notamment les transistors et les diodes, qui sont réalisés à partir du silicium depuis plus d'un demi-siècle. Remplacer le silicium par des semiconducteurs à grand gap comme le nitrure de gallium (GaN) est une rupture technologique qui est susceptible de présenter de nombreux avantages [3] [4]. Ainsi, cette nouvelle technologie permet de réduire les pertes par conduction et par commutation, diminuant les besoins de refroidissement. De plus, les transistors GaN commutent très rapidement (à l'échelle de 1-10 ns), ce qui permet d'augmenter la fréquence de travail des convertisseurs et donc de diminuer le volume des éléments passifs. Ainsi, l'adoption du GaN permettrait de réduire le poids des convertisseurs et d'augmenter leur rendement, autrement dit d'accroitre leur densité massique de puissance.

1 - Propriétés du nitrure de gallium

1.1 - Bref historique

Le Nitrure de Gallium (GaN) est un matériau semi-conducteur à grand gap ayant une structure cristalline hexagonale nommée wurzite [8] comme le montre la Figure 1. Cette structure confère au GaN des propriétés piézoélectriques autorisant à l'interface GaN/AlGaN une conductivité supérieure aux autres matériaux semi-conducteurs.

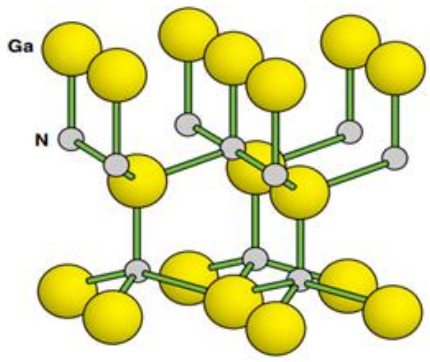


Figure 1 : structure hexagonale du GaN (tiré de [8]).

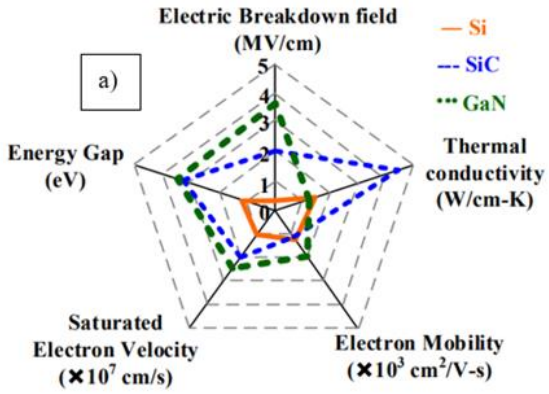
En 1969, le laboratoire RCA (New York, USA) parvient à déposer du GaN sur un substrat saphir en utilisant la méthode de l'épitaxie en phase vapeur hybride [9]. Il faudra attendre 1993 pour que les premières LED bleues au GaN soient fabriquées. Les transistors GaN à haute mobilité d'électrons (*High Electron Mobility Transistor* ou *HEMT*) sont apparus en 2004 dans le domaine RF (Eudyna Corporation, Japon). Ces transistors avaient un substrat SiC. En 2005, la société Nitronex parvient à faire croitre du GaN sur des wafers en silicium. Depuis 2009, on trouve des semiconducteurs de puissance moyenne tension en GaN, grâce aux progrès réalisés sur les matériaux et les procédés de fabrication [4]. A titre d'exemple, la société *EPC* commence à produire des transistors GaN de puissance normalement bloqués cette année-là. Depuis, de grands fondeurs comme Infineon et TI développent de tels transistors.

En comparaison, les *MOSFET* de puissance au silicium ont été introduits en 1976, et ils ont peu à peu remplacé les transistors bipolaires. Il aura fallu environ 30 années pour que les *MOSFET* atteignent leurs limites théoriques en termes de résistance spécifique [3]. La résistance passante spécifique (ou figure de mérite) $R_{DS,ON} \times S_{active}$ a diminué d'un facteur 50 environ entre l'IRF100 et un *MOSFET* moderne [8]. La fréquence de commutation des *MOSFET* est limitée à quelques centaines de kHz. Les transistors GaN, eux, permettent d'atteindre des fréquences de l'ordre du MHz (l'article [3] indique un facteur 30 entre la fréquence de commutation maximale du silicium et celle du GaN).

1.2 - Avantages du matériau GaN

Les composants à semiconducteur jouent un rôle central en électronique de puissance, et le silicium reste aujourd'hui le matériau le plus utilisé. Cependant, les semiconducteurs à grand gap comme le Carbure de Silicium (SiC) et le Nitrure de Gallium (GaN) possèdent des propriétés physiques supérieures [4] [10] [11] [12] [13] [14] [15] [16] [17], comme le montre la Figure 2. Ces semiconducteurs permettent de concevoir des convertisseurs de puissance à plus haut rendement, plus haute température de fonctionnement, plus haute tension et plus haute vitesse de commutation que ceux basés sur les composants au silicium [3] [18] [19] [20]. Aujourd'hui, le matériau GaN est principalement utilisé pour des tensions comprises entre $100\,V$ et $650\,V$. Au-delà de $1200\,V$, ce sont plutôt les composants SiC qui sont privilégiés, mais les transistors GaN pourraient bien les remplacer lorsque des composants $1200\,V$ apparaitront sur le marché, ce qui risque de prendre un certain temps selon [6].

a)



b)

Parameter		Silicon	GaN	SiC
Band Gap E_g	eV	1.12	3.39	3.26
Critical Field E_{Crit}	MV/cm	0.23	3.3	2.2
Electron Mobility μ_n	$cm^2/V{\cdot}s$	1400	1500	950
Permittivity ε_r		11.8	9	9.7
Thermal Conductivity λ	W/cm·K	1.5	1.3	3.8

Figure 2 : a) comparaison des propriétés physiques du silicium, du GaN et du SiC (tiré de [10]) ; b) valeurs typiques de ces grandeurs (tiré de [8]).

Concentrons-nous sur le GaN. La plupart de ses propriétés physiques sont supérieures à celles du SiC [3], sauf pour la conductivité thermique qui est nettement plus faible. La forte rigidité diélectrique du GaN permet de réduire la distance drain-source du transistor, ce qui diminue la résistance à l'état passant. Cette diminution est encore accrue par la forte mobilité des électrons, notamment dans le gaz d'électrons 2D des transistors au GaN. Ces derniers peuvent ainsi atteindre des résistances passantes spécifiques très inférieures à celles du Si et du SiC (Figure 3). La taille de la puce est ainsi environ dix fois plus faible pour un transistor GaN que pour un transistor Si [3] [21].

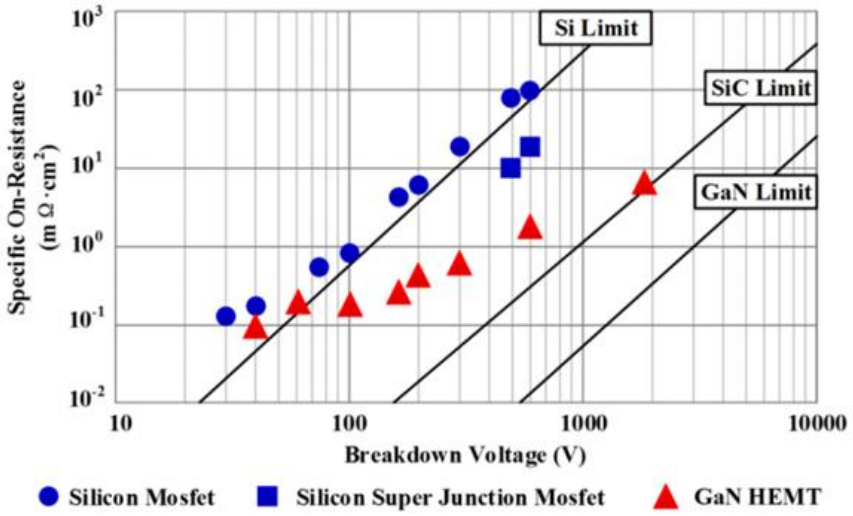


Figure 3 : résistance spécifique en fonction de la tension de claquage pour les MOSFETs silicium, les MOSFETs à superjonction et pour les HEMTs au GaN (tiré de [13]).

Cet effet de réduction de la taille de la puce (pour un calibre en courant donné) par rapport à un transistor au silicium, a pour conséquence de diminuer les capacités parasites. De plus, la structure latérale des transistors GaN leur procure naturellement une faible capacité de transfert C_{GD} et une faible capacité C_{GS} [3] [13]. Ainsi, la capacité d'entrée ($C_{ISS} = C_{GS} + C_{GD}$) est environ 30 fois plus faible pour un transistor GaN que pour un composant au silicium [3] de même caractéristiques et la charge totale de la grille, bien plus faible [13] (Figure 4), explique la diminution des pertes de commande et l'augmentation de la vitesse de commutation. Par ailleurs, la capacité de sortie C_{oss} est environ dix fois plus faible pour un transistor GaN que pour un *MOSFET* à superjonction [22]

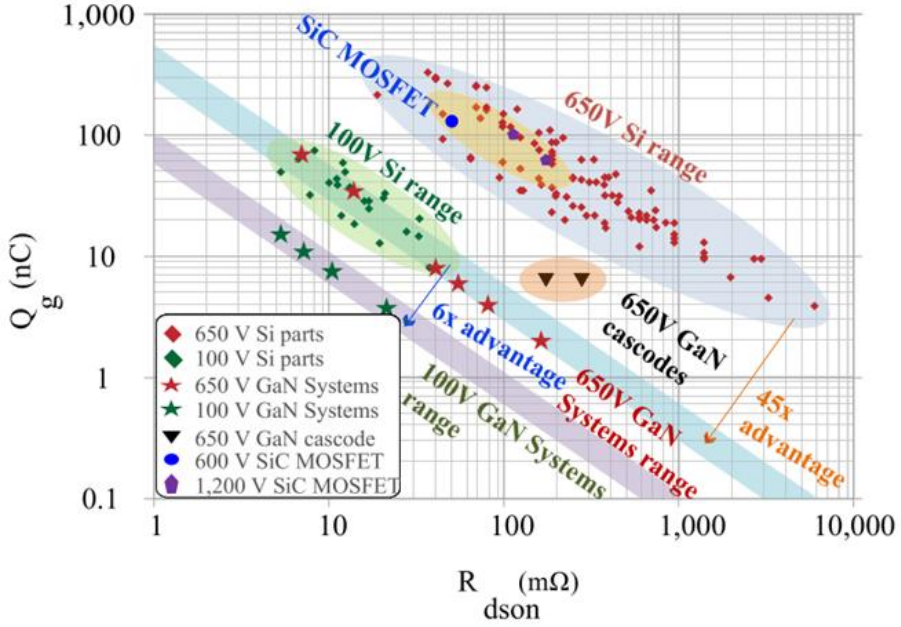


Figure 4 : charge de grille Q_g en fonction de la résistance passante pour des transistors Si, SiC et GaN (tiré de [3]).

La vitesse de saturation élevée dans le GaN, combinée avec la faible valeur des capacités parasites, autorise des commutations plus rapides que le silicium. A titre d'exemple, sous $400\ V$, il est possible de commuter en moins de $10\ ns$ [10]. Le fondeur GaN Systems indique que la vitesse de commutation d'un transistor GaN est quatre fois plus grande à l'amorçage, et deux fois plus grande au blocage par rapport à un transistor SiC [22] de calibre similaire.

La conduction inverse d'un transistor GaN diffère de celle des *MOSFETs* silicium. Ces derniers possèdent une diode de corps de type PIN, qui est lente et oblige dans la plupart des applications à leur adjoindre une diode antiparallèle de type Schottky ou à les utiliser en mode redresseur synchrone. Du fait de leur structure, les transistors GaN n'ont pas de diode de corps, et donc pas de pertes par recouvrement inverse. En revanche, ces transistors présentent une chute de tension source-drain en conduction inverse lorsque la tension grille-source V_{GS} est inférieure à la tension de seuil V_{TH}, comme le décrivent Jones et al. dans [10]. La chute de tension est égale à : $V_{SD} = V_{TH} - V_{GS-OFF}$. Utiliser une tension V_{GS} négative permet de bloquer efficacement le composant mais cela augmente donc les pertes en conduction inverse.

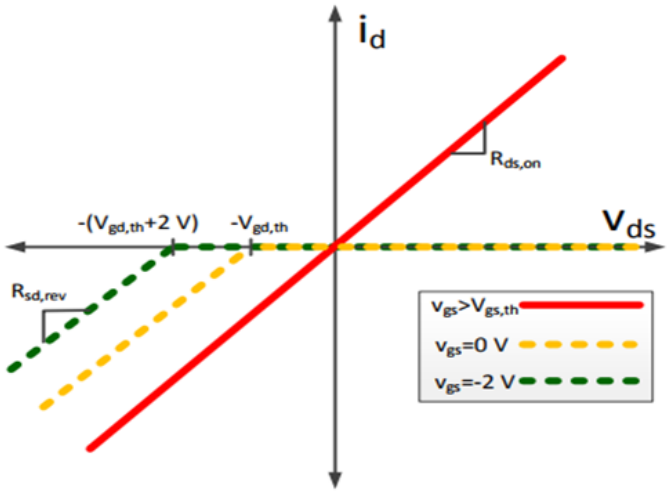


Figure 5 : caractéristique $I_D = f(V_{DS})$ pour un transistor GaN conduisant en inverse (tiré de [10]).

Enfin, du fait de leurs structures, la tension de seuil des différentes familles de transistors GaN est plus basse qu'un transistor *MOSFET* Si ($1{,}5\ V$ environ pour un transistor GaN $650\ V$, contre $4\ V$ pour un *MOSFET* Si $650\ V$), et la tension de grille optimale est de $5\ V$ à $6\ V$ seulement [22], ce qui diminue aussi les pertes de commande car les *IGBTs*, les *MOSFETs* silicium et SiC nécessitent des tensions de l'ordre de 15 à $20\ V$ (Tableau 1).

Tableau 1 : comparaison des tensions de commande des transistors GaN de GaN Systems avec les MOSFETs Si, les IGBT et les MOSFETs SiC [18].

Gate Bias Level	GaN Systems GaN E-HEMT	Si MOSFET	IGBT	SIC MOSFET
Maximum rating	-20/+10V	-/+20V	-/+20V	-8/+20V
Typical gate bias values	0 or-3/+5-6V	0/+10-12V	0 or -9/+15V	-4/+15-20V

Le GaN est donc un très bon candidat pour des applications à fort rendement et/ou de découpage à haute fréquence [23]. Dès 2013, l'article [24] indiquait que ces transistors commençaient déjà à se répandre sur le marché grâce à leurs propriétés supérieures à celles du silicium.

1.3 - Importance du substrat

Une étape cruciale dans la fabrication d'un transistor GaN est le dépôt du GaN sur un substrat. Ce dernier peut être du GaN (homo-épitaxie) ou bien un matériau étranger (hétéro-épitaxie).

L'homo-épitaxie sur substrat GaN est la solution idéale structurellement parlant [25] et assure une faible densité de défauts cristallins, mais sa fabrication est difficile. La méthode Czochralski (utilisée avec succès pour faire croitre du silicium) n'est pas adaptée car elle demanderait de très hautes température et pression [4]. L'épitaxie hybride en phase vapeur (*HVPE*) sur un matériau étranger comme le saphir reste la méthode la plus utilisée, par exemple dans [26]. Elle est décrite en détails dans [27] où les auteurs fabriquent un wafer de GaN de trois pouces. Le saphir est ensuite supprimé. D'autres techniques permettent de se passer de ce matériau étranger et font l'objet de recherches [10] : la méthode *HNPSG* (*High Nitrogen Pressure Solution Growth*), la méthode Na-Flux (solution à basse pression avec du sodium) et la méthode ammonothermale, qui semble prometteuse en termes de coût dans un futur proche. Dans l'article [28] de 2021, les auteurs présentent une technique innovante nommée FFC, mélange de la technique « Na Flux » et HVPE ; ils fabriquent ainsi un wafer GaN de deux pouces. La fabrication de ces derniers est encore un sujet actif de recherche.

L'hétéro-épitaxie est moins couteuse et c'est la solution retenue industriellement aujourd'hui. Le GaN est généralement déposé par hétéro-épitaxie sur un matériau étranger comme le SiC, le saphir ou le silicium, mais les distances atomiques et les propriétés thermiques du GaN sont différentes de ces derniers, ce qui engendre des contraintes et des dislocations [4] [3]. Il est donc nécessaire de placer une couche tampon sur le substrat, qui absorbe les contraintes mécaniques. Des couches alternées de GaN, d'AlGaN et d'AlN peuvent être utilisées [10], formant une épaisseur totale de quelques micromètres. La Figure 6 présente les caractéristiques physiques de trois matériaux pouvant être utilisés comme substrat.

	SiC	Sapphire	Si
Lattice constant mismatch	3.1%	15%	17%
Linear thermal expansion coefficient ($\times$ 10^{-6} K^{-1}) (GaN value = 5.6)	4.16 (*c*-axis)	7.5	2.6
Thermal conductivity (W cm^{-1} K^{-1})	3.8–4.9	0.25	1.56
Typical epitaxial GaN dislocation density	$>10^8$ cm^{-2}	$>10^8$ cm^{-2}	$>10^9$ cm^{-2}
Cost	Expensive	Less expensive	Cheap

Figure 6 : comparaison des substrats possibles pour un transistor GaN (tiré de [4]).

Le SiC est le meilleur choix, mais reste très cher. On le réserve aux très hautes densités de puissance. Le saphir est moins cher mais sa conductivité thermique est très inférieure au SiC. De plus, les wafers de saphir sont limités à une taille de deux pouces. Pour des applications commerciales sensibles au coût, le substrat Si est un bon compromis [3], et les wafers sont disponibles en grand diamètre. Pour rester compétitifs, les transistors GaN doivent être fabriqués sur des wafer de diamètre $150\ mm$ minimum (comme c'est le cas dans [29]), sachant que les *IGBTs* Si sont fabriqués sur des wafers de $200\ mm$ voire $300\ mm$ en pleine production. L'article [30] présente ainsi la fabrication de transistors GaN sur un wafer silicium de $300\ mm$. On réalise une hétéro-épitaxie par *MOCVD* (*Metal Organic Chemical Vapor Deposition*). La suite des étapes de fabrication du transistor dépend de la structure choisie (partie 2 -).

Les transistors GaN sont pour l'heure encore couteux, mais leur diffusion massive devrait permettre de réduire leur prix [3], peut-être au niveau de celui du silicium [19].

2 - Structures de transistors GaN de puissance

Les transistors GaN normalement passants peuvent être utilisés pour des applications basse tension à haute fréquence de commutation, mais le domaine de l'électronique de puissance exige un comportement normalement bloqué pour des raisons de sécurité [3] [10] [20] [23]. Cela permet aussi de simplifier la commande du transistor. Nous allons nous focaliser sur ce type de composants.

Les structures verticales sont également prometteuses, nous les aborderons brièvement dans la partie 2.2 - .

2.1 - *HEMT* (HFET)

Ces transistors seront désignés par le terme *HEMT (high electron mobility transistor)*, même si on trouve le terme *HFET* (*Heterojunction Field Effect Transistor*) dans la littérature, qui est équivalent.

2.1.1 - *HEMT* à grille gravée (*recessed gate*)

La structure *HEMT* est la plus étudiée et la plus répandue chez les industriels [10]. Elle est basée sur un phénomène spontané de polarisation piézoélectrique apparaissant à l'interface d'une hétérostructure AlGaN/GaN [8] [10]. La couche d'AlGaN a une épaisseur généralement comprise entre $20\,nm$ et $30\,nm$. Un gaz 2D d'électrons très dense et à haute mobilité ($1500 - 2000\,cm^2.V^{-1}.s^{-1}$ [8]) est alors créé. Les *HEMTs* sont ainsi de très bons candidats pour des applications de forte puissance et/ou de découpage haute fréquence. Cependant, le gaz 2D rend ces transistors normalement passants. Le canal existe en permanence, même en l'absence d'une polarisation de grille positive.

Il existe plusieurs façons d'appauvrir le gaz 2D et ainsi fabriquer un transistor normalement bloqué, chacune ayant ses limitations et ses compromis [4] [23]. La méthode la plus répandue est de graver partiellement la couche d'AlGaN pour y loger la grille (*recessed gate*). On amincit en général cette couche uniquement sous la grille, en gravant avec du plasma (méthode *ICP-RIE* pour *Inductive Coupled Plasma Reactive Ion Etching* [31]). La concentration en aluminium de la couche d'AlGaN peut aussi être modifiée. La gravure diminue l'effet piézoélectrique et coupe le gaz 2D. La structure *HEMT* est assez difficile à fabriquer car elle exige un contrôle précis de la profondeur de gravure, ce paramètre influant directement sur la tension de seuil [23]. Si la gravure est mal contrôlée, cela peut générer un courant de fuite de grille et impacter l'uniformité de la tension de seuil. La Figure 7 montre la structure d'un *HEMT* à grille gravée.

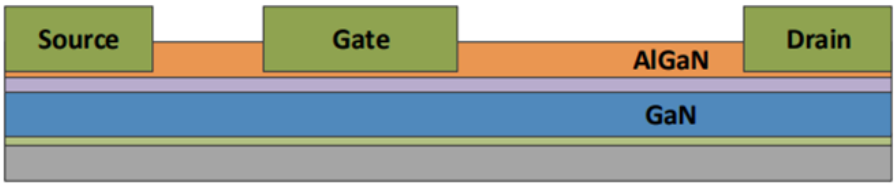


Figure 7 : structure HEMT à grille gravée, où la couche d'AlGaN est gravée partiellement pour recevoir la grille (tiré de [10]).

Les *HEMTs* de ce type présentent de très faibles valeurs de résistance à l'état passant.

2.1.2 - *MOS-HEMT* (ou *MIS-HEMT*)

Les termes *MOS-HEMT* (*Metal Oxyde Semiconductor High Electron Mobility Transistor*) et *MIS-HEMT* (*Metal Insulator Semiconductor High Electron Mobility Transistor*) désignent tous les transistors GaN *HEMT* possédant un isolant sous la grille. Si l'isolant est un oxyde, on parle de *MOS-HEMT*. Si ce n'est pas le cas, on parle de *MIS-HEMT*. Différentes structures de *MOS(MIS)-HEMT* existent.

La première structure possible est le *recessed gate MOS(MIS)-HEMT*. Cette structure est identique au *recessed gate HEMT*, à la différence près qu'il y a un isolant sous la grille (Figure 8).

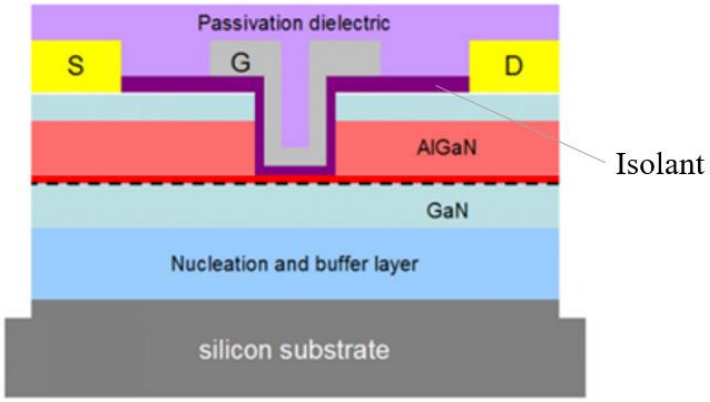


Figure 8 : structure recessed gate MOS(MIS)-HEMT, où la couche d'AlGaN est gravée partiellement (tiré de [4]).

Une deuxième structure possible est le *MOS(MIS)-HEMT* à gravure profonde, où la couche d'AlGaN est supprimée complètement sous la grille [4], comme présenté en Figure 9. Le gaz 2D est alors coupé. Le drain et la source doivent être connectés par une couche d'inversion pour rendre passant le composant [10] [32]. Le choix de l'isolant est critique : il impacte à la fois la mobilité du canal et la stabilité de la tension de seuil.

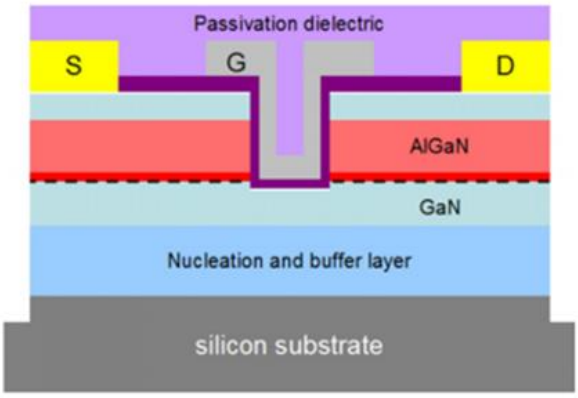


Figure 9 : structure MOS(MIS)-HEMT, où la couche d'AlGaN est gravée complètement (tiré de [4]).

La résistance spécifique des *MOS(MIS)-HEMT* à gravure profonde est supérieure à celle des *recessed gate HEMTs* puisque ces derniers bénéficient du gaz 2D d'électrons. Les premiers sont toutefois moins sensibles à la profondeur précise de gravure que les *HEMTs* et possèdent des tensions de seuil comprises entre $2\,V$ et $5\,V$, ce qui peut éviter les remises en conduction intempestives [23].

On trouve aussi une autre structure hybride, dont la grille comprend à la fois une couche de p-GaN et un isolant, décrite dans l'article [33] notamment (Figure 10). Le diélectrique est critique pour la fiabilité et mesure une dizaine de nanomètres d'épaisseur. La tension de seuil peut souffrir d'instabilité car des charges peuvent se piéger sur les bords du diélectrique [33]. La grille peut souffrir d'un effet tunnel qui contribue à augmenter la valeur du courant de grille.

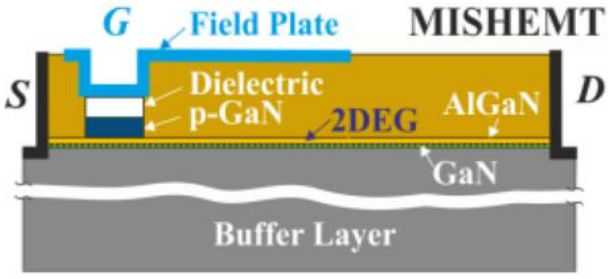


Figure 10 : structure d'un MIS-HEMT (tiré de [33]).

2.1.3 - *HEMT* avec grille p-GaN

Tout en restant dans la famille des *HEMT*, on peut appauvrir le gaz 2D en insérant une couche de p-GaN (ou p-AlGaN) sous la grille [8] [4]. Cette dernière possibilité reste la plus prometteuse, et est utilisée notamment chez les fondeurs GaN Systems et Panasonic [23]. Cela forme avec le métal de l'électrode de grille un contact métal/semiconducteur. Ce dernier peut se comporter comme un contact ohmique ou comme un contact schottky, en fonction du métal choisi et d'autres paramètres liés au procédé de fabrication. Si le contact est ohmique, il existe en permanence un courant de grille lorsque le transistor est à l'état passant (c'est le cas des transistors du fondeur Panasonic). Au contraire, si le contact est de type Schottky, ce courant est négligeable en dehors des phases de charge et de décharge de la grille (cette structure est celle du fondeur GaN Systems). Il faut noter que la fabrication d'un contact ohmique est délicate ; si celui-ci n'est pas parfaitement réalisé, la caractéristique électrique peut être de type contact Schottky. Il est possible qu'un fondeur annonce un contact d'un certain type (ohmique ou Schottky), mais qu'en réalité la grille se comporte différemment.

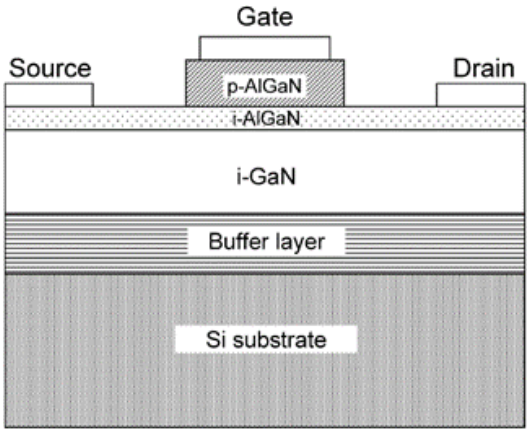


Figure 11 : Structure d'une grille de type P-AlGaN. Tiré de [34].

La grille p-GaN se comporte comme une diode, et est assez fragile. Un courant de grille trop important peut la détruire. La tension V_{GS} admissible est donc assez faible ($-10\,V/+7\,V$ pour le fondeur GaN Systems par exemple).

Certains fondeurs combinent plusieurs techniques pour fabriquer leurs transistors. Il est possible de graver légèrement la couche d'AlGaN et d'utiliser en plus une grille p-GaN (cas de GaN Systems).

2.1.4 - *HEMT* avec injection de fluor (implanted gate)

Une autre possibilité est d'injecter du fluor en dessous de la grille (*implanted gate*) avec un traitement par plasma [8] [4] [23]. La charge négative des ions fluorés appauvrit le gaz 2D. Cela a pour inconvénient de dégrader la stabilité de la tension de seuil à haute température. On peut aussi introduire une couche à base d'oxydes de nickel en dessous de la grille, mais cette approche est peu répandue.

2.1.5 - Cascode

Pour obtenir un transistor normalement bloqué, une autre solution est d'associer en série un transistor GaN normalement passant avec un *MOSFET* silicium basse tension : c'est la structure cascode [13] (Figure 12). Elle est adaptée aux tensions supérieures à $200\,V$, sinon la résistance passante ramenée par le *MOSFET* est trop importante [8]. Le transistor GaN peut être rendu passant ou bloqué en appliquant une tension de grille sur le *MOSFET* basse tension. La tension de seuil est de l'ordre de $5\,V$, la commande est donc moins sensible aux bruits [14] [33] et il est possible d'utiliser un driver de *MOSFET* silicium classique fournissant une tension de commande égale à $15\,V$. Cette structure est intéressante car la structure d'un transistor GaN normalement passant est

plus simple que celle d'un transistor normalement bloqué et qu'il est aisé d'ajouter en série un *MOSFET* silicium basse tension pour en faire un interrupteur normalement bloqué. Les pertes au blocage sont très faibles, ce qui en fait un bon choix pour des convertisseurs à fort courant [35]. La résistance passante spécifique est similaire à celle des structure p-GaN. Un autre avantage de cette structure est une meilleure robustesse car les grilles des transistors GaN à appauvrissement ne sont pas dopées et donc exemptes de défauts dans le cristal [36]. Les performances en commutation dépendent beaucoup des inductances parasites du package (connexion entre le *HEMT* et le *MOSFET* silicium) et de l'accord entre les capacités parasites des deux transistors [10].

Cette structure souffre cependant de certains inconvénients. Tout d'abord, la complexité du boitier est accrue [3] puisqu'il y a deux puces à connecter entre elles. Dans le circuit de grille du transistor GaN, les inductances parasites des fils de bounding et les capacités parasites forment un circuit résonant très peu amorti [35]. Cela peut générer des oscillations qui sont parfois auto-entretenues [3], et d'autant plus fortes que la tension V_{DS} est grande. Ceci peut affecter les commutations. En conduction inverse, la diode du *MOSFET* conduit et il y a un temps de recouvrement non nul, ce qui augmente les pertes [13]. Le fait d'utiliser deux transistors pose aussi des soucis de coût [4].

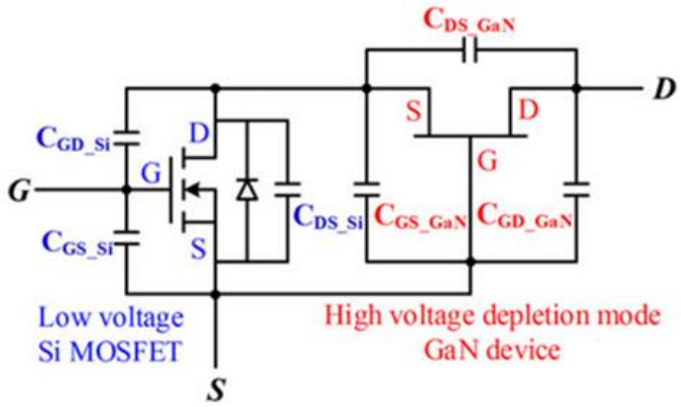


Figure 12 : structure cascode avec les capacités parasites (tiré de [37]).

J'ai réalisé des essais de court-circuit sur des transistors cascode du fondeur Transphorm.

2.2 - Structures verticales

Les structures verticales ont des caractéristiques supérieures aux structures latérales, mais exigent en général une homo-épitaxie, c'est-à-dire un substrat en GaN, qui est coûteux, difficile à fabriquer et encore peu répandu [10]. Ces structures ont beaucoup moins de défauts dans la structure cristalline, et n'ont aucune différence de paramètre de maille. Ceci augmente leur fiabilité [3] et supprime le phénomène de résistance passante dynamique (partie 3 -). Par ailleurs, cette structure est capable de résister au phénomène d'avalanche, contrairement à la structure latérale. En général, les niveaux de courant des transistors verticaux sont plus élevés que ceux de leurs homologues latéraux [3]. La taille de la puce sur le wafer est plus petite car il suffit d'augmenter l'épaisseur de la zone de drift pour augmenter la tenue en tension [38]. Les structures verticales seraient plutôt adaptées à de fortes puissances [3]. La gestion de la thermique est aussi facilitée par rapport aux structures latérales [3].

La société Avogy a déposé de nombreux brevets dans ce domaine et proposait des échantillons de diodes $600\,V$, $1200\,V$ et $1700\,V$. Plus récemment, l'entreprise Nexgen montre sur son site des datasheets de transistors verticaux avec une tension de claquage de $700\,V$ ou $1200\,V$ [39]. Cependant, ces composants ne sont pas encore industrialisés. Par ailleurs, un transistor vertical à nanostructure 3D est présenté dans [40]. D'autres structures utilisant du GaN comme le *CAVET* (*Current Aperture Vertical Electron Transistors*) [41] [38] ou le *MOSFET* à tranchées [42]

permettent d'atteindre des tensions de claquage supérieures à $1200\,V$, mais utilisent de coûteux substrats en GaN, et font parfois appel à des étapes de croissance de cristal spécifiques.

Il existe cependant des structures verticales à substrat silicium, étudiées notamment par le MIT (*Massachusetts Institute of Technology*, USA) et par l'entreprise Cambridge Electronics Inc [10]. Cependant, en 2020, l'article [3] indiquait qu'aucun transistor GaN vertical n'était encore commercialisé même si ces composants font l'objet de beaucoup de recherches [3].

3 - Phénomène d'effondrement du courant (*current collapse*)

Les transistors GaN latéraux souffrent généralement de défauts dans la couche de GaN ou d'AlGaN, comme des dislocations ou des atomes étrangers (impuretés). On parle de pièges, qui s'activent ou se désactivent selon la tension appliquée. Ces pièges correspondent à des niveaux d'énergie dans le *gap* du GaN ou de l'AlGaN. Des charges peuvent se piéger et se dépiéger. La thèse de Fadi Zaki [43] indique que le phénomène de piégeage de charges est réversible mais peut affecter le comportement statique et dynamique des *HEMTs*.

Aussi appelé « résistance passante dynamique », le phénomène d'effondrement du courant dans les structures GaN latérales est directement lié aux pièges et reste un frein à la diffusion des transistors GaN [15]. Cela consiste en une augmentation temporaire de la résistance à l'état passant, qui est fonction de la tension drain-source V_{DS} supportée à l'état bloqué [10] et de la durée d'application de cette tension. Quand le transistor est bloqué, des charges peuvent se piéger dans les défauts du cristal ou aux interfaces entre couches dans la puce [44]. Cela génère des zones chargées négativement au sein du transistor, ce qui a deux effets : la répartition du champ électrique dans le transistor est modifiée, et le gaz 2D est perturbé. La résistance à l'état passant augmente alors, ainsi que les pertes [15], pouvant aller jusqu'à la destruction du composant [45].

On distingue deux types de piégeage [10]. Tout d'abord, des charges situées en surface peuvent se piéger à côté du bord de la grille ou du côté du drain quand le composant est à l'état bloqué. Ces charges se comportent comme une grille virtuelle et affaiblissent le gaz 2D. Le second mécanisme est le piégeage d'électrons à haute énergie en profondeur, dans la couche tampon située au-dessus du substrat notamment. Cela affaiblit aussi le gaz 2D, augmente la résistance à l'état passant et peut endommager le transistor sur le long-terme. Le phénomène est dynamique car la résistance à l'état passant se met à augmenter au fil des commutations.

Les fondeurs ont fait beaucoup de progrès pour contrer le phénomène de courant de collapse, en utilisant par exemple des plaques de champ pour redistribuer le champ électrique entre le drain et la grille, ou en connectant le substrat à la source. Panasonic aurait éliminé le courant de collapse grâce à l'ajout d'une structure dopée P à côté du drain [10] [44] (structure *HD-GIT*). L'article [14] étudie les mécanismes de piégeage de charges causés par l'application d'une tension négative pulsée sur la grille de transistors cascode. Ils constatent une dégradation de la résistance passante et du courant de drain maximal ainsi qu'un déplacement de la tension de seuil. Les auteurs de l'article [15] caractérisent le phénomène de résistance passante dynamique sur des transistors GS66508T du fondeur GaN Systems. Ils conseillent de faire cette caractérisation à haute fréquence et haute température pour être dans des conditions proches d'un usage réel dans un convertisseur. La résistance passante dynamique augmente en effet fortement avec la température.

4 - Utilisation du GaN dans des convertisseurs

En 2017, il n'y avait que trois fondeurs qui proposaient des transistors GaN à enrichissement à la vente, dont une structure cascode [10]. L'article [3], paru en 2020, indiquait les applications industrielles possibles du GaN dans les secteurs des voitures électriques et hybrides, de l'avion plus

électrique, du navire plus électrique, des trains électriques et des véhicules tout terrain électrifiés. En 2021, on compte plus de neuf fondeurs mais les transistors GaN ne sont pas encore présents dans ces secteurs.

En 2013, l'article [13] montrait déjà l'utilisation de transistors GaN cascode dans un convertisseur résonnant LLC de $300\,W$ avec une fréquence de commutation de $1\,MHz$.

En 2014, le fondeur *EPC* publie un article [46] montrant l'utilisation de leurs transistors GaN basse tension dans un convertisseur DC/DC *POL* (*Point Of Load*) (entrée : $12-28\,V$; sortie : $1{,}2\,V$) travaillant à $1\,MHz$. La comparaison faite avec des transistors silicium donne un net avantage au GaN avec presque 4 points de rendement de plus pour ce dernier. Un peu plus tard, en 2015, les auteurs de l'article [47] intègrent des transistors cascode de Transphorm dans un convertisseur *PFC* (*Power Factor Corrector*). Ce dernier atteint un rendement maximal de $99\,\%$ et une densité de puissance égale à $220\,W/in^3$. Ensuite, courant 2016, l'article [48] présente l'utilisation de transistors du fondeur GaN Systems dans un onduleur triphasé de $10\,kW$. Le rendement atteint $98{,}8\,\%$ et le volume n'est que de $0{,}7\,L$. Puis, en 2017, l'article [49] présente un convertisseur résonant LLC à commutation douce à base de transistors GaN cascode du fondeur Transphorm. Leur tension de claquage vaut $600\,V$ et la tension d'entrée du convertisseur est de $400\,V$. Les auteurs comparent les pertes entre ces derniers et des transistors silicium et il apparait que le GaN présente des pertes $50\,\%$ plus faibles.

Plus tard, en 2019, dans l'article [50], les auteurs conçoivent et testent un convertisseur *NPC* (*Neutral Point Clamped*) à trois niveaux à base de transistors du fondeur GaN Systems (leur tension de claquage est de $650\,V$) et de diodes SiC. Leur prototype atteint un rendement de 98%. Ces transistors ont un boitier monté en surface, ce qui a impacté la façon de router le convertisseur ainsi que son refroidissement. Toujours en 2019, les auteurs de l'article [12] passent en revue plus de 160 papiers montrant une intégration de transistors GaN dans des convertisseurs de puissance de plus de $500\,W$ de différentes familles (DC/DC, DC/AC et AC/DC). Il ressort de tout cela que ces composants permettent de concevoir des convertisseurs à haut rendement (jusqu'à $99{,}8\,\%$) dans la gamme moyenne tension, de moyenne à forte puissance (jusqu'à $80\,kW$). La fréquence de travail des systèmes passés en revue s'échelonne de $100\,\mathrm{kHz}$ à plus de $1\,MHz$. Onze papiers font état d'une fréquence supérieure à $1\,MHz$. La grande majorité des convertisseurs étudiés a un rendement compris entre 95% et 99% ; plus des deux tiers des convertisseurs ont une densité de puissance supérieure à $100-200W/in^3$ et le maximum atteint est de $9{,}5\,kW/in^3$.

Enfin, en 2020, l'article [51] montre l'utilisation de transistors GaN du fondeur *EPC* dans un convertisseur DC/DC résonant à commutations douces à $1\,MHz$ qui pourrait être utilisé dans le domaine spatial. Une comparaison est faite avec des *MOSFETs* silicium commutant à $100\,kHz$. Il est montré que la densité de puissance est accrue d'environ $80\,\%$ avec le GaN, et que le rendement reste identique.

Si les transistors GaN sont pour le moment absents dans l'industrie des transports, on trouve une application industrielle dans le domaine des chargeurs USB. Depuis 2018, il existe dans le commerce des chargeurs intégrant des transistors GaN [5] pour alimenter un téléphone portable, une tablette ou un ordinateur. C'est la seule utilisation industrielle du GaN recensée en 2021.

5 - Conclusion et futures tendances

Les défis à relever pour permettre la diffusion massive du GaN dans l'industrie sont : la robustesse et la fiabilité en court-circuit qui sont encore limitées, la tension de claquage encore limitée, la complexité de la commande de la grille, la gestion de la thermique et la conception des boitiers

des transistors pour utiliser la pleine capacité de ces composants, et assurer une fiabilité à long terme [3]. Enfin, le coût élevé des composants freine encore leur adoption.

Concernant la robustesse des transistors GaN, elle est encore insuffisante en court-circuit lorsqu'on les utilise sous des tensions proches de la tension de claquage. La fiabilité est, elle aussi, encore perfectible même si certains composants comme le *HD*-GIT de Panasonic peuvent supporter de nombreux courts-circuits sous une tension de bus élevée.

La tension de claquage encore limitée des transistors GaN du commerce ($650\,V$) restreint aujourd'hui leur usage [50]. Il est encore difficile de fabriquer des transistors GaN haute tension ($1200\,V$) car cela nécessite des améliorations et des innovations dans le procédé de fabrication, au niveau du substrat. L'utilisation du nitrure d'aluminium polycristallin (AlN) comme substrat est prometteur pour fabriquer des composants $1200\,V$, car son coefficient d'expansion thermique est plus proche de celui du GaN par rapport aux substrats silicium classique [3]. Le substrat AlN autorise le dépôt d'une couche plus épaisse de GaN (avec des wafers de $200\,mm$), ce qui est nécessaire pour augmenter la tenue en tension.

Le routage du PCB d'un convertisseur à base de GaN est critique [46] à cause des vitesses de commutation élevées. Les inductances parasites peuvent avoir un impact important sur les pertes [13]. Les circuits de commande doivent être adaptés aux forts di/dt et dv/dt, ce qui impose un routage très soigné. Des oscillations peuvent apparaitre au niveau de la grille à cause des éléments parasites du circuit, la commande de la grille est donc critique [14]. Par ailleurs, du fait de leur vitesse de commutation, les transistors GaN posent des problèmes de pollution électromagnétique [3]. La grille des composants GaN discrets est vulnérable au bruit et aux surtensions causées par les commutations très rapides. La tension de seuil est très basse, ce qui peut causer des remises en conduction intempestives lors du blocage. A titre d'exemple, le choix de la valeur de la résistance de grille pour l'amorçage est critique : trop faible, elle entraine des oscillations et donc des pertes de commandes importantes ; trop grande, elle ralentit les commutations et augmente aussi les pertes. L'intégration monolithique du driver, des protections et du transistor GaN (voire d'un demi-pont complet) a de nombreux avantages [3] : un meilleur contrôle des inductances parasites, un meilleur contrôle du dv/dt, la possibilité d'intégrer des protections (au démarrage, contre les surcourants, contre un échauffement excessif et contre les courts-circuits). Cette intégration permet une conception plus modulaire du convertisseur, augmente sa densité de puissance et permet aussi de réduire les coûts.

Le boitier d'un transistor a de nombreuses fonctions : il protège la puce, permet de la monter sur un PCB et participe à l'évacuation de la chaleur. La conception d'un boitier et de ses connexions avec la puce est délicate car elle doit prendre en compte un besoin d'isolation contre de fortes tensions et une tenue à de forts courants, sans dégrader les performances de la puce. Le boitier joue un rôle important dans la fiabilité du transistor. Les transistors GaN nécessitent un boitier spécifique dédié, avec des connexions optimisées vers le radiateur. Le refroidissement est en effet critique du fait de la taille réduite de la puce.

La minimisation du poids et du volume des systèmes embarqués est un des objectifs principaux de l'industrie des transports. La grande densité de puissance et la fréquence de commutation élevée du GaN en font un très bon candidat pour les convertisseurs de puissance du secteur des transports. En effet, cela permet de diminuer la taille des composants passifs et donc d'alléger le véhicule. Si les différentes contraintes listées dans cette partie restent des freins à la diffusion des transistors GaN, il est probable que ces derniers soient massivement adoptés dans les années à venir [52].

Références :

[1] Ministère de la transition écologique et Institute for climate economics, « Chiffres clés du climat - France, Europe et Monde ». déc. 2020. [En ligne]. Disponible sur : https://www.statistiques.developpement-durable.gouv.fr/edition-numerique/chiffres-cles-du-climat

[2] Institut Français du Pétrole, « Les énergies pour le transport : avantages et inconvénients ». déc. 2008. [En ligne]. Disponible sur : https://inis.iaea.org/collection/NCLCollectionStore/_Public/42/016/42016175.pdf

[3] N. Keshmiri, D. Wang, B. Agrawal, R. Hou, et A. Emadi, « Current Status and Future Trends of GaN HEMTs in Electrified Transportation », IEEE Access, vol. 8, p. 70553-70571, 2020, doi: 10.1109/ACCESS.2020.2986972.

[4] M. Su, C. Chen, et S. Rajan, « Prospects for the application of GaN power devices in hybrid electric vehicle drive systems », Semicond. Sci. Technol., vol. 28, no 7, p. 074012, juin 2013, doi: 10.1088/0268-1242/28/7/074012.

[5] Ganfast, « Site internet de ganfast ». [En ligne]. Disponible sur : https://ganfast.com/products/

[6] H. Schefer, L. Fauth, T. H. Kopp, R. Mallwitz, J. Friebe, et M. Kurrat, « Discussion on Electric Power Supply Systems for All Electric Aircraft », IEEE Access, vol. 8, p. 84188-84216, 2020, doi: 10.1109/ACCESS.2020.2991804.

[8] A. Lidow, M. De Rooij, J. Strydom, D. Reusch, et J. Glaser, GaN Transistors for Efficient Power Conversion.

[9] S. Nakamura et M. R. Krames, « History of Gallium–Nitride-Based Light-Emitting Diodes for Illumination », Proc. IEEE, vol. 101, no 10, p. 2211-2220, oct. 2013, doi: 10.1109/JPROC.2013.2274929.

[10] E. A. Jones, F. F. Wang, et D. Costinett, « Review of Commercial GaN Power Devices and GaN-Based Converter Design Challenges », IEEE J. Emerg. Sel. Top. Power Electron., vol. 4, no 3, p. 707-719, sept. 2016, doi: 10.1109/JESTPE.2016.2582685.

[11] A. Castellazzi, « Opportunities and challenges of integrated WBG power electronics development », in 2021 Third International Symposium on 3D Power Electronics Integration and Manufacturing (3D-PEIM), juin 2021, p. 1-6. doi: 10.1109/3D-PEIM49630.2021.9497265.

[12] C.-T. Ma et Z.-H. Gu, « Review of GaN HEMT Applications in Power Converters over 500 W », Electronics, vol. 8, no 12, Art. no 12, déc. 2019, doi: 10.3390/electronics8121401.

[13] X. Huang, Z. Liu, Q. Li, et F. C. Lee, « Evaluation and Application of 600 V GaN HEMT in Cascode Structure », IEEE Trans. Power Electron., vol. 29, no 5, p. 2453-2461, mai 2014, doi: 10.1109/TPEL.2013.2276127.

[14] S. Elangovan, S. Cheng, et E. Y. Chang, « Reliability Characterization of Gallium Nitride MIS-HEMT Based Cascode Devices for Power Electronic Applications », Energies, vol. 13, no 10, Art. no 10, janv. 2020, doi: 10.3390/en13102628.

[15] Y. Li et al., « Evaluation and Analysis of Temperature-Dependent Dynamic $R_\mathrmDS,\mathrmON$ of GaN Power Devices Considering High-Frequency Operation », IEEE J. Emerg. Sel. Top. Power Electron., vol. 8, no 1, p. 111-123, mars 2020, doi: 10.1109/JESTPE.2019.2947575.

[16] J. Wu, W. Meng, F. Zhang, G. Dong, et J. Shu, « A Short-Circuit Protection Circuit With Strong Noise Immunity for GaN HEMTs », IEEE Trans. Power Electron., vol. 36, no 2, p. 2432-2442, févr. 2021, doi: 10.1109/TPEL.2020.3013984.

[17] J. Acuna, J. Walter, et I. Kallfass, « Very Fast Short Circuit Protection for Gallium-Nitride Power Transistors Based on Printed Circuit Board Integrated Current Sensor », in 2018 20th European Conference on Power Electronics and Applications (EPE'18 ECCE Europe), sept. 2018, p. P.1-P.10.

[18] X. Huang et al., « Experimental study of 650V AlGaN/GaN HEMT short-circuit safe operating area (SCSOA) », in 2014 IEEE 26th International Symposium on Power Semiconductor Devices IC's (ISPSD), juin 2014, p. 273-276. doi: 10.1109/ISPSD.2014.6856029.

[19] R. Mitova, R. Ghosh, U. Mhaskar, D. Klikic, M.-X. Wang, et A. Dentella, « Investigations of 600-V GaN HEMT and GaN Diode for Power Converter Applications », IEEE Trans. Power Electron., vol. 29, no 5, p. 2441-2452, mai 2014, doi: 10.1109/TPEL.2013.2286639.

[20] J. Sun, J. Wei, Z. Zheng, et K. J. Chen, « Short Circuit Capability Characterization and Analysis of p-GaN Gate High-Electron-Mobility Transistors Under Single and Repetitive Tests », IEEE Trans. Ind. Electron., vol. 68, no 9, p. 8798-8807, sept. 2021, doi: 10.1109/TIE.2020.3009603.

[21] D. Ueda et al., « Present and future prospects of gan-based power electronics », in 2008 9th International Conference on Solid-State and Integrated-Circuit Technology, oct. 2008, p. 1078-1081. doi: 10.1109/ICSICT.2008.4734738.

[22] GaN Systems, « GN001 Application Note "An introduction to GaN Enhancement-mode HEMTs ». avr. 16, 2020.

[23] G. Greco, F. Iucolano, et F. Roccaforte, « Review of technology for normally-off HEMTs with p-GaN gate », Mater. Sci. Semicond. Process., vol. 78, p. 96-106, mai 2018, doi: 10.1016/j.mssp.2017.09.027.

[24] C. Abbate, F. Iannuzzo, et G. Busatto, « Thermal instability during short circuit of normally-off AlGaN/GaN HFETs », Microelectron. Reliab., vol. 53, no 9, p. 1481-1485, sept. 2013, doi: 10.1016/j.microrel.2013.07.119.

[25] H. Hamza K., D. Nirmal, et L. Arivazhagan, « Impact of AlGaN Back Barrier in AlGaN/GaN HEMT on GaN substrate », in 2020 5th International Conference on Devices, Circuits and Systems (ICDCS), mars 2020, p. 290-293. doi: 10.1109/ICDCS48716.2020.243601.

[26] H. Ohta, T. Nakamura, et T. Mishima, « High quality free-standing GaN substrates and their application to high breakdown voltage GaN p-n diodes », in 2016 IEEE International Meeting for Future of Electron Devices, Kansai (IMFEDK), juin 2016, p. 1-2. doi: 10.1109/IMFEDK.2016.7521696.

[27] T. Yoshida et al., « Fabrication of 3-in GaN substrates by hydride vapor phase epitaxy using void-assisted separation method », J. Cryst. Growth, vol. 310, no 1, p. 5-7, janv. 2008, doi: 10.1016/j.jcrysgro.2007.10.014.

[28] M. Imanishi, S. Usami, M. Maruyama, M. Yoshimura, et Y. Mori, « Growth of a High Quality GaN Wafer from Point Seeds by the Na-Flux Method », in 2021 28th International Workshop on Active-Matrix Flatpanel Displays and Devices (AM-FPD), juin 2021, p. 70-72. doi: 10.23919/AM-FPD52126.2021.9499151.

[29] T. Nagahisa, H. Ichijoh, T. Suzuki, A. Yudin, A. O. Adan, et M. Kubo, « Robust 600 V GaN high electron mobility transistor technology on GaN-on-Si with 400 V, 5 µs load-short-circuit withstand capability », Jpn. J. Appl. Phys., vol. 55, no 4S, p. 04EG01, févr. 2016, doi: 10.7567/JJAP.55.04EG01.

[30] H. W. Then et al., « Gallium Nitride and Silicon Transistors on 300 mm Silicon Wafers Enabled by 3-D Monolithic Heterogeneous Integration », IEEE Trans. Electron Devices, vol. 67, no 12, p. 5306-5314, déc. 2020, doi: 10.1109/TED.2020.3034076.

[31] W. B. Lanford, T. Tanaka, Y. Otoki, et I. Adesida, « Recessed-gate enhancement-mode GaN HEMT with high threshold voltage », Electron. Lett., vol. 41, no 7, p. 449-450, mars 2005, doi: 10.1049/el:20050161.

[32] H. Kambayashi et al., « Over 100A operation normally-off AlGaN/GaN hybrid MOS-HFET on Si substrate with high-breakdown voltage », Solid-State Electron., vol. 54, no 6, p. 660-664, juin 2010, doi: 10.1016/j.sse.2010.01.001.

[33] M. Fernández et al., « Short-Circuit Study in Medium-Voltage GaN Cascodes, p-GaN HEMTs, and GaN MISHEMTs », IEEE Trans. Ind. Electron., vol. 64, no 11, p. 9012-9022, nov. 2017, doi: 10.1109/TIE.2017.2719599.

[34] Y. Uemoto et al., « Gate Injection Transistor (GIT)—A Normally-Off AlGaN/GaN Power Transistor Using Conductivity Modulation », IEEE Trans. Electron Devices, vol. 54, no 12, p. 3393-3399, déc. 2007, doi: 10.1109/TED.2007.908601.

[35] P. Xue, L. Maresca, M. Riccio, G. Breglio, et A. Irace, « Experimental Study on the Short-Circuit Instability of Cascode GaN HEMTs », IEEE Trans. Electron Devices, vol. 67, no 4, p. 1686-1692, avr. 2020, doi: 10.1109/TED.2020.2974518.

[36] S. Ben-Yaakov et L. Van de Perre, « A Novel Circuit Topology for Turning a 'Normally On' GaN Transistor into 'Normally Off' that Can be Driven by Popular Drivers ». juin 2018.

[37] X. Huang, W. Du, F. C. Lee, Q. Li, et W. Zhang, « Avoiding Divergent Oscillation of a Cascode GaN Device Under High-Current Turn-Off Condition », IEEE Trans. Power Electron., vol. 32, no 1, p. 593-601, janv. 2017, doi: 10.1109/TPEL.2016.2532799.

[38] S. Chowdhury, B. L. Swenson, M. H. Wong, et U. K. Mishra, « Current status and scope of gallium nitride-based vertical transistors for high-power electronics application », Semicond. Sci. Technol., vol. 28, no 7, p. 074014, juin 2013, doi: 10.1088/0268-1242/28/7/074014.

[39] Nexgen, « Site internet de Nexgen ». [En ligne]. Disponible sur: https://nexgenpowersystems.com/datasheets

[40] K. Strempel et al., « Vertical 3D gallium nitride field-effect transistors based on fin structures with inverted p-doped channel », Semicond. Sci. Technol., vol. 36, p. 9, nov. 2020, doi: 10.1088/1361-6641/abc5ff.

[41] I. Ben-Yaacov, Y.-K. Seck, U. K. Mishra, et S. P. DenBaars, « AlGaN/GaN current aperture vertical electron transistors with regrown channels », J. Appl. Phys., vol. 95, no 4, p. 2073-2078, févr. 2004, doi: 10.1063/1.1641520.

[42] T. Oka, T. Ina, Y. Ueno, et J. Nishii, « 1.8 mΩ·cm2 vertical GaN-based trench metal-oxide-semiconductor field-effect transistors on a free-standing GaN substrate for 1.2-kV-class operation », Appl. Phys. Express, vol. 8, no 5, p. 054101, avr. 2015, doi: 10.7567/APEX.8.054101.

[43] F. N. F. Zaki, « Characterization, modeling and aging behavior of GaN power transistors », phdthesis, Université Paris Saclay (COmUE), 2018. Consulté le : sept. 09, 2021. [En ligne]. Disponible sur : https://tel.archives-ouvertes.fr/tel-01794546

[44] « Panasonic GaN power transistors white paper ».

[45] G. Meneghesso et al., « Reliability and parasitic issues in GaN-based power HEMTs: a review », Semicond. Sci. Technol., vol. 31, no 9, p. 093004, août 2016, doi: 10.1088/0268-1242/31/9/093004.

[46] D. Reusch et J. Strydom, « Understanding the Effect of PCB Layout on Circuit Performance in a High-Frequency Gallium-Nitride-Based Point of Load Converter », IEEE Trans. Power Electron., vol. 29, no 4, p. 2008-2015, avr. 2014, doi: 10.1109/TPEL.2013.2266103.

[47] Z. Liu, F. C. Lee, Q. Li, et Y. Yang, « Design of GaN-based MHz totem-pole PFC rectifier », in 2015 IEEE Energy Conversion Congress and Exposition (ECCE), sept. 2015, p. 682-688. doi: 10.1109/ECCE.2015.7309755.

[48] H. Li, X. Li, Z. Zhang, C. Yao, et J. Wang, « Design consideration of high power GaN inverter », in 2016 IEEE 4th Workshop on Wide Bandgap Power Devices and Applications (WiPDA), nov. 2016, p. 23-29. doi: 10.1109/WiPDA.2016.7799904.

[49] W. Zhang, F. Wang, D. J. Costinett, L. M. Tolbert, et B. J. Blalock, « Investigation of Gallium Nitride Devices in High-Frequency LLC Resonant Converters », IEEE Trans. Power Electron., vol. 32, no 1, p. 571-583, janv. 2017, doi: 10.1109/TPEL.2016.2528291.

[50] P. Grzejszczak, A. Kulpa, et R. Barlik, « Design of a high-efficiency three-phase three-level NPC converter based on GaN HEMT and SiC SB diode », in 2019 Progress in Applied Electrical Engineering (PAEE), juin 2019, p. 1-6. doi: 10.1109/PAEE.2019.8788993.

[51] E. Maset et al., « Optimized Design of 1 MHz Intermediate Bus Converter Using GaN HEMT for Aerospace Applications », Energies, vol. 13, p. 6583, déc. 2020, doi: 10.3390/en13246583.

[52] D. Bisi et al., « Short-Circuit Capability Demonstrated for GaN Power Switches », in 2021 IEEE Applied Power Electronics Conference and Exposition (APEC), juin 2021, p. 370-375. doi: 10.1109/APEC42165.2021.9486987.

Course Voitures Autonomes Paris Saclay (CoVAPSy) : Travaux pratiques autour des voitures autonomes

Thomas BOULANGER[1] – Eve DÉLÈGUE[2] – Kévin HOARAU[3]
Anthony JUTON[4]

La Revue 3EI
Culture Sciences de l'Ingénieur

Édité le 06/11/2023

école normale supérieure paris–saclay

[1] *Élève en année de recherche pré-doctorale à l'étranger, ENS Paris-Saclay – DER Nikola Tesla*
[2] *Élève en année de recherche en intelligence artificielle, ENS Paris-Saclay – DER Nikola Tesla*
[3] *Élève en M2 Formation à l'Enseignement Supérieur, ENS Paris-Saclay – DER Nikola Tesla*
[4] *Professeur agrégé de physique appliquée au DER Nikola Tesla, ENS Paris-Saclay*

Cette ressource fait partie du N°111 de La Revue 3EI de janvier 2024.

Portés par l'essor de la recherche sur les véhicules autonomes, dans le but de travailler l'informatique embarquée et/ou l'intelligence artificielle, les établissements de l'Université Paris Saclay organisent depuis 2019 une course de voitures autonomes 1/10ème, ouverte à tous (voir la vidéo « *Course de voitures autonomes 2023* » [1]). Au fil de ces quatre années, les solutions ont mûri et sont devenues fiables, ce qui nous permet aujourd'hui de diffuser ces travaux.

Figure 1 : Voiture sur le simulateur webots

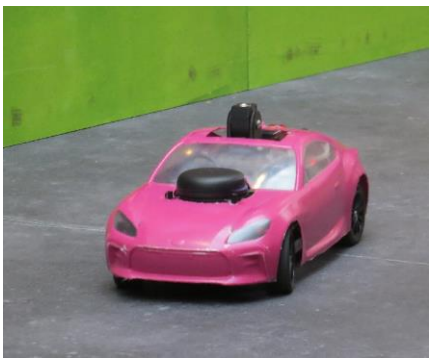


Figure 2 : Voiture réelle

Cette ressource présente et introduit deux supports de travaux pratiques et/ou projets, qui peuvent être travaillés indépendamment ou de façon complémentaire, en NSI, en SI, en BTS CIEL, en BUT GEII ou en école d'ingénieur :

- Une simulation sur le logiciel Webots où les étudiants travaillent sur le code de conduite autonome d'une voiture, en python ou en C, sur une piste virtuelle afin de parcourir la piste le plus rapidement possible [2]. Il est possible de faire concourir plusieurs voitures ensemble.
- Une partie expérimentale sur une voiture 1/10ème, mettant en œuvre l'acquisition des données du lidar, les commandes des moteurs et l'utilisation d'un nano-ordinateur raspberry pi [3] (python) ou d'un microcontrôler STM32 (langage C) [5]. L'algorithme de conduite autonome peut être travaillé directement sur la voiture réelle ou repris du code du simulateur compatible.
- Le travail à la fois sur simulateur et sur la voiture réelle permet d'étudier les limites de la simulation, l'amélioration du passage simulation → réalité...

Pour aller plus loin ou pour participer à la course de Paris Saclay, des ressources sont également disponibles sur le site de la course de voitures autonomes du plateau de Saclay [4] : https://ajuton-ens.github.io/CourseVoituresAutonomesSaclay/

La vidéo de l'édition 2023 est disponible sur Culture Sciences de l'Ingénieur [1].

Figure 3 : Grille de départ de la course de voitures autonomes de Paris Saclay 2023, [1]

Les pré-requis sont l'utilisation de python ou du C au choix (utilisation de fonctions, utilisation des tableaux) pour la simulation comme pour la partie expérimentale.

1 - Introduction

Depuis une vingtaine d'années, l'industrie automobile et les organismes de recherche ont pour objectif de faire rouler sur route ouverte des voitures autonomes, sans conducteur. Cette technologie offrirait la possibilité de voyager de manière plus sûre, en utilisant le temps du voyage pour d'autres activités de loisirs ou professionnelles.

Des voitures autonomes (Waymo, Renault notamment) sont déjà en phase de test dans le monde, mais leur adoption par le public soulève des questions importantes concernant notamment la sécurité des usagers (les Tesla ont notamment eu de graves problèmes de « freinage fantôme ») et la responsabilité en cas d'accident.

Par ailleurs, des voitures à l'autonomie partielle sont déjà commercialisées : parking automatique, maintien dans la voie sur autoroute sont des fonctionnalités courantes sur les voitures haut de gamme.

Figure 4 : Renault Zoe Cab, en phase de test (laboratoire Paris-Saclay Autonomous Lab), source Renault Group

Afin de pouvoir se déplacer sans conducteur, les voitures autonomes font appel à de nombreuses technologies intéressantes pour l'enseignement des sciences de l'ingénieur et de l'informatique. On peut les regrouper autour de trois fonctions principales :

- observer l'environnement autour de la voiture et s'y repérer à l'aide de capteurs (GPS, lidar, caméra, radar, télémètres ultrason, centrale inertielle, ...),
- décider de la direction et de la vitesse à l'aide d'un contrôleur embarqué,
- agir sur la direction et la propulsion (direction motorisée et moteur de la voiture).

Figure 5 : Robot Taxi Waymo, actuellement en circulation en phase de test à Phoenix, Arizona et Los Angeles, Californie, Source Waymo

2 - Présentation de la voiture

La voiture utilisée pour la course de voitures autonomes de Paris Saclay (CoVAPSy) est au format 1/10ème, ce qui permet un coût raisonnable (1000 à 1500 euros) et un risque nul.

Le châssis TT-02 choisi est robuste et bon marché. Il est possible de trouver des pièces détachées chez tous les fournisseurs de modélisme. La voiture complète permet l'usage de microcontrôleur (Arduino ou STM32) et/ou d'un nano-ordinateur (raspberry Pi) et/ou d'un GPU (Jetson Nano), de lidar et/ou caméra, de télémètres et d'asservissement de vitesse.

Les schémas des cartes électroniques, les plans des pièces mécaniques et les références des composants sont disponibles sur le dépôt git de la course [4].

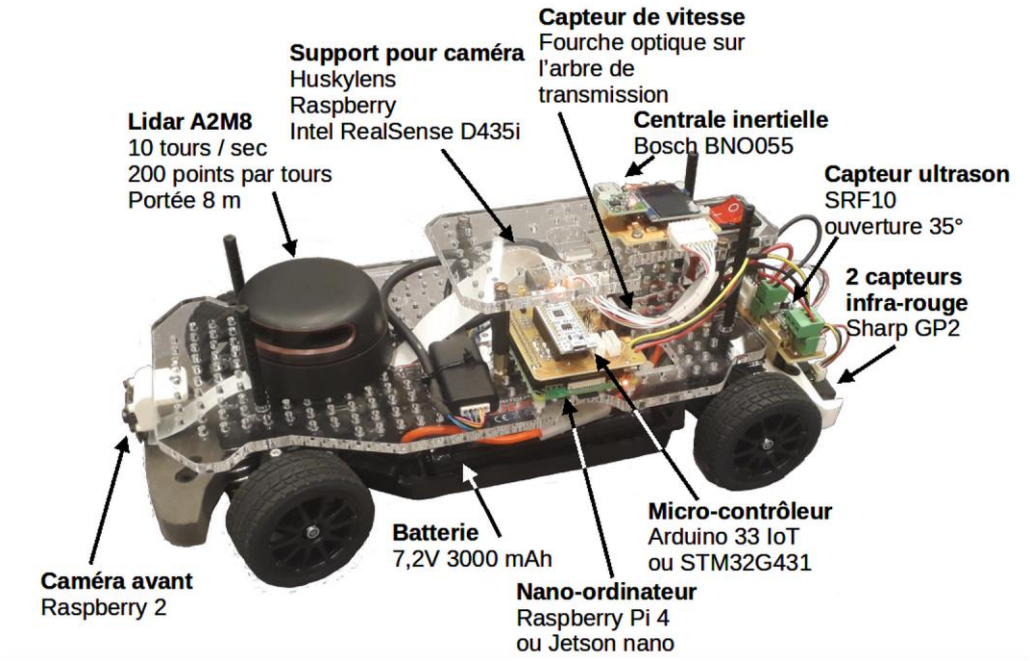


Figure 6 : Voiture autonome CoVAPSy 2023

Le diagramme SysML de blocs internes de la voiture CoVAPSy complète, sans caméra, est le suivant :

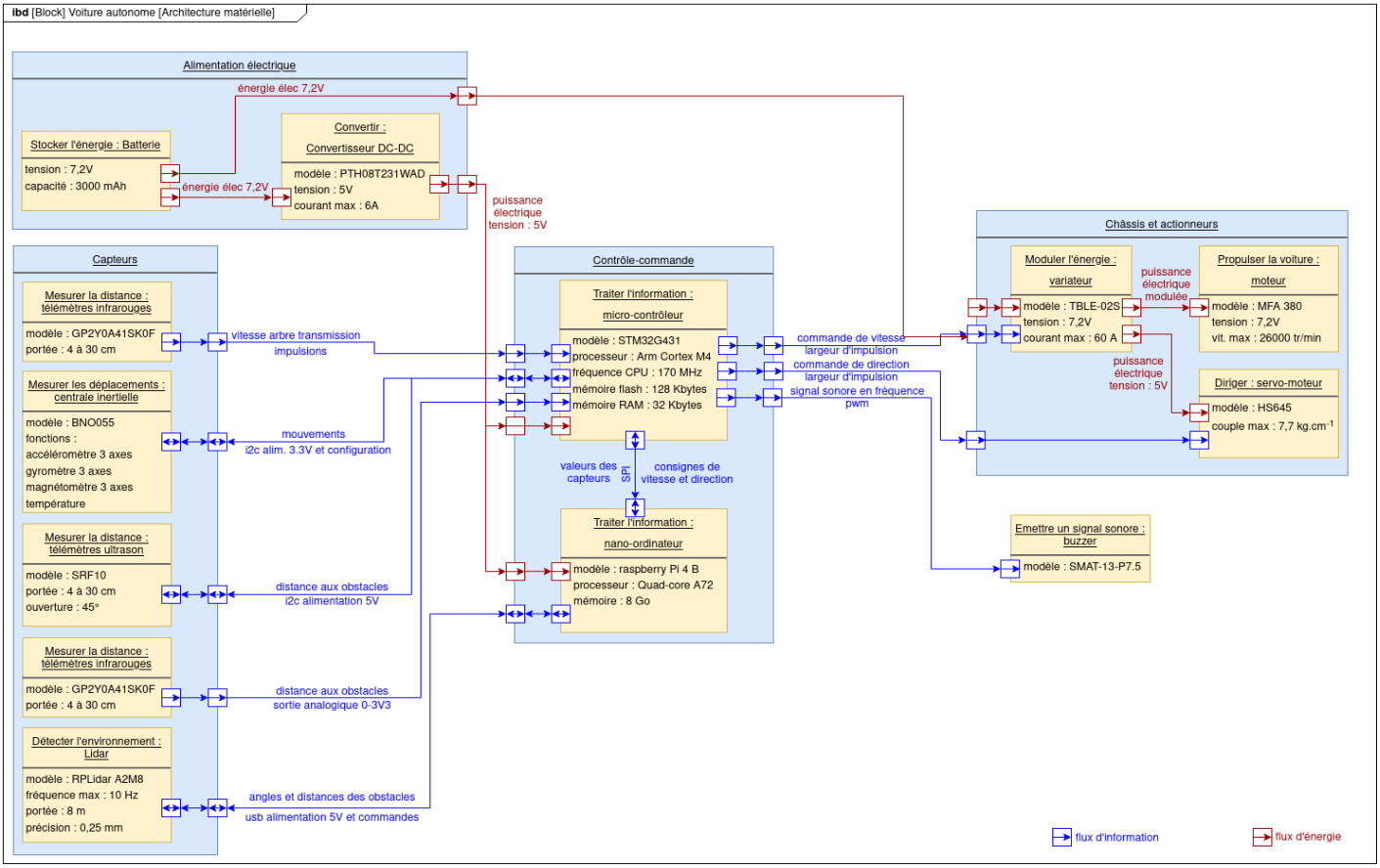


Figure 7 : Diagramme SysML de blocs internes de la voitures CoVAPSy 2023

Pour la première approche proposée par cette série de ressources (simulation et mise en œuvre expérimentale), une version simplifiée *CoVAPSy_RPIonly* est présentée, avec seulement un nano-ordinateur raspberry Pi et un lidar, pour un coût d'environ 650 euros.

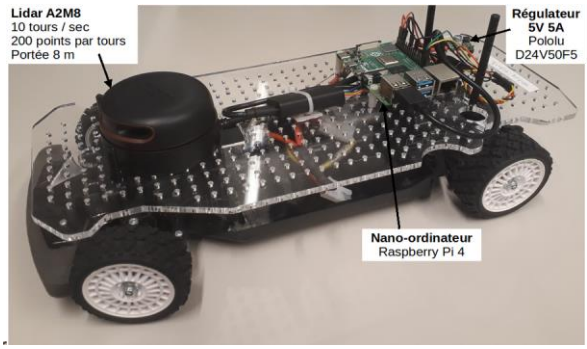


Figure 8 : Voiture autonome CoVAPSy RPIonly

Figure 9 : Voiture autonome CoVAPSy_RPIonly avec sa carrosserie

La liste des références des composants et de fournisseurs potentiels est fournie en annexe (§6).

Une version similaire de cette voiture simplifiée, contrôlée par un microcontrôleur STM32 programmé en langage C, est présentée dans la ressource « *CoVAPSy : Premiers programmes en langage C sur voiture réelle* » [5].

La même voiture a été conçue sur le simulateur Webots, avec des dimensions et des caractéristiques dynamiques les plus proches possible du châssis TT-02 utilisé. La carrosserie, issue d'un modèle de Chevrolet Camaro dessiné spécialement pour être simple à simuler, est un peu différente :

Figure 10 : La voiture CoVAPSy sur le simulateur

Les différents éléments de la voiture simplifiée sont présentés sur le diagramme de blocs internes suivant :

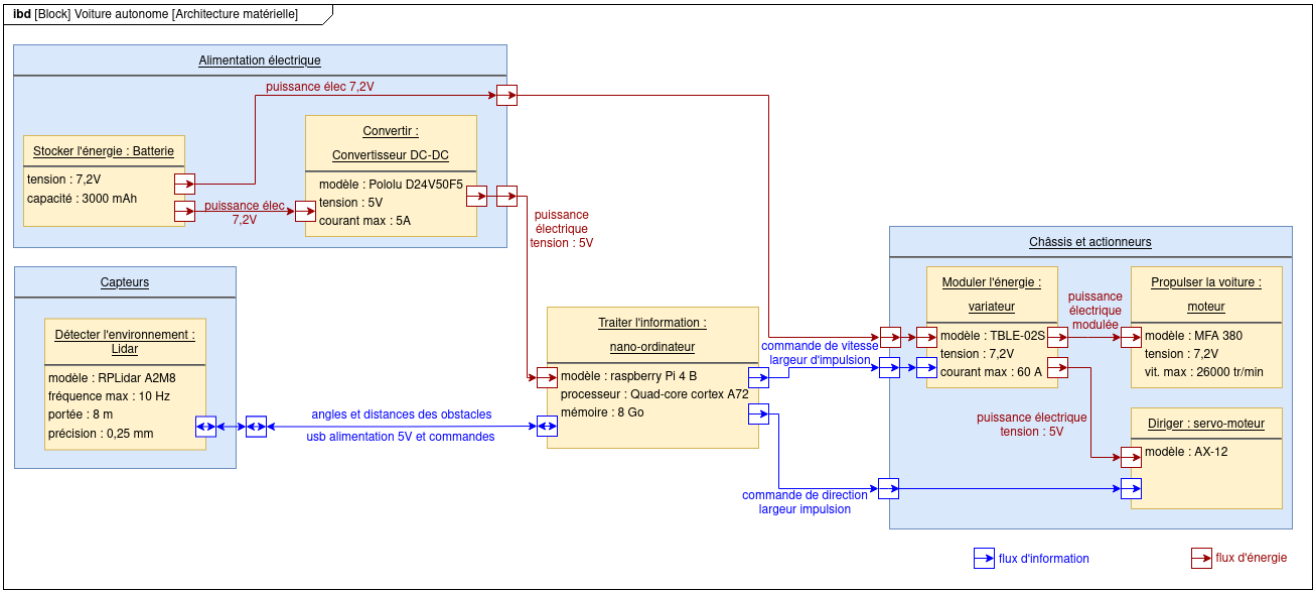


Figure 11 : Diagramme SysML de blocs internes de la voiture CoVAPSY_RPIonly

Un Lidar (*laser imaging detection and ranging* soit en français *détection et estimation de la distance par laser)* est un télémètre tournant. Une impulsion laser est émise par le Lidar, elle se réfléchit sur les objets qu'elle rencontre et revient sur un photorécepteur du Lidar. Le lidar A2M8 (ou son successeur A2M12) est un lidar bon marché utilisant la triangulation pour estimer la distance à l'objet ayant réfléchi l'impulsion laser.

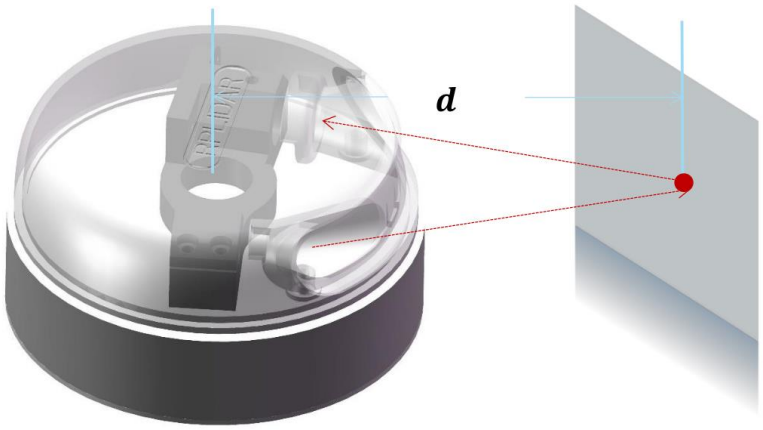


Figure 12 : Principe de la mesure de distance du Lidar Slamtec A2M8, source Slamtec

Les lidars plus précis (Slamtec S2 par exemple ou lidar professionnels Hokuyo) mesurent le temps de l'aller-retour (ToF Time of Flight) de l'impulsion laser pour estimer la distance à l'objet.

Le détecteur tourne sur lui-même afin de connaître la distance de l'ensemble des objets entourant la voiture. Les données du lidar sont fournies, sur le simulateur comme avec la voiture réelle, dans un tableau de 360 cases, chaque case correspondant à la mesure à 1 degré, l'angle 0 étant devant la voiture.

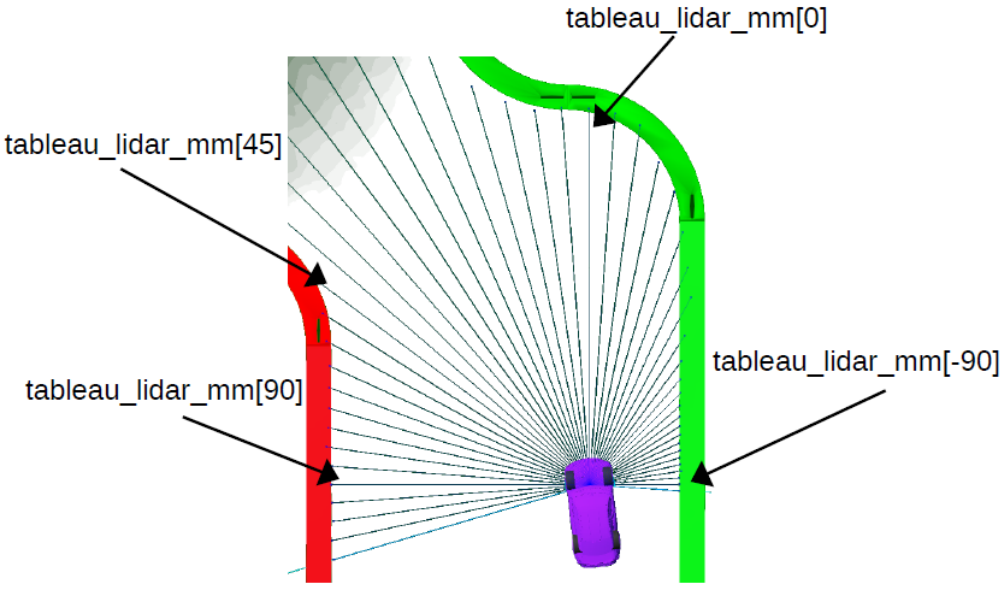


Figure 13 : Affichage sur le simulateur d'un rayon de mesure sur 4 du lidar

Deux moteurs sont utilisés sur la voiture, issus du modèle de voiture radiocommandée. Le premier sert à la propulsion de la voiture et est alimenté par un variateur de vitesse. Le second contrôle la direction de la voiture. C'est un servomoteur, un moteur asservi en position. Les deux reçoivent du nano-ordinateur ou du microcontrôleur une commande sous forme d'impulsion :

Figure 14 : Impulsion de commande du servomoteur et du variateur de propulsion

Suivant les marques de servomoteur et de variateur (Tamiya ne met pas toujours les mêmes variateurs dans ses kits de voitures 1/10ème et le servomoteur est acheté à part), les sens de rotation et les butées changent de valeurs (toujours entre 1 et 2 ms). Il est donc important de prendre le temps de calibrer ces butées, comme c'est expliqué dans la ressource « *CoVaPSy : Premiers programmes python sur la voiture réelle* » [2].

Une batterie NiMH 7,2V 3000 mAh fournit l'énergie à tous les éléments de la voiture (lidar, nano-ordinateurs, moteurs, …). Un régulateur Pololu D24V50F5 génère une tension 5V à partir de la tension 7,2V de la batterie pour l'alimentation du nano-ordinateur. Le variateur du moteur de propulsion fournit quant à lui la tension 5V d'alimentation du servomoteur de direction.

Un nano-ordinateur a la charge de piloter la voiture. La carte raspberry Pi, très populaire, regroupe une large communauté d'utilisateurs partageant leurs expériences, des guides de mise en œuvre, des forums de résolution de problème. Elle utilise une distribution linux pour laquelle existe une bibliothèque (python ou c++) pour l'utilisation du Lidar et une (également python ou c++) pour la génération d'impulsions en vue de commander les moteurs.

3 - Course de voitures simulées

La simulation utilise le logiciel Webots. La programmation peut être faite en python ou en C.

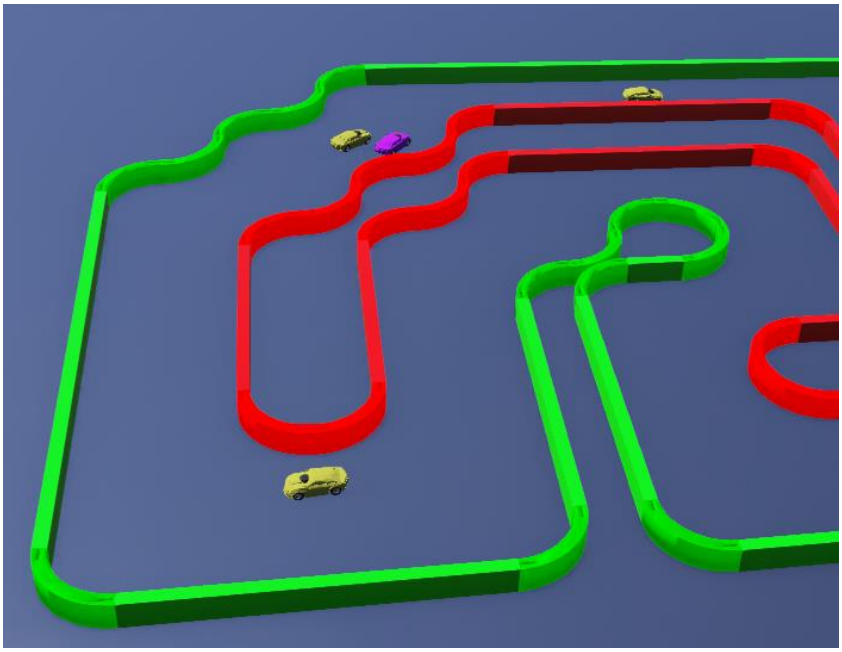


Figure 15 : Course de voitures autonomes sur le simulateur, entre une voiture étudiant (violette) et 3 voitures préprogrammées (« sparring partners ») en jaune

Le simulateur fourni propose une piste et un modèle de voiture proche de la voiture réelle (blocs jaunes de la figure 16). Les fonctions de traitement des données du lidar, de contrôle de la vitesse

et de la direction sont données (en violet sur la figure 16), ainsi qu'un algorithme de conduite basique (en rouge sur la figure 16), en C sur l'une des voitures et en python sur une autre.

Le travail des élèves / étudiants est dans un premier temps de faire un programme de conduite plus performant que celui des voitures préprogrammées (sparring partners).

Il est ensuite possible de positionner plusieurs voitures sur la piste pour faire une course entre les voitures programmées par différents étudiants. Les pistes de travail sont nombreuses : Côté programmation, on peut travailler sur les machines à état pour le contrôle de la voiture ou ajouter un nœud superviseur pour mesurer le temps au tour des voitures. Côté asservissement, on peut travailler sur les algorithmes d'asservissement classiques (PID) ou sur des méthodes avancées.

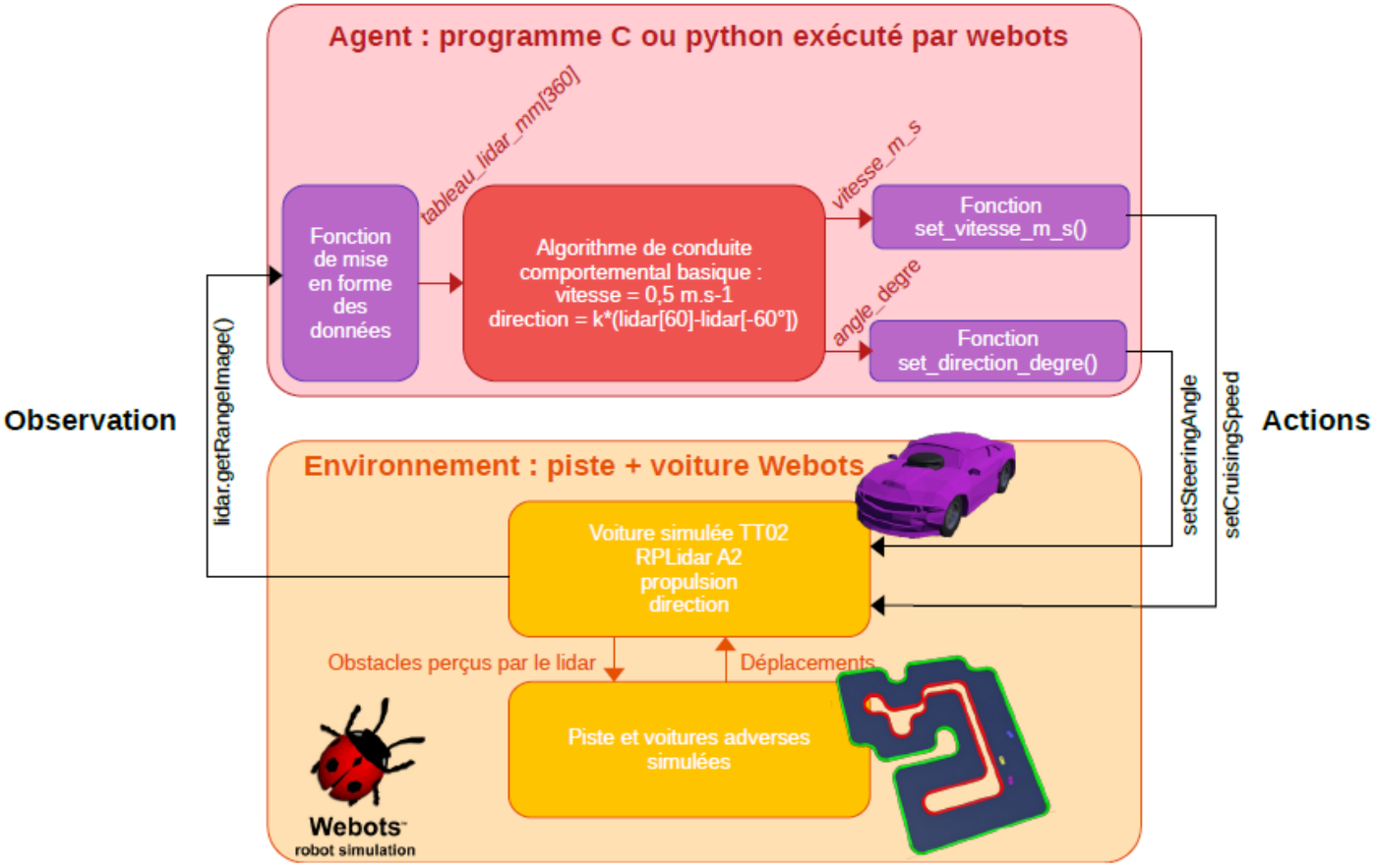


Figure 16 : Schéma des blocs logiciels fournis pour le travail sur le simulateur

Le ressource « *CoVapSy : Mise en œuvre du simulateur Webots* » [2] décrit en détail la prise en main du simulateur et les fonctions fournies, jusqu'au premier programme de conduite basique.

4 - Course de voitures réelles

La ressource « *CoVAPSy : Premiers programmes python sur la voiture réelle* » [3] traite de la mise en œuvre de la voiture CoVAPSy_RPIonly, programmable en python. La ressource « *CoVAPSy : Premiers programmes en langage C sur voiture réelle* » [5] propose la mise en œuvre avec la voiture CoVAPSy_STM32, programmée en C.

La voiture simulée ayant été conçue pour être la plus proche possible de la voiture réelle, des fonctions de traitement des données du lidar, de contrôle de la vitesse et de la direction (en rouge sur la figure 17) ont été développées avec un comportement proche de celles du simulateur. Le même algorithme de conduite basique (en rouge) peut donc être utilisé.

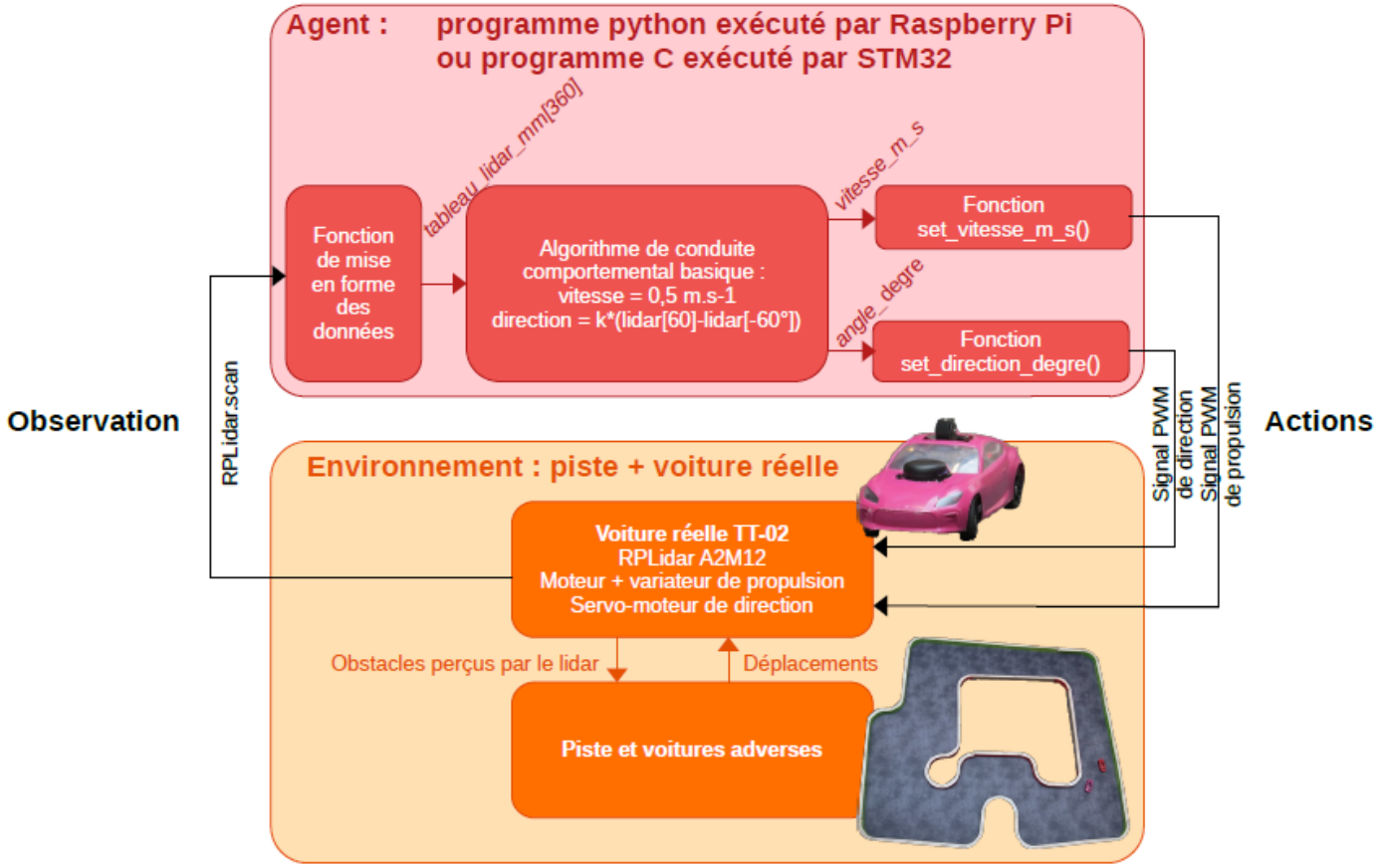


Figure 17 : Schéma des blocs logiciels fournis pour le travail sur la voiture réelle

5 - Conclusion

Ces deux activités permettent la simulation du fonctionnement d'une voiture autonome sur un circuit virtuel et la mise en pratique sur une piste réelle. Ces activités menées en parallèle permettent de mettre en évidence l'écart entre le domaine simulé et le domaine réel et de travailler à l'amélioration du modèle (prise en compte de la dynamique de la direction), à l'amélioration de la voiture réelle (asservissement de vitesse), pour rendre plus robuste le passage simulation → réalité.

Cette ressource s'appuie sur l'utilisation d'un seul Lidar pour la perception de l'environnement, bien adapté à une piste avec des bordures. Les voitures autonomes réelles circulent sur des routes sans bordure et utilisent notamment des caméras pour repérer le marquage au sol. L'utilisation de caméras implique des unités de calcul plus complexes et des algorithmes de traitement d'image avancés. L'ajout d'une caméra sur la voiture autonome 1/10ème permet d'approcher ces problématiques, mais au prix d'une complexité qui dépasse le public bac/bac+2 ciblé par ces ressources.

6 - Annexe

Désignation	Fabricant	Référence fabricant	Fournisseur	Réf fournisseur	Prix TTC	Remarque
Tamiya TT-02 Toyota GR 86 KIT 58694	Tamiya	58694	RCTeam	58694	134,90 €	disponible aussi chez Conrad
Orion Chargeur IQ801 1A ORI30197	Orion	IQ801	RCTeam	ORI30197	15,90 €	disponible aussi chez Conrad
Accu 7.2v Nimh 3000mah Tamiya T1006300	Tamiya	T1006300	RCTeam	T1006300	27,30 €	disponible aussi chez Conrad
Servomoteur Konect 9kg 0.13s KN-0913LVM	Konect	KN-0913LVMG	RCTeam	KN-0913LVMG	19,90 €	disponible aussi chez Conrad
Scanner Laser 360° RPLIDAR A2M12	Slamtec	RPLIDAR A2M12	Robotshop	RB-Rpk-22	263,90 €	disponible aussi chez Gotronic
Convertisseur DC-DC Pololu D24V50F5	Pololu	D24V50F5	Robotshop	RB-Pol-295	56,41 €	disponible aussi chez Gotronic
Raspberry Pi 4 Modèle B, 8Go	Raspberry	PI4-8GB	Kubii	PI48GB	95,40 €	appeler en tant qu'établissement
Carte Micro-SD SanDisk Classe 10 32GB	SanDisk	SDXC CI10 UHS-1	Kubii	SD_SANDISK	11,95 €	d'enseignement si rupture de stock.
				TOTAL TTC	625,66 €	

Tableau 1 : Références et fournisseurs pour la voiture CoVAPSy_RPIonly

Références :

[1]: Course de voitures autonomes 2023, mai 2023, https://eduscol.education.fr/sti/si-ens-paris-saclay/actualites/course-voitures-autonomes-2023-resultats

[2]: CoVAPSy : Mise en œuvre du Simulateur Webots, T. Boulanger, E. Délègue, K. Hoarau, A. Juton, https://eduscol.education.fr/sti/si-ens-paris-saclay/ressources_pedagogiques/covapsy-mise-en-oeuvre-du-simulateur-webots

[3]: CoVAPSy : Premiers programmes python sur la voiture réelle, T. Boulanger, E. Délègue, K. Hoarau, A. Juton, https://eduscol.education.fr/sti/si-ens-paris-saclay/ressources_pedagogiques/covapsy-premiers-programmes-python-sur-voiture-reelle

[4]: Dépôt git de la course de voitures autonomes de Paris Saclay : https://github.com/ajuton-ens/CourseVoituresAutonomesSaclay

[5]: CoVAPSy : Premiers programmes en langage C sur voiture réelle , A. Azan, A. Juton, https://eduscol.education.fr/sti/si-ens-paris-saclay/ressources_pedagogiques/covapsy-premiers-programmes-langage-c-sur-la-voiture-reelle

CoVAPSy : Premiers programmes python sur la voiture réelle

La Revue 3E.I
Culture Sciences de l'Ingénieur

Thomas BOULANGER[1] – Eve DÉLÈGUE[2]– Kévin HOARAU[3]
Anthony JUTON[4]

Édité le
16/11/2024

école normale supérieure paris–saclay

[1] *Élève en année de recherche pré-doctorale à l'étranger, ENS Paris-Saclay – DER Nikola Tesla*
[2] *Élève en année de recherche en intelligence artificielle, ENS Paris-Saclay – DER Nikola Tesla*
[3] *Élève en M2 Formation à l'Enseignement Supérieur, ENS Paris-Saclay – DER Nikola Tesla*
[4] *Professeur agrégé de physique appliquée au DER Nikola Tesla, ENS Paris-Saclay*

Cette ressource fait partie du N°111 de La Revue 3EI de janvier 2024.

Faisant suite à la ressource d'introduction « Course Voitures Autonomes Paris Saclay (CoVAPSy) : Travaux pratiques autour des voitures autonomes » [1] et parallèle à la ressource de simulation « CoVAPSy : Mise en œuvre du Simulateur Webots » [2], l'objectif de cette ressource est de mener le lecteur au démarrage, à la configuration et à la programmation des premiers pas d'une voiture autonome réelle, la plus simple possible (CoVAPSy RPIonly utilise juste un nano-ordinateur raspberry Pi et un lidar).

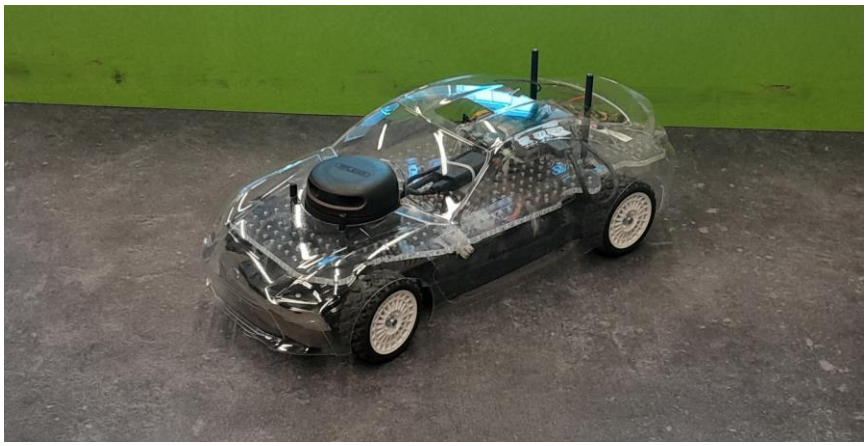


Figure 1 : CoVAPSy_RPIonly

Pour se faire, plusieurs signaux vont devoir être récupérés des capteurs et d'autres devront être générés afin d'activer les actionneurs. Les fonctions et tableaux utilisés pour cela sont les mêmes que sur le simulateur webots présenté dans la ressource « *CoVAPSy : Mise en œuvre du Simulateur Webots* »[2].

Les programmes de test sont disponibles sur le dépôt github :

https://github.com/ajuton-ens/CourseVoituresAutonomesSaclay/tree/main/Bibliotheques_logicielles/programmes_python_base_lidar_propulsion_direction_conduite

1 - Réalisation de la voiture

Une fois la voiture montée en suivant les instructions Tamiya, il faut lui ajouter l'électronique pour pouvoir la programmer.

1.1 - Montage mécanique de l'électronique

Pour fixer le nano-ordinateur raspberry Pi, le lidar et le convertisseur DC/DC, le plus simple est de réaliser une plaque découpée / percée au laser ou à la scie sauteuse / perceuse. Voici un exemple dont le plan est fourni en stp et dxf à l'adresse suivante : https://github.com/ajuton-ens/CourseVoituresAutonomesSaclay/tree/main/Materiel/Pieces_mecaniques_stp_dxf

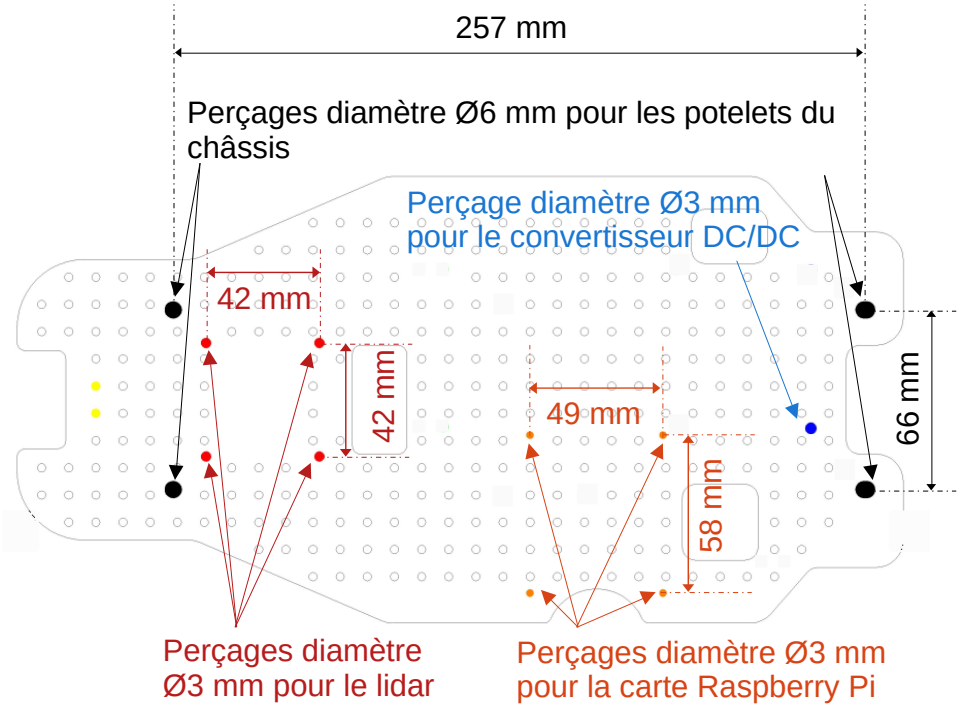


Figure 2 : Découpe et perçage pour le support du lidar et de l'électronique

La découpe de la carrosserie pour laisser passer le lidar peut se faire avec une mini-fraiseuse de type Dremel, ou avec une machine de découpe laser

Figure 3 : Découpe du passage du lidar au laser

1.2 - Câblage de l'électronique

Le câblage de l'électronique se fait par soudure pour la puissance et par soudure ou câbles de prototypage pour les signaux.

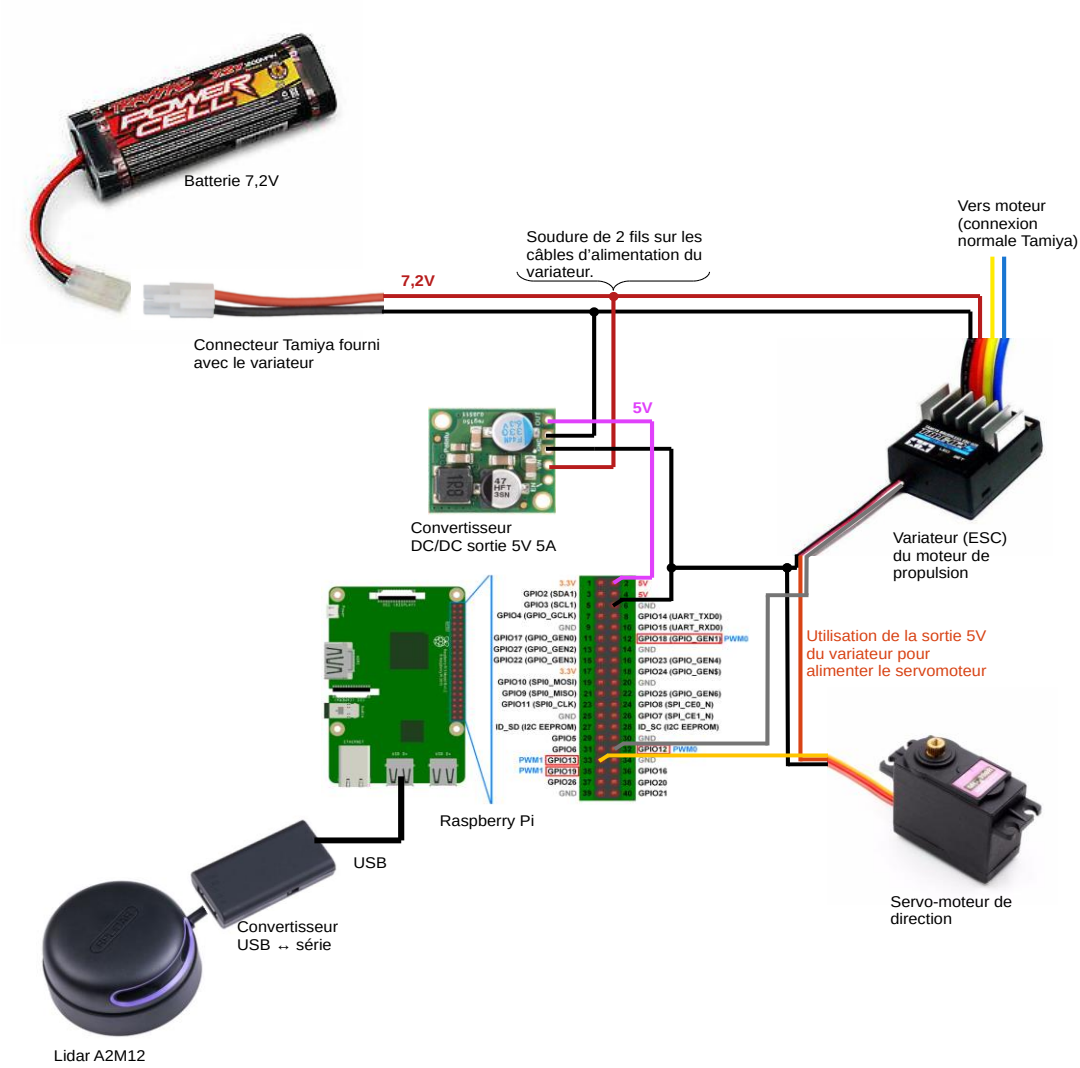


Figure 4 : Câblage des composants de la voiture CoVAPSY_RPIonly

2 - Installation de la carte Raspberry Pi

L'installation de Linux sur la carte Raspberry Pi 4 est expliquée dans le document « *Annexe : Installation de Raspberry OS sur Raspberry PI 4* » [5].

La suite de cette ressource suppose que, comme expliqué dans le guide d'installation cité ci-dessus, les bibliothèques sont installées et que l'on se connecte via VNC au nano-ordinateur de la voiture.

L'environnement de développement utilisé est l'interpréteur python Thonny, fourni avec Raspberry OS.

3 - Mise en œuvre des entrées/sorties

3.1 - Réception des données du Lidar

La réception des données du Lidar se fait par une transmission série puis, après une adaptation série vers USB, par liaison USB.

Attention : La version A2M8 (noir et rouge) du RPLidar utilise une communication à 115200 baud et la version A2M12 (noir et violet) une communication à 256000 baud. Il faut donc adapter ce paramètre de la fonction RPLidar() au lidar utilisé.

Ce premier code (raz_lidar.py) permet de vérifier la communication avec le lidar, grâce à la fonction lidar.get_info() et de remettre le lidar à 0, avec une déconnexion propre si la communication a été coupée brutalement.

```
from rplidar import RPLidar
import time


lidar = RPLidar("/dev/ttyUSB0",baudrate=115200)
lidar.disconnect()
time.sleep(1)
lidar.connect()
try :
print (lidar.get_info())
except :
print("la communication ne s'est pas établie correctement")
lidar.start_motor()
time.sleep(1)
lidar.stop_motor()
lidar.stop()
time.sleep(1)
lidar.disconnect()
```

Le code suivant (test_lidar_v2.py) permet de stocker les données captées par le Lidar dans un tableau python de 360 cases, l'indice indiquant l'angle du lidar (0 pour l'avant de la voiture puis dans le sens trigonométrique) et la valeur indiquant la distance en mm.

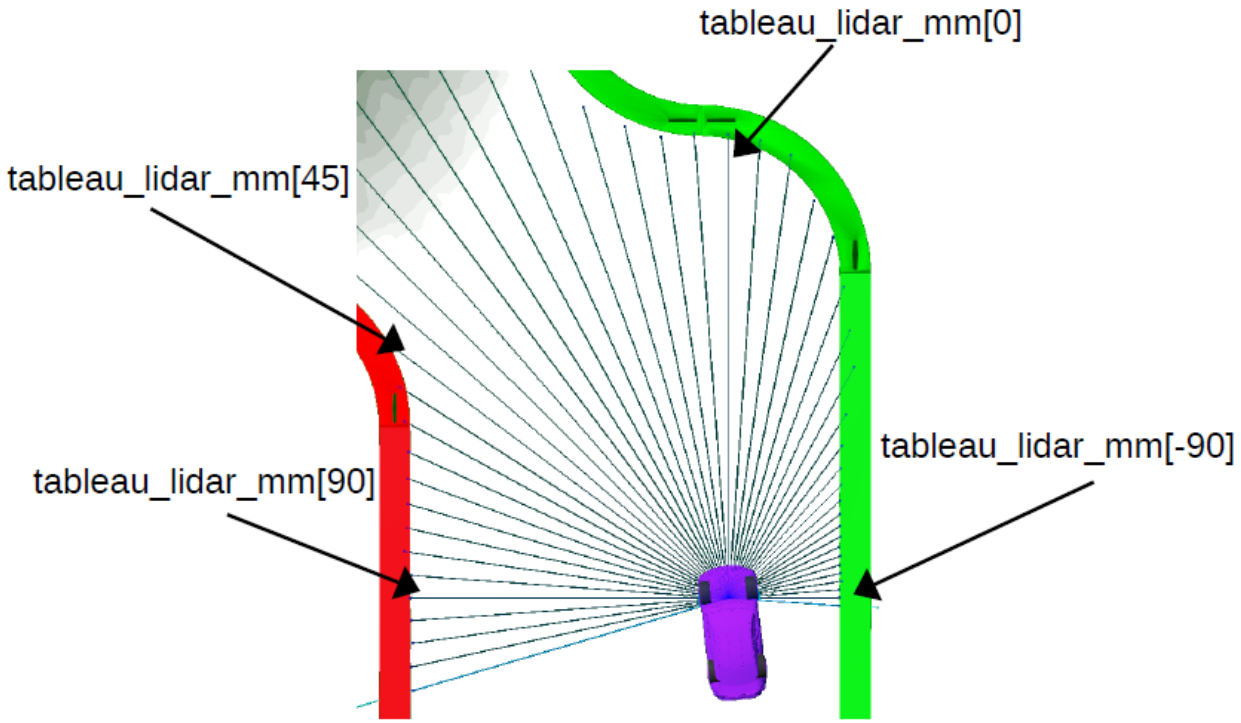


Figure 5 : Affichage sur le simulateur d'un rayon de mesure sur 4 du lidar

Le lidar capte tout autour de lui, mais les mesures vers l'arrière (entre 100 et 260° environ) sont perturbées par l'habitacle de la voiture

```
from rplidar import RPLidar
import numpy as np
import time
import matplotlib.pyplot as plt

#connexion et démarrage du lidar
lidar = RPLidar("/dev/ttyUSB0",baudrate=115200)
lidar.connect()
print (lidar.get_info())
lidar.start_motor()
```

```
time.sleep(1)

tableau_lidar_mm = [0]*360 #création d'un tableau de 360 zéros

try :
    for scan in lidar.iter_scans(scan_type='express') :
    #Le tableau se remplissant continuement, la boucle est infinie
        #affichage du nb de points récupérés lors du tour, pour les tests
        print("nb pts : " + str(len(scan)))
        #rangement des données dans le tableau
        for i in range(len(scan)) :
            angle = min(359,max(0,359-int(scan[i][1]))) #scan[i][1]:angle
            tableau_lidar_mm[angle]=scan[i][2] #scan[i][2]:distance

except KeyboardInterrupt: #récupération du CTRL+C
    print("fin des acquisitions")

#arrêt et déconnexion du lidar
lidar.stop_motor()
lidar.stop()
time.sleep(1)
lidar.disconnect()

#######################################################
#affichage des données acquises sur l'environnement
#pour les tests
#######################################################

teta = [0]*360 #création d'un tableau de 360 zéros

for i in range(360) :
    teta[i]=i*np.pi/180

fig = plt.figure()
ax = plt.subplot(111, projection='polar')
line = ax.scatter(teta, tableau_lidar_mm, s=5)
line.set_array(tableau_lidar_mm)
ax.set_rmax(8000)
ax.grid(True)
plt.show()
```

Une fois l'acquisition lancée, le programme affiche le nombre de points acquis par scan (moins que 360, ce qui signifie qu'il n'y a pas un point à chaque degré). CTRL+C permet de stopper l'acquisition et d'afficher le résultat sous forme d'un graphique Matplotlib.

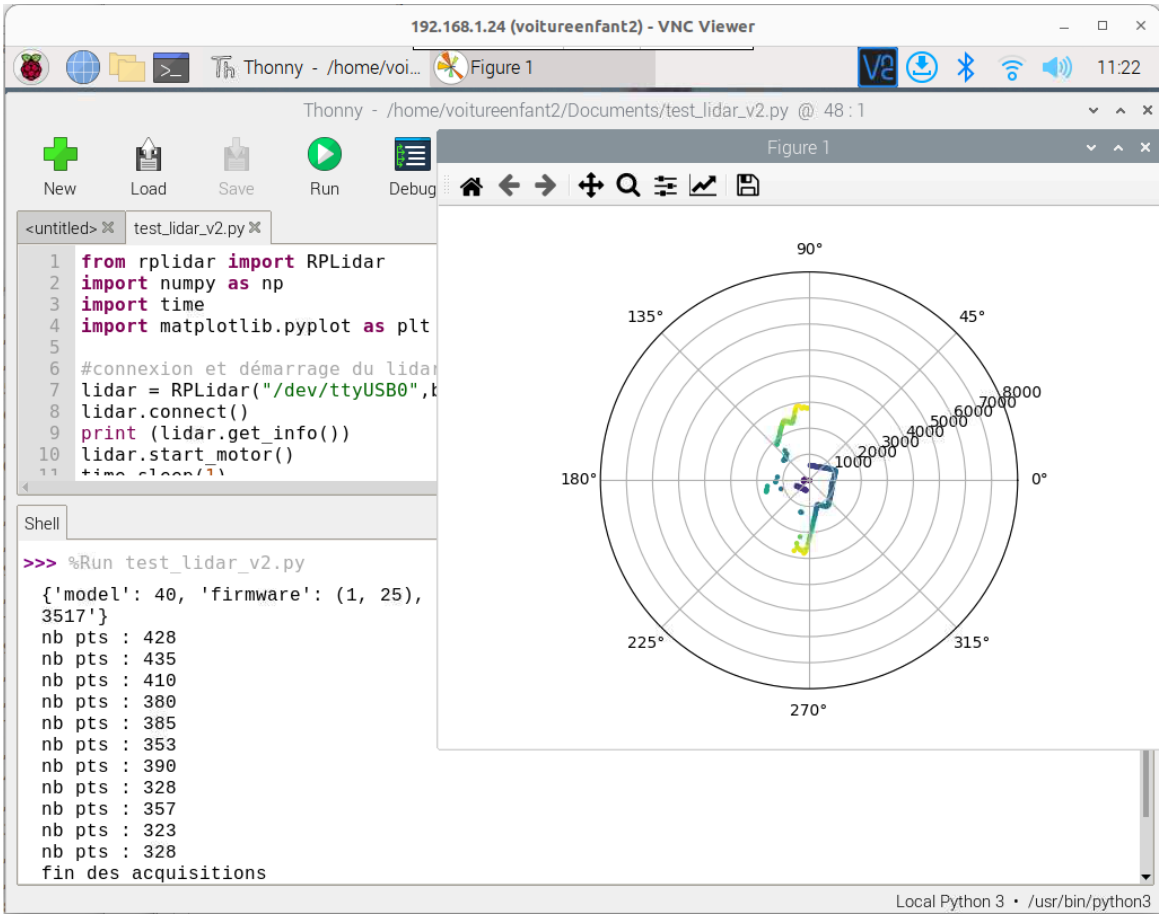
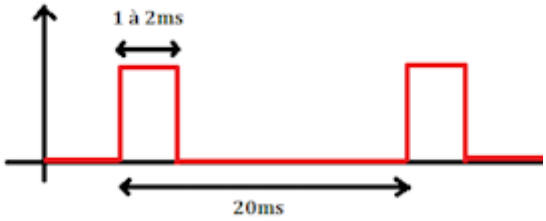


Figure 6 : Fenêtre VNC de la raspberry Pi après un arrêt de l'acquisition lidar et affichage des données

3.2 - Génération d'un signal de contrôle pour le servomoteur

Le servomoteur permet de contrôler la direction du véhicule. Le contrôle du servomoteur se fait par l'émission d'un signal PWM (pulse width modulation) à 50Hz avec un rapport cyclique variant entre 5% et 10%. Ce signal a la forme suivante :



- Un rapport cyclique de 5% sera équivalent à un état haut pendant 1ms, le servomoteur se positionnera alors tout à gauche (ou à droite suivant les modèles).
- Un rapport cyclique de 10% sera équivalent à un état haut pendant 2ms, le servomoteur se positionnera alors tout à droite (ou à gauche suivant les modèles).
- Un rapport cyclique de 7.5% sera équivalent à un état haut pendant 1.5ms, le servomoteur se positionnera alors au centre.

Attention : Le servomoteur étant relativement fragile, il est conseillé de rechercher progressivement les butées depuis la position centrale, d'autant plus que, suivant la marque de servomoteur installé, le sens de rotation est inversé...

Les sorties pwm de la Raspberry Pi n'ont parfois pas exactement la fréquence souhaitée, ce qui n'est pas gênant pour la période du signal mais demande d'ajuster la valeur centrale et les butées.

Voici donc un programme (test_pwm_direction.py) permettant de trouver les butées et le sens de rotation du servo-moteur pour compléter ensuite les paramètres de la fonction set_direction_degre().

```
from rpi_hardware_pwm import HardwarePWM
import time

#paramètres de départ, avec des butées très proche du centre
direction = 1 #1 pour angle_pwm_min a gauche, -1 pour angle_pwm_min à droite
angle_pwm_min = 6.7 #min
angle_pwm_max = 8.3 #max
angle_pwm_centre= 7.5

angle_degre_max = +18 #vers la gauche
angle_degre=0

pwm_dir = HardwarePWM(pwm_channel=1,hz=50)
pwm_dir.start(angle_pwm_centre)

def set_direction_degre(angle_degre) :
    global angle_pwm_min
    global angle_pwm_max
    global angle_pwm_centre
    angle_pwm = angle_pwm_centre + direction * (angle_pwm_max - angle_pwm_min) * angle_degre /(2 *
angle_degre_max )
    if angle_pwm > angle_pwm_max :
        angle_pwm = angle_pwm_max
    if angle_pwm < angle_pwm_min :
        angle_pwm = angle_pwm_min
    pwm_dir.change_duty_cycle(angle_pwm)

print("réglage des butées, Q pour quitter")
print("valeur numérique pour tester un angle de direction")
print("I pour inverser droite et gauche")
print("g pour diminuer la butée gauche et G pour l'augmenter")
print("d pour diminuer la butée droite et D pour l'augmenter")

while True :
    a = input("g, G, d, D ?")
    try :
        angle_degre=int(a)
        set_direction_degre(angle_degre)
    except :
        if a == "I" :
            direction = -direction
            print("nouvelle direction : " + str(direction))
        elif a == "g" :
            if direction == 1 :
                angle_pwm_max -=0.1
                print("nouvelle butée gauche : " + str(angle_pwm_max))
            else :
                angle_pwm_min +=0.1
                print("nouvelle butée gauche : " + str(angle_pwm_min))
                angle_pwm_centre = (angle_pwm_max+angle_pwm_min)/2
                set_direction_degre(18)
        elif a == "G" :
            if direction == 1 :
                angle_pwm_max +=0.1
                print("nouvelle butée gauche : " + str(angle_pwm_max))
            else :
                angle_pwm_min -=0.1
                print("nouvelle butée gauche : " + str(angle_pwm_min))
                angle_pwm_centre = (angle_pwm_max+angle_pwm_min)/2
```

```
                set_direction_degre(18)
            elif a == "d" :
                if direction == -1 :
                    angle_pwm_max -=0.1
                    print("nouvelle butée droite : " + str(angle_pwm_max))
                else :
                    angle_pwm_min +=0.1
                    print("nouvelle butée droite : " + str(angle_pwm_min))
                angle_pwm_centre = (angle_pwm_max+angle_pwm_min)/2
                set_direction_degre(-18)
            elif a == "D" :
                if direction == -1 :
                    angle_pwm_max +=.1
                    print("nouvelle butée droite : " + str(angle_pwm_max))
                else :
                    angle_pwm_min -=0.1
                    print("nouvelle butée droite : " + str(angle_pwm_min))
                angle_pwm_centre = (angle_pwm_max+angle_pwm_min)/2
                set_direction_degre(-18)
            else :
                break

print("nouvelles valeurs")
print("direction : "        + str(direction))
print("angle_pwm_min : "    + str(angle_pwm_min))
print("angle_pwm_max : "    + str(angle_pwm_max))
print("angle_pwm_centre : " + str(angle_pwm_centre))
```

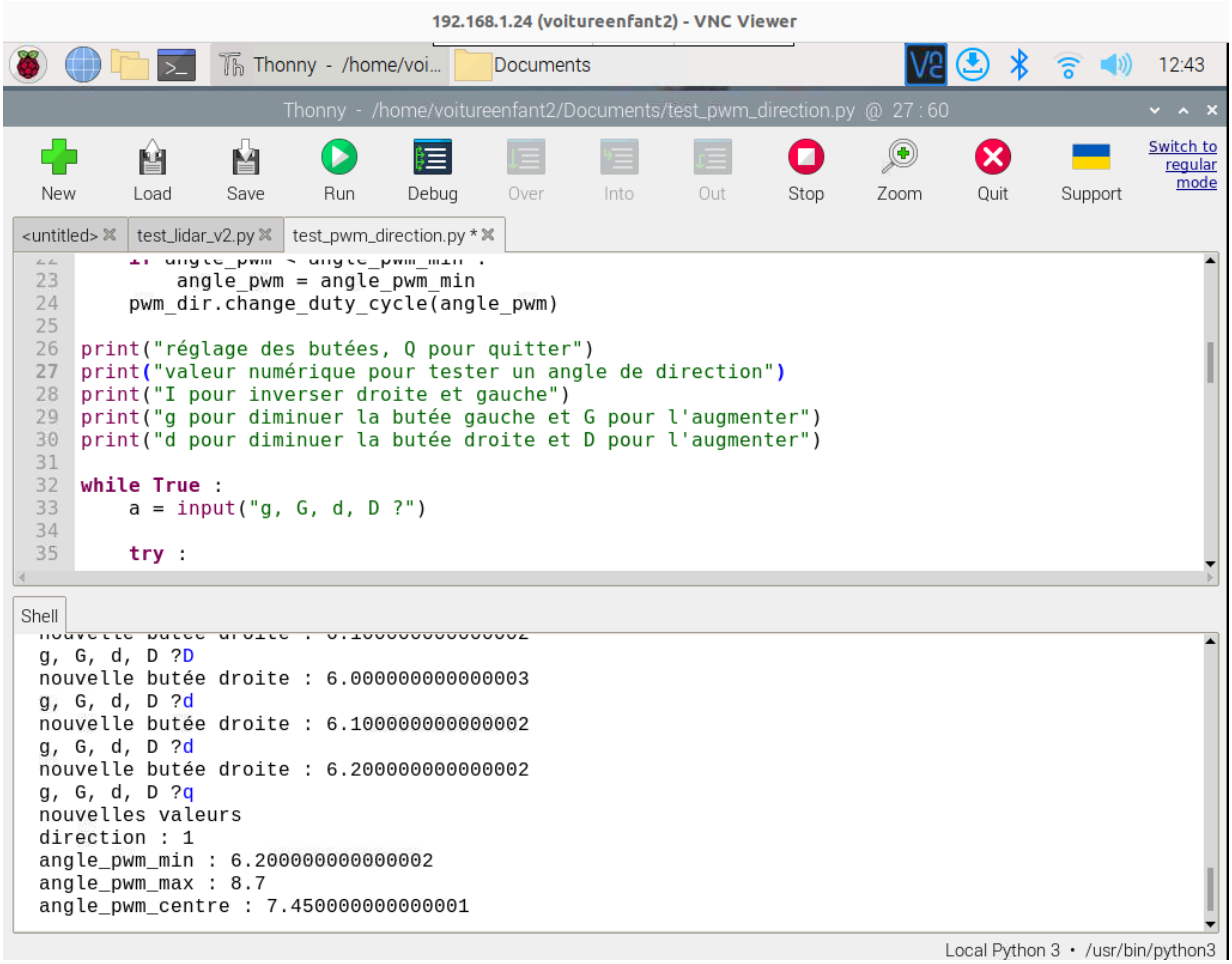


Figure 7 : Fenêtre VNC de la raspberry pi après un étalonnage des butées de la fonction set_direction

3.3 - Génération d'un signal de contrôle pour le variateur moteur

Le variateur de vitesse permet de contrôler la vitesse et la direction du moteur. Il se contrôle avec le même type de signal que le servomoteur de direction :

- Un rapport cyclique de 10% est équivalent à un état haut pendant 2 ms, le moteur sera à pleine vitesse en marche avant. Pour certains variateur, c'est un rapport cyclique à 5 % qui correspond à la pleine vitesse.
- Un rapport cyclique de 7.5% est équivalent à un état haut pendant 1.5 ms, le moteur est à l'arrêt.

Dans les faits, il y a un point mort autour de 7,5 % et la vitesse maximale est atteinte bien avant les 10 %. Il est donc important de trouver les butées pour avoir une commande de vitesse proche de la vitesse réelle, même si l'absence de mesure de la vitesse réelle empêche une bonne précision.

La marche arrière de la plupart des variateurs de voitures radiocommandés se fait en 2 temps :

- d'abord un freinage, rapport cyclique minimum (1 ms → 5%) pendant 0,3 seconde,
- un passage par le point mort (1,5 ms → 7,5%) pendant 0,3 seconde,
- la marche arrière à vitesse réglable (entre 1,2 et 1,4 ms → 6 à 7%)

Le code suivant (test_pwm_propulsion.py) permet de tester les fonctions set_vitesse_m_s() et recule() de la voiture et d'étalonner les butées, à l'image du code similaire test_pwm_direction.py

```
from rpi_hardware_pwm import HardwarePWM
import time

#paramètres de la fonction vitesse_m_s, à étalonner
direction_prop = -1 # -1 pour les variateurs inversés
pwm_stop_prop = 8.17
point_mort_prop = 0.13
delta_pwm_max_prop = 1.5 #pwm à laquelle on atteint la vitesse maximale

vitesse_max_m_s_hard = 8 #vitesse que peut atteindre la voiture
vitesse_max_m_s_soft = 2 #vitesse maximale que l'on souhaite atteindre

pwm_prop = HardwarePWM(pwm_channel=0, hz=50)
pwm_prop.start(pwm_stop_prop)

def set_vitesse_m_s(vitesse_m_s):
    if vitesse_m_s > vitesse_max_m_s_soft :
        vitesse_m_s = vitesse_max_m_s_soft
    elif vitesse_m_s < -vitesse_max_m_s_hard :
        vitesse_m_s = -vitesse_max_m_s_hard
    if vitesse_m_s == 0 :
        pwm_prop.change_duty_cycle(pwm_stop_prop)
    elif vitesse_m_s > 0 :
        vitesse = vitesse_m_s * (delta_pwm_max_prop)/vitesse_max_m_s_hard
        pwm_prop.change_duty_cycle(pwm_stop_prop + direction_prop*(point_mort_prop + vitesse ))
    elif vitesse_m_s < 0 :
        vitesse = vitesse_m_s * (delta_pwm_max_prop)/vitesse_max_m_s_hard
        pwm_prop.change_duty_cycle(pwm_stop_prop - direction_prop*(point_mort_prop - vitesse ))

def recule():
    set_vitesse_m_s(-vitesse_max_m_s_hard)
    time.sleep(0.2)
    set_vitesse_m_s(0)
    time.sleep(0.2)
    set_vitesse_m_s(-1)

print("réglage des butées, Q pour quitter")
print("valeur numérique pour tester une vitesse en mm/s")
print("R pour reculer")
print("I pour inverser droite et gauche")
print("p pour diminuer delta_pwm_max_prop et P pour l'augmenter")
print("z pour diminuer le point zéro 1,5 ms et Z pour l'augmenter")
```

```
print("m pour diminuer le point mort et M pour l'augmenter")

while True :
    a = input("vitesse en mm/s, R, I, p, P, z, Z, m, M")
    try :
        vitesse_mm_s=int(a)
        set_vitesse_m_s(vitesse_mm_s/1000.0)
    except :
        if a == "I" or a == "i" :
            direction_prop = -direction_prop
            print("nouvelle direction : " + str(direction_prop))
        elif a == "R" :
            recule()
            print("recule")
        elif a == "p" :
            delta_pwm_max_prop -=0.1
            print("nouveau delta_pwm_max_prop : " + str(delta_pwm_max_prop))
            pwm_prop.change_duty_cycle(pwm_stop_prop+direction_prop*(point_mort_prop+\
            delta_pwm_max_prop))
        elif a == "P" :
            delta_pwm_max_prop +=0.1
            print("nouveau delta_pwm_max_prop : " + str(delta_pwm_max_prop))
            pwm_prop.change_duty_cycle(pwm_stop_prop + direction_prop*(point_mort_prop+\
            delta_pwm_max_prop))
        elif a == "z" :
            pwm_stop_prop -=0.01
            print("nouveau pwm_stop_prop : " + str(pwm_stop_prop))
            pwm_prop.change_duty_cycle(pwm_stop_prop)
        elif a == "Z" :
            pwm_stop_prop +=0.01
            print("nouveau pwm_stop_prop : " + str(pwm_stop_prop))
            pwm_prop.change_duty_cycle(pwm_stop_prop)
        elif a == "m" :
            point_mort_prop -=0.01
            print("nouveau point_mort_prop : " + str(point_mort_prop))
            pwm_prop.change_duty_cycle(pwm_stop_prop + direction_prop*(point_mort_prop))
        elif a == "M" :
            point_mort_prop +=0.01
            print("nouveau point_mort_prop : " + str(point_mort_prop))
            pwm_prop.change_duty_cycle(pwm_stop_prop + direction_prop*(point_mort_prop))
        else :
            break

pwm_prop.change_duty_cycle(pwm_stop_prop)
print("nouvelles valeurs")
print("direction : "      +     str(direction_prop))
print("delta_pwm_max_prop : " + str(delta_pwm_max_prop))
print("point zero 1,5 ms : "  + str(pwm_stop_prop))
print("point mort : "         + str(point_mort_prop))
```

4 - Codage d'un suivi de ligne simple

4.1 - Code de base

Comme dans le simulateur, les données du lidar sont acquises dans un tableau **tableau_lidar_mm[360]** et la commande de la direction et de la propulsion se fait via les fonctions **set_direction_degre()** et **set_vitesse_m_s()**.

On peut importer la zone de code correspondant à l'algorithme de conduite du simulateur.

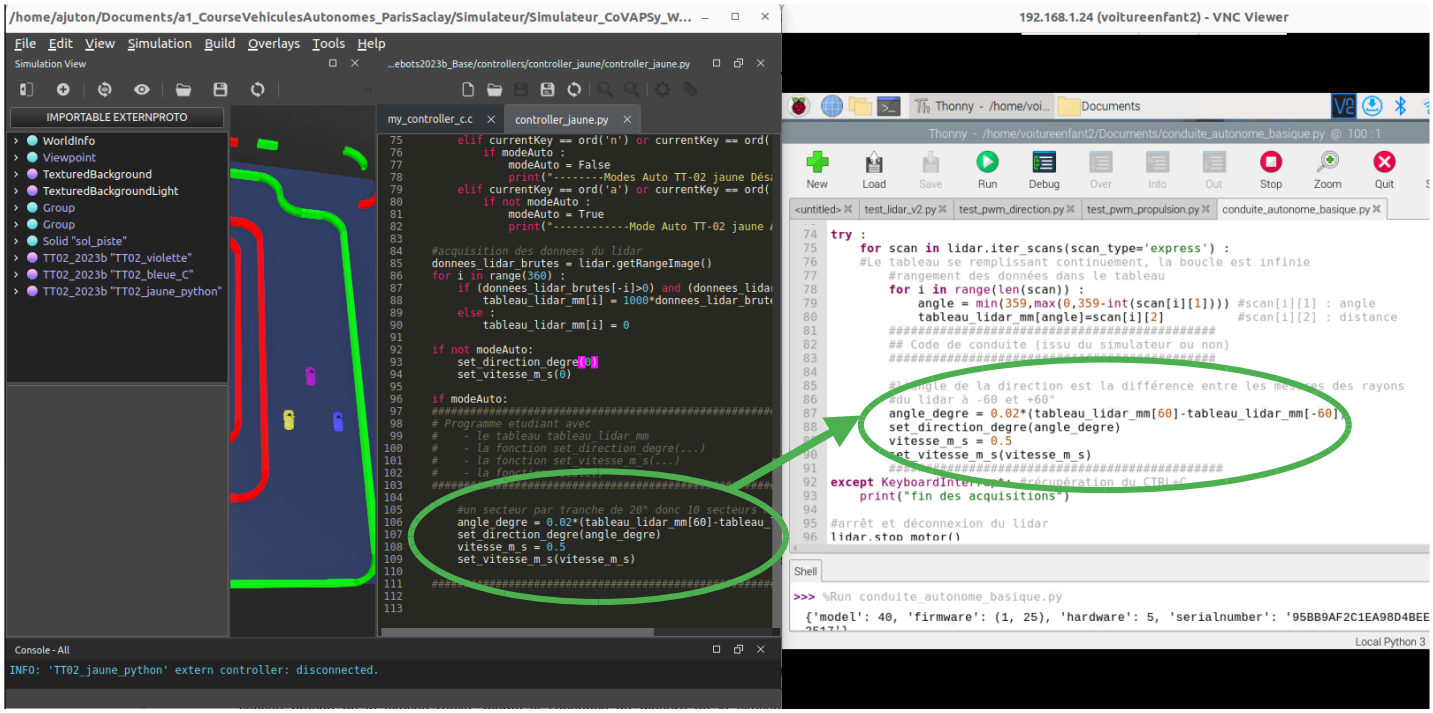


Figure 8 : Copie du code du simulateur vers le nano-ordinateur Raspberry Pi de la voiture réelle

Le code de base fourni (conduite_autonome_basique.py) est le suivant :

```
from rplidar import RPLidar
import time
from rpi_hardware_pwm import HardwarePWM

#paramètres de la fonction vitesse_m_s, issus des essais précédents
direction_prop = -1 # -1 pour les variateurs inversés
pwm_stop_prop = 8.17
point_mort_prop = 0.13
delta_pwm_max_prop = 1.5 #pwm à laquelle on atteint la vitesse maximale

vitesse_max_m_s_hard = 8 #vitesse que peut atteindre la voiture
vitesse_max_m_s_soft = 2 #vitesse maximale que l'on souhaite atteindre

#paramètres de la fonction set_direction_degre, issus des essais précédents
direction = 1 #1 pour angle_pwm_min a gauche, -1 pour angle_pwm_min à droite
angle_pwm_min = 6 #min
angle_pwm_max = 9 #max
angle_pwm_centre= 7.5

angle_degre_max = +18 #vers la gauche
angle_degre=0

pwm_prop = HardwarePWM(pwm_channel=0, hz=50)
pwm_prop.start(pwm_stop_prop)

def set_vitesse_m_s(vitesse_m_s):
    # cf définition de la fonction précédemment

def recule():
    # cf définition de la fonction précédemment

pwm_dir = HardwarePWM(pwm_channel=1,hz=50)
pwm_dir.start(angle_pwm_centre)

def set_direction_degre(angle_degre) :
    # cf définition de la fonction précédemment

#connexion et démarrage du lidar
lidar = RPLidar("/dev/ttyUSB0",baudrate=115200)
lidar.connect()
print (lidar.get_info())
lidar.start_motor()
```

```
time.sleep(1)

tableau_lidar_mm = [0]*360 #création d'un tableau de 360 zéros

try :
    for scan in lidar.iter_scans(scan_type='express') :
    #Le tableau se remplissant continuement, la boucle est infinie
        #rangement des données dans le tableau
        for i in range(len(scan)) :
            angle = min(359,max(0,359-int(scan[i][1]))) #scan[i][1] : angle
            tableau_lidar_mm[angle]=scan[i][2] #scan[i][2] : distance
        ################################################
        ## Code de conduite (issu du simulateur ou non)
        ################################################

        #l'angle de la direction est la différence entre les mesures
        #des rayons du lidar à -60 et +60°
        angle_degre = 0.02*(tableau_lidar_mm[60]-tableau_lidar_mm[-60])
        set_direction_degre(angle_degre)
        vitesse_m_s = 0.5
        set_vitesse_m_s(vitesse_m_s)
        ################################################
except KeyboardInterrupt: #récupération du CTRL+C
    print("fin des acquisitions")

#arrêt et déconnexion du lidar et des moteurs
lidar.stop_motor()
lidar.stop()
time.sleep(1)
lidar.disconnect()
pwm_dir.stop()
pwm_prop.start(pwm_stop_prop)
```

4.2 - Améliorations possibles

On donne ici les mêmes pistes que pour l'amélioration de l'algorithme du simulateur :

- Il est possible de regrouper les rayons en secteur de 10° pour chercher le secteur dont le rayon le plus court est le plus long parmi les plus courts des autres secteurs.
- Il est possible d'adapter sa vitesse à la distance de l'obstacle devant.
- Il est intéressant de détecter un obstacle pour réussir à l'éviter. On peut pour cela ajouter des morceaux de bordure de piste au milieu de la piste.
- A la différence de celui du simulateur, le lidar réel fait moins de 250 mesures par tour, donc il ne met pas à jour chaque valeur du tableau. Des valeurs moins « fraîches » restent dans le tableau. Il pourrait être intéressant de vider le tableau avant sa mise à jour et de compléter les valeurs manquantes par une estimation.
- Il est possible de reculer quand on est dans un mur, en surveillant une valeur minimale des rayons du lidar à l'avant et sur les côtés. Pour déterminer les situations de quasi-collision, on peut se baser sur 3 valeurs du Lidar : la mesure à 0°, celle à -30° et celle à 30°. Si les distances captées par le Lidar sur ces angles spécifiques sont inférieures à un certain seuil, on considère que la voiture s'est crashée. On a alors 3 possibilités, qui demande un peu de travail car il n'est pas possible d'utiliser time.sleep, cette fonction bloquante mettant aussi en pause le moteur physique (l'utilisation de time.time() peut alors être intéressante) :
    - Seuil franchi pour l'angle 0° (mur devant) : On replace les roues pour aller droit puis on recule jusqu'à ce que la valeur de lidar franchisse un seuil ou pendant 0,5s.
    - Seuil franchi pour l'angle 30° (mur à gauche) : On tourne complètement à gauche puis on recule jusqu'à ce que la valeur de lidar franchisse un seuil ou pendant 0,5s.
    - Seuil franchi pour l'angle -30° (mur à droite) : On tourne complètement à droite puis on recule jusqu'à ce que la valeur de lidar franchisse un seuil ou pendant 0,5s.

- Des méthodes avancées sont bien évidemment possible, en sortant du cadre du lycée, avec des trajectoires en forme de tentacules (https://doi.org/10.1002/rob.20256), ou avec de l'apprentissage par renforcement (voir « Apprentissage par renforcement et transfert simulation vers réalité pour la conduite de voitures autonomes » [3]).

L'utilisation de la fonction de recul ou d'une temporisation un peu longue n'est pas possible dans la boucle de scan, en effet, pendant la temporisation, le buffer du lidar se remplit. Il faut donc, comme sur le simulateur utiliser une machine à état avec la lecture du temps (time.time()) non bloquante.

Une autre solution consiste à utiliser les threads de python, comme dans le code suivant, avec un thread pour l'acquisition lidar et un thread pour la conduite autonome.

```
from rplidar import RPLidar
import numpy as np
import time
from rpi_hardware_pwm import HardwarePWM
import threading

#paramètres de la fonction vitesse_m_s, issus des essais précédents
direction_prop = -1 # -1 pour les variateurs inversés
pwm_stop_prop = 8.17
point_mort_prop = 0.13
delta_pwm_max_prop = 1.5 #pwm à laquelle on atteint la vitesse maximale

vitesse_max_m_s_hard = 8 #vitesse que peut atteindre la voiture
vitesse_max_m_s_soft = 2 #vitesse maximale que l'on souhaite atteindre


#paramètres de la fonction set_direction_degre, issus des essais précédents
direction = 1 #1 pour angle_pwm_min a gauche, -1 pour angle_pwm_min à droite
angle_pwm_min = 6 #min
angle_pwm_max = 9 #max
angle_pwm_centre= 7.5

angle_degre_max = +18 #vers la gauche
angle_degre=0

pwm_prop = HardwarePWM(pwm_channel=0, hz=50)
pwm_prop.start(pwm_stop_prop)

def set_vitesse_m_s(vitesse_m_s):
    # cf définition de la fonction précédemment

def recule():
    # cf définition de la fonction précédemment

pwm_dir = HardwarePWM(pwm_channel=1,hz=50)
pwm_dir.start(angle_pwm_centre)

def set_direction_degre(angle_degre) :
    # cf définition de la fonction précédemment

acqui_tableau_lidar_mm = [0]*360 #création d'un tableau de 360 zéros
tableau_lidar_mm = [0]*360
drapeau_nouveau_scan = False
Run_Lidar = False

def lidar_scan() :
    global drapeau_nouveau_scan
    global acqui_tableau_lidar_mm
    global Run_Lidar
    global lidar
    print ("tâche lidar_scan démarrée")
    while Run_Lidar == True :
        try :
            for scan in lidar.iter_scans(scan_type='express') :
            #Le tableau se remplissant continuement, la boucle est infinie
                #rangement des données dans le tableau
                for i in range(len(scan)) :
                    angle = min(359,max(0,359-int(scan[i][1]))) #scan[i][1]:angle
                    acqui_tableau_lidar_mm[angle]=scan[i][2] #scan[i][2]:distance
                    drapeau_nouveau_scan = True
                    time.sleep(0.01)
                    if(Run_Lidar == False) :
                        break
```

```
        except :
            print("souci acquisition lidar")

def conduite_autonome():
    global drapeau_nouveau_scan
    global acqui_tableau_lidar_mm
    global tableau_lidar_mm
    global Run_Lidar
    print ("tâche conduite autonome démarrée")
    while Run_Lidar == True :
        if(drapeau_nouveau_scan == False) :
            time.sleep(0.01)
        else :
            #récupération du tableau_lidar acquis par l'autre thread
            for i in range(-100,101) :
                tableau_lidar_mm[i] = acqui_tableau_lidar_mm[i]
            drapeau_nouveau_scan = False

            ######################################################
            # programme de conduite avec détection des murs et marche arrière
            ######################################################
            if tableau_lidar_mm[0]>0 and tableau_lidar_mm[0]<150:
                print("mur devant")
                set_direction_degre(0)
                recule()
                time.sleep(0.5)
            elif tableau_lidar_mm[-30]>0 and tableau_lidar_mm[-30]<150 :
                print("mur à droite")
                set_direction_degre(-18)
                recule()
                time.sleep(0.5)
            elif tableau_lidar_mm[30]>0 and tableau_lidar_mm[30]<150 :
                print("mur à gauche")
                set_direction_degre(+18)
                recule()
                time.sleep(0.5)
            else :
                #l'angle de la direction est la différence entre les mesures des rayons
                #du lidar à -60 et +60°
                angle_degre = 0.02*(tableau_lidar_mm[60]-tableau_lidar_mm[-60])
                set_direction_degre(angle_degre)
                vitesse_m_s = 0.5
                set_vitesse_m_s(vitesse_m_s)
            ######################################################

#connexion et démarrage du lidar
lidar = RPLidar("/dev/ttyUSB0",baudrate=115200)
lidar.connect()
print (lidar.get_info())
lidar.start_motor()
time.sleep(2)

#démarrage des 2 thread
Run_Lidar = True
thread_scan_lidar = threading.Thread(target= lidar_scan)
thread_scan_lidar.start()
time.sleep(1)
thread_conduite_autonome = threading.Thread(target = conduite_autonome)
thread_conduite_autonome.start()

while True :
    try :
        pass
    except KeyboardInterrupt: #récupération du CTRL+C
        print("arrêt du programme")
        Run_Lidar = False
        break

#attente de l'arrêt des tâches
thread_conduite_autonome.join()
thread_scan_lidar.join()

#arrêt et déconnexion du lidar
lidar.stop_motor()
lidar.stop()
time.sleep(1)
lidar.disconnect()
pwm_prop.stop()
pwm_dir.stop()
```

5 - Ouvertures

La programmation de la voiture 1/10ème avec seulement un nano-ordinateur Raspberry Pi et un lidar permet de manipuler un capteur avancé et la raspberry pi en programmation python, avec des tableaux et des fonctions. En programmation, avec du multi-tâche, du travail simulation / réalité, de l'apprentissage… le champ des possibles est immense.

Côté systèmes embarqués, profitant de la connectivité i2c de la raspberry pi, Il est possible d'ajouter des capteurs : ultrason (SRF10) à l'arrière, une centrale intertielle (BNO055) pour mesurer l'orientation ou les chocs, un afficheur OLED (TF051). Il est possible d'utiliser le bluetooth pour ajouter une manette de playstation, pour faire une course entre voiture autonome et voiture commandée, etc.

Enfin, côté asservissement, il est possible de mettre en place un correcteur PID ou des correcteurs plus avancés pour le contrôle de la voiture. En monovariable, la direction est calculée à partir de la différence entre les distances mesurées par le lidar à +60 et -60°.

Des exemples sont fournis sur le dépôt git [4] de la course et les nouvelles contributions sont les bienvenues.

Références :

[1]: Course Voitures Autonomes Paris Saclay (CoVAPSy) : Travaux pratiques autour des voitures autonomes, T. Boulanger, E. Délègue, K. Hoarau, A. Juton, https://eduscol.education.fr/sti/si-ens-paris-saclay/ressources_pedagogiques/covapsy-tp-autour-des-voitures-autonomes

[2]: CoVAPSy : Mise en œuvre du Simulateur Webots, T. Boulanger, E. Délègue, K. Hoarau, A. Juton, https://eduscol.education.fr/sti/si-ens-paris-saclay/ressources_pedagogiques/covapsy-mise-en-oeuvre-du-simulateur-webots

[3]: Apprentissage par renforcement et transfert simulation vers réalité pour la conduite de voitures autonomes , R. Bennani, K. Hoarau, A. Juton, https://eduscol.education.fr/sti/si-ens-paris-saclay/ressources_pedagogiques/apprentissage-renforcement-transfert-simulation-vers-realite-pourla-conduite-voitures-autonomes

[4]: Dépôt git de la course de voitures autonomes de Paris Saclay : https://github.com/ajuton-ens/CourseVoituresAutonomesSaclay

[5]: Annexe : Installation de Raspberry OS sur Raspberry Pi 4, A. Juton, https://eduscol.education.fr/sti/si-ens-paris-saclay/ressources_pedagogiques/covapsy-premiers-programmes-python-sur-voiture-reelle

Ressource publiée sur Culture Sciences de l'Ingénieur : https://eduscol.education.fr/sti/si-ens-paris-saclay

CoVAPSy : Mise en œuvre du simulateur Webots

La Revue 3E.I
Culture Sciences de l'Ingénieur
école normale supérieure paris—saclay

Thomas BOULANGER[1] – Eve DÉLÈGUE[2]– Kévin HOARAU[3]
Anthony JUTON[4]

Édité le
09/11/2023

[1] *Élève en année de recherche pré-doctorale à l'étranger, ENS Paris-Saclay – DER Nikola Tesla,*
[2] *Élève en année de recherche en intelligence artificielle, ENS Paris-Saclay – DER Nikola Tesla,*
[3] *Élève en M2 Formation à l'Enseignement Supérieur, ENS Paris-Saclay – DER Nikola Tesla,*
[4] *Professeur agrégé de physique appliquée au DER Nikola Tesla, ENS Paris-Saclay*

Cette ressource fait partie du N°111 de La Revue 3EI de janvier 2024.

Cette ressource fait suite à la ressource « *Course Voitures Autonomes Paris Saclay (CoVAPSy) : Travaux pratiques autour des voitures autonomes* » [1] pour présenter en détail le simulateur et aider à sa prise en main, afin d'organiser une course de voitures simulées avec des élèves ou étudiants.

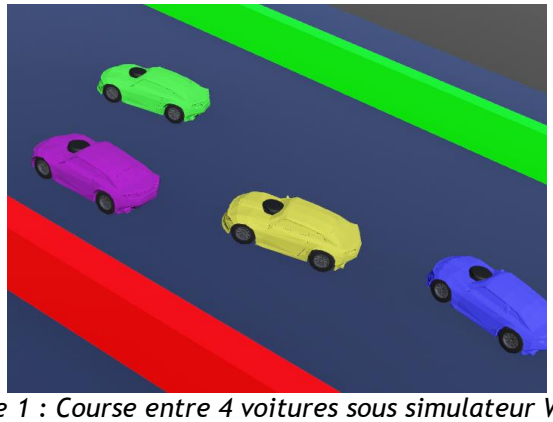


Figure 1 : Course entre 4 voitures sous simulateur Webots

Pour travailler sur les algorithmes en robotique, en s'affranchissant des problèmes matériels, il est très intéressant d'utiliser un simulateur, avant de travailler sur le robot réel. C'est d'autant plus vrai avec l'apprentissage automatique qui demande des milliers d'essais auxquels le robot physique ne résisterait pas. Dans ce cadre, pour la course de voitures autonomes de Paris Saclay, plusieurs équipes ont choisi le simulateur Webots et y ont développé un modèle de la voiture proche de la voiture 1/10ème utilisée pour la course, notamment pour faire de l'apprentissage par renforcement (voir « *Apprentissage par renforcement et transfert simulation vers réalité pour la conduite de voitures autonomes* » [2]). Webots est un simulateur open-source de robotique populaire dans la recherche et l'enseignement. Il utilise la bibliothèque ODE (Open Dynamics Engine) pour détecter des collisions et simuler la dynamique des corps rigides et des fluides. Il est bien documenté, multiplateforme (Linux, Windows, MacOS), utilisable sur un PC non doté d'un processeur graphique performant et permet la programmation en C, en java ou en python directement depuis le logiciel (le plus simple) ou à partir d'un environnement de développement tiers (plus efficace pour le débogage).

L'objectif de cette ressource est de guider le lecteur vers une course de voitures 1/10ème simulées. La programmation peut se faire en python ou en C. La ressource se limite à un algorithme très simple, les étudiants ayant en charge de travailler sur des algorithmes plus performants.

Le code issu du simulateur peut ensuite être réutilisé dans la voiture 1/10ème réelle (voir « *CoVAPSy : Premiers programmes python sur la voiture réelle* » [3]).

Prérequis : Les bases de la programmation Python (connaissance des classes non nécessaire) ou en C suivant le langage retenu : manipulation des tableaux, utilisation de fonctions.

1 - Installation de Webots

La version utilisée pour cet article est la R2023b. Il est recommandé de travailler sur cette version, le passage d'une version à l'autre de Webots demandant une bonne connaissance de l'outil.

La dernière version de Webots se télécharge à l'adresse : https://www.cyberbotics.com/

Les versions antérieures se téléchargent à l'adresse : https://github.com/cyberbotics/webots/releases

Sous Linux, mieux vaut ne pas utiliser l'installation par *snap*, celle-ci ayant un lien plus complexe avec les autres logiciels (notamment l'IDE python ou C) du PC. Une archive pour une installation classique est également proposée par Webots.

En plus du logiciel, pour travailler sur les voitures autonomes 1/10ème, il faut télécharger le projet de base Simulateur_CoVAPSy_Webots2023b_Base.zip à l'adresse suivante : https://github.com/ajuton-ens/CourseVoituresAutonomesSaclay/blob/main/Simulateur/Simulateur_CoVAPSy_Webots2023b_Base.zip

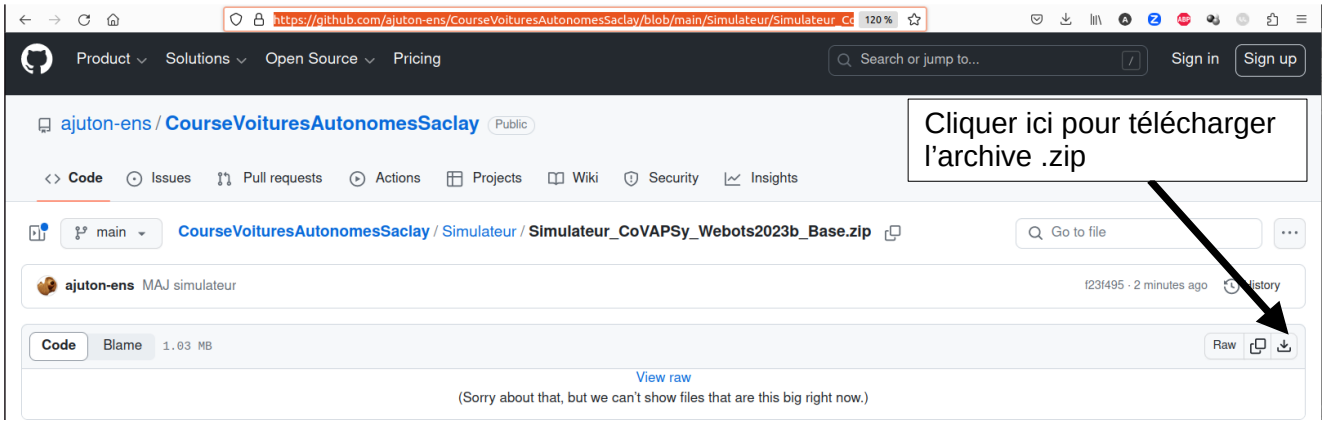
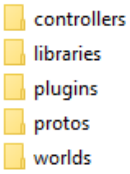


Figure 2 : Téléchargement de l'archive contenant le projet webots de base

2 - Présentation de l'environnement

2.1 - Éléments d'un projet

Dans le dossier *Simulateur_CoVAPSy_Webots2023b_Base*, une fois décompressé, on trouve les éléments d'un projet webots.



Tous les projets Webots sont composés des dossiers montrés ci-dessus. Voici un détail de leur utilité :

- **controllers** : Dossier contenant les programmes qui contrôlent les robots présents dans le projet
- **libraries** : non utilisé ici
- **plugins** : non utilisé ici

- **protos** : Dossier contenant les fichiers des modèles des robots qui ne sont pas présents dans la base de données de Webots (dont la voiture TT-02)
- **worlds** : Dossier contenant tous les mondes créés pour le projet (la piste ***Piste_CoVAPSy_2023.wbt*** notamment)

2.2 - Interface Webots

Une fois lancé, le logiciel se présente sous la forme de quatre panneaux :

- **L'arborescence des éléments** permet de configurer le monde (la piste ici) et les voitures. On peut y modifier les paramètres de la piste, y ajouter des voitures et y modifier leur couleur ou les paramètres du lidar.
- **L'environnement graphique** permet de visualiser l'évolution des voitures sur la piste et de modifier la position des voitures ou des éléments de la piste à la souris.
- **L'éditeur de texte** permet de modifier le code des contrôleurs des voitures, mais aussi des définitions des voitures (les PROTOS).
- **La console** affiche les avertissements et erreurs du logiciel (dont ceux liés à l'interprétation du code python ou à la compilation du code C) et les affichages (print en python, printf en C) lors de l'exécution du programme.

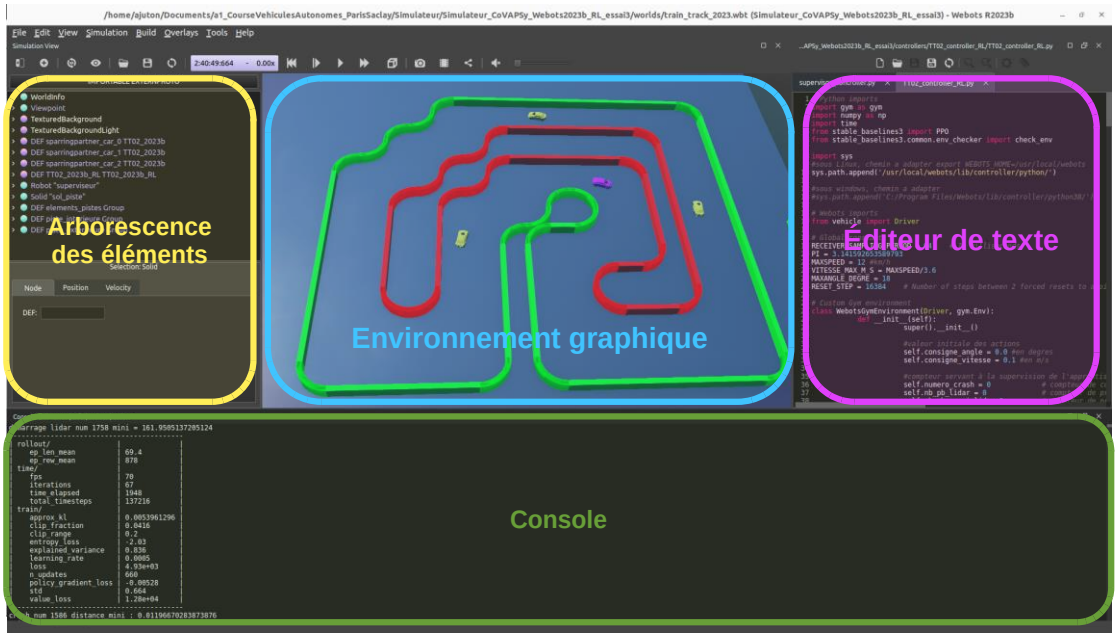


Figure 3 : Les différents panneaux de Webots

- Aller dans l'onglet File → Open World
- Choisir le fichier Piste_CoVAPSy_2023b.wbt qui se trouve dans le dossier : /<chemin_vers_le_dossier>/Simulateur_CoVAPSy_Webots2023b_Base/worlds/

2.3 - Modification de la piste

Le World ouvert précédemment contient une piste déjà construite. Mais il est possible de modifier le tracé avec les blocs individuels présents en haut de la piste.

Dans l'arborescence, ces blocs se trouvent dans « ***DEF Elements_individuels Group*** » dans le dossier ***children***. Il est possible de copier ces blocs et de les rajouter à la piste pour redéfinir le tracé.

Pour cela, il suffit de sélectionner le bloc voulu et de faire ***Clic droit*** →***Copy*** (ou ***Ctrl+C***). Il faut maintenant sélectionner dans l'arborescence « ***DEF Piste Group*** » dossier ***children*** pour coller le bloc avec ***Clic droit*** + ***Paste*** (ou ***Ctrl+V***).

Il y a trois manières de déplacer un bloc :

- Sélectionner le bloc. Maintenir la touche ***Shift***. Bouger la souris avec le ***clic gauche*** enfoncé pour translater le bloc et ***clic droit*** enfoncé pour la rotation.
- Utiliser les flèches verte, rouge et bleu pour déplacer et faire tourner le bloc. Une métrique est disponible pour aider en précision.
- Dans l'arborescence, dérouler le ***Solid*** correspondant. Modifier les champs ***Translation*** et ***Rotation*** pour modifier le positionnement du bloc. C'est la méthode la plus précise.

Figure 4 : Modification de la piste

3 - Programme des voitures

Cette partie aborde le cœur du travail : la programmation des voitures dans le simulateur, d'abord en python, puis en C.

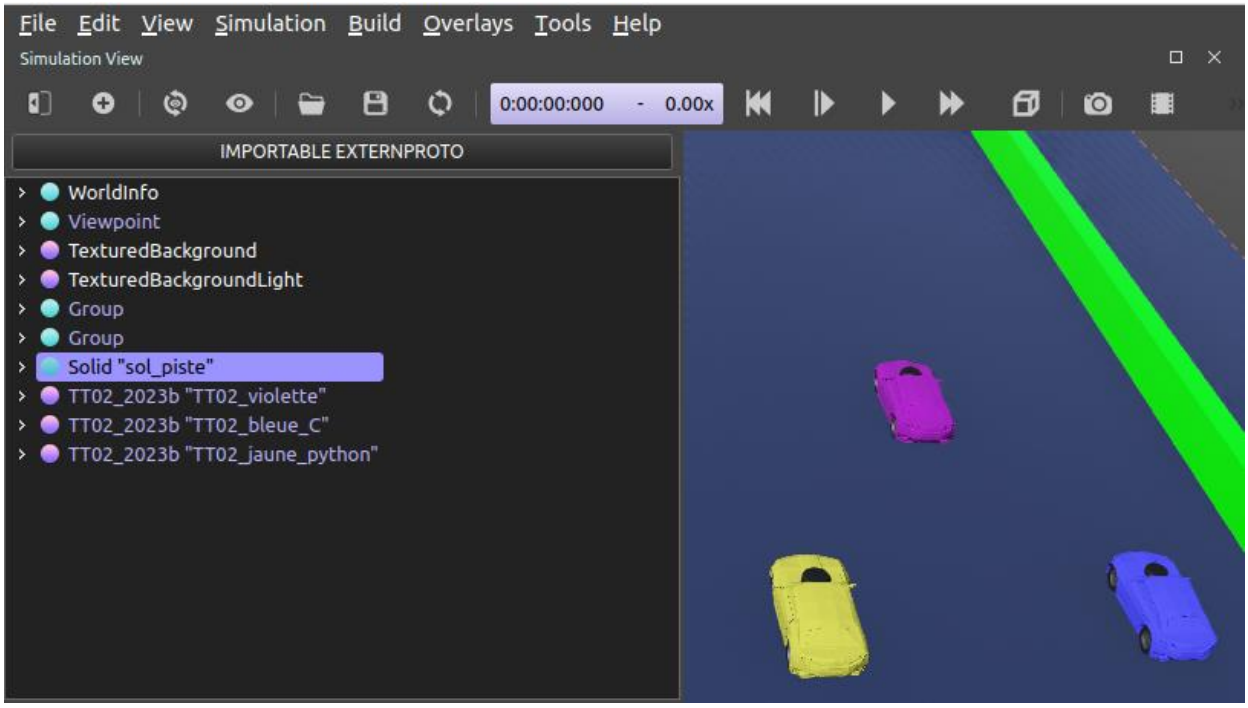


Figure 5 : 3 voitures, pour la programmation en C et en python

La voiture violette est la voiture préprogrammée (sparring partner). Une fois le programme exécuté, la voiture violette peut être conduite avec les flèches (cliquer dans l'environnement graphique et taper m). Il est possible d'afficher le tableau des données du lidar en mm (taper l). Pour bien comprendre le fonctionnement du lidar, il est intéressant d'afficher les faisceaux (View -> Optionnal Rendering -> Show Point Cloud).

La voiture jaune a un programme de base en python et la voiture bleue un programme de base en C (La vitesse est fixe, l'angle de braquage est déterminé avec les données captées par le Lidar au niveau des angles -60° et 60°). Il est facile de supprimer la voiture inutile dans l'arborescence des éléments.

En C comme en python, trois fonctions et un tableau sont utilisés pour contrôler la voiture :

- *set_vitesse_m_s()* : fonction qui permet de contrôler la voiture en indiquant sa vitesse en m/s
- *set_direction_degre()* : fonction qui permet de contrôler la voiture en indiquant sa direction en degré. Pour tourner à gauche, il faut indiquer un angle positif et un angle négatif pour tourner à droite.
- *recule()* : fonction qui permet à la voiture dans le simulateur de reculer à la demande.
- Les données du Lidar sont récupérées dans la variable *tableau_lidar_mm* (en python) ou *data_lidar_mm_main* (en C) qui est un tableau de 360 valeurs (1 par degré) dans laquelle sont stockées les distances en millimètres. L'indice 0 correspond à l'avant de la voiture. Attention, les valeurs entre 100 et 260 ne sont pas significatives car elles correspondent à des angles auxquels le lidar est face à l'habitacle de la voiture.

Que ce soit en C ou en python, l'objectif est d'abord de dépasser la voiture violette (pas très compliqué), puis d'aller le plus vite possible. Pour cela, plusieurs pistes sont proposées ici :

- Il est possible de regrouper les rayons en secteur de 10° pour chercher le secteur dont le rayon le plus court est le plus long parmi les plus courts des autres secteurs.
- Il est possible d'adapter sa vitesse à la distance de l'obstacle devant.
- Il est intéressant de détecter un obstacle pour réussir à l'éviter. On peut pour cela ajouter des morceaux de bordure de piste au milieu de la piste.
- Il est possible de reculer quand on est dans un mur, en surveillant une valeur minimale des rayons du lidar à l'avant et sur les côtés. Pour déterminer les situations de quasi-collision, on peut se baser sur 3 valeurs du Lidar : la mesure à 0°, celle à -30° et celle à 30°. Si les distances captées par le Lidar sur ces angles spécifiques sont inférieures à un certain seuil, on considère que la voiture a subi une collision. On a alors trois possibilités, qui demandent un peu de travail car il n'est pas possible d'utiliser time.sleep, cette fonction bloquante mettant aussi en pause le moteur physique (l'utilisation de time.time() peut alors être intéressante) :
  - Seuil franchi pour l'angle 0° (mur devant) : On replace les roues pour aller droit puis on recule jusqu'à ce que la valeur de lidar franchisse un seuil ou pendant 0,5s.
  - Seuil franchi pour l'angle 30° (mur à gauche) : On tourne complètement à gauche puis on recule jusqu'à ce que la valeur de lidar franchisse un seuil ou pendant 0,5s.
  - Seuil franchi pour l'angle -30° (mur à droite) : On tourne complètement à droite puis on recule jusqu'à ce que la valeur de lidar franchisse un seuil ou pendant 0,5s.
- Des méthodes avancées sont bien évidemment possible, en sortant du cadre du lycée, avec des trajectoires en forme de tentacules (https://doi.org/10.1002/rob.20256), des correcteurs issus de l'automatique (PID par exemple sur la différence entre le distance à 60 et -60°) ou avec de l'apprentissage par renforcement (voir « *Apprentissage par renforcement et transfert simulation vers réalité pour la conduite de voitures autonomes* » [2]).

Aucune connaissance sur la bibliothèque du simulateur n'est requise. Webots supporte aussi le java. Pour aller plus loin, il est possible de se référer à la documentation de Webots[1] .

3.1 - Programmer en python

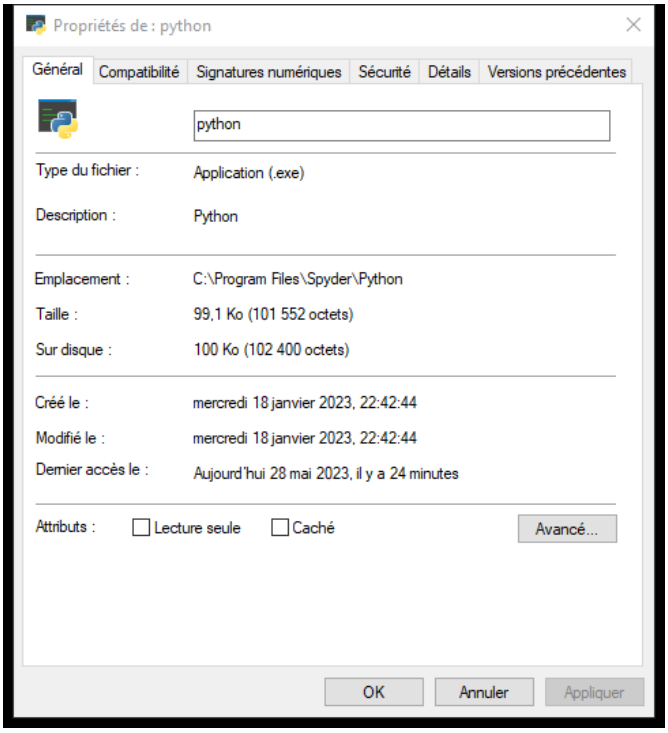
1) Configuration

Linux ne demande pas de configuration particulière. Passer directement à « Modifier un programme dans l'éditeur de texte du simulateur ».

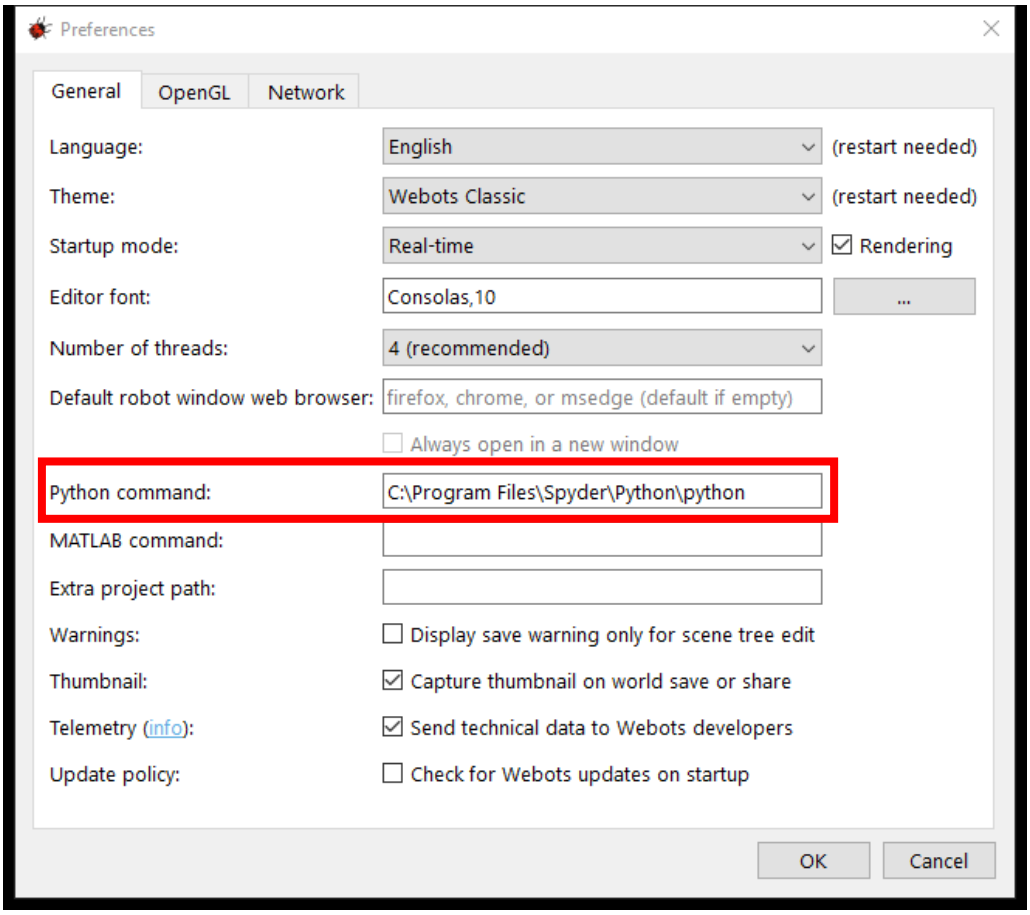
Sous windows, il faut indiquer l'interpréteur *Python* qui sera utilisé par le logiciel pour lancer les différents programmes.

- Chercher le chemin d'accès vers l'exécutable *python.exe*

1 https://www.cyberbotics.com/doc/guide/tutorials



- Dans Webots, cliquer sur ***Tools→Préférences***
- Copier le Chemin d'accès de l'exécutable dans la case *Python Command*



2) <u>Modifier un programme dans l'éditeur de texte du simulateur</u>

Webots permet de modifier les programmes *Python* directement.

- Sélectionner une des voitures dans l'arborescence et dérouler son arbre
- Sélectionner la case ***controller*** puis plus bas ***Edit***

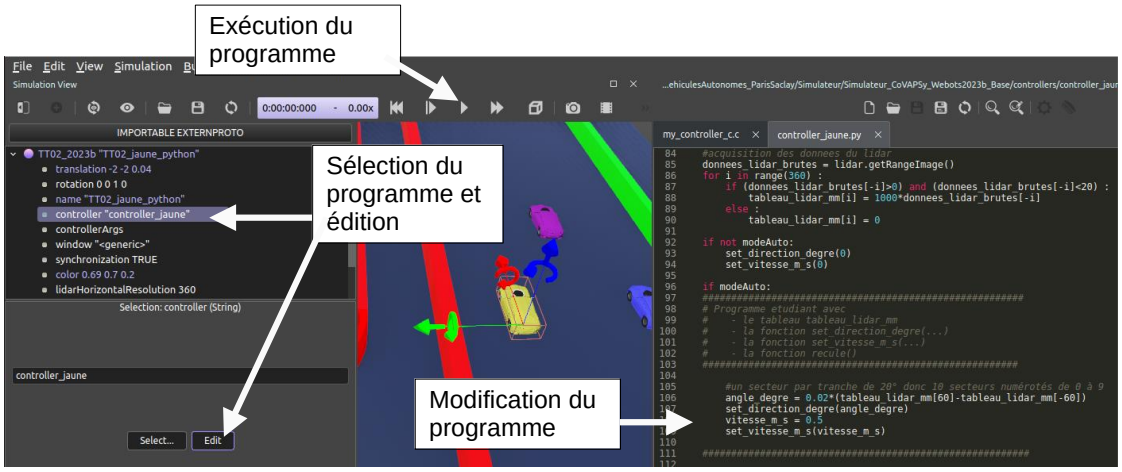


Figure 6 : Modification du programme controller de la voiture jaune

- Le code de la voiture s'affiche dans l'éditeur de texte. Le programme python de base est donné en annexe. Il faut cliquer dans la fenêtre de visualisation graphique puis taper 'a' au clavier pour lancer le mode automatique, une fois le programme en exécution.
- Il est possible qu'il faille supprimer la voiture bleue programmée en C sous windows (le code devant être compilé pour windows, comme expliqué dans la partie suivante).
- Attention, webots n'enregistre pas automatiquement le code python avant l'exécution. Il faut donc enregistrer puis lancer l'éxécution.

Remarque : Webots gère assez mal l'indentation. Celle-ci doit être réalisée uniquement avec des tabulations ou uniquement avec des espaces. Si ce n'est pas le cas, Webots affiche une erreur d'indentation, facile à corriger.

3.2 - Programmer en langage C

De la même manière, il est possible de programmer la voiture bleue en C. Il faut juste penser à enregistrer et compiler avant de lancer l'exécution. Le programme C de base est donné en annexe.

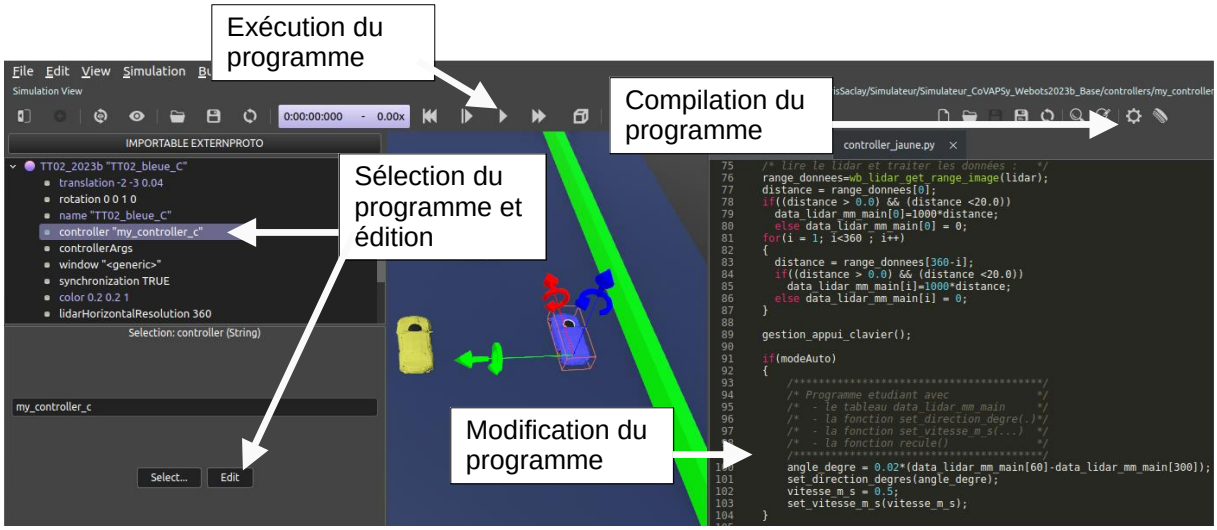


Figure 7 : Modification du programme my_controller_c.c de la voiture bleue.

3.3 - Programmation avec un environnement de développement tiers (utilisation avancée)

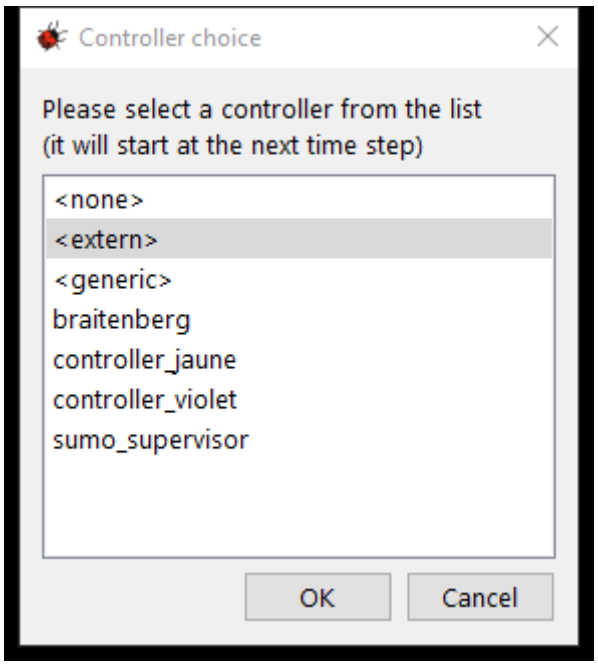
Pour programmer de manière approfondie, il est recommandé d'utiliser un environnement de développement (IDE) dédié comme *Spyder* ou *VSCode* par exemple pour python ou *Eclipse* pour le C, IDE qui permettent notamment l'usage du débogueur. Ceci ajoute cependant une couche de complexité. Deux méthodes sont possibles, une seule est exposée ici. Plus détails sont donnés dans les deux sections correspondantes de la documentation de Webots :

- https://cyberbotics.com/doc/guide/using-your-ide
- https://cyberbotics.com/doc/guide/running-extern-robot-controllers

L'utilisation d'un contrôleur externe, méthode retenue ici, permet d'avoir un contrôleur tournant sur un IDE complet (avec débogueur), y compris sur une machine distante, ce qui peut être intéressant pour une course à plusieurs voitures, chaque étudiant se connectant sur une des voitures de la piste, son tour venu.

Pour pouvoir faire le lien entre ces IDE et *Webots*, il est nécessaire de rajouter deux lignes de codes en début du programme.

- **Dans le programme python** *Spyder ou VSCode* : Rajouter les deux lignes de codes suivantes indiquant le chemin vers les bibliothèques webots, en adaptant le nom du dossier au PC et à la version de python utilisée.
  `import sys`
  Pour Windows :
  `sys.path.append('C:/Program Files/Webots/lib/controller/python38/')`
  Pour Linux : **`sys.path.append('/usr/local/webots/lib/controller/python38/')`**
  Il faut également créer la variable d'environnement indiquant le dossier d'installation de webots :
  export WEBOTS_HOME=/usr/local/webots/
- Dans *Webots*:
  - Sélectionner une voiture et dérouler l'arborescence
  - Sélectionner ***controller*** puis ***Select***
  - Choisir ***<extern>***



Il suffit par la suite de lancer la simulation puis d'exécuter le programme sur *Spyder* ou *VSCode*.

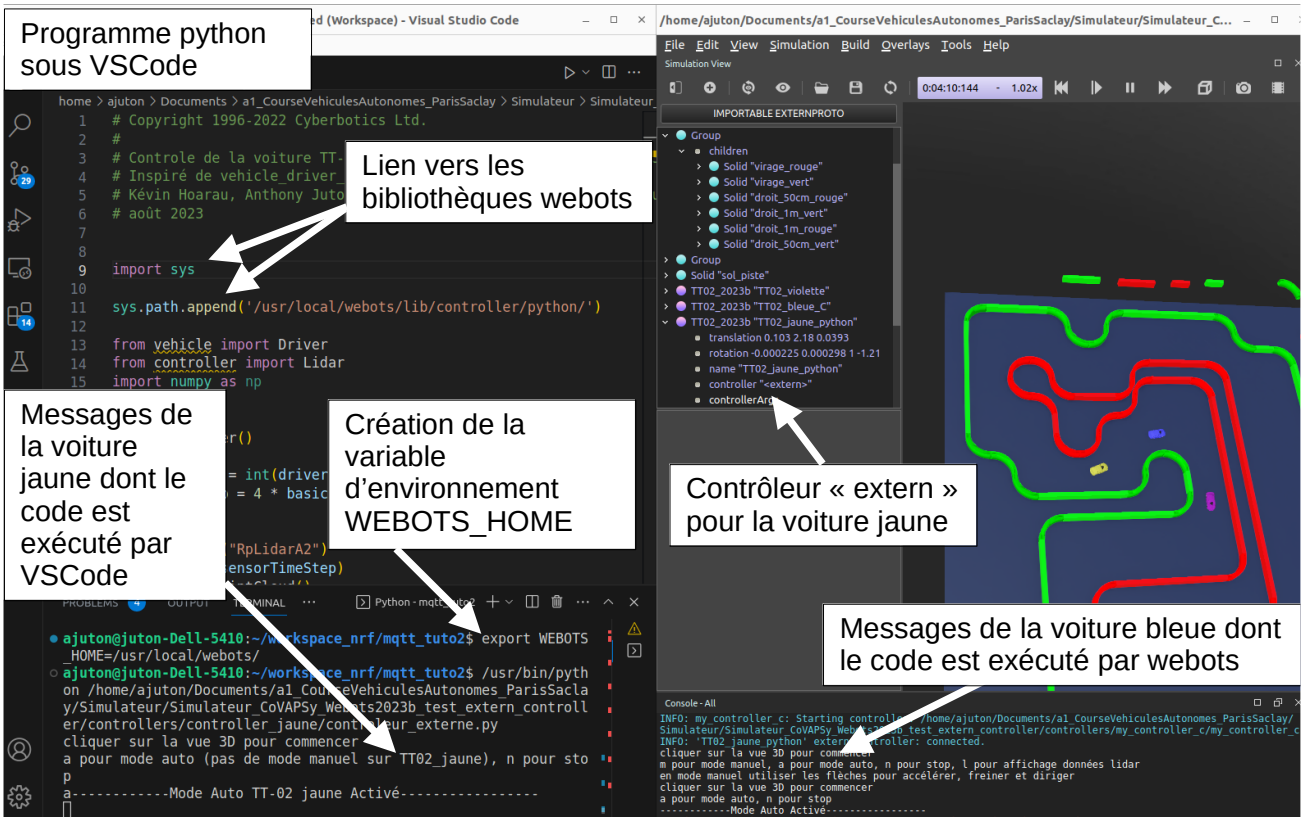


Figure 8 : Voiture jaune simulée sous Webots avec un contrôleur exécuté sous VSCode

Pour exécuter du code depuis des PCs extérieurs, il faut utiliser l'exécutable Webots, comme expliqué sur la documentation Webots.

```
$WEBOTS_HOME/webots-controller --protocol=tcp --ip-address=X.X.X.X
$WEBOTS_HOME/projects/robots/softbank/nao/controllers/nao_demo/nao_demo
```

4 - Course entre 4 voitures

Pour une course entre plusieurs voitures, il suffit d'ajouter d'autres voitures, de changer leur couleur et d'ajouter les contrôleurs.

4.1 - Ajout du contrôleur

Les contrôleurs (hors contrôleurs externes) étant situés obligatoirement dans le dossier *controllers* du projet, il est nécessaire d'ajouter le nouveau contrôleur à cet emplacement, soit en copiant un dossier *controller* envoyé par un étudiant, soit en copiant/collant un dossier *controller* existant. Le nom du fichier doit correspondre au nom du dossier.

Figure 9 : Ajout d'un contrôleur au projet pour une course à 4 voitures

4.2 - Ajout de la voiture

Une fois le contrôleur ajouté, il est possible d'ajouter la voiture et de modifier ses paramètres.

Attention, pour faire des modifications pérennes, il est préférable de se placer à t=0 (bouton « retour arrière » à gauche du bouton exécuter) et d'enregistrer ensuite le modifications avant de lancer l'exécution.

- Se placer à t=0
- Sélectionner la voiture jaune dans l'arborescence
- Faire ***Clic droit→Copy*** puis ***Clic droit→Paste***

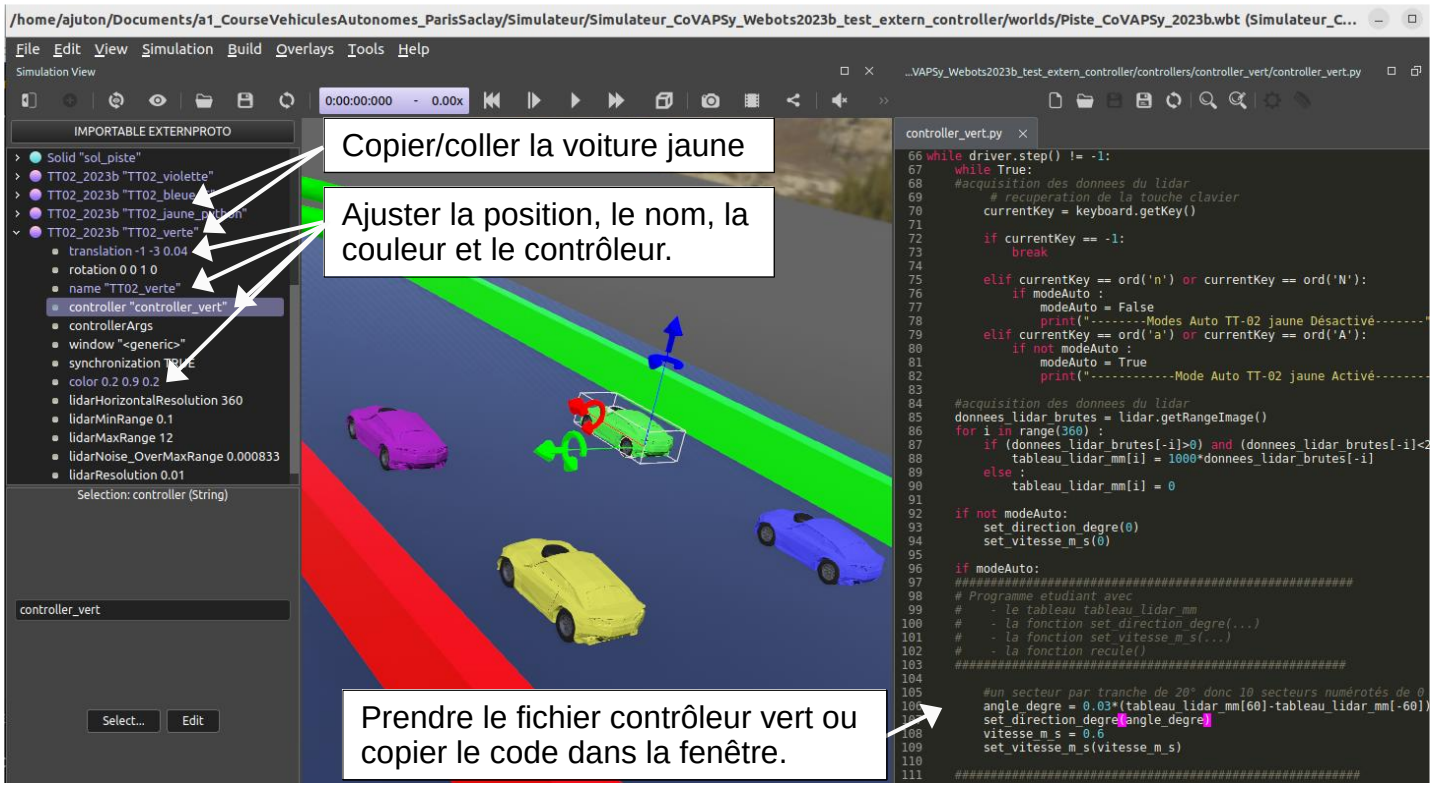


Figure 10 : Ajout d'une voiture au projet

Attention, penser à sauvegarder la nouvelle configuration de départ : *File* > *Save World*

5 - Ouvertures

Cette ressource amène à réaliser et tester des algorithmes permettant de faire concourir une ou plusieurs voitures de manière autonome sur le simulateur.

Côté informatique, la visualisation dynamique des données du lidar (graph en coordonnées cylindriques) et des actions de la voiture et leur archivage, comme la télémétrie pour les voitures réelles, peut être un axe de travail intéressant, dans l'objectif d'améliorer les performances de la voiture.

Sur Webots, l'ajout d'un robot « superviseur » , ayant l'accès aux données des autres robots permet de faire un chronométrage individuel de chaque voiture :
https://www.cyberbotics.com/doc/reference/supervisor

Les ressources « *CoVAPSy : Premiers programmes python sur la voiture réelle* » [3] et « *CoVAPSy : Premiers programmes en C sur la voiture réelle* » [4] consistent quant à elles à transférer cet algorithme sur la voiture réelle, disposant des mêmes fonctions de contrôle de vitesse et de direction et du même tableau d'acquisition des données lidar. Le transfert de la simulation à la réalité est alors intéressant pour mettre en évidence les limites du simulateur (prise en compte imparfaite de la dynamique de la voiture, de sa direction notamment) et des contraintes de la réalité (nécessité de prendre en compte la tension batterie ou de faire un asservissement de vitesse, absence de mesure renvoyée par le lidar en cas de mauvaise réflexion…).

6 - Annexes

6.1 - Programme python (contrôleur de la voiture jaune)

```
# Copyright 1996-2022 Cyberbotics Ltd.
#
# Controle de la voiture TT-02 simulateur CoVAPSy pour Webots 2023b
# Inspiré de vehicle_driver_altino controller
# Kévin Hoarau, Anthony Juton, Bastien Lhopitallier, Martin Raynaud
# août 2023

from vehicle import Driver
from controller import Lidar
import time

driver = Driver()

basicTimeStep = int(driver.getBasicTimeStep())
sensorTimeStep = 4 * basicTimeStep

#Lidar
lidar = Lidar("RpLidarA2")
lidar.enable(sensorTimeStep)
lidar.enablePointCloud()

#clavier
keyboard = driver.getKeyboard()
keyboard.enable(sensorTimeStep)

# vitesse en km/h
speed = 0
maxSpeed = 28 #km/h

# angle de la direction
angle = 0
maxangle_degre = 16


# mise a zéro de la vitesse et de la direction
driver.setSteeringAngle(angle)
driver.setCruisingSpeed(speed)

tableau_lidar_mm=[0]*360

def set_vitesse_m_s(vitesse_m_s):
    speed = vitesse_m_s*3.6
    if speed > maxSpeed :
        speed = maxSpeed
    if speed < 0 :
        speed = 0
    driver.setCruisingSpeed(speed)

def set_direction_degre(angle_degre):
    if angle_degre > maxangle_degre:
        angle_degre = maxangle_degre
    elif angle_degre < -maxangle_degre:
        angle_degre = -maxangle_degre
    angle = -angle_degre * 3.14/180
    driver.setSteeringAngle(angle)

def recule(): #sur la voiture réelle, il y a un stop puis un recul pendant 1s.
    driver.setCruisingSpeed(-1)

# mode auto desactive
modeAuto = False
print("cliquer sur la vue 3D pour commencer")
print("a pour mode auto (pas de mode manuel sur TT02_jaune), n pour stop")

while driver.step() != -1:
    while True:
    #acquisition des donnees du lidar
         # recuperation de la touche clavier
        currentKey = keyboard.getKey()

        if currentKey == -1:
            break

        elif currentKey == ord('n') or currentKey == ord('N'):
            if modeAuto :
                modeAuto = False
                print("--------Modes Auto TT-02 jaune Désactivé-------")
```

```
        elif currentKey == ord('a') or currentKey == ord('A'):
            if not modeAuto :
                modeAuto = True
                print("------------Mode Auto TT-02 jaune Activé----------------")

    #acquisition des donnees du lidar
    donnees_lidar_brutes = lidar.getRangeImage()
    for i in range(360) :
        if (donnees_lidar_brutes[-i]>0) and (donnees_lidar_brutes[-i]<20) :
            tableau_lidar_mm[i] = 1000*donnees_lidar_brutes[-i]
        else :
            tableau_lidar_mm[i] = 0

    if not modeAuto:
        set_direction_degre(0)
        set_vitesse_m_s(0)

    if modeAuto:
    ##########################################################
    # Programme etudiant avec
    #    - le tableau tableau_lidar_mm
    #    - la fonction set_direction_degre(...)
    #    - la fonction set_vitesse_m_s(...)
    #    - la fonction recule()
    ##########################################################

        #un secteur par tranche de 20° donc 10 secteurs numérotés de 0 à 9
        angle_degre = 0.02*(tableau_lidar_mm[60]-tableau_lidar_mm[-60])
        set_direction_degre(angle_degre)
        vitesse_m_s = 0.5
        set_vitesse_m_s(vitesse_m_s)

    ##########################################################
```

6.2 - Programme C (contrôleur de la voiture bleue)

```
/*
 * File:          my_controller_c.c
 * Date:          23 mai 2023
 * Description:
 * Author: Bruno Larnaudie, Anthony Juton
 * Modifications: 24 août 2023
 */

/*
 * You may need to add include files like <webots/distance_sensor.h> or
 * <webots/motor.h>, etc.
 */
#include <math.h>
#include <webots/robot.h>
#include <webots/vehicle/car.h>
#include <webots/vehicle/driver.h>
#include <webots/keyboard.h>
#include <stdio.h>
#include <webots/lidar.h>
/*
 * You may want to add macros here.
 */
#define TIME_STEP 32
#define SIZE_TABLEAU 200
#define MAX_SPEED 6.28  // Vitesse maximale des moteurs

const float* range_donnees;
//float range_donnees[SIZE_TABLEAU];
unsigned char gestion_appuie_clavier(void);
unsigned char modeAuto=0;

// prototype des fonctions
void affichage_consigne();
void set_direction_degres(float angle_degre);
void set_vitesse_m_s(float vitesse_m_s);
unsigned char gestion_appui_clavier(void);
void recule(void);

//vitesse en km/h
float speed = 0;
float maxSpeed = 28; //km/h

// angle max de la direction
```

```c
float maxangle_degre = 16;

  /* main loop
   * Perform simulation steps of TIME_STEP milliseconds
   * and leave the loop when the simulation is over
   */

int main(int argc, char **argv)
{
  unsigned int i;
  signed int data_lidar_mm_main[360];
  float angle_degre, vitesse_m_s;
  /* necessary to initialize webots stuff */
  //intialisation du conducteur de voiture
  wbu_driver_init();
  //enable keyboard
  wb_keyboard_enable(TIME_STEP);
  // enable lidar
  WbDeviceTag lidar = wb_robot_get_device("RpLidarA2");
  wb_lidar_enable(lidar,TIME_STEP);
  // affichage des points lidar sur la piste
  wb_lidar_enable_point_cloud(lidar);

  affichage_consigne();
  set_direction_degres(0);
  set_vitesse_m_s(0);
  while (wbu_driver_step() != -1)
  {
    float distance;
    /* lire le lidar et traiter les données :   */
    range_donnees=wb_lidar_get_range_image(lidar);
    distance = range_donnees[0];
    if((distance > 0.0) && (distance <20.0))
      data_lidar_mm_main[0]=1000*distance;
      else data_lidar_mm_main[0] = 0;
    for(i = 1; i<360 ; i++)
    {
      distance = range_donnees[360-i];
      if((distance > 0.0) && (distance <20.0))
        data_lidar_mm_main[i]=1000*distance;
      else data_lidar_mm_main[i] = 0;
    }
    gestion_appui_clavier();
    if(modeAuto)
    {
        /*****************************************/
        /* Programme etudiant avec               */
        /*  - le tableau data_lidar_mm_main      */
        /*  - la fonction set_direction_degre(.)*/
        /*  - la fonction set_vitesse_m_s(...)   */
        /*  - la fonction recule()               */
        /*****************************************/
        angle_degre = 0.02*(data_lidar_mm_main[60]-data_lidar_mm_main[300]); //distance à 60° ,-60°
        set_direction_degres(angle_degre);
        vitesse_m_s = 0.5;
        set_vitesse_m_s(vitesse_m_s);
    }
  }
  /* This is necessary to cleanup webots resources */
  wbu_driver_cleanup();
  return 0;
}

unsigned char gestion_appui_clavier(void)
{
  int key;
  key=wb_keyboard_get_key();
  switch(key)
  {
    case -1:
      break;

    case 'n':
    case 'N':
      if (modeAuto)
      {
        modeAuto = 0;
        printf("--------Mode Auto Désactivé-------");
      }
      break;

    case 'a':
```

```
    case 'A':
      if (!modeAuto)
      {
        modeAuto = 1;
        printf("------------Mode Auto Activé-----------------");
      }
      break;

    default:
      break;
  }
  return key;
}

void affichage_consigne()
{
  printf("cliquer sur la vue 3D pour commencer\n");
  printf("a pour mode auto, n pour stop\n");
}

void set_direction_degres(float angle_degre)
{
  float angle=0;
  if(angle_degre > maxangle_degre)
    angle_degre = maxangle_degre;
  else if(angle_degre < -maxangle_degre)
    angle_degre = -maxangle_degre;
  angle = -angle_degre * 3.14/180;
  wbu_driver_set_steering_angle(angle);
}

void set_vitesse_m_s(float vitesse_m_s){
  float speed;
  speed = vitesse_m_s*3.6;
  if(speed > maxSpeed)
    speed = maxSpeed;
  if(speed < 0)
    speed = 0;
  wbu_driver_set_cruising_speed(speed);
}

void recule(void){
    wbu_driver_set_cruising_speed(-1);
}
```

Références :

[1]: Course Voitures Autonomes Paris Saclay (CoVAPSy) : Travaux pratiques autour des voitures autonomes, T. Boulanger, E. Délègue, K. Hoarau, A. Juton, https://eduscol.education.fr/sti/si-ens-paris-saclay/ressources_pedagogiques/covapsy-tp-autour-des-voitures-autonomes

[2]: Apprentissage par renforcement et transfert simulation vers réalité pour la conduite de voitures autonomes , R. Bennani, K. Hoarau, A. Juton, https://eduscol.education.fr/sti/si-ens-paris-saclay/ressources_pedagogiques/apprentissage-renforcement-transfert-simulation-vers-realite-pourla-conduite-voitures-autonomes

[3]: CoVAPSy : Premiers programmes python sur la voiture réelle, T. Boulanger, E. Délègue, K. Hoarau, A. Juton, https://eduscol.education.fr/sti/si-ens-paris-saclay/ressources_pedagogiques/covapsy-premiers-programmes-python-sur-voiture-reelle

[4] : CoVAPSy : Premiers programmes en langage C sur la voiture réelle, A. Azan, A. Juton, https://eduscol.education.fr/sti/si-ens-paris-saclay/ressources_pedagogiques/covapsy-premiers-programmes-langage-c-sur-la-voiture-reelle

CoVAPSy : Premiers programmes en langage C sur la voiture réelle

La Revue 3E.I
Culture Sciences de l'Ingénieur

Antoine AZAN[1] – Anthony JUTON[2]

Édité le 21/11/2023

école normale supérieure paris–saclay

[1] *Professeur agrégé de génie électrique, Lycée Bergson, PARIS*
[2] *Professeur agrégé de physique appliquée au département Nikola Tesla de l'ENS Paris-Saclay*

Cette ressource fait partie du N°111 de La Revue 3EI de janvier 2024.

L'objectif de cette ressource est de présenter la configuration et programmation de la voiture autonome CoVAPSy en langage C à base d'un microcontrôleur de la famille des STM32 de STMicroelectronics.

La voiture type est présentée en détail dans la ressource « *Course Voitures Autonomes Paris Saclay (CoVAPSy) : Travaux pratiques autour des voitures autonomes* » [1]. Une version nommée CoVAPSy_STM32, sans nano-ordinateur, et une version encore plus simplifiée dite *CoVAPSy_STM32only, avec juste un lidar et un micro-contrôleur, sur une carte de prototypage, sont* présentée ici.

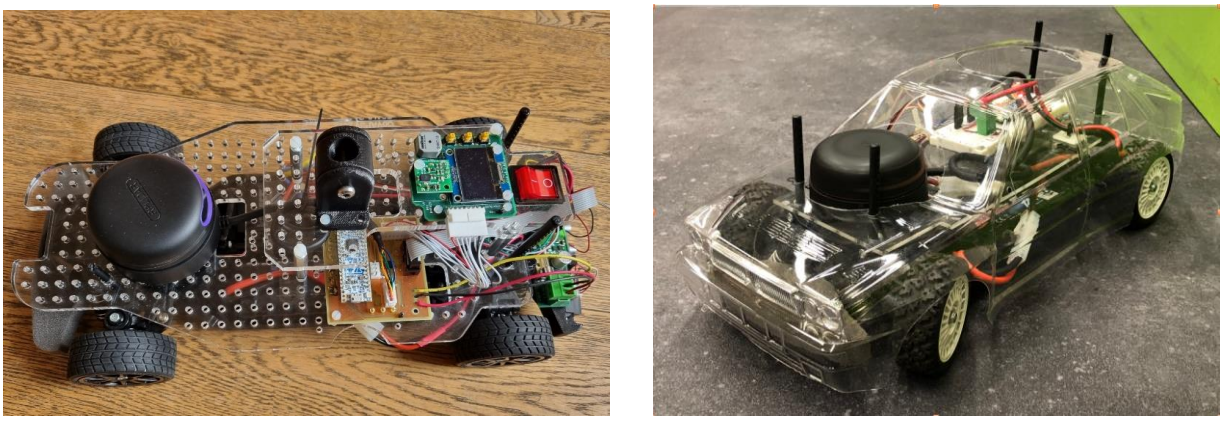


Figure 1 : Voitures CoVAPSy_STM32 et CoVAPSy_STM32only

Cette ressource fait écho à la ressource sur la programmation de la voiture en langage Python « *CoVAPSy : Premiers programmes python sur la voiture réelle* » [2]. Les fonctions utilisées ici sur la voiture réelle sont les mêmes que les fonctions en langage C utilisées sur le simulateur webots présenté dans la ressource « *CoVAPSy : Mise en œuvre du Simulateur Webots* » [3].

Les bibliothèques et programmes de test en langage C sont disponibles sur le dépôt github : https://github.com/ajuton-ens/CourseVoituresAutonomesSaclay/tree/main/Bibliotheques_logicielles/programmes_langageC_base_lidar_propulsion_direction_conduite.

1 - Architecture matérielle

Le diagramme de définition de bloc SysML ci-dessous représente la composition matérielle de la voiture autonome *CoVAPSy_STM32only*.

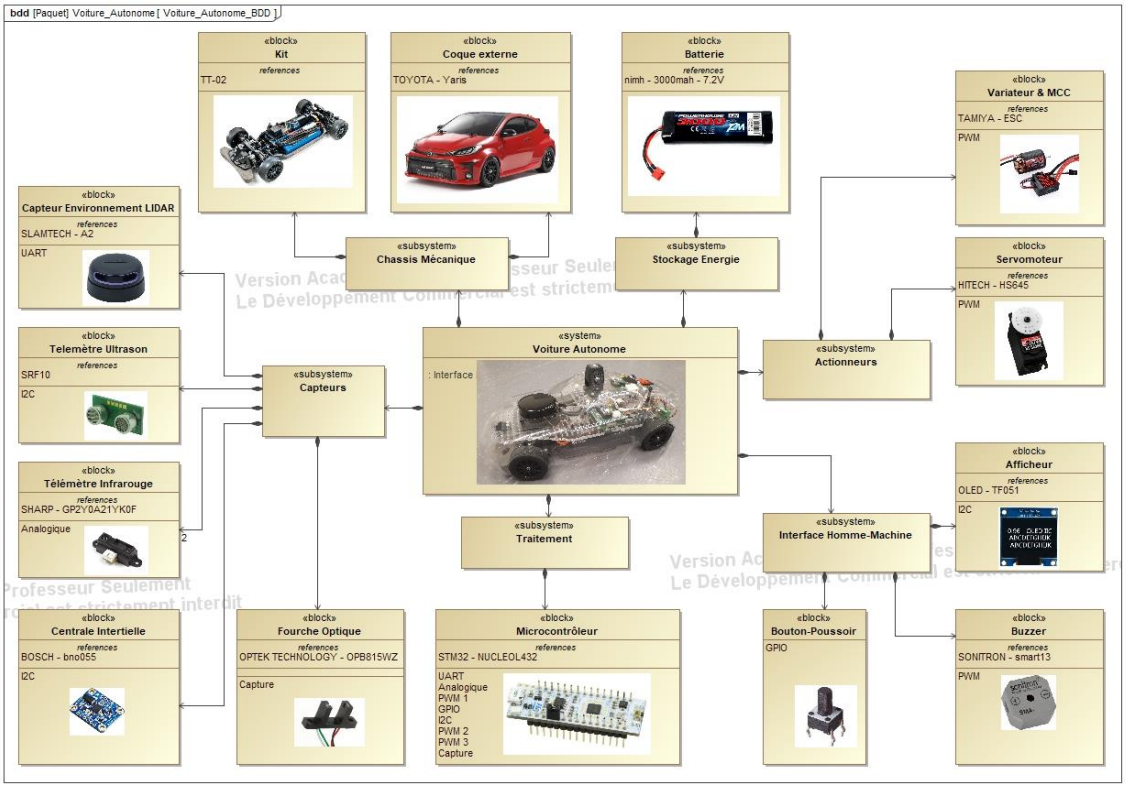


Figure 2 : Diagramme de définition de bloc SysML de la voiture autonome CoVAPSy_STM32only

Le diagramme de définition de bloc interne ci-dessous présente les flux de données entre les différents périphériques et le microcontrôleur.

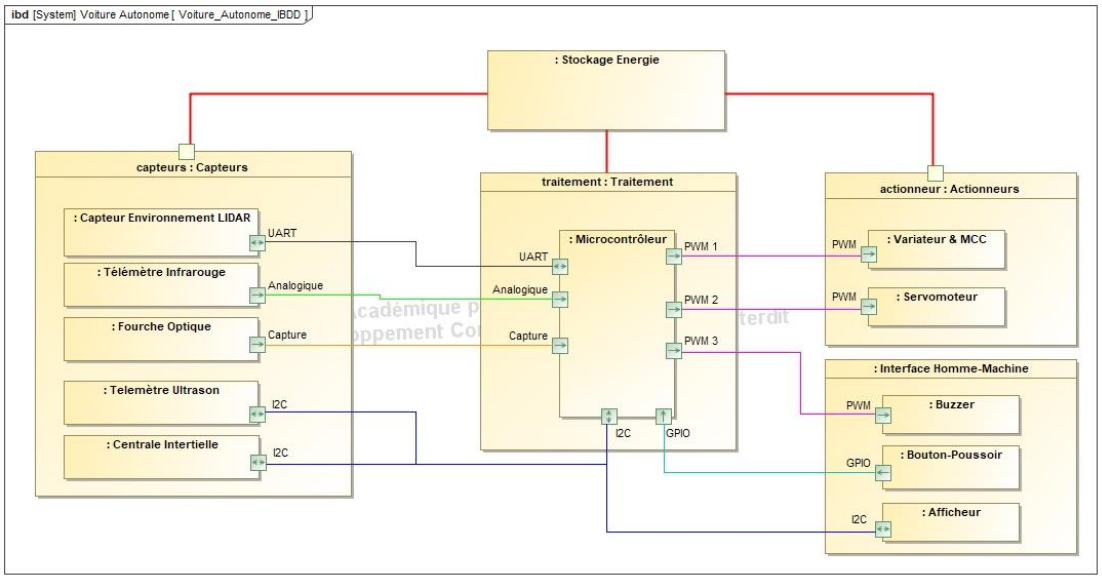


Figure 3 : Diagramme de définition de blocs internes de la voiture autonome CoVAPSy_STM32_only

Le microcontrôleur utilisé a donc besoin de gérer :

- Une liaison UART pour recevoir les données du Lidar
- Un bus I2C pour le télémètre à ultrason, la centrale inertielle et l'afficheur,
- Quatre signaux à modulation de largeur d'impulsion (Pulse Width Modulation PWM) pour la propulsion via l'ensemble variateur et moteur à courant continu, le servomoteur de direction, le Lidar et le buzzer
- Deux entrées analogiques pour les télémètres infrarouges
- Deux entrées Tout ou Rien (GPIO) pour les boutons-poussoirs
- Une entrée capture pour la fourche optique

Le microcontrôleur utilisé pour gérer l'ensemble des périphériques est un microcontrôleur STM32G431KB. Afin de faciliter l'intégration dans la voiture autonome CoVAPSy_STM32, la carte de développement NUCLEO-G431KB est utilisée. Cette carte de développement comporte un microcontrôleur STM32G431KB soudée sur une carte électronique au format Arduino nano. Le microcontrôleur STM32G431KB est basé sur un ARM Cortex-M4-32 bits, pouvant être cadencé jusqu'à une fréquence d'horloge de 170MHz avec une mémoire Flash de 128ko et un mémoire RAM de 32ko.

1.1 - Configuration matérielle minimale

Une configuration matérielle minimale comportant uniquement la partie châssis mécanique, l'ensemble variateur et moteur à courant continu, le servomoteur, le Lidar, la batterie et le microcontrôleur est envisageable. Cette configuration minimale permet, à un coût réduit de 472€ TTC, de mettre en œuvre la voiture autonome *CoVAPSy_STM32only*.

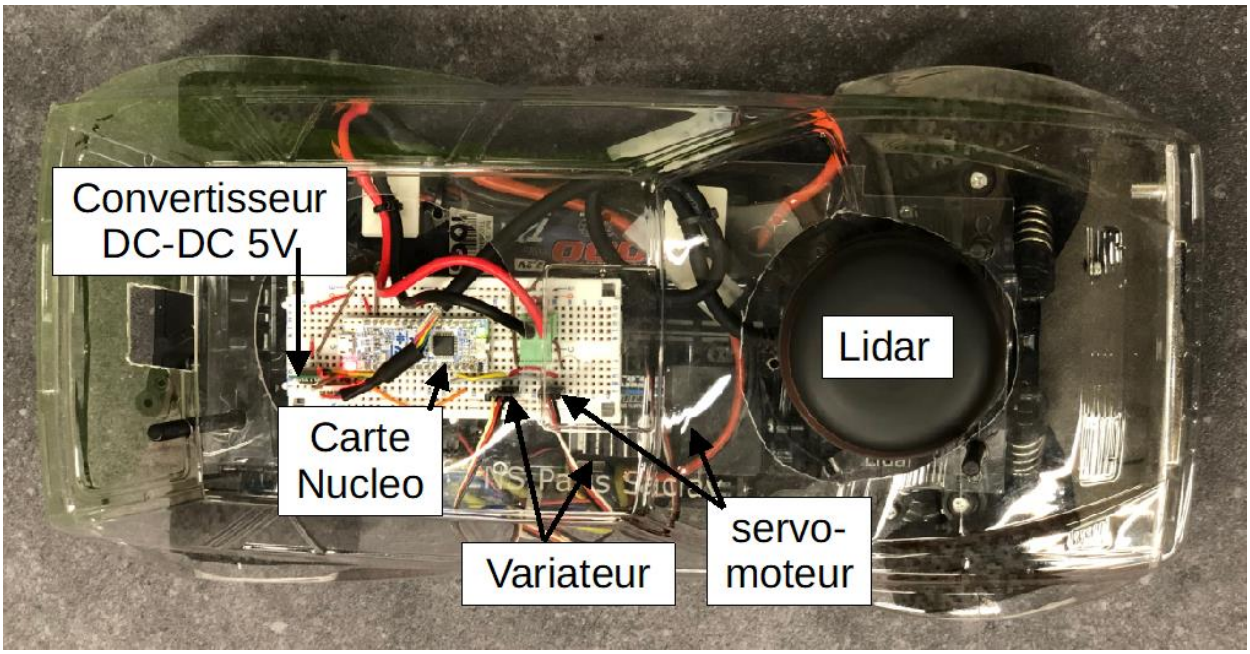


Figure 4 : Voiture CoVAPSy_STM32only, avec une carte de prototypage

Le tableau ci-dessous liste les éléments de la configuration minimale *CoVAPSy_STM32only*, un exemple de distributeur ainsi que le prix de chaque élément.

Matériel	Description	Exemple de distributeur - référence	Prix unitaire TTC (€)
Chassis mécanique	Tamiya TT-02 Toyota GR 86	RCTeam - 58694	134,90
Batterie	T2M Accu 7.2v Nimh 3000mah	RCTeam - T1006300	27,30
Servomoteur de direction	Konect Servo 9kg 0.13s Digital KN-0913LVMG	RCTeam - KN-0913LVMG	19,90
Microcontrôleur STM32	Carte de développement With Stm32g431kb Mcu	RS 196-2534 NUCLEO-G431KB	13,78
Convertisseur DC-DC 5V	Régulateur de commutation Murata, sortie 5V, 1.5A, 7.5W	RS 796-2132 OKI-78SR	6,89
Lidar	RPLIDAR A2M12 360 Slamtec A2-M12	Robotshop - RB-Rpk-22	269,03
TOTAL			472

La figure ci-dessous présente le schéma de câblage des différents composants de cette configuration minimale. La connexion entre les différents périphériques peut se faire à l'aide d'une platine d'essai (breadboard) et de fils. Un circuit imprimé supportera mieux les vibrations.

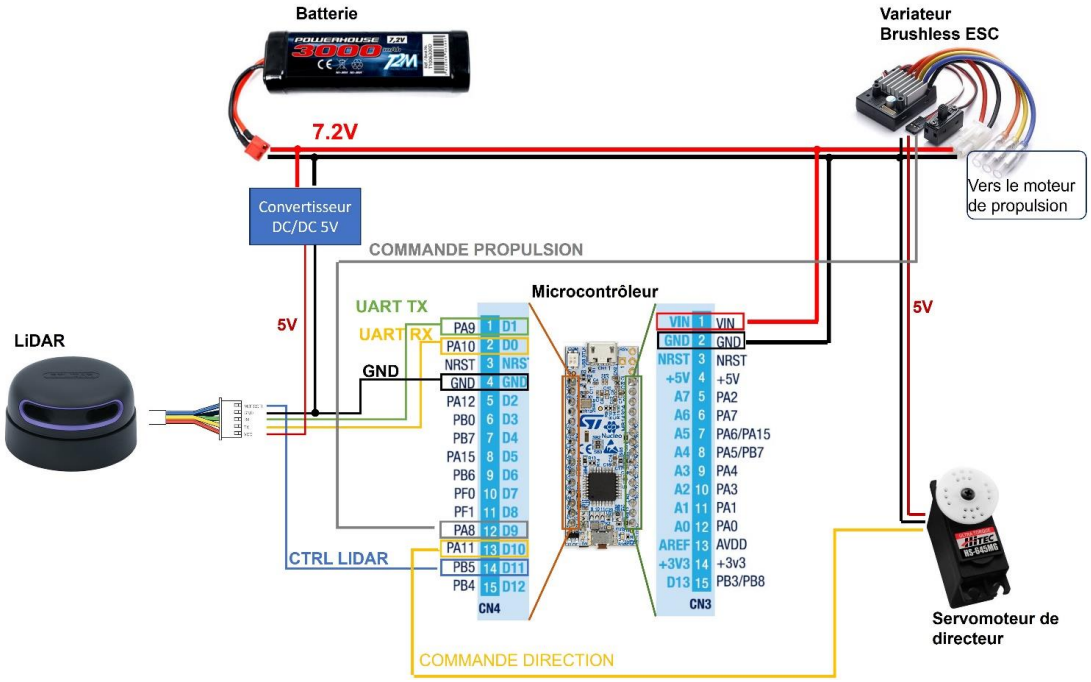


Figure 5 : Schéma de câblage pour la configuration minimale CoVAPSy_STM32only

1.2 - Configuration matérielle complète

Une configuration matérielle complète avec l'ensemble des périphériques listés dans le diagramme de définition de bloc SysML permet de mettre en œuvre l'ensemble des fonctionnalités de la voiture autonome *CoVAPSy_STM32*.

Afin d'intégrer l'ensemble de ces périphériques sur la voiture autonome *CoVAPSy_STM32*, une carte électronique a été conçue spécifiquement afin de permettre l'intégration de la carte NUCLEO-G321KB dans la voiture dotée des cartes interface (convertisseur DC/DC et connectiques vers les différents capteurs et actionneurs et la batterie) et mezzanine (écran, boutons, buzzer). Les schémas et PCB des cartes sont sur le dépôt git de la course : https://github.com/ajuton-ens/CourseVoituresAutonomesSaclay/tree/main/Materiel/Cartes_electroniques

Le schéma électronique ainsi que son circuit imprimé sont représentés ci-dessous.

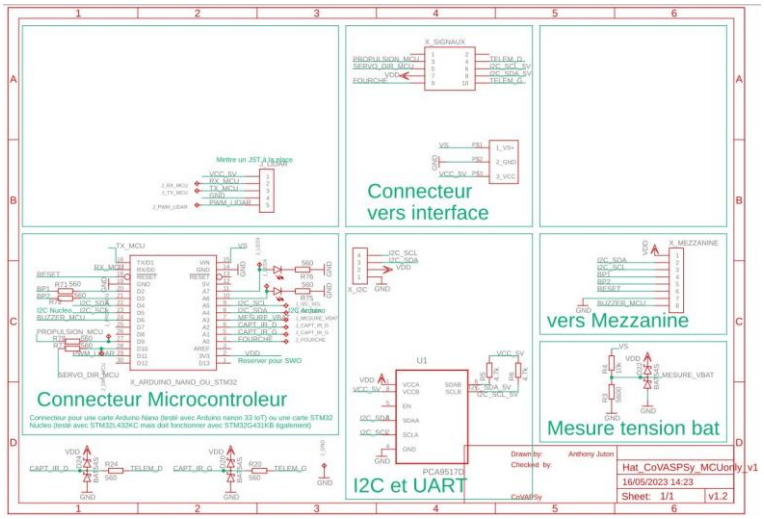


Figure 6 : Schéma électronique de la carte d'intégration du microcontrôleur STM32G431KB

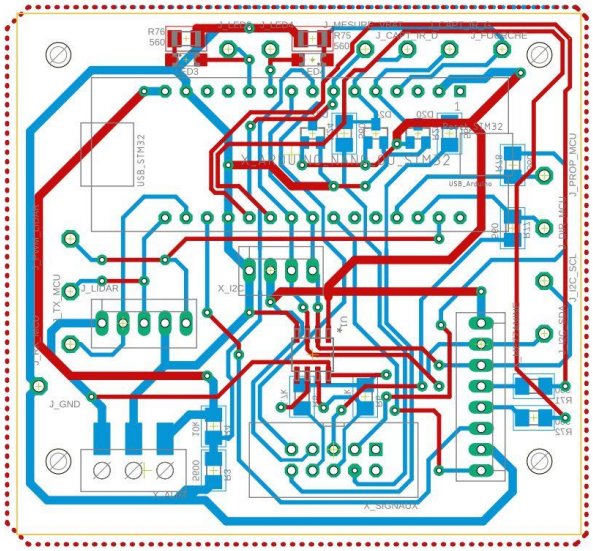


Figure 7 : Circuit Imprimé de la carte d'intégration du microcontrôleur STM32G431KB

Le microcontrôleur est programmé à l'aide du logiciel STM32CubeIDE version 1.13.1 développé par ST Microelectronics. Ce logiciel permet la configuration des périphériques, des broches, de l'horloge, la génération du code d'initialisation, la compilation et le téléversement du programme ainsi que le débogage.

2 - Mise en œuvre des entrées/sorties et bibliothèques associées

Dans cette partie, il est fait référence à de nombreuses bibliothèques et fonctions spécifiques à chaque périphérique. Toutes les bibliothèques et le programme de démonstration complet sont disponibles sur le dépôt github :
https://github.com/ajuton-ens/CourseVoituresAutonomesSaclay/tree/main/Bibliotheques_logicielles/programmes_C_base_lidar_propulsion_direction_conduite.

2.1 - Configuration des entrées/sorties du microcontrôleur STM32

Précédemment, l'ensemble des broches nécessaires pour commander les actionneurs et recevoir les données des capteurs a été listé.

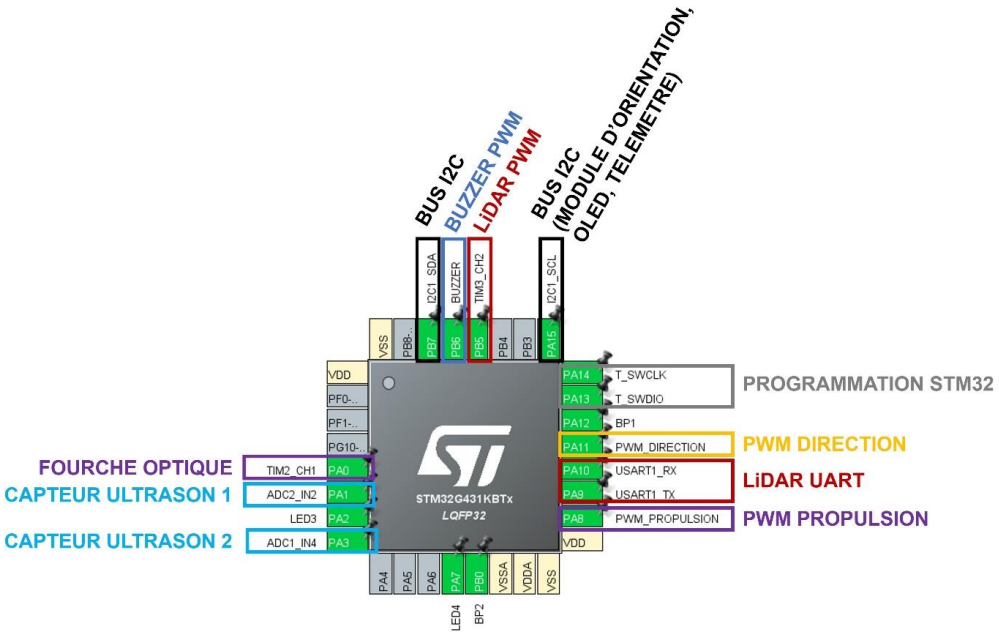


Figure 8 : Configuration complète des broches du STM32

Le tableau ci-dessous résume les broches du microcontrôleur STM32 utilisé pour chaque périphérique.

PERIPHERIQUES	BROCHES	TYPE
Lidar	PA9	Liaison série Tx
	PA10	Liaison série Rx
	PB5	PWM sur Timer 3 Channel 2
PROPULSION	PA8	PWM sur Timer 1 Channel 1
DIRECTION	PA11	PWM sur Timer 1 Channel 4
CENTRALE INERTIELLE	PA15	I2C SCL
	PB7	I2C SDA
AFFICHEUR OLED	PA15	I2C SCL
	PB7	I2C SDA
TELEMETRE ULTRASON	PA15	I2C SCL
	PB7	I2C SDA
TELEMETRE INFRAROUGE 1	PA1	Entrée Analogique
TELEMETRE INFRAROUGE 2	PA3	Entrée Analogique
BUZZER	PB6	PWM sur Timer 4 Channel 1
FOURCHE OPTIQUE	PA0	Timer 2 - Channel 1 en mode capture

Une présentation détaillée de la mise en œuvre des périphériques principaux : Lidar, Moteur à propulsion et servomoteur est proposée ci-dessous. La mise-en-œuvre des autres périphériques (Centrale inertielle, fourche optique, capteur ultrason, télémètre, afficheur OLED) sera présentée via des fiches individuelles qui seront publiées sur le dépôt github.

2.2 - Lidar

Le Lidar permet de sonder l'environnement de la voiture autonome *CoVAPSy_STM32* à 360°. Les mesures du Lidar vers l'arrière de la voiture sont perturbées par l'habitacle de la voiture.

Le Lidar utilisé est un Lidar de la marque SLAMTECH de référence A2M12. Le Lidar envoie les données via une liaison série avec un débit binaire de 256 000 bps. La version A2M8 encore disponible chez certains revendeurs fonctionne à 115200 bps, débit plus facile à gérer par la carte microcontrôleur.

La liaison série du microcontrôleur est configurée de la manière suivante :

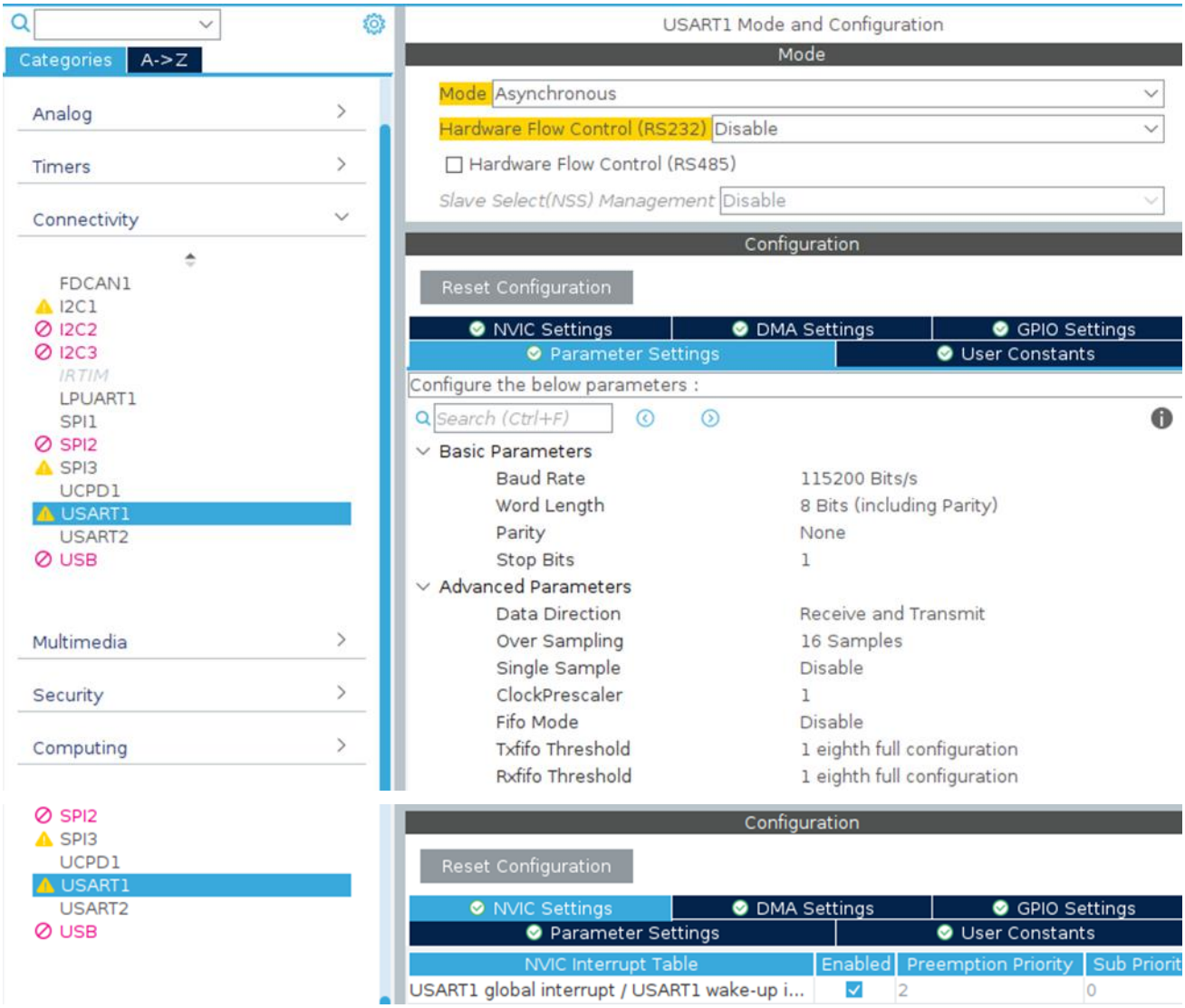


Figure 9 : Configuration de la liaison série associée à un Lidar A2M8 (115200 bit/s)

La librairie *CoVAPSy_Lidar* comporte une fonction permettant d'initialiser le Lidar et de fixer sa vitesse de rotation. La fonction *Lidar_init()* démarre et réinitialise le Lidar. Un signal PWM généré par le microcontrôleur permet de fixer la vitesse de rotation du Lidar. La fréquence du timer associé à cette PWM est fixée à 10 MHz. La période de la PWM est fixée à 40µs. La durée à l'état haut de la PWM est fixée à 30µs. La fonction *Lidar_init()* envoi la requête *START_SCAN* pour démarrer les acquisitions Lidar, puis cette fonction active l'interruption de la liaison UART associée au Lidar.

La réception des données du Lidar s'effectue via une liaison série UART sur interruption. A chaque réception d'un octet, la donnée reçue est stockée dans une variable *Data_RX_LIDAR* de type entier non-signé de 16 bits.

Les sept premiers octets émis par le Lidar suite à la requête *START SCAN* sont des octets dit *descriptor* ne contenant pas d'information de mesure d'angle ou de distance. Ces sept premiers octets reçus ne sont pas donc stockés dans le tableau *Data_Lidar_mm*.

Une fois le *descriptor* passé, les données sont envoyées par groupe de 5 octets contenant l'information de qualité de la mesure (codée sur 1 octet), l'information d'angle de la mesure (codée sur 2 octets) et l'information de distance mesurée (codée sur 2 octets).

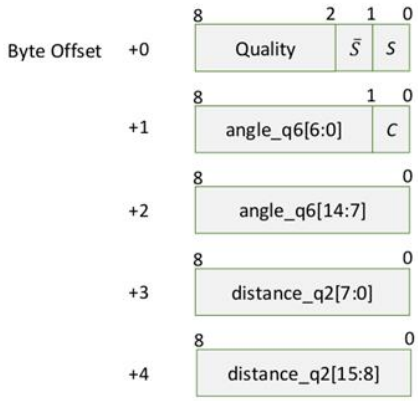


Figure 4-4 Format of a RPLIDAR Measurement Result Data Response Packet

Figure 10 : Format des paquets envoyés par le Lidar, extrait de la documentation slamtec_rplidar_protocol

À chaque nouveau groupe de 5 octets reçus, le programme extrait les informations d'angle et de distance puis stocke la distance dans un tableau de 360 lignes nommé *Data_Lidar_mm*. Chaque index du tableau *Data_Lidar_mm* correspond à un angle de mesure compris entre 0° et 359°. La donnée reçue est stockée dans le tableau *Data_Lidar_mm* si la qualité de la mesure associée est suffisante. Les angles dans le programme sont comptés dans le sens trigonométriques, comme sur le simulateur, alors que le lidar les mesure dans le sens horaire. C'est la raison de la courte conversion avant le stockage dans le tableau.

Le code ci-dessous présente le programme pour réceptionner les trames du Lidar à chaque interruption et compléter le tableau *Data_Lidar_mm*. Le tableau *Data_Lidar_mm* peut être utilisé dans le programme principal pour prendre la décision de vitesse et de direction de la voiture. La drapeau *drapeau_fin_tour*, mis à un par la fonction d'interruption indique que le lidar a fini son tour (passage à -100°). C'est le bon moment pour le programme principal pour récupérer des données fraîches et remettre le drapeau à zéro.

Le programme principal :

```
/* USER CODE BEGIN PV */
//////////////////////     Variables pour le Lidar          //////////////
uint8_t Data_RX_LIDAR;
uint16_t Data_Lidar_mm[360];
uint8_t drapeau_fin_tour = 0;
uint16_t data_lidar_mm_main[360];
…

int main(void)
{
       uint8_t drapeau_fin_tour_old = 0;    // Variable pour le Lidar
       uint32_t i = 0;
 (...)

  /// Initialisation du Lidar ///
  HAL_UART_Receive_IT(&huart1, &Data_RX_LIDAR, 1);

  while (1)
  {
  //Recopie du tableau lidar en fin de tour                                    ///////
        if((drapeau_fin_tour == 1) && (drapeau_fin_tour_old == 0))
              {
                for (i=0;i<360;i++)
                        data_lidar_mm_main[i]=Data_Lidar_mm[i];
              }
        drapeau_fin_tour_old = drapeau_fin_tour;
  }
}
```

La fonction d'interruption, avec une restriction aux trames de qualité maximale 15 (tableau_trame[0] = 0x3d pour la trame de début de tour et 0x3e pour les autres) :

```
/* USER CODE BEGIN 4 */
void HAL_UART_RxCpltCallback(UART_HandleTypeDef *huart) {
        static uint32_t index = 0;
        static uint8_t drapeau_demarrage = 0; // Variable indiquant pour réception des 7 1er octets
        static uint8_t tableau_trame[7]; // Variable de stockage des 7 premiers octets
        uint16_t angle;
        uint16_t distance;
          ////////////////////////////////////////////////////////////////////////////
          /////// Réception des sept octets correspondant au descriptor     ////////////
          ////////////////////////////////////////////////////////////////////////////
        if(drapeau_demarrage == 0){
                tableau_trame[index]=Data_RX_LIDAR;
                index++;
        }
        if ((index == 7) && (drapeau_demarrage == 0)) {
                index = 0;
                drapeau_demarrage = 1;                          // Fin des sept octets du descriptor
        }

          ////////////////////////////////////////////////////////////////////////////
          /////// Réception des données par groupe de 5 octets              /////////
          /////// Qualité (1 octet), Angle (2 octets) et Distance (2 octets)   /////////
          ////////////////////////////////////////////////////////////////////////////
        if (drapeau_demarrage == 1) {
                if (index <= 4) {
                        if ((Data_RX_LIDAR == 0x3e)||(Data_RX_LIDAR == 0x3d)) //début de trame de qualité 15
                                index = 0;
                        tableau_trame[index] = Data_RX_LIDAR;
                        if(tableau_trame[0] > 0x02)
                                index++;
                }

                  ////////////////////////////////////////////////////////
                  /////// Traitement de donnée pour convertir           /////////
                  /////// les octets reçus en données réelles       /////////
                  ////////////////////////////////////////////////////////
                if (index >= 5) {
                        index = 0;
                        angle = ((uint16_t) tableau_trame[2] << 1)
                                        + ((uint16_t) tableau_trame[1] >> 7);
                        distance = ((uint16_t) tableau_trame[4] << 6)
                                        + ((uint16_t) tableau_trame[3] >> 2);

                        // Stockage de la dist. mesurée dans le tableau à l'indice associé à l'angle
                        if (angle == 0)
                                Data_Lidar_mm[0] = distance;
                        if ((angle < 360) && (angle>0))
                                Data_Lidar_mm[360-angle] = distance;

                          ////////////////////////////////////////////////////////
                          ///////  Détection de la fin d'un tour            /////////
                          ////////////////////////////////////////////////////////



                if ((angle > 100) && (angle < 180) && (drapeau_fin_tour == 0)) {
                                drapeau_fin_tour = 1;
                        }
                        if ((angle > 270) && (angle < 360) && (drapeau_fin_tour == 1)) {
                                drapeau_fin_tour = 0;
                        }
                }
        }
        ///// Ré-activation de l'interruption UART RX
        HAL_UART_Receive_IT(&huart1, &Data_RX_LIDAR, 1);
}
/* USER CODE END 4 */
```

La figure 11 montre le programme de mise en œuvre du Lidar en cours d'exécution en mode debugage avec la visualisation en direct de la variable *Data_Lidar_mm*.

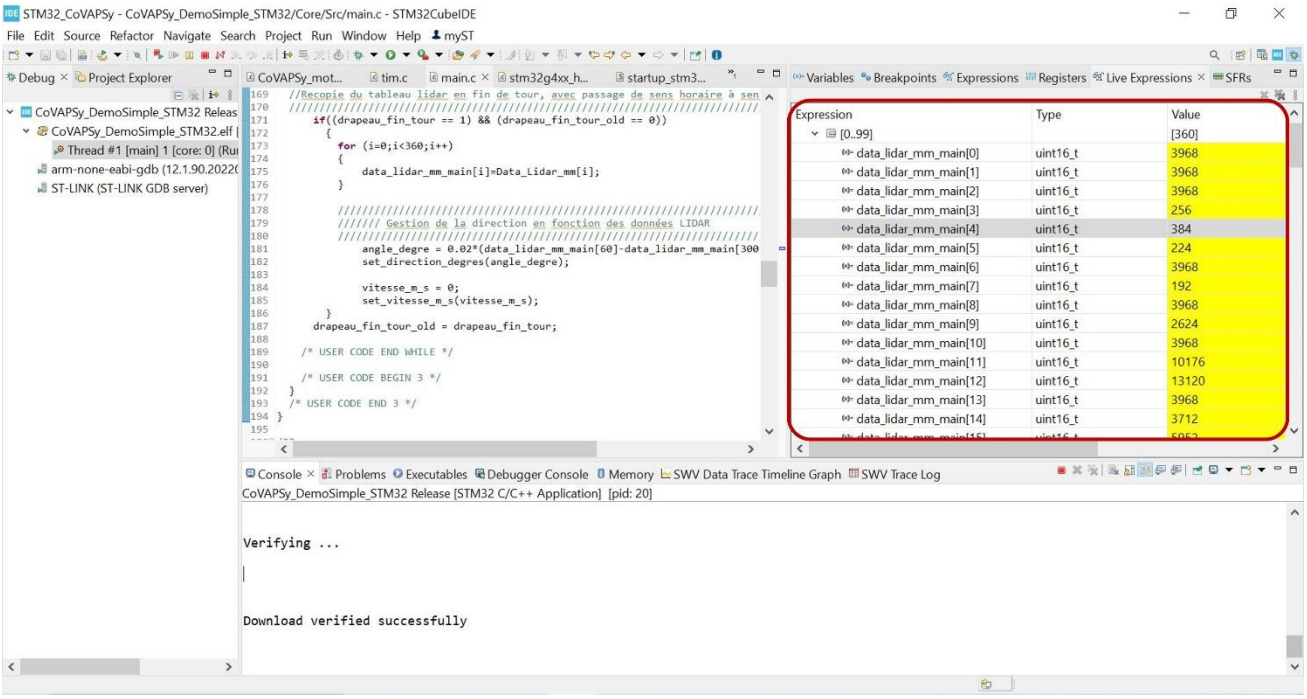


Figure 11 : Programme de mise en œuvre du Lidar avec visualisation de la variable Data_Lidar_mm

Il peut être nécessaire d'effectuer un reset hardware du microcontrôleur ou un ON/OFF général de la voiture autonome pour un bon fonctionnement du Lidar, celui-ci hésitant à redémarrer une connexion suite à une interruption de la connexion précédente.

2.3 - Moteur de propulsion

Le moteur de propulsion est géré par un signe de type PWM associé au timer 1 et channel 4. La fréquence de l'horloge du timer est fixé à 1MHz et le compteur est limité à 20 001. La configuration du timer est précisé dans la figure ci-dessous.

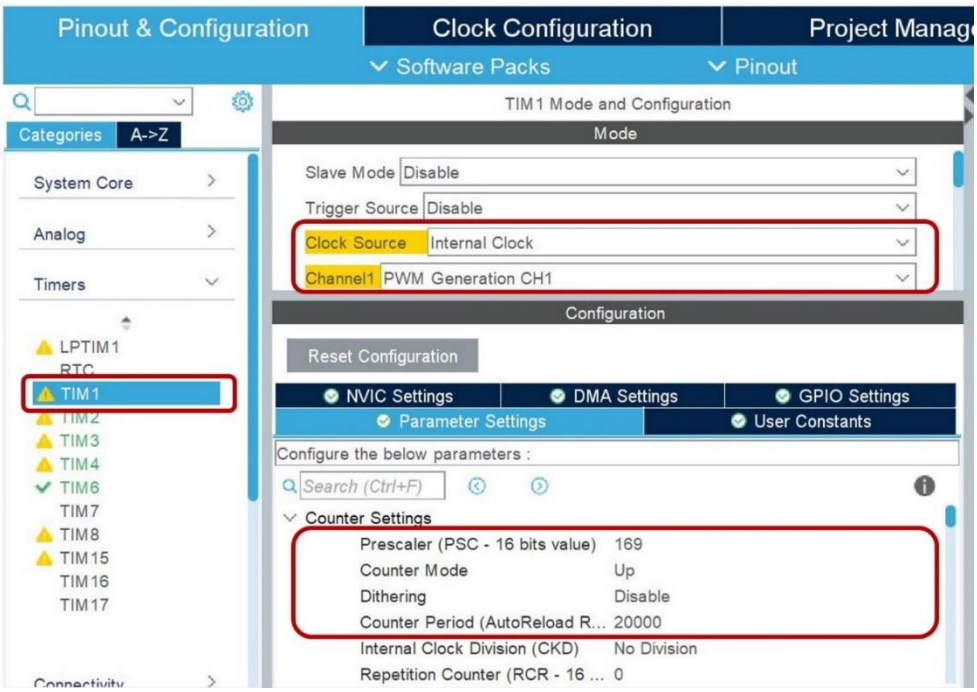


Figure 12 : Configuration du timer 1 associé à la propulsion

La librairie *CoVAPSy_moteurs* comporte des fonctions permettant de commander le variateur de vitesse et donc la propulsion de la voiture. Parmi les fonctions de la librairie, la fonction *Propulsion_init*() permet d'initialiser la commande du moteur de propulsion. La fonction

set_vitesse_m_s(vitesse_m_s) permet de commander le moteur de propulsion à une vitesse fixée en m/s. Le code ci-dessous est en exemple pour initialiser la commande du moteur puis le commander avec une vitesse approximative de +2.5m/s.

```
/* USER CODE END Header */
/* Includes ------------------------------------------------------------------*/
#include "main.h"
#include "tim.h"
#include "usart.h"
#include "gpio.h"

/* Private includes ----------------------------------------------------------*/
/* USER CODE BEGIN Includes */
#include "CoVAPSy_moteurs.h"
/* USER CODE END Includes */

( …)

/* Private function prototypes -----------------------------------------------*/
void SystemClock_Config(void);
 (…)
int main(void)
{
  (...)

  /* USER CODE BEGIN 2 */
  Propulsion_init();            // Initialisation de la propulsion
  set_vitesse_m_s(2.5);         // Propulsion fixée à 2.5m/s
  /* USER CODE END 2 */
  /* Infinite loop */
  /* USER CODE BEGIN WHILE */
  while (1)
  {
    /* USER CODE END WHILE */
    /* USER CODE BEGIN 3 */
  }
  /* USER CODE END 3 */
}
```

Le code ci-dessous détaille les fonctions *Propulsion_init*() et *set_vitesse_m_s()*.

```
void Propulsion_init(void){
        HAL_TIM_PWM_Start(&htim1, TIM_CHANNEL_1);
}



void set_vitesse_m_s(float vitesse_m_s){
        uint32_t largeur_impulsion_us;
        if (vitesse_m_s == 0)
        {
                largeur_impulsion_us = PROP_REPOS ;
        }
        else if (vitesse_m_s < 0){
                if(vitesse_m_s < -V_MAX_HARD)
                        vitesse_m_s  = -V_MAX_HARD;
                largeur_impulsion_us = PROP_POINT_MORT_NEG + (PROP_MAX - PROP_POINT_MORT) * vitesse_m_s /
V_MAX_HARD;
//version variateur bizarre
//largeur_impulsion_us = PROP_POINT_MORT - (PROP_MAX - PROP_POINT_MORT) * vitesse_m_s / V_MAX_HARD;
        }
        else if (vitesse_m_s > 0){
                if (vitesse_m_s > V_MAX_SOFT)
                        vitesse_m_s = V_MAX_SOFT;
                largeur_impulsion_us = PROP_POINT_MORT + (PROP_MAX - PROP_POINT_MORT) * vitesse_m_s /
V_MAX_HARD;
//version variateur bizarre
//largeur_impulsion_us = PROP_POINT_MORT_NEG - (PROP_MAX - PROP_POINT_MORT) * vitesse_m_s / V_MAX_HARD;
        }
        __HAL_TIM_SET_COMPARE(&htim1, TIM_CHANNEL_1, largeur_impulsion_us);
}
```

Il est important de noter que la vitesse maximale de la voiture est fixée à 8 m/s dû aux caractéristiques techniques du variateur. A noter que même Tamiya livrant ses voitures avec des variateurs de référence différente, il a été constaté que certains variateurs devaient se commander différemment des autres (l'impulsion est entre 1 et 1,5 ms pour la marche avant et entre 1,5 et 2 ms pour la marche arrière). Ainsi, un rapport cyclique de 1,8 ms peut faire tourner la voiture en marche avant ou en marche arrière ! C'est pour ces variateurs que le code de la bibliothèque *CoVAPSy_moteurs* comporte des lignes de code en commentaire. Il est donc important de prendre le temps de bien régler les paramètres de sa bibliothèque *CoVAPSy_moteurs*.

Dans la bibliothèque *CoVAPSy_moteurs*, La fonction *recule()* permet de faire reculer la voiture en cas de rencontre d'un obstacle.

2.4 - Servomoteur de direction

Le servomoteur de direction est géré par un signal de type PWM dont les caractéristiques sont les mêmes que celle de la direction. En effet, le même timer est utilisé pour générer la PWM de la direction et de la propulsion. La configuration de ce timer est précisé dans la partie associé à la propulsion.

La librairie *CoVAPSy_moteurs* comporte des fonctions permettant de commander le servomoteur de direction. Parmi les fonctions de la librairie, la fonction *init_direction*() permet d'initialiser la commande du moteur de propulsion. La fonction *set_direction_degres(angle_degre)* permet de commander le servomoteur de direction à un angle de direction défini en argument de cette fonction. Le code ci-dessous est en exemple pour commander le servomoteur de direction avec un angle de +10°.

```
/* USER CODE END Header */
/* Includes ------------------------------------------------------------------*/
#include "main.h"
#include "tim.h"
#include "usart.h"
#include "gpio.h"

/* Private includes ----------------------------------------------------------*/
/* USER CODE BEGIN Includes */
#include "CoVAPSy_moteurs.h"
/* USER CODE END Includes */

(...)

/* Private function prototypes -----------------------------------------------*/
void SystemClock_Config(void);
 (…)
int main(void)
{
  (...)

  /* USER CODE BEGIN 2 */
  Direction_init();             // Initialisation de la direction
  set_direction_degres(10);     // Direction fixée à 10°
  /* USER CODE END 2 */
  /* Infinite loop */
  /* USER CODE BEGIN WHILE */
  while (1)
  {
    /* USER CODE END WHILE */
    /* USER CODE BEGIN 3 */
  }
  /* USER CODE END 3 */
}
```

Le code ci-dessous détaille les fonctions *Direction_init*() et *set_direction_degres()*.

```
void Direction_init(void){
        HAL_TIM_PWM_Start(&htim1, TIM_CHANNEL_4);
}
```

```
void set_direction_degres(float angle_degre)
{
        uint32_t largeur_impulsion_us;
        if (angle_degre < -DIR_ANGLE_MAX)
                angle_degre = -DIR_ANGLE_MAX;
        else if (angle_degre > DIR_ANGLE_MAX)
                angle_degre = +DIR_ANGLE_MAX;
        largeur_impulsion_us = DIR_MILIEU + (DIR_BUTEE_GAUCHE - DIR_BUTEE_DROITE)*angle_de-
gre/(2*DIR_ANGLE_MAX);
        __HAL_TIM_SET_COMPARE(&htim1, TIM_CHANNEL_4, largeur_impulsion_us);

}
```

Attention, suivant les servo-moteurs, un rapport cyclique de 1,8 ms peut faire tourner la direction vers la gauche ou vers la droite ! Il est donc important de prendre le temps de bien régler les paramètres de sa bibliothèque *CoVAPSy_moteurs*. De plus, il est important de signaler que l'architecture mécanique de la voiture limite l'angle de braquage entre -18° et +18°.

3 - Exemple de code complet

Le programme de démonstration permet de mettre en œuvre l'ensemble des périphériques présentés précédemment. Il permet de :

- Faire tourner les roues à deux angles de directions différents (+10° puis -10°),
- Faire avancer la voiture à différentes vitesses (1m/s, 2m/s puis 3m/s),
- Faire reculer la voiture,
- Réceptionner les données du Lidar,
- Gérer la direction en fonction des données du Lidar.

Dans ce programme de démonstration, la propulsion est fixée à une vitesse lente de 0.5 m/s. La gestion de la direction dépend de la mesure des obstacles à +60° et -60°. En fonction de ces deux mesures de distance, le programme modifie l'angle de direction de la voiture.

Le programme complet de démonstration est disponible sur le dépôt github et présenté ci-dessous.

```
/* USER CODE END Header */
/* Includes ------------------------------------------------------------------*/
#include "main.h"
#include "tim.h"
#include "usart.h"
#include "gpio.h"

/* Private includes ----------------------------------------------------------*/
/* USER CODE BEGIN Includes */
//// Ajout des bibliothèques CoVAPSy_moteur et CoVAPSy_Lidar
#include "CoVAPSy_moteurs.h"
#include "CoVAPSy_Lidar.h"

/* USER CODE END Includes */

(...)
/* USER CODE BEGIN PV */
////////////////////////////////////////////////////////////////////////
//////////////////////     Variables pour le Lidar          //////////////
////////////////////////////////////////////////////////////////////////
uint8_t Data_RX_LIDAR;
uint16_t Data_Lidar_mm[360];
uint8_t drapeau_fin_tour = 0;
uint16_t data_lidar_mm_main[360];

////////////////////////////////////////////////////////////////////////
/////////     Variables pour la propulsion et direction          ////////
////////////////////////////////////////////////////////////////////////
float angle_degre,vitesse_m_s;

/* USER CODE END PV */

/* Private function prototypes -----------------------------------------------*/
```

```
/* USER CODE BEGIN PFP */
void SystemClock_Config(void);
/* USER CODE END PFP */
(...)
/**
  * @brief  The application entry point.
  * @retval int
  */
int main(void)
{
  (...)
  /* USER CODE BEGIN 2 */
    ////////////////////////////////////////////////////////////////////
  ///////// Initialisation de la propulsion et de la direction ///////
  ////////////////////////////////////////////////////////////////////
  Propulsion_init();
  Direction_init();
  /* USER CODE END 2 */

  /* Infinite loop */
  /* USER CODE BEGIN WHILE */

  ////////////////////////////////////////////////////////////////////
  ////////// Commande Direction à différents angles //////////////////
  ////////////////////////////////////////////////////////////////////
  set_direction_degres(0);
  HAL_Delay(500);
  set_direction_degres(-10);
  HAL_Delay(500);
  set_direction_degres(10);
  HAL_Delay(500);
  set_direction_degres(0);
  HAL_Delay(500);



  ////////////////////////////////////////////////////////////////////
  ////////// Commande Propulsion à différentes vitesses //////////////
  ////////////////////////////////////////////////////////////////////
  set_vitesse_m_s(1.0);
  HAL_Delay(1000);
  set_vitesse_m_s(2.0);
  HAL_Delay(1000);
  set_vitesse_m_s(3.0);
  HAL_Delay(1000);
  set_vitesse_m_s(0.0);
  HAL_Delay(1000);
  recule();
  HAL_Delay(1000);
  set_vitesse_m_s(0.0);
  HAL_Delay(1000);

  ////////////////////////////////////////////////////////////////////
  ////////////////////// Initialisation du Lidar ////////////////////
  ////////////////////////////////////////////////////////////////////
  Lidar_init();


  while (1)
  {
  ////////////////////////////////////////////////////////////////////////////////////////
  //Recopie du tableau lidar en fin de tour, avec passage de sens horaire à sens trigo//////
  ////////////////////////////////////////////////////////////////////////////////////////
          if((drapeau_fin_tour == 1) && (drapeau_fin_tour_old == 0))
                {
                  for (i=0;i<360;i++)
                  {
                          data_lidar_mm_main[i]=Data_Lidar_mm[i];
                  }
                  ////////////////////////////////////////////////////////////////////////
                  /////// Gestion de la direction en fonction des données LIDAR
                  ////////////////////////////////////////////////////////////////////////
                  //distance à 60° - distance à -60°
                  angle_degre = 0.02*(data_lidar_mm_main[60]-data_lidar_mm_main[300]);
                  set_direction_degres(angle_degre);
                  vitesse_m_s = 0.5;
                  set_vitesse_m_s(vitesse_m_s);
```

```
                }
            drapeau_fin_tour_old = drapeau_fin_tour;
    /* USER CODE END WHILE */
    /* USER CODE BEGIN 3 */
  }
  /* USER CODE END 3 */
}
```

4 - Axes d'amélioration

Outre le travail sur les algorithmes de conduite autonome, il existe plusieurs axes d'améliorations des performances de la voiture commandée par un microcontrôleur STM32. Tout d'abord la mise en place d'un asservissement numérique de vitesse permettrait de mieux contrôler la vitesse de la voiture. En remplaçant le servomoteur analogique par un servomoteur numérique de type Dynamixel XL430-W250-T, on gagne en dynamique et on peut lire la position réelle de la direction.

Actuellement, la voiture autonome *CoVAPSy_STM32only* peut présenter un défaut de réactivité à l'approche d'un obstacle. En effet, le Lidar envoie beaucoup de donnée au microcontrôleur ce qui engendre une forte activité pour cette tâche. Il pourrait être intéressant de mieux gérer ce flux de donnée. Un microcontrôleur plus rapide STM32H7 résoudrait le problème également.

5 - Conclusion

Cette ressource présente une voiture autonome *CoVAPSy_STM32only* gérée entièrement par un seul microcontrôleur programmé en C.

Les bibliothèques et programmes proposés permettent de mettre en œuvre rapidement les périphériques principaux afin de laisser place à la créativité pour la partie stratégie de course.

Cette voiture autonome destinée principalement aux étudiants d'école d'ingénieur ou d'IUT peut également être un support d'étude de projet pour les étudiants de BTS CIEL ou les élèves de STI2D. Les possibilités pédagogiques sont nombreuses. Un futur article présentera un projet de projet pour le BTS CIEL option ER basé sur cette voiture.

Références :

[1]: Course Voitures Autonomes Paris Saclay (CoVAPSy) : Travaux pratiques autour des voitures autonomes, T. Boulanger, E. Délègue, K. Hoarau, A. Juton, https://eduscol.education.fr/sti/si-ens-paris-saclay/ressources_pedagogiques/covapsy-tp-autour-des-voitures-autonomes

[2]: CoVaPSy : Premiers programmes python sur la voiture réelle, T. Boulanger, E. Délègue, K. Hoarau, A. Juton, https://eduscol.education.fr/sti/si-ens-paris-saclay/ressources_pedagogiques/covapsy-premiers-programmes-python-sur-voiture-reelle

[3]: CoVaPSy : Mise en œuvre du Simulateur Webots, T. Boulanger, E. Délègue, K. Hoarau, A. Juton, https://eduscol.education.fr/sti/si-ens-paris-saclay/ressources_pedagogiques/covapsy-mise-en-oeuvre-du-simulateur-webots

Contrôle de température via une cellule Peltier

La Revue 3EI
Culture Sciences de l'Ingénieur

Édité le 25/05/2023

école normale supérieure paris—saclay

Rémi AL AJROUDI[1,3] - Sylvie JAUVERT[2] - Morgan ALMANZA[3]

[1] Université Paris-Saclay, IUT GEII Cachan, [2] Education Nationale, Lycée Saint-Louis, [3] Université Paris-Saclay, ENS Paris-Saclay

Cette ressource fait partie du N°111 de La Revue 3EI de janvier 2024.

Dans de nombreux processus, la température joue un rôle important. Pour les systèmes de petites tailles, les cellules Peltier permettent de contrôler la température d'un système avec précision et rapidité. A partir de kits de développement facilement accessibles, nous proposons de mettre en œuvre un système complet de régulation de température en allant de la chaîne d'énergie à la chaîne d'information. Ce système offre un excellent support de cours et de TP pour plusieurs disciplines comme l'électronique de puissance (pont en H), l'informatique industrielle (codage à virgule fixe, bus série, threads, SPI), l'automatique (linéarisation, correcteur PI), implémentation sur microcontrôleur et des aspects plus logiciels avec une interface graphique et la gestion des threads sous python ou encore la communication par socket sur réseau TCP/IP. Ce travail se positionne avant tout dans le cadre de notions de niveau Licence.

1 - Introduction

Le contrôle de température est nécessaire dans de nombreuses applications allant de la biologie à l'ingénierie. Aujourd'hui deux types de stratégies sont utilisées pour contrôler la température, soit un thermocryostat qui utilise un fluide réfrigérant, soit un système à base d'une cellule Peltier. Ces derniers permettent un contrôle précis et rapide de la température mais les puissances froides disponibles sont faibles autour de quelques dizaines de Watts. Des sociétés comme *Stanford Research* proposent des systèmes de régulation de température utilisant une cellule Peltier, par exemple le modèle PTC10 offre un contrôle au milliKelvin avec la capacité de piloter une cellule Peltier jusqu'à 1A – 50V DC.

Dans cet article, un système [1] de contrôle de température via une cellule Peltier est développé, sa résolution en température est de 10 mK et il est capable de piloter des cellules Peltier de 7A-50V. L'accès à des puissances importantes permet d'obtenir des variations rapides de température.

La Figure 1 montre une photo du système avec trois parties, la partie thermique avec le Peltier (à gauche), le système d'asservissement de la température sur microcontrôleur (en bas à gauche) et le système de supervision avec une interface utilisateur (à droite).

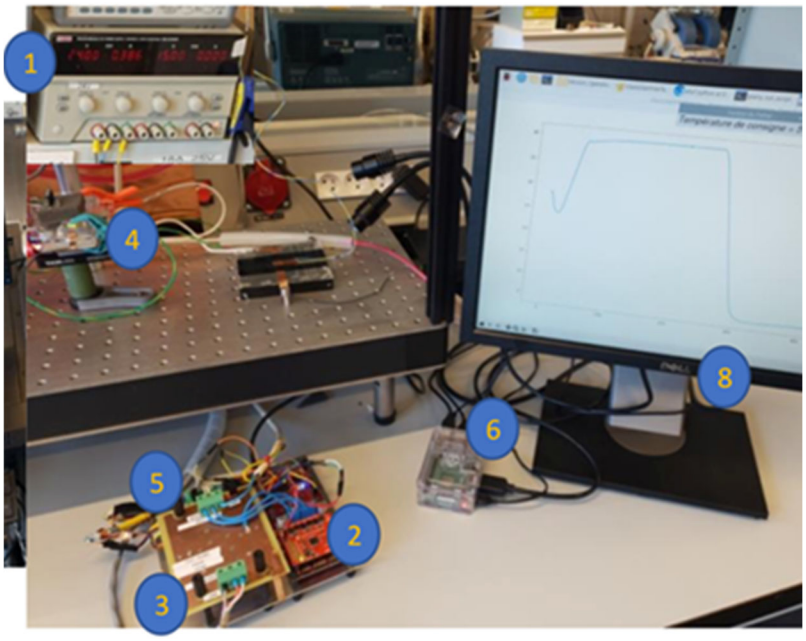


Figure 1 : Photo du système mis en œuvre avec en 1-l'alimentation, en 2-le microcontrôleur et le hacheur, en 3-le filtre LC associé au hacheur, en 4-la cellule Peltier, en 5 les conditionneurs des sondes de température PT100, en 6 et 8 le système de supervision

La Figure 2 donne le synoptique du système avec en 2- le module Hacheur (kit DRV8301 de TI, Texas Instrument) qui se clipse sur la carte de développement du microcontrôleur (F28379D de TI). Puisque le module hacheur dispose d'un abaisseur de tension, il est en mesure d'alimenter le microcontrôleur. Les deux points milieu des bras du module hacheur sont ensuite connectés à la cellule Peltier au travers d'un filtre LC. Ce filtre élimine l'ondulation due au découpage pour ne récupérer que la composante continue tout en évitant de possibles interférences électromagnétiques avec d'autres équipements. Pour mesurer la température, une sonde PT100 est connectée à un conditionneur de signaux (Max31865), celui transmet la température au microcontrôleur via un bus SPI. Après une phase d'initialisation des capteurs, le microcontrôleur exécute à pas de temps fixé (20ms) la lecture du capteur de température, scrute la consigne reçue sur le port série, transmet la température mesurée sur le port série et exécute la boucle de régulation qui agit sur le rapport cyclique du hacheur.

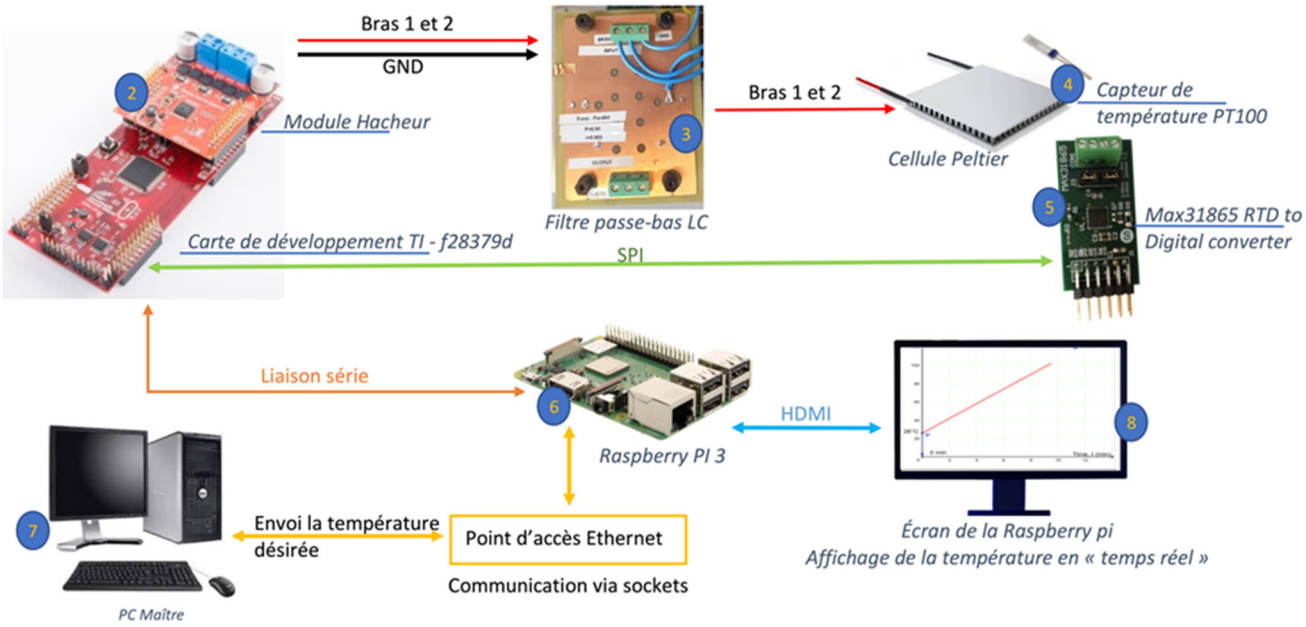


Figure 2 : Schéma synoptique où les numéros indiquent la correspondance avec la Figure 1, le 7-correspond au PC maitre

Ce système avec un pas de temps fixé est connecté à un système de supervision (Raspberry Pi), lequel dispose d'un affichage graphique et d'une interface utilisateur. Bien que son pas de temps soit souple, le recours à un système d'exploitation (Linux et des librairies Python) simplifie la réalisation d'interface graphique et la gestion de la communication client/serveur. Le système de supervision dialogue avec le microcontrôleur de la boucle de régulation via une communication série alors que la communication avec un PC maitre se fait via une communication par socket sur réseau Ethernet TCP/IP.

Ces illustrations ne montrent pas les éléments avec lesquels la cellule Peltier interagit. Sur une face, nous avons une température fixe qui est maintenue par flux d'eau généré par un thermocryostat, sur l'autre face il y a le bloc d'aluminium que nous devons réguler en température.

Dans la partie 2, nous décrivons la constitution et la modélisation du système, puis la modélisation de notre système, soit les modèles thermique et électrique de la cellule Peltier et du hacheur. Dans la partie III, nous décrirons l'asservissement de la cellule Peltier[3]. Enfin dans la partie IV, nous décrivons l'interface utilisateur permettant d'envoyer la température de consigne et d'observer en temps réel l'évolution de la température mesurée sur un écran.

2 - Éléments du système et modélisation

Cette partie établit un modèle thermique et électrique du système. Ce modèle nous permettra ensuite de définir la stratégie de contrôle à utiliser.

2.1 - Cellule Peltier et modèle thermique

La Figure 3 montre comment la cellule Peltier est intriquée dans le système thermique. Cet assemblage nécessite quelques précautions qui sont disponibles chez le fabricant de cellule Peltier[1]. Puisque le coefficient de performance d'une cellule Peltier est autour de 1, alors qu'il est de 6 pour des systèmes à compression détente de gaz, il est indispensable d'évacuer efficacement la chaleur de la source de chaude. Il faut donc refroidir cette face afin que la cellule Peltier ne se détériore pas. Ici nous avons recours à un échangeur à eau sur un thermo cryostat (ou un mini chiller), cette solution va nous permettre d'imposer la température de la source chaude, néanmoins un simple échangeur à eau ou à air aurait été suffisant. Le système étant destiné à refroidir un échantillon, les puissances moyennes nécessaires sont faibles.

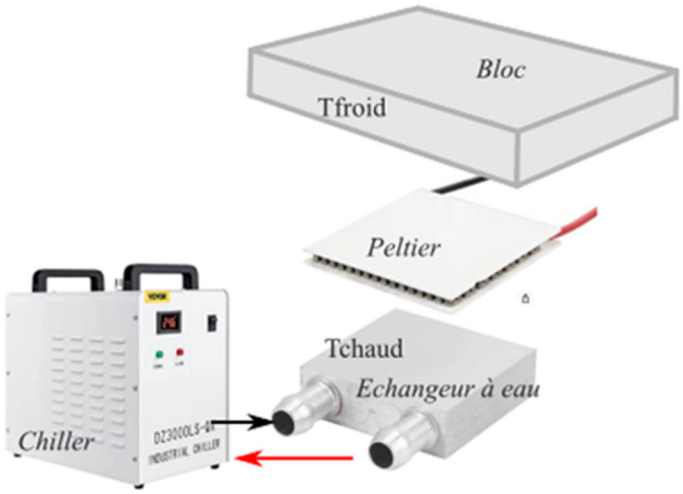


Figure 3 : Intégration de la cellule Peltier dans le système. La cellule est pressée à l'aide de deux vis entre le bloc où l'on souhaite contrôler la température et un échangeur à eau.

[1] https://tetech.com/wp-content/uploads/2013/11/tem_thermoelectric_module_mounting_procedure.pdf

Goupil (Goupil, Ouerdane, et Apertet 2013) propose un modèle d'une cellule Peltier, l'Éq. 1 décrit la partie thermique tandis que l'Éq. 2 est associée à la partie électrique. Pour cette dernière, nous avons trois effets, dans l'ordre de l'équation. L'effet thermoélectrique qui donne un flux de chaleur proportionnel au courant où P en W/A est le coefficient Peltier, un terme de dissipation liée à la résistance R et un terme de conduction thermique liée à la conductance thermique et à l'écart de température entre les deux faces notées T_{froid} et T_{chaud}. Le flux de chaleur prélevé du côté froid et celui donné du côté chaud, sont respectivement notés Φ_f et Φ_c. Il est important de distinguer la quantité de chaleur Q en Joule, du flux de chaleur Φ en Watt, qui est dQ/dt.

$$\Phi_f = P \cdot i - \frac{R}{2} \cdot i^2 - K \cdot (T_f - T_c)$$
$$\Phi_c = P \cdot i + \frac{R}{2} \cdot i^2 - K \cdot (T_f - T_c) \qquad \text{Éq. 1}$$

Du point de vue électrique, la cellule Peltier se comporte comme une résistance R en série avec une source de tension dont la tension est proportionnelle à la différence de température, où α est lié au coefficient Seebeck et au nombre de jonction thermoélectrique de la cellule. Le modèle électrique est en convention récepteur.

$$U = \alpha \Delta T + R \cdot i \qquad \text{Éq. 2}$$

La résistance R, 1.4 Ω, est directement donnée par le constructeur de la cellule Peltier. Les autres paramètres du modèle doivent être identifiés à partir des caractéristiques fournies. La Figure 4 et la Figure 5 permettent d'identifier la conductance K, soit environ 2.7 W/K et le coefficient Peltier, soit environ 25.5 W/A. La Figure 7 quant à elle permet de déterminer le coefficient α de l'Éq. 2, soit 0.075 V/°C.

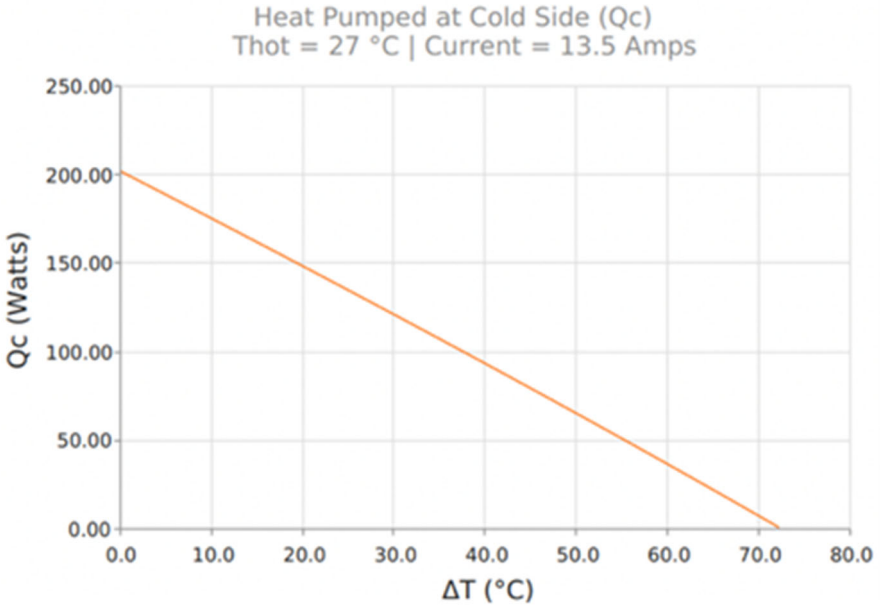


Figure 4 : Flux thermique Qc en fonction de la différence de température Th - Tc (Laid 387005665 extrait de la documentation)

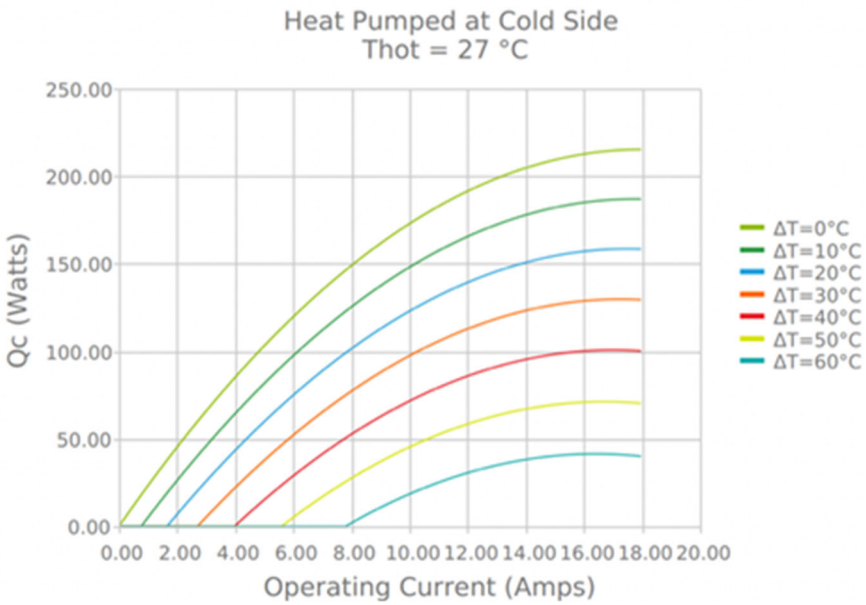


Figure 5 : Caractéristiques électrique et thermique de la cellule Peltier (Laid 387005665 extrait de la documentation)

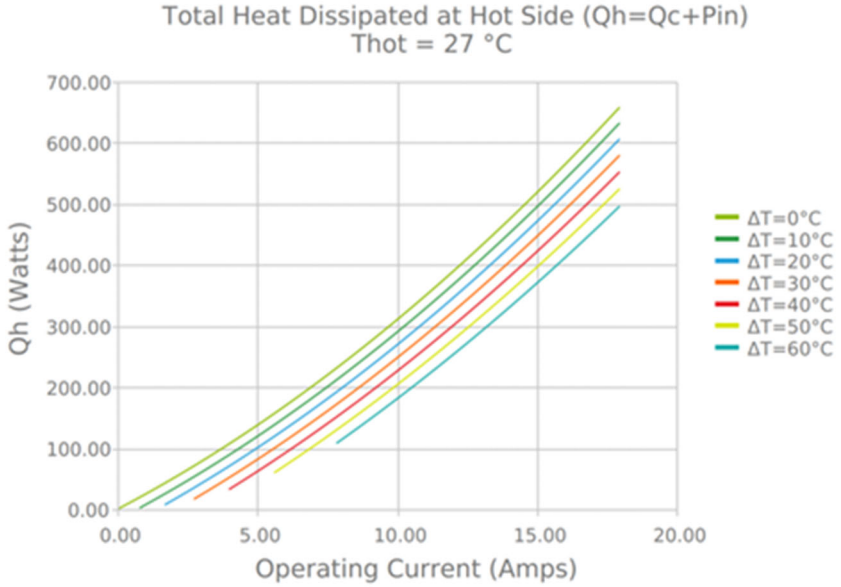


Figure 6 : Puissance dissipée sur la face chaude de la cellule Peltier (Laid 387005665 extrait de la documentation)

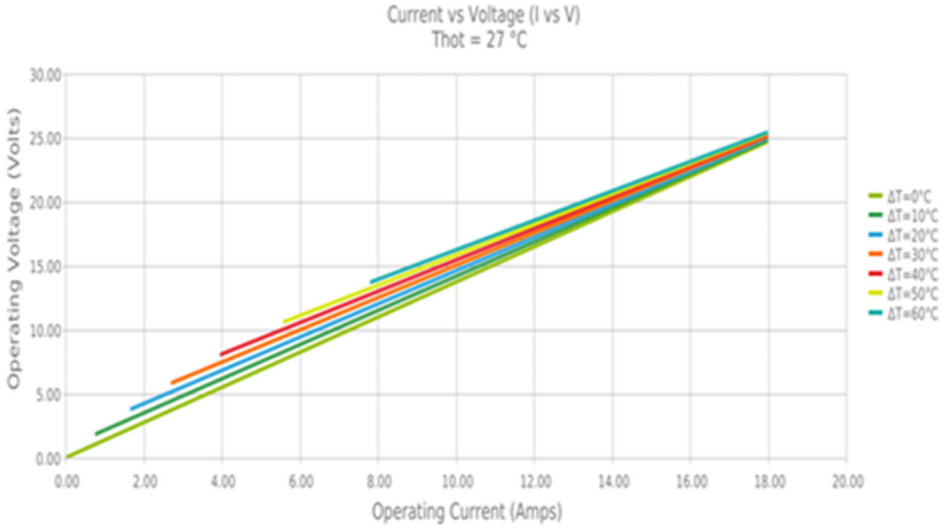


Figure 7 : Caractéristique électrique de la cellule Peltier (Laid 387005665 extrait de la documentation)

À partir d'un premier principe sur le bloc de masse m et de capacité thermique c_v, l'Éq. 3 donne l'évolution de la température du bloc à partir du flux de la cellule Peltier. On suppose ici qu'il n'y a pas de fuite et que la température du bloc est homogène.

$$c_p m \frac{dT_f}{dT} = \Phi_f \qquad \text{Éq. 3}$$

La température du côté chaud résulte de l'interaction entre la cellule qui tend à le chauffer et le fluide du thermocryostat (*Huber Minichiller 300 OLÉ*) qui tend à le refroidir. Dans le pire cas, un courant de 5 ampères et une différence de température entre la face froide et chaude de 30°C, la puissance dissipée sur la face chaude est d'environ 90 W. Le thermocryostat quant à lui est capable d'absorber 140 W pour une différence de température entre la face froide et chaude de 30°C. Donc même en imposant -5°C sur le côté chaud, le thermocryostat sera capable de maintenir une température constante. Dans cette modélisation on supposera que T_c est constant,

$$T_c = 20°C \qquad \text{Éq. 4}$$

2.2 - Alimentation à découpage et modèle électrique

La Figure 8 présente la structure à découpage utilisée pour alimenter la cellule Peltier. Par rapport à une structure linéaire, cette solution offre un bon rendement de conversion, néanmoins elle génère plus de perturbations électromagnétiques. Afin de limiter ces perturbations liées aux forts dv/dt engendrés par le découpage, le filtre est placé au plus près du pont en H. Afin de respecter les règles d'association des sources, un temps mort est introduit par le hardware du module PWM (pulse width modulation) du microcontrôleur. Différentes stratégies de modulation sont possibles, ici nous avons simplement pris une modulation bipolaire. Le circuit de puissance (Boost XL de Texas) que l'on enfiche sur la carte du microcontrôleur (F28379d) possède la commande rapprochée nécessaire aux contrôles des interrupteurs MOSFETs.

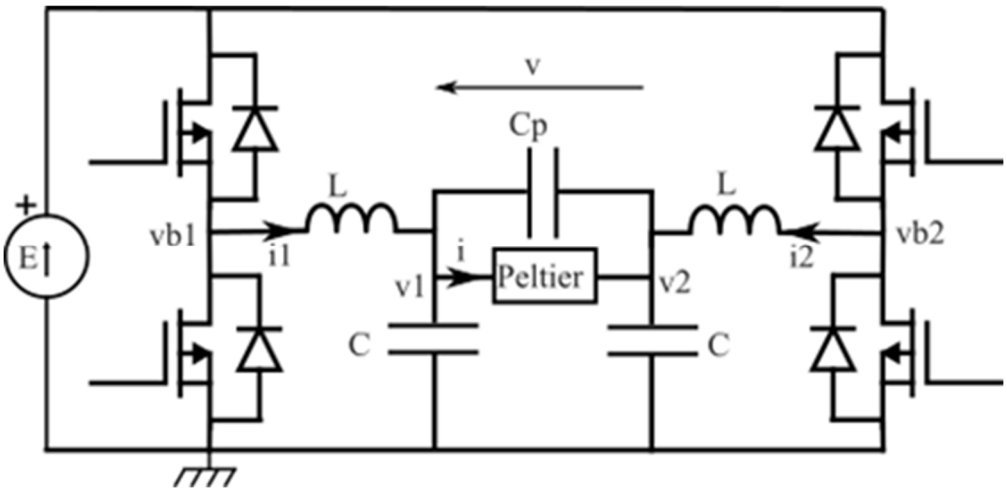


Figure 8 : Schéma électrique du pont H et du filtre placé sur la cellule Peltier avec $\boldsymbol{L = 100\ \mu H, C = 10\mu F\ et\ C_p = 100\mu F}$

En écrivant les relations électriques sur les éléments, on a :

$$L \cdot \frac{di_1}{dt} = V_{b1} - V_1$$
$$L \cdot \frac{di_2}{dt} = V_{b2} - V_2 \qquad \text{Éq. 5}$$

$$C \frac{dv_1}{dt} = i_1 - i_p - i_{cp} \qquad \text{Éq. 6}$$

$$C\frac{dv_2}{dt} = i_2 + i_p + i_{cp}$$
$$C_p\frac{d(v_1 - v_2)}{dt} = i_{cp}$$

À partir des Éq. 5 et Éq. 6, le système est représenté par sa composante en mode différentiel Éq. 7 et sa composante en mode commun Éq. 8.

$$\frac{d(v_1 - v_2)}{dt} = \frac{i_1 - i_2}{C + 2C_p} - 2i_p$$
$$\frac{d(i_1 - i_2)}{dt} = -\frac{v_1 - v_2}{L} + v_{b1} - v_{b2} \qquad \text{Éq. 7}$$

$$\frac{d(v_1 + v_2)}{dt} = \frac{i_1 + i_2}{C}$$
$$\frac{d(i_1 + i_2)}{dt} = -\frac{v_1 + v_2}{L} + v_{b1} + v_{b2} \qquad \text{Éq. 8}$$

Sur le mode commun $(v_1 + v_2)$ et le mode différentiel $(v_1 - v_2)$, il y a des éléments de filtrage, filtre LC, qui annihilent/réduisent les ondulations de tension liées aux découpages à 20 kHz. Le modèle s'intéresse donc à la valeur moyenne sur une période de découpage comme décrit dans l'Éq. 9 à partir de la fonction de modulation f_m que l'on décompose en f_m^0 et Δf_m.

$$< v_{b1} >= (f_m^0 + \Delta f_m)E \qquad \text{Éq. 9}$$
$$< v_{b2} >= (f_m^0 - \Delta f_m)E$$

Sur le mode commun, l'amortissement étant uniquement dû aux éléments parasites des éléments, le point de polarisation à 0.5, soit f_m^0, devra être positionné progressivement afin d'éviter des oscillations trop importantes. Le mode différentiel, soit la tension sur la cellule U, est piloté avec le paramètre Δf_m.

L'utilisation du kit de Texas Boost xl offre une plateforme fiable. On observe aussi que lors de la connexion de l'alimentation de la partie puissance, un fort appel de courant nécessaire à charger les condensateurs sur le point de fonctionnement, vient saturer l'alimentation.

2.3 - Mesure de la température et communication SPI

Afin d'asservir en température le système, il faut mesurer la température T_c, pour cela une sonde de type PT100 est utilisée. La température est déduite de la résistance d'un élément de platine, lequel est proportionnel à la température. La dépendance à la température étant faible (<1% par degré) un circuit électrique doit conditionner le signal, par exemple à l'aide d'un point de Wheatstone suivi d'un amplificateur d'instrumentation et d'un ADC. Ici, nous utilisons un circuit de conditionnent tout intégré comme le MAX31865. Le conditionneur et le microcontrôleur vont communiquer au travers du bus de donnée, ici un bus de type SPI « Serial Peripheral Interface ». Ce bus dispose de quatre fils clock, Master Output Slave Input, Master Input Slave Output, Chip Select, notés respectivement CLK, MISO, MOSI, CS. Dans un premier temps, le microcontrôleur (uC) transmet des données pour configurer le conditionneur, par exemple le uC transmet 8 bits par exemple 80h pour écrire sur le registre de configuration (Figure 9) puis l'octet à placer dans le registre.

REGISTER NAME	READ ADDRESS (HEX)	WRITE ADDRESS (HEX)	POR STATE
Configuration	00h	80h	00h
RTD MSBs	01h	—	00h
RTD LSBs	02h	—	00h
High Fault Threshold MSB	03h	83h	FFh
High Fault Threshold LSB	04h	84h	FFh
Low Fault Threshold MSB	05h	85h	00h
Low Fault Threshold LSB	06h	86h	00h
Fault Status	07h	—	00h

Figure 9 : Registre disponible du circuit de conditionnement de la PT100 (cf la documentation du MAX31865 pour plus de détail)

Après cette étape de configuration, toutes les 20 ms, le uC écrit sur le bus 01h pour lire l'octet RTD MSB, et juste après il écrit 02h puis lire l'octet RTD LSB

REGISTER	RTD MSBS (01h) REGISTER								RTD LSBS (02h) REGISTER							
Bit	D7	D6	D5	D4	D3	D2	D1	D0	D7	D6	D5	D4	D3	D2	D1	D0
RTD Resistance Data	MSB	—	—	—	—	—	—	—	—	—	—	—	—	—	LSB	Fault
Bit Weighting	2^{14}	2^{13}	2^{12}	2^{11}	2^{10}	2^{9}	2^{8}	2^{7}	2^{6}	2^{5}	2^{4}	2^{3}	2^{2}	2^{1}	2^{0}	—
Decimal Value	16384	8192	4096	2048	1024	512	256	128	64	32	16	8	4	2	1	—

Figure 10 : Registre qui stocke la mesure de température

À partir de ces deux octets et du codage de la température donnée dans la documentation, nous allons reconstruire la température sur un nombre codé en virgule flottante, c'est cette grandeur que nous allons appeler T_m et qui s'actualise toutes les 20 ms. Il y a donc un retard pur sur la mesure mais au vu de la dynamique du système, supérieur à 20 s, il ne sera pas nécessaire de prendre en compte ce retard pour le dimensionnement du correcteur.

Par ailleurs, nous disposons d'une autre sonde de température sur le côté chaud afin de faire un monitoring du système ou de détecter une défaillance, par exemple pour détecter que le thermo cryostat ne fonctionne pas.

2.4 - Le microcontrôleur

Le microcontrôleur interface les capteurs (communication SPI), pilote le hacheur et dialogue avec le superviseur (communication série). L'implémentation du code sur le microcontrôleur F28379D de Texas Instrument (TI) se fait en langage C dans l'outil de développement de TI, code composer studio. Néanmoins, le recours à des outils de génération de code automatique comme Embedded Coder de Matlab-Simulink permet de faciliter l'implémentation en se concentrant sur le correcteur plus que sur le code C. De plus grâce à des outils tierces de TI, « Embedded coder », nous avons directement accès à la configuration des périphériques du microcontrôleur.

3 - Asservissement

Le modèle du système et la mesure de la température vont permettre de mettre en place un asservissement.

3.1 - Linéarisation du modèle

Les équations de Éq. 1 à Éq. 4 et Éq. 6 forment un système multivariable d'ordre 3 non linéaire avec comme variables d'état T, $v_1 - v_2$, $i_1 - i_2$.

Le système pourrait nécessiter plusieurs correcteurs, soit dit autrement un retour d'état, afin de contrôler chaque variable d'état. Néanmoins la mise en place d'un correcteur PI sur la variable de la grandeur d'intérêt, la température, est plus appropriée pour le niveau visé. Le correcteur PI se place donc au niveau de l'écart entre la température de consigne et la température de mesure. Puisque qu'il n'y a pas de boucle de contrôle sur le courant, le taux de variation de la tension, soit de Δf_m, va être limité afin d'éviter la saturation en courant des inductances à 7.5A. En limitant le rate, les équations électriques sont dans un état quasi-stationnaire où $U = v_1 - v_2$, soit en moyenne $v_{b1} - v_{b2}$.

Du côté des non-linéarités associées au comportement de la cellule Peltier, nous proposons de linéariser le système car c'est l'approche habituelle pour aborder les systèmes non linéaires. Le système s'écrit, in fine, sous la forme de l'Éq. 10.

$$\dot{T}_f = f(T_f, U) \qquad \textit{Éq. 10}$$

Prenons une linéarisation autour de U = 0V et l'on décompose la température comme illustré à l' Éq. 11 avec T_f^0 le point polarisation et $\widetilde{T_f}$ la variation de température autour de ce point.

$$T_f = T_f^0 + \widetilde{T_f} \qquad \textit{Éq. 11}$$

Le point de polarisation étant pour $\boldsymbol{U = 0}$, la température de polarisation $\boldsymbol{T_f^0}$ est telle que $\boldsymbol{f(T_f^0, 0)}$ égale 0, soit $\boldsymbol{T_f^0 = T_c}$.

En linéarisant le système, Éq. 10, soit un système à deux variables autour du point de fonctionnement $\boldsymbol{T_f^0}$, on obtient le système linéaire de l'équation

$$\dot{\widetilde{T_F}} = 0 + \frac{\partial f(T_f^0, 0)}{\partial T_f} \cdot \widetilde{T_f} + \frac{\partial f(T_f^0, 0)}{\partial U} \cdot \widetilde{U} \qquad \textit{Éq. 12}$$

La mise en place de correcteur PI se faisant traditionnellement dans le domaine de Laplace, l'**Éq. 12** est réécrite sous la forme de Éq. 13, ici bien que les notations soient identiques, les fonctions $\widetilde{T_f}(p)$ et $\widetilde{U(p)}$ dépendent de la variable de Laplace, notée p.

$$\frac{\widetilde{T_F}(p)}{\widetilde{U}(p)} = \frac{\frac{\partial f}{\partial U}}{p - \frac{\partial f}{\partial T_F}} \qquad \textit{Éq. 13}$$

soit

$$\frac{\widetilde{T_F}(p)}{\widetilde{U}(p)} = \frac{\frac{P}{R}}{\frac{P \cdot \alpha}{R} + K} \Big/ \left(\frac{cm}{\left(\frac{P \cdot \alpha}{R} + K\right)} \cdot p + 1\right) \qquad \textit{Éq. 14}$$

Il s'agit donc d'un premier ordre qui peut se faire sous la forme classique de l'**Éq. 15** avec $\boldsymbol{\tau = 22\ s}$ et $\boldsymbol{\beta = 4,5\ °C/V}$.

$$\frac{\widetilde{T_F}(p)}{\widetilde{U}(p)} = \frac{\beta}{1 + \tau \cdot p} \qquad \textit{Éq. 15}$$

Pour valider le modèle, le système (en boucle ouverte) est soumis à plusieurs réponses indicielles selon le rapport cyclique des PWM. La Figure 11 montre l'évolution de la température en fonction du temps.

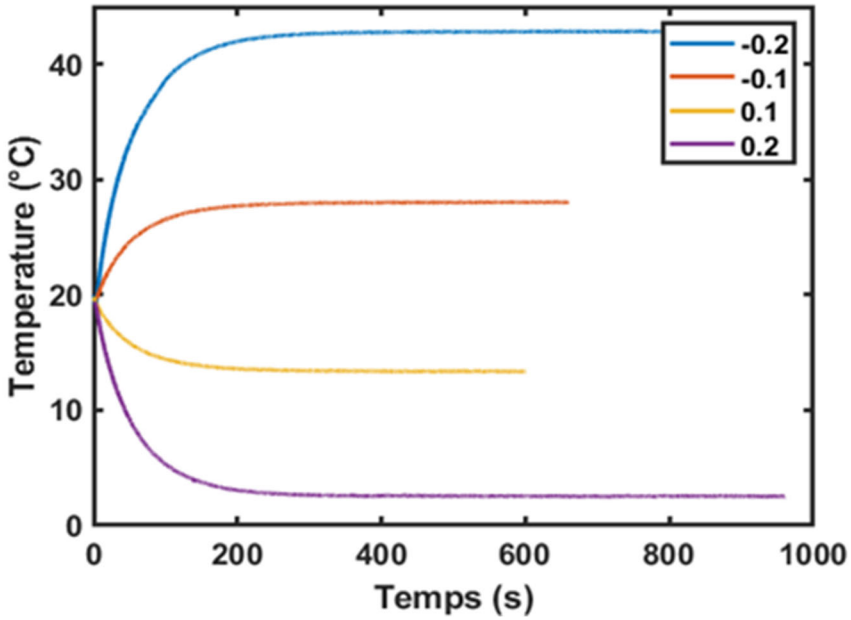


Figure 11 : Réponse indicielle du système pont en H et de la cellule Peltier mesurée. La légende indique le rapport cyclique choisi autour du point de polarisation, soit Δf_m. La tension d'alimentation du bus était autour de 12V (au final on utilisera 24V).

En première approximation, ce système s'apparente bien à un premier ordre. Néanmoins, les non-linéarités apparaissent car il y a une dissymétrie entre les rapports cycliques positifs et négatifs ($\pm\Delta f_m$.) et une non proportionnalité entre l'essai à 0.1 et 0.2 de rapport cyclique. On peut alors déterminer la constante de temps moyenne τ et le gain moyen β du système. Dans ce cas, on identifie $\boldsymbol{\tau = 40\ s}$ et $\boldsymbol{\beta = 5\ °C/V}$, sachant que toutes les courbes ne donnent pas exactement les mêmes paramètres à cause des non-linéarités du système. Disons que les paramètres identifiés sont représentatifs des zones de fonctionnement visées.

3.2 - Dimensionnement du correcteur - Correcteur Proportionnel Intégral

L'asservissement en boucle fermée avec le correcteur proportionnel intégral (PI), va permettre d'annuler l'erreur statique et d'augmenter la dynamique de réponse du système. La forme du PI est celle proposée à Éq. 16.

$$C(p) = K_i \cdot \left(1 + \frac{1}{\tau_i \cdot p}\right) \qquad \text{Éq. 16}$$

Le schéma bloc du système en boucle fermé est représenté dans la Figure 12.

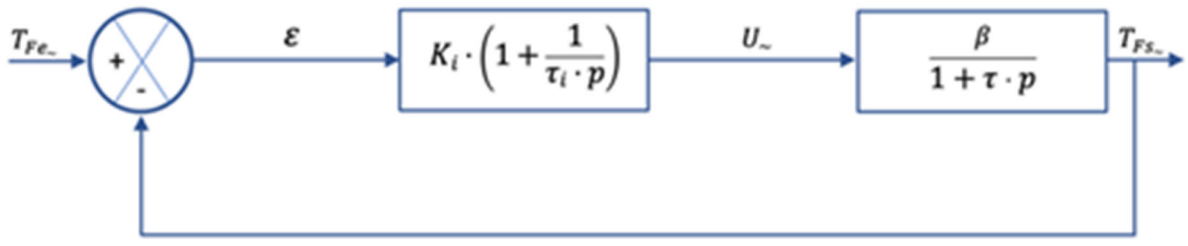


Figure 12 : Schéma bloc du système en boucle fermée avec T_{Fe} pour la température de consigne et T_{Fs} pour la température du système, soit T_F

Le correcteur est dimensionné de façon à compenser le pôle dominant, soit $\tau_i = \tau$. On règle ensuite la bande passante à l'aide du gain Ki. On rappelle que la bande passante de la FTBF correspond à la fréquence où la FTBO est égale à 0dB.

Puisque le retour de la boucle est unitaire, la fonction de transfert en boucle fermée s'exprime comme :

$$\frac{\widetilde{T_F}(p)}{\widetilde{T_{F_e}}(p)} = \frac{1}{1 + \frac{\tau \cdot p}{\beta \cdot K_i}} \qquad \text{Éq. 22}$$

La constante de temps du système en boucle fermée est $\tau_c = \frac{\tau}{\beta \cdot K_i}$, si l'on souhaite augmenter légèrement la dynamique du système, prenons $\tau_c = \frac{\tau}{5}$, on obtient $K_i = \frac{5}{\beta} = 0.06$ et $\frac{K_i}{\tau_i} = 0.001$.

4 - Implémentation « Embedded Coder » de Matlab-Simulink

Le microcontrôleur est programmé sous Matlab-Simulink à l'aide d'outil de génération de code C tel que « embedded coder ». Les codes Matlab-Simulink étant partagés, nous détaillons uniquement la structure du code. Le code se décompose en plusieurs parties : 1) la partie communication avec le capteur de température (communication SPI) ; 2) la partie réception/transmission des données avec le superviseur (communication série) ; 3) le correcteur avec la gestion des modules PWM du microcontrôleur ainsi que d'un affichage/contrôle de l'état du système à l'aide de LEDs/interrupteur.

4.1 - Communication avec le capteur de température (communication SPI)

La Figure 13 décrit la communication avec le capteur de température. Il y a une phase d'initialisation du capteur puis une phase de lecture répétitive de la température.

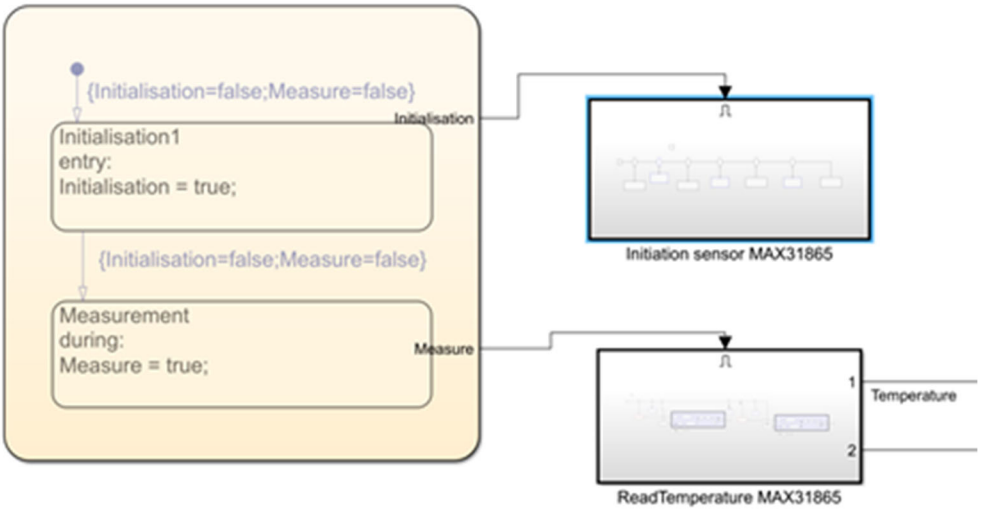


Figure 13 : Initialisation du capteur de température et lecture de la température au travers d'une machine d'état ("state flow chart") activée au pas de temps du système (20 ms). Les blocs « initialization » et « read » exécutent la communication SPI du microcontrôleur pour configurer le capteur puis pour lire la température.

4.2 - Réception/transmission des données avec le superviseur (communication série)

La Figure 14 décrit la transmission sur communication série des différentes températures. Un codage à virgule fixe (fixed point number) a été utilisé, 8 bits pour l'entier signé et 8 bits pour la

virgule fixe. On fait attention dans le bloc « convert » de Matlab- Simulink, soit pour maintenir la valeur représentée, soit son mot binaire.

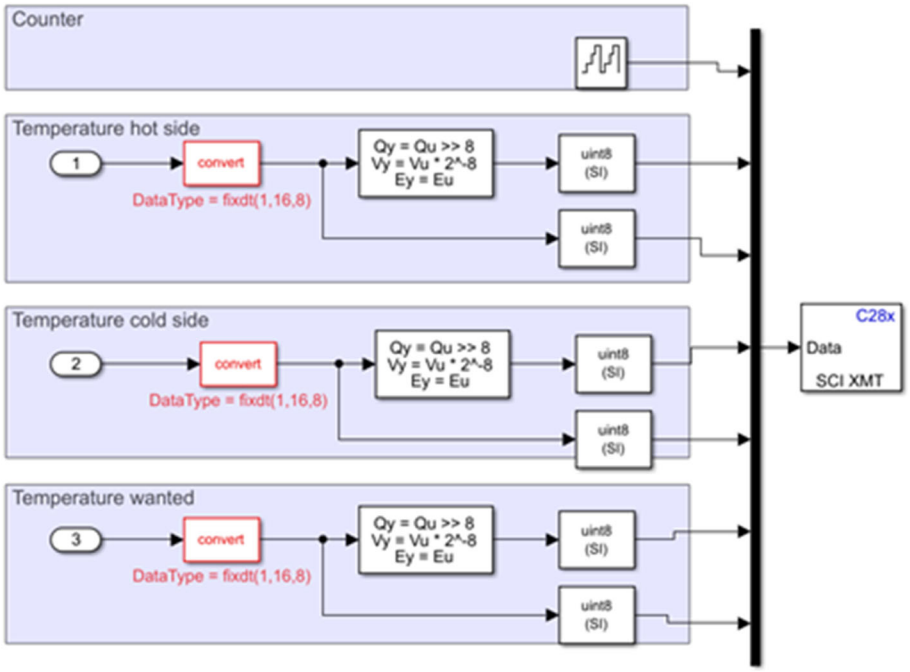


Figure 14 : Communication des températures au travers du port série « SCI », la trame série se compose de série d'octets qui sont un compteur, puis 2 octets pour chaque température. La température est transmise en virgule fixe avec un octet pour la partie entière et un octet pour la partie fractionnaire.

La Figure 15 correspond à la température de consigne reçue par le microcontrôleur via le port série. La transmission se fait via 4 octets, pour le moment seul deux octets sont exploités pour coder la température (8 bits pour l'entier signé et 8 bits pour la partie fractionnaire). Le microcontrôleur via le code en Figure 15, reconstitue la consigne de température.

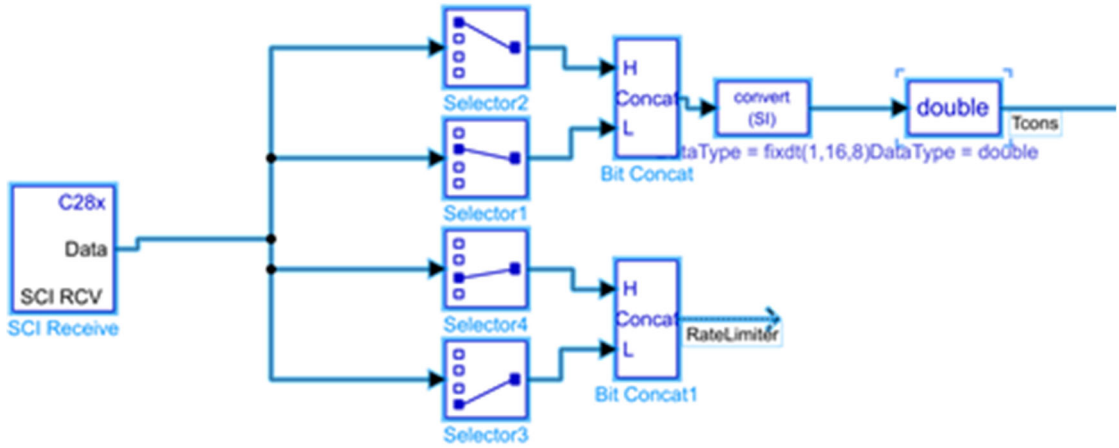


Figure 15 : Code simulink pour reconstituer la consigne de température via les octets reçus sur le port série

La partie entière de la température, codée comme un entier signé sur 8 bits, donne une plage de -128.000°C à 127,996°C, et la partie fractionnaire donne une résolution de 1/2^8, soit 0.0039 °C. Le codage choisi est donc adapté à notre plage de travail ainsi qu'à la résolution du convertisseur de température de 0.01 °C (MAX31865). Néanmoins dans l'ensemble des programmes sur microcontrôleur et sur RaspBerry nous travaillerons en flottant, seule la communication série se fera en virgule fixe.

4.3 - Correcteur et gestion des entrées/sorties

La Figure 16 intègre le correcteur PI avec la consigne (en haut à gauche) et la mesure (au niveau du rond vert). La partie à droite est liée aux PWM pour le contrôle de la partie puissance, tandis que la partie en bas à gauche est liée à la partie affichage ou interrupteur de démarrage. L'effet intégral du correcteur est actif uniquement lorsque la partie puissance n'est pas inhibée. Il y a aussi une protection basée sur la température de l'échangeur à eau, en effet si celui-ci devient trop chaud alors la partie puissance est désactivée et la LED rouge s'allume.

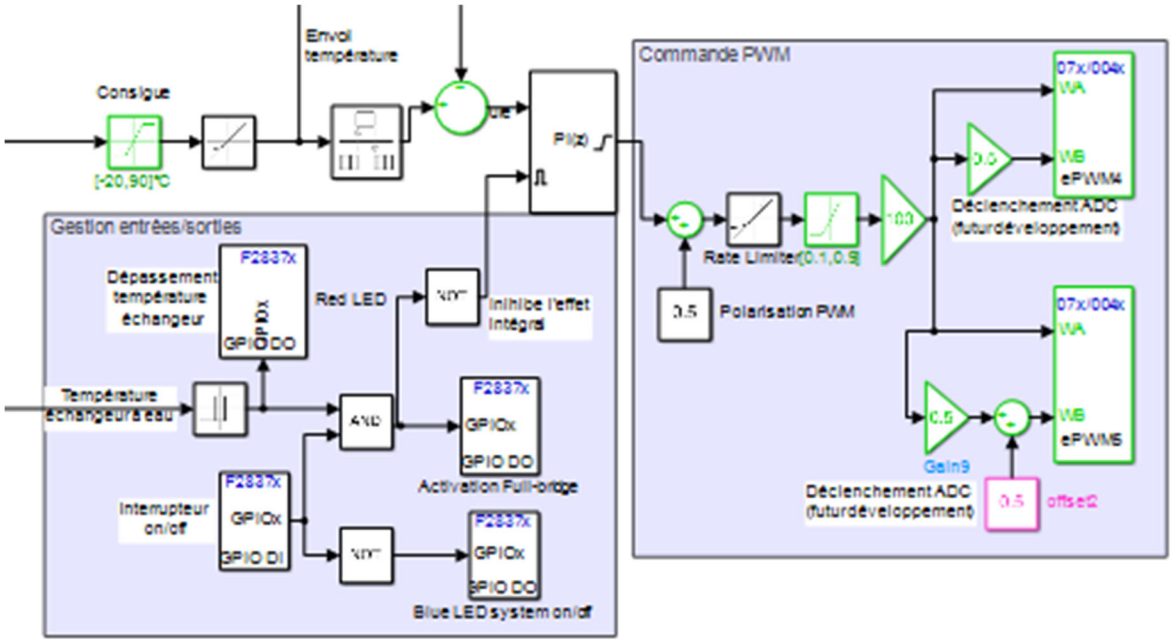


Figure 16 : Correcteur avec la consigne, la mesure, la commande et la gestion des entrées/sorties (LEDs, interrupteur)

5 - Interface utilisateur

L'interface utilisateur hébergée sur le module RaspBerry permet de superviser le système, soit de :

- Visualiser en temps réel l'évolution des températures sur un écran connecté en HDMI ;
- Il peut agir sur la consigne mais dans le fonctionnement normal il ne fait que retransmettre la consigne qu'il a lui-même reçu de l'ordinateur maître via un socket TCP/IP.

Les librairies utilisées pour se faire sont :

- Tkinter pour la création de l'interface ;
- Sockets pour la communication avec le PC maître via un protocole client-serveur sur sockets TCP/IP ;
- PySerial pour la communication avec le microcontrôleur via une liaison série sur USB ;
- Fixed Point pour la conversion des nombres de virgule fixe à virgule flottante ;
- Matplotlib pour afficher la courbe.

Le tableau, ci-dessous, illustre simplement l'algorithme de l'interface utilisateur.

server_program() ->Mise en place du serveur TCP/IP, réception de la consigne de température du PC (on utilisera un codage ascii)
plotTemperature() ->Réception de la consigne du uC, conversion des données de virgule fixe vers virgule flotante et affichage d'un graphique

sendTemperature() -> Saisie de la valeur ou transfert de celle reçue par le serveur TCP/IP et conversion de la température de consigne en virgule fixe
Lancement des Threads liés aux fonctions server_program(), plotTemperature() et sendTemperatuer().

Tableau 1 : Schéma illustrant le fonctionnement du programme

Les fonctions sont exécutées dans des boucles infinies, les threads sont ordonnancés selon le « Round Robbin » c'est-à-dire que chacun des threads se donne la parole à tour de rôle après un certain temps si l'exécution de la fonction du thread est plus longue que le temps imparti.

La mise en place de l'interface s'est faite avec Tkinter afin d'offrir une prise en main rapide.

Le programme est codé à partir des fonctions, l'utilisation d'objets permettrait une meilleure intégration du code mais demande aussi des notions de programmation plus avancées.

Les deux sous-parties suivantes introduisent brièvement la communication par socket et les conversions utilisées pour coder la température à chaque étape.

5.1 - Communication de la Raspberry avec le PC maître au travers de socket par réseau TCP/IP

Pour envoyer la température de commande, l'utilisateur fait appel aux protocoles TCP/IP et aux sockets (objet liant une adresse IP à un port). On établit une communication client-serveur, sur le réseau local du laboratoire. Le diagramme d'état ci-dessous illustre le principe.

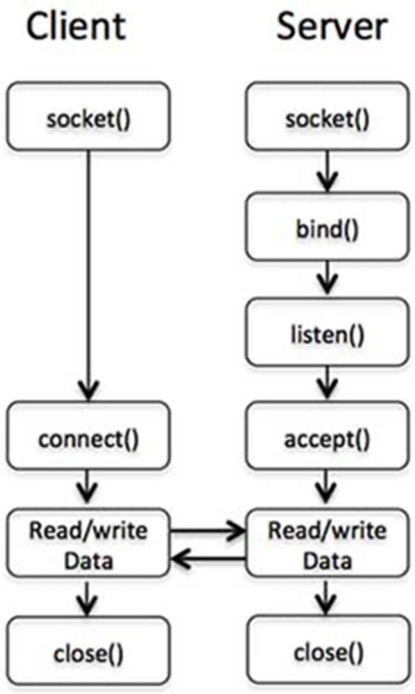


Figure 17 : Diagramme d'état de la communication client-serveur [4]

Le serveur (le Raspberry PI) se met en mode « écoute » sur un port spécifié puis le client (le PC sur lequel l'utilisateur travaille) se connecte au serveur et les deux communiquent jusqu'à ce que le client ferme la connexion. Au cours de leur échange, le client envoie la température de consigne désirée. Lorsque le Raspberry PI reçoit la température de consigne, il la transmet au microcontrôleur via la liaison série, puis le microcontrôleur se charge de l'asservissement de la cellule Peltier pour atteindre la température de consigne.

5.2 - Gestion des types

Le transfert des données requiert de bien manipuler les codages utilisés pour coder l'information. Nous allons illustrer quelques passages clefs du code implémenté dans le Raspberry qui décode la trame envoyée par le uC.

dataSerialRec = ser.read(140), permet de lire 140 octets stockés dans le buffer du bus série. La variable dataSerialRec est de type bytes de Python et ***dataSerialRec[i+3:i+5].hex()*** affiche les bytes de i+3 à i+4 en hexadécimal sous la forme d'une chaine de caractère.

C'est ici que l'on utilise la manière dont le microcontrôleur envoie la température, c'est à dire que le 3e et 4e octet représentent la température codée en virgule fixe signé (8 bits pour la partie entière et 8 bits pour la partie fractionnaire). Alors à partir d'une chaîne de caractère qui représente le mot binaire en hexadécimal (***'0x'+dataSerialRec[i+3:i+5].hex()***) et du format utilisé (***signed=1, m= 8,n=8***), on crée un type « fixed point » de la librairie fixedPoint, avec la commande ***FixedPoint('0x'+dataSerialRec[i+3:i+5].hex(), signed=1, m= 8,n=8)***.

Cet élément de type fixed point est ensuite converti en flottant et ajouté à la liste qui stocke toutes les températures reçues. On obtient alors la commande suivante : ***tempOrderFromuC.append(float(FixedPoint('0x'+dataSerialRec[i+5:i+7].hex(), signed=1, m= 8,n=8)))***

L'utilisation de la librairie permet de rendre plus lisible et plus robuste la conversion.

6 - Conclusion & perspective

Cet article propose un système de régulation de température à base d'une cellule Peltier. Il est d'un point de vue pédagogique assez complet dans le domaine du génie électrique et de l'informatique industrielle. Une partie requiert des compétences en protocole de communication d'informatique industrielle et d'analyse de signaux, pour établir la communication SPI entre le capteur de température et le microcontrôleur ou la communication série entre le uC et le Raspberry. Il y a évidemment une grande partie liée à l'automatique, l'asservissement de système non linéaire. Des connaissances sur l'électronique de puissance sont aussi nécessaires pour la mise en œuvre d'une alimentation à découpage avec une très faible ondulation. Pour faire communiquer le PC maître avec le Raspberry PI nous utilisons les réseaux et établissons une communication via sockets. Enfin, le Raspberry PI aborde la notion de thread avec un ordonnancement des tâches en « Round Robin » (Real-time operating system) mais aussi de la conversion des types et de l'interface graphique.

L'ensemble des sous parties peuvent être améliorées, stratégie de modulation sur la MLI, type d'asservissement, utilisation de la notion d'objet pour la programmation de l'interface sur python, etc. Ainsi pour des élèves de niveau Master, il offre la possibilité d'approfondir des notions spécifiques d'automatique, de programmation, d'électronique, etc. L'ensemble des documents sources sont disponibles sur Culture Sciences de l'Ingénieur - eduscol (https://eduscol.education.fr/sti/si-ens-paris-saclay/ressources_pedagogiques/controle-de-temperature-via-une-cellule-peltier).

Remerciement & Annexe

Un grand merci à Christophe SALLÉ et Christopher DOUGLAS pour leur aide dans l'entreprise de ce système.

Annexe - Coût du système

Notre système de contrôle de température relève d'un certain coût pour sa conception et sa mise en œuvre. On propose ici de budgéter l'ensemble des composants nécessaires au fonctionnement du système (cf. image ci-dessous).

Matériel	Prix H.T	T.V.A	Prix T.T.C
Alimentations			
1 x Module hacheur Texas Instruments BOOSTXL-DRV8305EVM	83,18 €	20,00%	99,82 €
1 x Module d' Alimentation linéaire régulée Double DC LT1083 24V 5A	24,92 €	20,00%	29,90 €
Filtre passe-bas			
6 x Condensateur EPCOS B32562 10µF 63 V ac	24,78 €	20,00%	29,74 €
2 x Inductance Murata 60B104C 100µH - 7.5 A	9,80 €	20,00%	11,76 €
2 x Bornier 3 contacts GMKDS 3/3-7.62	7,60 €	20,00%	9,12 €
Échangeurs thermiques			
1 x Cellule Peltier Laird Thermal Systems 387005665	83,33 €	20,00%	100,00 €
1 x Thermocryostat Huber Minichiller 300 OLÉ	3 458,33 €	20,00%	4 150,00 €
Mesure de température			
2 x Resistance to digital converter MAX31865	57,92 €	20,00%	69,50 €
2 x Sonde de température PT100 Capteur RTD PT100 362-9812	15,36 €	20,00%	18,44 €
Systèmes embarqués			
1 x Raspberry PI 3 - 1gb	27,72 €	20,00%	33,26 €
1 x Carte de développement Ti LAUNCHXL-F28379D	43,17 €	20,00%	51,80 €
Affichage & autres			
1 x Écran	83,33 €	20,00%	100,00 €
1 x Carte microSD 16 gb	8,33	20,00%	10,00 €
1 x Câble hdmi vers DVI	6,67	20,00%	8,00 €
1 x Adaptateur secteur Raspberry PI	9,75	20,00%	11,70 €
	Total H.T :		Total T.T.C
	3 944,19 €		4 733,04 €

Figure 18 Gestion budgétaire du système de contrôle de température

On remarque que le prix, élevé, du système est notamment dû au thermocryostat dont le coût hors taxe s'élève à 3 458, 33 €, on pourrait refroidir la cellule Peltier avec un ventilateur fixé sur la face chaude, cela diminuerait grandement le coût du système pour environ 500 €.

Références :

[1]: Goupil, Christophe, Henni Ouerdane, et Yann Apertet. 2013. « Thermoélectricité - Thermodynamique et applications ». Consulté le 10 juin 2022. *Ressources énergétiques et stockage*, avril. Disponible sur : https://doi.org/10.51257/a-v1-be8080

[2]: Cellule Peltier [Documentation constructeur]. Consulté le 10 juin 2022. Disponible sur : https://www.mouser.fr/datasheet/2/988/datasheet_PCX15_6_19_F1_4040_TA_RT_W6-2298292.pdf

[3]: Thermocryostat. [Documentation constructeur] Consulté le 11 juin 2022. https://www.huber-online.com/download/pdf_datasheets/englisch/3006.0089.98.PDF

[4]: Diagramme d'état de la communication client-serveur, consulté le 16 juin 2022, disponible sur : https://www.tutorialspoint.com/perl/perl_socket_programming.htm

[5]: Convertisseur de température. Consulté le 12 juin 2022. Disponible sur : https://datasheets.maximintegrated.com/en/ds/MAX31865.pdf

[6]: Raspberry PI 4. Consulté le 14 juin 2022. Disponible sur : https://www.planete-domotique.com/blog/2019/09/16/raspberry-pi-4-modele-b-il-se-met-en-quatre-pour-vous/

Ressource publiée sur Culture Sciences de l'Ingénieur : https://eduscol.education.fr/sti/si-ens-paris-saclay

Les combustibles métalliques renouvelables, en route vers une nouvelle ère énergétique ?

La Revue 3E.I
Culture Sciences de l'Ingénieur

Édité le 05/02/2024

école normale supérieure paris—saclay

Driss LARAQUI[1]

[1]*Expert en combustibles métalliques, Co-fondateur de FENIX Energy*

Cette ressource fait partie du N°111 de La Revue 3EI de janvier 2024.

Cette ressource présente la nouvelle filière énergétique des combustibles métalliques renouvelables dans le contexte actuel de la décarbonation des activités humaines. Ce vecteur énergétique innovant alliant forte densité énergétique, recyclabilité des ressources métalliques, simplicité et sûreté de transport peut permettre le stockage longue durée et le transport longue distance des énergies renouvelables. Les combustibles métalliques, comme la poudre de fer, ont la capacité, comme leur nom l'indique, de brûler en libérant de la chaleur sans émettre de CO2. L'équation de réaction exothermique du fer avec l'oxygène qui a pour produit l'hématite solide (ou oxydes de fer III) est la suivante :

Réaction de production d'énergie verte par le fer :

$$2\,Fe + 3/2\,O_2 -> Fe_2O_3 + \ chaleur\ (\Delta rH_0 = -824\ kJ/mol) \qquad \text{(Eq. 1)}$$

Cette solution s'intègre dans une logique de répartition des énergies renouvelables à l'échelle mondiale, en facilitant les flux entre les pays excédentaires en renouvelable et les pays fortement consommateurs en énergie. La circularité de ces métaux réside en effet dans la possibilité de capturer les oxydes métalliques issus de leur combustion (lors de la production d'énergie) pour, a posteriori, réaliser la régénération du métal pur à l'aide d'énergies renouvelables et de réaliser ainsi leur stockage longue durée dans ce carburant facile et sûr à stocker. La régénération peut se faire par voie électrolytique, comme c'est souvent le cas dans la métallurgie de l'aluminium ou du magnésium (idéalement avec des anodes inertes), ou par voie thermochimique comme plus classiquement dans l'exploitation des minerais de fer (présents sous forme d'oxydes dans la nature). La production de fer vert dans la métallurgie ou dans le cadre du recyclage du carburant de fer suit l'équation chimique endothermique suivante :

Réactions de stockage d'énergie renouvelable dans le fer :

Etape 1 : $$H_2O + électricité\ verte -> H_2 + 1/2\ O_2\ (\Delta rH_0 = \ 241{,}6\ kJ/mol) \qquad \text{(Eq. 2)}$$

Etape 2 : $$Fe_2O_3 + 3H_2 + chaleur\ verte -> 2Fe + 3H_2O\ (\Delta rH_0 \ = \ 98{,}76\ kJ/mol) \qquad \text{(Eq. 3)}$$

À la vue des deux équations de réaction précédentes, il est très clair que le fer et l'eau (cette dernière utilisée pour produire de l'hydrogène) peuvent être utilisés en boucle quasi-fermée pour préserver nos ressources (éviter le stress hydrique dans les pays producteurs déjà très secs).

Capturer et régénérer 99% des oxydes produits dans la combustion à chaque cycle permet de réutiliser 100 fois le stock initial de poudre métallique. S'ajoute à cet avantage majeur le fait que 5 litres de cette poudre (dans le cas du fer), contient autant d'énergie que 10000 litres d'hydrogène, ce qui en fait un carburant 3 fois moins coûteux à transporter sur de longues

distances. Sa stabilité à basse température et l'absence de fuite en raison de son état solide en fait également un combustible 15 fois moins coûteux à stocker que l'hydrogène. Cette solution arrive à point nommé dans une période où la nécessité d'importer de l'énergie verte dans les zones fortement consommatrices par l'intermédiaire de l'hydrogène s'avère être une tâche complexe, risquée, coûteuse en infrastructures et en eau.

1 - Limites des vecteurs énergétiques bas carbones actuels : batteries, biomasse, hydrogène vert et autres carburants issus du stockage *"Power to X"*

Si on prend la définition stricto sensu, selon la norme ISO 13600, un vecteur énergétique est une méthode permettant de transporter de l'énergie d'un endroit à un autre pour être transformée sous forme de chaleur ou de travail mécanique, ou être utilisée dans des processus physiques ou chimiques. Un vecteur énergétique sert donc au stockage et au transport des énergies intermittentes, pour les rendre pilotables et transportables. Un anglicisme décrit une catégorie de vecteurs énergétiques par le terme « Power-to-X ». Ces vecteurs énergétiques prennent tout leur sens dans le contexte énergétique où l'Europe, déficitaire en énergie renouvelable à faible coût, doit importer des énergies vertes depuis les pays fortement excédentaires en renouvelable. Il faut donc stocker cet excédent dans la production d'un vecteur énergétique qui sera transporté et consommé à des milliers de kilomètres de sa zone de production et à des périodes également décalées éventuellement de plusieurs mois (stockage saisonnier). L'Allemagne est un bon exemple de pays ayant défini un plan massif d'importation d'hydrogène vert dès 2030 depuis les pays tiers producteurs d'énergies renouvelables. Pour citer un article de l'ambassade de France en Allemagne, "L'Allemagne ne pourra pas se passer d'importations : d'ici 2030, la stratégie prévoit la production de 14 TWh d'hydrogène vert, or la demande nationale est estimée entre 90 et 110 TWh (…) parmi les pays prioritaires pour l'exportation d'hydrogène vert "des « nouveaux » entrants comme le Chili, le Brésil, l'Afrique du Sud, le Maroc, le Portugal ou encore l'Australie, disposant de conditions favorables aux projets renouvelables à grande échelle" [1].

Le premier exemple qui nous vient à l'esprit pour stocker de l'énergie : les batteries (NMC, LFP). Ces dernières exploitent beaucoup de ressources minières et impliquent de fortes tensions sur ces dernières. De plus, il n'est pas possible d'envisager un système d'échanges planétaires d'énergie utilisant des batteries.

La catégorie des biocarburants (à partir de biomasse) est une forme de Power-to-X dans le sens où ils sont issus du captage naturel de CO2 et de rayonnement solaire (biomasse) ils sont des systèmes naturels de stockage d'énergie solaire et de capture de CO2. Néanmoins la déforestation qui devient plus importante que la reforestation par cette filière, la concurrence sur les ressources à l'agriculture, les émissions de particules et autres polluants (COV, HC, suies) restent encore les freins importants au développement à grande échelle des biocarburants.

Ceux qui sont issus du stockage par voie technologique d'énergies renouvelables par captage ou non de CO2 (e-fuel, hydrogène vert, ammoniac) présentent chacun leurs difficultés.

La difficulté de transport et de stockage hydrogène et la consommation d'eau locale importante pour les molécules présentant de l'hydrogène comme l'ammoniac ou les e-fuel (9L/kg juste pour la réaction chimique sans compter les pertes d'eau liées au procédé). En effet, cette problématique se pose d'autant plus que cette eau sera extraite dans les pays producteurs de renouvelable pour rejoindre les pays très consommateurs d'énergie. Pour un ordre d'idée, produire 100 TWh d'hydrogène pour l'exporter en Europe (une quantité d'énergie répondant au besoin actuel en hydrogène de l'Allemagne par exemple) correspond à une consommation de 28,1 millions de m^3

d'eau annuellement (consommation annuelle d'1M d'habitants) ! Et ceci est valable pour l'ammoniac et les e-fuels. Certes cette eau se retrouve dans la pluie… mais diffuse dans la zone de destination. Allez donc trouver cette quantité d'eau dans le désert marocain où le Soleil est lui, abondant. Le stress hydrique est une vraie préoccupation de ces pays producteurs d'énergie solaire [2]. D'où la nécessité de trouver un procédé qui permette d'utiliser en boucle la même quantité d'eau afin de préserver les zones de production de renouvelable où l'eau est aussi critique qu'ailleurs pour la biodiversité et les populations locales. Et la production d'eau douce à partir d'eau salée est un procédé très énergivore et possible qu'à proximité de la zone de consommation de l'eau. Sans oublier les risques de sécurité liés au transport et au stockage d'hydrogène qui nécessitent encore un développement important des systèmes actuels pour s'y prémunir [3,4,5].

La toxicité de l'ammoniac est un élément bloquant à son transport longue distance. En effet, il est possible de sentir l'ammoniac à une concentration de 2 à 55 parties par million (ppm). La molécule est considérée comme un danger grave pour la santé en raison de sa toxicité, une exposition à 300 ppm représente un danger immédiat [6].

Toutes ces problématiques peuvent être levées par l'utilisation de métal vert régénérable décrit dans la partie suivante.

2 - Les combustibles métalliques renouvelables comme solution au stockage et au transport à bas coût d'énergie verte

2.1 - Apparition et principe du concept de combustible métallique circulaire

2.1.a - Premières utilisations du potentiel énergétique des métaux

L'idée d'utiliser les métaux pour leur forte densité énergétique n'est pas nouvelle [7,8], les premiers secteurs à s'être intéressés au sujet sont, comme souvent, l'armement et le spatial. En effet, la propulsion spatiale par poudre métallique existe depuis les années 1950 et est encore utilisée aujourd'hui dans les fusées les plus modernes (Ariane, etc.) pour booster le propergol qui sert à la première phase du lancement pour l'extraction à la gravité terrestre (particules d'aluminium de 100 microns). Certains missiles sont équipés également de poudre métallique en tous genres, sans oublier d'évoquer les feux d'artifices qui sont aussi un bel exemple de l'utilisation du potentiel énergétique des particules métalliques (souvent sous forme de sels métalliques, de limailles ou de granules).

Figure 1 : À gauche, lancement d'Ariane 5, à droite, un feu d'artifices, source :http://spaceblog.over-blog.com/

2.1.b - Apparition du concept pour une application à la production d'énergie

Le concept du métal renouvelable pour subvenir aux besoins énergétiques sur terre a été détaillée et proposée pour la première fois par le chercheur Japonais Yabe dans son livre « The Magnesium Civilization: An Alternative New Source of Energy to Oil »(2008) [9], et a ensuite été portée dans la recherche par JF bergthorson à McGill [10] et poussée à l'échelle industrielle par le chercheur Philip De Goey initiant le projet étudiant Team Solid en 2017 qui a permis l'apparition de la Start up RIFT en 2020.

En France, nous pouvons être fiers d'avoir eu des chercheurs dont les travaux ont été financés par un grand groupe automobile français dès 2013, pour ma part en 2016, afin d'évaluer la faisabilité technique d'applications énergétiques utilisant la combustion métallique [11,12,13]. La start-up Fenix Energy© a été créée à Lyon en 2023, par l'auteur de l'article, pour perpétuer les travaux sur la combustion métallique en France en leur faisant atteindre un niveau de maturité industrielle pour des applications dans le chauffage et la production d'électricité pour l'industrie et les collectivités [14].

L'idée d'utiliser les métaux comme combustibles découle d'une simple sélection établie en partant du tableau périodique des éléments. En effet, il faut un élément qui puisse réagir avec l'oxygène avec un taux de réaction suffisant pour de la combustion, qui ne soit pas rare (élimine Titane, Platine, Bore…), qui ne soit pas lourd (élimine le bas du tableau périodique), dont les oxydes ne soient pas toxiques (élimine Béryllium…), qui soient stables à basse température en contact d'air ou d'eau (élimine Sodium, Lithium…). Il reste donc 3 métaux aux côtés de l'hydrogène (en mettant évidemment de côté le carbone). Parmi eux, l'aluminium, le fer et le magnésium sont parmi les 8 éléments les plus présents de la croûte terrestre.

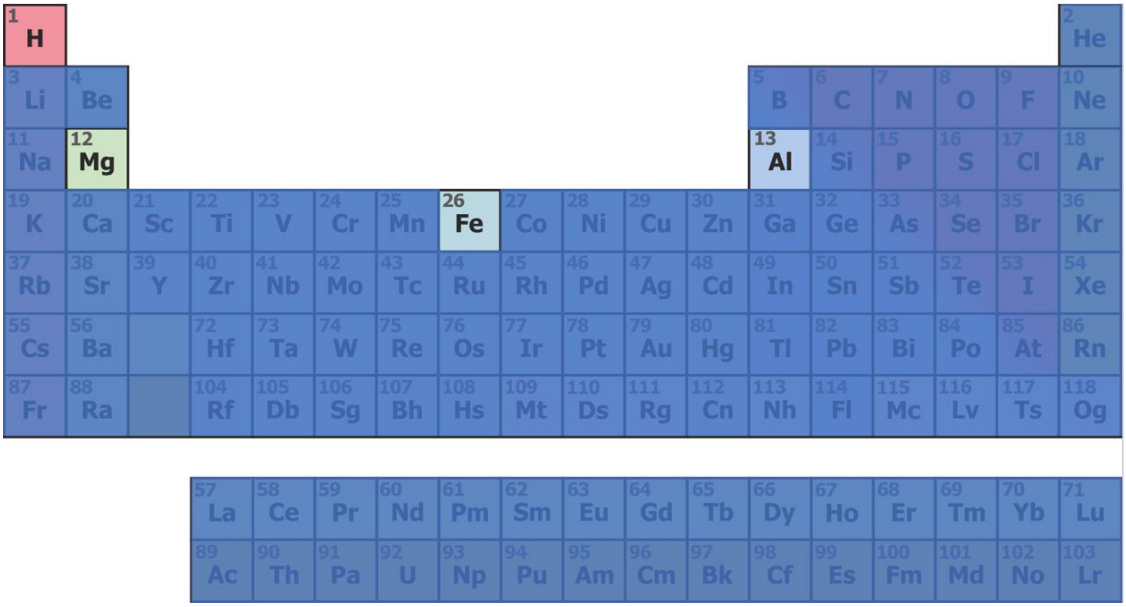


Figure 2 : Tableau périodique des éléments suite à la sélection de combustibles potentiels pour la production d'énergie à grande échelle

2.2 - Les avantages compétitifs des métaux en tant que combustibles et vecteurs d'énergie

2.2.a - Les métaux sont carburants naturellement denses énergétiquement

Le potentiel énergétique de ces métaux est comparable à ceux des carburants fossiles, et supérieur à celui des alternatives actuelles de stockage d'énergie verte, comme le montre la figure ci-dessous. Cet aspect permet d'envisager ces poudres métalliques pour remplacer les carburants

fossiles dans toutes les applications énergétiques où ils sont aujourd'hui exploités (centrales électriques, chaufferies, groupes électrogènes, moteurs automobiles).

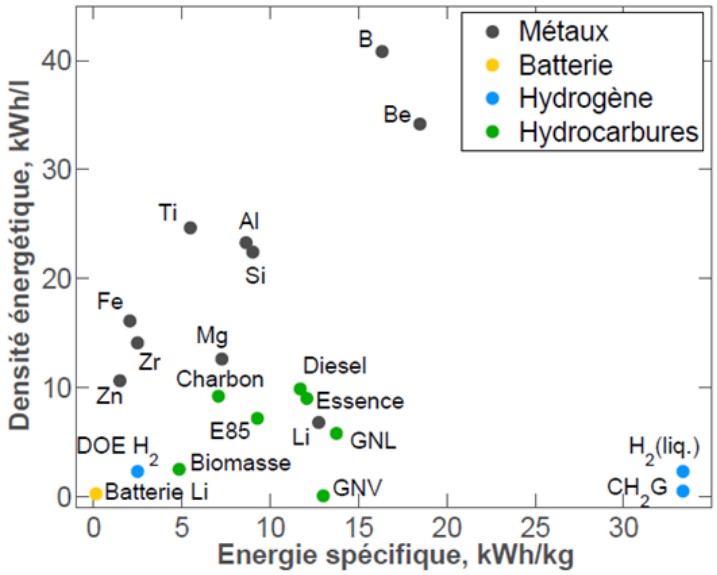


Figure 3 : Densités énergétiques de différentes catégories de carburants (métaux, hydrocarbures…)

2.2.a - Les métaux sont des carburants recyclables

Avec l'avantage supplémentaire, détaillé dans la partie suivante, que les produits de combustion de ces métaux génèrent des oxydes solides (contrairement aux carburants fossiles) ce qui facilite grandement leur récupération pour un recyclage ultérieur vers le métal pur initial.

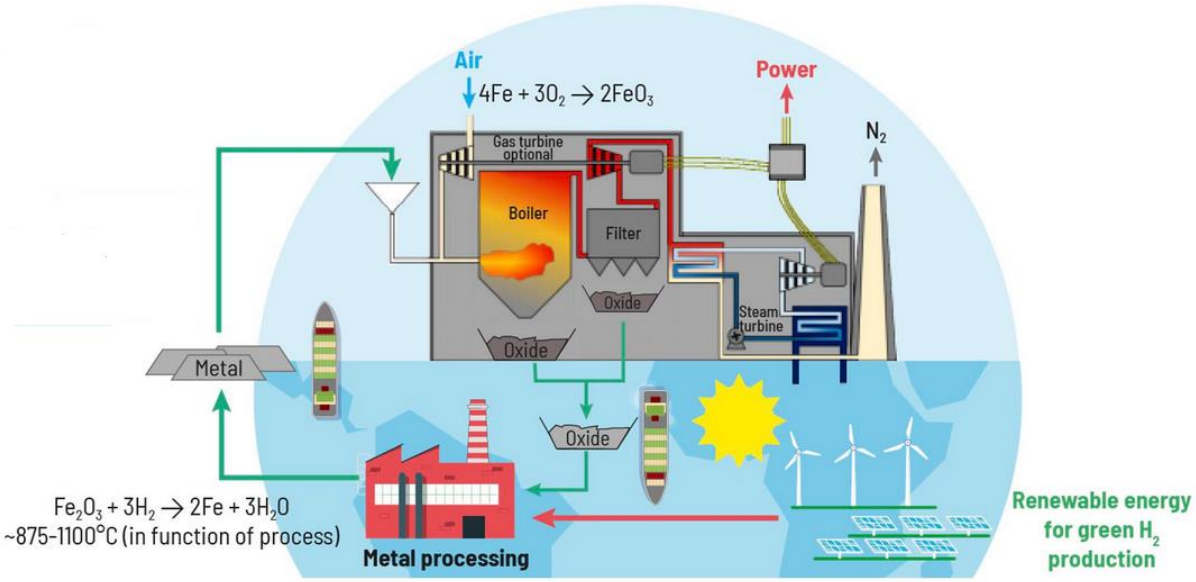


Figure 4 : Schéma de principe du cycle du fer, de sa production dans des zones excédentaires en renouvelables jusqu'à son transport et sa consommation dans des zones très énergivores de la planète, Source : inspiré Engie

Comme le montre le schéma précédent, une fois le combustible utilisé pour produire de l'énergie, les produits peuvent être acheminés vers des zones de recyclage alimentées en énergie renouvelable, pour réaliser la régénération des oxydes métalliques vers le métal pur. Ainsi, la réaction de réduction étant consommatrice d'énergie, elle permet de stocker cette énergie dans la production d'un carburant qui pourra être brûlé des jours plus tard et à des milliers de kilomètres, au lieu et au moment où le besoin en énergie verte se fait ressentir (exemple : la nuit, dans les zones faiblement productrices en renouvelables, dans les zones fortement consommatrices en énergies, etc.). Le rendement du cycle total avoisine les 70% pour la production de chaleur (du

panneau solaire à la chaleur en sortie de chaufferie), principalement limité par le rendement des électrolyseurs, et pour la production d'électricité le rendement global de ce cycle avoisine les 35%, de manière équivalente à tout combustible étant soumis au rendement de Carnot pour la production d'électricité (hydrogène, gaz, pétrole, charbon). Le rendement est important pour minimiser le coût final de l'énergie utile mais ce qui compte également, et d'une manière plus forte, c'est le coût initial de l'énergie renouvelable et de son transport/stockage. Plus le carburant à produire est facilement transportable, plus on peut se permettre de le produire dans zones à faible coût de renouvelable.

2.2.c - Ils sont peu consommateurs en ressources : métaux et eau

La circularité de la ressource permet de réduire l'impact sur l'environnement en éliminant le besoin en minage des métaux. La criticité en ressources métalliques n'est plus d'actualité dans le cadre de ce concept. En effet, si l'on réalise 99% de rendement de recyclage il est possible de réutiliser 100 fois le métal initial. De plus, l'utilisation en boucle de l'hydrogène sans besoin de quantité d'eau à renouveler totalement est un avantage certain dans des zones où présence abondante d'énergie solaire rime avec sécheresse et pénuries d'eau. Prenons l'exemple d'un pays qui souhaiterait exporter 10 Mt d'hydrogène par an, cela nécessiterait d'exporter en réalité 90 000 000 m^3/an d'eau équivalents à la consommation de 2 millions de personnes. Cette problématique se pose pour l'hydrogène vert, les e-fuel, l'ammoniac et les autres carburants de synthèse produits à partir d'eau, mais pas pour les carburants métalliques.

H2O -> H2 + ½ O2 (Pour 1 kg de H2 produit il faut consommer 9kg d'eau, sans compter le rendement réel)

2.2.d - La simplicité et le faible coût logistique du stockage saisonnier et du transport longue distance de la poudre métallique en font des vecteurs d'énergie idéaux.

Afin de réaliser efficacement cette réduction thermochimique, qui est l'objet de la partie 2.3, l'apport d'hydrogène vert et d'énergie renouvelable pour le chauffage du mélange est nécessaire. Il convient donc de réaliser le recyclage de ces oxydes métalliques dans des zones abondantes en renouvelable pour minimiser le coût de revient du combustible produit afin de le rendre compétitif avec les combustibles actuels (gaz, pétrole, charbon). Le prix actuel de l'hydrogène vert en Europe avoisine les 10€/kg alors que dans les pays fortement producteurs de renouvelables l'hydrogène peut atteindre des coûts inférieurs à 1,5 €/kg. De nombreux pays, comme cité dans la partie précédente, ont actuellement de très gros projets de production d'hydrogène vert, et la filière métallique peut facilement s'intégrer dans cette dernière pour apporter sa complémentarité : la simplicité de stockage et de transport des métaux (voir carte Figure 5).

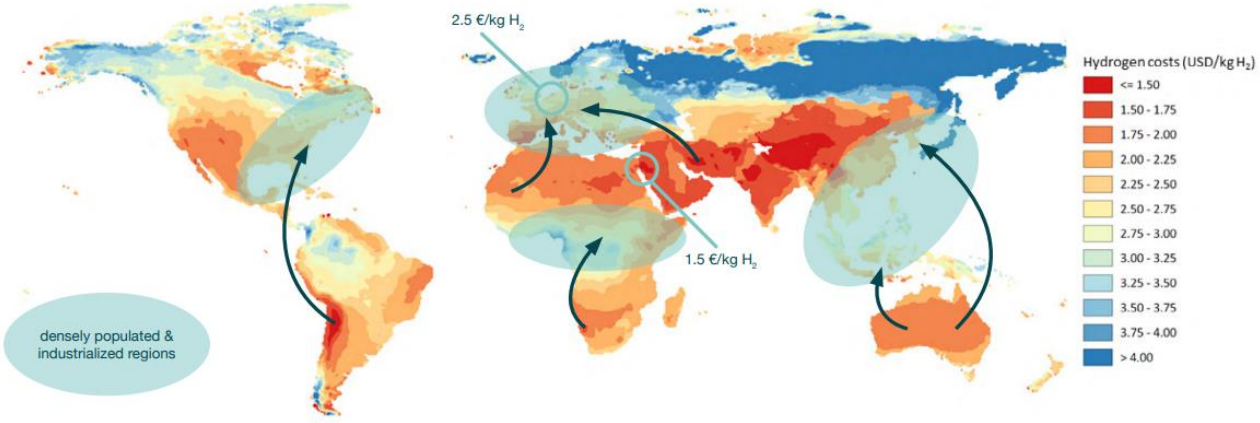


Figure 5 : Projection des coûts de production d'hydrogène dans le monde, Source : Metalot [15]

Cet aspect permet d'envisager des flux d'énergie verte entre les zones productrices de renouvelable et les zones consommatrices, comme décrit précédemment, sans avoir à développer des infrastructures très coûteuses et risquées pour les populations environnantes (pipelines, transport de réservoirs sous pressions). Comme le montre le schéma ci-dessous, le métal peut être transporté en camion, bateaux ou trains, dans des containers simples ou des big bags suivant les quantités. La raison à cela est simple, il suffit de 5 litres de poudre de fer pour transporter l'énergie contenue dans 10 000 Litres d'hydrogène. De plus, le point d'inflammation est bien plus élevé (supérieur à 1000 degrés), et le carburant étant solide, les risques de fuites sont négligeables.

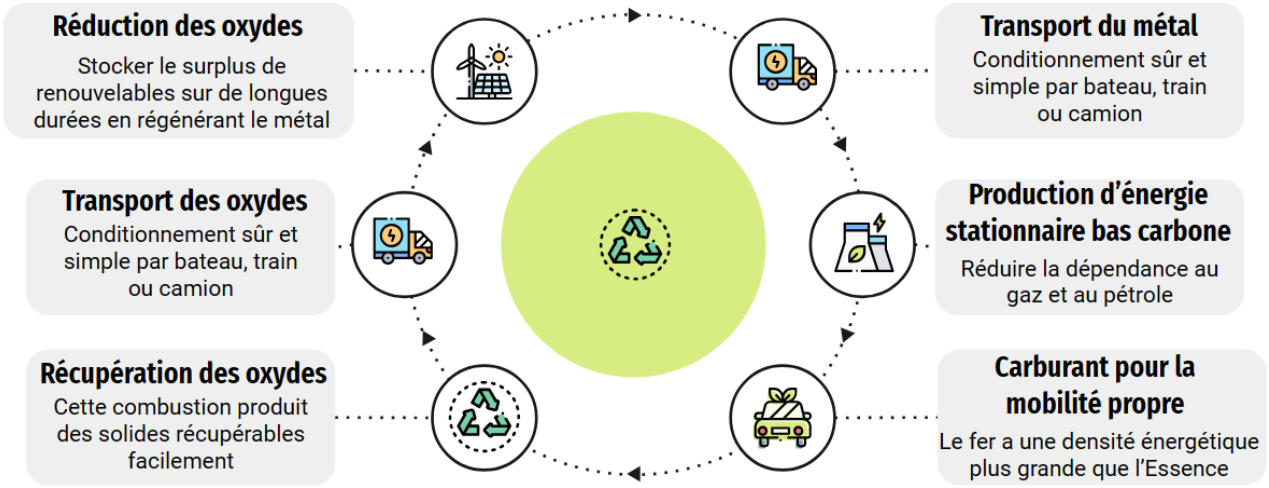


Figure 6 : Cycle logistique des métaux régénérables par énergies renouvelables pour la production d'énergie verte

La simplicité de stockage et de transport du métal est un argument supplémentaire qui justifie l'utilisation des métaux comme vecteurs énergétiques vis-à-vis de l'hydrogène et des autres carburants de synthèse. Sans oublier d'évoquer les investissements initiaux bien plus importants pour les infrastructures liées au transport sous pression ou par liquéfaction de l'hydrogène contrairement à des « big bags » de fer qui utilisent les modes de transports actuels sans transformation (bateau, train et camions).

La Figure 7 montre que dans un futur proche (d'ici 10-15 ans), un hydrogène produit dans un pays fortement producteur en renouvelables coûtant 1,5 €/kg permettra d'atteindre un coût au kWh thermique de fer inférieur à 0,08€/kWh, le rendant compétitif avec le gaz naturel et bien moins cher que l'essence ou le diesel. Avec la croissance des taxes carbones, la compétitivité de ce vecteur énergétique deviendra d'autant plus évidente. L'intérêt vis à vis d'une utilisation de l'hydrogène se présente dès qu'il s'agit de transporter ou de stocker sur plusieurs semaines le vecteur énergétique. Ce qui est le cas dans le cadre du stockage saisonnier, des réserves stratégiques et dans le marché global de l'énergie verte qui commence à prendre de l'ampleur.

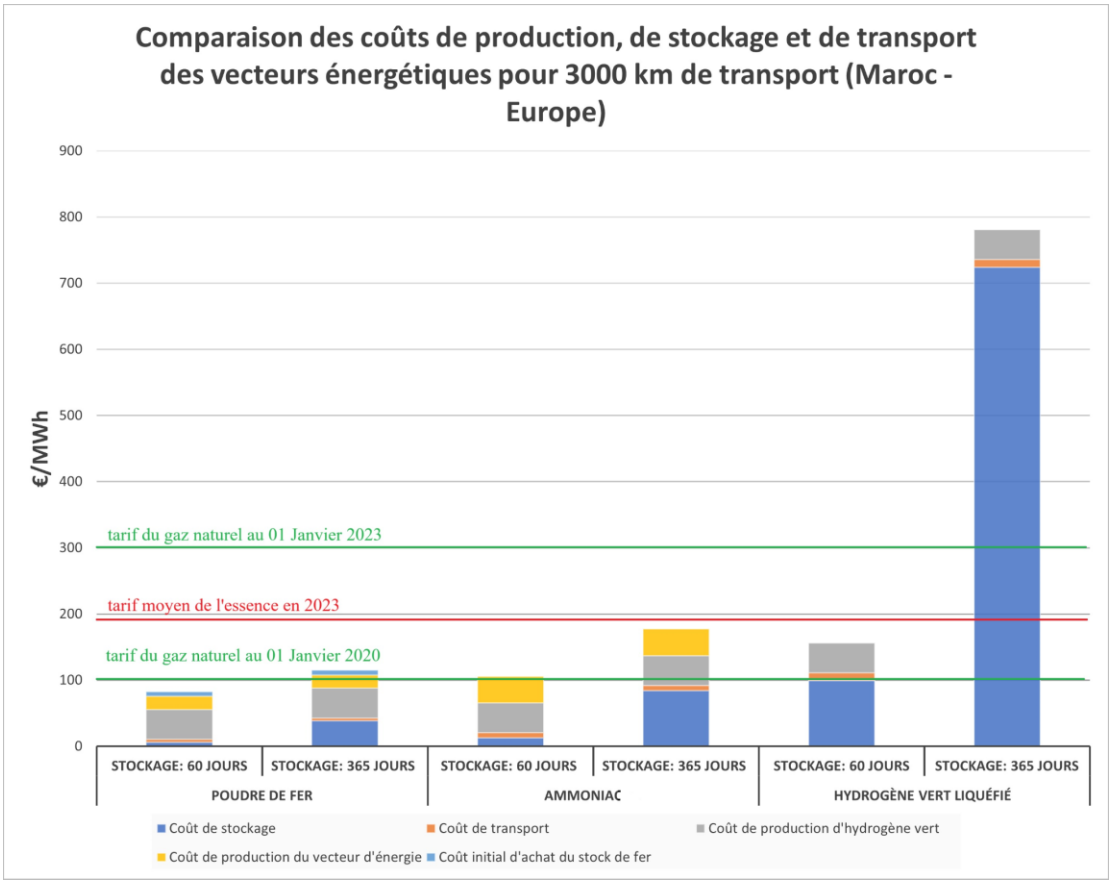


Figure 7 : Projections de coûts de revient de 3 vecteurs énergétiques pour le stockage saisonnier et le transport longue distance d'énergies vertes (données de Metalot [15] et Fenix Energy© [14], hypothèse de 1,5€/kg de H2)

2.3 - La combustion métallique : une façon innovante mais non nouvelle de générer de la chaleur sans émettre de CO2

La combustion est un phénomène physico-chimique présent au sein de notre société depuis des millénaires et dont le but premier est de convertir l'énergie chimique contenu dans un combustible en énergie thermique utile à plusieurs fins. Les 3 conditions nécessaires et suffisantes à l'obtention de cette réaction chimique sont une source de chaleur pour amorcer la réaction, un combustible et un comburant. L'innovation dans la combustion métallique porte sur le combustible. La grande majorité des combustibles utilisés sur Terre sont carbonés, mais comme évoqué précédemment, il est en réalité possible de brûler les métaux comme c'est le cas dans certaines applications.

D'un point de vue combustion fondamentale, plusieurs paramètres doivent être maîtrisés (vitesse de combustion, délai d'allumage, température de flamme, stabilité…) pour caractériser ces combustibles et les comparer aux combustibles classiques pour mieux adapter les systèmes existants. Il est en effet envisagé de "rétrofit" les centrales électriques à charbon vers le fer [16].

La combustion appliquée qui nous intéresse, est pluridisciplinaire, à l'intersection de la mécanique des fluides, la thermodynamique, la chimie, la thermique, ce qui en fait une discipline passionnante et peu monotone !

La particularité de la combustion métallique est que le phénomène de combustion est présent à la fois à petite et grande échelle. Le carburant étant solide, chaque particule peut être vu comme un microréacteur quasi-indépendant composé de sa micro-flamme [17] séparant le réducteur de l'oxydant. En plus de cela, la combustion de ces microréacteurs est possible sous deux modes différents : homogène et hétérogène. La réaction homogène, en phase gaz, présente l'inconvénient

de ne pas maîtriser la taille des particules d'oxydes formés (majoritairement des nanoparticules), qui posent des problèmes de filtration en sortie du brûleur et de régénération du métal initial vers la granulométrie initiale. Ce mode est prépondérant pour l'aluminium et le magnésium [11,12,13,18].

Le critère de Glassman permet d'estimer si le métal peut brûler en phase vapeur, et donc en phase homogène, à partir de ses propriétés thermodynamiques (et celles de son oxyde). Il part du principe que la température de flamme est limitée à son maximum par la température de vaporisation de l'oxyde. Pour que la réaction se fasse en phase vapeur il faut que la température de vaporisation de l'oxyde soit bien au-dessus de la température d'ébullition du métal, autrement dit il faut qu'a minima la température de vaporisation de l'oxyde soit au-dessus de celle d'ébullition du métal :

$$T_{vaporisation\ oxydes} > T_{ébullition\ métal} \text{ (condition nécessaire à un mode homogène)}$$

Pour savoir si une combustion homogène est possible, il suffit donc de vérifier si la condition citée ci-dessus est vérifiée pour le métal considéré. C'est la seule façon d'obtenir une température de réaction au-dessus de la température de vaporisation du métal et donc de le vaporiser. Le Tableau 1 rassemble quelques métaux et oxydes correspondants et les températures de vaporisation associées :

Metal	T_b (K)	Oxyde	T_{vap} (K)	$\Delta H^0_{f(298)}$ (kj/mol)	$\Delta H_{vap} + (H^0_{T_{vol}} - H^0_{298})$ (kj/mol)
Al	2791	Al_2O_3	4000	-1676	2550
B	4139	B_2O_3	2340	-1272	640
Be	2741	BeO	4200	-608	1060
Cr	2952	Cr_2O_3	3280	-1135	1700
Fe	3133	FeO	3400	-272	830
Hf	4876	HfO_2	5050	-1088	1420
Li	1620	Li_2O	2710	-599	680
Mg	1366	MgO	3430	-601	920
Si	3173	SiO_2	2860	-904	838
Ti	3631	Ti_3O_5	4000	-2459	2970
Zr	4703	ZrO_2	4280	-1097	1320

Tableau 1 : Tableau de températures de vaporisation des métaux et leurs oxydes pour l'application du critère de Glassman.

Il est possible d'observer, grâce à ce tableau, que le fer a une température de vaporisation de son oxyde I très proche de la température de vaporisation du métal, ce qui implique que le mode homogène (combustion en phase gaz) est peu probable. En effet, le fer brûle sous sa forme solide (ou en partie liquide) tout au long de la réaction, c'est le mode hétérogène. L'implication est une température de réaction plus faible, un taux de réaction plus faible et des oxydes de fer de la même taille que les particules de métal initiales. La plus faible température de la combustion du fer entraîne également une réduction des émissions de NOx par le mécanisme thermique de Zeldovich relativement au magnésium ou à l'aluminium [19].

À l'échelle de la flamme, les difficultés technologiques à lever dans le cadre de la combustion métallique appliquée à la production d'énergie portent en partie sur le contrôle en continu d'une flamme métallique afin de maîtriser son utilisation. Ce qui n'est pas le cas dans les boosters à poudre de fusées ou encore dans les feux d'artifices où la combustion est volontairement non contrôlée. Des études ont montré la faisabilité d'une telle source fixe de chaleur à puissance relativement élevée. L'idée étant de doser la vitesse de l'écoulement relativement à la vitesse de la flamme, d'apporter suffisamment de chaleur pour démarrer la réaction (chose plus complexe

lorsque l'on brûle un solide qu'un gaz) et contrôler son extinction lorsque l'on sait que chaque particule peut maintenir sa combustion indépendamment du reste du mélange.

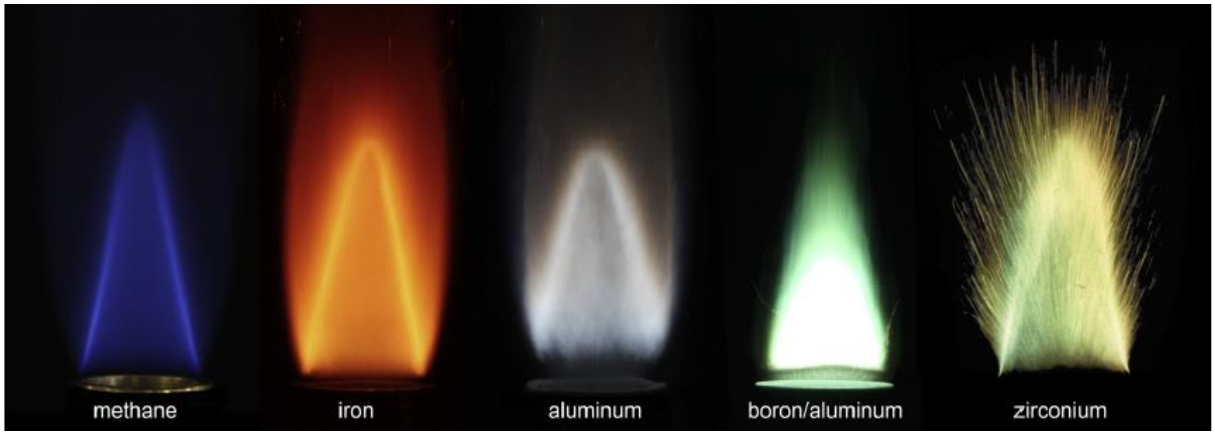


Figure 8 : Photographies de flammes métalliques (à partir de la seconde)

La stabilisation de flamme laminaire consiste à avoir un écoulement de gaz frais prémélangé dans une direction opposée à la propagation de l'onde de combustion et à une vitesse égale à celle de cette dernière, ce que permet stricto sensu le bec bunsen.

Dans le cas des métaux, une stabilisation par bec bunsen implique une vitesse d'écoulement de l'ordre de 100 cm/s maximum à faible richesse. Ce maximum de vitesse peut être difficile à respecter sachant les contraintes de puissance (correspondant à un débit d'air) et de faible encombrement nécessité par les applications énergétiques. Il faut envisager des techniques de stabilisation d'écoulements turbulents.

Figure 9 : Extrait de la bande dessinée "Sciences en Bulles", 2020

Une autre difficulté technique concerne la filtration des oxydes métalliques, qui peuvent être nanométriques dans le cas du Magnésium et de l'Aluminium, c'est beaucoup plus simple pour le Fer qui n'émet que de la rouille micrométrique de la taille de la particule initiale.

Mis à part cela, la combustion métallique est bien évidemment propre car elle ne produit aucun composé organique volatile, hydrocarbures ou suies (par définition), elle peut dans certains cas produire des oxydes d'azote mais en quantité beaucoup plus faible que les hydrocarbures. Comme évoqué précédemment, la raison principale de la faible émission de NOx du fer vis-à-vis des autres métaux [20] est liée à la plus faible température de flamme de ce métal. Il faut compter aux environs de 2000 K pour le fer contre 3500 K pour l'aluminium et le magnésium [11,12,13]. Les NOx thermiques sont liés exponentiellement à la température de la zone réactionnelle.

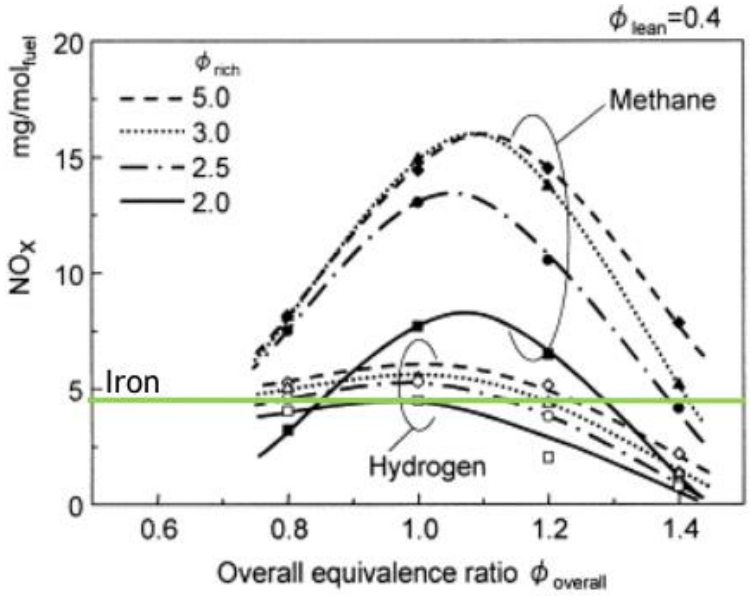


Figure 10 : Comparaison des émissions de NOx de l'hydrogène, du fer et du méthane [21] avec ajout d'un point d'émissions de NOx du fer [15]

Les coûts compétitifs au kWh du fer renouvelable (vis-à-vis du gaz, du pétrole et du nucléaire à plus long terme), combiné à des technologies classiques utilisables pour brûler ces métaux, rendent l'ensemble des applications énergétiques compatibles avec ce carburant circulaire. Parmi les applications urgentes et plus accessibles à décarboner nous pouvons citer le chauffage urbain, les centrales de cogénération, les groupes électrogènes, le rétrofit de centrales au charbon dont la compétitivité dépendra de la volonté étatique de décarboner la production d'électricité du réseau pour assumer un léger surcoût du carburant (vis-à-vis d'un charbon peu coûteux). En général, pour la production d'électricité « on-grid » le fer présente des avantages certains, dont économiques et de sûreté, vis-à-vis de l'hydrogène (cités dans l'article).

Les moteurs automobiles peuvent être également envisagés, mais à plus long terme à cause des technologies plus complexes et des points d'approvisionnement en carburant trop dispersés. Mais les gestes de sobriété combinés à une électrification verte seront probablement suffisants les pays décarbonant leur électricité stationnaire (potentiellement par le fer). Même si la problématique liée aux ressources en minerai restera le point bloquant de cette solution. Il y a certainement encore un marché dans les pays émergents qui augmentent leur dépendance au transport individuel durant leur croissance économique.

2.4 - La réduction des oxydes métalliques : la réaction inverse permettant de stocker de l'énergie renouvelable dans la production du métal vert.

Comme dans une batterie on peut inverser le sens de la réaction en apportant de l'énergie au système. Sauf qu'avec la production thermochimique (ou électrochimique) des métaux on la stocke dans un élément qui se transporte sur de longues distances, dans de faibles volumes et pour de longues durées.

Il faut imaginer que comme dans un barrage hydraulique on pompe l'eau pour remonter la pente d'une montagne, ici on remonte la pente énergétique (Gibbs) des éléments en les faisant passer d'un état bas en énergie à un état haut en énergie, c'est ce qu'on appelle le stockage thermochimique. Produire du e-fuel, de l'ammoniac, du fer et de l'hydrogène c'est du stockage thermochimique de l'énergie. On part d'éléments stables (les oxydes métalliques) et on apporte de l'énergie pour remonter la pente énergétique vers des éléments moins stables mais susceptibles par la suite de pouvoir réagir de nouveau avec l'oxygène pour libérer leur énergie (voir la partie

précédente sur l'oxydation). Le stockage thermochimique a pour avantage certain, vous l'aurez compris, qu'il ne nécessite bien sûr pas de montagne avec une réserve d'eau à proximité.

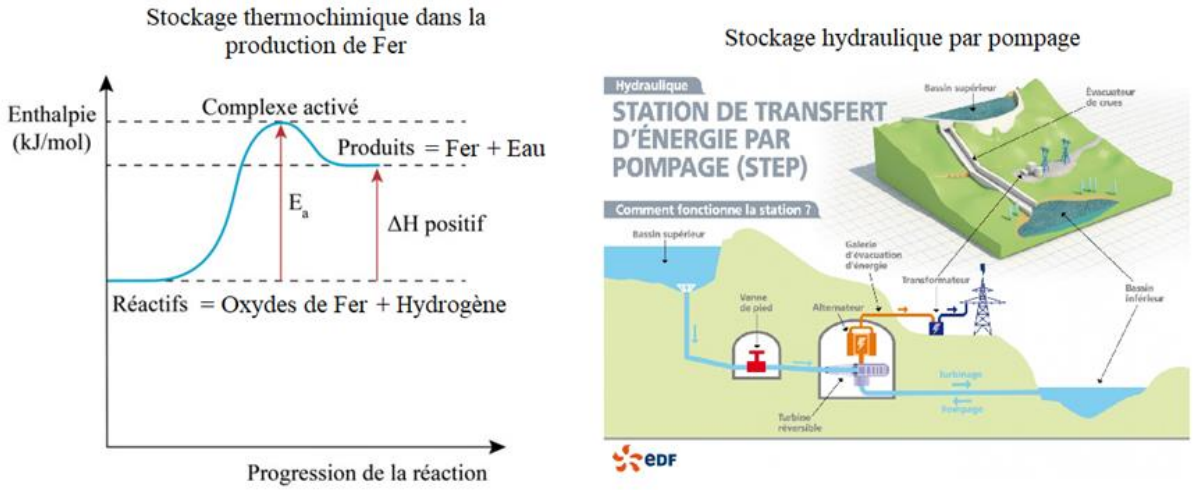


Figure 11 : Analogie entre le stockage thermochimique et le stockage hydraulique

La réaction de réduction des oxydes de fer est un exemple de stockage thermochimique possible d'énergies renouvelables. La réaction de réduction à l'aide d'un agent réducteur tel que l'hydrogène s'écrit de cette façon :

$$Fe_2O_3 + 3H_2 -> 2Fe + 3H_2O \qquad \text{(Eq. 3)}$$

Comme évoqué dans le graphe d'énergie potentiel précédent, la réaction est endothermique et nécessite une enthalpie standard de réaction de 98,76 kJ/mol.

La réaction de réduction sans agent réducteur est la suivante :

$$Fe_2O_3 -> 2Fe + 1{,}5O_2 \qquad \text{(Eq. 4)}$$

Cette réaction nécessite 824,2 kJ/mol et la température à partir de laquelle l'équilibre chimique se déplace dans le sens de la réduction est supérieure à 1800°C. Contrairement à la réaction précédente avec un agent réducteur où la température de réduction baisse à 600°C. L'intérêt de l'agent réducteur est double, réduire l'énergie à apporter au système sous forme de chaleur et abaisser la température de réduction pour rendre la technologie moins complexe et plus robuste.

Si l'envie et le courage vous prend de vérifier ces températures : le diagramme d'Ellingham en Figure 13 permet d'estimer, grâce aux droites représentant l'enthalpie libre des réactions en fonction de la température, le sens préférentiel de l'équilibre chimique : vers l'oxydation si ΔG négatif et vers la réduction si ΔG positif. En ajoutant un agent réducteur, comme le carbone, le monoxyde de carbone ou l'hydrogène, la température à laquelle la réduction est privilégiée diminue à 600°C (intersection entre les courbes d'oxydo-réduction des métaux avec les droites des agents réducteurs). Dans le diagramme d'Ellingham, il est aussi visible que la voie thermique pour la production d'aluminium et de magnésium requiert des températures bien plus élevées que pour le fer.

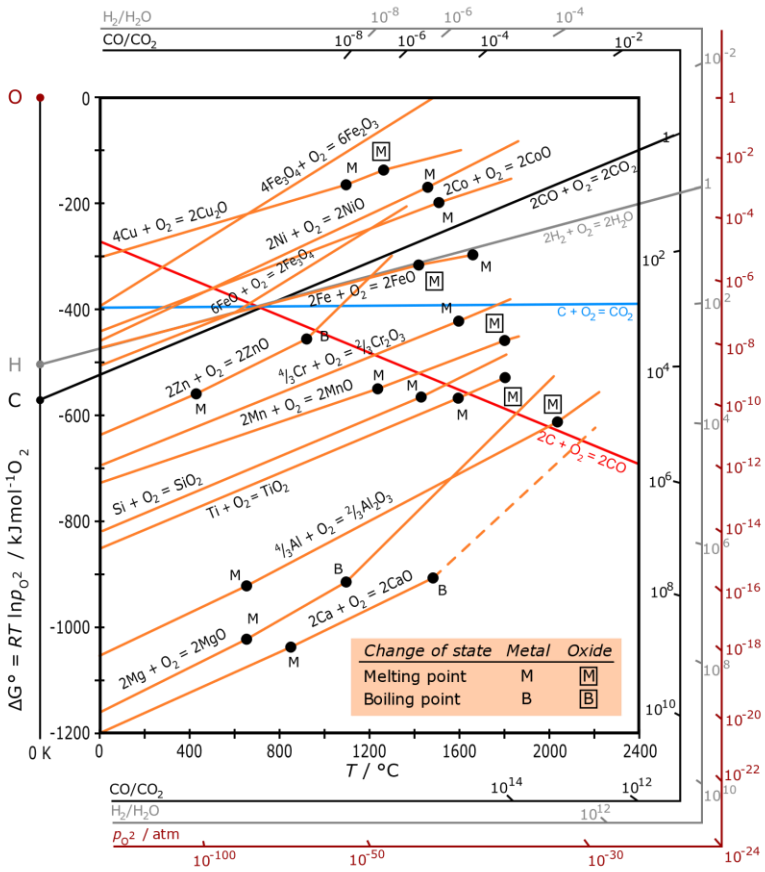


Figure 12 : Diagramme d'Ellingham

Ce diagramme est très utilisé par les métallurgistes pour dimensionner leurs procédés. Pour cause, dans la nature, les métaux se trouvent sous forme oxydées, il y a donc nécessairement d'ores et déjà des procédés de réduction dans la métallurgie pour produire le métal pur que nous retrouvons dans nos bâtiments ou pour nos canettes.

Un des procédés métallurgiques de l'acier vert utilise de la réduction directe à hydrogène pour "purifier" le minerai de fer comme évoqué précédemment et un schéma explicatif du procédé industriel est présenté en figure 14.

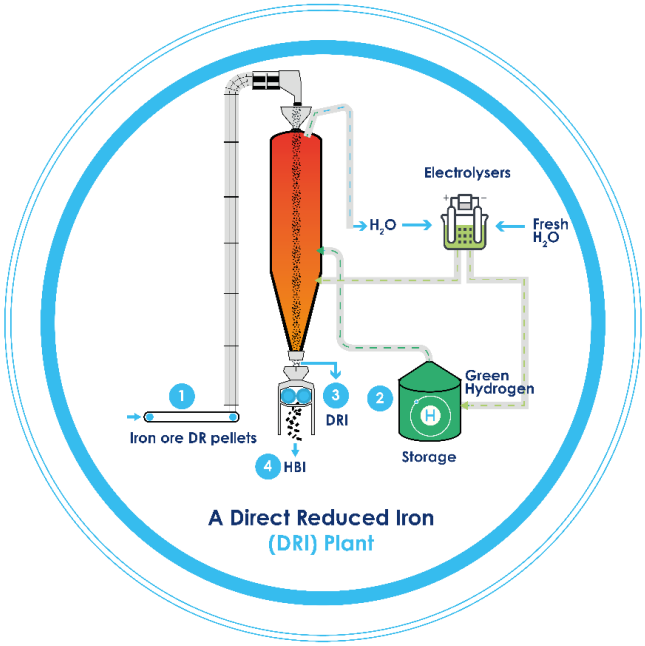


Figure 13 : Procédé DRI produisant du pré-réduit de fer à partir de minerai et d'hydrogène vert

La voie électrolytique à anode inerte est privilégiée pour l'aluminium et le magnésium. On peut citer des projets comme Elysis qui produisent de l'aluminium vert. Pour le fer, la technologie d'électrolyse est encore en phase de maturation technologique, mais plusieurs projets dans le monde essaient de porter cette technologie à l'échelle industrielle (Boston Metals, etc.). Cette technologie, si elle s'avère économiquement viable, pourrait permettre de s'affranchir du besoin en hydrogène dans le cycle de régénération des oxydes de fer. Néanmoins, l'hydrogène vert devient de plus en plus répandu dans le monde, là où les énergies renouvelables abondent, et son utilisation dans la production de fer vert permettrait de faciliter son stockage et son transport.

3 - Conclusion

Parmi les métaux possibles pour servir de vecteur énergétique alternatif (magnésium, aluminium, fer, etc.), le fer s'avère être le plus adapté pour une utilisation énergétique stationnaire à l'échelle mondiale grâce à sa forte production au sein de la bien établie filière de l'acier (2 Gt/an). Une analyse de la Supply Chain circulaire du fer a été réalisée par un groupe de recherche allemand pour évaluer la faisabilité de l'importation de fer vert [22]. La conclusion est en faveur du fer vert pour du transport de renouvelable à moyennes/ longues distances ou pour du stockage à partir de plusieurs semaines (stockage saisonnier, stockage d'autonomie industrielle, réserves stratégiques nationales…).

Ci-dessous, un tableau listant l'ensemble des avantages à utiliser le fer comme vecteur énergétique en écho aux problématiques évoquées dans la partie 1 pour les autres alternatives.

Domaine	Caractéristiques
Impact écologique	Pas d'émissions directe de CO2 sur un cycle
	Très faibles émissions particulaires (captage des oxydes simplifiées par leur taille)
	Pas d'émissions de SOx
	Très faibles émissions de NOx (relativement basses températures de flamme)
	Carburant entièrement recyclable, énergie circulaire, très faible tension sur les ressources métalliques
Sûreté d'utilisation et de transport	Non volatile, et non explosif
	Non corrosif et pas nocif pour l'environnement
	Non toxique pour le corps humain
Impact Économique	Faible coût opérationnel
	Faible coût d'infrastructure de stockage et de transport
	Possibilités de Retrofit de centrales à charbon, chaufferies…
	Faible coût des applications énergétiques
	Valorisation par l'export des ressources en renouvelable des pays en voie de développement
	Faible coût de l'énergie pour les consommateurs
	Stabilisation des prix de l'énergie par la diversification des sources de fer vert
Avantages techniques	Hautes températures (200 -1000°C)
	Pertes de stockage négligeables
	Forte densité énergétique (2 kWh/kg, 5kWh/L)
	Faibles pertes énergétiques, et faibles pertes en masse des ressources
	Utilisation de techniques thermodynamiques très communes

Figure 14 : Tableau rassemblant les avantages certains d'une utilisation du fer, Source : Metalot [15]

Il est clair que la direction la plus rationnelle et équilibrée serait de se diriger vers plus de sobriété énergétique avec en parallèle un développement des solutions de transport et de stockage d'énergies vertes pour contrecarrer le risque d'une décroissance trop lente à l'échelle mondiale (peu envisageable dans les pays en voie de développement).

Il semblerait que les experts de l'énergie soient au moins tous d'accord sur une conclusion un mix énergétique rassemblant plusieurs solutions faisant sens à l'échelle nationale si ce n'est régionale est la meilleure solution. La biomasse là où cela fait sens, dans des quantités qui préservent notre biodiversité et ne concurrence pas notre agriculture, de l'hydrogène consommé à proximité de zones de production, et dans les zones peu fournies en renouvelable abordable, il faut utiliser des vecteurs énergétiques à faible coût de stockage et de transport tel que la poudre de fer.

Des discussions sont en cours avec les institutionnels européens pour développer la filière métallique en Europe et promouvoir son aspect circulaire, sûr et bas carbone. Et plusieurs projets sont en cours pour développer cette filière industriellement à l'échelle globale (Fenix Energy en France, RIFT en Hollande, Ferron Energy en Australie, Hyron Energy au Maroc…), sans oublier les clusters scientifiques Metalot en Hollande et Clean Circles en Allemagne qui aident au développement technologique et économique de la filière. Une entreprise aux Etats-Unis nommée Form Energy a levé des centaines de millions d'euros pour commercialiser des batteries au fer qui utilisent les mêmes réactions en boucle citées précédemment mais uniquement pour de la production d'électricité pour piloter des champs de renouvelables (solution de batteries stationnaires). Les technologies de combustion et de réduction développées par Fenix Energy diffèrent par leur objectif de décarbonation des procédés thermiques de chauffage, de production d'électricité et de chaleur industrielle. Elles permettent également de réaliser des flux d'énergie verte à l'échelle mondiale par le transport de fer comme vecteur énergétique à bas coût.

Références :

[1] https://www.tresor.economie.gouv.fr/Articles/c3a6d87f-5d74-4f2d-b9fc-fe8f4c250511/files/b6dc74cb-5d37-4013-ba99-8f81eaa8b792

[2] https://www.rechargenews.com/energy-transition/vast-majority-of-green-hydrogen-projects-may-require-water-desalination-potentially-driving-up-costs/2-1-1070183

[3] Nouman Rafique Mirza, Sven Degenkolbe, Werner Witt, Analysis of hydrogen incidents to support risk assessment, 2011

[4] Daniel A. Crowl, Young-Do Jo, The hazards and risks of hydrogen, 2007

[5] P.H.C. Lins, A.T. de Almeida, Multidimensional risk analysis of hydrogen pipelines, 2012

[6] https://scfp.ca/lammoniac#:~:text=L'ammoniac%20est%20consid%C3%A9r%C3%A9%20comme,les%20yeux%20et%20les%20poumons.

[7] Irvin Glassman, METAL COMBUSTION PROCESSES, 1969

[8] Cassel HM, Das Gupta AK, Guruswamy S. Factors affecting flame propagation through dust clouds. Symp Combust Flame Explos Phenom 1948

[9] Yabe,« The Magnesium Civilization An Alternative New Source of Energy to Oil »,2008

[10] Direct combustion of recyclable metal fuels for zero-carbon heat and power, J.M. Bergthorson, S. Goroshin, M.J. Soo, P. Julien, J. Palecka, D.L. Frost, D.J. Jarvis, 2015

[11] Ricardo Lomba, Utilisation de la combustion métallique dans les machines thermiques, 2016

[12] Driss Laraqui, Production d'énergie par combustion de magnésium. Caractérisation des émissions gazeuses et particulaires, 2020

[13] Pascal Laboureur, Caractérisation expérimentale d'une flamme prémélangée aluminium / air, 2023

[14] https://www.fenixenergy.fr/

[15] https://www.metalot.nl/pdf/Vision%20document%20Iron%20Power%20juni2022.pdf

[16] J. Janicka, P. Debiagi, A. Scholtissek, A. Dreizler, B. Epple, R. Pawellek, A. Maltsev, C. Hasse, The potential of retrofitting existing coal power plants a case study for operation with green iron, 2023

[17] François-David Tang, Samuel Goroshin, Andrew J. Higgins, Modes of particle combustion in iron dust flames, 2011

[18] Driss Laraqui, Gontrand Leyssens, Cornelius Schonnenbeck, Olivier Allgaier, Ricardo Lomba, Clément Dumand, Jean-François Brilhac, Heat recovery and metal oxide particles trapping in a power generation system using a swirl-stabilized metal-air burner, 2020

[19] Zeldovich Y, Frank-Kamenetskii D, Sadovnikov P. Oxidation of Nitrogen in Combustion. Publishing House of the Acad of Sciences of USSR; 1947.

[20] Driss Laraqui, Olivier Allgaier, Cornelius Schönnenbeck, Gontrand Leyssens, Jean-François Brilhac, Ricardo Lomba, Clément Dumand, Olivier Guézet, Experimental study of a confined premixed metal combustor Metal flame stabilization dynamics and nitrogen oxides production, 2018

[21] Toshio Shudo, Takashi Mizuide, NOx emission characteristics in rich–lean combustion of hydrogen, 2002

[22] JANNIK NEUMANN et al. TECHNO-ECONOMIC ASSESSMENT OF LONG-DISTANCE SUPPLY CHAINS OF ENERGY CARRIERS COMPARING HYDROGEN AND IRON FOR CARBON-FREE ELECTRICITY GENERATION, 2023