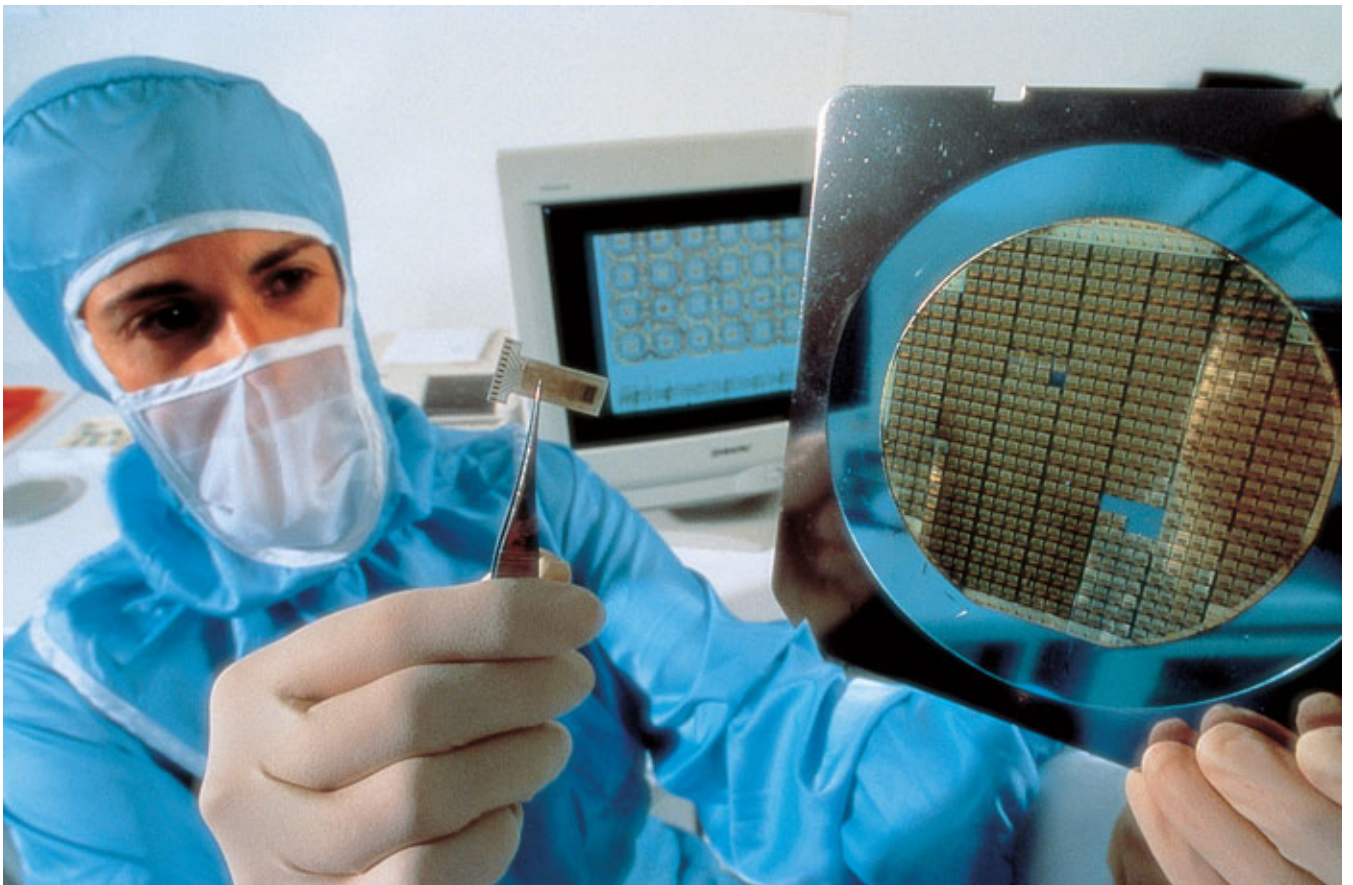


LA SIMULATION DES SYSTÈMES BIOLOGIQUES

Il est désormais possible, avec les puces à ADN, de mesurer l'expression des gènes d'un organisme vivant. Reconstituer à partir de ces données, par une étape de modélisation, le réseau de gènes ainsi exprimés permettrait de le simuler et, par exemple, de prédire le comportement d'une cellule ou le phénotype d'un individu, c'est-à-dire l'expression de ses gènes. La simulation numérique se révèle également cruciale dans le développement technologique des puces elles-mêmes.



CEA/REA/Allard

Puces Micam® multiplexées (format 128 plots), en wafer (couche de silicium, au premier plan à droite) et en montage individualisé (à gauche).



Récemment, les puces à ADN⁽¹⁾ (encadré 1) ont permis une mesure simultanée de l'**expression** de milliers de **gènes** dans une cellule ou un tissu (figure 1). Chez la **levure** par exemple, il est aujourd'hui possible d'analyser l'expression de la totalité du **génom**e (environ 6 300 gènes); chez l'homme, on parvient à analyser simultanément près de 20 000 gènes, soit environ les deux tiers du génome. Cette masse d'informations pourrait permettre de mieux comprendre l'architecture des réseaux biochimiques ou génétiques au sein d'une cellule (figure 2) et la logique de leurs mécanismes régulateurs, qui définissent *in fine* le comportement d'une cellule ou l'établissement d'un phénotype⁽²⁾. Il est ainsi possible de mesurer l'expression *différentielle* d'une partie importante des gènes et de comparer, par exemple, des cellules cancéreuses à des cellules saines.

Si l'on réussit à déduire de ces données le réseau de gènes sous-jacent par une étape de **modélisation**, on pourra simuler ce réseau et prédire l'état de la cellule (cancéreuse ou sain) selon le degré d'expression du gène régulateur principal.

S'il reste important d'étudier individuellement les gènes et les **protéines**⁽¹⁾, il devient possible et intéressant d'étudier la structure et la dynamique des systèmes biologiques par des approches globales. Un réseau génétique étant

(1) Voir encadré 1 dans *La modélisation des macromolécules biologiques*.

(2) Ensemble des caractères observables chez un individu. Pour une cellule, ce sera l'ensemble des caractères apparents résultant de l'expression de ses gènes (type cellulaire, état sanitaire de la cellule, etc.).



Qu'est ce qu'une puce à ADN ?

1

Une puce à **ADN**, ou biopuce (en anglais : *biochip* ou *genechip*) est un dispositif permettant dans son principe de détecter la présence d'un brin d'ADN, la molécule qui sert de support à l'information génétique de tous les êtres vivants, par appariement de ce brin avec son complémentaire, appelé sonde, fixé sur la puce.

Son principe, décrit à la fin des années 80, repose en effet sur la propriété d'**hybridation** de l'ADN, c'est-à-dire la capacité des **bases** d'un brin de cet ADN (thymine, adénine, cytosine et guanine) de reconnaître spontanément des bases complémentaires (thymine et adénine d'une part, cytosine et guanine de l'autre) pour s'apparier comme les deux parties d'une fermeture à glissière et former une double hélice.

La puce peut ainsi identifier une séquence donnée de **nucléotides**, c'est-à-dire l'enchaînement des bases d'un fragment d'ADN en mettant ce dernier en présence d'autres brins dont la séquence est connue. La technique est utilisable dans de nombreuses applications, qui vont du diagnostic et du criblage de nouveaux médicaments et de leurs sites d'action à la détection de contaminants et de polluants en passant par la recherche **génomique** (étude des **mutations** notamment). Une puce à ADN permet de visualiser très rapidement les différences d'expression entre les **gènes** et ceci même à l'échelle d'un génome complet.

Des puces à ADN sont aujourd'hui capables d'identifier en une seule opération jusqu'à des dizaines de milliers d'échantillons d'ADN, voire des centaines de milliers.

Si le principe est simple, la réalisation pratique fait appel à la combinaison de technologies avancées dans les domaines de la micro-électronique, de la chimie des **acides nucléiques**, de l'analyse d'images et de la bio-informatique. Pour la fabrication de la puce, les fragments d'ADN sont amplifiés par la technique de la **polymérisation** en chaîne (PCR) puis fixés (*via* des interactions **électrostatiques**) sur un support de verre, de polymère, de silicium ou de métal pour créer autant de sites d'hybridation.

Lors de la fabrication de la sonde, l'ADN de l'échantillon est **marqué** par un **fluorophore**. Une fois l'hybridation réalisée, chaque plot est excité par un laser, la fluorescence émise, signature de l'hybridation, étant détectée par un microscope à fluorescence. L'analyse des données est en fait une analyse d'image qui permet non seulement de repérer mais aussi de quantifier le signal fluorescent émis par chaque fragment d'ADN.

Le CEA a pour sa part développé avec la société CIS bio International la technique, dite Micam[®], de l'adressage électrochimique⁽¹⁾ dans laquelle la puce est composée d'un support en silicium recouvert à chaque plot d'une mini-électrode en or. Cette technologie, visant le marché des puces à hautes performances, va être mise sur le marché par la société Apibio (créée en 2001 par le CEA et bioMérieux), qui en détient l'exclusivité.

(1) Voir note (3) de *Modélisation phonotique des puces à ADN*.

plus qu'un assemblage de gènes et de protéines, la seule étude de l'architecture des connexions ne permettra pas d'en comprendre toutes les propriétés. L'évolution du système au cours du temps est aussi importante et la dynamique d'interaction des composants doit être examinée.

Deux principes de base

La compréhension d'un réseau génétique repose sur deux principes.

Le premier concerne le *flux d'information génétique*, qui consiste à définir les relations

existantes de l'espace **séquence** vers l'espace *fonction*. Le génome contient les informations qui permettent la construction d'"objets" aussi complexes... qu'un homme. En termes d'information, la complexité d'un organisme complètement développé est contenue dans celle de son génome. Mais quels sont les codes qui traduisent une séquence en structure et en fonction ? Ces codes doivent être représentés sous une forme compréhensible pour qu'il soit possible de les appliquer à la construction de modèles. Les biologistes sont donc à la recherche de méthodes qui permettent de

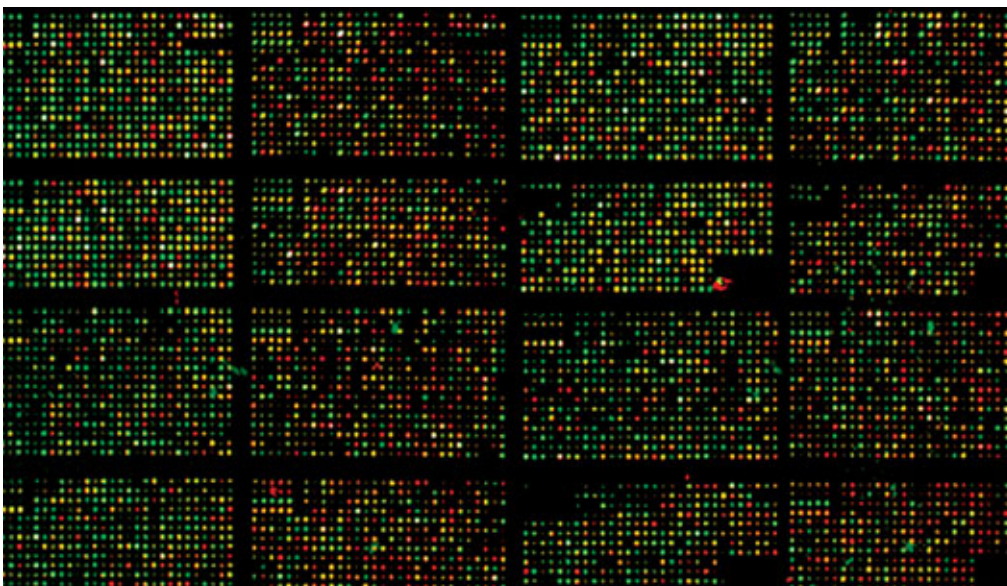


Figure 1. Profil d'expression du génome complet de la levure obtenu avec une puce à ADN. Chaque spot correspond à un gène. Les gènes induits (dont l'expression est activée dans une condition biologique donnée) sont en rouge, les gènes réprimés (dont l'expression est réduite) en vert et les gènes invariants en jaune.

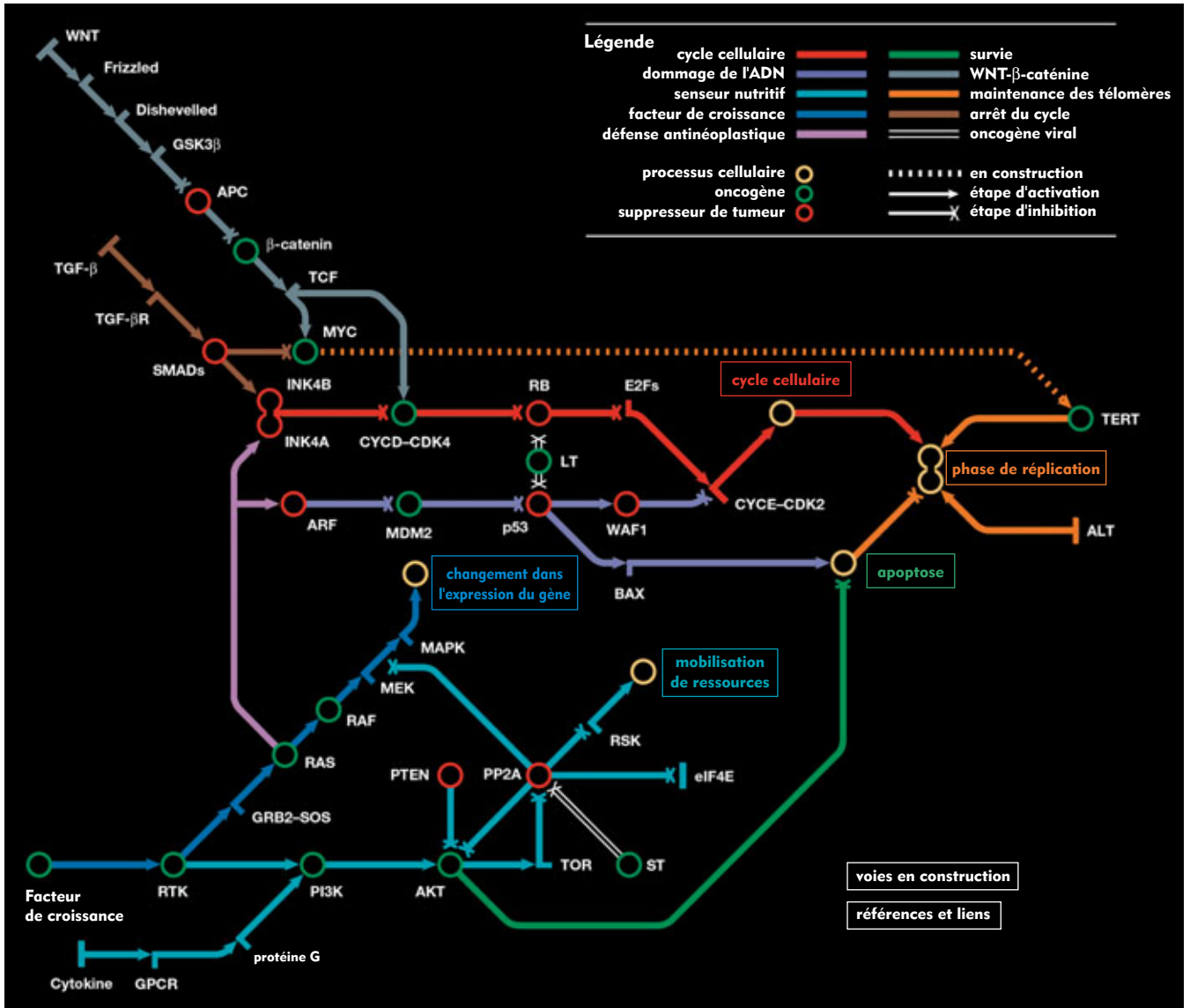


Figure 2. Encore très incomplète, cette carte en "plan de métro" répertorie l'ensemble des gènes impliqués à ce jour dans la cancérisation d'une cellule ainsi que leurs interactions. Il est probable que la génomique permettra d'en découvrir de nombreux autres, à la fonction aujourd'hui inconnue. Un des enjeux de la simulation des réseaux génétiques est de placer ces gènes sur ce réseau et de déterminer la nature de leurs interactions avec le reste de ce réseau (carte établie par William C. Hahn, Robert A. Weinberg et Claudia Bentley, reproduite avec l'autorisation de Nature Reviews Cancer, Vol. 2, n° 5, mai 2002; © 2002 Macmillan Magazines Ltd).



les trouver à partir de la séquence des gènes et des données d'activité obtenues par les puces à ADN. Le second principe est l'étude des *systèmes dynamiques complexes*. Une fois comprise l'architecture d'un réseau génétique, il devient possible d'en étudier la dynamique, c'est-à-dire le comportement du système au cours du temps dans des conditions variées. Ce problème est abordé par la mesure des variations temporelles de l'expression des gènes. L'objectif est de pouvoir prédire l'attracteur d'un réseau génétique, c'est-à-dire le phénotype stable d'une cellule, pour éventuellement diriger ce réseau vers un attracteur choisi : d'une cellule cancéreuse vers une cellule bénigne, d'une cellule vieillissante vers une cellule juvénile.

Cinq paramètres à prendre en compte

Pour l'analyse dynamique d'un réseau génétique, il est nécessaire de créer des modèles. Le choix du modèle est souvent déterminé par la question à laquelle l'expérimentateur

essaie de répondre et le niveau d'abstraction qu'il peut accepter. Cinq paramètres doivent être pris en compte.

Le niveau de détail biochimique

Les modèles peuvent être très abstraits, tels que les réseaux booléens⁽³⁾ ou au contraire très concrets, tels que des modèles d'interactions biochimiques complètes à l'aide des paramètres cinétiques. La première approche permet l'analyse de systèmes de taille considérable. La seconde correspond mieux à la réalité biochimique mais, plus complexe, elle reste limitée à l'étude de petits systèmes. Le besoin se fait

(3) Réseau génétique basé sur la logique booléenne permettant une simplification considérable des interactions entre les gènes. Chaque gène est considéré comme une variable binaire (0 pour un gène exprimé, 1 dans le cas contraire) régulée par d'autres gènes selon des fonctions booléennes.



donc sentir de méthodes pouvant manipuler un très grand nombre de données de manière globale, tout en ayant un niveau de détail acceptable, sans aller jusqu'à la réaction exacte.

Booléen ou continu ?

Les modèles booléens partent du principe que la réponse d'**activation** ou de **répression** des gènes est très abrupte. Or, l'expression des gènes tend à être continue plutôt que binaire. En outre, des concepts indispensables pour les mécanismes de régulation des gènes ne peuvent pas être représentés par des variables booléennes. À titre d'exemple, la rétroinhibition⁽⁴⁾ stabilise un réseau en permettant le contrôle de l'homéostasie⁽⁵⁾ d'une protéine ou d'un gène et réduit la sensibilité aux variations extérieures. Or, dans un circuit booléen, la rétroinhibition provoquera des oscillations plutôt qu'une augmentation de la stabilité.

(4) Inhibition d'une voie de synthèse métabolique par son produit final.

(5) Stabilisation, dans un organisme vivant, des différentes constantes physiologiques en termes de composition chimique, de température et de volume.

Déterministe ou stochastique ?

Une supposition implicite des modèles continus est que les fluctuations d'une molécule unique peuvent être ignorées. Il existe cependant dans les réseaux génétiques de nombreux exemples montrant que la présence d'une seule copie d'**ARN** messager (ARNm)⁽¹⁾ peut jouer un rôle essentiel dans certains processus biologiques qu'il peut être alors impossible de modéliser avec des modèles purement **déterministes**. L'analyse de l'abondance et de la dégradation des ARNm et des protéines pourrait permettre d'identifier les gènes pour lesquels une modélisation **stochastique** s'avérerait nécessaire.

La dimension spatiale

La dimension spatiale peut jouer un rôle majeur au niveau de la compartimentation intracellulaire (noyau, cytoplasme, mitochondrie, membrane), mais aussi à celui des interactions entre cellules. La plupart des processus biologiques chez les organismes multicellulaires (au cours du développement, par exemple), nécessitent des interactions entre différents types cellulaires. La notion d'espace ajoute un niveau de complexité considérable aux modèles. Quelques informations peuvent être extraites de modèles

"non spatiaux", mais à terme il sera nécessaire d'en développer qui intègrent cette dimension.

La disponibilité des données

Un modèle biologique exhaustif devrait prendre en compte la concentration d'ARN, mais aussi la concentration en protéine, sa localisation, etc., puisque chaque variable moléculaire véhicule une information unique sur le fonctionnement cellulaire. Les limitations technologiques de mesure compliquent l'obtention de ces informations, même si les contraintes et la redondance dans les réseaux biologiques suggèrent une possible appréhension du fonctionnement d'un système biologique sans une modélisation de tous les paramètres. À terme, il sera important de développer des outils innovants de mesure simultanée et à haut débit des paramètres moléculaires évoqués plus haut. Bien entendu, la simulation des réseaux génétiques n'aura de sens que si elle permet de faire des prédictions sur les processus biologiques et l'évolution des maladies et, *in fine*, le développement de thérapies efficaces au bénéfice des patients. ●

Xavier Gidrol
Génopole
Évry (Essonne)

Modélisation photonique des puces à ADN

Dans la conception des puces à ADN, la complexité de l'architecture optique et chimique des capteurs impose une phase de simulation numérique importante autour de la technique du marquage fluorescent.

Sur les puces à **ADN** (encadré 1), la reconnaissance de l'**hybridation** de deux brins complémentaires est réalisée par marquage⁽¹⁾ **fluorescent** (encadré 2). Cette méthode de détection, très sensible, est largement appliquée en biologie et en chimie analytique. Dans le cadre du développement technologique des puces à ADN, la reproductibilité et la quantité des signaux émis par les plots est un facteur à prendre en compte pendant la phase de conception et lors de la lecture de la puce. Ce souci revêt une importance majeure dans la mesure où l'analyse de la quantité de lumière fluorescente émise par

(1) Ajout d'un groupement chimique ou d'un atome radioactif à une molécule afin de la suivre et de la localiser par l'intermédiaire de ce marqueur.

La fluorescence

2

Le principe de la fluorescence est simple : une molécule (un **fluorophore**) absorbe un photon d'énergie donnée puis réémet en général très rapidement un photon d'énergie inférieure. Cette perte d'énergie s'explique par l'agitation thermique induite ou encore par la modification de la structure de la molécule.

En conséquence, la couleur de la lumière d'émission se différencie de celle de l'excitation : ceci permet le **marquage** en séparant ces couleurs par un dispositif optique approprié (un prisme étant l'exemple le plus simple).

La détection par fluorescence est très sensible, puisque, avec les progrès opérés au cours de ces dernières années dans le domaine des détecteurs et des capteurs d'images, il est possible de distinguer une molécule unique.

Malheureusement, la fluorescence est extrêmement sensible à des paramètres comme l'environnement chimique des fluorophores pouvant éteindre la fluorescence (ce phénomène est généralement appelé le "*quenching*") ou la trop forte exposition à la lumière excitatrice qui peut détruire les marqueurs (phénomène de "*photoblanchiment*").

chacun des plots de la biopuce doit être reliée au nombre de brins d'ADN s'étant hybridés, dans le but de déterminer le taux d'expression des gènes à analyser. Bien qu'un modèle analytique permette de comprendre la physique du problème, la complexité de l'architecture optique et chimique des capteurs à concevoir impose une phase de simulation numérique importante (encadré A, *Qu'est-ce qu'une simulation numérique ?*). Celle-ci tient ses informations de campagnes de mesures réalisées avec des outils spécifiques afin de disposer de produits fiables, sensibles et réalisables de manière industrielle.

Un modèle analytique pour comprendre les tendances

Il s'agit de modéliser de la manière la plus simple et la plus pertinente possible le comportement de la fluorescence sur des surfaces planes constituées d'un substrat (lame de microscope, *wafers*⁽²⁾ de silicium) pouvant être revêtu de couches minces minérales (silices par exemple) ou organiques. Ces couches ont notamment pour but d'assurer un bon greffage de l'ADN sur la surface de la "biopuce".

L'objectif est d'identifier les paramètres intervenant au premier ordre dans le comportement optique des biopuces. À cette fin, les fluorophores (encadrés 1 et 2) sont assimilés à de petites sources de lumière ayant la propriété de posséder une orientation particulière : des dipôles. La lumière qui excite le marqueur fluorescent induit une oscillation du nuage électronique et la direction de cette oscillation est imposée par la structure atomique de la molécule. Le couplage entre la source et la surface est traité de manière analytique par un formalisme reposant sur une approche électromagnétique du rayonnement lumineux : on peut assimiler le dipôle à une antenne émettant près du sol, comme l'avait fait le physicien allemand Arnold Sommerfeld, il y a... 90 ans !

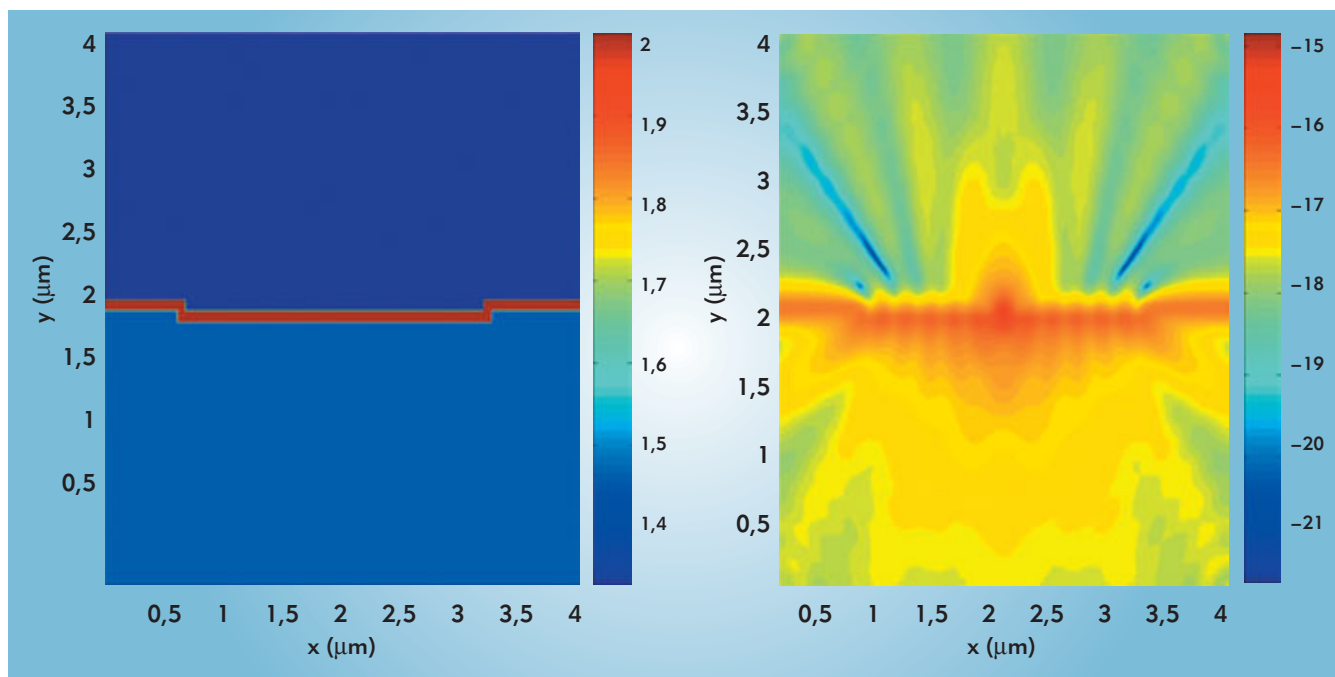
Cette étude préliminaire a permis aux chercheurs de découpler les phénomènes d'émission de la lumière, liés aux caractéristiques optiques et géométriques de la biopuce, des effets liés à son environnement chimique et biologique. L'observation des signaux émis par des biopuces réalisées sur verre ou sur silicium a pu le confirmer expérimentalement.

Prévoir les signaux sur des systèmes complexes

L'évolution technologique des biopuces tend, de manière de plus en plus affirmée, vers des microsystèmes complexes. Une demande toujours plus grande en termes de miniaturisation, d'intégration de fonctionnalités comme l'adressage des plots par des transistors CMOS⁽³⁾ (technologie Micam[®]) l'impose. Comprendre et prédire les signaux de fluorescence sur ces composants devient une tâche nettement plus délicate que sur des puces classiques. Étudier les caractéristiques de la fluorescence à proximité de nombreux matériaux sur des surfaces structurées nécessite, de par la complexité du problème, une approche numérique. Pour répondre à ce besoin, les chercheurs du CEA ont réalisé un outil de simulation permettant de dresser le bilan photonique d'une biopuce : ils résolvent sans approximations les équations de l'électromagnétisme décrivant l'interaction du fluorophore, toujours assimilé à un dipôle, avec un environnement quelconque.

(2) Tranche de matériau (le plus souvent du silicium) dans laquelle sont élaborés avant d'être découpés les circuits intégrés et autres composants et dispositifs micro-électroniques.
 (3) Sur les biopuces de type Micam[®], il est associé à chacun des plots un transistor CMOS (circuit intégré à haute densité utilisant à la fois des transistors de type N et de type P qui leur permet de ne pratiquement pas consommer d'énergie en dehors des commutations). Il est alors possible de les activer et, par un phénomène proche d'une réaction d'électrodéposition, de disposer spécifiquement sur ces plots les brins d'ADN voulus.

● ● ● ● ●
 Figure 1. Analyse de la fluorescence émise sur une piste Lightscan[®]. La figure de gauche indique l'intensité de la lumière fluorescente (échelle logarithmique) dans la structure dont la géométrie est représentée à droite.



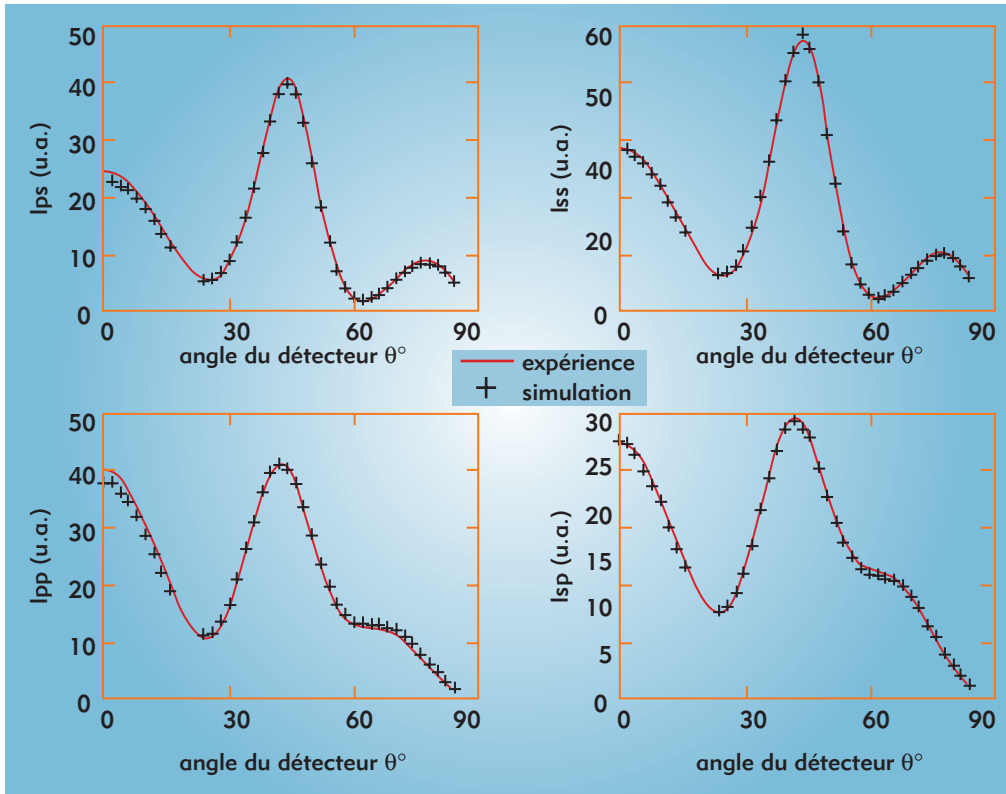


Figure 2. Comparaison théorie/expérience de l'intensité lumineuse émise par des biopuces à ADN (réalisées sur un substrat de silicium présentant une couche mince de silice de quelques centaines de nanomètres d'épaisseur) en fonction de l'angle de mesure. Les courbes correspondent à une analyse de la polarisation de la lumière fluorescente selon deux directions orthogonales ("s" et "p") après éclairage de la biopuce avec une lumière polarisée de manière rectiligne selon ces deux directions.



Application à un cas technologique concret

Une des grandes applications de cet outil a été la modélisation du comportement de la fluorescence sur les biopuces Lightscan[®] codéveloppées par la société bioMérieux et le Leti. Ce système revient à réaliser un lecteur de biopuce inspiré... d'un lecteur de compact disc, solution qui permet de réduire le coût d'un facteur 5 à 10 par rapport à un dispositif classique. Comme un disque compact, les biopuces doivent présenter des pistes afin d'assurer un positionnement très précis de la tête de lecture pendant son déplacement. Les scientifiques ont ainsi analysé le comportement de la fluorescence sur les pistes de ces biopuces qui présentent alors une structuration du verre (quelques dizaines de nanomètres de profondeur pour une largeur de plusieurs microns) sur laquelle est déposée une

couche mince de quelques dizaines de nanomètres (figure 1). La simulation photonique permet de montrer que la lumière émise par le fluorophore est piégée dans la couche mince sous forme d'ondes guidées (comme dans les fibres optiques) et en partie libérée au niveau des structurations.

Une approche expérimentale indispensable

La modélisation seule ne peut garantir la maîtrise des procédés technologiques. Il est indispensable d'en caractériser les résultats avec un outil métrologique dédié. Le développement d'un goniofluorimètre⁽⁴⁾ a permis en l'occurrence de valider les travaux de simulation (figure 2). L'examen critique des *indicatrices* mesurées en fluorescence, en prenant compte des effets de polarisation⁽⁵⁾ de la lumière, a permis de mettre en cause l'approche analytique et d'améliorer la modélisation. Les chercheurs disposent à présent d'une plate-forme de caractérisation optique qui les rend aptes à envisager une approche de quantification et proposer des standards de fluorescence.

Il leur reste à prendre en compte des phénomènes de couplage énergétique entre le fluorophore et le rayonnement incident, ainsi que des phénomènes environnementaux tels que le "photoblanchiment" ou le "quenching" (encadré 2). ●

(4) Un goniofluorimètre (néologisme) permet de déterminer l'intensité de la lumière de fluorescence pour une direction donnée dans l'espace. Le dispositif développé au CEA-Leti permet aussi d'analyser l'état de polarisation de la lumière fluorescente.

(5) La théorie de la polarisation permet de décrire l'orientation (en termes d'amplitude et de variation temporelle) du vecteur champ électrique décrivant une vibration lumineuse. Dans le cas de la fluorescence, l'étude de la polarisation de la lumière permet d'analyser des phénomènes comme le changement de la structure de la molécule entre l'excitation et la réémission de lumière.

Stéphane Gétin

Direction de la recherche technologique
CEA centre de Grenoble

Qu'est-ce qu'une simulation numérique ?

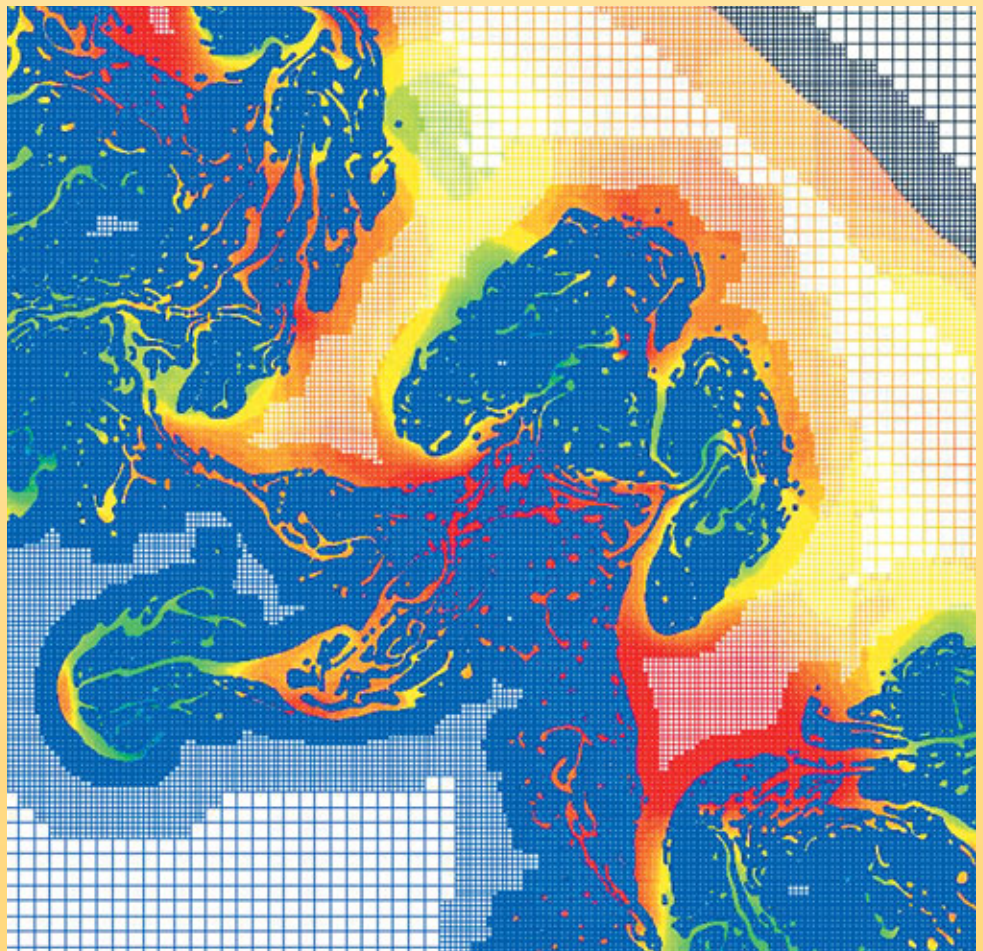
La **simulation numérique** consiste à reproduire par le calcul le fonctionnement d'un système, préalablement décrit par un ensemble de **modèles**. Elle s'appuie sur des méthodes mathématiques et informatiques spécifiques. Les principales étapes de la réalisation d'une étude par simulation numérique sont communes à de nombreux secteurs de la recherche et de l'industrie, en particulier le nucléaire, l'aérospatial ou l'automobile.

En chaque point de l'"objet" considéré, plusieurs grandeurs physiques (vitesse, température...) décrivent l'état et l'évolution du système étudié. Celles-ci ne sont pas indépendantes, mais reliées et régies par des **équations**, généralement **aux dérivées partielles**. Ces équations constituent la traduction mathématique des lois de la physique qui modélisent le comportement de l'objet. Simuler l'état de ce dernier, c'est déterminer – idéalement en tout point – les valeurs numériques de ses **paramètres**. Comme il y a un nombre infini de points, donc une infinité de valeurs à calculer, cet objectif est inaccessible (sauf dans des cas bien particuliers où l'on peut résoudre les équations de départ à l'aide de formules analytiques). Une approximation naturelle consiste donc à ne considérer qu'un nombre fini de points. Les valeurs des paramètres à calculer sont ainsi en nombre fini et les opérations nécessaires deviennent abordables grâce à l'ordinateur. Le nombre effectif de points traités dépendra bien sûr de la puissance de celui-ci : plus il sera élevé, meilleure sera finalement la description de l'objet. À la base du calcul des paramètres comme à la base de la simulation numérique, il y a donc la réduction de l'infini au fini, la **discrétisation**.

Comment opère-t-on précisément à partir des équations mathématiques du modèle ? Deux méthodes sont très souvent utilisées, respectivement représentatives des méthodes de **calcul déterministe**, qui résolvent les équations régissant les phénomènes étudiés après avoir discrétisé les variables, et des méthodes de **calcul statistique** ou **probabiliste**.

Le principe de la première, connue sous le nom de **méthode des volumes finis**, est antérieur à l'usage des ordinateurs. Chaque point de l'objet est assimilé simplement à un petit volume élémentaire (un cube par exemple), d'où le nom de *volume fini*. Un plasma, par exemple, est ainsi vu comme un ensemble ou un réseau de volumes contigus qui, par analogie avec la trame d'un tissu, sera dénommé **maillage**. Les paramètres de l'état de l'objet sont maintenant définis dans chaque maille du maillage. Pour chacune d'elles, en reformulant les équations mathématiques du modèle par des moyennes volumiques, il sera alors possible de construire des *relations algébriques* entre les paramètres de la maille et ceux de ses voisins. Au total, il y aura autant de relations que de paramètres inconnus et ce sera à l'ordinateur de résoudre le *système* de relations obtenu. Il faudra pour cela recourir aux techniques de l'**analyse numérique** et programmer des **algorithmes** spécifiques.

L'accroissement de la puissance des ordinateurs a permis d'augmenter la finesse de discrétisation, permettant de passer de quelques dizaines de mailles dans les années soixante à plusieurs dizaines de milliers dans les années quatre-vingt, à des millions dans les années quatre-vingt-dix et jusqu'à la dizaine de milliards de mailles aujourd'hui (machine Tera de la Direction



Exemple d'image d'une simulation 2D d'instabilités réalisée avec le supercalculateur Tera du CEA. Le calcul a fait appel au maillage adaptatif, qui se fait plus fin dans les zones où les phénomènes sont les plus complexes.

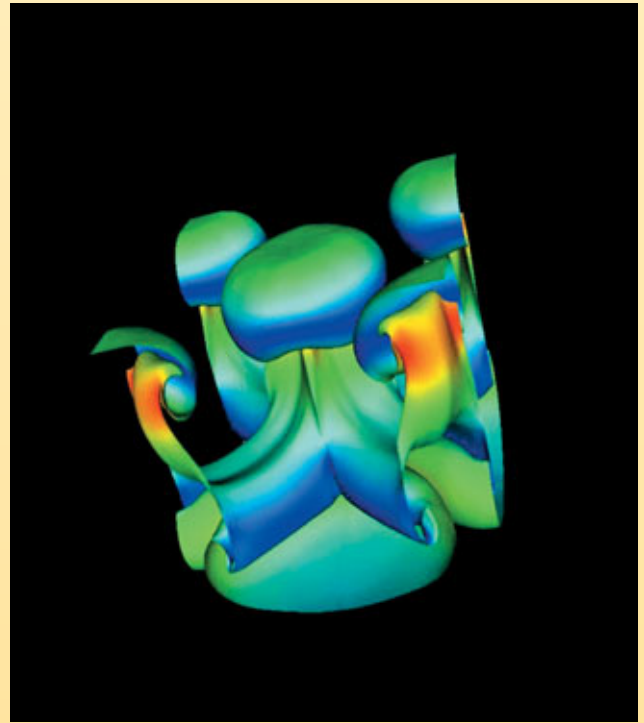
des applications militaires du CEA), chiffre qui devrait décupler à la fin de la décennie.

Un raffinement du maillage, le **remaillage adaptatif**, consiste à ajuster la taille des mailles en fonction des circonstances, par exemple en les rendant plus petites et plus serrées aux interfaces entre deux milieux, là où les phénomènes physiques sont les plus complexes, ou là où les variations sont les plus importantes.

La méthode des volumes finis s'applique dans des contextes physiques et mathématiques très variés. Elle autorise toute forme de maille (cube, hexaèdre, tétraèdre...) et le maillage peut être modifié durant le calcul, en fonction de critères géométriques ou physiques. Enfin, elle est aisée à mettre en œuvre dans le contexte des **ordinateurs parallèles** (encadré B, **Les moyens informatiques de la simulation numérique hautes performances**), le maillage pouvant en effet faire l'objet d'un découpage pour des calculs sur ce type de machines (exemple figure B, p. 13).

Appartiennent à la même famille la **méthode des différences finies**, cas particulier de la **méthode des volumes finis** où les côtés des mailles sont orthogonaux, et la **méthode aux éléments finis**, qui peut juxtaposer divers types de mailles. La deuxième grande méthode, dite de **Monte-Carlo**, est particulièrement adaptée pour simuler le *transport de particules*, par exemple des neutrons ou des photons d'un **plasma** (voir *Les simulations en physique des particules*). Un tel transport est en fait caractérisé par une succession d'étapes lors desquelles chaque particule peut subir différents événements (diffusion, absorption, émission...) possibles *a priori*. Les probabilités élémentaires de chacun de ces événements sont connues individuellement pour chaque particule.

Il est alors naturel d'assimiler un point du plasma à une particule. Un ensemble de particules, en nombre fini, va constituer un échantillon représentatif de l'infinité de particules du plasma, comme lors d'un sondage statistique. D'étape en étape, l'évolution de l'échantillon sera déterminée grâce à des tirages au hasard (d'où le nom de la méthode). L'efficacité de cette méthode, mise en œuvre à Los Alamos dès les années 1940, dépend bien sûr de la qualité statistique des tirages au hasard.



Simulation 3D réalisée à l'aide du supercalculateur Tera installé fin 2001 au centre CEA/DAM Île-de-France à Bruyères-le-Châtel (Essonne).

Il existe pour cela des méthodes de nombres aléatoires, bien adaptées au traitement par un ordinateur.

Les méthodes des volumes finis et de Monte-Carlo ont suscité et suscitent de nombreuses études mathématiques. Ces études s'attachent notamment à préciser la convergence de ces méthodes, c'est-à-dire comment la précision de l'approximation varie avec le nombre de mailles ou de particules. Cette question est naturelle lors de la confrontation des résultats de la simulation numérique à ceux de l'expérience.

Comment se déroule une simulation numérique ?

Il est souvent question d'*expérience numérique* pour souligner l'analogie entre la pratique d'une simulation numérique et la conduite d'une expérience de physique.

Brièvement, cette dernière utilise un dispositif expérimental, configuré selon des conditions initiales (de température, de pression...) et des paramètres de contrôle (durée de l'expérience, des mesures...). Durant l'expérience, le dispositif produit des points de mesures qui sont enregistrés. Ces enregistrements sont ensuite analysés et interprétés.

Dans une simulation numérique, le dispositif expérimental consiste en un ensemble de programmes informatiques exécutés sur des ordinateurs. Les **codes** ou **logiciels de calcul** sont la traduction, à travers des algorithmes numériques, des formulations mathématiques des modèles physiques étudiés. En amont et en aval du calcul, les *logiciels d'environnement* effectuent la gestion de plusieurs opérations complexes de préparation des calculs et de leur dépouillement.

Les données initiales de la simulation comporteront d'abord la délimitation du domaine de calcul à partir d'une représentation approchée des formes géométriques (produite par le dessin et la CAO, conception assistée par ordinateur), suivie de la discrétisation de ce

domaine de calcul sur un maillage, ainsi que les valeurs des paramètres physiques sur ce maillage et les paramètres de contrôle du bon déroulement des programmes... Toutes ces données (produites et gérées par les logiciels d'environnement) seront saisies et vérifiées par les codes. Les résultats des calculs proprement dits, c'est-à-dire les valeurs numériques des paramètres physiques, seront sauvegardés au fur et à mesure. En fait, un protocole spécifique structurera les informations produites par l'ordinateur afin de constituer une base de données numériques.

Un protocole complet organise l'échange informatique des informations requises (dimensions notamment) suivant des formats prédéfinis : *modeleur*⁽¹⁾, *mailleur*⁽²⁾, *découpeur de maillage*, *code*

- (1) Le *modeleur* est un outil qui permet la création et la manipulation de points, courbes et surfaces en vue par exemple de la création d'un maillage.
- (2) Les formes géométriques d'un maillage sont décrites par des ensembles de points reliés par des courbes et des surfaces (de Bézier par exemple) qui en représentent les frontières.



de calculs, logiciel de visualisation et d'analyse. Les études de *sensibilité* des résultats (au maillage et aux modèles) font partie des "expériences" numériques.

À l'issue des calculs (résolution numérique des équations décrivant les phénomènes physiques qui se déroulent dans chaque maille), l'analyse des résultats par des spécialistes reposera sur l'exploitation de la base de données numériques. Elle comportera plusieurs étapes : extraction sélective des données (selon le paramètre physique recherché) et visualisation, extraction et transfert des données pour calculer et visualiser des diagnostics.

Le parallèle entre la conduite d'un cas de calcul, d'une expérience numérique et la conduite d'une expérience physique ne s'arrête pas là : les résultats numériques seront comparés aux résultats expérimentaux. Cette analyse comparative, effectuée sur la base de critères quantitatifs standardisés, fera appel et à l'expérience

et à l'art de l'ingénieur, du physicien, du mathématicien. Elle débouchera sur de nouvelles améliorations des modèles physiques et des programmes informatiques de simulation.

Bruno Scheurer

Direction des applications militaires
CEA centre DAM-Ile de France

Frédéric Ducros et Ulrich Bieder

Direction de l'énergie nucléaire
CEA centre de Grenoble

L'exemple d'un calcul de thermohydraulique

La mise en œuvre d'un protocole de simulation numérique peut être illustrée par les travaux réalisés par l'équipe de développement du logiciel de calcul **thermohydraulique** Trio U. Ces travaux se sont déroulés dans le cadre d'une étude faite en collaboration avec l'Institut de radioprotection et de sûreté nucléaire (IRSN). L'objectif était d'obtenir des données très précises pour fournir à l'ingénieur les sollicitations en température à la paroi des composants d'un réacteur à eau sous pression dans le cas d'un accident grave impliquant une circulation naturelle turbulente de gaz chauds. Cette étude requiert la modélisation simultanée d'effets "système" à grande échelle et de phénomènes **turbulents** à petite échelle (encadré F, **Modélisation et simulation des écoulements turbulents**).

Elle débute par la définition du modèle de calcul global (figure A), suivie de la réalisation du modèle CAO et du maillage correspondant avec des logiciels du commerce (figure B). Les maillages de plus de cinq millions de mailles exigent l'utilisation de puissantes stations graphiques. Dans cet exemple, le maillage d'un générateur de vapeur (figures C et D) a été découpé pour répartir les calculs sur huit processeurs d'un ordinateur parallèle du CEA : chaque couleur symbolise une zone affectée à un processeur particulier. Les calculs, dont les conditions aux limites sont données par un calcul "système" (Icare-Cathare), produisent des résultats qu'il appartient aux spécialistes d'interpréter. En l'occurrence, les visualisations sur des stations graphiques des valeurs instantanées des champs de vitesse montrent l'impact d'un panache chaud sur la plaque tubulaire du

générateur de vapeur (coupe dans le champ de vitesses à gauche de la figure E) et la température instantanée dans la boîte à eau (à droite).

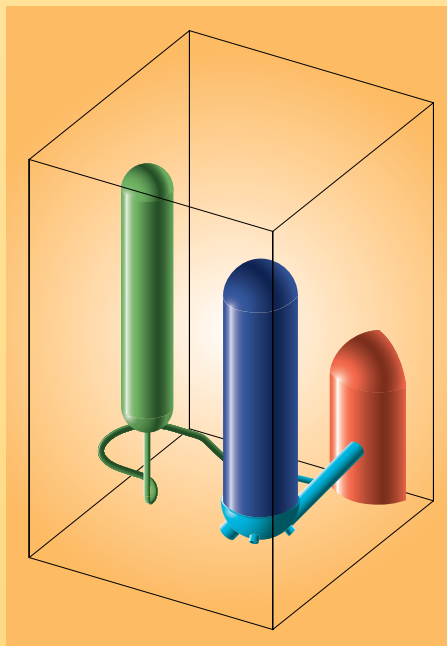


Figure A. Domaine de calcul global incluant une partie de la cuve réacteur (rouge), la conduite de sortie (branche chaude en bleu clair), le générateur de vapeur (bleu foncé) et le pressuriseur (vert).

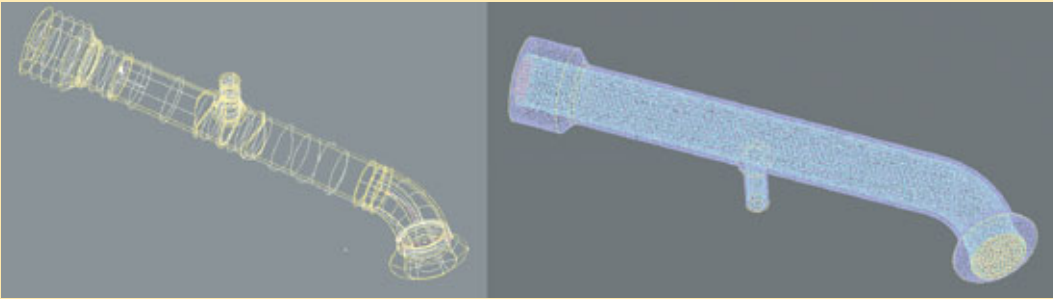
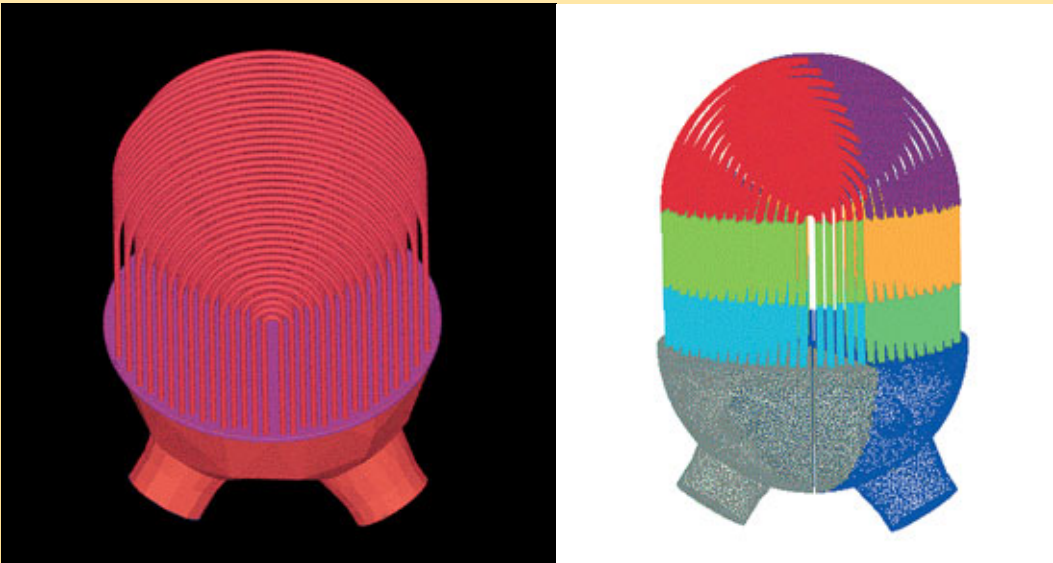


Figure B. Modèle CAO de la branche chaude en sortie de la cuve réacteur (à gauche) et son maillage non structuré (à droite).



Figures C et D.

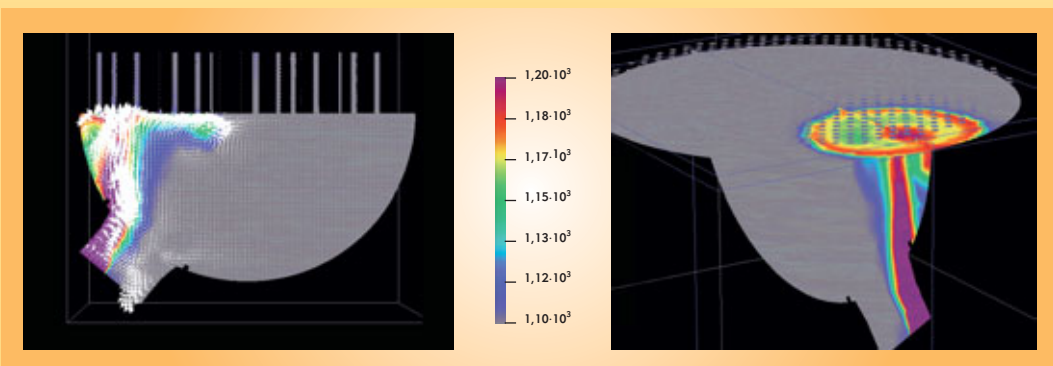


Figure E.

Les moyens informatiques de la simulation

Effectuer des **simulations numériques** plus précises impose de mettre en œuvre des **modèles** physiques et numériques eux-mêmes plus précis portant sur des descriptions plus fines des objets simulés (encadré A, *Qu'est-ce qu'une simulation numérique ?*). Tout ceci nécessite des progrès dans le domaine des logiciels de simulation mais aussi une augmentation importante de la capacité des équipements informatiques sur lesquels ces logiciels sont utilisés.

Processeurs scalaires et vectoriels

Au cœur de l'ordinateur, le processeur est l'unité de base qui, déroulant un programme, effectue les calculs. Il en existe deux grands types, les **processeurs scalaires** et les **processeurs vectoriels**. Les premiers exécutent des opérations portant sur des nombres élémentaires (scalaires), par exemple l'addition de deux nombres. Les seconds exécutent des opérations portant sur des ensembles de nombres (vecteurs), par exemple additionner deux à deux les nombres composant deux ensembles de 500 éléments. À ce titre, ils sont particulièrement adaptés à la simulation numérique : lors de l'exécution d'une opération de ce type, un processeur vectoriel peut fonctionner à une vitesse proche de sa performance maximale (crête). La même opération avec un processeur scalaire exige de nombreuses opérations indépendantes (opérations par composante des vecteurs) qui s'exécutent à une vitesse bien inférieure à sa vitesse crête. L'avantage principal des processeurs scalaires est leur prix : il s'agit de microprocesseurs généralistes dont les coûts de conception et de fabrication peuvent être amortis sur de larges marchés.

Forces et contraintes du parallélisme

Les processeurs récents permettent de hautes performances, d'une part en utilisant une fréquence de fonctionnement plus élevée, d'autre part en cherchant à exécuter en même temps

plusieurs opérations : c'est un premier niveau de **parallélisme**. L'accélération de la fréquence est limitée par l'évolution de la technologie micro-électronique, tandis que les dépendances entre instructions à exécuter par le processeur limitent le parallélisme possible. La mise en œuvre simultanée de plusieurs processeurs constitue un second niveau de parallélisme, qui permet d'obtenir des performances accrues à condition de disposer de programmes capables d'en tirer parti. Alors que le parallélisme au niveau des processeurs est automatique, celui *entre processeurs* dans un ordinateur parallèle est à la charge du programmeur, qui doit découper son programme en morceaux indépendants et prévoir entre eux les communications nécessaires. On procède souvent par un découpage du domaine sur lequel porte le calcul, chaque processeur étant chargé de simuler le comportement d'un domaine, et par l'établissement de communications régulières entre processeurs afin de garantir la cohérence d'ensemble du calcul. Pour obtenir un programme parallèle efficace, il faut s'assurer de l'équilibrage de charge entre processeurs et chercher à limiter le coût des communications.

Les différentes architectures

Les équipements informatiques ont différentes fonctions. À partir de son ordinateur de travail sur lequel il prépare ses calculs et en analyse les résultats, l'utilisateur accède à des moyens de calcul, de stockage, et de visualisation partagés, mais beaucoup plus puissants que les siens propres. L'ensemble de ces équipements sont reliés par des réseaux informatiques permettant de faire circuler les informations entre eux avec des débits compatibles avec le volume de données produites, pouvant atteindre 1 **téraoctet** (1 To = 10^{12} octets) de données pour une seule simulation.

Les grands équipements de calcul sont généralement appelés **supercalculateurs**. Ils atteignent aujourd'hui des puissances qui se chiffrent en **téraflows** (1 Tflops = 10^{12} opérations de calcul par seconde).

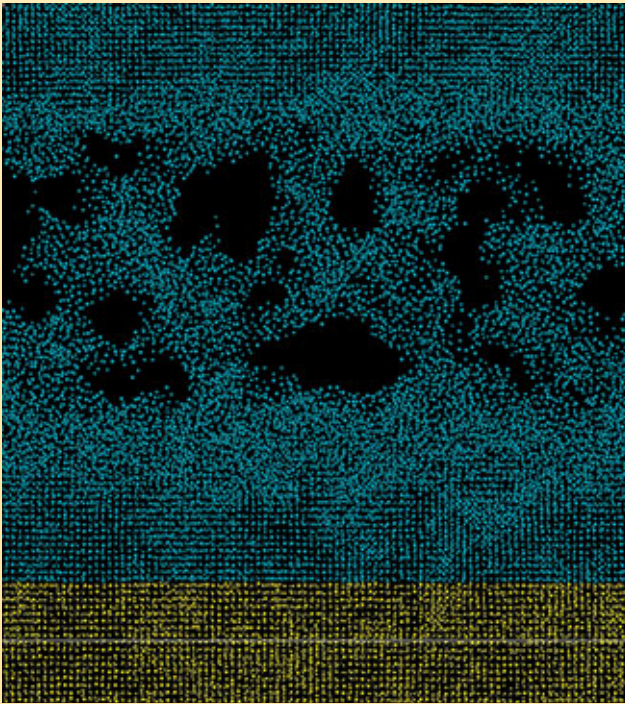
Il existe aujourd'hui trois grands types de supercalculateurs : les supercalculateurs vectoriels, les grappes de mini-ordinateurs à mémoire partagée et les grappes de PC (l'ordinateur que chacun possède chez soi). Le choix entre ces architectures dépend largement des applications et de l'utilisation visées. Les supercalculateurs vectoriels disposent de processeurs très performants mais dont il est difficile d'augmenter la puissance en ajoutant des processeurs. Les grappes de PC sont peu coûteuses mais mal adaptées à des environnements où de nombreux utilisateurs font beaucoup de calculs très gourmands en puissance machine, en mémoire et en entrées-sorties.

Ce sont ces considérations qui ont en particulier conduit la Direction des applications militaires (DAM) du CEA à choisir pour son programme simulation (voir *Le programme Simulation : la garantie des armes sans essais nucléaires*) les architectures de type grappe de mini-ordinateurs à mémoire partagée, encore appelées **clusters de SMP** (Symmetric Multi-Processor). Un tel système utilise comme brique de base un mini-ordinateur com-



CEA

Installée en décembre 2001 au CEA (centre DAM-Ile de France) et conçue par Compaq (devenue depuis HP), la machine Tera a pour élément de base un mini-ordinateur à 4 processeurs Alpha à 1 GHz partageant une mémoire de 4 Go et fournissant une puissance totale de 8 Gflops. Ces éléments de base sont interconnectés par un réseau rapide conçu par la société Quadrics. Une opération de synchronisation sur l'ensemble des 2 560 processeurs s'effectue en moins de 25 microsecondes. Le système de fichiers global offre un espace de stockage de 50 téraoctets pour les entrées-sorties avec une bande passante agrégée de 7,5 Go/s.



CEA

Les calculateurs parallèles sont adaptés aux méthodes numériques basées sur des maillages (encadré A, **Qu'est-ce qu'une simulation numérique ?**) mais aussi au traitement de calculs ab initio comme cette simulation par dynamique moléculaire de l'endommagement par choc de deux plaques de cuivre à 1 km/s (voir La simulation des matériaux). Le système considéré est constitué de 100 000 atomes de cuivre représentant un parallélépipède de section carrée (0,02 µm de côté) à densité normale. Les atomes interagissent suivant un potentiel EAM (embedded atom potential) pendant 4,6 picosecondes. Le calcul, effectué sur 18 processeurs du supercalculateur Tera de Bruyères-le-Châtel à l'aide du logiciel Stamp développé au CEA, a représenté une dizaine de minutes de temps "utilisateur" (calcul réalisé par B. Magne). Des tests impliquant jusqu'à 64 millions d'atomes ont été réalisés, mobilisant 256 processeurs pendant une centaine d'heures.

portant plusieurs microprocesseurs qui partagent une mémoire commune (figure). Ces mini-ordinateurs étant largement diffusés dans des domaines variés allant de la banque au serveur web

en passant par les bureaux d'études, ils offrent un excellent rapport performance/prix. Ces "briques" de base (encore appelées *nœuds*) sont reliées entre elles par un réseau d'interconnexion hautes performances : la puissance cumulée de plusieurs centaines de ces "briques" peut atteindre plusieurs téraflops. On parle alors d'**ordinateur massivement parallèle**.

Cette puissance peut être disponible pour une seule application parallèle utilisant toutes les ressources du supercalculateur mais aussi pour de multiples applications indépendantes, parallèles ou non, utilisant chacune une partie des ressources.

Si la caractéristique mise en avant pour décrire un supercalculateur est en général sa puissance de calcul, il ne faut pas négliger l'aspect entrées-sorties. Ces machines capables d'effectuer des simulations de grande taille doivent disposer de systèmes de disques avec des capacités et des performances adaptées. Dans les *clusters* de SMP, chaque mini-ordinateur dispose d'un espace disque local. Il n'est néanmoins pas judicieux d'utiliser celui-ci pour les fichiers utilisateurs, ce qui obligerait l'utilisateur à explicitement déplacer ses données entre les différentes phases de ses calculs. Pour cette raison, il est important de disposer d'un espace disque accessible par l'ensemble des mini-ordinateurs du supercalculateur. Cet espace est en général constitué de batteries de disques reliées à des nœuds dont la fonction principale est de les gérer. Comme pour le calcul, c'est le parallélisme pour les entrées-sorties qui permet d'offrir des performances élevées. Il faut, pour ce faire, disposer de systèmes de fichiers globaux parallèles permettant un accès rapide et sans contraintes à l'espace disque partagé.

Offrant des puissances de calcul considérables, les *clusters* de SMP posent néanmoins plusieurs défis. Parmi les plus importants, outre la programmation de logiciels de simulation capables de tirer parti du grand nombre de processeurs, il faut mettre au point des systèmes d'exploitation et les logiciels associés compatibles avec de telles configurations et tolérants vis-à-vis des pannes.

François Robin

Direction des applications militaires
CEA centre DAM-Ile de France

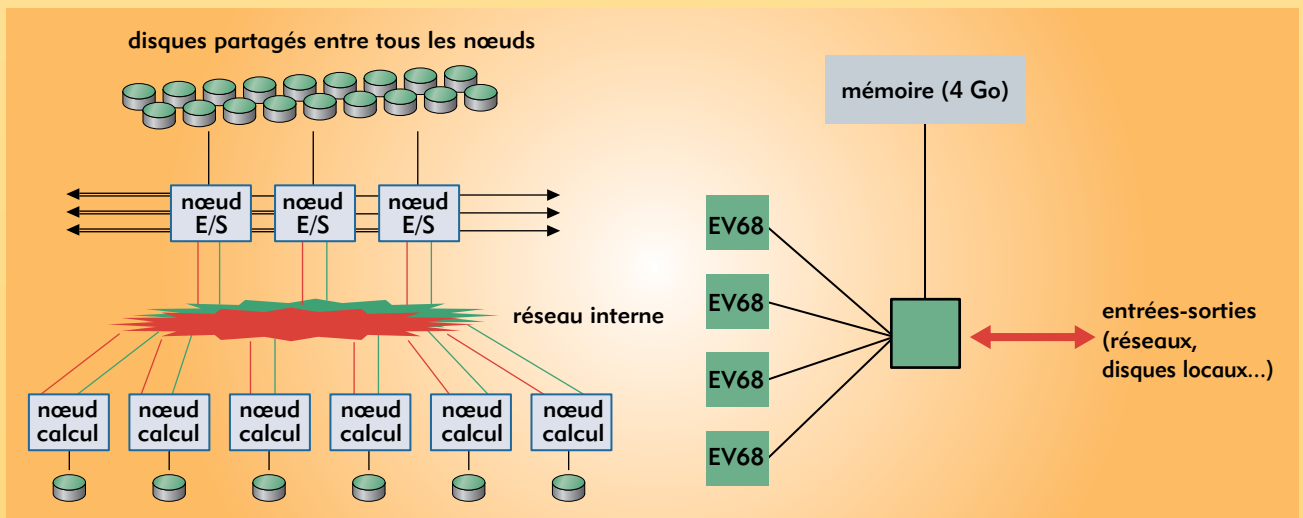


Figure. Architecture d'une machine du type "cluster de SMP". À gauche, l'architecture générale (E/S = entrée/sortie), à droite celle d'un nœud avec quatre processeurs Alpha EV68 cadencés à 1 GHz.

Modélisation et simulation des écoulements turbulents

La **turbulence**, ou l'agitation de l'écoulement dit turbulent, se développe dans la plupart des écoulements qui conditionnent notre environnement immédiat (rivières, océan, atmosphère). Elle se révèle être aussi un, sinon le, paramètre dimensionnant dans un bon nombre d'écoulements industriels (liés à la production ou la conversion d'énergie, à l'aérodynamique...). Il n'est donc pas étonnant que soient entrepris des efforts visant sa prédiction – fût-elle encore imprécise – surtout lorsqu'elle se trouve combinée à des phénomènes qui la compliquent : stratification, combustion, présence de plusieurs phases... C'est que, paradoxalement, même s'il est possible d'anticiper la nature turbulente d'un écoulement et même, d'un point de vue théorique, de dégager certaines caractéristiques communes et apparemment universelles aux écoulements turbulents⁽¹⁾, leur prédiction dans

des cas précis reste délicate. Celle-ci doit en effet prendre en compte l'importante gamme d'échelles spatiales et temporelles⁽²⁾ impliquées dans tout écoulement de ce type.

Les chercheurs ne sont pourtant pas démunis, aujourd'hui, pour aborder ce problème. En premier lieu, les équations qui régissent l'évolution spatio-temporelle des écoulements turbulents (équations de Navier-Stokes⁽³⁾) sont connues. Leur résolution complète, dans des cas très favorables, a conduit à des descriptions prédictives. Mais l'emploi systématique de cette méthode de résolution se heurte à deux difficultés rédhibitoires : d'une part, il nécessiterait la connaissance complète et simultanée de toutes les variables attachées à l'écoulement et des forçages s'exerçant sur lui⁽⁴⁾ et, d'autre part, il mobiliserait des moyens de calculs irréalistes pour encore des décennies.

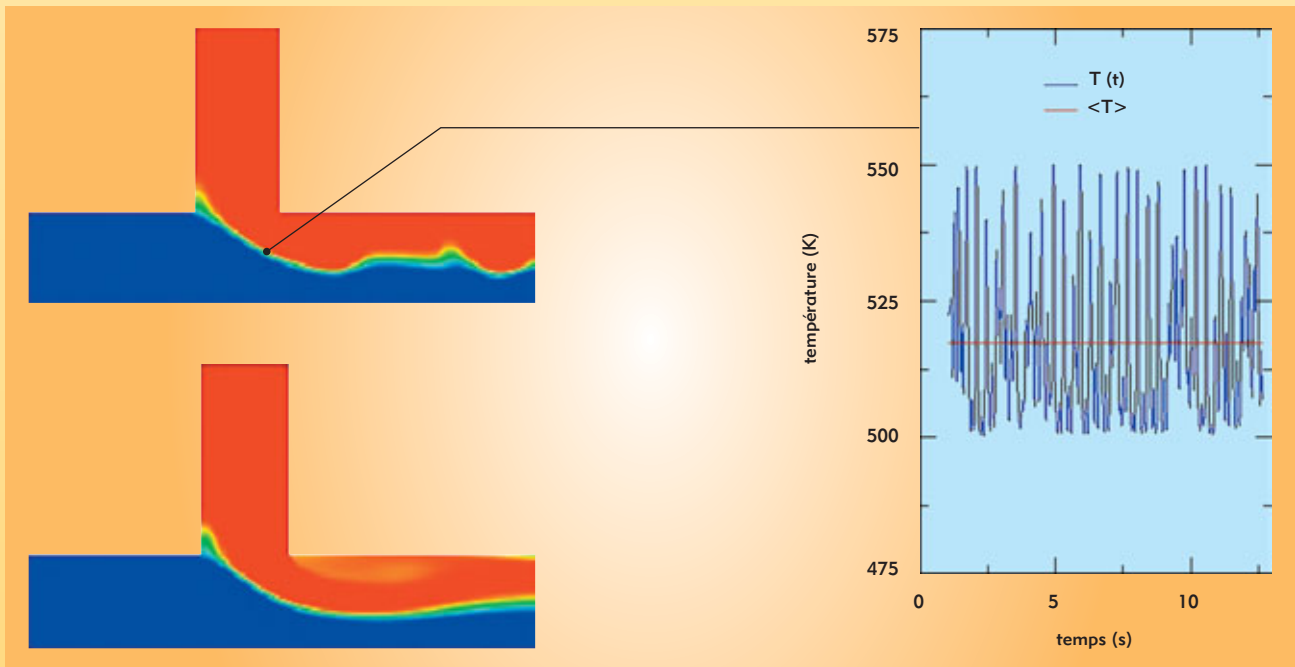


Figure. Champ de température instantané (haut) et moyenné (bas) dans une situation de mélange. La courbe donne l'historique de la température en un point : valeur instantanée fluctuante en bleu et moyenne en rouge (d'après la thèse d'Alexandre Chatelain [DEN/DTP/SMTH/LDTA]).

Il faut donc se résoudre, en s'appuyant sur le caractère fluctuant dû à l'agitation turbulente, à définir et utiliser des moyennes. Une des approches les plus répandues consiste à aborder le problème sous un angle statistique. Les moyennes d'ensemble de vitesse, de pression, de température... dont la distribution caractérise l'écoulement turbulent sont définies comme les variables principales de l'écoulement qu'on cherche à qualifier par rapport à ces moyennes. Ceci conduit à une décomposition du mouvement (dite de Reynolds) en champs moyen et fluctuant, ce dernier mesurant l'écart instantané et local entre chaque grandeur réelle et sa moyenne (figure). Ces fluctuations représentent la turbulence et couvrent une partie importante du spectre de Kolmogorov⁽¹⁾.

Cette opération réduit considérablement le nombre de degrés de liberté du problème et le rend « manipulable » informatiquement. Elle comporte aussi de nombreuses difficultés : il faut tout d'abord constater que, précisément en raison des non-linéarités des équations du mouvement, toute moyenne fait surgir des termes nouveaux et inconnus qu'il faut estimer. En fermant la porte à la description complète et déterministe du phénomène, on ouvre celle de la modélisation, c'est-à-dire à la représentation des effets de la turbulence sur les variables moyennes.

Beaucoup de progrès ont été accomplis depuis les premiers modèles (Prandtl, 1925). Les modélisations n'ont cessé d'évoluer vers plus de complexité, se basant sur le fait généralement vérifié que toute nouvelle extension permet de conserver les propriétés antérieurement acquises. Il faut aussi constater que, même

si de nombreux développements remettent en avant la nécessité de traiter les écoulements en respectant leur caractère *instationnaire*, les modélisations les plus populaires ont été développées dans le cadre des écoulements *stationnaires*, pour lesquels on n'accède donc qu'à une représentation de la moyenne temporelle de l'écoulement : dans le modèle mathématique final, les effets de la turbulence proviennent ainsi intégralement de la modélisation.

Il est également remarquable que, malgré de nombreux travaux, aucune modélisation n'est aujourd'hui capable de rendre compte de l'intégralité des phénomènes qui influencent la turbulence ou sont influencés par elle (transition, instationnarité, stratification, compression, etc.). Ce qui semble pour l'instant empêcher les modélisations statistiques de nourrir une ambition d'universalité.

Malgré ces limitations, la plupart des modélisations statistiques courantes sont maintenant disponibles dans les codes commerciaux et les outils des industriels. Il n'est pas possible de prétendre qu'elles permettent des calculs prédictifs dans toute situation. Leur précision est variable, offrant des résultats utiles pour l'ingénieur dans des situations maîtrisées et favorables (prédiction de la trainée avec une précision de 5 % à 10 % d'erreur [parfois mieux] sur certains profils), mais parfois faux dans des situations qui se révèlent, après coup, en dehors du champ de validité du modèle. Tout emploi maîtrisé d'une modélisation repose donc sur une qualification particulière au type d'écoulement à traiter. Des modélisations alternatives, répondant au besoin d'une plus grande précision sur des gammes d'échelles spatiales et temporelles plus étendues et donc basées sur un opérateur de "moyenne" d'une nature différente, sont actuellement en développement et représentent des voies nouvelles.

Le paysage des modélisations de la turbulence est aujourd'hui très complexe et l'unification des points de vue et des divers concepts de modélisation est une gageure. La tentation de l'universalité des modélisations reste donc hors de propos. Leur mise en œuvre réelle relève la plupart du temps de compromis généralement guidés par le savoir-faire de l'ingénieur.

(1) On peut faire référence à la répartition spectrale de l'énergie cinétique turbulente, connue comme le "spectre de Kolmogorov", qui illustre de manière très simple la hiérarchie des échelles, des grandes échelles porteuses d'énergie aux échelles de plus en plus petites et de moins en moins énergétiques.

(2) Cette étendue est le résultat des non-linéarités des équations du mouvement qui donne naissance à une gamme étendue d'échelles spatiales et temporelles. Cette gamme est une fonction croissante du nombre de Reynolds, Re , mesurant le rapport entre force d'inertie et force visqueuse.

(3) L'hypothèse selon laquelle la résolution complète des équations de Navier-Stokes permet la simulation de la turbulence est généralement admise, tout du moins dans la gamme des écoulements sans choc.

(4) Il s'agit d'un problème régi par des conditions initiales et aux limites.

Les macromolécules biologiques

Les cellules sont les entités fondamentales de tous les organismes vivants. En dehors de l'eau, les **macromolécules** biologiques sont les constituants majeurs de la cellule, où elles remplissent des fonctions multiples. Une macromolécule biologique est composée de sous-unités de faible poids moléculaire, ajoutées les unes aux autres pour donner un long **polymère** en forme de chaîne. Habituellement, chaque chaîne n'est formée que d'une seule famille de sous-unités dont l'enchaînement précis est essentiel à la fonction de la macromolécule. Les grandes catégories de macromolécules sont au nombre de quatre.

Les **protéines** sont probablement les macromolécules les plus importantes parce qu'elles jouent un rôle prédominant dans la plupart des processus biologiques. Par exemple, les **enzymes** sont les protéines qui **catalysent** la majorité des réactions chimiques dans la cellule. D'autres classes de protéines ont plutôt un rôle structural ou sont impliquées dans la signalisation⁽¹⁾, la régulation du **métabolisme** ou la défense immunitaire. Les protéines sont des polymères d'**acides aminés** – dont une vingtaine de types différents sont communément trouvés – et une

protéine peut être constituée de quelques chaînes, chacune comportant quelques centaines d'acides aminés. Les protéines sont souvent associées à d'autres molécules qui les assistent dans leurs tâches biologiques. Les structures tridimensionnelles des protéines sont très compliquées mais critiques pour leur fonctionnement.

Les **acides nucléiques** – l'acide désoxyribonucléique (ADN) et l'acide ribonucléique (ARN) – sont des polymères de **nucléotides**. L'ADN, dans sa forme double brin (deux chaînes de nucléotides organisées en double hélice), est le matériel génétique qui, entre autre, code les instructions pour les séquences (l'enchaînement des acides aminés) de toutes les protéines synthétisées par la cellule. L'ARN, habituellement dans une forme simple brin (une chaîne de nucléotides), est essentiel pour la synthèse des protéines.

Les **lipides**, constituants fondamentaux des membranes cellulaires, exercent également une fonction importante dans le métabolisme et comme réservoirs énergétiques. Parmi les classes de lipides se trouvent les phospholipides, les triglycérides et les stéroïdes.

Les **polysaccharides** sont des polymères de sucres simples, tels que le fructose ou le glucose. Ils jouent un rôle structural, notamment dans les plantes (la cellulose est un polysaccharide), dans la reconnaissance moléculaire, et peuvent servir de réservoirs énergétiques.

(1) Transmission de signaux permettant aux cellules de communiquer entre elles.