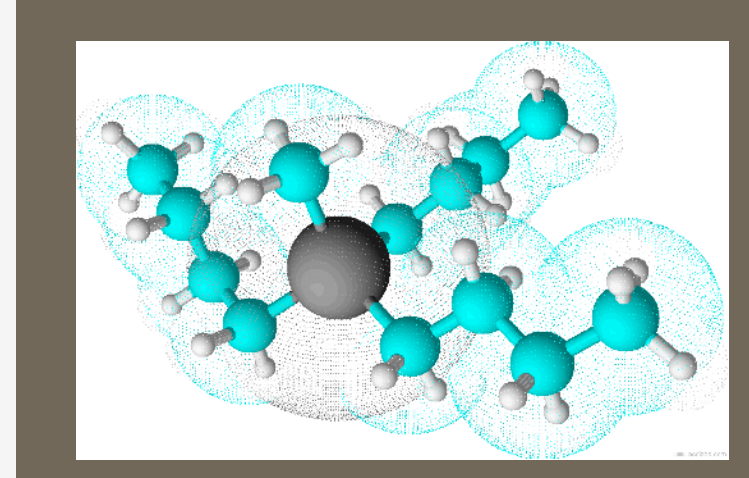


# Théorie des invariants et application en Chimie



Modélisation des molécules par des **orbitales** sous **une action finie, invariants** associés et prédiction des **propriétés chimiques** : Exemple des **graphes moléculaires**

L'algèbre des invariants au cœur de la physique d'un système.

# Plan :

- **Contexte actuel de la Chimie** : besoin de **modélisations moléculaires** efficaces pour calculer des grandeurs physiques par **QSARs**.
- **Principe** rapide des **QSARs**.
- **QSARs** et **modèles** inspirés des **graphes** (usage des invariants sous un groupe fini de permutations).
- **Sommes orbitales générales**
- **Connexité** et décomposition en **composantes connexes**
- **N-Connexité** et décomposition associée
- Résultat de **simplification** de la **base** classique d'**invariants** pour les **graphes chimiques**.

# Contexte réglementaire de la chimie actuelle

situation actuelle du traitement du risque chimique

## ○ Directive REACH de l'UE :

- Renversement de la charge de la preuve
- Nombreux paramètres à estimer :
  - physico-chimiques
  - Toxicité
  - Ecotoxicité

Méthodes : *in vivo*, *in vitro*, *in silico*.

## ○ Approches *in silico* à privilégier :

- Rapides et peu coûteuses
- Ethiquement plus satisfaisantes
- Parfois plus efficaces scientifiquement (ex : toxicité des hydrophobes vis-à-vis des poissons)

PREDIMOL et CAESAR ont montré la faisabilité de cette approche.

REACH

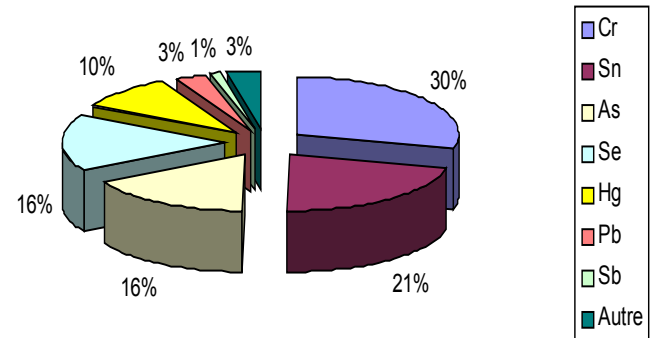


# Contexte économique

## ○ Tendance actuelle de l'industrie chimique

- Utilisation d'organométalliques
- Certains éléments sont problématiques : généralement très toxiques et utilisés à grande échelle

**mercure, Etain,...**



**Parts de marché**

## ○ Besoins des acteurs :

- Les entreprises chimiques ont des besoins d'expertise pour monter des dossiers REACH.
- Des acteurs institutionnels ont des besoins d'outils d'évaluation des risques chimiques.

# Contexte scientifique

- Des institutions se sont mobilisées pour apporter des solutions :

- OCDE (QSAR Toolbox)
- EPA
- ECHA



FEDERCHIMICA  
CONFINDUSTRIA



EUROPEAN CHEMICALS AGENCY

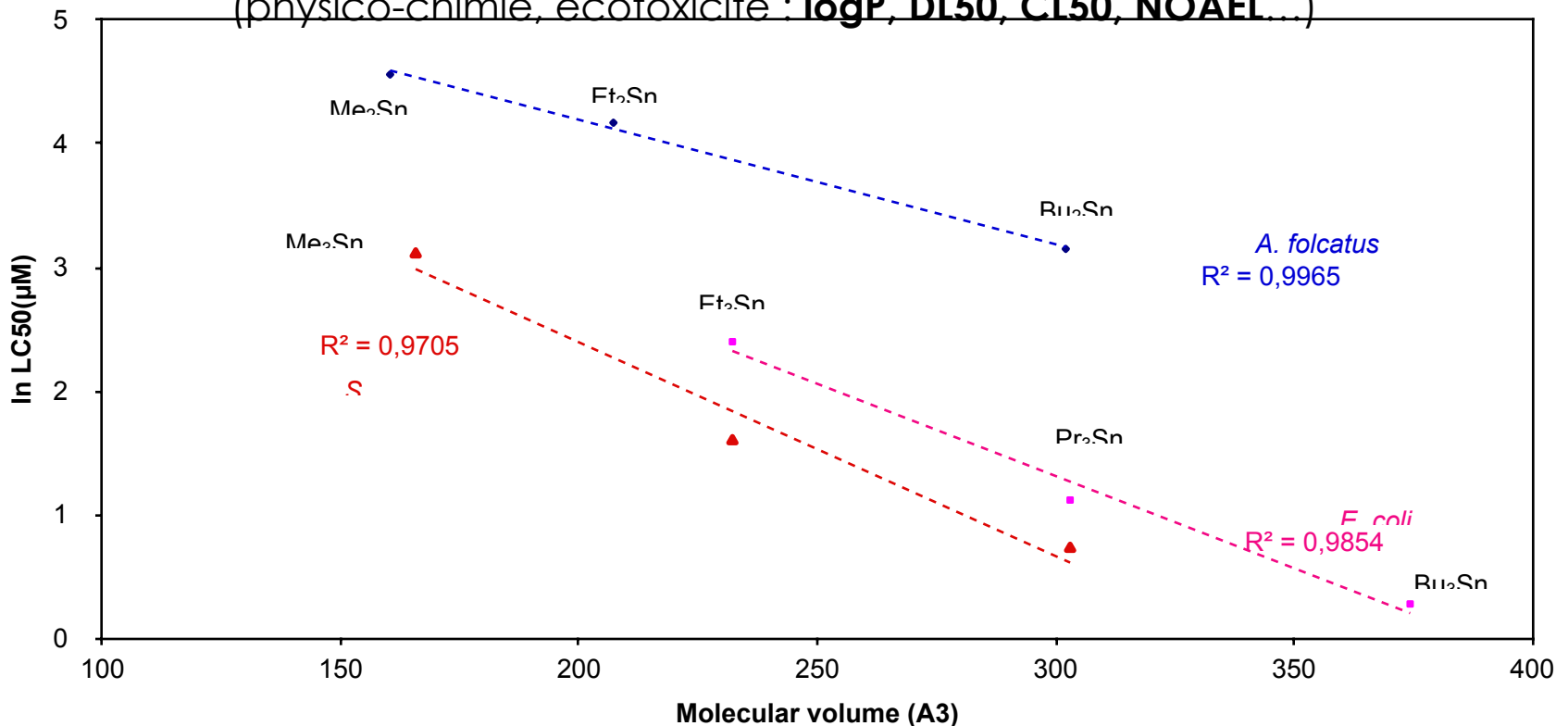


maîtriser le risque  
pour un développement durable

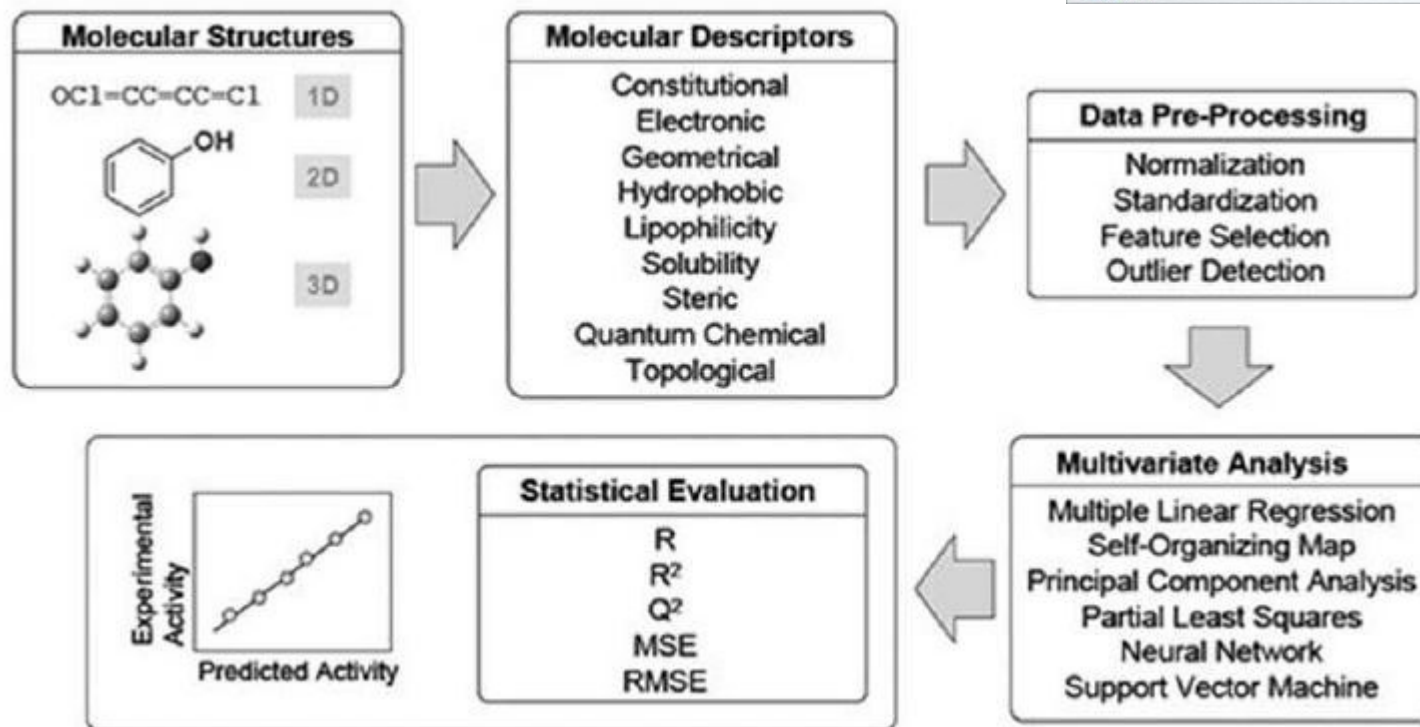
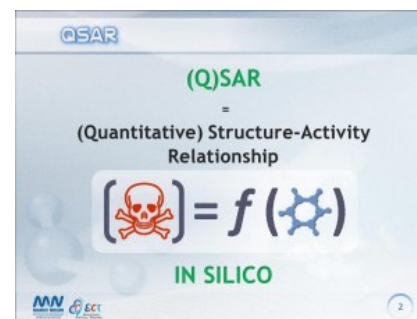
- **Mais**, actuellement, les méthodes adaptées aux **organométalliques** sont **peu développées**. Les méthodes adaptées à la chimie organique sont éparpillées et **non unifiées**.

Mon travail vise à **développer des QSARs** pour répondre à ce besoin de prédiction de certains

Toxicity of di- and tri-organotin compounds (chlorides) on different organisme as a function of the molecular volume (after Luedke et al., 1991, Main Group Metal Chem.)  
(physico-chimie, écotoxicité : **logP, DL50, CL50, NOAEL...**)



# Principe des QSARs



# Approche des descripteurs topologiques :

tenseurs et actions de groupes de permutations

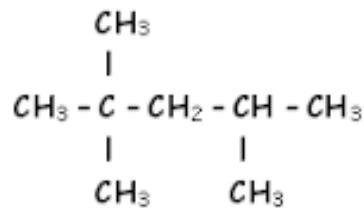
On se donne un modèle des molécules sous la forme d'un **tenseur** contenant des nombres ou des étiquettes (alors en **nombre fini**) pour stocker l'information reflétant la molécule au sein d'une famille fixée de composés.

Ces modèles tensoriels ont une particularité commune : **l'action du groupe des permutations** agissant sur les lignes ou les colonnes, voire les deux, **laisse le composé** traduit par le tenseur **invariant**.

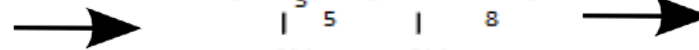
$$\begin{matrix} & & & & & & & \sigma \\ & & & & & & & \curvearrowright \\ \sigma & \curvearrowleft & \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 1 & 0 & 0 & 1 & 0 \end{pmatrix} \end{matrix}$$



# Exemple de ce type d'approches : un codage pour les alcanes - 1



molécule

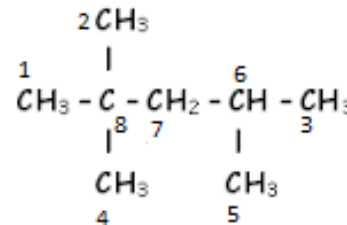


Étiquetage arbitraire

$$\begin{pmatrix}
 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
 1 & 1 & 0 & 1 & 1 & 0 & 0 & 0 \\
 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 \\
 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0
 \end{pmatrix}$$

Codage associé

On passe d'une matrice à l'autre en permutant les numéros des carbones



Autre étiquetage arbitraire

$$\begin{pmatrix}
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\
 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\
 1 & 1 & 0 & 1 & 0 & 0 & 1 & 0
 \end{pmatrix}$$

Codage associé

## Exemple de ce type d'approches : un codage pour les alcanes - 2

La molécule d'alcane est donc modélisée par une **matrice bien construite de 0 et de 1 à permutation près des lignes et des colonnes.**

$$\begin{matrix} & & & & & & & & \sigma \\ & & & & & & & & \curvearrowright \\ \sigma \curvearrowleft & \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \end{pmatrix} & & & & & & & & \end{matrix}$$

En général, on modélise les **molécules** au sein d'une famille de composés par des **orbitales** sous l'action de **groupes** de permutations partitionnant un ensemble **fini** de **tenseurs**.

# Application des résultats généraux sur les invariants

Les observables polynomiales de ces objets vivent dans des **algèbres de Cohen–Macaulay**, caractérisées par :

- des **familles génératrices minimales d'invariants** de base
- et leurs **syzygies** (liaisons polynomiales) qui les contraignent.

$$A = \mathbb{R}[I_1, I_2, I_3, \dots, I_N] / \text{Id}(S_1, S_2, S_3, \dots, S_p)$$

En Chimie, ceci a pour conséquence que toute quantité observable attachée au modèle de molécule s'exprime comme une fonction aussi voisine que l'on désire (en un sens à préciser) d'un polynôme des invariants de base, ce qui fait que

**les invariants de base contiennent toute la physique du composé.**

# L'algèbre des invariants polynomiaux

du modèle pour les alcanes

- On démontre que l'algèbre des invariants est **engendrée linéairement** (et non algébriquement) **par la famille** suivante :

$$\sum_{\sigma \in S_n} \prod_{i < j} X_{ij}^{m_{\sigma(i)\sigma(j)}}, m_{ij} \in \{0,1\}^{[[1,n]]^2}$$

où  $X_{ij}$  est le nombre inscrit aux coordonnées  $(i,j)$  de la matrice.

- On associe donc à chaque matrice  $(m_{ij})$  un **invariant**. Or ces matrices ne contiennent que des 0 et des 1 et sont définies à invariance près sous l'action de  $S_n$  sur les lignes et les colonnes : ce sont à nouveau les **graphes**. On a donc une base d'invariants définie par une **dualité** (non symétrique) **entre graphes**.

## Principe des sommes orbitales : Une heuristique pour de nouveaux modèles

- Cette somme peut alors s'écrire plus généralement :

$$I(G,H) = \text{nombre de sous-modèles de } G \text{ isomorphes à } H,$$

ou  $G$  et  $H$  sont des modèles moléculaires (ex : graphes)

Les  $(I(G,H), H$  parcourant les modèles possibles) forment alors une **base linéaire** de l'algèbre des invariants polynomiaux.

- Dans le cas des graphes, on montre que les  $(I(G,H), H$  parcourant les graphes seulement connexes) engendrent algébriquement l'algèbre des invariants
- Enjeu : **simplification de la famille des  $H$ .**



# Autres applications de l'heuristique des sommes orbitales

**$I(G,H)$  = nombre de sous-modèles de  $G$  isomorphes à  $H$ ,**

où  $G$  et  $H$  sont des modèles moléculaires (ou pas)

ex.: - graphes (simples (ou non), non orientés sans boucles)

- graphes biétiquetés (de valence d'atome ou de liaison bornée ou pas)

- hypergraphes biétiquetés (de valence des points blancs majorée par 2)

- autre modèle tensoriel ou plus général (jeu de taquet, carte, hypercarte).

**Les  $(I(G,H), H$  parcourant les modèles possibles) forment alors une **base linéaire** de l'algèbre des invariants polynomiaux.**

**On a aussi le résultat sur les modèles **connexes**.**

# Un point fort du modèle des graphes et des modèles tensoriels

La base précédente permet de **caractériser tous les modèles tensoriels**. En particulier l'algèbre des observables réelles des graphes permet de caractériser les graphes. Ce n'est pas le cas dans la théorie générale des invariants



# Dualités entre modèles réels et modèles potentiels

Avec le système générateur d'invariants ( $I(G,H)$ ,  $H$  parcourant les modèles possibles), on considère beaucoup de graphes  $H$  qui ne peuvent pas être des sous modèles viables d'une molécule (cycles trop courts, trop de liaisons, etc...).

**Donc on enlève de la famille des  $H$  tous les modèles qui ne peuvent pas apparaître comme sous-structure d'une molécule réelle.**

# Notion de modèle matriciel connexe

- Les **graphes connexes** se caractérisent par le fait qu'ils **ne peuvent pas** être coupés en deux paquets d'atomes AUA' et de liaisons LUL' tels que si l'on écrit leur matrices dans les ordres sur les liaisons et atomes AUA' et LUL' la matrice obtenue soit **diagonale par blocs** :

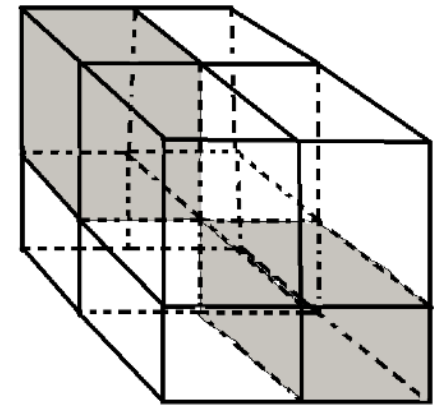
$$\begin{pmatrix} M(A, L) & 0 \\ 0 & M(A, L') \end{pmatrix}$$

# Modèle matriciel connexe

Lorsqu'un modèle matriciel peut s'écrire sous la forme d'une telle matrice, **l'algèbre de ses invariants est engendrée algébriquement par les invariants des deux sous-graphes  $(A,L)$  et  $(A',L')$ .**

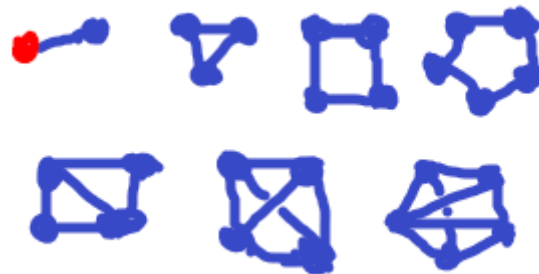
L'algèbre des invariants d'un modèle matriciel est donc engendrée algébriquement par **les invariants de ses composantes connexes.**

La généralisation aux modèles tensoriels permet de **définir des modèles tensoriels connexes** et de conserver la propriété précédente.



## Une avancée sur la famille algébriquement génératrice de l'algèbre des invariants

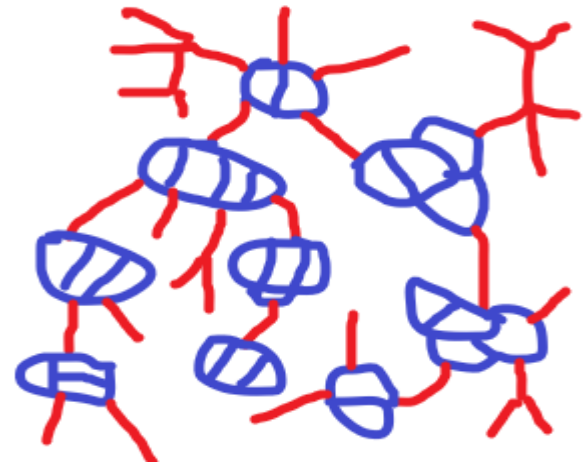
- On sait que l'algèbre des invariants d'un graphe est **algébriquement engendrée** par une famille de polynômes invariants, **un pour chaque graphe connexe**.
- En fait on a le même résultat en restreignant cette famille aux seuls polynômes associés aux **graphes 2-connexes**, et aux **hypergraphes à une arête**.



# Notion de $n$ -connexité

- Un graphe est **au moins  $n$ -connexe** s'il n'est pas possible de le rendre non connexe en lui **enlevant  $n-1$  arêtes** bien choisies.
- S'il est au moins  $n$ -connexe et pas au moins  $(n+1)$ -connexe, il est **exactement  $n$ -connexe**.
- Les graphes exactement 1-connexes sont les **arbres**.

**On montre que tout graphe connexe** (= au moins 1-connexe) **est un arbre de sous-graphes pleins au moins 2-connexes ou de sommets.**

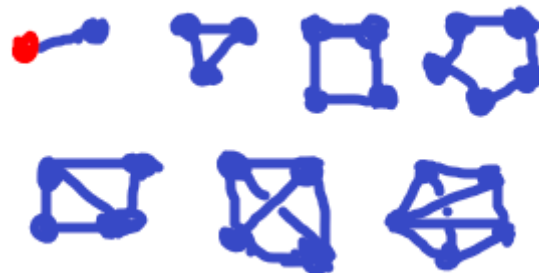


## Forme de la matrice de graphe dans la décomposition précédente

- On regroupe les sommets des zones au moins 2-connexes dans des **blocs** et on obtient une **matrice** de la forme :

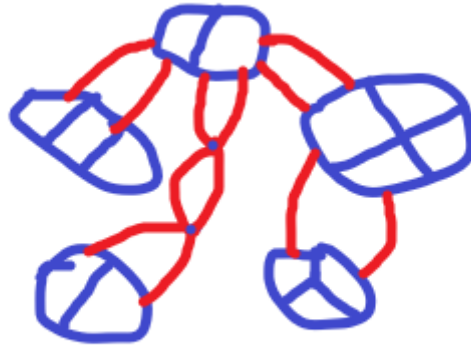
zone	bloc hypergraphe	bloc hypergraphe
2-connexe	à 0 ou 1 arête	à 0 ou 1 arête
bloc hypergraphe	zone	bloc hypergraphe
à 0 ou 1 arête	2-connexe	à 0 ou 1 arête
bloc hypergraphe	bloc hypergraphe	zone
à 0 ou 1 arête	à 0 ou 1 arête	2-connexe

- On montre alors que **tout invariant du graphe** est un polynôme des invariants des **blocs diagonaux** et des invariants des **blocs hypergraphes** (il n'y en a qu'un).
- On peut donc restreindre la famille des graphes que l'on utilise pour tester dans le calcul des invariants. On se limite aux **graphes 2-connexes** et au simple **hypergraphe à une arête**.



## Suite de la récurrence le long des $n$ de la notion de $n$ -connexité.

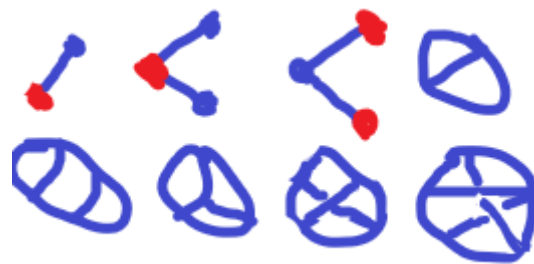
- On montre que tout graphe au moins 2-connexe est un **2-arbre** de sous-graphes pleins au moins **3-connexes** ou de sommets.



- On montre encore que **tout invariant du graphe** est un polynôme des invariants des **blocs diagonaux** et des invariants des **blocs hypergraphes à 2 arêtes** (il y en a 6).



- On peut donc à nouveau restreindre la famille des graphes que l'on utilise pour tester dans le calcul des invariants. On se limite aux **graphes 3-connexes** et aux simples **hypergraphes à une ou 2 arêtes**.



- On peut potentiellement continuer cette récurrence sans fin et trouver de nouvelles décompositions.
- Mais **en Chimie, la n-connexité a une limite**. En effet un graphe n-connexe n'a que des sommets de valence n au moins, et la nature limite la valence des atomes à 9. En outre, les molécules admettant des sous-graphes 4-connexes sont très rares.



Merci pour votre attention