Processus asynchrones provenant d'une grammaire stochastique ou quantique appliqués en génomique

Dimitri PETRITIS

Institut de Recherche Mathématique

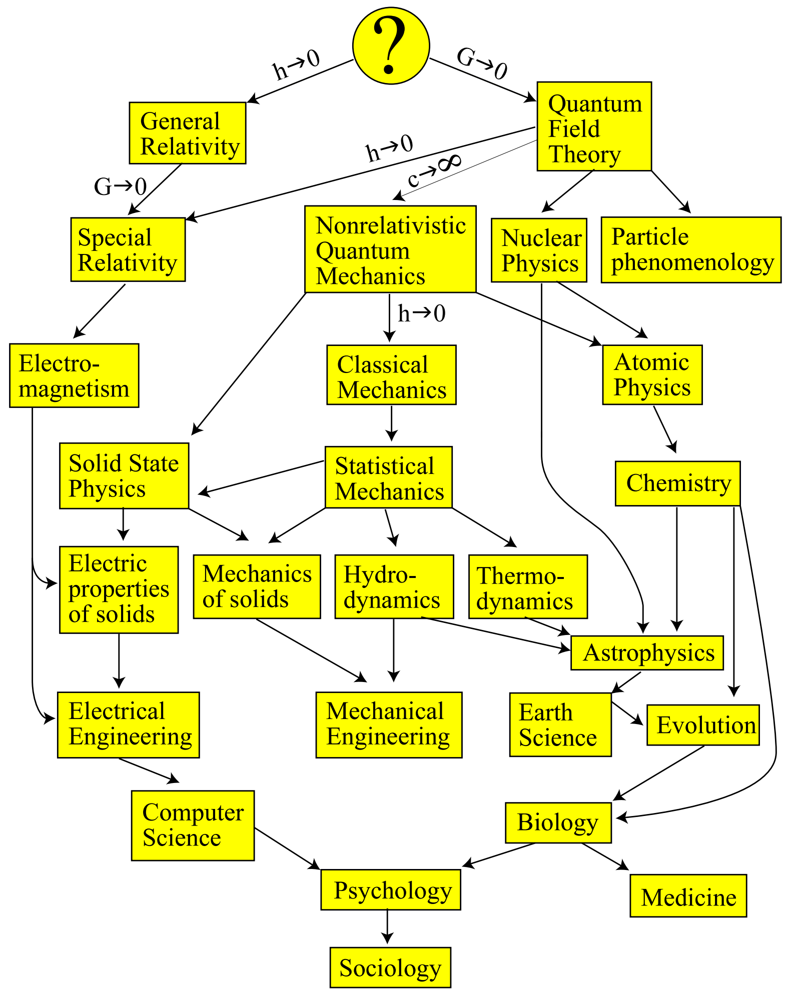
Université de Rennes 1 et CNRS (UMR 6625)

France

dimitri.petritis@univ-rennes1.fr

http://name.math.univ-rennes1.fr/dimitri.petritis

Interdépendence des disciplines scientifiques



Max Tegmark, John Archibald Wheeler: 100 years of the Quantum, Scientific American

Un résumé de biologie moléculaire

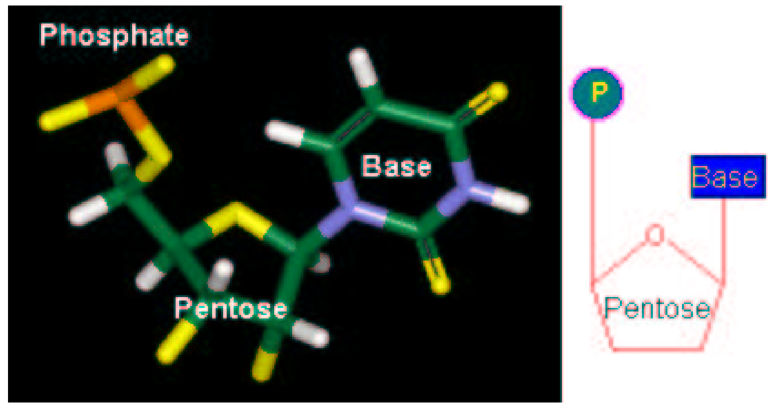
- Dans la cellule il y a 2 types de molécules:
  - les courtes: sucres, acides aminés, nucléotides, etc.
  - les longues: ADN (acide desoxyribonucléïque), ARN (acide ribonucléïque) et les protéines
- ADN code la représentation des protéines
- Dogme central de la biologie: l'information se transmet suivant les flèches:

$$\circlearrowleft \text{ADN} \longrightarrow \text{ARN} \longrightarrow \text{PROTÉINE}$$

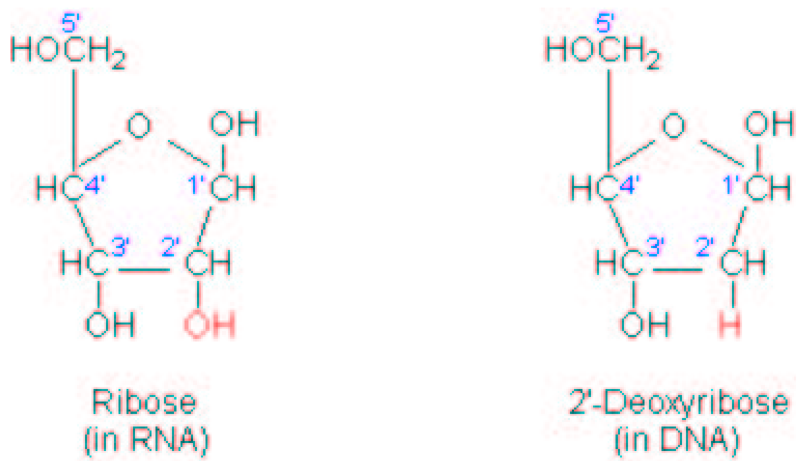
appeées *replication, transcription et traduction.*

La structure de la molécule d'ADN

- Nucléotides: *E. coli* 5×10^6; *H. sapiens* 3×10^9

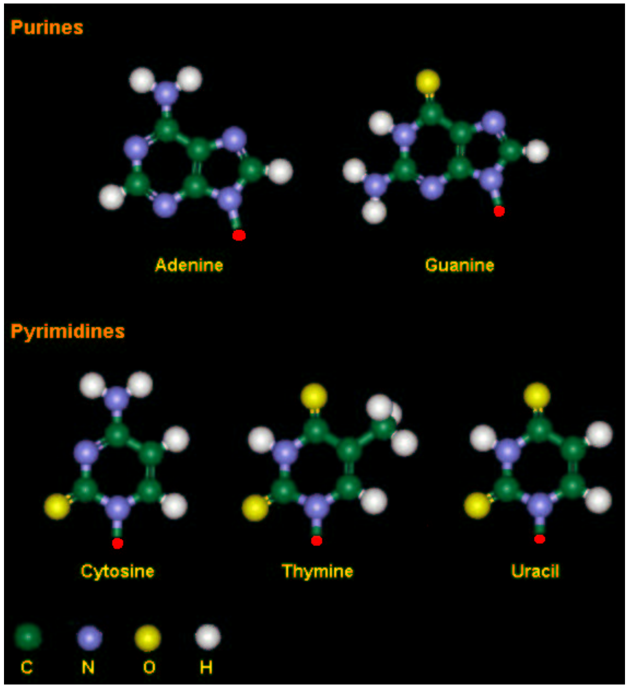


- Squelette de pentoses



La structure de la molécule d'ADN

- Bases

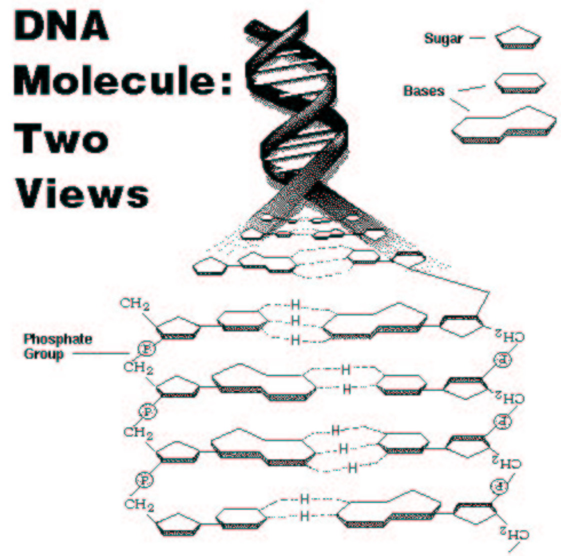


$\mathbb{B} = \{A, C, G, T(U)\}$; G-C et A-T(U) complémentaires

- L'ensemble $B = \{A, C, G, T(U)\}$ indexe les nucléotides

La structure de la molécule d’ADN

- ADN long polymère branché. Stéreo-chimiquement: double hélice. Distance entre monomères 0.34 nm



Le code génétique

- $\mathbb{A}$ ensemble de 20 acides aminés

Acide	nnn	n	Acide	nnn	n
Alanine	Ala	A	Leucine	Leu	L
Arginine	Arg	R	Lysine	Lys	K
Acide aspartique	Asp	D	Méthionine	Met	M
Asparginine	Asn	N	Phénylanine	Phe	F
Cystéine	Cys	C	Proline	Pro	P
Acide glutique acid	Glu	E	Sérine	Ser	S
Glutamine	Gln	Q	Thréonine	Thr	T
Glycine	Gly	G	Tryptophane	Trp	W
Histine	His	H	Tyrosine	Tyr	Y
Isoleucine	Ile	I	Valine	Val	V

- 3 nucléotides, le *codon*, pour coder un acide aminé
- Direction de lecture $5' \rightarrow 3'$

Le code génétique

Code génétique: $g : \mathbb{B}^3 \to \mathbb{A} \cup \{\textsc{stop}\}$. Codon début: AUG

1st	2nd				3rd	1st	2nd				3rd
	U	C	A	G			U	C	A	G	
U	Phe	Ser	Tyr	Cys	U	A	Ile	Thr	Asn	Ser	U
U	Phe	Ser	Tyr	Cys	C	A	Ile	Thr	Asn	Ser	C
U	Leu	Ser	STOP	STOP	A	A	Ile	Thr	Lys	Arg	A
U	Leu	Ser	STOP	Trp	G	A	Met	Thr	Lys	Arg	G
C	Leu	Pro	His	Arg	U	G	Val	Ala	Asp	Glu	U
C	Leu	Pro	His	Arg	C	G	Val	Ala	Asp	Glu	C
C	Leu	Pro	Gln	Arg	A	G	Val	Ala	Glu	Glu	A
C	Leu	Pro	Gln	Arg	G	G	Val	Ala	Glu	Glu	G

Gènes et génome

- Gène: région d'ADN impliquée dans synthèse d'une protéine
- Gène séparé en régions codantes (*exons*) et non codantes (*introns*)
  - 5% ADN "utile" (contient information pour synthèse des protéines)
  - 95% ADN "poubelle" (signature de l'individu)
- Génome: ADN contenu dans la cellule. Longueur de la molécule d'ADN $= 10^5$ diamètres cellulaires!

“Vous visitez un musée de peinture, et vous vous promenez parmi les toiles françaises du début de XXe siècle. Vous voyez ici un Renoir somptueux, là sans erreur possible c’est un Modigliani, là encore ce sont de fleurs peintes par van Gogh, ou des fruits de Cézanne. Plus loin vous appercevez un Picasso, à moins que ce ne soit un Braque. Sans doute est-ce la première fois que vous voyez ces peintures, mais vous n’avez le plus souvent aucun doute quant à l’artiste à qui on les doit.[. . .] Comment les distinguez vous? Eh bien la peinture n’est pas appliquée de la même manière, et les sujets sont différents. Mais il y a quelque chose d’autre, qui est plus difficile à exprimer et que cependant on saisit immédiatement, un quelque chose qui dépend du choix des formes et de l’équilibre de couleurs.”

–David RUELLE, Hasard et chaos

Quelques observations

- **Première observation:** Une séquence de symboles codant message avec contenu relatif à un individu contient la signature statistique de l'individu.
- "Zij $\omega(\cdot)$ een lineaire afbeelding van een reële Hilbertruimte $\mathcal{K}$ naar de ruimte van stochasten op een kansruimte met eindige momenten."
- **Seconde observation:** Confiance en la signature statistique renforcée lorsque message concerne large classe d'individus.

Qeulques observations

- **Troisème observation:** Au delà de variété combinatoire des phrases individuelles d'une classe, il existe une structure commune sous-jacente, une *grammaire*, permettant la distinction entre phrases grammaticallement correctes et fausses et nous aidant à donner un sens:
  - Un crocodile mord un homme (grammaticalement correcte)
  - Un homme mord un crocodile (grammaticalement correcte)
  - Un crocodile un homme mord (grammaticalement incorrecte)
- **Quatrième observation:** Une *grammaire* est un *ensemble fini* de règles permettant formation et compréhenseion d'une infinité de phrases.

Dernières observations et une thèse

- **Cinquième observation:** L'ADN est une molécule chimique standard. Covalence des liens et distance interbase typique indiquent qu'interactions quantiques ne sont pas négligeables.
- **Sixième observation:** L'ADN de chaque individu est une collection de phrases ayant un sens précis.
- **Une thèse:** Il existe un grammaire sous-jacente (probablement quantique) pour le langage de l'ADN.
  - Formalisme mathématiquement cohérent
  - Inventer d'algorithmes et expériences *in vivo* pour vérifier pertinence biologique de la thèse

Théorie des probabilités et physique classique

- **Axiomatisation de Kolmogorov**
  - Espace de probabilité: $(\Omega, \mathcal{F}, \mathbb{P})$
  - Espace mesurable: $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$
  - Variable aléatoire réelle: $X : \Omega \to \mathbb{R}$, measurable
  - Loi de X:

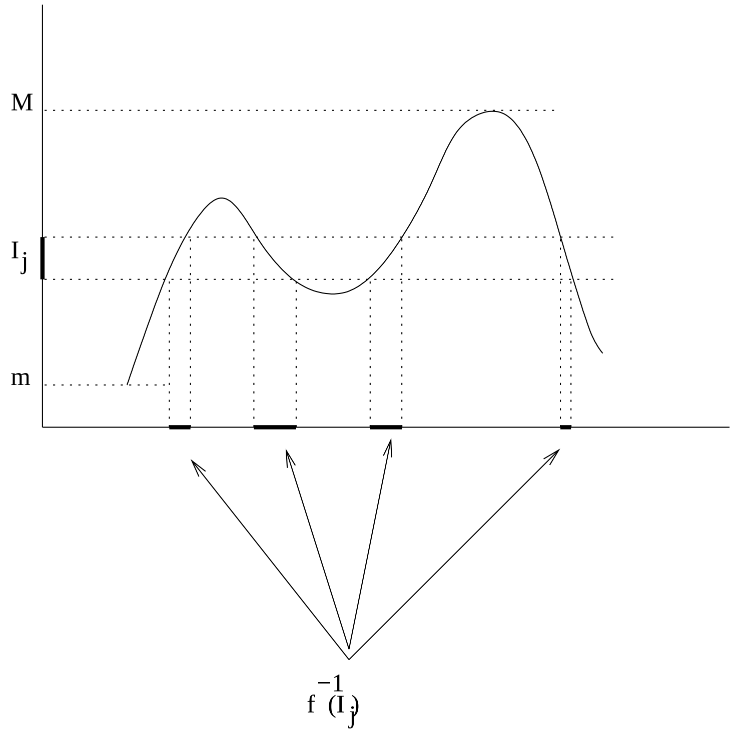
$$\mathbb{P}_X(B) \equiv \mathbb{P}(\{\omega \in \Omega : X(\omega) \in B\}) = \mathbb{P}(X \in B), B \in \mathcal{B}(\mathbb{R}).$$

- **Quelle est la signification de cette axiomatisation?**
  - Kolmogorov: Ensemble de propositions expérimentalement vérifiables = σ-algebra booléenne
  - Théorème de Loomis-Sikorski: toute σ-algebra booléenne est image σ-homomorpe d'une tribu $\mathcal{F}$ des parties d'un Ω non-vide.

Dictionnaire

	Probabilités	Physique
Ω	Espace des essais	Espace des phases
$\omega \in \Omega$	réalisation	microétat
X	variable aléatoire	observable
$\mathbb{P}$	probabilité	(macro) état
$\mathbb{P}_X$	loi de X	fonction d'état X dans état $\mathbb{P}$
$(\omega_t)_t$	trajectoire	flot dynamique
$X(\omega_t)$	processus stochastiaue	évolution temporelle de X

Approximation par des fonctions simples



Une approche différente

- Continue, bornée $X : \Omega \to \mathbb{R}$;
  $m = \inf X(\omega)$; $M = \sup X(\omega)$; $\mathsf{spec}(X) = [m, M]$.
- Objets importants: $\mathcal{F}, \mathcal{B}(\mathbb{R})$. (Convention: $\mathbb{1}_F \equiv F$.)
- Mesure à valeurs indicatrices $P : \mathcal{B}(\mathbb{R}) \ni B \to \mathbb{1}_{X^{-1}(B)}$.

$$\forall \omega : \left| X(\omega) - \sum_j x_j P(I_j)(\omega) \right| < \epsilon.$$

$$\lim \sum_j x_j P(I_j) = \int x P(dx) = X.$$

$$\begin{aligned} \mathbb{E}X &= \int X(\omega)\mathbb{P}(d\omega) \\ &= \int_\Omega (\int_m^M xP(dx)(\omega))\mathbb{P}(d\omega) = \int_{\mathsf{spec}\, X} x\mathbb{P}_X(dx). \end{aligned}$$

- $P^2 = P$; $P(B \cap C) = P(B)P(C)$; $\mathsf{supp}\, P = \mathsf{spec}\, X = \mathsf{Ran} X$.
- P est la *mesure spectrale* de X. Contient la même information que X.
- En mécanique quantique de nouveau: $X = \int xP(dx)$ mais P projection à un sous-espace hilbertien.

Mécanique quantique

- **Postulat 1:** Espace des phases = espace hilbertien complexe séparable $\mathbb{H}$; rayons $\psi \in \mathbb{H} : ||\psi|| = 1$ correspondend à des états (purs).
- Exemple $\mathbb{H} = \mathbb{C}^2$:
$||\psi||^2 = |\psi_1|^2 + |\psi_2|^2 = 1$. Composants de ψ correspondend à des amplitudes de probabilité.
- **Postulat 2:** Évolution temporelle d'un système quantique isolé dérive d'un opérateur unitaire sur $\mathbb{H}$.
- ϕ rayon; $\psi = U\phi$ est encore un rayon i.e. un état pure.
- $\phi = U^*\psi$: l'évolution temporelle est réversible.

Mécanique quantique

Postulat 3: Observables physiques associés à des opérateurs auto-adjoints bornés sur $\mathbb{H}$. Mesure (physique)de X dans état ψ signifie de déterminer mesure (borélienne) dur $\mathbb{R}$ induite par mesure spectrale de X.

$$X = \begin{pmatrix} 1 & 2i \\ -2i & 2 \end{pmatrix}.$$

Val. propres x	vec. propres $u(x)$	Projecteurs $P(\{x\})$
-3	$\frac{1}{\sqrt{5}}\begin{pmatrix} -i \\ 2 \end{pmatrix}$	$\frac{1}{5}\begin{pmatrix} 1 & -2i \\ 2i & 4 \end{pmatrix}$
2	$\frac{1}{\sqrt{5}}\begin{pmatrix} 2i \\ 1 \end{pmatrix}$	$\frac{1}{5}\begin{pmatrix} 4 & 2i \\ -2i & 1 \end{pmatrix}$

$$\begin{aligned} X &= \sum_{x\in\{-3,2\}} xP(\{x\}) \\ &= (-3)\frac{1}{5}\begin{pmatrix} 1 & -2i \\ 2i & 4 \end{pmatrix} + 2\frac{1}{5}\begin{pmatrix} 4 & 2i \\ -2i & 1 \end{pmatrix}. \end{aligned}$$

Mécanique quantique

$$\begin{aligned}
\psi &= \alpha_{-3}u(-3) + \alpha_2 u(2). \\
\langle\, \psi \,|\, X\psi \,\rangle &= \sum_{x,x',x''} \alpha_x^* \alpha_{x''} x' \langle\, u(x) \,|\, P(\{x'\}) u(x'') \,\rangle \\
&= \sum_{x \in \mathsf{spec}(X)} x |\alpha_x|^2,
\end{aligned}$$

où $|\alpha_x|^2 = \langle\, \psi \,|\, P(\{x\})\psi \,\rangle$. $\mathbb{E}_\psi X = \langle\, \psi \,|\, X\psi \,\rangle$.
Observables élémentaires: P questions "oui-non".
Nouvel état après mesure $P(\{x\})$: $u(x)$.

Calcul propositionnel

- Chercher description unifée pour systèmes classiques et quantiques.
- Il suffit d'examiner seuelment les mesures spectrales P.
- Trouver carctéristiques communes entre ensembles mesurables de $\mathcal{F}$ et sous-espaces fermés de $\mathbb{H}$.
- Intérêt théorique: unification du formalisme
- Intérêt pratique: l'extraction de l'information comme opération géométrique à partir de logiques conditionnelles, floues, quantiques.
- $\Rightarrow$ **treillis de propositions**

Treillis des propositions

Définition: Λ est un *treillis* si $\mathbf{0} \in \Lambda$ et $\mathbf{1} \in \Lambda$ et

1. idempotence: $a \wedge a = a = a \vee a$,
2. commutativité: $a \wedge b = b \wedge a$ and $a \vee b = v \vee a$,
3. associativité: $a \wedge (b \wedge c) = (a \wedge b) \wedge c$ et $a \vee (b \vee c) = (a \vee b) \vee c$,
4. identité: $a \wedge \mathbf{1} = a$ and $a \vee \mathbf{0} = a$,
5. absorption: $a \wedge (a \vee b) = a = a \vee (a \wedge b)$.

- $a \leq b \Leftrightarrow a \wedge b = a$: Treillis = ensemble partiellement ordonné (poset).
- a' complément de a si $a \wedge a' = \mathbf{0}$ et $a \vee a' = \mathbf{1}$.

Treillis de propositions

- Distributivité: $a \vee (b \wedge c) = (a \vee b) \wedge (a \vee c)$ (et duale).
- Modularité: $a \leq c$: $a \vee (b \wedge c) = (a \vee b) \wedge c$
- Orthomodularité: $a \leq c$: $a \vee (a' \wedge c) = c$

Définition: *Orthocomplémentation* $\perp$: $\Lambda \ni a \mapsto a^{\perp} \in \Lambda$, vérifiant pour $a, b \in \Lambda$:

1. $\perp$ injective,
2. $a \leq b \Rightarrow b^{\perp} \leq a^{\perp}$,
3. $(a^{\perp})^{\perp} = a$,
4. $a \wedge a^{\perp} = \mathbf{0}$.

Logique

Definition: Une *logique* Λ est un treillis orthocomplémenté t.q.

1. pour toute suite dénombrable $(a_n)_{n\in\mathbb{N}}$ d'éléments de Λ, $\vee_{n\in\mathbb{N}} a_n$ et $\wedge_{n\in\mathbb{N}}$ existent dans Λ,
2. si $a_1, a_2 \in \Lambda$ et $a_1 \leq a_2$, alors il existe $b \in \Lambda$, tel que $b \leq a_1^\perp$ et $b \vee a_1 = a_2$.

Spg orthomodularité avec $a' = a^\perp$.
Si distributivité vraie alors logique = σ-algèbre (tribu).

Reformulation (système)

Postulat 1': Dans tout système physique (classique ou quantique), l'ensemble des propositions expérimentalement vérifiables forme une logique (classique ou quantique standard).

Reformulation (observables)

Définition: Soit Λ une logique. Une *observable* réelle est une application

$$x : \mathcal{B}(\mathbb{R}) \to \Lambda, \ \text{t.q.}$$

1. $x(\emptyset) = \mathbf{0}; x(\mathbb{R}) = \mathbf{1}$,
2. $B_1 \cap B_2 = \emptyset \Rightarrow x(B_1) \perp x(B_2)$.
3. Si $(B_n)_n$ mutuellement disjoints alors $x(\cup_{n\in\mathbb{N}} B_n) = \vee_{n\in\mathbb{N}} x(B_n)$.

Ensemble des observables sur Λ: $\mathcal{O}(\Lambda)$.

$$\mathsf{spec}(x) = \cap_{C \text{ ferm}:x(C)=\mathbf{1}} C.$$

Postulat 2': Ensemble des observables d'un système physique (classique ou quantique), décrit par logique Λ est donné par $\mathcal{O}(\Lambda)$.

Reformulation (états)

Definition: Un *état* sur la logique Λ est une application $p : \Lambda \to [0,1]$ t.q.

1. $p(\mathbf{0}) = 0; p(\mathbf{1}) = 1$,
2. si $(a_n)_n$ propositions mutuellement orthogonales et $a = \vee_n a_n$ alors $p(a) = \sum_n p(a_n)$.

Ensemble des états noté $\mathcal{S}(\Lambda)$. Une *fonction d'état* est une applications

$$\sigma : \mathcal{O}(\Lambda) \ni x \mapsto \sigma(x) \in \mathcal{M}_1^+(\mathbb{R}, \mathcal{B}(\mathbb{R})).$$

Reformulation (états)

Théorème: $\mathcal{S}(\Lambda)$ est convexe (mais non nécessairement un simplex de Choquet.)
Théorème: Soit $p \in \mathcal{S}(\Lambda)$.

1. Si $\sigma_p : \mathcal{O}(\Lambda) \to \mathcal{M}_1^+(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ défini par

$$\sigma_p(x)(B) = p(x(B)), \forall B \in \mathcal{B}(\mathbb{R}),$$

   alors σ_p est une fonction d'état.

2. Si σ fonction d'état, alors $\exists ! p \in \mathcal{S}(\Lambda)$, t.q. $\forall x \in \mathcal{O}(\Lambda)$ et $\forall B \in \mathcal{B}(\mathbb{R})$,

$$\sigma(x)(B) = p(x(B)).$$

Reformulation (états)

$$\begin{aligned}\mathbb{E}_p(x) &= \int_{\mathsf{spec}(x)} t\sigma_p(x)(dt) \\ &= \int_{\mathsf{spec}(x)} tp(x(dt)).\end{aligned}$$

Postulat 3': Ensemble des états de tout système physique est $\mathcal{S}(\Lambda)$.

Postulat 4': Measurer physiquement une observable x signifie de déterminer sa fonction d'état $p(x(\cdot))$.

Reformulation (automorphismes et symmétries)

Lemme: Soit $\mathsf{Aut}(\Lambda)$ l'ensemble d'automorphismes de Λ et $\alpha \in \mathsf{Aut}(\Lambda)$. Application induite $\tilde{\alpha}$ sur $\mathcal{S}(\Lambda)$ par

$$\tilde{\alpha}(p)(a) = p(\alpha^{-1}(a)), a \in \Lambda, p \in \mathcal{S}(\Lambda),$$

est un automorphisme convexe de $\mathcal{S}(\Lambda)$.

Définition: G groupe topologique localement compact. L'application $\pi : G \to \mathsf{Aut}(\mathcal{S}(\Lambda))$ est une *représentation* si

1. $\pi(g_1 g_2) = \pi(g_1)\pi(g_2)$ pour tout $g_1, g_2 \in G$,
2. $\forall a \in \Lambda$, $\forall p \in \mathcal{S}(\Lambda)$, l'application $g \mapsto \pi(g)(p)(a)$ est $\mathcal{B}(G)$-mesurable.

Reformulation (automorphismes et symmétries)

Postulat 5': Évolution temporel de système isomlé engendrée par représentation du groupe abelien $(\mathbb{R}, +)$ sur l'ensembe d'automorphismes convexes de $\mathcal{S}(\Lambda)$.

Toute symétrie physique correspondant à G localement compact, induit une représentation sur les automorphismes convexes de $\mathcal{S}(\Lambda)$.

Logique quantique standard

- Logique: $\Lambda = \{$sous-espaces ferms du sparable $\mathbb{H}\}$.
Identifier $\Lambda \ni M$ avec projection P_M.
- Observables: mesures spectrales, i.e. projections x.
Identifier opérateurs auto-adjoints X avec x via
$X = \int tx(dt)$.
- États purs: correspondend aux rayons $\psi \in \mathbb{H}$ en définissant

$$\Lambda \ni M \mapsto p_\psi(M) \equiv \langle \psi \,|\, P_M \psi \,\rangle = \|P_M \psi\|^2.$$

- **Définition:** Un opérateur ρ sur $\mathbb{H}$ est une *metrice densité* s'il est borné, auto-adjoint, positif, de classe trace avec
$\mathsf{tr}\,\rho = 1$.
- Ensemble des matrices densité noté $\mathcal{D}(\mathbb{H})$.

Logique quantique standard

- **Exemple 1:** Soit $\psi \in \mathbb{H}$. Alors P_ψ projecteur sur $\mathbb{C}\psi$ est une matrice densité.
- **Exemple 2:** Soit $(\psi_n)_n$ suite arbitraire de rayons dans $\mathbb{H}$. Alors $\rho = \sum_n c_n P_{\psi_n}$ avec $\sum_n c_n = 1, c_n \geq 0$ est une matrice densité.
- **Définition:** Soit $\rho \in \mathcal{D}(\mathbb{H})$. Alors p, défini par

$$\Lambda \ni M \mapsto p(M) = \mathsf{tr}(\rho P_M),$$

est un état, appelé *état tracial.* Nous avons $\mathbb{E}_p(X) = \mathsf{tr}(\rho X)$.

Logique quantique standard

- Symétries: $\langle U\phi \,|\, U\psi \rangle = \langle \phi \,|\, \psi \rangle, \forall \phi, \psi \Rightarrow UU^* = U^*U = \mathbb{1}$
- Automorphisme $\alpha \in \mathsf{Aut}(\Lambda)$: $\alpha(M) = UM$
- Induitautomorphisme convexe $\tilde{\alpha}$ sur états traciaux par:

$$\tilde{\alpha}(p)(M) = p(\alpha^{-1}(M)) = \mathsf{tr}(\rho U^* P_M U) = \mathsf{tr}(U \rho U^* P_M).$$

- Se traduit à la transformation

$$\rho \mapsto U \rho U^*$$

sur matrices densité.

Logique quantique standard

- État après mesure de l'observable $X = \sum_j \lambda_j P_j$:
  - Classiquement:

$$\begin{aligned} \mathbb{P}(\cdot) &\mapsto \mathbb{P}(\cdot | X = \lambda_j), \quad \text{avec filtage} \\ \mathbb{P}(\cdot) &\mapsto \sum_j \mathbb{P}(\cdot | X = \lambda_j)\mathbb{P}(X = \lambda_j), \text{sans filtrage.} \end{aligned}$$

  - Quantiquement:

$$\begin{aligned} \rho &\mapsto \frac{P_j \rho P_j}{\mathrm{tr}(\rho P_j)}, \text{avec filtrage} \\ \rho &\mapsto \sum_j \frac{P_j \rho P_j}{\mathrm{tr}(\rho P_j)} \mathrm{tr}(\rho P_j), \text{sans filtrage.} \end{aligned}$$

Mesures à valeurs oprérateurs positifs

MVOP: $\Phi : \mathcal{D}(\mathbb{H}) \ni \rho \mapsto \sum_{i\in I} S_i \rho S_i^*$, avec $(S_i)_i$ isometries partielles vérifiant $\sum_{i\in I} S_i S_i^* \leq \mathbb{1}$.

Connexion avec C^* *algèbre de Cuntz* rencontrée en

- Espaces de décalage Non-commutatif
- Chaînes de Markov topologiques
- Graphes orientés
- Ondelettes
- Pavages
- ...

Grammaires et langages

- $\mathbb{A}_t$ alphabet fini de *symboles terminaux*
  $\mathbb{A}_{sc}$ alphabet fini de *catégories syntaxiques*
  $\mathbb{A}_t \cap \mathbb{A}_{sc} = \emptyset$, $\mathbb{A} = \mathbb{A}_t \cup \mathbb{A}_{sc}$, $\mathbb{A}^* = \cup_{n\geq 0}\mathbb{A}^n$
- $\Pi \subset (\mathbb{A}^+ \setminus \mathbb{A}_t^*) \times \mathbb{A}^*$ *productions*
- $s_0 \in \mathbb{A}_{sc}$ symbole initial
- *Grammaire:* $\Gamma = (\mathbb{A}_t, \mathbb{A}_{sc}, \Pi, s_0)$
- *Forme sententielle:* tout mot de $\mathbb{A}^*$
- Si $\alpha\beta\gamma \in \mathbb{A}^+$ et $\beta \to \delta \in \Pi$, alors $\alpha\delta\gamma \in \mathbb{A}^*$
- *Phrase* de Γ: tout mot sans catégories syntaxiques
- *Langage* $L(\Gamma)$: ensemble de phrases de Γ.

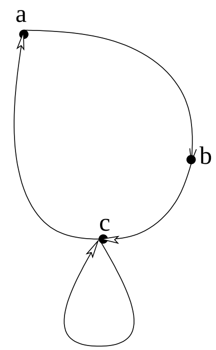
Ex. de grammaire réguilère

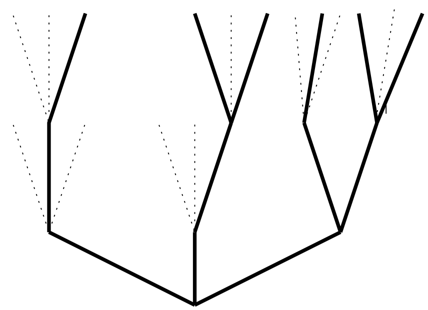
- $\mathbb{A}_t = \{0, 1, \ldots, 9, .\}$
- $\mathbb{A}_{sc} = \{n, f, d, i\}$
- $s_0 = n$
- Productions sous forme de Backus-Naur

$$
\begin{array}{rcl}
n & ::= & i \mid f \mid i\ f \\
f & ::= & .\ i \\
i & ::= & d \mid d\ i \\
d & ::= & 0 \mid 1 \mid \cdots \mid 9
\end{array}
$$

- Immediat que 137.28 est **reconnu** comme n par la grammaire mais pas 137.28.57.204.

Graphe orienté ⇔ grammaire





$$\begin{aligned} \mathbb{A}_{sc} &= \{a, b, c\} \\ a &::= b \\ b &::= c \\ c &::= c \mid a \end{aligned}$$

Ensemble de mots de longueur 6 commençant par a: abcccc, abccca, abccab, abcabc

Matrice d'adjacence des mots

En général graphe orienté ayant

- $\mathbb{A}^*$ (dénombrable) comme ensemble de vertex
- ensemble d'arêtes défini par *matrice d'adjacence de mots* $W : \mathbb{A}^* \times \mathbb{A}^* \to \{0, 1\}$ définie par

$$W(\alpha, \beta) = \begin{cases} 1 & \text{si } \exists (\alpha' \leq \alpha, \beta' \leq \beta) : (\alpha', \beta') \in \Pi \\ 0 & \text{sinon} \end{cases}$$

- Mots accessibles de α

$$D_\alpha = \{\beta \in \mathbb{A}^* : W(\alpha, \beta) = 1\}, \alpha \in \mathbb{A}^*$$

Types d'évolution en informatique théorique

- *Déterministe:* $\forall \alpha \in \mathbb{A}^* : \sharp D_\alpha = 1$. Alors $\mathbb{A}^*$ se sépare en plusieurs parties disjointes de trajectoires calculatoires.
- *Non-déterministe:* $\forall \alpha \in \mathbb{A}^* : 1 < \sharp D_\alpha < \infty$. Plusieurs trajectoires atteignent le même résultat.
- *Stochastique:* non-déterministe et pour tout α vecteur probabilité $\mathbf{p}_\alpha = (p_{\alpha,\beta}, \beta \in D_\alpha)$ avec $p_{\alpha,\beta} \geq 0$ et $\sum_{\beta \in D_\alpha} p_{\alpha,\beta} = 1$. Chaque trajectoire acquiert poids de probabilité.
- *Quantique:* non-déterministe et pour α vecteur unitaire $\mathbf{U}_\alpha = (U_{\alpha,\beta}, \beta \in D_\alpha)$ with $U_{\alpha,\beta} \in \mathbb{C}$ et $\sum_{\beta \in D_\alpha} |U_{\alpha,\beta}|^2 = 1$. Chaque trajectoire acquiert amplitude complexe de probabilité.

Hiérarchie des grammaires de Chomsky

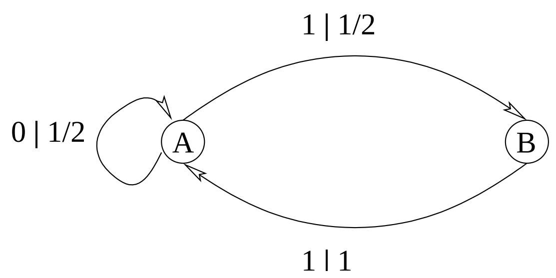
Type	Grammaire	Productions de la forme	Réconnaissance
0	recursivement énumerable	$\beta \to \beta'$	eTM
1	contextuelle	$\alpha b\gamma \to \alpha\beta\gamma,\ b \in \mathbb{A}_{sc}, \beta \neq \kappa$	eLBA
2	non-contextuelle	$b \to \beta,\ b \in \mathbb{A}_{sc},\ \beta \in \mathbb{A}^*$	ePDA
3	régulière	$b ::= c\beta \vert \beta,\ b, c \in \mathbb{A}_{sc},\ \beta \in \mathbb{A}_t^*$	eDFA

Préfixe e pour type d'évolution:

$$e \in \{D, N, S, Q\}.$$

Machine (automate) = système dynamique

L'exemple de chaînes de Markov cachées



$$\begin{array}{c|cccc} & 0A & 1A & 0B & 1B \\ 0A & 1/2 & 0 & 0 & 1/2 \\ 1A & 1/2 & 0 & 0 & 1/2 \\ 0B & 0 & 0 & 1 & 0 \\ 1B & 0 & 1 & 0 & 0 \end{array}$$

Phrases du type: $0011011000111011011110\ldots$

Théorème: CMC et grammaires régulières stochastiques sont équivalentes.

Ex. grammaire régulière stochastique

```
                1         11        21        31        41        51                  71        81        91        101       111       121       131
   Subdomain    .........<-I-----*-*-----*------------->.........<-II---------*----------------------->.<-III------------*-------------->....<-IV---
   PROSITE      ................AAAAAAAA.A.............................................................................................................
   X-ray        .........BBBB.BB.........BBBBB..BB..................B......BBBBBBB.BAAA..AA.A..........A..A.AAA.........A.A.AA......AAA..AA.A............
   X-ray        .........1111.11.........22222..22..................3......3333333.3BBB..BB.B..........B..C.CCC.........C.C.CC......CCC..CC.C............
 1 CAPK-ALPHA   .........FERI.KTLGTGSFGR.VMLVK..HK.....E..........TG.N......HYAMKIL.DKqk..VV.K..........LK.Q.IEH.........T.L.NE......KRI..LQ.AV....N...F...
 2 WEE1+        .........FRNV.TLLGSGEFSE.VFQVE..DP.....V..........EK.Tl.....KYAVKKL.KVk...FS.G..........PK.E.RNR.........L.L.QE......VSI..QR.AL....Kg..H...
 3 TIK          .........FEDI.EEIGLGGFGQ.VFKAK..HR.....I..........DG.K......RYAIKRV.KY....--.N..........TE.K.AE-.........-.-.HE......VQA..LA.EL....N...H...
 4 SPK1         df.......SIID.EVVGQGAFAT.VKKAI..ER.....T..........TG.K......TFAVKII.SKrk..VI.G..........NM.DgVT-.........-.-.RE......LEV..LQ.KL....N...H...
 5 RSK1-N       .........FELL.KVLGQGSFGK.VFLVR..KVtrp..D..........SG.H......LYAMKVL.KKa...TL.K..........VR.D.RVR.........T.K.ME......RDI..LA.DV....N...H...
 6 PYT          .........YSIL.KQIGSGGSSK.VFQVL..NE.....K..........K-.Q......IYAIKYV.NLe...EA.D..........NQ.T.LDS.........Y.R.NE......IAY..LN.KL....Qqh.S...
 7 PKC-ALPHA    .........FNFL.MVLGKGSFGK.VMLAD..RK.....G..........TE.E......LYAIKIL.KKdv..VI.Q..........DD.D.VEC.........T.M.VE......KRV..LA.LL....Dk..P...
 8 PDGFR-B      .........LVLG.RTLGSGAFGQ.VVEAT..AH.....G..........LS.Hsqatm.KVAVKML.KS....TA.R..........SS.E.KQA.........L.M.SE......LKI..MS.HL....Gp..H...
 9 PBS2         .........LEFL.DELGHGNYGN.VSKVL..HK.....P..........TN.V......IMATKEV.RL....EL.D..........EA.K.FRQ.........I.L.ME......LEV..LH.KC....N...S...
10 MIK1         .........FQQV.KPIHESDFSF.VYHVS..SInp...P..........TE.T......VYVVKML.KKn...AA.K..........FT.G.KER.........H.L.QE......VSI..LQ.RL....Qa..C...
11 MCK1         .........VKEY.RKIGRGAFGT.VVQAY..LT.....Q..........DK.Knwlg..PFAIKKV.PA....--.H..........TE.Y.KS-.........-.-.RE......LQI..LR.IA....D...H...
12 INS.R        .........ITLL.RELGQGSFGM.VYEGN..AR.....D..........II.Kgeaet.RVAVKTV.NE....SA.S..........LR.E.RIE.........F.L.NE......ASV..MK.GF....T...C...
13 HSVK         .........FTIH.GALTPGSEGC.VFDSS..HP.....D..........YP.Q......RVIVKAG.WY....-T.S..........TS.-.---.........-.-.HE......ARL..LR.RL....D...H...
14 ERK1         .........YTQL.QYIGEGAYGM.VSSAY..DH.....V..........RK.T......RVAIKKI.SP....FE.H..........QT.Y.CQR.........T.L.RE......IQI..LL.GF....R...H...
15 EGFR         .........FKKI.KVLGSGAFGT.VYKGL..WI.....P..........EG.Ekvki..PVAIKEL.RE....AT.S..........PK.A.NKE.........I.L.DE......AYV..MA.SV....D...N...
16 ECK          .........VTRQ.KVIGAGEFGE.VYKGM..LKt....S..........SG.Kkev...PVAIKTL.KA....GY.T..........EK.Q.RVD.........F.L.GE......AGI..MG.QF....S...H...
17 DPYK1        ne.......LEFG.QTIGKGFFGE.VKRGY..WR.....E..........T-.-......DVAIKII.YRdq..FK.T..........KS.S.LVM.........F.Q.NE......VGI..LS.KL....R...H...
18 CLK          ar.......YEIV.DTLGEGAFGK.VVECI..DH.....K..........VGgR......RVAVKIV.KN....--.V..........DR.Y.CEA.........A.Q.SE......IQV..LE.HL....Ntt.D...
19 CDC2HS       .........YTKI.EKIGEGTYGV.VYKGR..HK.....T..........TG.Q......VVAMKKI.RLe...SE.E..........EG.V.PST.........A.I.RE......ISL..LK.EL....R...H...
20 CAMII-ALPHA  .........YQLF.EELGKGAFSV.VRRCV..KV.....L..........AG.Q......EYAAKII.NTk...KL.S..........AR.D.HQK.........L.E.RE......ARI..CR.LL....K...H...
21 C-SRC        .........LRLE.VKLGQGCFGE.VWMGT..WN.....G..........T-.T......RVAIKTL.KP....--.G..........TM.S.PEA.........F.L.QE......AQV..MK.KL....R...H...
22 C-RAF        .........VMLS.TRIGSGSFGT.VYKGK..WH.....G..........D-.-......-VAVKIL.KVv...DP.T..........PE.Q.FQA.........F.R.NE......VAV..LR.KT....R...H...
23 KLSK_HUMAN   mgcgcss244LKLV.ERLGAGQFGE.VWMGY..YN.....G..........H-.T......KVAVKSL.KQ....--.G..........SM.S.PDA.........F.L.AE......ANL..MK.QL....Q...H...
24 KLSK_MOUSE   mgcvcss244LKLV.ERLGAGQFGE.VWMGY..YN.....G..........H-.T......KVAVKSL.KQ....--.G..........SM.S.PVP.........F.L.AE......ANL..MK.QL....Q...H...
25 ARKB_HUMAN   madleav190FSVH.RIIGRGGFGE.VYGCR..KR.....D..........TG.K......MYAMKCL.DKk...RI.K..........MK.Q.GET.........LaL.NE......RIM..LS.LVstg.D...C...
26 ARKB_BOVIN   madleav190FSVH.RIIGRGGFGE.VYGCR..KA.....D..........TG.K......MYAMKCL.DKk...RI.K..........MK.Q.GET.........LaL.NE......RIM..LS.LVstg.D...C...
27 BYR1_SCHPO   mfkrrrnp65LEVV.RHLGEGNGGA.VSLVK..HR.....-..........-N.I......FMARKTV.YV....GS.D..........SK.L.QKQ.........I.L.RE......LGV..LH.HC....R...S...
28 CYGR_ARBPU   mattrll579QQIF.ATIG------.----T..YR.....G..........T-.-......ICAIHAVhKN....--.H..........ID.L.TRA.........V.R.TE......LKL..MR.DM....R...H...
29 ANPA_RAT     mpgsrrv536GSLL.TT--EGQF-Q.VFAKT..AY.....Y..........KG.N......LVAVKRVnRK....--.R..........IE.L.TRK.........V.L.FE......LKH..MR.DV....Q...N...
30 ANPA_HUMAN   mpgprrp471LEVL.ALVGSLS---.-LLGI..LI.....Vsffiyrkm70KG.N......LVAVKRVnRK....--.R..........IE.L.TRK.........V.L.FE......LKH..MR.DV....Q...N...
31 ANPB_HUMAN   malpsll512SRLT.LSLRGSSYGS.LMTAH..GKyqifaN..........TG.Hfkgn..VVAIKHV.NK....-K.R..........IE.L.TRQ.........V.L.FE......LKH..MR.DV....Q...F...
32 ANPA_MOUSE   mprsrrv536GSLL.TT--EGQF-Q.VFAKT..AY.....Y..........KG.N......LVAVKRVnRK....--.R..........IE.L.TRK.........V.L.FE......LKH..MR.DV....Q...N...
33 ANPB_RAT     malpsll512SRLT.LSLRGSSYGS.LMTAH..GKyqifaN..........TG.Hfkgn..VVAIKHV.NK....-K.R..........IE.L.TRQ.........V.L.FE......LKH..MR.DV....Q...
```

Ex.grammire non-contextuelles stochastique

Y. Sakakibara. M. Brown, R. Hughey, I. S. Mian, K. Sjölander, R. C. Underwood, and D. Haussler, Stochastic Context-Free Grammars for tRNA modeling", Nucleic Acids Research, 22:5112-5120 (1994)

a. Productions

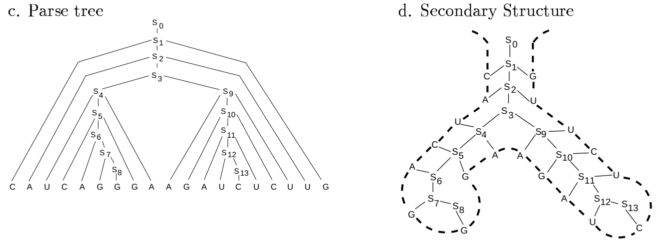
$$P = \{ \begin{array}{ll} S_0 \to S_1, & S_7 \to \mathtt{G}\ S_8, \\ S_1 \to \mathtt{C}\ S_2\ \mathtt{G}, & S_8 \to \mathtt{G}, \\ S_1 \to \mathtt{A}\ S_2\ \mathtt{U}, & S_8 \to \mathtt{U}, \\ S_2 \to \mathtt{A}\ S_3\ \mathtt{U}, & S_9 \to \mathtt{A}\ S_{10}\ \mathtt{U}, \\ S_3 \to S_4\ \ S_9, & S_{10} \to \mathtt{C}\ S_{10}\ \mathtt{G}, \\ S_4 \to \mathtt{U}\ S_5\ \mathtt{A}, & S_{10} \to \mathtt{G}\ S_{11}\ \mathtt{C}, \\ S_5 \to \mathtt{C}\ S_6\ \mathtt{G}, & S_{11} \to \mathtt{A}\ S_{12}\ \mathtt{U}, \\ S_6 \to \mathtt{A}\ S_7, & S_{12} \to \mathtt{U}\ S_{13}, \\ S_7 \to \mathtt{U}\ S_7, & S_{13} \to \mathtt{C} \end{array} \}$$

b. Derivation

$$\begin{array}{ll} S_0 & \Rightarrow S_1 \Rightarrow \mathtt{C}S_2\mathtt{G} \Rightarrow \mathtt{CA}S_3\mathtt{UG} \Rightarrow \mathtt{CA}S_4S_9\mathtt{UG} \\ & \Rightarrow \mathtt{CAU}S_5\mathtt{A}S_9\mathtt{UG} \Rightarrow \mathtt{CAUC}S_6\mathtt{GA}S_9\mathtt{UG} \\ & \Rightarrow \mathtt{CAUCA}S_7\mathtt{GA}S_9\mathtt{UG} \Rightarrow \mathtt{CAUCAG}S_8\mathtt{GA}S_9\mathtt{UG} \\ & \Rightarrow \mathtt{CAUCAGGGA}S_9\mathtt{UG} \Rightarrow \mathtt{CAUCAGGGAA}S_{10}\mathtt{UUG} \\ & \Rightarrow \mathtt{CAUCAGGGAAG}S_{11}\mathtt{CUUG} \\ & \Rightarrow \mathtt{CAUCAGGGAAGA}S_{12}\mathtt{UCUUG} \\ & \Rightarrow \mathtt{CAUCAGGGAAGAU}S_{13}\mathtt{UCUUG} \\ & \Rightarrow \mathtt{CAUCAGGGAAGAUCUCUUG}. \end{array}$$

c. Parse tree

d. Secondary Structure



Processus asynchrones classiques

Compilation de plusiers articles sur des sujets différents:
Gravitation quantique: Malyshev (98), Malyshev (00)
Chaos multiplicatif: Kahane-Peyrière (78), Collet-Koukiou (88), Liu (98), Menshikov-P (01)

Rwre sur des arbres: Comets-Menshikov-Popov (98), Menshikov-P (01),

Menshikov-P-Popov (02)

- Productions $\beta a\gamma := \delta \to \delta' := \beta\alpha\gamma$
- Chaîne de Markov à temps continu

$$\mathbb{P}(X_t = \delta' | X_0 = \delta) = (\exp(-tH))_{\delta\delta'}, \delta, \delta' \in \mathbb{A}^*$$

- Processus induit sur $\mathbb{A}^{\mathbb{Z}}$ par limite thermodynamique
- **Théorème:** Sous conditions sur productions, la limite thermodynamique pour le processus sur $\mathbb{A}^{\mathbb{Z}}$ existe.

Processus asynchrones classiques

$$\Omega = \{\omega : [0, \infty[\to \mathbb{A}^{\mathbb{Z}} \text{ t.q. } \omega \text{ admissible } \}$$

Admissible signifie:

- continu à droite
- compatible avec productions, *i.e.* si $\omega(s^-) = \alpha\beta\gamma$ and $\omega(s) = \alpha\beta'\gamma$ alors $(\beta, \beta') \in \Pi$

Pour $\alpha \in \mathbb{A}^{\mathbb{Z}}$ noter $\alpha_{[-N,N]}$ sa restriction dans $\{-N, -N+1, \ldots, N\}$. Soit $-N \leq i < 0 \leq j \leq N$ et

$$\begin{aligned} \Omega^{(N)}(\alpha) &= \{\omega \in \Omega : \omega(0)_{[-N,N]} = \alpha_{[-N,N]}\} \\ \Omega^{(N)}_{ij;t}(\alpha) &= \{\omega \in \Omega^{(N)}(\alpha) : \ \alpha_i \text{ et } \alpha_j \text{ ne changent pas pour } s \in [0, t]\} \end{aligned}$$

Processus asynchrones classiques

$$\lim_{N\to\infty} \mathbb{P}(\Omega_{ij;t}^{(N)}(\alpha)) = P(t;i,j), \qquad \sum_{(i,j):i<0\leq j} P(t;i,j) = 1$$

$$\begin{array}{ccccccccccccccc}
s_4 & \ldots & A & G & G & C & \cdot & T & T & C & A & T & A & C & G & \ldots \\
s_3 & \ldots & A & G & G & C & \cdot & T & T & C & T & T & A & C & G & \ldots \\
s_2 & \ldots & A & G & G & C & \cdot & T & T & C & A & T & A & C & G & \ldots \\
s_1 & \ldots & A & G & G & C & T & T & T & C & T & T & A & C & G & \ldots \\
0 & \ldots & A & G & G & C & T & A & & C & T & T & A & C & G & \ldots \\
 & & -N & & & i & & & & j & & & & & N &
\end{array}$$

Processus asynchrones quantiques

- Espace de Hilbert $\mathbb{H} = \ell^2(\mathbb{A}^*)$, avec base vérifiant $\langle\, e_\alpha \,|\, e_\beta \,\rangle = \delta_{\alpha,\beta}$
- Productions

$$\pi := (\beta a \gamma := \zeta \to \zeta' := \beta \alpha \gamma), |\beta| = j - 1$$

engendrées par des opérateurs

$$A_\pi(j) e_\zeta = e_{\zeta'}; \quad A_\pi^*(j) e_{\zeta'} = e_\zeta$$

- Hamiltonien formel

$$H = \sum_{\pi \in \Pi} \sum_{j \in \mathbb{N}} (\lambda_\pi A_\pi(j) + \overline{\lambda}_\pi A_\pi^*(j)).$$

Processus asynchrones quantiques

- Volume fini: $\mathbb{H}_N = \mathsf{span}\{e_\alpha : |\alpha| \leq N\}$.
- Operateurs en volume fini: $A_{\pi,N}(j) = P_N A_\pi(j) P_N$
- $\mathfrak{A}_N$ est la C^*-algèbre agissant sur $\mathbb{H}$ engendrée par $(P_N A_\pi(j) P_N, \pi \in \Pi; j \in \mathbb{N})$
- $\mathfrak{A}_N \simeq \mathcal{M}_{\frac{r^{N+1}-1}{r-1}}$, où $r = |\mathbb{A}|$
- $\phi_N : \mathfrak{A}_N \to \mathfrak{A}_{N+1}$ défini par $\phi_N : B \mapsto \begin{pmatrix} B & 0 \\ 0 & 0 \end{pmatrix}$
- *Algèbre locale:* $\mathfrak{A}_\infty = \varinjlim(\mathfrak{A}_N, \phi_N)$.
- *Algèbre quasi-locale:* $\mathfrak{A} = \overline{\mathfrak{A}_\infty}$

Processus asynchrones quantiques

- Groupe d'automorphismes sur $\mathfrak{A}$:

$$\tau_t^{(N)}(B) = \exp(itH_N)B\exp(-itH_N).$$

- **Théorème:** $\exists t_0 > 0$: $B \in \mathfrak{A}$ et $\forall t$ avec $|t| < t_0$ la norme $\lim_{N\to\infty} \tau_t^{(N)}(B)$ existe et définit unique groupe $(\tau_t)_t$ d'automorphismes sur $\mathfrak{A}$

États KMS

- **Définition:** Soit $(\mathfrak{A}, \tau)$ un C^*-système dynamique. L'état ϕ sur $\mathfrak{A}$ est un **état τ-KMS à température inverse** β si

$$\phi(A\tau_{i\beta}(B)) = \phi(AB).$$

- Naïvement états KMS:

$$\langle\, B \,\rangle_\beta = \lim_N \langle\, B \,\rangle_{\beta,N} = \lim_N \frac{\mathsf{tr}_N(B \exp(-\beta H_N))}{Z_N}.$$

Mais

- **Lemme:** Pour β petit
  1. $\langle\, B \,\rangle_\beta = 0$
  2. La pression existe

États KMS

- Introduire subsititutions triviales $a \to a, a \in \mathbb{A}$ et noter $A_a(j)$ les opérateurs de production correspondant
- Soit $P_{=N}$ la projection orthogonale sur $\mathbb{H}_N \ominus \mathbb{H}_{N-1}$
- Pour $\mu > 0$, $\mu H^0 = \mu \sum_{N=0}^{\infty} N P_{=N} = \mu \sum_{a \in \mathbb{A}} \sum_j A_a(j)$
- Utiliser l'Hamiltonien formel $H' = H + \mu H^0$ pour définir états KMS
- Il est naturel de s'attendre à une transition de phase à μ_{cr}

Algorithmes

- Inversion de Möbius
- Recuit simulé
- Résaux neuronaux
- Algorithmes génétiques
- ...

"I think of my lifetime in physics as divided into three periods. In the first period [. . .] I was in the grip of the idea that Everything is Particles. [. . .] I call my second period Everything is Fields. [. . .] Now I am in the grip of a new vision, that Everything is Information."

– John Archibald WHEELER

Geons, Black Holes, and Quantum Foam