



Algorithmes génétiques



Optimisation Combinatoire

- S : ensemble de solutions potentielles de cardinal au plus *dénombrable* (souvent fini de grande taille)
- Problème combinatoire :
- Trouver la ou les solutions de S convenable
- Optimisation combinatoire :
- $f : S \rightarrow \mathbb{R}$ fonction à optimiser (ou de coût)
- Trouver la ou les solutions de S donnant la ou les plus grandes (ou plus petites) valeurs pour f .



Optimisation Combinatoire

- Exemples :
 - Affectation de fréquence en téléphonie
 - Le problème du sac-à-dos
 - Couverture d'ensemble
 - Découpage de verre sans perte
 - Routage de véhicules
 - Le voyageur de commerce
 - Yield management : gestion de ressource
 - Horaire de train
 - et même plus.....

Métaheuristique

Du grec *heuriskein* : trouver/découvrir (heureka)



C'est un ensemble d'algorithmes visant à résoudre des problèmes d'optimisation difficile pour lesquels **on ne connaît pas de méthode classique plus efficace**.

Les métaheuristiques sont généralement des algorithmes **stochastiques itératifs**, qui progressent vers l'optimum global d'une fonction, par **échantillonnage** d'une fonction objectif.

Elles se comportent comme des algorithmes de recherche, tentant **d'apprendre** les caractéristiques d'un problème afin d'en trouver une **approximation** de la meilleure solution.

Ce sont des solutions **génériques** pouvant s'adapter à une grande variété de problèmes.

Métaheuristiques



- Classification :
 - Méthodes exactes de construction :
 - branch and bound, simplex
 - Algorithmes évolutifs :
 - Algo. Génétiques
 - Programmation Génétique
 - Stratégies d'évolution
 - Recherche locale
 - Méthode de descente (Hill-Climbing)
 - Recuit Simulé
 - Tabou
 - ...

Utilisation et domaines d'application



- Voir l'évolution d'une population : Simulation
- Création d'un pool pour tests
- Enrichissement automatique de bases de connaissances

Sommaire



Algorithmes génétiques

Stratégies évolutives

Classifieurs génétiques

Le modèle ECHO de J. Holland

Introduction : les systèmes évolutifs



S'inspire du Néo-Darwinisme, synthèse de la théorie de la sélection naturelle (Darwin) et de la théorie de l'hérédité (Mendel) :

- Evolution : résultat d'une altération progressive des êtres vivants au cours des générations
- Reproduction basée sur le caractère génétique qui subit au cours des générations des recombinaisons et des mutations
- Mécanisme de sélection naturelle

Algorithmes génétiques



- Darwin : survie du plus robuste
- (1973) Holland : évolution ordinateurs / humains
- (1989) Goldberg : individu = environnement + ADN

Principe

1 paramètre \approx 1 chromosome \approx 1 individu

1 solution = 1 ensemble de chromosomes

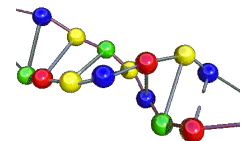
Représentation binaire

Opérateurs génétiques

Mutation

Croisement

- Programmation génétique





Définition

AG = Procédé d'optimisation permettant de trouver le minimum/maximum d'une fonction à partir d'un ensemble d'éléments choisis au hasard!

Introduction : les systèmes évolutifs



Un individu n'a pas été conçu pour réaliser des actions spécifiques, mais il peut évoluer

- Il détermine comment évoluer (génétiquement, en adaptant son corps et son esprit)
- Il organise ses actions en fonction de ses buts
- Il construit son « prototype »

Principe des algorithmes génétiques



- S'adapter c'est évoluer pour être au mieux en adéquation avec son environnement
- Simuler cette évolution en 3 phases importantes :
 - Reproduction
 - Croisement
 - Mutation



Principe des algorithmes génétiques

- Le processus de sélection est basé sur une fonction d'évaluation (ou fonction d'objectif) qui correspond à une performance de l'individu

On en déduit une probabilité pour :

- chaque individu de se reproduire ou de générer des clones (cette probabilité s'appelle *fitness*)
- Finalement, ce processus contribue à produire une population ayant une meilleure adaptabilité (convergence vers l'optimum)

Fonctionnement

- Imaginons une population d'individus et son rythme d'évolution
- Comment la modéliser : 1 objet/entité = 1 individu
- Comment les différencier : 1 individu = 8 bits (exemple des chromosomes et des allèles)



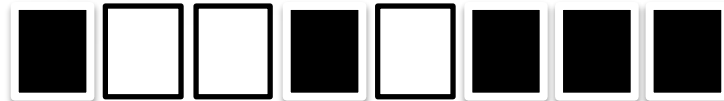
Individu 1 = Bob

Clonage

- Individu 1 = Bob



- Individu 2 = John



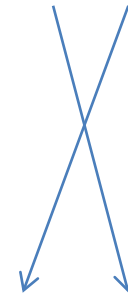
- Question : John est le clone de Bob mais est-il Bob?

Croisement

- Individu 1 = Bob



- Individu 2 = John



John n'est pas Bob

Mutation

- Individu 1 = Bob



- Individu 2 = John



John n'est pas Bob



Exemple avec $f(x)=x^2$

- Population totale possible : nombre compris entre 0 et 31
- Codage binaire sur 5 bits (00000 à 11111)
- On recherche donc le maximum de la fonction

On choisi 4 individus au hasard

- 13 de code 01101
- 24 de code 11000
- 8 de code 01000
- 19 de code 10011

Fonction de performance pour chaque individu

- $f(x) = x^2$
- L'évolution, la fonction « fitness » est une répartition par pourcentage de ces performances

Nbr. individus	Performances	Fitness
13	169	0,14
24	564	0,49
8	64	0,06
19	361	0,31

Génération suivante : reproduction



- On choisi au hasard 4 nouvelles chaînes en tenant compte de la valeur de répartition
- On obtient, par exemple : 13, 24, 24 et 19.

Génération suivante : croisement

- On prend les individus 2 par 2
- On coupe les chromosomes à une position aléatoire et on croise les parties coupées

13 : 0110 || 1 → 01100 : 12

24 : 1100 || 0 → 11001 : 21

24 : 11 || 000 → 11011 : 27

19 : 10 || 011 → 10000 : 16

Génération suivante : mutation



- On prend aléatoirement à partir d'une probabilité de mutation définie initialement
- On prend ici 0,05 donc sur les 20 bits des 4 chromosomes, on suppose qu'un seul allèle a été sélectionné et mute

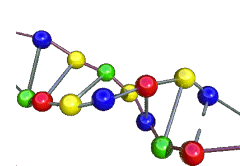
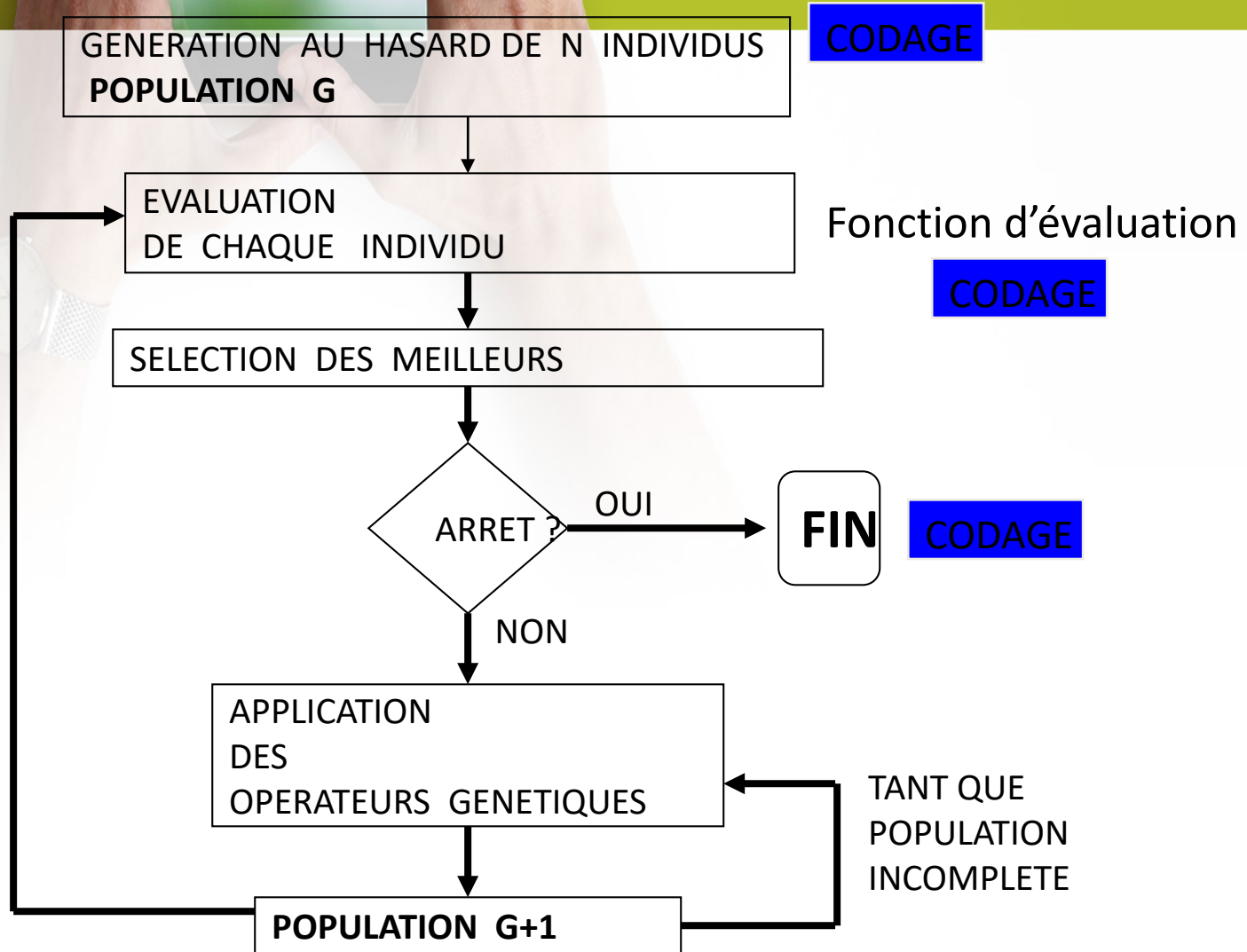
12 : 01100 → 01000 : 8



Conclusion

- La nouvelle génération est donc : 8, 21, 27, 16.
- Si on ajoute les performances (nombre données par les bits qui codent les individus), on obtient 1490 au lieu de 1170

Algorithme





Résumé et !!!

- On choisi une population et son rythme d'évolution
- On défini la modélisation des individus
- On paramètre l'utilisation des différentes méthodes
- On défini le but à atteindre

Pas forcément tout dans cet ordre!



Améliorations possibles

- Les améliorations possibles peuvent être nombreuses ...
- Autres opérateurs que le croisement et la mutation :
 - inversion de sous-chaînes, croisement à partir d'un masque binaire (dont les valeurs indiquent les positions de croisement), etc ...
- Variante inspirée de la théorie des niches écologiques et des espèces : on conserve en partie la population de la génération précédente

Evolution Différentielle (DE)



Storn & Price (Berkeley - 1995)

AG pauvres de concepts

1 chromosome = 1 ensemble de paramètres

Représentation binaire / réelle

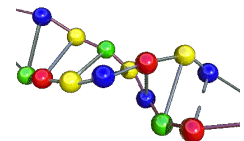
Croisement : addition / soustraction / perturbation de vecteurs

Sélection :

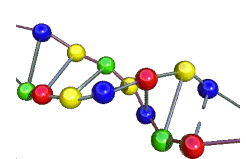
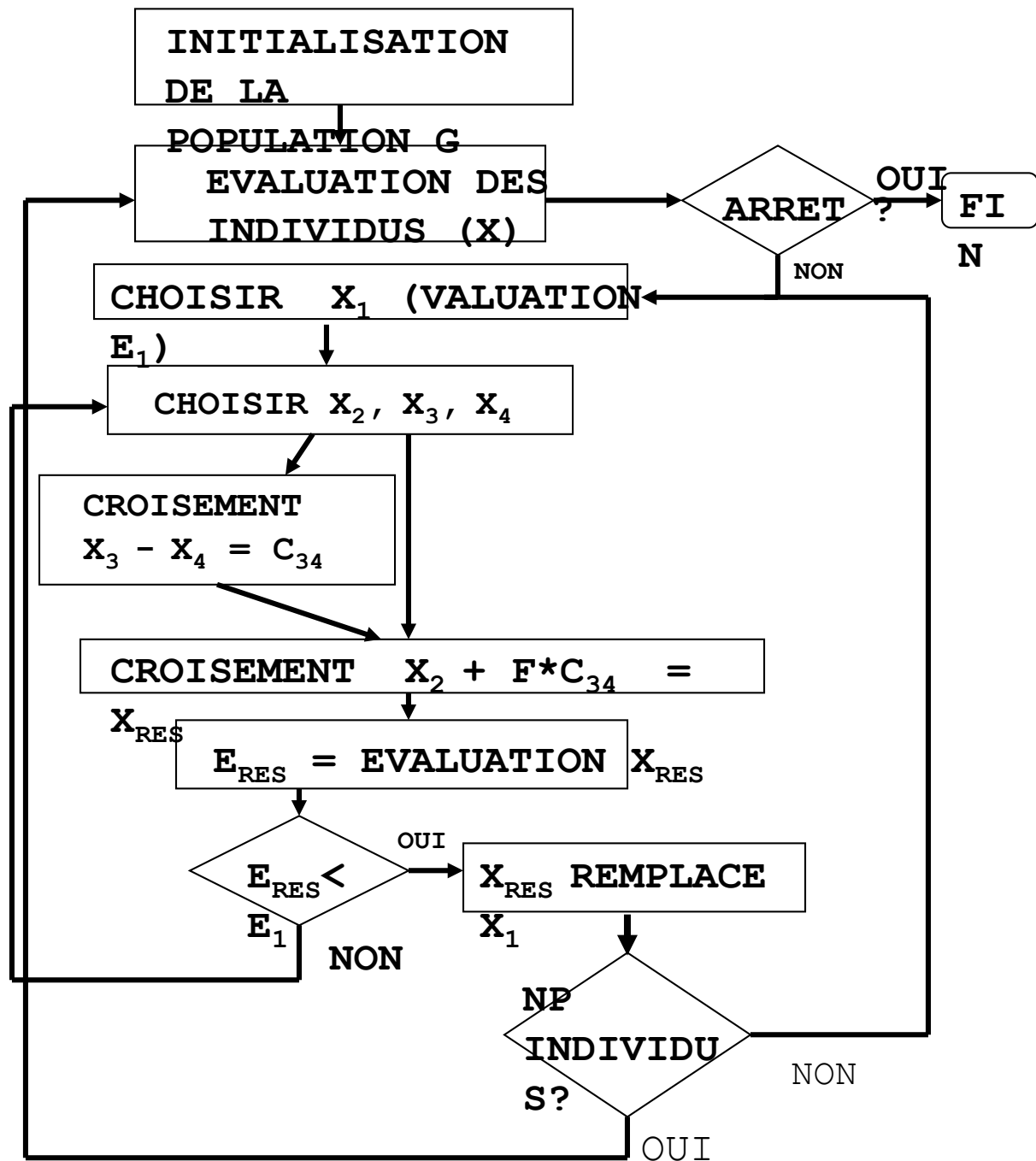
- l'élitisme
- la population sans double

Nouveaux mécanismes de reproduction

AG avancés



Algorithme DE



Méthodes de croisement

DE / RAND / 1

$$V_{iG+1} = X_{1G} + F * (X_{2G} - X_{3G})$$

F constante de croisement

DE / RAND / 2

$$V_{iG+1} = X_{1G} + F * (X_{2G} + X_{3G} - X_{4G} - X_{5G})$$

DE / BEST / 1

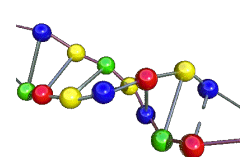
$$V_{iG+1} = X_{\text{best } G} + F * (X_{2G} - X_{3G})$$

DE / BEST / 2

$$V_{iG+1} = X_{\text{best } G} + F * (X_{2G} + X_{3G} - X_{4G} - X_{5G})$$

Algorithmes évolutionnaires différentiels hybrides (HDE)

F est choisi aléatoirement



Modèles utilisés pour l'évaluation des AG

Modèles biologiques

Systemes non linéaires

Grand nombre de paramètres

Basés sur la réalité

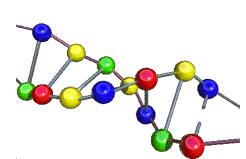
Contraintes

Intervalles de recherche

Données réelles / Fiabilité du modèle

Simulation avec des fonctions mathématiques

Fonction d'évaluation



Evaluation sur des données simulées

Modèle

$$\frac{dX}{dt} = \mu * X$$

$$\frac{dS}{dt} = -\mu * \frac{X}{Y1}$$

$$\frac{dP}{dt} = -\mu * X * \frac{Y2}{Y1}$$

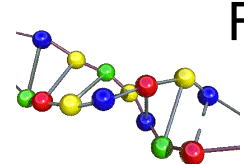
$$\mu = \mu_M * \frac{S}{K_S + S} * \left(1 - \frac{P}{P_M}\right)$$

	EXACT	Recherche
$\mu_M (h^{-1})$	0.38	[0 - 1]
$K_S (g.L^{-1})$	5	[0 - 10]
$P_M (g.L^{-1})$	100	[80 - 120]
Y1	0.07	[0 - 1]
Y2	0.44	[0 - 1]

A estimer : μ_M K_S P_M Y1 Y2

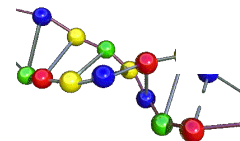
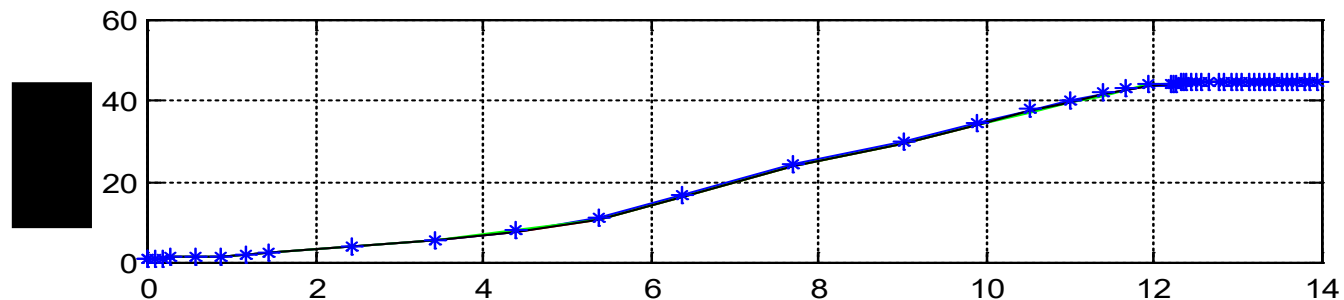
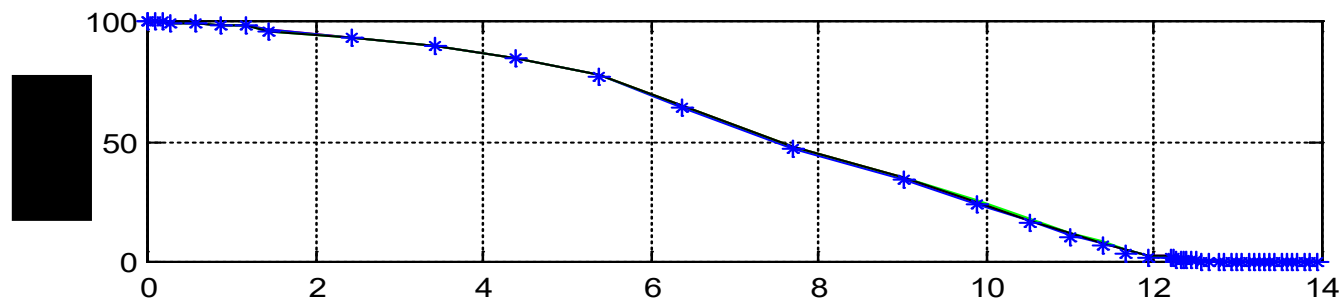
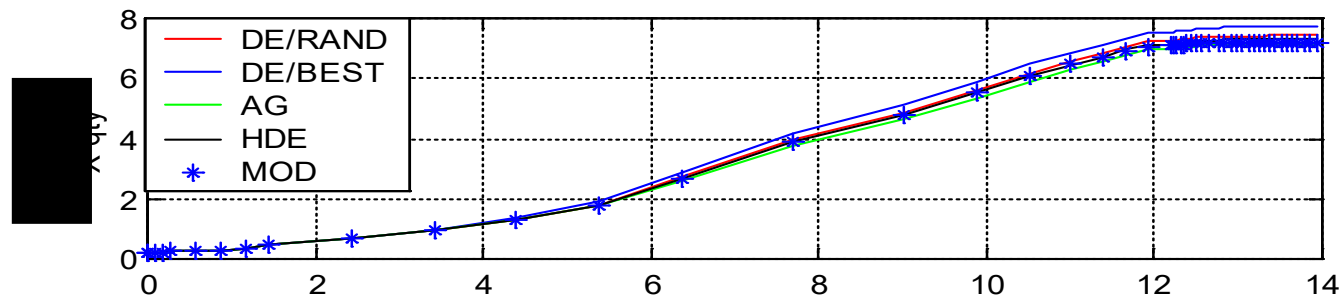
X, S et P seront simulés et comparés aux valeurs de base

Fonction objectif : minimiser l'erreur quadratique sur S, P et X



Evaluation sur des données simulées

Résultats obtenus



Evaluation sur des données expérimentales

Modèle

$$\frac{dXt(t)}{dt} = \mu * Xv(t)$$

$$\frac{dXv(t)}{dt} = (\mu - \mu_D) * Xv(t)$$

$$\frac{dS(t)}{dt} = q_s * Xv(t)$$

$$\frac{dP(t)}{dt} = q_p * Xv(t)$$

$$\frac{dG(t)}{dt} = q_g * Xv(t)$$

$$\mu = \mu_M + \frac{S}{K_{SX} + S} * \left(1 - \frac{P}{K_{PINX}}\right) * \left(1 - \frac{Xt(t)}{K_{XTIN}}\right)$$

$$\mu_D = \alpha * \mu + \beta$$

$$q_p = q_{pmax} * \frac{S}{K_{SP} + S} * \left(1 - \frac{P}{K_{PINP}}\right)$$

$$q_g = \gamma + \mu + \varepsilon$$

$$q_s = \frac{\mu}{Y_{SX}} + \frac{q_p}{Y_{SP}} + \frac{q_g}{Y_{SG}}$$

15 paramètres à estimer

	Littérature	Recherche
μ_M (h ⁻¹)	0.2 - 0.52	[0 - 1]
K_{SX} (g.L ⁻¹)	0.22 - 5	[0 - 10]
K_{SP} (g.L ⁻¹)	0.33 - 2.75	[0 - 10]
q_{pmax} (g.g ⁻¹ .h ⁻¹)	1.4 - 3.2	[0 - 10]
K_{PINX} (g.L ⁻¹)	76 - 108	[60 - 120]
K_{PINP} (g.L ⁻¹)	93 - 140	[80 - 160]
maint	0 - 2.16	[0 - 10]
$\gamma, \beta, Y_{SX}, Y_{SG}, Y_{SP}, \varepsilon$	-	[0 - 1]

Fonction objectif basée sur :

Erreur quadratique

$$\text{Err1} = (X_t - X_{t_0})^2 + (X_v - X_{v_0})^2 + (S - S_0)^2 + (P - P_0)^2 + (G - G_0)^2$$

Erreur de Wang

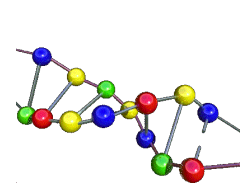
$$\text{Err2} = \frac{(X_t - X_{t_0})^2}{X_{t_{\max}}^2} + \frac{(X_v - X_{v_0})^2}{X_{v_{\max}}^2} + \frac{(S - S_0)^2}{S_{\max}^2} + \frac{(P - P_0)^2}{P_{\max}^2} + \frac{(G - G_0)^2}{G_{\max}^2}$$

Erreur relative

$$\text{Err3} = \frac{(X_t - X_{t_0})^2}{X_{t_0}^2} + \frac{(X_v - X_{v_0})^2}{X_{v_0}^2} + \frac{(S - S_0)^2}{S_0^2} + \frac{(P - P_0)^2}{P_0^2} + \frac{(G - G_0)^2}{G_0^2}$$

X_t = données réelles

X_{t_0} : données estimées



Evaluations sur des données expérimentales

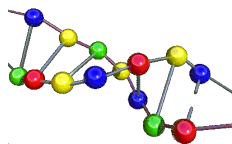
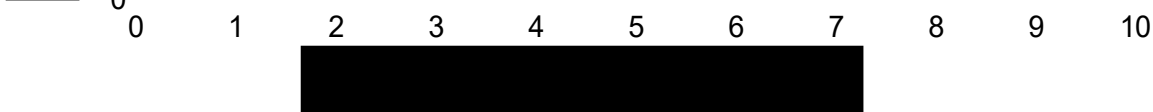
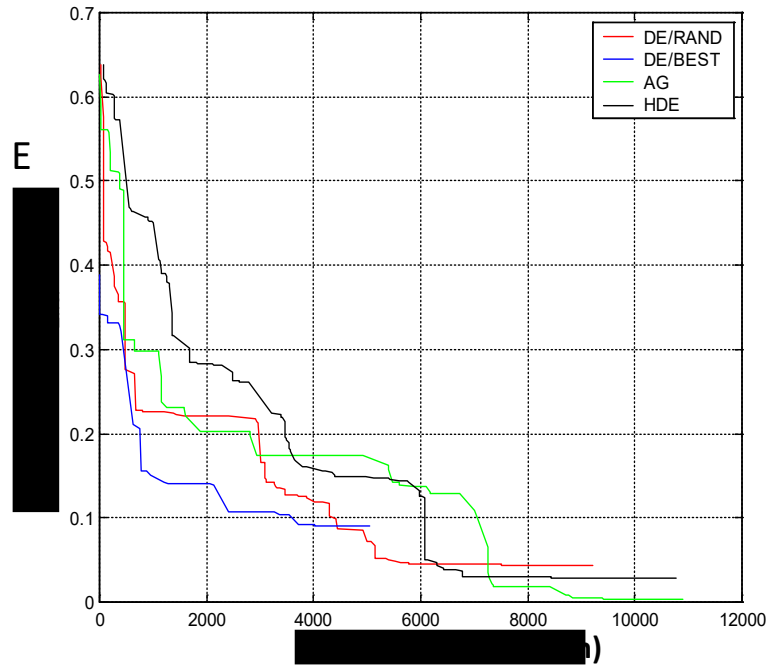
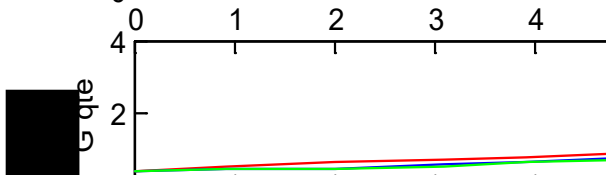
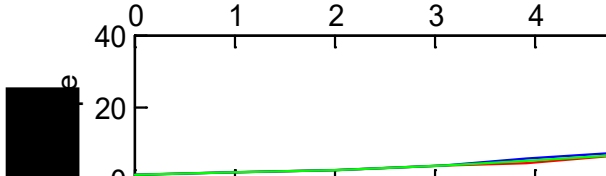
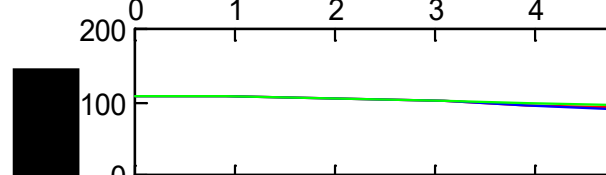
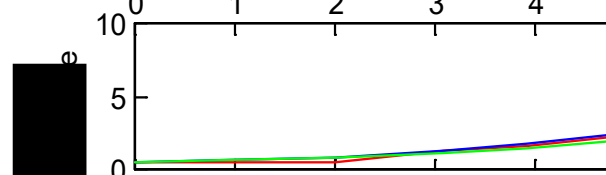


Re

Bo

Pa

Pe

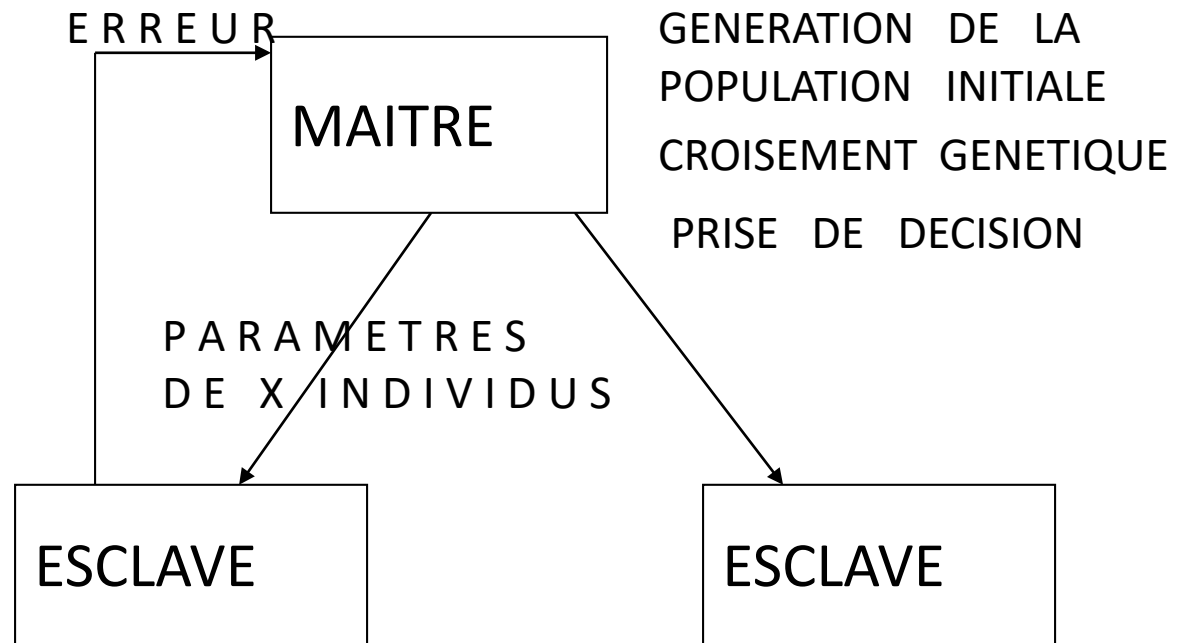


Parallélisme

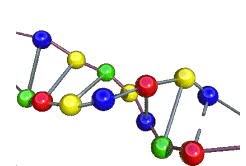
Lampinen (1999)

MPI : Message Passing Interface

Maître / Esclave



POUR CHAQUE INDIVIDU
SIMULATION D'UN MODELE
CALCUL DE L'ERREUR



Résultats obtenus

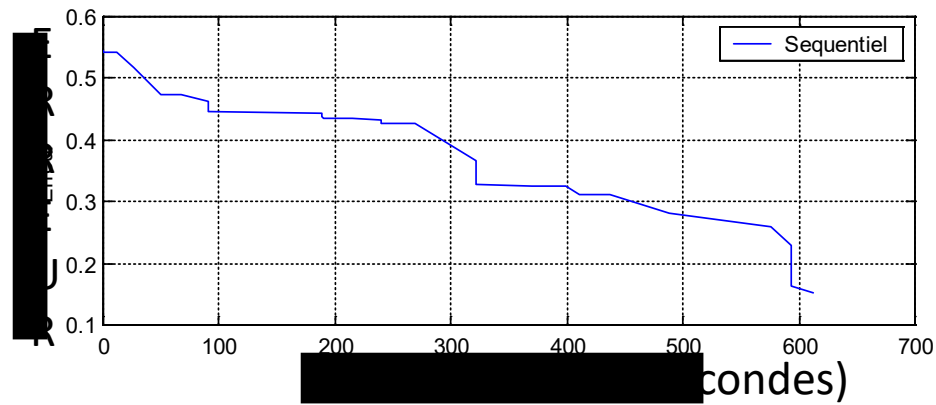
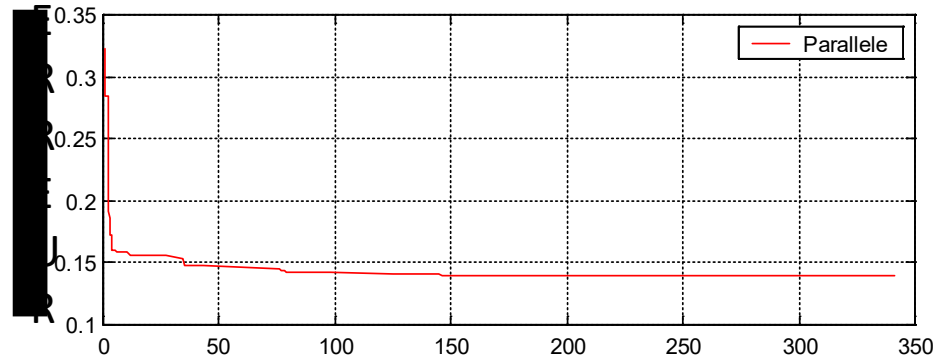
Temps séquentiel / parallèle

$$\text{Grain} = \frac{T_{\text{SEQ}}}{T_{\text{PAR}}}$$

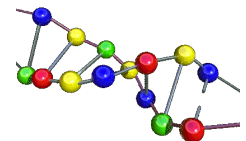
Précautions

Charges

Problème



(secondes)





AG = Méthodes d'optimisation efficaces

Optimisation des AG

Sensibilité au bruit

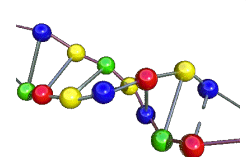
Temps de calcul

Fonction objectif

Nouveaux opérateurs

Comparaison avec d'autres méthodes

AG / DE pour l'aide à la création de modèles



Améliorations possibles



- Présélection : un enfant remplace un parent si sa performance est meilleure, sinon le parent reste
- Conservation et sélection d'individus suivant des critères de similarités entre eux
- ...

Stratégies évolutives



D'après Ingo Rechenberge (Allemagne) elles sont proches des algorithmes génétiques (J. Holland / USA).

Les différences essentielles sont :

- Les individus ne sont pas codés par des chromosomes mais par des valeurs réelles
- Moins d'aléatoire : la reproduction est déterministe
- L'opérateur de mutation agit principalement et l'opérateur de croisement est moins important : il agit simplement pour assurer une diversité suffisante



Classifieurs génétiques

Système capables d'apprendre des règles pour optimiser leurs performances dans un environnement perçu sous forme de messages

- Il est constitué de 3 parties principales :
 - Un système de règles
 - Un système d'attribution de crédit
 - Un algorithme génétique



Systemes de règles

- Une règle peut s'écrire sous la forme :
 <cond> : <resultat>
- Exemple :
 01#0 : 0111
- Sélection d'un message binaire de 4 bits commençant par 01 et se terminant par 0

Si acceptation d'un message, on renvoie le message
0111



Systemes d'attribution de credits

- Systeme de competition entre les regles qui possedent chacune un credit variable

Lorsqu'un message arrive :

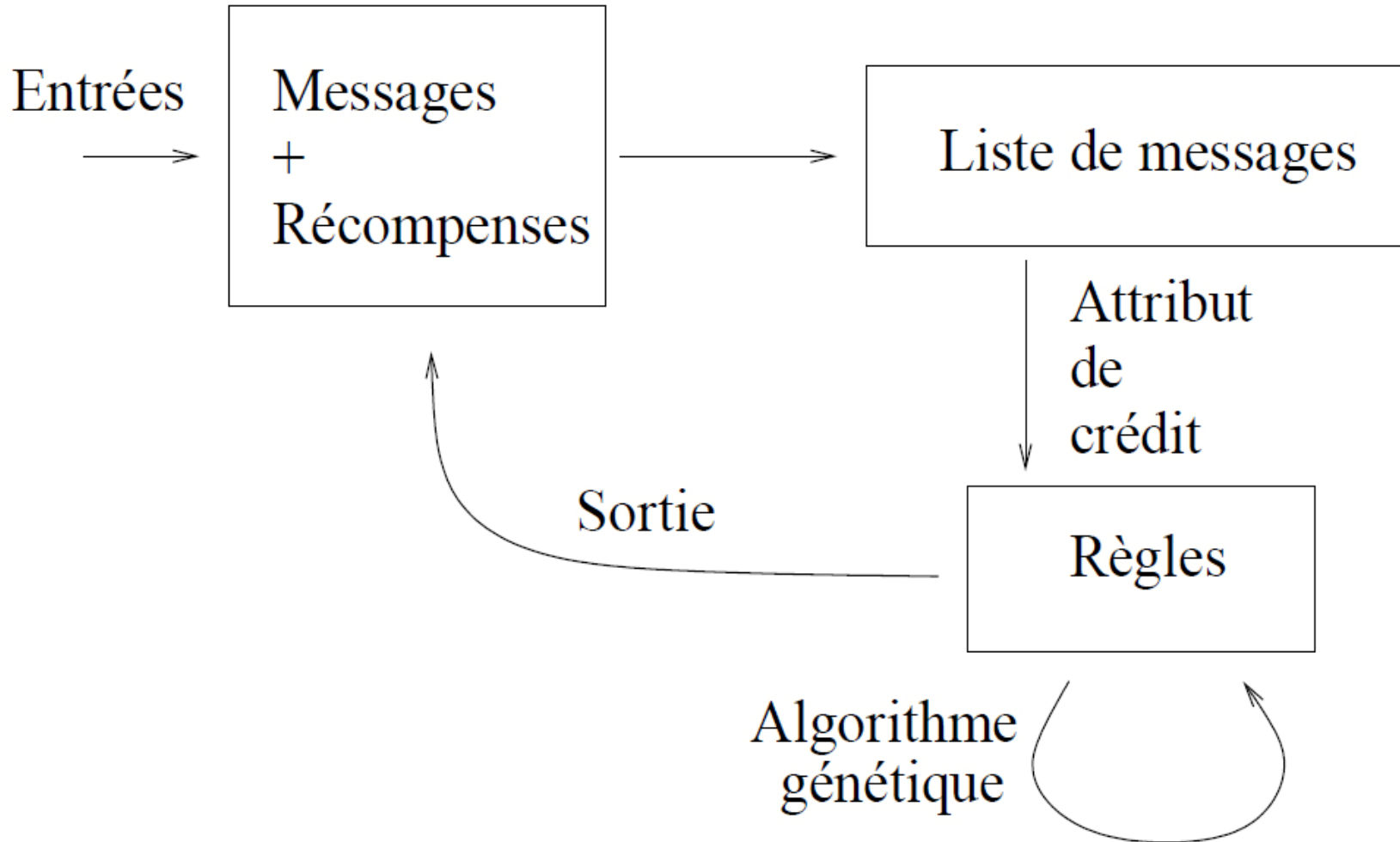
- Les regles qui peuvent le selectionner font une offre proportionnelle a leur credit actuel
- Le systeme d'attribution selectionne les regles qui offrent le plus et les active
- Une regle activee renvoie son message de resultat et offre une recompense correspondant a son credit pour les regles qui auront selectionne son message emis



Algorithmes génétiques

- Un algorithme génétique est mis en place sur la population des règles
- La fonction de performance est directement le crédit associé à chaque règle
- On sélectionne, reproduit, croise et fait muter les règles les plus performantes

Schéma d'une fonction d'un classifieur génétique





Le modèle ECHO J. Holland

- Ensemble d'agents situés dans un environnement nutritif
- Comportement des agents déduit de leur génome
- Interactions entre agents de nature variée :
 - Reproduction
 - Compétition
 - Coopération
- Performance/fitness des agents : leur capacité de survie



Caractéristiques du modèle ECHO

- Temps discret et monde représenté par une grille torique
- Déplacement possible d'un agent dans une case voisine à chaque pas de temps
- Interactions possibles entre agents que s'ils sont sur des cases voisines
- Les comportements des agents sont contrôlés par des gènes spécifiques appelés *conditions*
- Les agents ont des signes extérieurs apparents ou "tags" représentatifs de leur état social.
Ces tags sont représentés par des gènes spécifiques



Caractéristiques du modèle ECHO

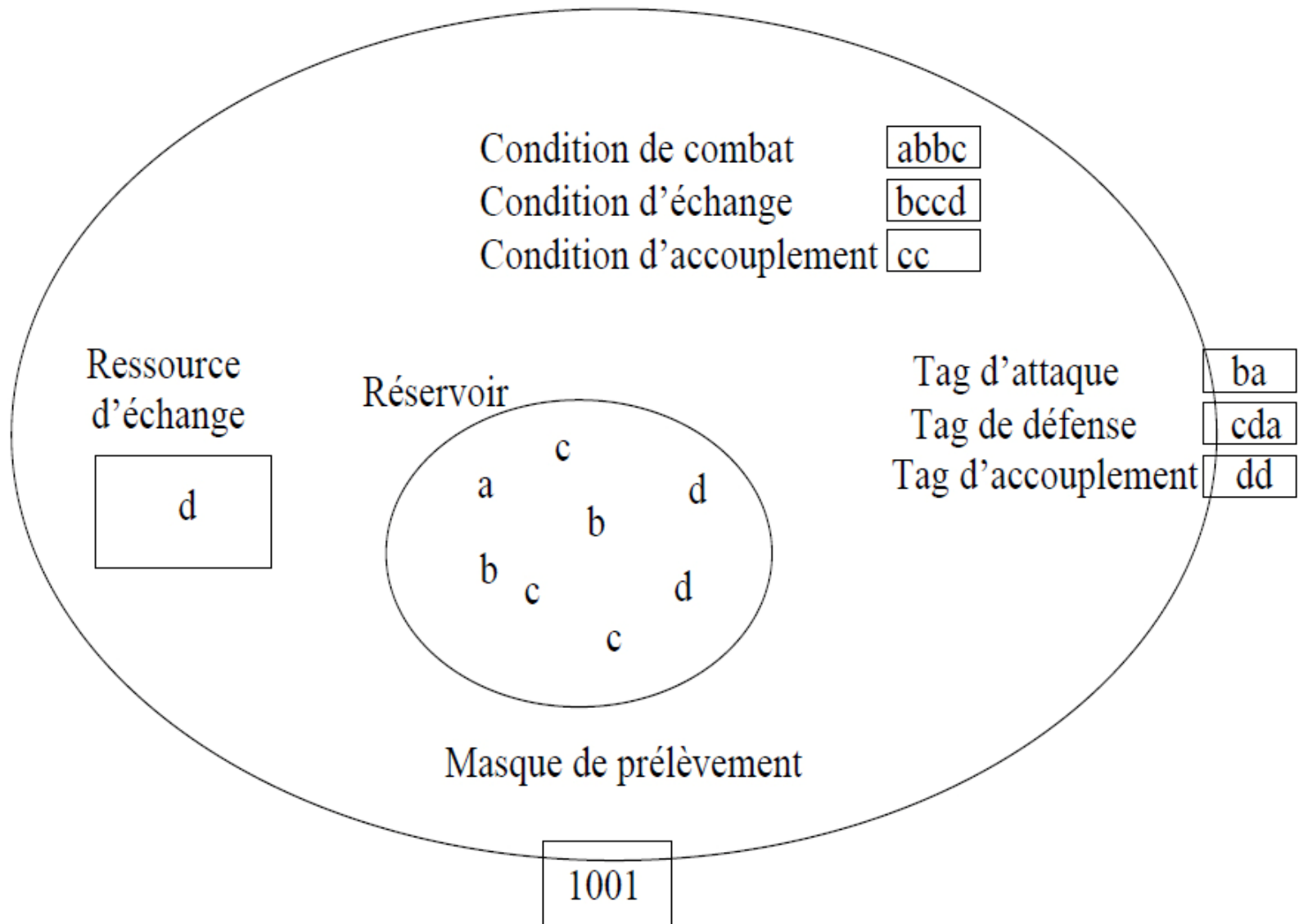
- Reproductions sexuelles ou asexuelles permettant l'évolution du génotype
- Chaque site produit une quantité de ressources spécifiques à chaque pas de temps. Il y a 4 variétés de ressources : a, b, c et d sur lesquelles la survie de chaque agent est basée
- Ce sont les “molécules” de base du monde Echo.
Elles sont les constituants du génome des agents et forment pour cela des chaînes (simple chaînage, pas de réaction chimique).
Il n'y a pas d'interaction entre les gènes d'un génome



ECHO : Les agents

- Constitués de 2 éléments :
 - Chromosome codant le comportement et l'apparence (tag) des agents : ce sont des chaînes de ressources (appelés aussi *nucléotide*)
 - Un réservoir de ressources stockées dans chaque agent

ECHO : Les agents





ECHO : Chromosome d'un agent

- Constitués de 8 gènes au total :
 - 3 gènes de tag :
 - Tag d'attaque
 - Tag de défense
 - Tag d'accouplement



ECHO : Chromosome d'un agent

- 3 gènes de conditions (comportementales et sociales)
 - Condition de combat
 - Condition de négociation
 - Condition d'accouplement
- 2 gènes de tags pour les échanges de ressources



Echanges de ressources

- Ressources d'échanges : type de ressource que peut donner l'agent à un autre au cours d'une négociation
- Masque de prélèvement : masque binaire d'un bit par type de ressources. Une ressource peut être acquise dans l'environnement que si le bit correspondant vaut 1
- D'autres possibilités d'échanges de ressources : suite à un combat ou par transmission filiale



ECHO : Auto reproduction

- Lorsque la quantité de ressources du réservoir de l'agent est suffisamment importante, il y a génération d'un clone auquel on transmet une quantité des ressources
- Le chromosome du clone est obtenu par copie et mutation



ECHO : Mutation

- Dans Echo, les mutations agissent sur les gènes, sur les tags et les conditions.

3 types de mutation sont permis :

- Suppression du dernier nucléotide du gène
- Insertion d'un nucléotide à la fin du gène
- Modification d'un nucléotide dans le gène



ECHO : Mutation

- Sur les ressources d'échanges, seule la modification est permise
- Sur le masque de prélèvement, la mutation correspond à la modification d'un bit



ECHO : Interactions entre agents

- Elles sont contrôlées par les tags et les conditions

Deux agents proches peuvent interagir :

- Par indifférence
- Par échange
- Par reproduction et génération de nouveaux agents avec croisements génétiques



ECHO : Interactions entre agents

- Les interactions sont contrôlées par un mécanisme de préfixe :
 - Elles ne peuvent se produire que si la condition est le préfixe du tag associé



Agent A	Agent B
Condition de combat : <u>ab</u>	Tag d'attaque : <u>abb</u>
Tag d'attaque : <u>bcd</u> Condition d'échange : <u>ab</u>	Condition d'échange : <u>b</u> Tag d'attaque : <u>abb</u>
Tag d'accouplement : <u>dcd</u> Condition d'accouplement : <u>a</u>	Condition d'accouplement : <u>dc</u> Tag d'accouplement : <u>adb</u>

L'agent A peut attaquer l'agent B

Les deux agents peuvent procéder à des échanges

Les deux agents peuvent s'accoupler



ECHO : Les combats

Les combats se produisent si les conditions sont réalisées et avec une certaine probabilité

- Dans le cas d'un combat, on compare 1 à 1 les nucléotides du tag d'attaque d'un agent avec le tag de défense de l'autre afin d'obtenir un score :
 - La probabilité d'être le vainqueur correspond au rapport de son score sur la somme des 2 scores
- Le vainqueur remporte toutes les ressources du vaincu, y compris celles qui constituent son génome
- Le vaincu est retiré du monde

ECHO : Accouplement

- Reproduction par croisement de deux sous-chaînes :
 - Agent A : ab || bcbd || cd ab || bd || cd
 - Agent B : b || bd || ca b || bcbd || ca
- ca
- Les longueurs des gènes des ressources d'échange et du masque de prélèvement doivent être constantes



ECHO : Cycle de base

- Interaction entre des agents sélectionnés
- Alimentation des agents par distribution des ressources de l'environnement
- Mort accidentelle des agents
- Renouvellement des ressources de l'environnement

- Les algorithmes génétiques : avantages

- Ils permettent de simuler un environnement, sa population et son évolution

- Les algorithmes génétiques : limites

- Il faut bien analyser le problème, et le paramétrage est long et complexe pour coller à la réalité




Question ?

Peut-on utiliser les AG dans le marketing et l'analyse de systèmes d'information?



Question ?



Les algorithmes génétiques et les avancées biotechnologiques respectent-elles l'éthique de la science ou de la frénésie de l'évolution technologique (par exemple pour le domaine médical, l'agriculture et OGM...) ?