

Système bio-inspiré pour la comparaison de séquences génétiques

Le problème :

Le but de la comparaison de séquences génétique est de déterminer si les séquences ont d'une part la même fonction, et d'autre part si elles ont le même ancêtre génétique. Cette comparaison se fait via un alignement, qui peut être fait sur des séquences d'ADN ou d'ARN, mais les protéines sont plus souvent utilisées.

Une protéine peut compter plusieurs milliers d'acides aminés, et l'alignement de même deux séquences nécessite l'utilisation de l'informatique. Le problème devient réellement compliqué quand il est nécessaire d'aligner une vingtaine de séquences. Il n'existe pas d'algorithme qui délivre la solution dans un temps acceptable, on peut cependant obtenir de bons résultats avec des algorithmes approximatifs tels que les méta-heuristiques.

```
-- D C E - A D F - - - C
- - - - E C A D - A E F -
```

Un alignement entre deux séquences, les lettres représentent les acides aminés, et les tirets des espaces introduits pour les aligner.

La solution envisagée :

Cette solution est en deux parties :

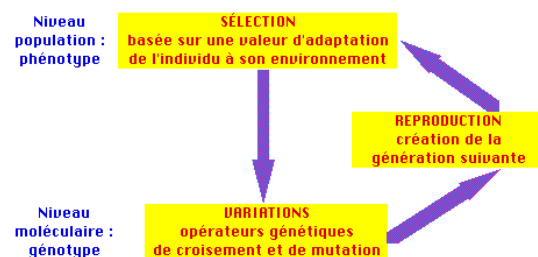
- Le développement d'un algorithme génétique pour délivrer des bons alignements de séquences
- Le passage d'une partie de cet algorithme génétique en matériel afin de l'accélérer.

Passer seulement une partie de l'algorithme en matériel permet de raccourcir le temps de développement.

L'algorithme génétique

Cet algorithme bio-inspiré effectue des étapes qui simulent le système évolutionniste : on génère un ensemble de solutions aléatoires,

appelé population, qu'on améliore avec le cycle suivant (commençant par la sélection) :

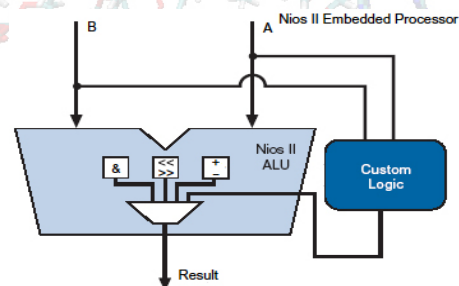


L'accélérateur matériel

La partie matérielle est écrite en VHDL, et est implémentée sur une FPGA Altera-Stratix.

La fonction développée en matériel est le calcul du score (dans l'étape de sélection), qui détermine la valeur d'une solution potentielle. C'est cette fonction qui prend la plus grande partie du temps processeur lors de l'exécution logicielle de l'algorithme.

La partie logicielle tourne sur un processeur « soft » Nios II (fourni par Altera), pour lequel il existe un compilateur C. Le calcul du score est rattaché au processeur au niveau de l'ALU, et est accessible comme une de ses fonctions (avec code assembleur, et une macro sous C).



Bien qu'il faille une FPGA plus grande pour que l'accélérateur délivre de meilleurs résultats qu'un bon processeur, les résultats obtenus valident l'approche employée.