

Directeur de thèse : D. Lavenier, Prof. ENS Cachan / Bretagne

Lieu : Equipe Symbiose Irisa/Inria Rennes

Contact : lavenier@irisa.fr – 02 99 84 72 17

<http://irisa.lavenier.net>

Contexte

Les avancées récentes et spectaculaires des biotechnologies permettent aujourd’hui de générer à bas coût d’énormes quantités de données génomiques. Les capacités de production des séquenceurs de dernière génération, par exemple, ont été multipliées par 1000 en peu de temps et obligent à repenser complètement les traitements associés. Les ordres de grandeur ne sont plus les mêmes, de nouvelles applications émergent, et les besoins en calcul explosent.

Simultanément, la stagnation des fréquences d’horloge des microprocesseurs oriente les architectures vers des systèmes multi cœurs (et probablement *manycores* dans un futur proche) pour maintenir une augmentation continue des performances. De plus, des accélérateurs exhibant un parallélisme massif à travers les cartes graphiques (GPU) ou les circuits reconfigurables (FPGA) sont désormais commercialement disponibles, offrant ainsi une puissance de calcul conséquente pour un coût raisonnable.

Objectifs de la thèse

La situation actuelle est la suivante : d’un côté les besoins de calcul en bioinformatique se font de plus en plus pressants ; de l’autre, on dispose de systèmes parallèles accessibles et très puissants. Malheureusement, la mise en œuvre des algorithmes de bioinformatiques sur ces plates formes n’est pas directe et ne conduit pas immédiatement à un gain significatif. La principale raison est que la plupart des algorithmes ont été conçus pour des machines séquentielles et n’offrent pas les propriétés requises pour une parallélisation efficace.

L’un des objectifs majeur de la thèse est donc de revisiter les principaux algorithmes du domaine et d’en proposer des versions parallèles ciblant les technologies de type GPU et/ou FPGA, sans oublier le potentiel qu’offrent intrinsèquement les architectures des processeurs actuels (instructions SIMD, multi cœurs).

A titre d’exemple, le logiciel BLAST, la référence du domaine pour comparer des séquences génomiques, a été réécrit, au sein de notre équipe, en une version parallèle (PLAST, thèse de V.H Nguyen, 2009) de manière à exprimer un parallélisme massif exploité soit par les instructions SIMD des processeurs actuels, soit par la déportation des calculs sur accélérateurs matériels (GPU et/ou FPGA). Les gains, par rapport à la version séquentielle, sont importants et motivent pour continuer dans ce sens.

Aujourd’hui, la bioinformatique, discipline encore jeune, doit impérativement se tourner vers l’algorithmique parallèle pour être en adéquation avec les ressources matérielles de demain. Les enjeux sont de taille. La thèse se positionne clairement sur ce créneau et se doit d’explorer une algorithmique bioinformatique nouvelle reposant expressément sur le parallélisme.

Bibliographie

PLAST: parallel local alignment search tool for database comparison, Nguyen VH, Lavenier D, *BMC Bioinformatics* 2009, 10:329 (12 October 2009)