

DATA



hackAtech

Shake science. Shape innovation.

#hardware

#bio-info

#génomique

#datascience

GENOPIIM

Accélérer le temps de calcul génomique

CARACTÉRISTIQUES

Les données étant toujours plus volumineuses, les centres de calculs ont un réel besoin d'optimisation pour gagner en temps de calcul. Et pour répondre aussi aux enjeux de réduction de consommation énergétique.

Ainsi GenScale parallélise des algorithmes fondamentaux de bio-informatique et de génomique en ciblant des serveurs à haute capacité mémoire. Cette technologie exploite le concept PIM (Processing-in-Memory).

La solution de PIM consiste à ajouter des unités de calcul dans la DRAM, pour minimiser le temps d'accès aux données et maximiser la bande passante, de manière à accélérer significativement les algorithmes gourmands en données.

Des applications coûteuses en temps ont été testées, le gain d'accélération en temps est non négligeable pour ces applications telles que la segmentation, le mapping, ou encore le logiciel BLAST (*Basic Local Search Alignment Tool*). Ce dernier est bien connu en biologie moléculaire pour scanner les banques de séquences protéiques ou d'ADN. Il est utilisé par de milliers de personnes chaque jour. Etant donné une séquence requête, BLAST retourne toutes les séquences de la banque où des similarités ont été détectées.

TRAITEMENT DES DONNÉES

Données génomiques (ADN, protéines), soit de gros volumes.

QUELS AVANTAGES ?

- Plus rapide : une accélération moyenne des temps de calcul de 20 en utilisant ces mémoires,
- Plus écologique : réduit fortement la consommation énergétique.



USE CASES

- Santé
- Agronomie
- Environnement

Collaboration avec l'entreprise grenobloise, UPMEM. La technologie PIM développée propose des mémoires DRAM de 16Go qui contiennent chacune 256 processeurs multithreadés. Genscale a parallélisé des algorithmes de bio-informatique sur ces composants



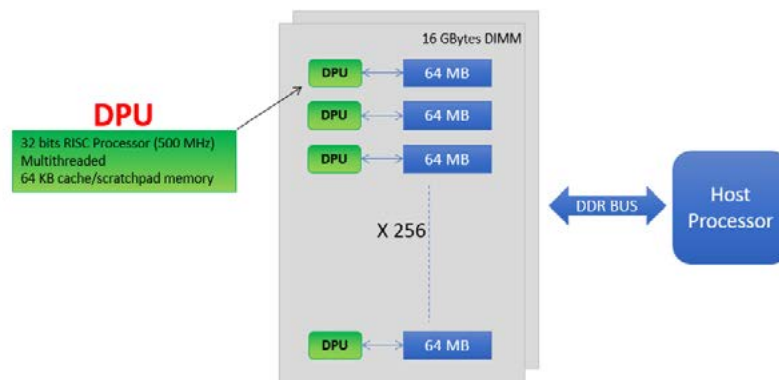
FICHE IDENTITÉ

- Langage de programmation : C
- Propriété intellectuelle : Inria (prototype commun avec Genscale et UPMEM)
- Équipe projet : GENSCALE

FONCTIONNALITÉS GÉNÉRIQUES

Cette technologie est actuellement en phase de prototype, il semblerait difficile, qu'un développement d'ingénierie soit fait lors du Hackatech. L'idée pendant cet évènement est d'évaluer et de démontrer un potentiel de la parallélisation ; par exemple une future start-up pourrait commercialiser les processeurs implémentés dont le potentiel hardware a déjà été démontré.

Les composants PIM qui sont développés sont d'abord des mémoires qui peuvent être incluses dans n'importe quel système. L'originalité de ces mémoires est qu'elles contiennent des dizaines de mini processeurs. De plus, des serveurs équipés de mémoire PIM et supportant les principaux logiciels de bioinformatique et de génomique hautement parallélisés seront mis au point. Ces serveurs seraient logés dans les centres de calcul en bio-informatique, ou les laboratoires et entreprises qui manipulent de gros volumes de données génomiques. Des serveurs déjà existants peuvent également être équipés avec PIM.



CONNAISSANCES MINIMUM REQUISES

Être spécialiste des outils de l'algorithmie parallèle.

READ ME

<https://hal.inria.fr/IRISA/hal-01327511v1>

<https://hal.archives-ouvertes.fr/hal-01294345v2>

Référents :

Dominique Lavenier
& Pierre Peterlongo

* Genscale est une équipe-projet commune à Inria, au CNRS, ENS Rennes et Université de Rennes 1.

