Contents

```
HP Charles : Merci pour vos questions ! J'ai répondu dans ce petit
document, mes réponses sont en fontes TT, les questions en fontes
en enpatement
```

User1:05 Comment on écrit du code qui va un coup sur big cœur, un coup sur petit ?

User1: L'archi des deux coeurs est similaire, c'est juste l'équilibrage de charges qui est folklorique

User1:08 Oui, mais il faut se connecter dès le début "avec micro". Je poserai la question à la fin, si c'est encore pertinent. (C'est un détail, et il y a peut-être plus intéressant comme question.) sy

Sylvain Jubertie14:08 @Dirk: Il faut faire de l'affinité si on a la main sur le governor, sur Android, c'est l'OS qui active, désactive et fait le placement tout seul.

```
HP Charles : Pour le big.LITTLE c'est l'OS qui
effectue le changement. Les codes sont les memes et peuvent
passer de l'un a l'autre sans recompilation
```

User2:19 Q: Est-il vraiment si facile d'empiler les chips de silicium, quand on sait qu'à un moment il faut évacuer la chaleur issue de chacun ?

```
HP Charles : Ca n'est pas facile : il y a plusieurs technologies
en cours. Soit le 3D stacking ou l'on construit le chip couches
par couches soit une approche "chiplet" ou l'on vient "coller"
des chips sur un réseau d'interconnection. Dans les 2 cas la
technologie est complexe et effectivement rend difficile
l'extraction de chaleur.
https://en.wikipedia.org/wiki/Three-dimensional_integrated_circuit
```

User3: - quid des ALU en pure optique (idem pour les Bus) ? (intérêt des faibles longueur d'ondes dans les lignes de transmission)

```
HP Charles: l'optique est plutôt envisagée pour la
communication. C'est ce qui est envisagé par HP pour son projet
The Machine https://www.labs.hpe.com/memory-driven-computing
```

- et la RAM statique ? Elle est plus rapide (L1, L2, L3...) et plus chère, mais énergétiquement serait-ce rentable ?

  HP Charles : la RAM statique n'est pas assez dense et consomme trop d'énergie

- est-ce que les post-optimiseurs (salto, sofan,...) ont-ils de l'avenir ? (post-optimiseurs : pour Itanium par exemple, ils permettaient de réarranger les instructions binaires pour optimiser le temps d'exécution)... des post-optimiseurs qui testent l'architecture de la machine au moment de la pré-exécution (hardware...) et terminent l'optimisation de code intermédiaire ?

  HP Charles : pour ce domaine, je travaille sur un projet de génération de code dynamique (au moment de l'appel des fonctions) pour tirer partie de la connaissance des données. On devrait avoir des publications bientôt :-)

User4 14:26 est ce que l'horizon à atteindre n'est pas la programmation matérielle via les FPGA ou l'ouverture des architectures CPU type RISC-V ? (pour les calculs spécialisés *)

HP Charles : Pour les FPGA, c'est intéressant s'il y a des acquisitions de données, sinon pour le calcul c'est trop complexe à mettre en oeuvre (outils de synthèse de circuit)

Pour le RISCV c'est une initiative intéressante, mais qui mettra du temps pour arriver a de bonnes performances. Il y a un projet européen qui utilise du RISCV.
https://www.european-processor-initiative.eu/

Il a fallut 10 ans a ARM pour arriver à faire des processeurs équivalents en performance a Intel, mais son évolution est plus rapide... https://www.anandtech.com/show/16226/apple-silicon-m1-a14-deep-dive/4 (Intel vs Apple top performance graph)

User5 14:31 Q : dans la 'guerre' AMD/ Intel, un des arguments d'AMD pour le non support de AVX512 était le coup son utilisation : optimisation => plus de chaleur => clock qui diminue, Plus généralement a trop optimiser, le risque n'est il pas d'être pénalisé sur la clock et donc paradoxalement en performance ?

HP Charles : Je suis d'accord pour le cout de la vectorisation, pas pour le pb de chaleur. Ca fonctionne bien pour ARM...

User6 14:31 Q: Ne faudrait-il pas abandonner toutes ces abstractions matérielles et revenir sur des architectures simples et faire de l'assembleur ?

HP Charles : On ne peut pas revenir à des "abstractions simples" pour le matériel, parce que la demande en performance oblige l'utilisation de notions complexes : mémoire cache, pipeline, prefetch etc. Sans ces abstractions on n'aura pas la performance / fréquence d'utilisation

User7 14:32 Q : les SoC (system on a chip) assemblent plusieurs circuits, donc une moins bonne isolation. La plupart utilisent des binaires fermés (exemple : le modem d'un téléphone). Comment assurer la sécurité des données dans ce cadre ?

HP Charles : C'est un problème ouvert et largement exploité par les pirates ! Meldown / Spectre utilisent des fuites d'informations liées a des dispositifs matériels (le cache principalement)

User8 14:34 Q. Quelle sont les perspectives du Machine learning pour l'optimisation énergétique des code ?

HP Charles : je n'y crois pas trop personnellement. Il y a des approches de ML pour essayer de trouver les meilleures optimisations, mais qui sont limitées a 1 code et 1 jeu de données, sans donner de réponse au "pourquoi ça marche"

User9 15:04 Q: Peut-être plus une question à retardement à HP Charles, mais si c'est aussi cher de "tirer" des données depuis la RAM, est-ce que ça veut dire que rapprocher la RAM des processeurs comme font les GPU et les SoC change la consommation d'énergie liée à la mémoire de façon notable ? Ou est-ce que la consommation liée à la RAM se situe au niveau des puces mémoires elle-même plus que des fils?

HP Charles : oui ça change la consommation ... pour le GPU seul,

mais comme je l'ai montré, il faut également transférer les
données de DRAM vers la mémoire GPU