



**Master Informatique
Années 2019-2020**

Veille technologique: Processeurs multi-cœurs

Combien de cœurs par processeur dans les années à venir ? Pourquoi, comment ?

Jules AUDURIER

jules.audurier@gmail.com

Master Informatique, parcours Application Interactives et Données Numériques (AIDN)

20 Décembre 2019

Table des matières

<u>Introduction</u>	3
1. <u>Analyse historique des processeurs multi-cœur</u>	5
2. <u>Présentation des processeurs muldi-die</u>	7
3. <u>Les futures technologies pour les processeurs MCM</u>	9
3.1. <u>Les Interposer</u>	9
3.2. <u>Intel EMIB</u>	10
3.3. <u>Intel Lakefield</u>	10
4. <u>Le cas des processeurs waferscale</u>	11
<u>Conclusion</u>	12
<u>Bibliographie</u>	13

Introduction

La loi de Moore, énonçant que le nombre de transistor par processeur doublerai tout les 2 ans, et l'augmentation des fréquence possible dans les micro processeurs résultante de cette loi s'approchent toutes les deux de leurs propre fin [2].

Le suivi de cette loi aura cependant permis à l'informatique de manière générale de garder un rythme d'innovation soutenue par le passé, notamment grâce à des roadmaps (ou feuilles de route) comme l'ITRS (acronyme de International Technology Roadmap For Semiconductors), créé et mise à jour par le Technology Working Group et le International Roadmap Committee [3].

L'ITRS c'est d'ailleurs avérée être un outil fondamental pour permettre à l'ensemble des acteurs du domaine informatique de coordonner et maintenir leur cadence d'innovation.

En effet, son but étant à la fois de servir comme un outil d'orientation de recherches sur les équipements et méthodes de fabrications, et de permettre aux différents acteurs du domaine de signaler des complications en cours ou à venir vis à vis de son suivi (il est bon de rappeler que ces acteurs ne sont pour la majorité soit pas dans le même segment de marché soit en concurrence directe), elle a permis à ces acteurs de profiter d'un écosystème en amélioration constante mais en plus de participer à l'évolution de cet écosystème [2][3].

Il est probable que sans la loi de Moore, et donc potentiellement l'ITRS, l'évolution observé sur les microprocesseurs depuis les 70 en terme de performance, d'efficacité et de rapport performance/prix n'aurait jamais été observé, ou du moins pas à cette échelle.

En effet, plus on améliore la technique de fabrication d'un processeur (i.e. diminuer la taille de chaque transistor ainsi que le gap entre plusieurs transistors à l'aide d'une lithographie plus avancée), moins ce dernier coûtera cher à produire et moins ce dernier aura besoin d'énergie à fréquence et nombre de transistor égale, on peut alors, sous certaine limites, généralement liés à la quantité d'énergie thermique dégagé par un transistor [2], augmenter la fréquence des processeur à consommation équivalente [3].

Cependant, il s'avère qu'augmenter les fréquences des processeurs n'est pas le seul moyen d'améliorer les performance, il s'avère même qu'augmenter les fréquences des processeurs est à la fois compliqué et problématique, car, en effet, il s'avère que l'augmentation de la fréquence d'un transistor demande une augmentation du voltage demandé par ce dernier, qui en retour cause une forte augmentation de sa production d'énergie thermique.

Ainsi, il est possible d'ajouter ou optimiser des instructions ou encore d'améliorer les temps d'accès à des instructions ou données nécessaire au fonctionnement du processeur, cependant ceci ne fait pas l'objet de ce document.

Par conséquent, un autre moyens d'augmenter les performance par processeurs est d'augmenter le nombre de cœurs dans ce processeur. Ainsi, si un problème peut être parfaitement divisé en plusieurs thread, alors il sera exécuté aussi rapidement sur N cœurs cadencés à une fréquence F qu'il ne le serait sur 1 cœur cadencé à la fréquence donné par $N \cdot F$ [2].

Multiplier le nombre de cœurs à la même fréquence augmente donc les performances et augmente la consommation de manière linéaire, là où multiplier de manière équivalente la fréquence du processeur causerait une augmentation massive de la consommation.

L'augmentation du nombre de cœurs par processeur s'avère aussi dépendante de la loi de Moore, en effet, multiplier le nombre de cœurs veut dire augmenter le nombre de transistor utilisé par ce même processeur, or, un processeur, ou du moins sont die, doit idéalement rester dans des dimensions raisonnables, ainsi, pour des raisons de rendement de die (on dira « yield » par la suite), de consommation mais aussi d'intégrabilité, la surface d'un die seule doit idéalement rester en dessous de 310 mm² [3], d'où la dépendance avec la loi de Moore.

Par ailleurs, l'ITRS s'est jusqu'à récemment basé sur la doctrine dite « More Moore »[\[3\]](#), visant à faire évoluer la miniaturisation de transistors, et par conséquent la performance et l'efficacité énergétique brute des semi-conducteurs, à l'aide de meilleures technologies de lithographie [\[3\]](#).

Cependant, des limitations arrivent, par exemple, plus de transistors sur une même surface cause une plus forte concentration thermique à voltage égale, mais encore, la limites de distance entre deux circuits avant l'arrivée de problèmes liés à la physique quantique (tel que le déplacement des particules se faisant de manière discrètes et non plus continue [cours de M. Menier]) est sur le point d'être atteinte à partir de la lithographie dite 3nm [\[2\]](#). Même sans ce genre de problèmes, il ne vas pas être possible de créer des circuits où les câblages ont une épaisseurs inférieurs à celle des atomes qui les compose.

Il s'avère que la loi de Moore est aussi freinée par des problèmes d'ordre économique. En effet, pour chaque nouvelle lithographie, la nouvelle génération de machines nécessaires à la production de processeurs gravé sur cette lithographie s'avère être un investissement de plusieurs milliards de dollars [\[2\]](#). C'est en grande partie pour cette raison qu'il ne reste plus que 4 acteurs majeurs dans la fabrication de micro processeur qui sont TSMC, Samsung, Intel et GlobalFoundries.

Il s'avère tout de même qu'il existe un moyen, déjà utilisé, pour augmenter le nombre de transistor sur une même surface et sans dépendre d'une plus fine lithographie, qui est de produire des architectures de silicone en 3 dimension, cependant, si cette solution s'avère effectivement réaliste pour de la DRAM, elle ne l'a pas été jusqu'à récemment pour des microprocesseurs, ces derniers consommant bien plus que de la DRAM, les conditions thermique associé n'ont pas rendu la chose possible [\[2\]](#), nous verrons cependant par la suite qu'Intel va dans un future proche être en capacité de produire un processeur utilisant une architecture de silicone 3D.

C'est là qu'entre en jeu une nouvelle doctrine appelée « More Than Moore », cherchant à faire intégrer de nouveaux composants, allant de différents types de capteurs analogiques à des modules régulateurs de tension (VRM) en passant par des modem 5G, directement dans le processeur [\[3\]](#). Les actuels roadmap se réfèrent déjà à cette nouvelle vision.

Il faut insister sur le fait que cette doctrine n'est pas en concurrence avec la doctrine « More Moore », il s'avère même qu'elle est en symbiose avec cette dernière, en effet, l'intégration, et donc la miniaturisation, de plus de composant jusque là externes aux processeurs pourrait permettre de fournir plus de surface à ces derniers dans les système, et donc permettre l'intégration de plus de cœurs dans ces derniers. De plus, comme on le verra par la suite, la surface totale en silicone d'un processeur n'est plus vraiment un problème.

1 - Analyse historique des processeurs multi-cœur

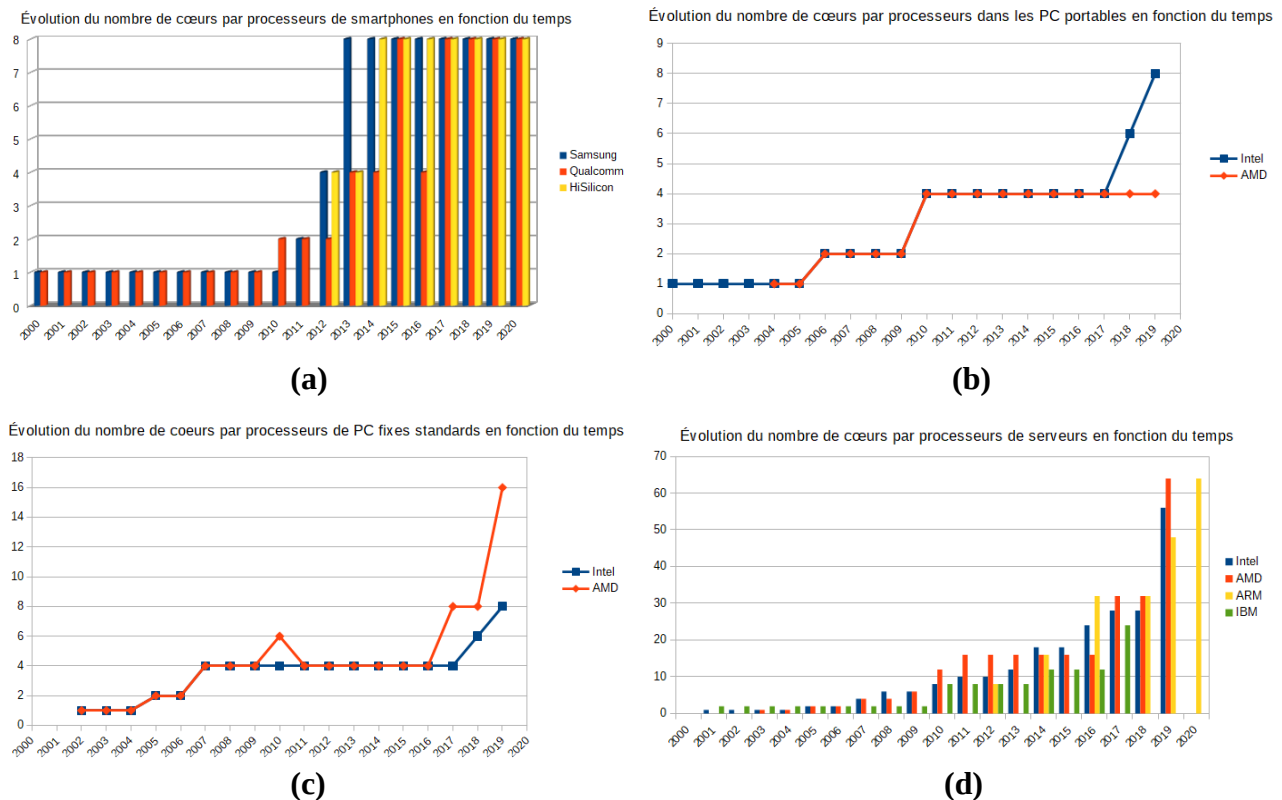


Figure 1 : Observation de l'évolution du nombre de cœurs par processeur dans les principaux domaine de l'informatique [sources]

Il est important de noter avant toute chose que pour cette partie et les suivantes, je ne tiendrai pas compte des systèmes utilisant plusieurs sockets, pour plusieurs raisons.

Premièrement, l'utilisation de système à plusieurs socket cause des problèmes de performance liés aux accès mémoire non uniforme (Non-Uniform Memory Access ou NUMA). Ce qui veut dire qu'à cause du fait que, dans un système multi-socket moderne, des parties de la mémoire du système est directement connecté aux contrôleurs mémoire de chaque processeur, si une donnée se trouve sur la mémoire connecté à un processeur qui n'est pas celui qui a besoin de la donnée, alors ce dernier doit faire une requête envers le processeurs correspondant, pour que ce dernier accède à cette donnée et la fournisse au premier, causant ainsi des temps d'accès à la mémoire bien plus (et bien trop) élevé, d'où le terme NUMA [12]. Un moyen de passer outre NUMA serait d'utiliser un contrôleur mémoire commun aux différents socket, un peu comme un North Bridge connecté aux sockets via un Front Side Bus (FSB) [7], cependant dans ce cas on aurait alors un temps de réponse accru pour tout les accès mémoires de tout les processeurs du système.

Deuxièmement, le développement de micro-architecture pouvant utiliser des configurations multi-socket s'avère être d'une grande complexité [7].

La seconde est que le titre du sujet sous entends que mon analyse doit se baser sur le nombre de cœurs par processeur et non par systèmes.

Il est aussi à préciser que la figure 1-c représente l'évolution du nombre de cœurs par processeur sur des plateforme dite « mainstream », c'est à dire utilisant des socket tel que les socket AM3, FM2 ou AM4 pour AMD et pour Intel des socket tel que LGA1150 ou LGA1151. Les plateformes dites « HEDT » (acronyme de High End DeskTop) ne sont donc pas prises en compte.

L'arrivée des processeurs multi cœurs a été marquée par l'arrivée en 2002 du premier processeur double cœurs créé par IBM nommé POWER4, comme on peut le voir sur la figure 1-d, qui aura donc eu une avance de plus de 2 ans sur Intel et AMD en se référant à la figure 1-b, 1-c et 1-d.

Les premiers processeurs à quatre cœurs seront eux arrivés en 2007 à la fois pour les serveurs (figure 1-d) et pour les ordinateurs grand public (figure 1-c).

C'est en 2010 que la barre des 12 cœurs a été atteinte par AMD (figure 1-d).

18 cœurs ont été atteints en 2014 par Intel (figure 1-d).

En 2016, un processeur ARM conçu par Huawei a atteint 32 cœurs (figure 1-d).

Enfin, cette année, en 2019, AMD a atteint 64 cœurs (figure 1-d).

Un phénomène intéressant à observer est l'évolution des processeurs de smartphone sur la figure 1-a, en effet, on peut observer que depuis 2013, ces processeurs semblent avoir atteint une limite à 8 cœurs.

Ce phénomène peut s'expliquer pour les raisons suivantes.

Tout d'abord, il ne faut pas oublier qu'un smartphone n'est pas un ordinateur ayant les mêmes objectifs de performances qu'un ordinateur fixe ou un serveur. Par exemple, il est difficile de trouver un intérêt à un smartphone possédant la puissance de calcul d'un processeur à 18 cœurs.

De plus, ces processeurs doivent respecter des dimensions, budget de consommation et enveloppe thermique extrêmement stricte, étant donné qu'un smartphone doit pouvoir tenir dans une main et avoir une durée de vie sur batterie acceptable !

Un autre phénomène observable sur les figures 1-b et 1-c est le blocage des processeurs d'ordinateurs grand public à 4 cœurs de 2007 à 2016.

Une explication possible à ce phénomène serait que, dès 2006, les parts de marché d'AMD face à Intel ont commencé à s'écrouler, reflétant un manque de d'intérêt des produits d'AMD vis-à-vis de ceux d'Intel sur cette période, ces parts restant jusqu'en 2019 à une fraction des parts d'Intel, laissant Intel avec une faible compétition [13].

Le saut soudain de la courbe sur la figure 1-c en 2017 correspond à l'arrivée des processeurs « Ryzen » d'AMD, une série de processeurs bien plus performante par rapport à la précédente architecture « Bulldozer » d'AMD.

Par ailleurs, la série de processeurs « fx » basée sur l'architecture « Bulldozer » était commercialisée comme allant jusqu'à un total de 8 cœurs, il s'avère cependant que ces processeurs utilisaient une unité de récupération et de décodage d'instructions ainsi qu'une unité de traitement des flottants en commun pour deux unités de calcul sur des entiers [lawsuit]. C'est pour cette raison que pour la figure 1-c il a été fait le choix de maintenir le nombre maximal de cœurs à 4 pour les données correspondant à AMD de 2011 à 2016.

Enfin, l'évolution du nombre de cœurs des processeurs de serveurs montre une évolution bien plus prévisible, ainsi, la courbe augmente de manière exponentielle dès le début des années 2000. Ceci pourrait s'expliquer par le fait que ces processeurs répondent à des besoins bien plus exigeants que celui des clients particuliers, en plus de se vendre à des tarifs bien supérieurs. De plus, selon les recommandations de l'ITRS, les serveurs et datacentres auront besoin dans les dix prochaines années d'assister un quadruplement du nombre de cœurs de 2017 à 2027 [10].

Un événement remarquable est aussi le pic important observé en 2019 sur la figure 1-d, il s'avère qu'elle correspond à l'arrivée des processeurs « Epyc » de 2nd génération d'AMD, utilisant une approche multi-die, identique conceptuellement à celle utilisée dans les processeurs « Epyc » de 1^{ère} génération lancés en 2017 et au processeur « Ryzen » de 3^{ème} génération lancé en 2019 et correspondant au dernier pic observé pour AMD en 2019 sur la figure 1-c. Il s'avère en effet que ces nouveaux processeurs « Epyc » de 2nd génération et « Ryzen » de 3^{ème} génération, utilisant l'architecture « Zen 2 », utilisent une approche surprenante, non seulement il ne s'agit pas de

processeur utilisant un seul die, mais en plus de cela ils séparent les cœurs et leur cache sur des dies différent de celui contenant l'ensemble de la logique d'I/O tel que les contrôleurs mémoire [15]. Intel utilise le même genre de design utilisant plus d'un die par processeur offrant ainsi un processeur à 56 cœurs en 2019 comme observé sur la figure 1-d [16].

Nous avons donc vue l'évolution du nombre de cœurs par processeurs du début des années 2000 jusqu'à 2019, nous allons voir dans les partie qui suivent l'ensemble des moyens utilisable pour pouvoir augmenter ces chiffres davantage.

2 - Présentation des processeurs muldi-die

Comme il a été dit précédemment, la loi de Moore arrive à sa propre fin et le coup de renouvellement du matériel de lithographie lié à la production de microprocesseurs est extrême. De plus, comme vu dans la partie précédente, les besoins en nombre de cœurs des serveurs et des datacentres seront exigeant dans un futur proche, et ne pourront pas, ou très difficilement, être assouvi à l'aide des designs classique des processeurs actuel [10], notamment pour des raisons de yield, de binning (sélection d'un die de qualité supérieur pour un même design) ainsi que de coup de production.

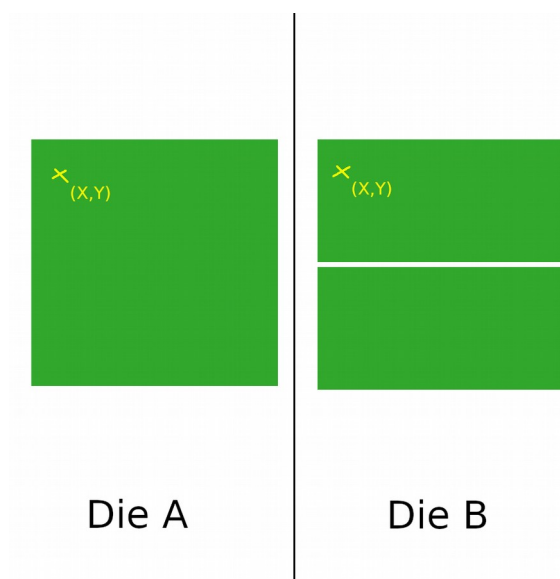


Figure 2 : Observation de la différence d'impact d'un défaut d'impression identique pour deux design de Die différents en taille

En effet, augmenter le nombre de cœurs d'un processeur veut dire augmenter son nombre de transistor, et donc, sa taille. Or, la fabrication d'un die de plus grande taille signifie qu'il coûtera plus cher à produire pour plusieurs raisons.

La première est que plus un die est gros et plus il y aura d'espace gaspillé sur le wafer (le wafer est le disque de Silicium sur lequel est imprimé un ensemble de die) sur lequel il est produit [11].

La seconde est que plus un die est gros et plus la probabilité qu'une erreur d'impression soit faite sur ce dernier est grande [11]. Prenons par exemple la situation expliquée dans la figure 2, si on prend un die A faisant la même surface que le total de 2 dies B1 et B2 côte à côte, et qu'une erreur d'impression est faite à la même coordonnées (X,Y) se trouvant à la moitié supérieure du die A et sur le die supérieur B1, alors dans le cas du die A 100 % du die devient soit inutilisable soit de moins bonne qualité, alors que pour le design B seul B1 est touché et non B2.

Un excellent outil pour observer ce phénomène est le Die Yield Calculator de Caly Technologies [6]. Ainsi, si l'on compare les résultats observés sur la figure 3-a à ceux observés sur la figure 3-b, on s'aperçoit que, pour des caractéristiques (externes aux dimensions des dies) identiques, le design de die de la figure 3-a, ayant la même largeur mais le double de hauteur comparé au design de die de la figure 3-b, produit des rendements significativement moindres à ceux observés dans la figure 3-b. On peut donc observer que le design de la figure 3-a permet la création de 215 dies « parfaits » de dimension $L \times H$ tandis que le design de la figure 3-b en permet la création de 519 de dimension $L \times H/2$. Pour $H=10\text{mm}$ et $L=10\text{mm}$, cela se traduit donc par une surface totale de dies « parfaits » de 21500 mm^2 pour le design de la figure 3-a contre 25950 mm^2 pour la figure 3-b.

Il est tout de fois bon de signaler que l'élaboration d'un design modulaire de dies (où certaines parties du die peuvent être désactivées, comme par exemple des cœurs ou une partie du cache) permet tout de même de pouvoir utiliser des dies ayant subi ces défauts d'impression [10]. Donc, si l'on prend cela en compte et que l'on considère, de manière idéaliste, que chacun des dies touchés par un défaut d'impression est touché au niveau d'une de ces parties désactivables dans la figure 3, alors un total de 31600 mm^2 de die utilisable serait obtenu pour les designs des figures 3-a et 3-b.

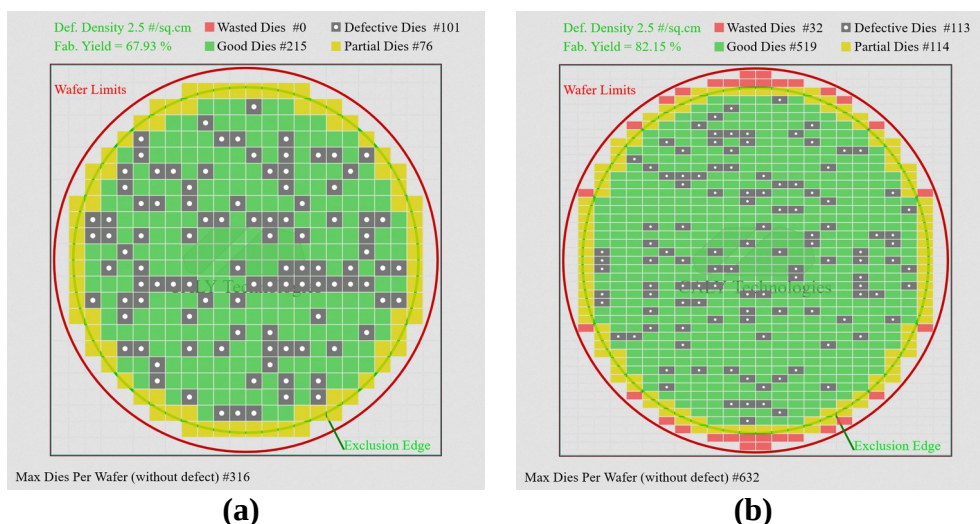


Figure 3 : Observation des différences de yields entre 2 design de taille différente sur une lithographie identique, ces images sont obtenues à l'aide de « Die Yield Calculator » [6]

Un plus petit design de die peut donc produire à la fois une quantité similaire de volume de die utilisable, plus de die de meilleure qualité et gaspille moins d'espace sur un même wafer. Or, si l'on produit des processeurs de manière classique, il faut de plus gros dies pour produire des processeurs avec plus de cœurs.

Cependant, il existe différents moyens d'utiliser plusieurs dies pour la conception d'un même processeur.

Des dies de DRAM sont notamment déjà empilés à l'aide de TSV (acronyme de Through Silicon Via). Cette technique, appelée die stacking, permet notamment de connecter entre eux des dies fabriqués à l'aide de différentes lithographies [11], cependant, cette technique n'a jusqu'à présent pas été utilisée pour des processeurs classiques principalement pour des raisons de dissipations thermiques.

Ainsi, il est par exemple possible de créer un processeur possédant 64 cœurs où ces 64 cœurs sont réparties sur 4 dies contenant 16 cœurs chacun [11]. Il s'avère même que la comparaison de rendement entre un design de processeur composé de dies dit monolithique contenant 64 cœurs face

à un design utilisant 4 die contenant chacun 16 cœurs démontre la supériorité en terme de rendement de l'approche multi-die, ainsi, un rendement de 84,5 % est observé pour le design monolithique contre 95,7 % pour le design multi-chip [11].

Cela permet donc de faire des économies importantes grâce à de meilleurs rendement, moins de gaspillage de surface pour chaque wafer, mais aussi des économies sur la conception et la fabrication des masques d'impression des die si l'on utilise plusieurs fois un même die sur chaque processeur [10]. Ces gain économique supposent tout de fois l'utilisation d'un test Known Good Dies (KGD) pour trier les dies non fonctionnels avant l'assemblage de ces processeurs [10].

Il s'avère de plus que cette approche n'a pas pour seule intérêt de permettre la création de design plus complexe, elle permet aussi des gains de performances. En effet, il est possible de faire du binning sur les différents die composant le design final d'un processeur [11] dans le but de produire des processeur possédant des die de qualité suffisantes pour que chacun des cœurs les composant puissent atteindre une fréquence dite « all core » qui seraient statistiquement bien plus rare avec un design monolithique.

La partie suivante va chercher à étudier les différentes technologie utilisable, et utilisé, pour la fabrication de modules à multiple puces, ou MCM (acronyme de Multi-Chip Module).

3 – Les futures technologies pour les processeurs MCM

3.1 – Les Interposer

Dans notre cas, les interposer sont en Silicone et servent de connecteurs entre plusieurs die d'un processeur ainsi que vers le reste du système. Ces derniers se composent de TSV et de circuits gravé [10], généralement dans une lithographie ayant atteint une certaine maturité et donc un fort rendement, comme la lithographie 65nm par exemple, le but étant de créer des liaison entre différents point et non de créer des architectures de calcul sauf pour un cas particulier qui sera traité dans cette partie.

Les interposer sont un excellent moyens, et techniquement l'un des seuls, d'atteindre un grand nombre d'interconnexions [4], au-delà de 10^5 , par processeurs. Un intérêt évident pour les interposer est pour les interconnexion entre dies dans un design MCM [10].

Cependant, il s'avère que, sans optimisation, un design MCM engendre des latences supplémentaire entre cœurs ainsi qu'entre les cœurs et la mémoire, et donc des pertes de performance, comparé à un design monolithique [11].

Cependant des optimisations peuvent être apporté à l'architecture interne d'un interposer pour minimiser les latences et maximiser les bandes passantes, le but étant d'éviter le plus possible que les communications passent trop exclusivement par un même segment de l'interposer tout en évitant de multiplier le nombre de lignes dans l'interposer pour à la fois éviter d'augmenter sa consommation et de diminuer sa fréquence [11].

Ainsi, il existe des modèles d'architecture d'interconnexions (ou plus simplement topologie) utilisable dans la conception d'un interposer qui offrent des performances supérieur à des architecture communément trouvés dans les processeur actuels tel que le « ring bus » ou le « mesh bus » et ce avec un nombre équivalent de nœud et parfois moins d'interconnexions [11]. Des

exemples de ces topologies sont le « Double Butterfly », le « Folded Torus » ou encore le « Butterdonut » qui est une combinaison des deux premiers [11]. L'utilisation dans l'interposer d'une topologie dite « non aligné », telle que certains nœuds connectant l'interposer aux dies se trouvent entre les dies, permet d'améliorer encore plus les performances du réseau [11].

Cet ensemble d'optimisation dans l'interposer permet de minimiser drastiquement les latences lié au design MCM, à tel point que les latences mesurées peuvent en fait être inférieures, pour un processeur à 64 cœurs utilisant un design MCM et un interposer optimisé, face à un processeur monolithique équivalent [11].

Le type d'interposer qui a été traité ici est le type passif, il existe donc naturellement un autre type d'interposer dit actif, qui désigne un interposer intégrant des transistors et ayant donc un rôle actif. L'intérêt d'un interposer actif face à un équivalent passif est qu'il peut permettre de fiabiliser ou d'améliorer des interconnexions, il est aussi possible de lui faire jouer un rôle de routeur interne au processeur, permettant ainsi de ne plus avoir besoin d'intégrer cette logique aux dies, voir même d'y intégrer du cache [10]. Ceci est dû au fait qu'un interposer actif peut être créé à l'aide de technologies de lithographie classique et peut donc incorporer des transistors CMOS classiques. Cependant, un défaut des interposers actifs face à leur équivalent passif est leur désavantage en terme de yield, en effet là où un interposer passif n'est essentiellement que des TSV et des lignes, ils sont donc constitués en majorité d'espace « vide », un défaut d'impression n'aura donc que rarement d'impact sur ces derniers (l'ajout de liaisons redondantes diminue ce problème encore plus), tandis qu'un interposer actif intégrera des zones avec des transistors, ce qui aura un impact plus ou moins grand sur les yields de ces derniers en fonction de la taille occupée par ces zones [10]. Pour cette raison il est préférable d'utiliser le moins de logique possible sur un transistor actif pour ne pas faire baisser ses yields et donc éviter de faire monter son coût, car il ne faut pas oublier qu'un interposer doit englober l'entièreté des dies du processeur, et, comme on l'a vu précédemment, un plus gros design de die veut dire un moins grand volume de die « parfait ».

3.2 - Intel EMIB

Acronyme de Embedded Multi-die Interconnect Bridge, il s'agit d'une technologie de packaging utilisant un ensemble de liaisons en silicone (similaire aux interposers vu précédemment) pour interconnecter des dies, ces liaisons en silicone se trouvent directement intégrées dans le package et entre les dies à interconnecter [9]. Un intérêt d'EMIB face à l'utilisation d'un interposer englobant l'ensemble des dies d'un processeur est qu'il n'a justement pas besoin d'un grand interposer pour faire l'ensemble des interconnexions mais il peut en utiliser plusieurs, ce qui, comme on l'a vu précédemment, peut permettre d'augmenter les yields de manière significative.

3.3 – Intel LakeField

LakeField est une architecture de processeur utilisant plusieurs concepts intéressants.

Tout d'abord, cette architecture utilise du 3D stacking, ainsi, un premier die intégrant de la logique n'ayant pas besoin de beaucoup de puissance ainsi que des éléments nécessaires à l'alimentation du processeur est connecté à l'aide de TSVs à un die placé au dessus. Ce dernier contient les éléments plus actifs du processeur, comme les cœurs et l'unité de calcul graphique intégrée (IGP). Enfin, de la DRAM est ensuite empilée par dessus cette dernière couche [5].

De plus, il est intéressant de voir que cette architecture utilise une configurations non homogène de cœurs, ainsi, de manière similaire à la majorité des processeurs de smartphone, un cœurs plus large et plus puissant, mais consommant plus, est couplé à 4 autres cœurs à plus faible consommation [5].

4 – Le cas des processeurs waferscale

Il existe plusieurs type de processeurs waferscale, l'un utilise directement un design monolithique de die utilisant l'entièreté d'un wafer, allant donc à l'exact opposé de ce qui a été présenté dans ce document, l'autre utilise un design utilisant un ensemble de plus petits die interconnecté sur un wafer [1]. Pour des raisons de yields qui commence à être évidentes, le premier design n'a jamais vraiment aboutit. Le second design tel qu'il est présenté ici utilise des interconnexion appelé Silicon-Interconnection Fabric (ou Si-IF). Il s'agit encore là d'une technologie permettant d'interconnecté plusieurs die entre eux ainsi qu'avec le reste du système, et cette technologie s'avère très similaire aux interposer vu précédemment.

Un waferscale n'a pas les même restriction que EMIB ou un interposer, EMIB ne pouvant aller au-delà d'une dizaine de puce par package et un interposer ne pouvant dépasser 1230mm² de surface [1].

De plus, un waferscale aura besoin de moins de surface totale à nombre de die égale avec des architecture utilisant d'autres technologie, en plus d'utiliser des communications inter die plus performante et moins énergivore [1].

Il s'avère que des optimisation de topologie d'un waferscale utilisant Si-IF similaire à celles faites pour un interposer peuvent être faite , cependant si une topologie comme le Torus peuvent être utilisé, toute les topologies ne peuvent l'être à cause de la taille et de la complexité de ce genre de designs [1].

La technologie Si-IF permet de plus d'intégrer des éléments d'habitude externe à un processeur telle que de la DRAM, des VRM ou même des condensateurs et des inducteurs, permettant ainsi de diminuer l'espace occupé dans un système par des composants externes au processeur.

Enfin, les yields d'un processeur waferscale utilisant Si-IF vont dépendre de trois choses : le yield des die intégrés au waferscale, le yield des TSV qui sont ici des micro pillier en cuivre, ainsi que le yield du substrat Si-IF. A noter que les yield des points de connexions de Si-IF sont proche de 100 % tandis que le yield du substrat est au dessus de 90 %, celui-ci étant composé de lignes passive conceptuellement similaire à un interposer passif.

Conclusion

Ce document a fait l'étude de l'évolution historique du nombre de cœurs dans les processeurs sur différent domaine majeur de l'informatique. Il a également fait l'étude des techniques et technologies qui permettront l'évolution du nombre de cœurs dans le futur, tout en ayant au préalable fait l'explication de l'intérêt de ces nouveautés.

Ainsi, au vu des problèmes se posant sur le chemin du progrès des processeurs multi-cœurs, et d'après les différents éléments de solution observés, il apparaît que le futur des processeurs multi-cœurs est représenté par une approche multi-die.

Pour conclure ce document, il ne semble pas aberrant, au vu des différentes observations faites, de présager des processeurs contenant une centaine, voire plusieurs centaines, de cœurs dans un futur proche.

Autocritique du document

Si je pense que la figure 1 possède un grand intérêt en terme de synthétisation de l'information qu'elle présente, je ne peux garantir une exactitude parfaite de l'entière des informations qu'elle présente, surtout pour les parties plus anciennes de ces graphes. Il est en effet possible que certaines sources utilisées n'aient pas pris en compte certaines caractéristiques intrinsèques à certains designs de processeurs qui auraient eu une influence sur le nombre de cœurs réel de ces derniers. Je pense notamment au design de certains processeurs utilisant l'architecture « Bulldozer », présentés comme ayant 8 cœurs mais, selon le fruit de recherche approfondis, se rapprochant plus de 4 cœurs réels.

Bibliographie

- (1) [Saptadeep Pal, Daniel Petrisko, Matthew Tomei, Puneet Gupta, Subramanian S. Iyer, and Rakesh Kumar. "Architecting Waferscale Processors - A GPU Case Study."](#)
- (2) [Waldrop, M. The chips are down for Moore's law. *Nature* **530**, 144–147 \(2016\). doi:10.1038/530144a](#)
- (3) [Wolfgang Arden, Michel Brillouët, Patrick Cogez, Mart Graef, Bert Huizing, and Reinhard Mahnkopf. "More-than-Moore."](#)
- (4) [R. R. Tummala \(2019\). Moore's Law for Packaging to Replace Moore's Law for ICS. In *2019 Pan Pacific Microelectronics Symposium \(Pan Pacific\)* \(pp. 1-6\).](#)
- (5) [S. Khushu, & W. Gomes \(2019\). Lakefield: Hybrid cores in 3D Package. In *2019 IEEE Hot Chips 31 Symposium \(HCS\)* \(pp. 1-20\).](#)
- (6) Die Per Wafer Calculator. <https://caly-technologies.com/die-yield-calculator/>.
- (7) [R. Zhao, "Intel enterprise server processor packaging challenge and future trend," 2014 15th International Conference on Electronic Packaging Technology, Chengdu, 2014, pp. 1516-1521. doi: 10.1109/ICEPT.2014.6922942](#)
- (8) Server Training Basics and Hardware Configuration, slide n° 5. <https://www.slideshare.net/itassistantdahanu/server-training>.
- (9) [R. Mahajan, R. Sankman, N. Patel, D. Kim, K. Aygun, Z. Qian, Y. Mekonnen, I. Salama, S. Sharan, D. Iyengar, & D. Mallik \(2016\). Embedded Multi-die Interconnect Bridge \(EMIB\) – A High Density, High Bandwidth Packaging Interconnect. In *2016 IEEE 66th Electr*](#)
- (10) [D. Stow, Y. Xie, T. Siddiqua, & G. H. Loh \(2017\). Cost-effective design of scalable high-performance systems using active and passive interposers. In *2017 IEEE/ACM International Conference on Computer-Aided Design \(ICCAD\)* \(pp. 728-735\).](#)
- (11) [Ajaykumar Kannan, Natalie Enright Jerger, and Gabriel H. Loh. "Enabling Interposer-based Disintegration of Multi-core Processors." \(2015\).](#)
- (12) [Sergey Blagodurov, Sergey Zhuravlev, Mohammad Dashti, and Alexandra Fedorova. "A Case for NUMA-aware Contention Management on Multicore Systems \(p.11\)." \(2011\).](#)
- (13) AMD vs Intel Market Share. https://www.cpubenchmark.net/market_share.html.
- (14) [R. C. Frye \(1994\). Physical scaling and interconnection delays in multichip modules *IEEE Transactions on Components, Packaging, and Manufacturing Technology: Part B*, **17**\(1\), 30-37.](#)

- (15) Dr. Ian Cutress. AMD Zen 2 Microarchitecture Analysis: Ryzen 3000 and EPYC Rome. <https://www.anandtech.com/show/14525/amd-zen-2-microarchitecture-analysis-ryzen-3000-and-epyc-rome/>.
- (16) Dr. Ian Cutress. Hands on with the 56-core Xeon Platinum 9200 CPU: Intel's Biggest CPU Package Ever. <https://www.anandtech.com/show/14182/hands-on-with-the-56core-xeon-platinum-9200-cpu-intels-biggest-cpu-package-ever>.

Les sources suivantes sont les sources correspondants au données affichées dans la figure 1 :

- **figure 1-a**
 - Qualcomm
 - <https://www.anandtech.com/show/4565/qualcomms-updated-brand-introducing-snapdragon-s1-s2-s3-s4-processors>
 - http://phonedb.net/index.php?m=processor&id=82&c=qualcomm_msm7200
 - http://phonedb.net/index.php?m=processor&id=172&c=qualcomm_msm7200a
 - http://phonedb.net/index.php?m=processor&id=226&c=qualcomm_msm7627
 - http://phonedb.net/index.php?m=processor&id=278&c=qualcomm_snapdragon_s3_msm8260
 - http://phonedb.net/index.php?m=processor&id=299&c=qualcomm_snapdragon_s4_msm8960
 - http://phonedb.net/index.php?m=processor&id=609&c=qualcomm_snapdragon_810_msm8994_v2.1
 - http://phonedb.net/index.php?m=processor&id=686&c=qualcomm_snapdragon_821_msm8996ac_pro
 - http://phonedb.net/index.php?m=processor&id=702&c=qualcomm_snapdragon_835_msm8998
 - http://phonedb.net/index.php?m=processor&id=787&c=qualcomm_snapdragon_855plus_sm8150-ac_hana
 - <https://www.anandtech.com/show/15178/qualcomm-announces-snapdragon-865-and-765-5g-for-all-in-2020-all-the-details>
 - HiSilicon
 - http://phonedb.net/index.php?m=processor&id=326&c=hisilicon_k3v2_hi3620
 - http://phonedb.net/index.php?m=processor&id=496&c=hisilicon_k3v2e
 - http://phonedb.net/index.php?m=processor&id=541&c=hisilicon_kirin925_hi3830
 - http://phonedb.net/index.php?m=processor&id=690&c=hisilicon_honor_kirin960_hi3660
 - http://phonedb.net/index.php?m=processor&id=727&c=hisilicon_honor_kirin970
 - http://phonedb.net/index.php?m=processor&id=768&c=hisilicon_honor_kirin980
 - http://phonedb.net/index.php?m=processor&id=799&c=hisilicon_honor_kirin990_5g
 - Samsung
 - <http://phonedb.net/download/s3c2410.pdf>
 - <http://www.bdtic.com/datasheet/SAMSUNG/S3C2412.pdf>
 - http://phonedb.net/index.php?m=processor&id=162&c=samsung_s5l8900
 - http://phonedb.net/index.php?m=processor&id=219&c=samsung_s5pc100

- http://phonedb.net/index.php?m=processor&id=269&c=samsung_s5pc210_exynos_4_dual_4210_orion
- http://phonedb.net/index.php?m=processor&id=337&c=samsung_exynos_4_quad_4412
- http://phonedb.net/index.php?m=processor&id=420&c=samsung_exynos_5_octa_5420
- http://phonedb.net/index.php?m=processor&id=579&c=samsung_exynos_7_octa_7420
- http://phonedb.net/index.php?m=processor&id=791&c=samsung_exynos_9_octa_9825
- <https://www.anandtech.com/show/15021/samsung-announces-exynos-990-7nm-euv-m5-g77-lpddr5-and-5g-modem-flagship-soc>
- figure 1-b
 - Intel
 - <https://ark.intel.com/content/www/us/en/ark/products/27412/mobile-intel-pentium-iii-processor-850-mhz-256k-cache-100-mhz-fsb.html>
 - <https://ark.intel.com/content/www/us/en/ark/products/27338/mobile-intel-celeron-processor-lv-650-mhz-256k-cache-100-mhz-fsb.html>
 - <https://ark.intel.com/content/www/us/en/ark/products/27359/mobile-intel-pentium-4-processor-m-2-20-ghz-512k-cache-400-mhz-fsb.html>
 - <https://ark.intel.com/content/www/us/en/ark/products/29761/intel-core-2-duo-processor-t7500-4m-cache-2-20-ghz-800-mhz-fsb.html>
 - <https://ark.intel.com/content/www/us/en/ark/products/37033/intel-core-2-quad-processor-q9100-12m-cache-2-26-ghz-1066-mhz-fsb.html>
 - <https://ark.intel.com/content/www/us/en/ark/products/53470/intel-core-i7-2675qm-processor-6m-cache-up-to-3-10-ghz.html>
 - <https://ark.intel.com/content/www/us/en/ark/products/97185/intel-core-i7-7700hq-processor-6m-cache-up-to-3-80-ghz.html>
 - <https://ark.intel.com/content/www/us/en/ark/products/134906/intel-core-i7-8750h-processor-9m-cache-up-to-4-10-ghz.html>
 - <https://ark.intel.com/content/www/us/en/ark/products/192990/intel-core-i9-9980hk-processor-16m-cache-up-to-5-00-ghz.html>
 - AMD
 - <http://www.cpu-world.com/CPUs/K8/AMD-Mobile Athlon 64 3000%2B - AMN3000BIX5AP.html>
 - <http://www.cpu-world.com/CPUs/K8/AMD-Turion 64 X2 Mobile technology TL-60 - TMDTL60HAX5DM.html>
 - <http://www.cpu-world.com/CPUs/K8/AMD-Turion X2 Ultra ZM-88 - TMZM88DAM23GG.html>
 - <http://www.cpu-world.com/CPUs/K10/AMD-Phenom II Quad-Core Mobile N950 - HMN950DCR42GM.html>
 - <http://www.cpu-world.com/CPUs/K10/AMD-A8-Series A8-3500M.html>
 - <http://www.cpu-world.com/CPUs/Jaguar/AMD-A6-Series A6-5200.html>
 - <http://www.cpu-world.com/CPUs/Zen/AMD-Ryzen 7 Mobile 2700U.html>
 - <http://www.cpu-world.com/CPUs/Zen/AMD-Ryzen 7 PRO Mobile 3700U.html>

- figure 1-c

- Intel

- <https://ark.intel.com/content/www/us/en/ark/products/27439/intel-pentium-4-processor-2-50-ghz-512k-cache-400-mhz-fsb.html>
- <https://ark.intel.com/content/www/us/en/ark/products/27460/intel-pentium-4-processor-520j-supporting-ht-technology-1m-cache-2-80-ghz-800-mhz-fsb.html>
- <https://ark.intel.com/content/www/us/en/ark/products/27514/intel-pentium-d-processor-840-2m-cache-3-20-ghz-800-mhz-fsb.html>
- <https://ark.intel.com/content/www/us/en/ark/products/30790/intel-core-2-quad-processor-q6700-8m-cache-2-66-ghz-1066-mhz-fsb.html>
- <https://ark.intel.com/content/www/us/en/ark/products/52214/intel-core-i7-2600k-processor-8m-cache-up-to-3-80-ghz.html>
- <https://ark.intel.com/content/www/us/en/ark/products/97129/intel-core-i7-7700k-processor-8m-cache-up-to-4-50-ghz.html>
- <https://ark.intel.com/content/www/us/en/ark/products/126684/intel-core-i7-8700k-processor-12m-cache-up-to-4-70-ghz.html>
- <https://ark.intel.com/content/www/us/en/ark/products/186605/intel-core-i9-9900k-processor-16m-cache-up-to-5-00-ghz.html>
- <https://ark.intel.com/content/www/us/en/ark/products/192943/intel-core-i9-9900ks-processor-16m-cache-up-to-5-00-ghz.html>

- AMD

- [http://www.cpu-world.com/CPUs/K7/AMD-Athlon XP 2800+ - AXDA2800DKV4D.html](http://www.cpu-world.com/CPUs/K7/AMD-Athlon%20XP%202800+-AXDA2800DKV4D.html)
- [http://www.cpu-world.com/CPUs/K8/AMD-Athlon 64 4000%2B - ADA4000DAA5BN \(ADA4000BNBOX\).html](http://www.cpu-world.com/CPUs/K8/AMD-Athlon%2064%204000%2B-ADA4000DAA5BN(ADA4000BNBOX).html)
- [http://www.cpu-world.com/CPUs/K8/AMD-Athlon 64 X2 4800%2B - ADA4800DAA6CD \(ADA4800CDBOX\).html](http://www.cpu-world.com/CPUs/K8/AMD-Athlon%2064%20X2%204800%2B-ADA4800DAA6CD(ADA4800CDBOX).html)
- [http://www.cpu-world.com/CPUs/K10/AMD-Phenom X4 9600 - HD9600WCJ4BGD \(HD9600WCGDBOX\).html](http://www.cpu-world.com/CPUs/K10/AMD-Phenom%20X4%209600-HD9600WCJ4BGD(HD9600WCGDBOX).html)
- [http://www.cpu-world.com/CPUs/K10/AMD-Phenom II X6 1035T - HDT35TWFK6DGR.html](http://www.cpu-world.com/CPUs/K10/AMD-Phenom%20II%20X6%201035T-HDT35TWFK6DGR.html)
- https://regmedia.co.uk/2015/11/06/amd_bulldozer_lawsuit.pdf
- [http://www.cpu-world.com/CPUs/Bulldozer/AMD-FX-Series FX-8150.html](http://www.cpu-world.com/CPUs/Bulldozer/AMD-FX-Series%20FX-8150.html)
- [http://www.cpu-world.com/CPUs/Bulldozer/AMD-FX-Series FX-8370.html](http://www.cpu-world.com/CPUs/Bulldozer/AMD-FX-Series%20FX-8370.html)
- [http://www.cpu-world.com/CPUs/Zen/AMD-Ryzen 7 1800X.html](http://www.cpu-world.com/CPUs/Zen/AMD-Ryzen%207%201800X.html)
- [http://www.cpu-world.com/CPUs/Zen/AMD-Ryzen 7 2700X.html](http://www.cpu-world.com/CPUs/Zen/AMD-Ryzen%207%202700X.html)
- [http://www.cpu-world.com/CPUs/Zen/AMD-Ryzen 9 3950X.html](http://www.cpu-world.com/CPUs/Zen/AMD-Ryzen%209%203950X.html)

- figure 1-d

- Intel

- <https://ark.intel.com/content/www/us/en/ark/products/27264/intel-xeon-processor-2-00-ghz-256k-cache-400-mhz-fsb.html>
- <https://ark.intel.com/content/www/us/en/ark/products/27280/intel-xeon-processor-3-20-ghz-2m-cache-533-mhz-fsb.html>
- <https://ark.intel.com/content/www/us/en/ark/products/27202/intel-xeon-processor-2-80-ghz-4m-cache-800-mhz-fsb.html>

- <https://ark.intel.com/content/www/us/en/ark/products/27287/intel-xeon-processor-7140m-16m-cache-3-40-ghz-800-mhz-fsb.html>
- <https://ark.intel.com/content/www/us/en/ark/products/30797/intel-xeon-processor-x3230-8m-cache-2-66-ghz-1066-mhz-fsb.html>
- <https://ark.intel.com/content/www/us/en/ark/products/36947/intel-xeon-processor-x7460-16m-cache-2-66-ghz-1066-mhz-fsb.html>
- <https://ark.intel.com/content/www/us/en/ark/products/46499/intel-xeon-processor-x7560-24m-cache-2-26-ghz-6-40-gt-s-intel-qpi.html>
- <https://ark.intel.com/content/www/us/en/ark/products/53580/intel-xeon-processor-e7-8870-30m-cache-2-40-ghz-6-40-gt-s-intel-qpi.html>
- <https://ark.intel.com/content/www/us/en/ark/products/75283/intel-xeon-processor-e5-2697-v2-30m-cache-2-70-ghz.html>
- <https://ark.intel.com/content/www/us/en/ark/products/81061/intel-xeon-processor-e5-2699-v3-45m-cache-2-30-ghz.html>
- <https://ark.intel.com/content/www/us/en/ark/products/93790/intel-xeon-processor-e7-8890-v4-60m-cache-2-20-ghz.html>
- <https://ark.intel.com/content/www/us/en/ark/products/120498/intel-xeon-platinum-8180m-processor-38-5m-cache-2-50-ghz.html>
- <https://ark.intel.com/content/www/us/en/ark/products/194146/intel-xeon-platinum-9282-processor-77m-cache-2-60-ghz.html>
- AMD
 - <http://www.cpu-world.com/CPUs/K8/AMD-Opteron 848 - OSA848CEP5AM.html>
 - [http://www.cpu-world.com/CPUs/K8/AMD-Dual-Core Opteron 880 - OSA880FAA6CC \(OSA880CCWOF\).html](http://www.cpu-world.com/CPUs/K8/AMD-Dual-Core Opteron 880 - OSA880FAA6CC (OSA880CCWOF).html)
 - [http://www.cpu-world.com/CPUs/K8/AMD-Second Generation Opteron 8218 - OSA8218GAA6CR \(OSA8218CRWOF\).html](http://www.cpu-world.com/CPUs/K8/AMD-Second Generation Opteron 8218 - OSA8218GAA6CR (OSA8218CRWOF).html)
 - <http://www.cpu-world.com/CPUs/K10/AMD-Third Generation Opteron 8347 - OS8347WAL4BGE.html>
 - [http://www.cpu-world.com/CPUs/K10/AMD-Six-Core Opteron 8435 - OS8435WJS6DGN \(OS8435WJS6DGNWOF\).html](http://www.cpu-world.com/CPUs/K10/AMD-Six-Core Opteron 8435 - OS8435WJS6DGN (OS8435WJS6DGNWOF).html)
 - [http://www.cpu-world.com/CPUs/K10/AMD-Opteron 6168 - OS6168WKTCEGO \(OS6168WKTCEGOWOF\).html](http://www.cpu-world.com/CPUs/K10/AMD-Opteron 6168 - OS6168WKTCEGO (OS6168WKTCEGOWOF).html)
 - <http://www.cpu-world.com/CPUs/Bulldozer/AMD-Opteron 6272 - OS6272WKTGGGU.html>
 - <http://www.cpu-world.com/CPUs/Bulldozer/AMD-Opteron 6370 HE.html>
 - <http://www.cpu-world.com/CPUs/Zen/AMD-EPYC 7601.html>
 - <http://www.cpu-world.com/CPUs/Zen/AMD-EPYC 7742.html>
- IBM
 - <https://iee.ezproxy.univ-ubs.fr/document/8875663>
 - <https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&cad=rja&uact=8&ved=2ahUKEwjxmrTCgqbmAhXvzoUKHUjhDH0QFjAAegQIAhAC&url=ftp%3A%2F%2Fpublic.dhe.ibm.com%2Fsoftware%2Fmktsupport%2Ftechdocs%2Fpower4.pdf&usg=AOvVaw1dUMnqt6Giw4dbExqE5lZ5>
 - https://www.archive.ece.cmu.edu/~ece740/f13/lib/exe/fetch.php?media=le_power6.pdf

- ARM

- https://www.fujitsu.com/global/Images/post-k_supercomputer_with_fujitsu%27s_original_cpu_a64fx_powered_by_arm_isa.pdf
- https://www.hotchips.org/wp-content/uploads/hc_archives/hc26/HC26-11-day1-epub/HC26.11-4-ARM-Servers-epub/HC26.11.430-X-Gene-Singh-AppMicro-HotChips-2014-v5.pdf
- https://amperecomputing.com/wp-content/uploads/2019/01/eMAG8180_PB_v0.5_20180914.pdf
- <https://www.anandtech.com/show/13620/huawei-server-efforts-hi1620-and-arms-big-server-core-ares>
- <https://www.anandtech.com/show/15189/amazon-announces-graviton2-soc-along-with-new-aws-instances-64core-arm-with-large-performance-uplifts>