

Les modules Power-on-Package fournissent jusqu'à 1000A à des puces de processeurs IA

Paul Yeaman - Directeur, ingénieur d'applications

Les récentes avancées en matière d'intelligence artificielle ont entraîné une augmentation significative des besoins en matière de haute puissance de calculs. Les processeurs XPU axés sur l'IA, qui utilisent des circuits intégrés haut de gamme dotés de multiples cœurs et de milliards de transistors, consomment plusieurs centaines d'ampères et posent de nombreux défis pour les concepteurs de systèmes d'alimentation. Les ingénieurs qui créent des systèmes sur puce haute densité et gourmands en énergie disposent désormais d'un nouvel outil permettant de simplifier la partie fourniture de l'alimentation de leur projet: Power-on-Package (PoP), qui amène les dernières étapes critiques de la chaîne d'alimentation directement sur le substrat du XPU tel que l'illustre la figure 1, et libère des avantages pour la performance et l'efficacité. ("XPU" est ici un terme générique pour CPU, GPU, FPGA complexe ou ASIC à l'échelle du système.)

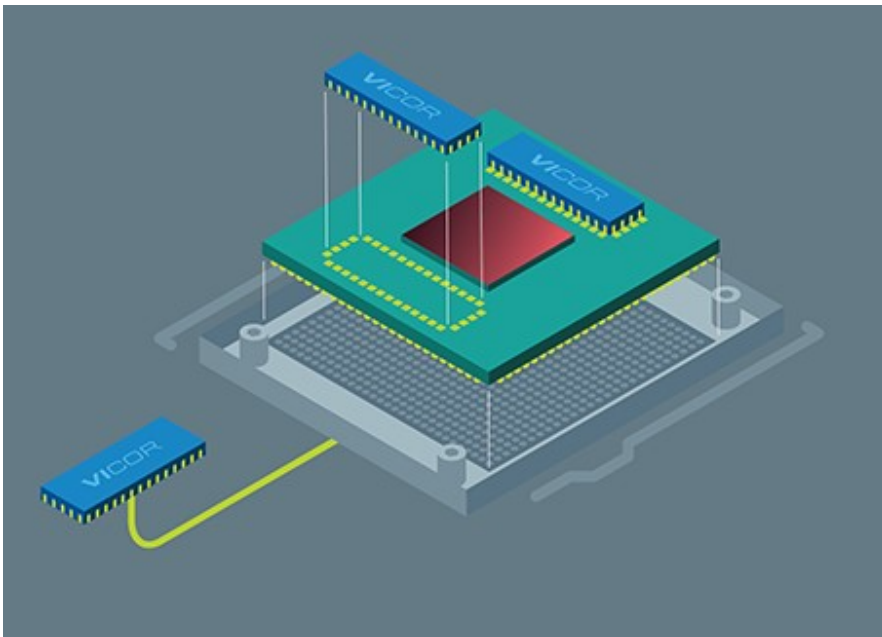


Figure 1. La solution Power-on-Package restructure les étapes finales de la fourniture d'énergie au XPU.

Au fur et à mesure de l'évolution de la technologie des semi-conducteurs, les besoins en énergie de périphériques tels que les cœurs de processeur ont considérablement évolué. Les tensions des noyaux sont passées de 5 V à 3,3 V, puis à 1,8 V, pour atteindre désormais des niveaux inférieurs à 1 V. La réduction correspondante de la taille des fonctionnalités permet de placer de plus en plus de logique sur une puce, si bien que de nombreux dispositifs complexes sont limités thermiquement: la quantité de chaleur que le boîtier peut dissiper détermine la limite supérieure de ce qui peut être placé dans la puce.

Avec une puissance de conception thermique (TDP) souvent bien supérieure à 100 W, la quantité de courant d'entrée requise a augmenté de façon spectaculaire avec la chute de la tension d'alimentation et les spécifications actuelles de plusieurs centaines d'ampères ne sont plus exceptionnelles. La pratique conventionnelle est d'alimenter les XPU. La pratique conventionnelle est d'alimenter les XPU avec une régulation multi phase à partir d'un bus de distribution 12V, mais fournir des centaines d'ampères de cette manière est extrêmement difficile.

Si les tensions d'alimentation ont diminué, la qualité requise pour l'alimentation fournie a augmenté dans le même temps. Puisque les marges de commutation représentent une proportion plus importante d'une tension d'alimentation très faible, l'alimentation elle-même doit être très stable et produire très peu de bruit afin d'éliminer la possibilité que des transitoires de tension entraînent des déclenchements intempestifs. Avec des courants de l'ordre de plusieurs centaines d'ampères, la moindre résistance dans le chemin d'alimentation peut entraîner des chutes de tensions inacceptables (I^2R). Les mêmes effets se produisant sur l'ensemble du chemin d'alimentation peuvent entraîner d'autres pertes indésirables. En outre, les grilles des XPU complètent souvent les noyaux des processeurs ou d'autres segments importants du circuit logique en les activant/désactivant à l'échelle des nanosecondes, générant de grands échelons en matière de courant exigé, que le schéma d'alimentation doit être en mesure de satisfaire sans effets transitoires indésirables.

La solution PoP représente la convergence des tendances, différentes mais intimement liées, en matière de conception d'alimentations : la séparation de la transformation de la tension et des fonctions de régulation, et le déplacement progressif de la dernière étape des systèmes d'alimentation au plus près du point de charge. En recherche perpétuelle de systèmes d'alimentation toujours plus efficaces et toujours plus performants, Vicor a mis au point des modules encore plus avancés, en faisant évoluer son architecture FPA à alimentation factorisée (*Factorized Power Architecture*) au travers de plusieurs étapes, débouchant sur la solution PoP proposée aujourd'hui, rendant possible l'utilisation de l'architecture FPA sur les applications de supercalculateur hautes performances.

L'alimentation factorisée permet la distribution de l'alimentation à des tensions plus élevées, typiquement 48 V, dans l'ensemble d'un système et sur la carte mère, avec une conversion à un seul étage à partir de cette valeur au point de charge. Sur les solutions industrielles classiques, il est difficile d'atteindre un taux de conversion aussi élevé. Vicor a mis au point une topologie appelée « Sine Amplitude Converter » (SAC) ou convertisseur d'amplitude sinusoïdale en français, qui permet de réaliser des taux de conversion très élevées, avec des pertes et un bruit faibles. Dans la topologie SAC, les transistors MOSFET d'alimentation commutent uniquement au point de passage à zéro de la tension et du courant, éliminant les pertes par commutation. La fréquence de commutation peut également être élevée, minimisant le besoin en condensateur de filtrage en sortie.

Sur la solution PoP, un module appelé « multiplicateur de courant modulaire » ou MCM, est placé directement sur le substrat qui supporte la puce du processeur. Comme son nom l'indique, le module MCM réalise la fonction d'un convertisseur DC/DC à rapport fixe (en réalité un « transformateur DC-DC »). Il génère la basse tension et le courant élevé nécessaires à la puce à partir d'une alimentation à plus faible courant et à plus haute tension, qui arrive via les broches du substrat/boîtier du XPU, et qui est délivrée dans le substrat ou la carte électronique sur le ou laquelle le processeur est monté. Cette solution offre plusieurs avantages clairs : le courant de crête le plus élevé est le chemin allant du MCM au XPU ; la très faible résistance sur ce chemin produit de faibles pertes par effet Joule (I^2R) ; la très faible inductance sur ce chemin améliore le comportement en régime transitoire du MCM.

C'est le domaine que l'on appelle le « dernier centimètre », qui couvre les quelques millimètres du chemin d'alimentation allant au processeur, comme on peut le voir sur la figure 2. Avec une solution classique, les pistes en cuivre du dernier centimètre et les broches du boîtier doivent présenter quelques centaines de micro-ohms de résistance. Pour un XPU haut de gamme consommant 200 A, un chemin de 500 $\mu\Omega$ entraînerait une chute de tension de 100 mV, soit plus de 10 % d'un rail d'alimentation inférieur à 1 V. Encore plus préoccupant, les pertes par effet Joule dans cet exemple impliquent 20 W supplémentaires de chaleur générés à proximité du XPU. Pour contourner ces obstacles, l'approche du PoP place le MCM sur le boîtier du composant, ce qui autorise toute l'alimentation fournie au processeur par la carte mère à rester à une tension plus élevée et à un niveau de courant réduit. Les pertes sur la carte mère elle-même sont ainsi diminuées, et le cuivre peut être économisé sur les plans d'alimentation de la carte électronique.

Le seul moyen d'offrir un chemin avec une impédance suffisamment faible, en l'absence de convertisseur sur le substrat, consiste à affecter de nombreuses broches à la fourniture d'alimentation du noyau. En plaçant le MCM directement sur le substrat du XPU, cependant, la fourniture d'alimentation depuis la carte électronique jusqu'au boîtier du processeur entraîne des courants typiquement de l'ordre de 1/50ème de la valeur nécessaire à l'alimentation du noyau. Le nombre de broches dédié à l'alimentation sur le boîtier du XPU est ainsi divisé par 10.

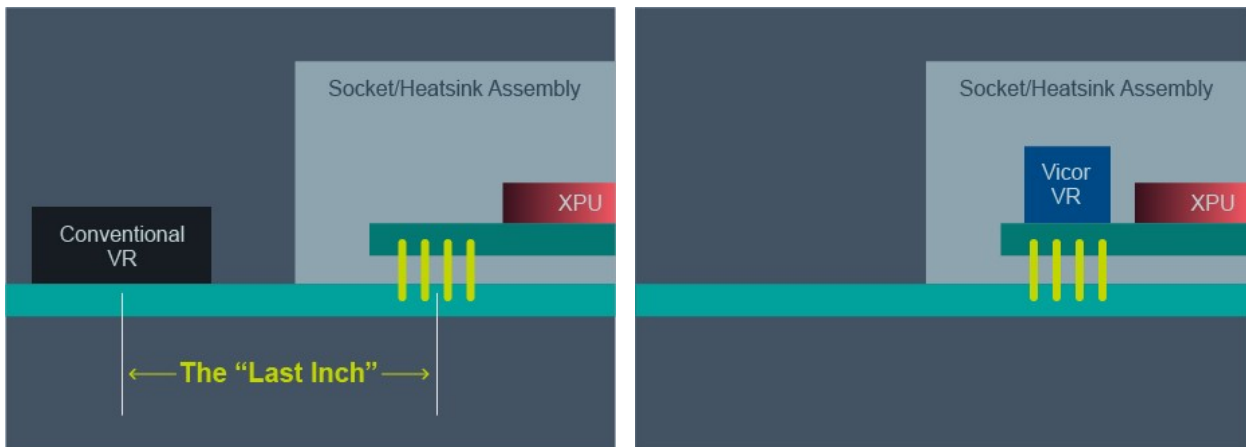


Figure 2. a) Étape finale classique d'une régulation de tension de noyau, avec un régulateur de point de charge proche du composant cible, mais monté sur la carte électronique hôte ; b) La solution PoP, où la régulation et la conversion sont déplacées sur le substrat du XPU.

Tout comme sur un système FPA classique, le système PoP place une étape de pré-régulation en amont du multiplicateur de courant : un module de pilotage du MCM, ou MCD (MCM Driver), qui est monté sur le circuit imprimé de la carte mère.

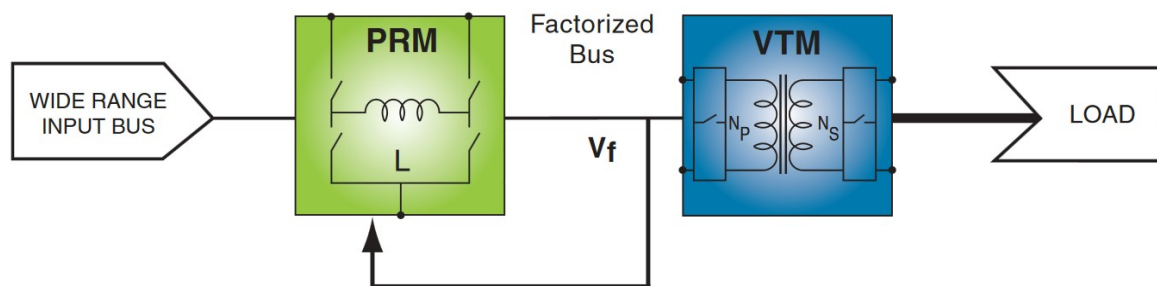


Figure 3. Version de l'architecture FPA à alimentation factorisée telle que déployée à l'origine dans des modules pré-régulateurs (PRM) et modules de transformation de la tension (VTM) par Vicor. La solution PoP réintègre les fonctions PRM et VTM dans les modules MCD et MCM contenant l'architecture FPA. Pour une régulation encore plus précise, on peut utiliser le retour de la sortie du dernier étage de conversion pour contrôler le module pré-régulateur (PRM).

Afin de fournir un format physique pouvant être monté sur le substrat du XPU, Vicor a configuré le boîtier SM-ChiP (figure 4), en réutilisant les techniques employées dans ses modules VI Chip. Le SM-ChiP utilise une construction de montage en surface avec des bornes destinées au substrat (ou à la carte mère) dotées d'une impédance minimale. Le module MCM3208S59Z01A6C00 mesure 32 x 8 mm pour 2,7 mm de hauteur. Deux modules suffisent pour fournir 2 x 160 A en continu, ou 2 x 320 A en courant de crête, alimentés par un module de pilotage MCM MCD3509S60E59D0C01 en amont (35 x 9 x 4,9 mm) sur la carte électronique hôte, qui est capable d'une puissance moyenne de 400 W et d'une puissance crête de 600 W. Pour les courants encore plus élevés, des modules MCM4608S59Z01B5T00 jumeaux (46 x 8 x 2,7 mm), chacun affichant un courant nominal de 300 A en continu et de 500 A en courant de crête, permettent à un système d'atteindre un courant de crête de 1000 A. Le module de pilotage MCM MCD4609S60E59H0T00 correspondant (46 x 9 x 4,9 mm) se monte sur la carte mère et fournit une puissance moyenne de 650 W et une puissance de crête de 1000 W.

Les exigences des courants élevés, imposés par les dernières générations de supercalculateurs, poussent à repenser la fourniture de l'alimentation. Aujourd'hui, alors qu'un petit centimètre de distance entre le régulateur et le processeur s'avère un facteur déterminant dans la perte de puissance, les modules Power-on-Package offrent une solution qui permet une fourniture d'énergie ultra efficace à très haute densité, critique pour les XPU haute puissance. Les modules Power-on-Package éliminent les barrières présentes avec les solutions d'alimentation classiques, en utilisant une tension distribuée plus élevée (48 V) et en permettant que la transformation de tension de l'étape finale ait lieu au plus près du XPU, supprimant réellement ce dernier centimètre.



Figure 4. Le boîtier SM-ChiP utilise une construction feuilletée et une interconnexion directe en montage de surface sur le substrat et la grille de connexion boîtier du XPU.



Figure S1. Le supercalculateur à refroidissement liquide par immersion Gyoukou d'ExaScaler/PEZY ZettaScaler-2.2.

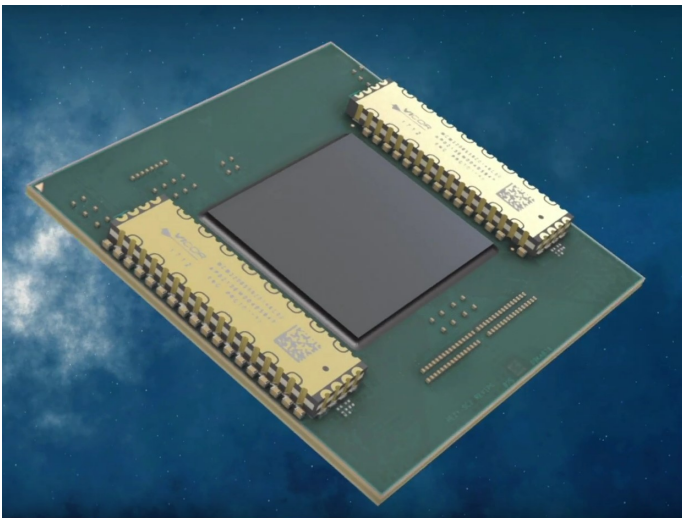


Figure S2. Les multiplieurs de courant MCM de Vicor, à côté de la puce du processeur sur le noyau de calcul d'ExaScaler du Gyoukou.

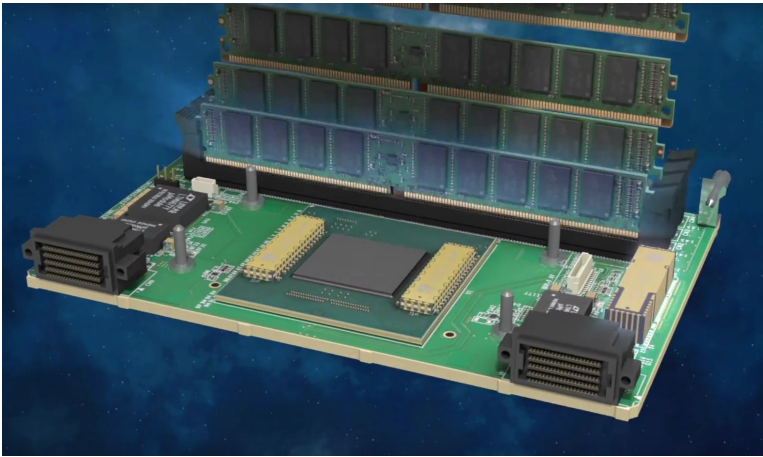


Figure S3. Le module de pré-régulation MCD Vicor (à droite) fournit l'alimentation à 48 V au substrat du XPU, accompagné d'un réseau de cellules de mémoire DDR4, sur chaque carte individuelle de l'ordinateur Gyoukou d'ExaScaler.

Le supercalculateur Gyoukou d'ExaScaler/PEZY ZettaScaler-2.2 (figure S1) se trouve dans l'institut de recherche de l'Agence japonaise de science et de technologie marine (JAMSTEC) de Yokohama au Japon. Le supercalculateur Gyoukou a été classé en novembre 2017 à la 5^{ème} place du Top500, le classement des machines les plus puissantes du monde, en termes de puissance de calcul, pour 19,1 PFlops/s (soit 10^{15} opérations à virgule flottante par seconde) et à la 4^{ème} place dans le Green500, palmarès prenant en compte l'énergie consommée par rapport à la puissance de calcul, avec un taux de performance de 14,1 GFlops/Watt.

Tout comme pour de nombreux autres supercalculateurs, la conception thermique constitue un point essentiel de l'architecture de la machine: il est refroidi par immersion, à l'aide de plusieurs bacs contenant les grappes de processeurs, dans lesquels un liquide diélectrique réfrigérant circule, transportant la chaleur vers des échangeurs de chaleur externes. Chaque bac contient 256 processeurs, utilisant une architecture FPA 48 V, et des CPU co-encapsulés avec des multiplicateurs de courant modulaires (MCM) Power-on-Package (PoP) de Vicor, qui réalisent la multiplication de courant directe de 48 V à moins d'1 V (figure S2).

Chaque processeur possède 2048 cœurs discrets, avec une horloge cadencée à 1 GHz et une mémoire cache intégrée à 3 niveaux. Le circuit intégré est encapsulé dans un module possédant sa propre mémoire DDR4 à haute bande passante (voir figure S3), pour une demande d'alimentation de 130 W. Trente-deux modules sont installées par châssis, formant une « brique » ; chaque bac de liquide contenant huit briques, plus de 33 kW à 48 V sont nécessaires pour chaque bac. Au total, la machine est constituée de 26 bacs de ce type.

L'architecture FPA mise en œuvre dans la solution Power-on-Package garantit que la plus grande partie de la chaleur générée par le superordinateur représente les opérations de calcul, et que la plus faible quantité possible est évacuée en pertes le long de la chaîne d'alimentation.