

PCIe400 : Estimation de puissance PTC



Julien Langouët,
Kévin Arnaud, Paul Bibron, Jean-Pierre Cachemiche,
Renaud Le Gac, CPPM

Plan

Contexte

Organisation

Choix technologiques

Statut

Conclusion

Méthodologie

Objectif

- Flow de conception inverse : HW avant FW
 - ▶ Besoin d'estimer les ressources du FPGA utilisées pour dimensionner les alimentations et le refroidissement

Mise en œuvre

- Utilisation de l'outil Power and Thermal Calculator (PTC) de Intel
 - ▶ Modèle préliminaire AGMF039R47A disponible depuis octobre 2022
- Utilisation des firmware LHCb développés pour PCIe40 et mise à l'échelle de l'utilisation des ressources
 - ▶ Power Analysis flow partiellement disponible pour Arria 10 sur Q18.1
 - Power Analysis ~ PTC après compilation de firmware
- Étude des paramètres par catégorie de ressource :
 - ▶ Logic, RAM, DSP, Clock, XCVR, HPS, NoC, HBM
 - ▶ >70 paramètres à estimer mais identification des paramètres influents

Cas considéré

- Cas typique : liens sériels à vitesse 10GbE
- Pire cas : ressources logiques supplémentaires, liens sériels à 25Gbps, ajout de l'interface 400GbE

Comparaison Agilex / Arria10

	Agilex (PCIe400)	Arria (PCIe40)	
	AGMF039R47A1E2V_R0	10AX115S3F45E2SG	
ALM	1 305 600	427 200	x3
FF (logic register)	5 222 400	854 000	x6
M20K	18 960	2 713	x7
DSP	12 300	1 518	x8
HPS	ARM Cortex A53 MPCore	-	
HBM2e	2x16GB	-	
NoC	256 initiators and targets	-	

- ▶ On utilise les données des firmware LHCb PCIe40 pour estimer l'utilisation des ressources en ALM, FF, M20K, DSP

Mise à l'échelle Agilex / Arria

Logic

	velo	calo	muon	scifi_ff	scifi_fv	plume	rich	AVERAGE	STD	test
ALM (%)	51	45	74	72	44	50	62	57	12	70
FF (%)	42	39	67	60	37	44	53	49	11	62

- On choisit un cas typique équivalent à la moyenne μ et un pire cas à $\mu+3\sigma$
 - ▶ Le LLI occupe 5 % de la logique et fonctionne à basse fréquence (~100MHz et un toggle rate $\ll 12.5\%$)
 - ▶ Le ACL (Application Core Logic) fonctionnerait à 640MHz (facteur 2) avec un toggle rate ~12.5 %

RAM

	velo	calo	muon	scifi_ff	scifi_fv	plume	rich	AVERAGE	STD	test
RAM (%)	68	67	78	97	96	76	80	80	12	100

- On choisit un cas typique équivalent à la moyenne μ et un pire cas à 100 %

DSP

	velo	calo	muon	scifi_ff	scifi_fv	plume	rich	AVERAGE	STD	test
DSP (%)	0.0	0.3	0.0	0.0	0.0	0.3	0.0	0	0.1	0

- On choisit un cas typique de 0.3 % et 1 % pour le pire cas et le cas typique
 - ▶ On choisit le type de DSP le plus énergivore : « Sum of 2 18x18 with accumulator »

Tableau récapitulatif

Paramètre commun

- Température de jonction à 85°C

TYPICAL CASE	Occupation (%)	Puissance (W)
ALM (%) LLI	5	26
FF (%) LLI	5	
ALM (%) ACL	55	0.5
FF (%) ACL	45	
RAM (%)	80	36
DSP (%)	0.3	0.3

WORST CASE	Occupation (%)	Puissance (W)
ALM (%) LLI	5	42
FF (%) LLI	5	
ALM (%) ACL	85	0.5
FF (%) ACL	75	
RAM (%)	100	35
DSP (%)	1	1

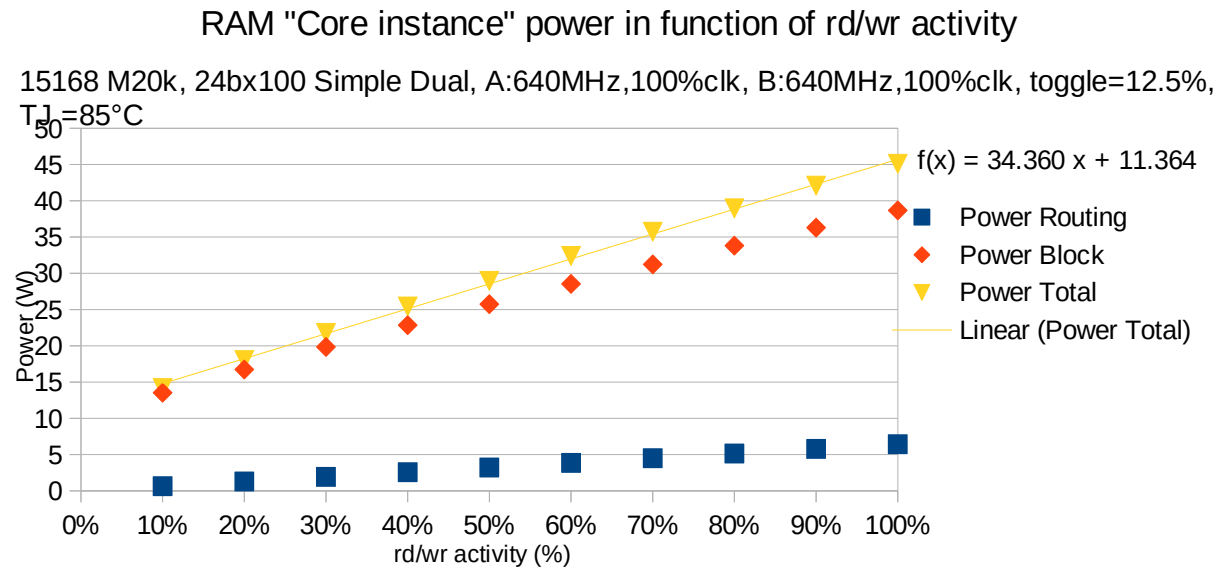
Paramètre commun

- Les ressources d'autre nature que les ALM ne sont pas négligeables

Discussion de certains paramètres

Taux d'utilisation en écriture et lecture des RAM

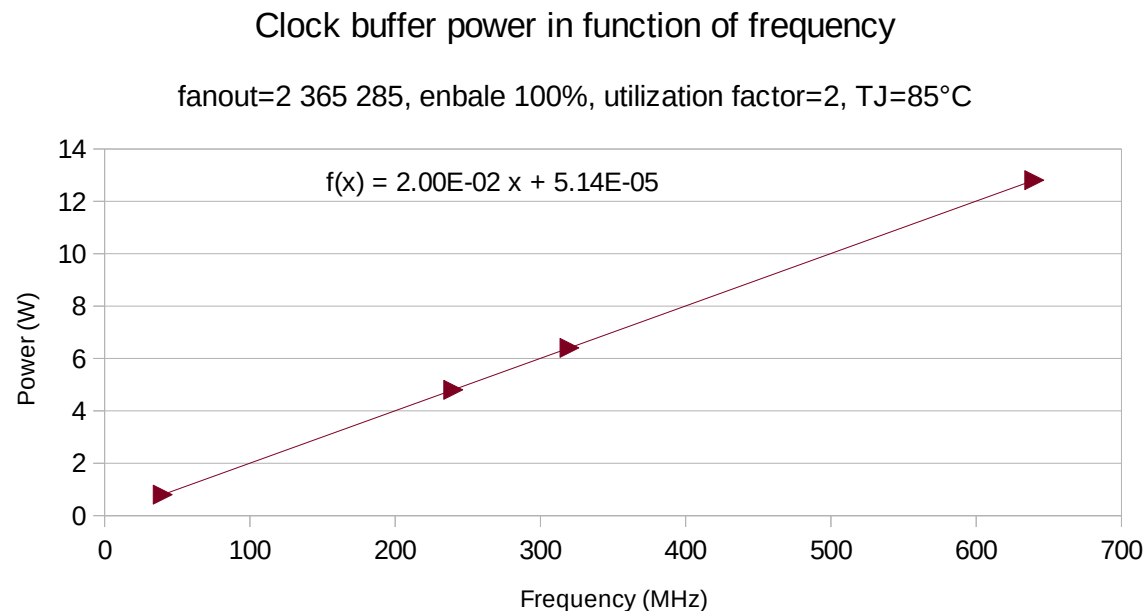
Valeur utilisée : 70 %



Discussion de certains paramètres

Fréquence

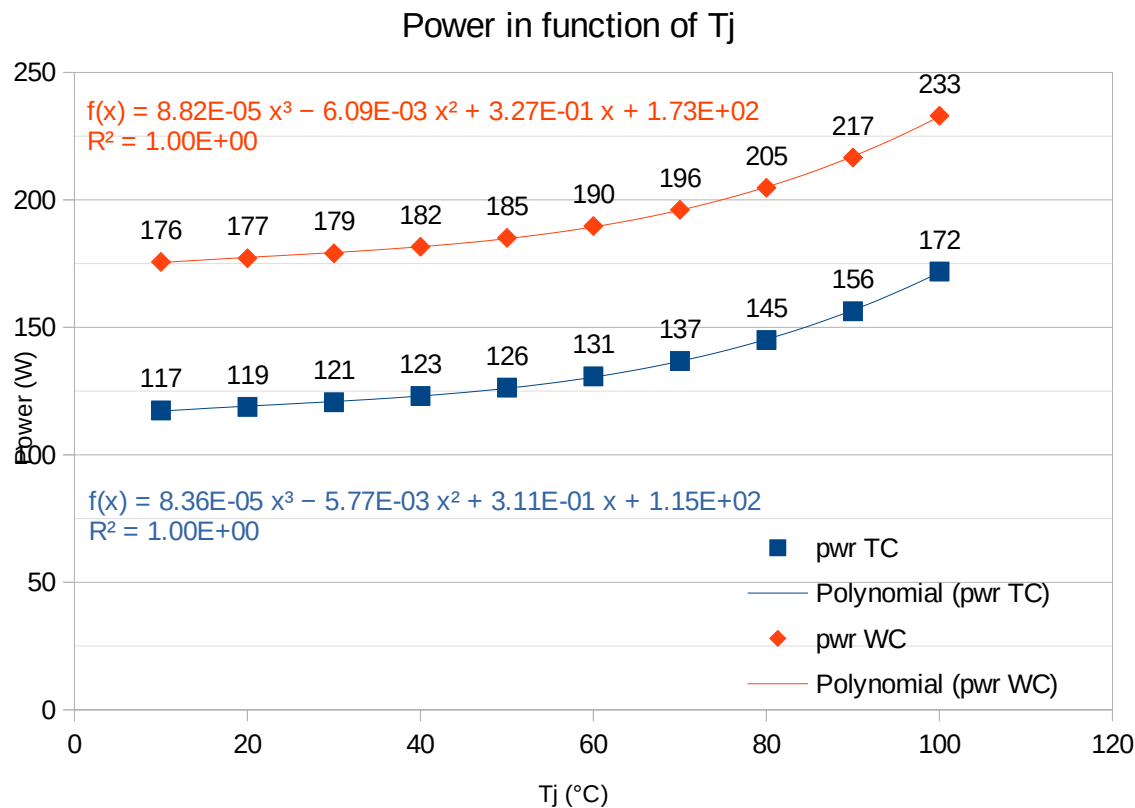
- On choisit 640MHz puisque le FPGA peut fonctionner jusqu'à 1000MHz et que la fréquence nominale dans l'Arria est de 320MHz. **Mais** toute la logique ne fonctionne pas à cette fréquence, il faudrait plutôt avoir une fréquence moyenne
 - ▶ Variation de la puissance linéaire
 - ▶ Si Toggle Rate absolu des firmware LHCb donné par le Power Analyzer, on peut déduire la fréquence moyenne de la logique



Bilan de puissance

Différence statique / dynamique

- La puissance statique est due à l'alimentation des cellules de ressource
 - Elle dépend de la température de jonction (évolution parabolique)
- La puissance dynamique est due à la commutation des cellules de ressource
 - Elle dépend de la présence d'une horloge, du toggle rate absolu (transitions par seconde)



Ressource type	Power dynamic (W) TC	Power dynamic (W) WC	multiplicative factor
Logic	31.0	47.1	1.52
RAM	31.2	42.3	1.36
DSP	0.3	1.0	3.06
Clock	13.0	21.5	1.65
PLL	0.0	0.0	-
IO	0.0	0.0	-
Transceiver	29.8	51.3	1.72
HPS	0.4	0.6	1.33
Crypto	0.0	0.0	-
NOC	0.0	0.0	-
HBM	0.0	0.0	-
Miscellaneous	1.5	1.5	1.00
TOTAL	107	165	1.54

Synthèse et prochaines étapes

Évaluation des paramètres

- Évaluation d'une fréquence moyenne à partir du toggle rate absolu
- Confirmation des ressources occupées par le LLI et baisser le TR $\ll 12.5\%$
- Confirmation des taux d'activité lecture écriture de la RAM (et aussi HBM)
- Étude d'un exemple design pour le 400GbE network interface/HBM/NoC

Solution de refroidissement

- Calcul de la résistance thermique nécessaire en fct de la température ambiante
 - ▶ Aide à déterminer dans quel cas solution à air ou liquide
 - ▶ Nécessite PTC Q22.4 (accès en avance de phase grâce à Intel via IPS case)

Vérification du dimensionnement de l'arbre d'alimentation