

Examen final – INF2500 Architecture des ordinateurs, Automne 2001

– Toute documentation et la calculatrice sont permises.

Écrire vos résultats de façon claire et précise. Pour les questions à développement, montrer les calculs, les réponses seules ne sont pas suffisantes.

Question 1 – (3 points)

On a besoin de concevoir un ordinateur spécialisé pour reconnaître automatiquement et à très grande vitesse un objet par reconnaissance des formes. L'approche suggérée est de faire des opérations sur chaque pixel de l'image vidéo. Quelle est la meilleure architecture parallèle pour ce type d'opérations?

- a) SISD
- b) SIMD
- c) MISD
- d) MIMD
- e) MIMS

Solution : SIMD (note: MIMS n'existe pas)

Question 2 – (3 points)

On vous demande de recommander un média de communication entre deux ordinateurs sur une distance de 1 km (sans amplificateur le long du trajet) et à une vitesse de transmission de 150 Mb/s. Nous devons avoir un média qui a la plus grande immunité aux interférences électriques externes. Comme deuxième priorité, nous devons également minimiser le prix d'achat. Que nous recommandez-vous?

- a) Paire torsadée (twisted pair);
- b) Cable coaxial;
- c) Fibre optique de type « multimode »;
- d) Fibre optique de type monomode « single-mode »;
- e) Communication sans-fil;

Solution : c.

Question 3 – (3 points)

Nous avons besoin d'un système fiable de lecture/écriture pour l'implantation d'une mémoire virtuelle dans un ordinateur de contrôle à bord d'un avion supersonique. Le coût n'est pas important mais le système doit être très rapide avec des temps d'accès relativement constants et pouvoir fonctionner dans un environnement qui a beaucoup de vibrations. Que recommandez-vous?

- a) Disque magnétique;
- b) Disque à semiconducteurs SSD (Solid State Disk);
- c) CDROM;
- d) Bande (tape) magnétique;
- e) Disque souple (floppy disk);

Solution : b.

Question 4 – (3 points)

Pour une mémoire virtuelle implantée sur un disque magnétique (hard drive), nous devons optimiser l'efficacité de transfert des données entre la mémoire virtuelle et la mémoire principale. Que recommandez-vous?

- a) Augmenter la dimension des pages et utiliser une stratégie d'écriture simultanée « write-through » au disque;
- b) Augmenter la dimension des pages et utiliser une stratégie de ré-écriture « write-back » au disque;
- c) Diminuer la dimension des pages et utiliser une stratégie d'écriture simultanée « write-through » au disque;
- d) Diminuer la dimension des pages et utiliser une stratégie de ré-écriture « write-back » au disque;
- e) Les réponses du haut n'ont rien à faire avec l'efficacité de transfert des données entre la mémoire virtuelle et la mémoire principale.

Solution : b.

Question 5 – (4 points)

Vous êtes impliqué dans l'implantation de la mémoire principale d'un nouvel ordinateur que votre compagnie veut concevoir et mettre sur le marché. Pour être très compétitif, la compagnie a accès à des DRAMs à très faible coût de toutes sortes mais qui ont des temps d'accès assez longs de 6 périodes horloge du processeur sur le RAS et 6 périodes horloge du processeur sur le CAS. L'ordinateur sera utilisé principalement dans des applications où l'accès à la mémoire principale sera hautement aléatoire. Avant de débiter le design, la compagnie vous demande de recommander la meilleure solution parmi ces alternatives possibles:

- a) Implanter 6 bancs (banks) entrelacées « interleaved » de mémoire avec des DRAM « nibble mode »;
- b) Implanter 6 bancs (banks) entrelacées « interleaved » de mémoire avec des DRAM « nibble mode » ou « page mode »;
- c) Implanter le maximum de bancs (banks) entrelacées « interleaved » de mémoire avec des DRAM « page mode » ou « static column »;
- d) Implanter le maximum de bancs (banks) entrelacées « interleaved » de mémoire avec des DRAM « nibble mode » ou « static column »;
- e) Implanter le minimum de bancs (banks) entrelacées « interleaved » de mémoire avec des DRAM « nibble mode » ou « static column »;
- f) Implanter le minimum de bancs (banks) entrelacées « interleaved » de mémoire avec des DRAM « nibble mode » ou « static column »;
- g) Éviter absolument le mode entrelacé « interleaved »

Solution : c.

Question 6 – (3 points)

Votre compagnie a eu un contrat pour le développement d'un superordinateur de contrôle en temps réel de très haute performance où la priorité du design doit minimiser les différences entre les temps d'accès à la

mémoire pour garantir une performance très constante. Le programme et les données ont une dimension d'environ 120K. Le prix d'implantation n'est pas une priorité mais le design du pipeline est déjà très complexe et la compagnie veut éviter si possible une solution qui risquerait de compliquer davantage le design du pipeline. Que recommandez-vous **par ordre du plus recommandé au moins recommandé** parmi ces options :

1. Utiliser uniquement que du SRAM rapide sans mémoire cache;
 2. Utiliser une mémoire cache associative par ensembles de 2 blocs « 2-way set associative » de 32Ko;
 3. Utiliser une mémoire cache avec un degré d'associativité de 4 « 4-way set associative » de 64Ko;
 4. Utiliser une mémoire cache de 32Ko pseudo-associative avec un degré d'associativité de 2 « 2-way pseudo-associative ».
- a) 1, 2, 3, 4
 - b) 4, 3, 2, 1
 - c) 1, 4, 3, 2
 - d) 1, 3, 2, 4
 - e) 4, 1, 3, 2

Solution : d.

Question 7 – (6 points)

Vous êtes consulté pour faire une recommandation pour une nouvelle architecture d'ordinateur aux ingénieurs impliqués dans le design du cache et à ceux responsables de la configuration de la mémoire. L'ordinateur doit être conçu pour exécuter des programmes ou tâches informatiques qui changent dans des intervalles de temps extrêmement courts pour supporter des applications multitâches « multi-tasking » à très haute fréquence. À chacun des changements de tâche, la mémoire cache doit être vidée (flushed) de son contenu. En discutant autour de la table de réunion, vous entendez plusieurs théories venant du personnel de la compagnie.

7.1 Parmi ces théories, laquelle est **vraie**? (3 points)

- a) Ceci semble être des échecs de type obligatoire (compulsory) et donc dépendants de la dimension de la cache;
- b) Ceci semble être des échecs de type capacité et donc dépendants de la dimension de la cache;
- c) Ceci semble être des échecs de type obligatoire (compulsory) et sont indépendants de la dimension de la cache;
- d) Ceci semble être des échecs de type capacité et donc indépendants de la dimension de la cache;

Solution : c.

7.2 Que recommandez-vous? (3 points)

- a) Augmenter la dimension du bloc;
- b) Diminuer la dimension du bloc;
- c) Augmenter le degré d'associativité dans la mémoire cache;
- d) Diminuer le degré d'associativité dans la mémoire cache.

Solution : a.

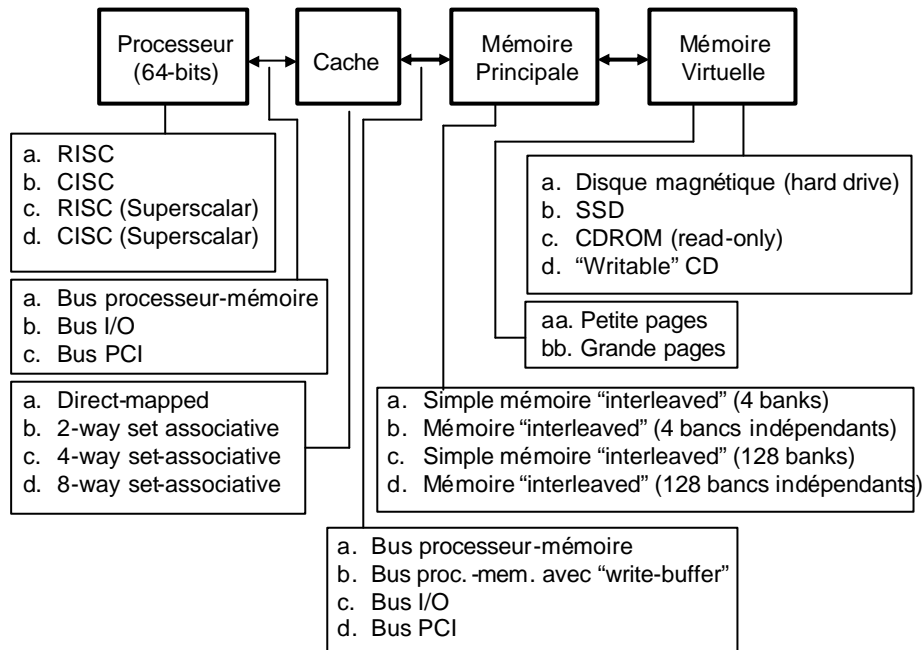
Question 8 – (8 points)

Vous devez concevoir 2 ordinateurs. Le premier modèle (Modèle CHEAP PRO) doit être le moins cher possible, tandis que le deuxième modèle (Modèle EMPTY WALLET) doit être conçu pour les professionnels qui recherchent la plus haute performance.

Dans les deux modèles, la mémoire cache d'instructions doit être la plus petite possible.

À partir du schéma suivant, concevez les deux modèles.

Note : degré d'associativité = « set associative »; totalement associative = « fully-associative »; correspondance directe = « direct-mapped »; tampon d'écriture = « write-buffer »; superscalaire = « superscalar »; entrelacé = « interleaved »; accès proches répétés = « nibble mode »; banc de mémoire = « memory bank »



Modèle « CHEAPPRO » (4 points)

8.1 (0.5 point) Processeur : a, b, c, ou d. **b**

8.2 (0.5 point) Bus Processeur-cache : a, b, ou c. **a**

8.3 (1 point) Cache : a, b, c, or d. **a**

8.4 (0.5 point) Bus cache-mémoire principale : a, b, c, ou d. **a**

8.5 (0.5 point) Mémoire principale : a, b, c, ou d. **a**

8.6 (0.5 point) Mémoire virtuelle : a, b, c, ou d. **a**

8.7 (0.5 point) Configuration de la mémoire virtuelle : aa ou bb. **aa (moins de perte espace disque)**

Modèle « EMPTY WALLET » (4 points)

- 8.8 (0.5 point) Processeur : a, b, c, ou d. **d**
- 8.9 (0.5 point) Bus Processeur-cache : a, b, ou c. **a**
- 8.10 (1 point) Cache : a, b, c ou d. **d (pas d'avantage à aller plus loin)**
- 8.11 (0.5 point) Bus cache-mémoire principale : a, b, c, ou d. **b**
- 8.12 (0.5 point) Mémoire principale : a, b, c, ou d. **d**
- 8.13 (0.5 point) Mémoire virtuelle : a, b, c, ou d. **b**
- 8.14 (0.5 point) Configuration de la mémoire virtuelle : aa ou bb. **bb (plus performant)**

Question 9 – (5 points)

Lequel des énoncés suivants est **vrai**?

- Pour optimiser le temps d'exécution d'un programme, il faut porter attention au nombre d'instructions statiques du programme parce qu'un programme passe 90% de son temps d'exécution sur 10% des instructions.
- Pour un système de mémoire conçu pour une mise à niveau au cours des trois prochaines années, il faudrait que la conception puisse accepter des DRAM dont les temps de cycle sont diminués d'environ 40% par rapport aux DRAM installés à la conception initiale.
- Lorsque la surface d'une puce diminue de 10%, le coût de la puce diminue du tiers
- Si une amélioration est utile seulement pour une fraction d'une tâche, on ne peut accélérer cette tâche par un facteur supérieur à l'inverse de la fraction améliorée.
- Il faut concevoir la mémoire virtuelle d'un ordinateur en tenant compte des suspensions dues aux défauts de page.

Solution :

- faux : c'est le nombre d'instructions dynamiques qui compte
- faux : les temps de cycle diminuent de 33% en 10 ans (p. 7)
- vrai : p. 14 le coût est une fonction de la surface puissance 4
- faux : p.33 énonce le corollaire correctement
- faux : p. 44 l'UC n'est pas suspendue sur un défaut de page

Question 10 – (4 points)

Lequel des énoncés suivants est **faux**?

- Le nombre d'adresses mémoire pouvant être utilisées dans une instruction est toujours inférieur ou égal au nombre au nombre maximum d'opérandes autorisé.
- Lorsqu'on conçoit un système de communications, l'ordre de rangement des octets dans un mot (Little et Big Endian) doit être défini pour l'interface entre ces machines.
- Dans une instruction de branchement conditionnel du DLX, la valeur immédiate ne représente pas la valeur de comparaison du registre.
- Pour les accès à des tableaux, les modes d'adressage Indexé, Auto Incrémenté, Auto-décrémenté et Indexé étendu peuvent être utiles.
- Un codage sur 8 bits du déplacement relatif à CP permet de faire des branchements de 256 instructions, relativement à l'instruction de branchement

Solution :

- vrai
- vrai (p. 80)

- c) vrai : cette valeur représente le déplacement vers l'adresse de destination (p. 136)
- d) vrai (p. 82)
- e) faux (p. 91)

Question 11 – (4 points)

Par rapport à l'instruction qui précède, indiquer la dépendance des instructions par une des réponses suivantes :

- (a) pas de dépendance;
- (b) dépendance LAE nécessitant une suspension par verrouillage du pipeline;
- (c) dépendance LAE évitée par l'envoi
- (d) dépendance EAE évitée par l'envoi
- (e) dépendance EAL évitée par l'envoi
- (f) dépendance avec accès ordonné par les phases première moitié et deuxième moitié du cycle
- (g) pas de dépendance

Question	Instructions du Programme	Dépendance par rapport à l'instruction	Réponse
3.1	Boucle : LW R1, 0(R2)	BNEZ R4, Boucle	
3.2	ADDI R1, R1, #1	Boucle : LW R1, 0(R2)	
3.3	SW 0(R2), R1	ADDI R1, R1, #1	
3.4	ADDI R2, R2, #4	SW 0(R2), R1	
3.5	SUB R4, R3, R2	ADDI R2, R2, #4	
3.6	BNEZ R4, Boucle	SUB R4, R3, R2	

Solution:

Question	Instructions du Programme	Dépendance par rapport à l'instruction	Réponse
3.1	Boucle : LW R1, 0(R2)	BNEZ R4, Boucle	g
3.2	ADDI R1, R1, #1	Boucle : LW R1, 0(R2)	b
3.3	SW 0(R2), R1	ADDI R1, R1, #1	c
3.4	ADDI R2, R2, #4	SW 0(R2), R1	g
3.5	SUB R4, R3, R2	ADDI R2, R2, #4	c
3.6	BNEZ R4, Boucle	SUB R4, R3, R2	c

Question 12 – (4 points)

Vous référant à la figure 5.9, Lequel des énoncés suivants est **faux** :

- a) Le taux d'échec d'un cache unifié de 2 Ko à correspondance directe est de 9,8 %
- b) Le taux d'échec d'un cache de 32 Ko de degré d'associativité = 2 est le même que celui d'un cache de 64K à correspondance directe
- c) Un cache à degré d'associativité plus élevée a un taux d'échecs de capacité plus élevé
- d) Les taux d'échec obligatoires ne dépendent pas de la taille du cache ni de son degré d'associativité
- e) Un cache de 2 Ko associatif par ensemble a plus d'échecs de capacité que d'échecs de conflit

Solution :

- a) Vrai : Le taux d'échec d'un cache unifié à correspondance directe de 2 Ko est de 9,8% (0.098)
- b) Vrai : 0,014 dans les 2 cas
- c) Faux : le taux d'échec de capacité est indépendant du degré d'associativité
- d) Vrai : toujours 0,002
- e) Vrai : (mais ce n'est pas vrai pour celui à correspondance directe)

Question 13 – (4 points)

Lequel des énoncés suivants est **faux** ?

- a) Le DMA permet de libérer l'UC des tâches de transfert de données lors des entrées-sorties.
- b) Le temps de réponse moyen d'un système d'E/S divisé par le nombre moyen de tâches dans le système est égal à l'intervalle moyen d'arrivée entre les tâches
- c) Lorsque l'utilisation d'un serveur approche les 90%, le temps de file est d'environ 900% du temps serveur.
- d) Lorsque l'utilisation serveur est nul, le temps de file est égal au temps serveur.
- e) Dans une matrice de disques, l'inconvénient est que la fiabilité chute de façon inversement proportionnelle au nombre de disques.

Solution :

- a) vrai (p.526)
- b) vrai : reformulation de la loi de Little (p. 531)
- c) vrai (p. 537)
- d) faux (p. 537)
- e) vrai (p. 545)

Question 14 – (3 points)

Lequel des énoncés suivants est **vrai**?

- a) La latence des communications dans un WAN dépend peu de la distance physique entre les nœuds de communication.
- b) L'avantage d'un routage «coupe à travers» par rapport à un routage « ranger et propager » est qu'il diminue la latence
- c) Dans un lien Ethernet, l'arbitrage est déterminé par un arbitre central supplémentaire, pour qu'un seul message soit envoyé à la fois.
- d) La suppression de paquets est peu utilisée pour contrôler la congestion car le logiciel de haut niveau ne réalise que rarement la gestion des erreurs de transmission.

- e) Les entêtes TCP sont séparées des données IP dans un datagramme, pour encapsuler les octets de messages à l'intérieur d'une fenêtre coulissante.

Solution

- a) faux (voir exemple p. 595)
 b) vrai (page 621)
 c) faux (p. 605)
 d) faux (p. 623)
 e) faux ; charabia ! (p. 639)

Question 15 – (10 points)

On a installé une passerelle pour interconnecter des segments de réseau Ethernet. Les paquets arrivent de manière exponentielle à 1500 paquets par seconde, et la passerelle les sert de façon exponentielle, en un temps moyen de 600 microsecondes.

- a) Quelle est l'utilisation de la passerelle? (2 points)
 b) Quel est le temps moyen passé dans la passerelle? (3 points)
 c) Quel est le nombre moyen de paquets dans la passerelle ? (3 points)
 d) (2 points) Quelle est la probabilité d'un débordement de la FIFO si elle peut contenir 20 paquets?

Utilisez cette loi : La probabilité de trouver n tâches ou plus dans un système M/M/1 est de $Utilisation^n$

Solution

Ce problème est basé sur l'exercice 7.13 du livre.

a)

$$Utilisation = \frac{Débit_{arrivé}}{Débit_{service}} = \frac{1500}{1/(600 \cdot 10^{-6})} = 90\%$$

b) et c) le temps moyen passé dans le serveur est de 6,0 ms.

Il faut tenir compte du serveur et de la file d'attente en utilisant la loi de Little

$$Temps_{file} = Temps_{serveur} \cdot \frac{Utilisation_{Serveur}}{1 - Utilisation_{Serveur}}$$

$$= 600 \cdot 10^{-6} \cdot \frac{90\%}{10\%} = 5,4ms$$

$$Longueur_{file} = Taux_{arrivé} \cdot (Temps_{file} + Temps_{serveur})$$

$$= 1500 \cdot (5,4ms + 0,6ms) = 1500 \cdot 6 \cdot 10^{-3} = 9$$

- d) La probabilité d'avoir 21 tâches ou plus dans le système est de $90\%^{21} = 11\%$

Question 16 – (5 points)

Un disque typique annonce un temps de recherche moyen (seek time) de 8 ms et tourne à 7200 tours par minute. Le débit de transfert est de 1 Mo par seconde et le surcoût du contrôleur est de 2 ms. Quel sera la taille minimale d'un secteur si l'on désire qu'au moins 10 % du temps d'accès moyen soit utilisé au transfert d'un secteur ? (Note : La taille d'un secteur doit être une puissance de 2!)

Solution

Ce problème est basé sur l'exemple au bas de la page 511

On doit calculer le rapport (Temps de transfert / Temps d'accès moyen)

Temps d'accès moyen = Temps de recherche moyen + Retard de rotation moyen + Temps de transfert
+ Surcoût du contrôleur

Temps d'accès moyen = 8ms + 0.5 / (7200 tr / minute) + Temps de transfert + 2ms
= 8ms + 4.2 ms + Temps de transfert + 2ms
= 14.2 ms + Temps de transfert

Donc

Temps de transfert
----- > 0.1
Temps de transfert + 14.2ms

On a finalement Temps de transfert > 1.578 ms

Ou encore (Taille_{Secteur} / 1 Mo/s) > 1.578 ms

Donc la Taille_{Secteur} > 1654 octets et la première puissance de 2 au-dessus de 1654 est **2048**.

Question 17 – (5 points)

Un programme test exécute un total de 1000 instructions. Le CPI des instructions entières est de 1 et le CPI des instructions flottantes est de 4. Si le CPI global est de 1.9, quel est le nombre d'instructions entières ?

Réponse

$CPI_{global} = (CPI_{entier} * NI_{entier} + CPI_{flottant} * NI_{flottant}) / NI$

Et $NI = NI_{entier} + NI_{flottant}$

Donc $1.9 = (1 * NI_{entier} + 4 * (1000 - NI_{entier})) / 1000$

Donc $NI_{entier} = 700$

Question 18 – (3 points)

Si l'on suppose que 80% des branchements d'un programme test du DLX, sont pris et que le compilateur peut remplir 70% des délais de branchement.

Quelle technique minimisera l'impact des aléas de contrôle sur la performance du pipeline?

- a) Suspension du pipeline à chaque branchement
- b) Prédiction pris
- c) Prédiction non-pri
- d) Branchement différé

Solution

- a) Perd 1 cycle a chaque branchement
- b) Perd 1 cycle a chaque branchement (car il faut attendre que le CP soit calculé!)
- c) Perd 1 cycle 80% des branchements
- d) Perd 1 cycle 30% des branchements

Donc la réponse est d)

Question 19 – (3 points)

Lequel des énoncés suivants est **faux** ?

Les opérations multicycles (pipeline flottant) ...

- a) Augmentent la probabilité (ou la fréquence) des aléas de données LAE.
- b) Introduisent des aléas de données EAE dans le pipeline.
- c) Amènent des conflits de lecture des registres.
- d) Amènent des conflits d'écriture des registres.
- e) Introduisent de nouveaux conflits structurels.

la réponse est c), les lectures registres sont toujours au début!

Question 20 – (3 points)

Lequel des énoncés suivants est **faux** ?

Un bus de haute performance...

- a) A des lignes d'adresses et de données séparées.
- b) A une horloge asynchrone.
- c) A une largeur de données supérieure au mot.
- d) A plusieurs maîtres du bus.
- e) Utilise des transactions éclatées (split transactions).

la réponse est b), les bus de haute performance ont des horloges synchrones (voir p. 520).

Question 21 – (3 points)

Lequel des énoncés suivants est **faux** :

- a) Une taille de bloc optimale réduit le taux d'échec d'une cache.
- b) Le programmeur peut contribuer à réduire le taux d'échec d'une cache.
- c) L'instruction *prefetch* (lecture anticipée) n'est utile que dans le cas d'une cache non-bloquante.
- d) Une cache dont l'associativité est plus élevée a un temps d'accès réussi plus bas.

- e) Une cache totalement associative de 1024 octets a un taux d'échec plus bas qu'une cache à correspondance directe de 512 octets.

la réponse est d) c'est le contraire, il y a plus d'étiquettes a vérifier... au pire c'est pareil!

Question 22 – (5 points)

Une cache de 4096 octets utilise la correspondance directe et la taille des blocs est de 256 octets. La mémoire principale (RAM) peut contenir 2^{16} blocs.

Quelle sera la structure de l'adresse provenant de l'UC (CPU)?

- a)

16 bits	8 bits
Index	Déplacement
- b)

12 bits	4 bits	8 bits
Étiquette	Index	Déplacement
- c)

4 bits	12 bits	8 bits
Étiquette	Index	Déplacement
- d)

16 bits	8 bits
Étiquette	Déplacement

la réponse est b) car il y a $4096/256=16 (2^4)$ blocs dans la cache donc 4 bits d'index et le reste pour l'étiquette

Question 23 – (3 points)

Lequel des énoncés suivants est **faux** ?

- La mémoire virtuelle permet de simuler une plus grande capacité mémoire.
- La traduction de l'adresse virtuelle en adresse physique ce fait grâce à une règle de correspondance fixe (exemple : No Page Physique = No Page Virtuelle MODULO Nombre de Page Physiques).
- Le remplacement de pages utilise presque toujours l'algorithme LRU (least recently used).
- La pénalité d'un échec cache est négligeable par rapport à la pénalité d'une faute de page.
- Le système d'exploitation fait la gestion de la mémoire virtuelle.

la réponse est b) La traduction utilise une table des pages! Ce n'est pas comme une cache!

Question 24 – (3 points)

Lequel des énoncés suivants ne contribue pas à réduire la longueur moyenne des instructions d'un programme?

- Utiliser un déplacement relatif par rapport au compteur programme (CP)
- Utiliser plusieurs formats d'instructions.
- Diminuer le nombre de modes d'adressages.
- Concevoir une machine orthogonale.
- Diminuer le nombre de registres.

la réponse est d) Une machine orthogonale permet tous les modes d'adressage avec toutes les opérations et tous les registres.