

Module 8: Entrées/sorties

1 Introduction

Les entrées/sorties représentent l'interface entre l'ordinateur et le monde extérieur. Sans les entrées/sorties, l'ordinateur devient inutile. Les périphériques communiquent avec le processeur via des bus de communication standard véhiculant des signaux d'interruption.

Parmi les périphériques les plus courants, se trouvent les systèmes de stockage, les écrans, et les imprimantes. Dans ce texte, nous allons mettre l'accent sur les communications entre les différents composants d'entrées/sorties et le processeur et la mémoire principale.

À la fin de ce texte, nous allons présenter les différents types de bus utilisés ainsi que leurs caractéristiques.

2 Bus

Le bus est le moyen principal d'échange de l'information entre les différents composants de l'ordinateur à savoir le processeur, la mémoire principale, la mémoire cache, la mémoire virtuelle, etc. Le bus est un ensemble de fils permettant de relier des circuits entre eux. La figure 1 montre un exemple de bus de communication¹.

1. Figure tirée de Wikipédia.

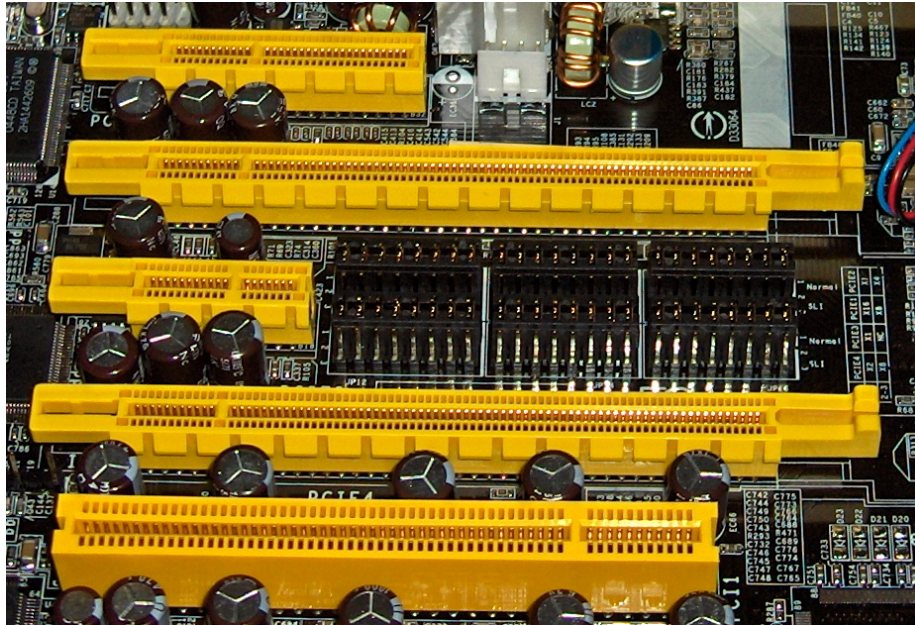


FIGURE 1 – Bus de communication.

nous pouvons avoir plusieurs bus de communication dans les ordinateurs actuellement. Bien évidemment, chaque type de bus a ses propres caractéristiques. Par exemple, un bus qui lie le processeur à la mémoire cache doit être assez efficace pour pouvoir supporter la vitesse du processeur. De la même façon, les bus d'entrées/sorties doivent être standards pour accommoder les différents types d'entrées/sorties.

Le bus relié au processeur s'appelle le **bus système** ou FSB (Front Side Bus). C'est l'unique moyen de communication du processeur. À cet effet, ce bus doit être :

- le plus rapide possible pour accommoder la vitesse du processeur, et
- le plus large possible pour gérer les quantités de données transférées.

Selon le texte du livre de architecture de l'ordinateur, le bus système est relié à un circuit intégré spécialisé, le pont nord (northbridge), chargé de répartir les informations entre processeur, mémoire, carte vidéo et le reste des composants via un deuxième circuit, le pont sud (southbridge), en charge des différentes entrées/sorties. La figure 2 montre comment le bus système est relié à différents composants².

2. Figure tirée du livre de l'architecture de l'ordinateur.

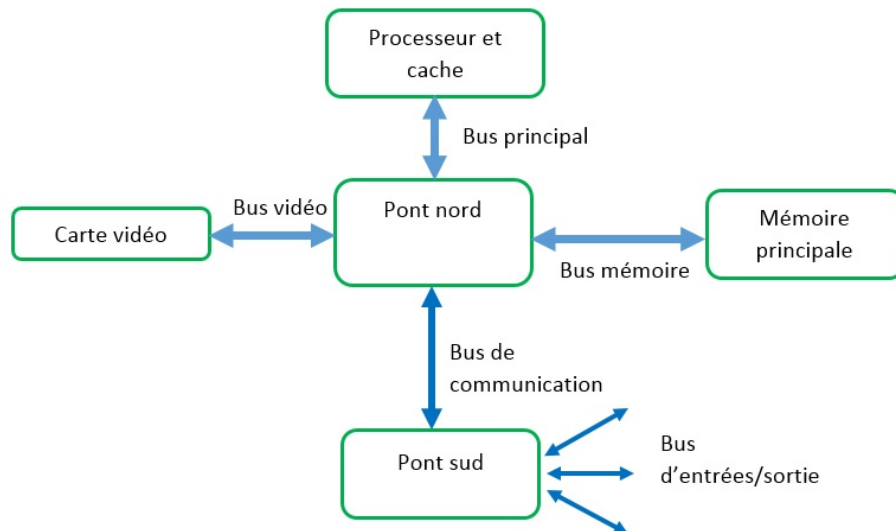


FIGURE 2 – Architecture d'un ordinateur récent.

Les bus se divisent généralement en trois catégories³ :

- **Bus d'adresse (Address Bus)** : c'est un ensemble de fils permettant de véhiculer les adresses afin de désigner un élément ou une case mémoire. Sa largeur (son nombre de bits) donne une indication sur l'espace mémoire adressable. De 16 bits à la fin des années 70 (soit 64 Ko adressables (2^{16})), il est passé à 20 bits puis à 32 bits. Sur les processeurs récents, suivant les modèles, il atteint les 36 à 42 bits. Il véhicule des adresses physiques pour communiquer avec la mémoire principale. Sa taille est donc indépendante des adresses virtuelles utilisées par un processus comme nous l'avons vu dans le module précédent sur les mémoires virtuelles. Il est souvent unidirectionnel car c'est le processeur qui indique une adresse aux autres composants.
- **Bus de données (Data Bus)** : c'est un ensemble de fils permettant de transférer les informations (données et instructions) d'un composant à un autre. Sa largeur influence directement le débit d'informations disponible sur le bus. S'il est de 8 bits, il permet d'échanger 1 octet à la fois, soit huit fois moins de données qu'un bus de 64 bits (valeur standard actuelle) fonctionnant à la même vitesse.
- **Bus de contrôle (Control Bus)** : appelé aussi bus de commande, c'est un ensemble de fils permettant de transporter les signaux de contrôle indiquant le type d'opération désiré (lecture ou écriture mémoire, accès vidéo, opération d'entrées/sorties, etc.), dont les éléments (adresse, don-

3. Texte copié avec quelques modifications du livre de l'architecture de l'ordinateur pour les besoins du module.

née) sont disponibles sur les deux autres bus. Il est bidirectionnel de sorte que les composants de l'ordinateur qu'il relie peuvent communiquer si besoin.

La figure 3⁴ montre un exemple des différentes catégories de bus et leurs interconnexions.

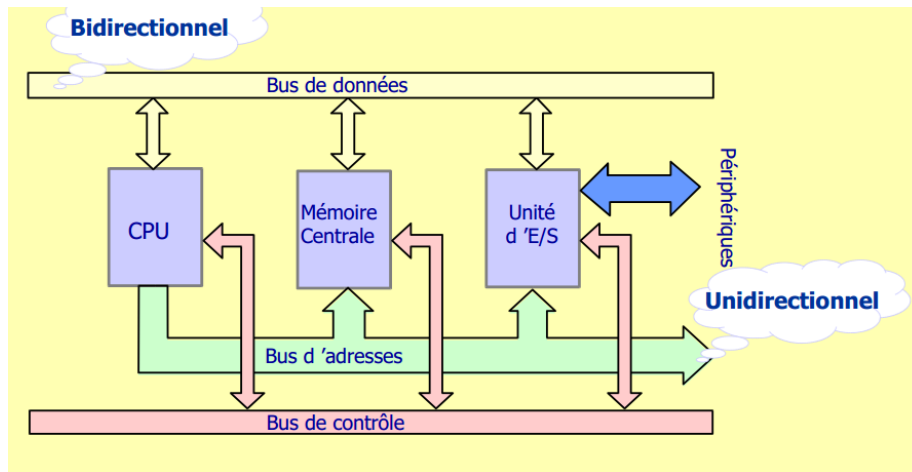


FIGURE 3 – Différentes catégories de bus.

3 Horloge de bus

Les opérations de transfert d'information sont fondamentalement liées au paramètre temporel. Notez bien que les différents composants dans l'ordinateur travaillent à des rythmes différents. À cet effet, la synchronisation de ces composants est indispensable. Il existe deux techniques utilisées :

3.1 Bus asynchrone

Le bus asynchrone, comme son nom l'indique, permet de relier deux composants sans aucune contrainte temporelle. C'est-à-dire, chaque composant peut travailler à son propre rythme sans se soucier des autres.

Le bus asynchrone utilise le mécanisme de poignée de main (handshaking) par signaux de requête et d'acquiescement pour permettre de signaler la présence d'une demande ou d'une réponse.

4. Image tirée de Google images

3.2 Bus synchrone

Contrairement au bus asynchrone, le bus synchrone est fortement lié au paramètre temporel.

La particularité du bus synchrone est qu'il intègre un signal d'horloge sur un de ses fils de la partie contrôle. Le rôle de ce signal est de permettre aux différents composants connectés au bus de se synchroniser lors d'un échange de l'information.

4 Gestion des entrées/sorties

Il existe trois méthodes de gestion des entrées/sorties :

- La liaison programmée.
- Les entrées-sorties pilotées par les interruptions.
- L'accès direct à la mémoire, DMA (Direct Memory Access).

4.1 La liaison programmée

Dans cette méthode, le contrôleur d'entrées/sorties est connecté via un bus à une paire de registres d'entrées/sorties, soit un registre pour les adresses et un autre pour les données comme illustré dans la figure 4⁵. Les échanges avec les périphériques sont contrôlés et pilotés par le processeur.

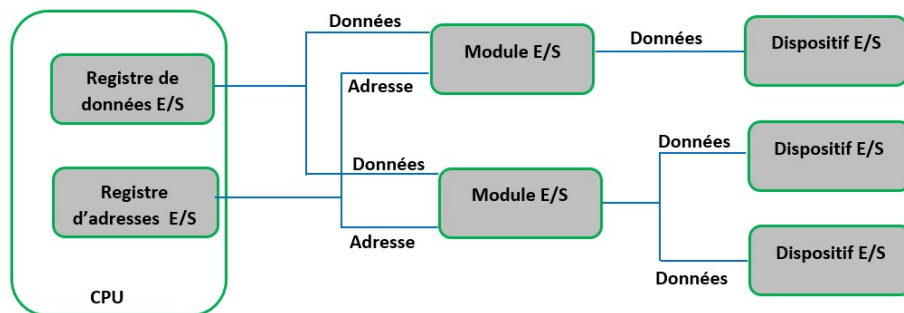


FIGURE 4 – La liaison programmée.

La limite de la liaison programmée est qu'elle est très lente. Ce qui affecte donc les performances du système. De plus, le processeur attend que le périphérique soit libre en exécutant d'autres instructions.

5. Figure inspirée du <https://www.iro.umontreal.ca/~monnier/1215/notes-io.pdf>

4.2 Entrées/sorties pilotées par les interruptions

Les interruptions sont des événements imprévus qui se produisent pendant que le processeur exécute des instructions. Ces interruptions nécessitent donc la réaction du processeur. Pour cela, le processeur doit interrompre temporairement l'exécution de l'instruction en cours pour répondre aux interruptions.

Les interruptions peuvent se produire dans beaucoup de situations comme des entrées imprévues, des situations anormales, des instructions illégales, lors de l'exécution multitâche et multiprocesseurs, etc.

Les périphériques peuvent utiliser le mécanisme des interruptions pour indiquer au processeurs qu'ils sont prêts. Notez que le système d'exploitation est doté d'un ensemble de programmes de gestion des interruptions afin de répondre aux interruptions.

Pour répondre à une interruption, le processeur arrête l'exécution du programme en cours en faveur du programme de gestion d'interruptions.

4.2.1 Gestion des interruptions

La gestion des interruptions est un mécanisme assez compliqué qui expose beaucoup de problèmes. Par exemple :

- Comment l'ordinateur identifie le périphérique qui déclenche l'interruption ?
- Que se passe-t-il si plusieurs interruptions se déclenchent en même temps ?
- Que se passe-t-il si l'ordinateur reçoit une interruption pendant qu'il traite une autre interruption ?
- Existe-t-il des interruptions plus prioritaires sur d'autres ?

Ce sont toutes des questions importantes liées à la gestion des interruptions. Nous allons voir dans ce qui suit, les différentes solutions proposées pour répondre à toutes ces questions.

L'ordinateur peut identifier le périphérique qui déclenche l'interruption par plusieurs méthodes différentes :

- Principe de scrutation : C'est-à-dire, l'ordinateur demande cette information auprès de tous les périphériques, et celui qui a déclenché l'interruption répond positivement à la requête.
- Interruption "vectored" : dans cette méthode, le périphérique peut, après avoir déclenché l'interruption, placer sur le bus de communication l'information d'identification du périphérique. Dans ce cas l'ordinateur peut savoir le périphérique déclenchant l'interruption.

Priorité des interruptions

Il est clair que les interruptions n'ont pas le même niveau d'importance. Les interruptions sont donc organisées selon une structure propre à chaque processeur, dans une pile d'interruptions par exemple. Cette structure est toujours associée à une échelle de priorité. Les interruptions prioritaires sont traitées en premier, comme les exceptions fatales par exemple qui sont généralement plus

graves et qui nécessitent un traitement prioritaire. De la même façon, un périphérique qui déclenche une interruption et qu'il ne peut pas fonctionner sans avoir une réponse de cette interruption est considéré prioritaire.

De cette façon, si une interruption est en cours de traitement, et une autre interruption survient, le processeur met alors en attente l'interruption la moins prioritaire pour traiter l'interruption la plus prioritaire.

Accès direct à la mémoire

Pour que les interruptions soient gérées dans un délais raisonnable, le contrôleur d'entrées/sorties sont dotés d'un mécanisme d'accès direct à la mémoire. Ce mécanisme, implémenté sous forme d'un petit contrôleur dans le contrôleur d'entrées/sorties, s'appelle contrôleur DMA (Direct Memory Access).

Le contrôleur DMA ajoute un certain niveau d'intelligence au contrôleur d'entrées/sorties en prenant le contrôle du bus pour pouvoir envoyer des données directement à la mémoire centrale depuis les périphériques et inversement sans avoir le besoin de passer par le processeur.

Pour pouvoir contrôler les bus d'adresses et de données, le contrôleur DMA est doté d'un registre d'adresses qui relié au bus d'adresses, et un registre de données relié à la fois au périphérique et au bus de données. Le contrôleur DMA inclut également un compteur qui est utilisé pour la gestion des opérations de lecture de données par le contrôleur DMA et l'écriture dans la mémoire principale peut être résumée dans les points suivants :

1. Le processeur envoie l'adresse de la première case mémoire où les données doivent être stockées.
2. L'adresse est stockées dans le registre d'adresses du contrôleur.
3. Le nombre d'octets à transférer est chargé dans le compteur du contrôleur.
4. La commande de lecture des données est envoyée au contrôleur par le processeur via le bus de contrôle.
5. Le contrôleur récupère donc les données du périphérique et les stocke dans une mémoire interne.
6. Le contrôleur procède au transfert des données à la mémoire principale en envoyant la première adresse sur le bus d'adresses avec la première données (premier octet par exemple) accompagné d'un signal d'écriture.
7. Après l'écriture de la première donnée en mémoire, le registre d'adresses est incrémenté pour préparer le transfert de la prochaine donnée.
8. Le compteur à son tour sera décrément après le transfert de la première données. Si le nombre d'octets à transférer est 4, alors la valeur du compteur devient $4-1 = 3$.
9. Si le compteur est nul, le transfert est fini, sinon, la procédure continue avec la prochaine donnée.
10. Une fois le transfert est complété, le contrôleur envoi un signal d'interruption au processeur pour l'aviser de la fin d'opération d'écriture.

La figure 5⁶ résume le fonctionnement du contrôleur DMA.

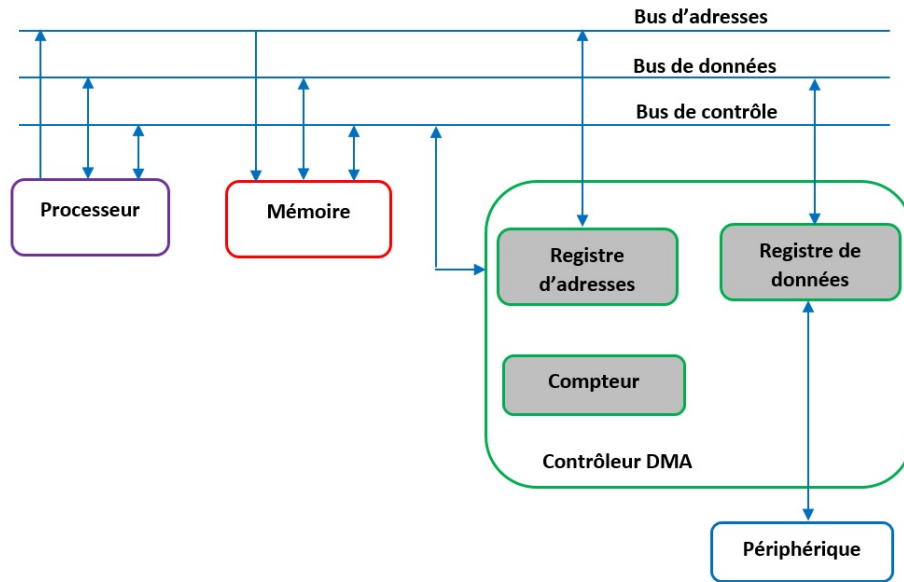


FIGURE 5 – Contrôleur DMA.

5 Caractéristiques des bus

Nous avons vu, au début de ce module, que nous avons différents types de bus chacun véhiculant des signaux représentant des informations particulières comme les données, les adresses et les commandes.

Les bus se distinguent les uns des autres par plusieurs caractéristiques que nous pouvons résumer dans les points suivants :

- La largeur (nombre de bits) des données.
- Le débit : le taux de transfert de données en bits par seconde.
- La longueur du bus.
- La topologie.

5.1 Catégories des bus

Les bus peuvent être parallèles ou série. Chaque catégorie possède des caractéristiques spécifiques qui sont résumées dans les points suivants :

6. Cette figure est tirée et modifiée du livre de l'architecture de l'ordinateur.

Bus parallèles

Dans cette catégorie de bus, le nombre de fils est égal au nombre de bits à transporter. Donc, dans ce cas, les bits sont transférés simultanément. Ces bus sont très fiables pour transporter des données sur des distances courtes. Cependant, ils ne sont pas fiables pour transporter des données sur des longues distances. Nous pouvons utiliser ces bus pour relier le processeur et la mémoire.

Bus série

Dans les bus série, les bits sont transférés les uns à la suite des autres, contrairement aux bus parallèles. Ces bus sont utilisés lors des transferts sur des grandes distances.



La grande majorité des bus à l'intérieur des ordinateurs modernes adoptent la catégorie des bus série.

5.1.1 Ports et pilotes

Les ports sont des connecteurs qui se trouvent à l'extrémité d'un bus dans lequel est connecté un périphérique comme le montre la figure 6.

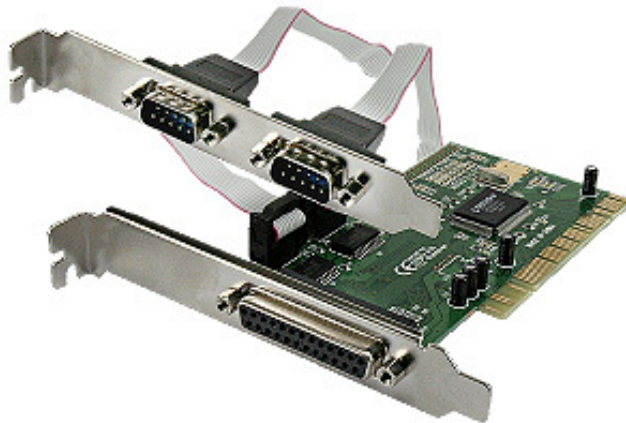


FIGURE 6 – Ports.

Les ports sont généralement contrôlés par un composant qui s'appelle le contrôleur de port. Le contrôleur de port est connecté à un bus de l'ordinateur.

Le contrôleur de port ainsi que les entrées/sorties et les autres périphériques sont contrôlés par un programme (logiciel) qui s'appelle le pilote (driver). Les pilotes peuvent être installés par l'utilisateur comme ils peuvent être inclus dans les systèmes d'exploitation.

5.1.2 Bus d'extensions

Un bus d'extensions permet de connecter des contrôleurs d'extensions, comme les cartes par exemple, à un ordinateur grâce à des connecteurs spéciaux. Ces connecteurs spéciaux peuvent être des entrées (slots) sur la carte mère par exemple. Les bus d'extensions sont appelés également des bus d'entrées/sorties.

Il existe donc plusieurs standards de bus d'extensions que nous pouvons résumer les plus utilisés dans les points suivants :

- **ISA (Industry Standard Architecture)** : c'est l'un des premiers bus utilisés dans les ordinateurs, et c'est le type de bus le plus répandu dans la construction des ordinateurs. Malgré le progrès technologique, le bus ISA, qui passe de la version 8 bits à 16 bits, fonctionne avec des fréquences relativement basses (8 Mhz). Ce bus est utilisé pour relier des cartes relativement lentes comme les modemes, les cartes son, etc.
- **MCA (Micro Channel Architecture)** : il a été introduit par la compagnie IBM. C'est un bus de 32 bits qui atteint un débit maximum de 40 Mo/s. C'est un bus coûteux ce qui l'a rendu moins populaire.
- **EISA (Extended Industry Standard Architecture)** : comme son nom l'indique, c'est une extension du bus ISA. C'est un bus de 32 bits avec une vitesse de 8.33 MHz.
- **PCI (Peripheral Component Interconnect)** : le bus PCI est introduit pour faire face à la vitesse limitée des autres bus. Le bus PCI de 32 bits (jusqu'à 64 bits) fonctionne avec des fréquences relativement grandes 33 Mhz (66 Mhz dans le cas de 64 bits). La version de 32 bits est la plus répandue. Ce bus est utilisé pour connecter des cartes son, cartes graphiques (anciennes cartes), cartes réseaux, etc. Le bus PCI peut atteindre un débit de 132 Mo/s (32 bits) et jusqu'à 264 Mo/s (64 bits).
- **PCI-E (Peripheral Component Interconnect Express)** : est un standard développé par Intel en 2004, c'est une évolution du bus PCI. Il peut atteindre un débit de 500 Mo/s de performance dans des nouvelles versions.
- **USB (Universal Serial Bus)** : les bus USB sont développés pour simplifier les interfaces d'accès avec de nombreux périphériques. Il existe plusieurs versions du bus USB comme USB-1 (12 Mo/s), USB-2 (480 Mo/s) et USB-3 (5 Go/s et même plus).

Les directions de transmission dans les bus peuvent être de type :

- **Simplex** : transmission unidirectionnelle.
- **Half duplex** : transmission bidirectionnelle non simultanée.
- **Full duplex** : transmission bidirectionnelle simultanée.