

Entrées-Sorties: Architectures de bus



GIF-1001 Ordinateurs: Structure et Applications, Hiver 2015
Jean-François Lalonde

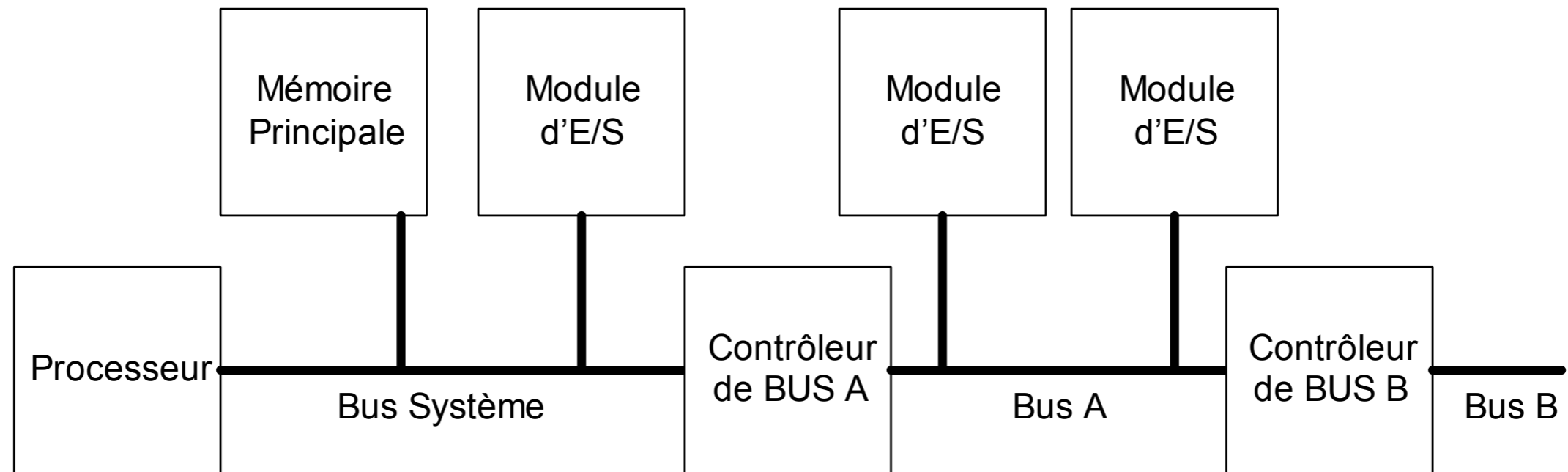
Architectures CPU/Mémoire/E-S

- Un ordinateur a cinq composantes de base:
 - le(s) processeur(s)
 - la (les) mémoire(s)
 - les modules entrées/sorties
 - les périphériques
 - les bus
- Il existe une multitude de façon d'agencer tous ces éléments entre eux. Chacune de ces façons est une architecture qui a ses avantages et ses inconvénients.
- Deux architectures sont utilisées de nos jours
 - **l'architecture de bus** (avec hiérarchie de bus, bus I/O)
 - **l'architecture de canaux** (Channel I/O).
- L'architecture de bus se retrouve habituellement dans vos PCs. L'architecture de canaux est principalement utilisée par IBM pour les serveurs.

Architecture de Bus

- Rappel: Un bus est un ensemble de lignes électriques.
- Dans une architecture de bus, les modules d'E/S sont connectés sur des bus, reliés à d'autres bus par des contrôleurs de bus (souvent appelés interfaces ou ponts). Le processeur et la mémoire sont eux-mêmes sur des contrôleurs de bus.
- Les contrôleurs de bus et les lignes de bus doivent supporter les entrées-sorties programmées, par interruptions, ou par DMA.
- Il existe une hiérarchie de bus. Plusieurs bus de plus en plus lents vont du processeur aux E/S principalement parce que:
 - les E/S ont des taux de transfert de données différents;
 - les bus rapides doivent être courts afin de limiter les délais de propagation et les pertes d'énergie;
 - il ne faut pas surcharger un bus d'appareils.

Architecture de Bus



Description des bus

- Les bus peuvent être différents, mais ils ont tous:
 - une largeur (un nombre de lignes!) comprenant un nombre de lignes de données et d'adresses;
 - une vitesse de communication;
 - un type de connecteur;
 - et un protocole qui décrit la façon dont sont échangées les données sur le bus.

Quelques définitions

- **Ligne dédiée**: ligne de bus réservée à une fonction unique.
- **Ligne multiplexée**: ligne de bus qui peut avoir plusieurs rôles (ex: multiplexage temporel)
- **Méthode d'arbitrage**: la méthode d'arbitrage détermine qui contrôle le bus. Dans les systèmes simples que nous avons vus, il n'y avait que le CPU qui pouvait contrôler le bus et l'arbitrage était tout simple: le CPU contrôle toujours le bus! Toutefois, lors de DMA ou lorsqu'il y a plusieurs CPU, il faut une convention pour savoir qui contrôle le bus: c'est la méthode d'arbitrage. Il y en a deux types:
 - centralisée (un arbitre décide qui contrôle le bus);
 - distribuée (les contrôleurs de bus s'échangent le contrôle à tour de rôle).

Quelques définitions

- **Bus synchrone ou asynchrone:** dans un bus synchrone, une ligne du bus sert d'horloge. Dans un bus asynchrone, la vitesse de communication est préétablie et l'occurrence d'un évènement dépend des évènements survenus auparavant.
- **Point à point ou multipoint:** un bus point à point est un bus sur lequel deux appareils seulement sont connectés. Il y aura plusieurs appareils sur un bus multipoint.
- **Bloc ou Données en rafales (burst):** les données peuvent être échangées par blocs de taille fixes ou en blocs de tailles variables, en rafale.
- **Protocole Série ou parallèle** (ne pas confondre avec port série et port parallèle): un bus série est un bus où les bits sont transmis sur une seule ligne, un après l'autre. Sur un bus parallèle, plusieurs bits sont transmis simultanément.

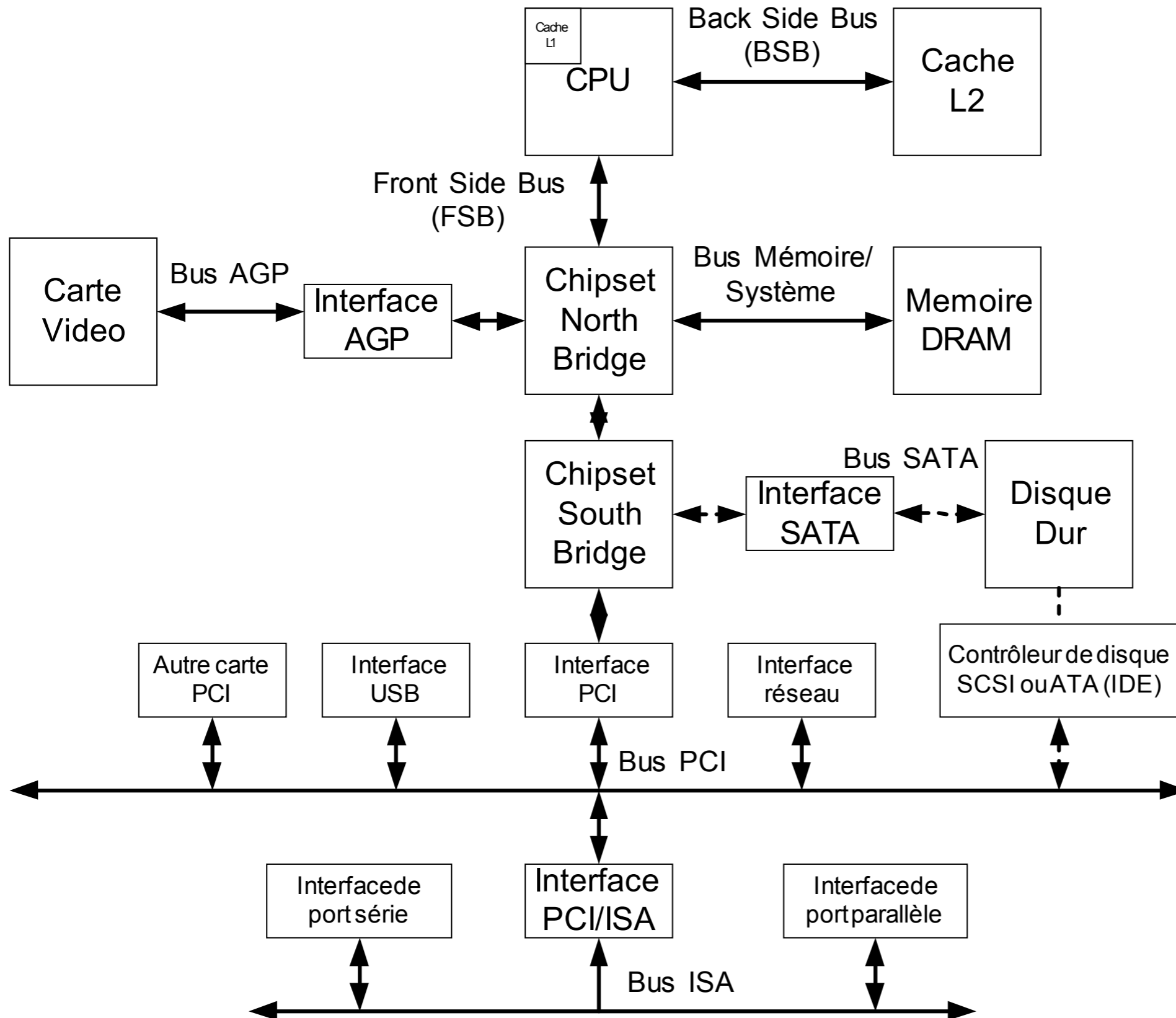
Vitesse de communication

- Pour tous les bus existants, une vitesse de communication maximale est toujours annoncée.
- Certaines vitesses sont en bits/secondes (bps). D'autres sont en octets ou bytes par secondes.
 - La vitesse en bps est 8 fois plus grande que celle en bytes!
 - Laquelle est la meilleure vendeuse?
- Ce chiffre est toujours supérieur à la vitesse maximale réelle à laquelle peut se transférer les données d'un usager!
- Pourquoi?

Vitesse effective vs vitesse maximale

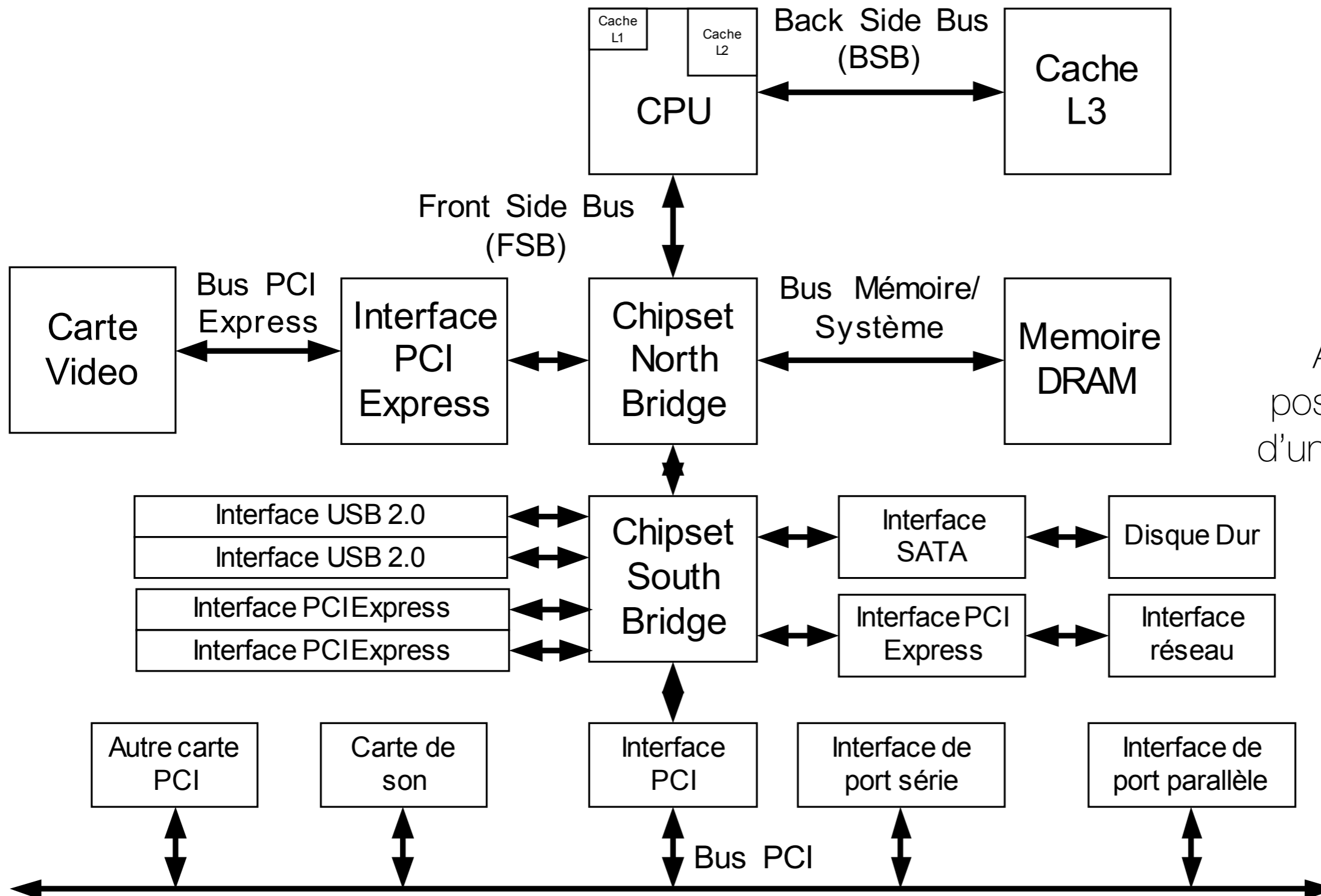
- Pourquoi la vitesse “effective” est inférieure à la vitesse maximale?
 - Des pertes de temps surviennent afin d’arbitrer les bus.
 - Des bits peuvent être ajoutés aux données brutes afin de:
 - détecter et/ou corriger des erreurs.
 - d’assurer la synchronisation des messages
 - d’éviter de longues séquences de 1 ou 0
 - de limiter la puissance des harmoniques secondaires du signal transmis.
 - Il y a toujours un temps de latence au début d’une communication entre deux appareils.
 - Il est possible et même très probable que la communication ne puisse pas se faire à vitesse maximum pendant un intervalle de temps très long.
 - Les données des usagers sont presque toujours encapsulées dans plusieurs couches de protocole de communication. Les octets de protocole ajoutés ont plusieurs fonctions comme identifier la source et la destination du message ou donner un numéro de référence au message.

Les Bus de votre Ordinateur (1/4)



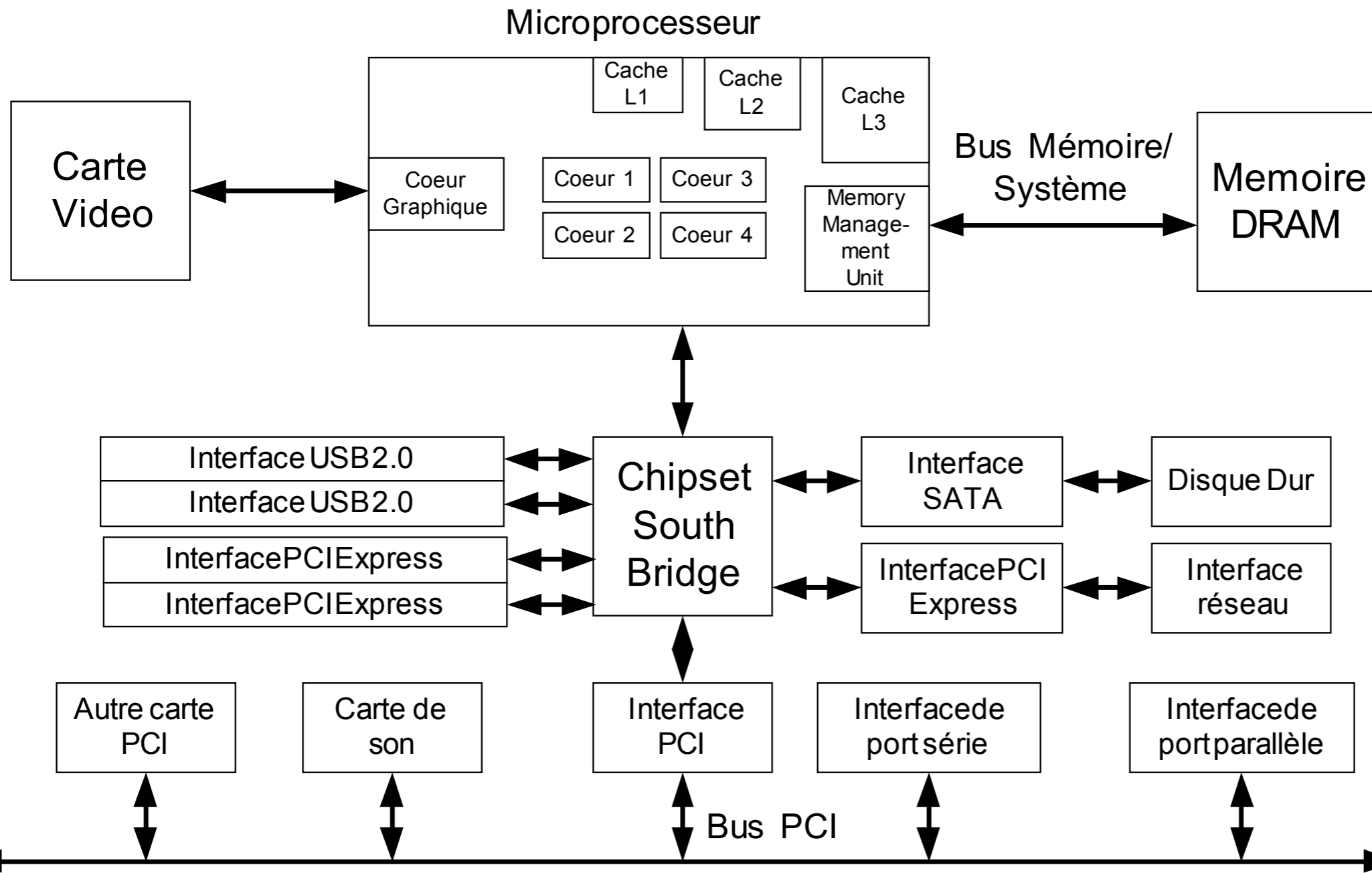
Architecture possible des bus d'un PC acheté en 2003

Les Bus de votre Ordinateur (2/4)



Architecture possible des bus d'un PC acheté en 2007

Les Bus de votre Ordinateur (3/4)



Architecture possible des bus d'un PC acheté en 2011

Les Bus de votre Ordinateur

- Habituellement, le CPU communique avec ses caches sur un bus dédié, le back side bus (BSB). Il communique également avec la mémoire, la carte video et les E/S à partir du front side bus (FSB). Le chipset, du moins la partie nord, est responsable de router les échanges entre la mémoire et le CPU, la mémoire et la carte vidéo et le CPU et la carte vidéo. De larges quantités de données sont rapidement échangées dans cette partie de l'ordinateur, habituellement concentrée sur la carte mère. Le chipset est également responsable de la communication entre le CPU (ou la mémoire dans le cas de DMA) et les autres E/S par sa partie sud (pont ou bridge). Ces échanges sont plus lents et moins volumineux que dans la partie nord.
- Le North Bridge et même un processeur graphique sont parfois intégrés dans les microprocesseurs modernes au même titre que la cache L3.
- Le bus PCI ou le bus PCI express est le principal bus sur lequel se connecte tous les modules d'E/S. Il s'y connecte habituellement le disque dur, les cartes réseaux, le port USB et d'autres bus plus lents, s'il y en a. Dans les figures précédentes, on retrouve le clavier et la souris sur un bus ISA, plus lent que le bus PCI auquel il est connecté.
- Les figures montrant l'architecture typique d'un PC ne sont que des exemples qui ne s'appliquent pas à tous les ordinateurs! Ces figures ne s'appliquent généralement pas aux serveurs, ni aux ordinateurs ayant deux coeurs, ni aux ordinateurs n'ayant pas de bus ISA, ni aux ordinateurs n'ayant pas une cache externe, ni aux ordinateurs...

Améliorations des bus

- La majorité des bus à l'intérieur des ordinateurs modernes adoptent des protocoles de type série
- Les bus récents sont des bus avec plusieurs liens point à point plutôt que multi-points.
- Les bits transmis sont encodés en mode différentiel.

Améliorations des bus

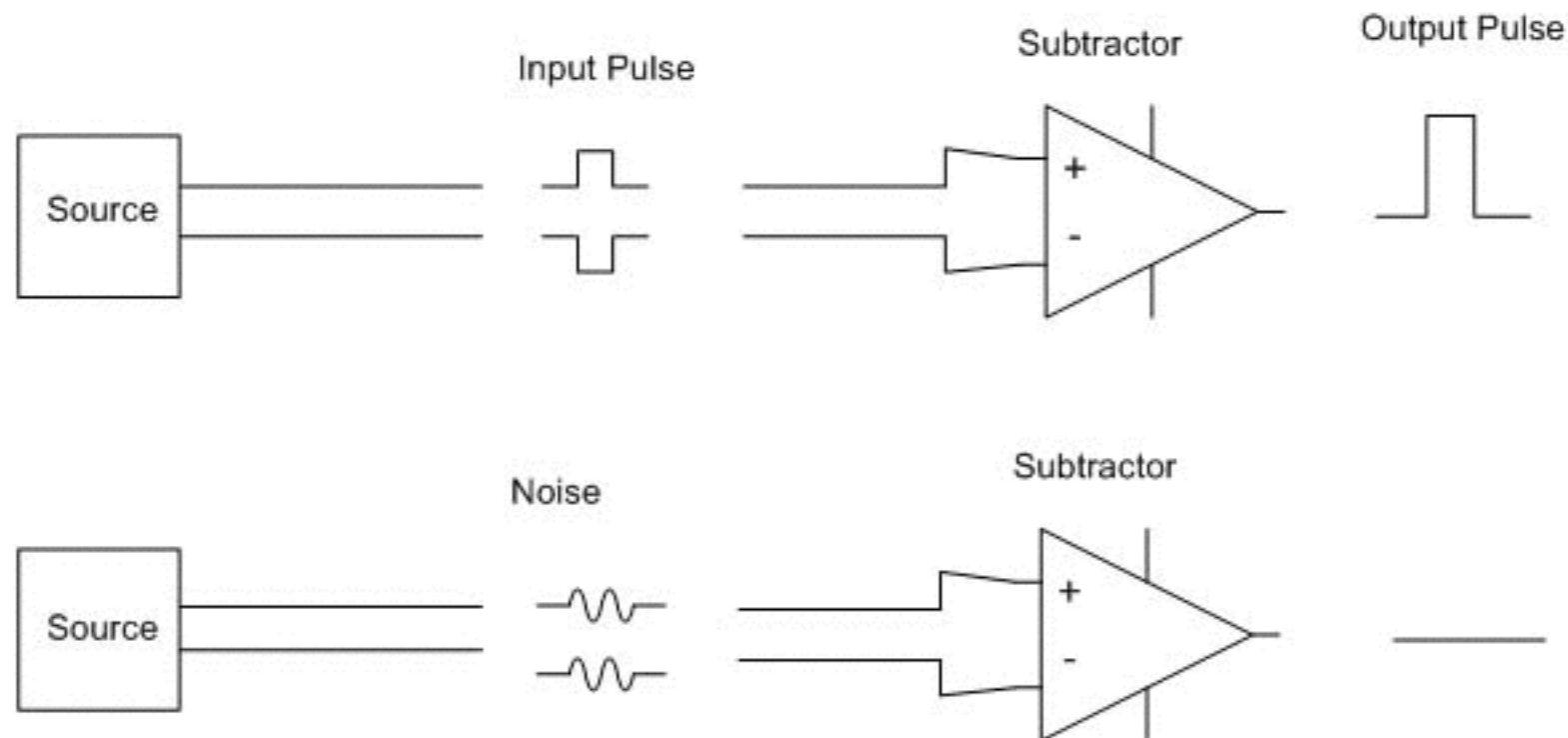
- La majorité des bus à l'intérieur des ordinateurs modernes adoptent des protocoles de type série. Pourquoi?
 - À très haute fréquence, les délais de transmission d'un signal électrique dans un fil court peuvent devenir considérable par rapport à la période d'un bit.
 - Lorsque plusieurs traces (fils électriques sur des PCBs) sont en parallèle, elles ne peuvent pas avoir la même longueur pour des raisons purement mécaniques.
 - Des signaux électriques qui partent en même temps d'un transmetteur n'atteignent pas en même temps le receveur si les fils qui propagent les signaux n'ont pas la même longueur. À haute fréquence, il est possible que des bits d'un fil arrivent décalés par rapport au bits des autres fils. De ce fait, un bus parallèle est peu envisageable pour des fréquences supérieures à 1 GHz.

Améliorations des bus

- Les bus récents sont des bus avec plusieurs liens point à point plutôt que multi-points. Pourquoi?
 - Si l'appareil A est relié à l'appareil B et que l'appareil B est reliée à l'appareil C, les messages de l'appareil A à l'appareil C seront reçus et retransmis par l'appareil B.
 - La bande passante d'un bus multipoint doit être répartie entre tous les appareils connectés sur le bus. Avec plusieurs liens point à point, la communication entre plusieurs paires d'appareils peut se faire simultanément.

Améliorations des bus

- Les bits transmis sont encodés en mode différentiel. Pourquoi?
 - La différence de tension entre deux signaux propagés sur deux lignes différentes détermine la valeur d'un bit transmis. Des symboles différents sont transmis si la différence est positive ou négative.
 - Le bruit commun sur les deux lignes propageant le signal est éliminé lorsque la différence est effectuée. Très robuste.
 - Lorsque la différence est nulle, le bit est invalide ou une autre information peut être transmise.



Quelques types de bus internes

- ISA
- PCI, PCI express
- ATA et SATA
- AGP

BUS PCI

- Le Bus PCI (Peripheral Component Interface) a été créé au début des années 1990 par Intel. Il a été rapidement mis dans le domaine public (Open Source) afin d'en permettre la diffusion à large échelle. Depuis, il s'agit d'un standard qui est/sera remplacé par le PCI Express.
- Le bus PCI a 32 lignes d'adresses et 32 ou 64 lignes de données. Pour avoir 64 bits de données, du multiplexage temporel est fait avec les 32 lignes d'adresses. Incluant ces lignes de données/adresses, le bus PCI a 120 lignes.
- Le bus PCI a une horloge de 66MHz (version 2.2), indépendante de l'horloge du CPU. Les données y sont transférées de façon synchrone à un débit pouvant aller jusqu'à $66\text{MHz} \times 64\text{bits}$ en parallèle, soit 4.224Gbit/s. Notez que la version originale du PCI avait une horloge de 33MHz et qu'il existe des versions à 133Mhz (PCI-X).
- Il y a un arbitre central sur le BUS PCI déterminant qui contrôle bus. Il est possible de connecter plusieurs processeur à un seul bus PCI. Il est également possible de faire du DMA.
- Le bus PCI supporte les interruptions. Il a quatre lignes pour y connecter des contrôleurs d'interruptions et supporter le transfert d'Interrupt Acknowledge du CPU vers un périphérique.
- Les données sont transmises par bloc sur le bus PCI. Elles sont emmagasinées dans une cache qui permet d'adapter la vitesse des modules d'E/S à celle du CPU.
- Il existe des chips simples et peu coûteuses permettant d'attacher d'autres bus comme le bus ISA au bus PCI. Le bus a été conçu en tenant compte des autres bus existants.
- Pour plus d'information sur le bus PCI ou sur le PCI-X, lire:
 - <http://www.pcstats.com/articleview.cfm?articleID=1087>

BUS PCI EXPRESS

- Le bus PCI EXPRESS a été créé par Intel en 2004 afin de remplacer le bus AGP et le bus PCI trop encombré (même sans la carte graphique, le bus PCI n'est pas capable de fournir une carte Ethernet Gigabyte, des périphériques USB2.0, du FireWire et plusieurs autres périphériques).
 - De nos jours, le PCI Express a remplacé l'AGP sur presque toutes les cartes mères, mais le PCI demeure pour les périphériques ayant peu de besoins en terme de bande passante (peu de données à échanger par unité de temps).
- Le bus PCI Express est un bus série transmettant en mode différentiel à 5Gbps. Il est possible de doubler, quadrupler la paire de fils utilisés pour passer à 10Gbps ou 20Gbps. Il est aussi possible de multiplier le nombre de fils par 8 ou 16...
- Les bytes de données de l'utilisateur sont codés sur 10 bits. 2 bits sont ajoutés essentiellement afin de détecter les erreurs. On parle d'encodage 8b/10b. Le PCI Express a aussi de nombreuses couches de communication qui ajoute un en-tête aux données de l'utilisateur.
- Le bus PCI Express est Point à point. Plusieurs connections point à point étant potentiellement disponibles à partir du Chipset (ou du microprocesseur directement).
- Pour plus de détail sur ce bus, voir des sites internet comme:
 - http://www.dell.com/content/topics/global.aspx/vectors/en/2004_pcieexpress?c=us&l=en&s=gen
 - www.intel.com/technology/pciexpress/devnet/docs/WhatIsPCIExpress.pdf

BUS ISA, AT, EISA, et VLBUS

- Le BUS ISA (Industry Standard Architecture) a été développé par IBM dans les années 1980 à 1981. Ce bus 8 bits (nombre de lignes de données!) devait fournir un bus plus puissant que les quelques bits disponibles sur le port parallèle des PCs. Toutefois, il a été remplacé rapidement, en 1984, par une version 16bits, le bus AT. Le bus EISA (Extended ISA), également une version plus récente du bus ISA, est sur 32 bits.
- Le VLBUS (Video Electronics Standard Association Local BUS) a été créé dans les années 1990 pour supporter des cartes graphiques plus performantes.
- Ces bus sont désuets, car ils ont tous été supplantés par le bus PCI. Seules quelques vieilles interfaces pratiques demeurent sur ces bus (port série et port parallèle sur le bus ISA!).

BUS ATA (IDE), SCSI et SATA

- Les bus ATA (IDE), SCSI et SATA ont été créés afin de communiquer avec des disques durs.
- Le bus IDE (Integrated Disk Electronics) est le plus vieux, le moins performant et... le moins cher. Il s'agit du câble à ruban plat de 40 pins que l'on retrouve encore dans certaines machines.
- Le bus ATA (Advanced Technology Attachment) est du même type que le bus IDE, mais en plus performant. Il utilise le même câble, mais à de plus hautes vitesses. Il existe plusieurs versions d'ATA, ATA-1 étant la même chose qu'IDE.
- Le bus SCSI (Small/Smart Computer Standard Interface) a été créé afin de suppléer aux limitations du bus IDE (nombre de disques limité, peu de support pour le RAID, etc.).
- Le bus SATA (Serial Advanced Technology Attachment) utilise un protocole série, contrairement au ATA ou au SCSI. Il est sorti en 2003-2004.
- Vous pourrez trouver davantage d'information sur chacun de ces bus sur internet. Voici deux liens à titre d'exemples:
 - <http://sophia.dtp.fmph.uniba.sk/pchardware/idescsi.html>
 - http://www.interfacebus.com/Design_Connector_Serial_ATA.html

BUS AGP

- Le bus AGP (Accelerated Graphics Port) a été créé par Intel en 1996 afin d'améliorer les vitesses de communication avec les cartes graphiques. Les transferts de données nécessaires aux besoins graphiques modernes ne peuvent pas être supportés par le bus PCI qui est trop lent et trop encombré (presque tous les périphériques sont connectés au bus PCI!!!).
- Il n'y a qu'un seul périphérique connecté sur le bus AGP: la carte graphique!
- Pour plus de détail sur ce bus, voir des sites internet comme <http://www.hardwaresecrets.com/article/155>.

Interfaces externes et périphériques

- En plus de tous les bus présentés dans ce cours, il existe plusieurs bus qui sortent de l'ordinateurs. Ces bus sont appelées interfaces externes. En voici une liste non exhaustive qui sera le sujet du prochain cours: ports parallèle et série, port PS/2 (clavier et souris), USB, FireWire, Ethernet, InfiniBand...

Références et exercices

- Références
 - Irv Englander: Chapitres 9, 11.1 et 11.2
 - Stallings: Chapitres 3.3, 3.4, 7