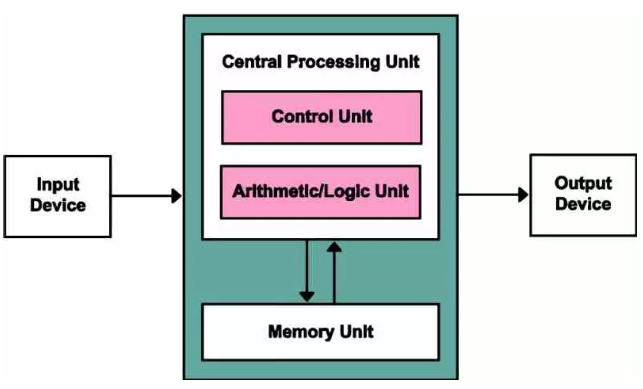
**Von Neumann, la arquitectura común de todos los procesadores**

La arquitectura Von Neumann es la arquitectura común de todos los procesadores para PC. Todas y cada una de las CPU, desde ARM a x86, desde 8086 al Ryzen pasando por los Pentium. Todas ellas son arquitecturas Von Neumann y todas ellas heredan cierto problema común.

John Von Neumann fue un matemático de origen húngaro que es famoso por dos cosas. La primera es por haber trabajado en el Proyecto Manhattan donde se desarrolló la bomba atómica que los Estados Unidos lanzo contra Japón a finales de la Segunda Guerra Mundial. La segunda es el desarrollo de la arquitectura base que hoy en día utilizan nuestros PC sean del tamaño que sean, así como también la forma de los programas que estos ejecutan.

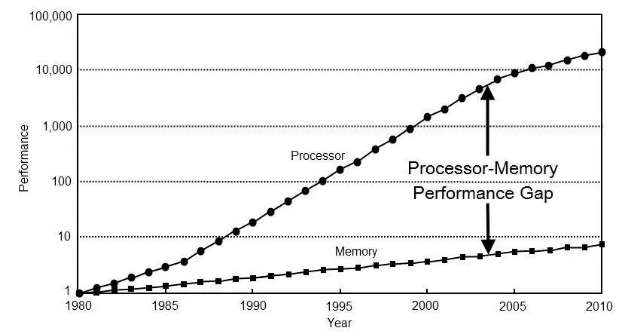
**¿Qué es la arquitectura Von Neumann?**



La arquitectura Von Neumann es en la que se basan todos los procesadores para PC, ya que todos ellos están organizados con una serie de componentes comunes, los cuales son los siguientes:

* **Unidad de Control:** Encargada de las etapas de captación y descodificación del ciclo de instrucción.
* **Unidad lógico-aritmética o ALU:** Encargada de realizar las operaciones matemáticas y de lógica que requieren los programas.
* **Memoria:** La memoria en la que se almacena el programa, la cual la conocemos como memoria RAM
* **Dispositivo de entrada:** Desde el que nos comunicamos con el ordenador.
* **Dispositivo de Salida:** Desde el que el ordenador se comunica con nosotros.

Como podéis ver se trata de la arquitectura común en todos los procesadores y es por ello que no tiene más secreto, pero existe otro tipo de arquitectura conocida como arquitectura Harvard en el que la memoria RAM se encuentra dividida en dos pozos distintos, en uno de ellos se almacenan las instrucciones del programa y en la otra memoria los datos, teniendo buses separados tanto para el direccionamiento de la memoria como para las instrucciones.

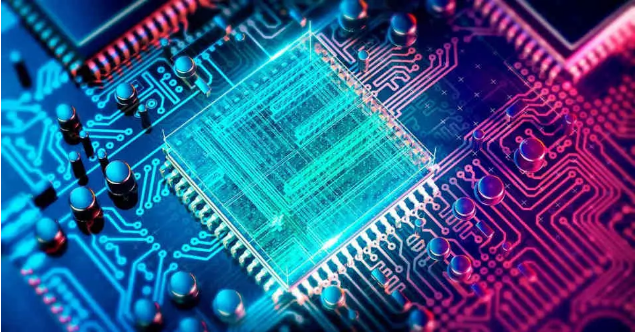
**¿Cuáles son las limitaciones de la arquitectura Von Neumann?  
**

La principal desventaja es que la memoria RAM que es donde se encuentran las instrucciones y los datos que han de ser procesados se encuentran unificados y compartidos a través un mismo bus de datos y direccionamiento común. Por lo que las instrucciones y los datos han de ser captados de manera secuencial desde la memoria. Este cuello de botella es el llamado cuello de botella de Von Neumann. Es por ello que los diferentes microprocesadores tienen la caché más cercana al procesador dividida en dos tipos, una para datos y otra para instrucciones.

En los años recientes, las velocidades de los procesadores ha ido aumentando de manera mucho más rápida que la memoria RAM, por lo que se ha aumentado el tiempo en que los datos tardan en ser comunicados desde hacía la memoria. Lo que ha obligado a desarrollar soluciones para paliar este problema, producto del cuello de botella de Von Neumann.

En los procesadores donde normalmente se utiliza la arquitectura Harvard es en los que estos se encuentran autocontenidos y por tanto no tienen acceso a la memoria RAM en común del sistema, sino que ejecutan su propia memoria y programa de manera aislada a la CPU principal. Estos procesadores reciben la lista de datos e instrucciones en dos ramales de datos distintos. Uno para la memoria de instrucciones y el otro para la memoria de datos de dicho procesador.

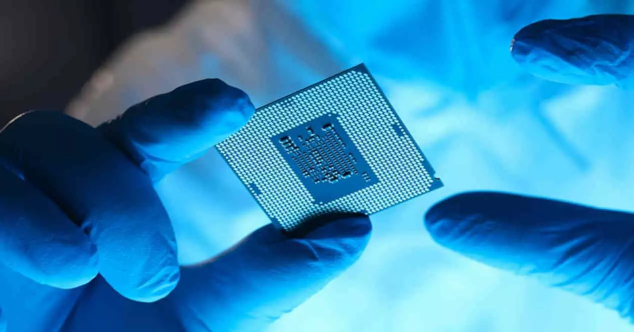
## ¿Por qué es la utilizada en CPU y GPU?



El motivo principal es el hecho que aumentar la cantidad de buses significa aumentar el perímetro del propio procesador, ya que para comunicarse con la memoria externa es necesario que la interfaz se encuentre en la parte exterior del mismo. Esto lleva a procesadores mucho más grandes y mucho más caros. Por lo que el principal motivo por el cual la arquitectura Von Neumann se ha estandarizado es por los costes.

El segundo motivo es que se necesita que los dos pozos de memoria estén sincronizados para que un una instrucción no se aplique a un dato erróneo. Lo que lleva a tener que crear sistemas de coordinación entre ambos pozos de memoria. Eso si, buena parte de los cuellos de botella se eliminarían al separar ambos buses. Pero tampoco reduciría del todo el cuello de botella de Von Neumann.

Esto se debe a que el cuello de botella de Von Neumann pese a ser una consecuencia del almacenamiento de datos e instrucciones en una misma memoria también se puede dar en una arquitectura Harvard si esta no es lo suficientemente rápida como para alimentar al procesador. Es por ello que las arquitecturas Harvard se ha reducido en especial a microcontroladores y DSP. Mientras que Von Neumann es común en CPU y GPU  
  
La caché de primer nivel, el escenario en el que se rompe el modelo



Habréis observado que tanto en CPU como en GPU la caché de nivel más bajo está siempre dividida en instrucciones y datos, lo cual rompe la definición de arquitectura de lo que es una arquitectura Von Neumann. El motivo de hacer esto es muy simple y es que esto tiene que ver con el objetivo de facilitar la descodificación de las instrucciones, por lo que habitualmente cuando se copia de la caché del segundo nivel más bajo al más bajo lo que se hace es separar el dato de la instrucción.

Esto lo podemos ver tanto en CPU como en GPU, y todo esto sirve para tener una unidad de descodificación de las instrucciones más simple en la que los primeros bits de cada una son separados e interpretados. Además esto permite la implementación de instrucciones del tipo vectorial o SIMD de manera más sencilla en el procesador, las cuales son clave en el caso de los procesadores gráficos.

Ya como resumen y para terminar, entender como funciona la arquitectura Von Neumann significa entender como funciona cualquier procesador, ya que todos se basan en estos principios. Es como comparar un motor Ferrari con un McLaren, ambos son de marcas distintas, pero se basan en una serie de principios y normas comunes y con el mundo de los procesadores ocurre lo mismo.