



El bus PCI-Express

Javier Valcarce García, javier.valcarce@gmail.com

Instrumentación Electrónica
curso 2005/2006

Resumen

La aparición de nuevos dispositivos que exigen cada vez más velocidad de E/S ha puesto de manifiesto la obsolescencia del bus PCI. Aunque el estándar ha sufrido sucesivas revisiones con el fin de aumentar su velocidad, no es posible aumentarla ya más usando un bus de acceso múltiple y transmisión paralela.

Una solución mejor y más flexible al problema es usar una red conmutada (enlaces punto a punto + conmutadores) y transmisión serie en los enlaces. El nuevo bus PCI-Express representa este nuevo enfoque a un coste razonable manteniendo compatibilidad software con el bus PCI anterior. Ofrece además características muy interesantes como velocidad escalable en los enlaces, QoS (Quality of Service) y muchas más cosas.

En el primer apartado de este trabajo de clase se resumen las características más importantes del bus PCI-Express, el segundo apartado examina la norma “a vista de pájaro” en su versión 1.1 y, por último, en el tercer apartado se muestra la situación actual de la tecnología en el mercado.

(PÁGINA EN BLANCO)

Índice de contenido

1. Introducción al bus PCI-Express.....	5
1.1 Enlace básico x1.....	6
1.2 Especificaciones mecánicas.....	7
1.3 Arquitectura de red.....	9
1.4 Calidad de servicio (QoS).....	12
1.5 Comparación con otros buses.....	13
2. La norma “PCI-Express Base Specification Revisión 1.1”.....	16
2.1 Nivel de transacción.....	17
2.2 Nivel de enlace.....	19
2.3 Nivel físico.....	20
2.3.1 Bloque lógico.....	20
2.3.2 Bloque físico.....	21
3. PCI-Express en el mercado.....	23
3.1 Estaciones de trabajo.....	23
3.2 Ordenadores portátiles. ExpressCard.....	26
3.3 Instrumentación PCI-Express y PXI-Express.....	27

Índice de tablas

Tabla 1: La velocidad del enlace es escalable incrementando el número de lanes.....	6
Tabla 2: Interoperabilidad entre tarjetas y conectores (slots).....	7
Tabla 3: Pines del conector x1 (36 pines).....	8

Índice de ilustraciones

Ilustración 1: Enlace básico de la red PCI-Express (1 sólo lane) punto a punto. Un LANE son dos pares de hilos (cada flecha gruesa es un par), uno de TX y otro de RX. La información se transmite en serie y en paquetes (tramas).....	7
Ilustración 2: Diferentes anchos de conector en función del número de lanes.....	7
Ilustración 3: Conector x1 (36 pines).....	8
Ilustración 4: Conector x16 (164 pines).....	8
Ilustración 5: Esquema eléctrico un enlace básico PCI-Express, la longitud del cable en general debe ser menor de 50.8cm.....	9
Ilustración 6: Ancho de banda/pin de diferentes buses.....	10
Ilustración 7: Red punto a punto (PCI-Express).....	10
Ilustración 8: Red de acceso múltiple (PCI).....	10
Ilustración 9: Topología típica de un bus PCI-Express conectado al módulo de mayor nivel en la jerarquía de E/S del sistema (Root Complex).....	11
Ilustración 10: Detalle de un conmutador PCI-Express.....	11
Ilustración 11: Arquitectura en niveles o capas del bus PCI-Express.....	12
Ilustración 12: Encapsulación (tunneling) de tramas en cada nivel.....	13
Ilustración 13: Tipos de tráfico de E/S.....	13
Ilustración 14: En PCI-Express, la velocidad de transmisión no se comparte (a diferencia de lo que	

ocurre en PCI).....	14
Ilustración 15: Ámbito de aplicación de los buses PCIe, HiperTransport, e Infiniband.....	14
Ilustración 16: Trama HT vs PCI-Express.....	15
Ilustración 17: Evolución de los buses del PC, de ethernet y USB. PCI-Express se presenta en el mercado como bus PCI para la próxima década.....	15
Ilustración 18: Velocidad vs latencia de PCI-Express respecto a otros buses y redes de comunicaciones.....	16
Ilustración 19: Niveles de la arquitectura de red.....	17
Ilustración 20: Cabeceras añadidas en cada nivel.....	17
Ilustración 21: Formato genérico de las tramas TLP.....	18
Ilustración 22: Dirección de 32 bits asignada a la tarjeta destino, usada por el conmutador para encaminar el TLP.....	18
Ilustración 23: Concepto de VC (Virtual Channel).....	20
Ilustración 24: Disciplina de cola para los distintos VCs.....	20
Ilustración 25: Trama DLLP para el control y monitorización del enlace.....	21
Ilustración 26: Circuito para calcular el CRC.....	21
Ilustración 27: Código bloque 8b/10b.....	22
Ilustración 28: Orden de transmisión para un conector x1.....	23
Ilustración 29: Registro LFSR usado como scrambler.....	23
Ilustración 30: Señal LVDS correspondiente a la secuencia 1001000011 de izquierda a derecha (aplicando -3.5dB de deénfasis en el canal).....	24
Ilustración 31: Restricciones sobre el diagrama de ojo que tiene que cumplir la señal transmitida (diferencial).....	24
Ilustración 32: Arquitectura del PC actual.....	26
Ilustración 33: Arquitectura del PC en el futuro. Un conmutador PCI-Express reemplaza al SouthBridge por completo.	26
Ilustración 34: Arquitectura de transición del PC para el año 2005 y siguientes. Coexisten PCI-Express, PCI y otras interfaces todas unidas al SouthBridge.....	27
Ilustración 35: Tarjeta gráfica nVidia PCI-Express (x16) XFX 7900GT PCX DDR3 HDTV, con 256MB de memoria de vídeo y salida digital DVI y HDTV (335€).....	27
Ilustración 36: Slots PCI-Express en azul junto a los viejos slots PCI en blanco, la placa base es de formato BTX.....	27
Ilustración 37: Conector hembra y macho de una tarjeta ExpressCard.....	28
Ilustración 38: Tamaño (form factor) de las tarjetas ExpressCard (segunda y tercera por la izquierda) comparadas con una antigua tarjeta CardBus (la primera por la izquierda).....	29
Ilustración 39: Chasis PXI con el controlador (tarjeta más a la izquierda) y las tarjetas (osciloscopios, multímetros, etc).....	29
Ilustración 40: Backplane de un chasis preparado para PXI-Express. Se pueden insertar tanto tarjetas PXI como PXI-Express.....	31
Ilustración 41: Dos tarjetas con distinto conector (PCI-Express y PXI-Express) y un mismo software para las dos.....	31
Ilustración 42: Frame grabber NI PCIe-1429 captura la señal de cualquier cámara que cumpla la norma CameraLink (2800€).....	32
Ilustración 43: Tarjeta de adquisición NI PCIe-6251 con conector x1 (250MB/s), 16 entradas y 4 salidas analógicas a 16bits 1.25 MSample/s.....	32
Ilustración 44: Chasis PCI-Express NI PXIe-1062Q.....	32

1. Introducción al bus PCI-Express

PCI-Express¹ es la evolución del bus PCI, es una reformulación radicalmente distinta de este aunque mantiene totalmente la compatibilidad software. PCI-Express es un bus, i.e, una red de comunicaciones de alta velocidad en distancias cortas y con baja latencia para interconectar dispositivos y tarjetas entre sí dentro de un chasis. Sus características más importantes son:

1. Es una red punto a punto de conmutación de paquetes². La red se construye mediante enlaces punto a punto full-dúplex + conmutadores (). Cada enlace está compuesto por *lanes*. Cada *lane* es full-dúplex y consiste en dos pares de hilos, un par de transmisión y otro de recepción como muestra la Ilustración 1. Cada par es un canal simplex en el que la señal de tensión transmitida es diferencial (LVDS - Low Voltage Differential Signaling), lo que mejora la inmunidad al ruido
2. La arquitectura de la red está dividida en capas. El transceptor PCIe implementa en hardware los 3 niveles inferiores del modelo OSI. Estos niveles son: Nivel de transacción (Transaction Layer), Nivel de enlace (Link layer), Nivel físico (Physical Layer) y se explican en el Apartado 2
3. La transmisión es serie³. En cada *lane* la transmisión de bits en ambos sentidos es en serie lo que evita el problema del jitter (el desfase entre las señales que viajan en paralelo a través del bus) en el receptor, que se produce siempre en alta velocidad
4. La velocidad del bus es escalable añadiendo más *lanes* al enlace. Cada *lane* proporciona (especificación inicial) una velocidad bruta⁴ de 250MB/s/sentido. La norma admite 1 (x1), 2 (x2), 4 (x4), 8 (x8), 16 (x16) y 32 (x32) *lanes* por enlace. El ancho del conector varía en función de este número como muestra la Ilustración 2
5. Calidad de servicio (QoS). Es capaz de diferenciar distintos tipos de tráfico⁵ y de ofrecer cierto nivel de calidad de servicio (QoS – Quality of Service) como se explicará más adelante
6. Compatibilidad software con PCI. La interfaz software del bus (el “driver model”) permanece igual, no hay que cambiar los drivers del sistema operativo ni el software de sistema, todo queda exactamente igual. Se trata por tanto de una evolución y no de una revolución, algo lógico teniendo en cuenta la inercia del mercado.
7. Conexión y cambio de tarjetas en caliente. Especialmente importante en servidores.

1 Abreviado como PCIe, que no tiene nada que ver con PCI-X que es simplemente una versión más rápida de PCI

2 PCI, al igual que Ethernet o FDDI, utiliza un canal compartido de acceso múltiple. Este tipo de redes no son escalables y las prestaciones se degradan rápidamente al aumentar el número de usuarios/dispositivos.

3 PCI transmite los bits en paralelo, lo que impide aumentar más la velocidad debido al jitter

4 Descontando la redundancia introducida por los protocolos, queda una velocidad ~ 200MB/s/sentido, aproximadamente el doble que PCI

5 PCI no tiene ningún tipo de QoS (Quality of Service)

1.1 Enlace básico x1

Un enlace básico del bus (Ilustración 1) está compuesto por un *lane*, que, como se ha dicho, son 2 pares de hilos. Cada par es un canal símplex que transportan una señal LVDS (Low Voltage Differential Signaling). La velocidad básica de un *lane* es 250MB/s/sentido (200MB/s/sentido)

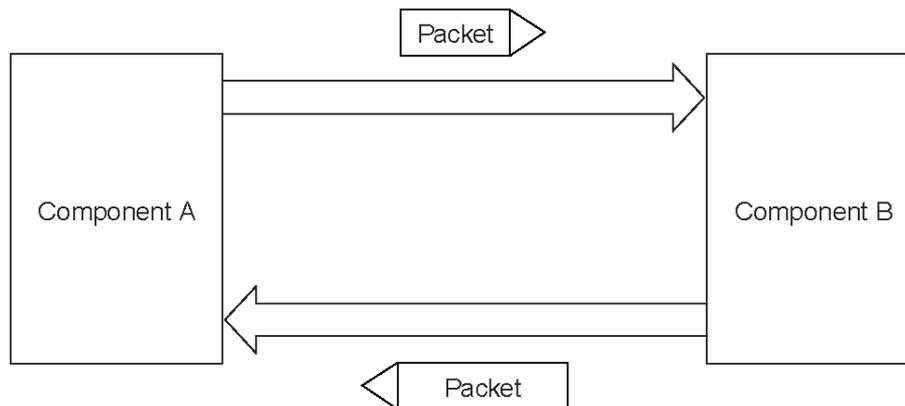


Ilustración 1: Enlace básico de la red PCI-Express (1 sólo lane) punto a punto. Un LANE son dos pares de hilos (cada flecha gruesa es un par), uno de TX y otro de RX. La información se transmite en serie y en paquetes (tramas)

Un enlace puede tener más de un *lane*, permitiendo así aumentar su velocidad, los conectores varían su ancho en función del número de *lanes*, como se puede ver en la Ilustración 2.

Bus	Velocidad (MB/s)
PCI (32bit-33MHz)	132
PCIe, conector x01	250 (básica)
PCIe, conector x04	1000 (=250x4)
PCIe, conector x08	2000 (=250x8)
PCIe, conector x16	4000 (=250x16)

Tabla 1: La velocidad del enlace es escalable incrementando el número de lanes

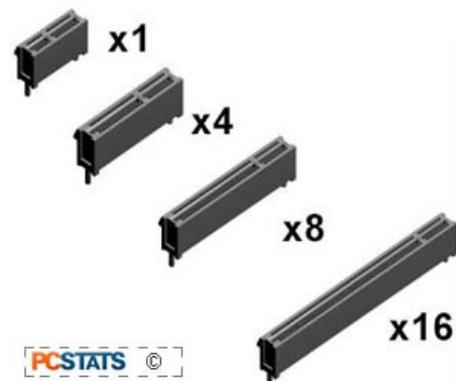


Ilustración 2: Diferentes anchos de conector en función del número de lanes

La norma PCI-Express asegura que una tarjeta x1 puede conectarse en todos conectores más anchos y negociar la velocidad básica⁶ x1. No ocurre lo mismo con tarjetas x2 y superiores en las que se permite que podamos introducir la tarjeta en el conector, pero no que la tarjeta funcione, tal como se muestra en la Tabla 2

⁶ La velocidad básica en bruto es 250MB/s/sentido, en la práctica queda en unos 200MB/s/sentido

<i>Tarjeta\Slot</i>	<i>x1</i>	<i>x4</i>	<i>x8</i>	<i>x16</i>
x1	Garantizado	Garantizado	Garantizado	Garantizado
x4	No	Garantizado	Permitido	Permitido
x8	No	No	Garantizado	Permitido
x16	No	No	No	Garantizado

Tabla 2: Interoperabilidad entre tarjetas y conectores (slots)

Si tenemos dos tarjetas x4 insertadas en zócalos también x4, eso no significa necesariamente que el enlace sea x4, puede ser menor. Durante la fase de inicialización, el nivel físico de las tarjetas negocia el número de lanes que se usan, y este puede ser menor si una de las partes por alguna razón no admite esa velocidad

1.2 Especificaciones mecánicas

Existen diferentes especificaciones mecánicas para el conector en función del tamaño de la tarjeta (form factor). Para los conectores estándar x1 y x4 en placa base, sus dimensiones físicas se muestran en la Ilustración 3 y la Ilustración 4, y los pines del conector x1 en la Tabla 3. Como ya hemos dicho, hay 4 tamaños diferentes de conector, todos ellos poseen una ranura en el extremo de 1.78mm que impide conectar la tarjeta al revés y pines espaciados 1.00mm. El conector x1 es el más pequeño y tiene 36 pines, el x4 tiene 64 pines, el x8 tiene 98 y el x16 tiene 164 pines. La altura del conector sobre la placa es de 11mm. La profundidad de todos ellos es de 8.70mm

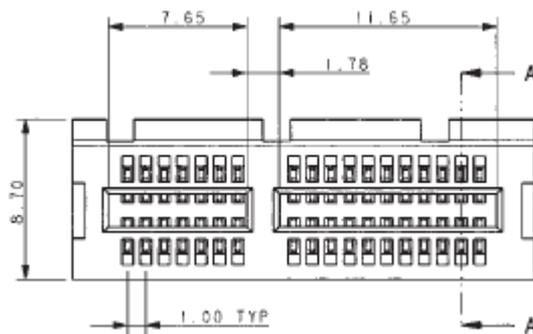


Ilustración 3: Conector x1 (36 pines)

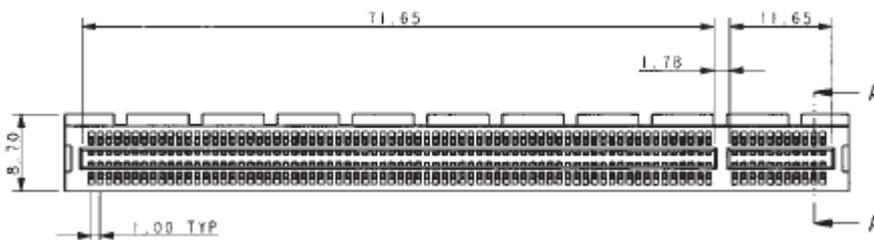


Ilustración 4: Conector x16 (164 pines)

La Ilustración 5 muestra el esquema eléctrico de un enlace LVDS (un lane). Aunque la norma dice

que la longitud de los enlaces debe ser menor de 50.8cm, bajo ciertas condiciones de impedancia en el *backplane*, se pueden permitir distancias mayores.

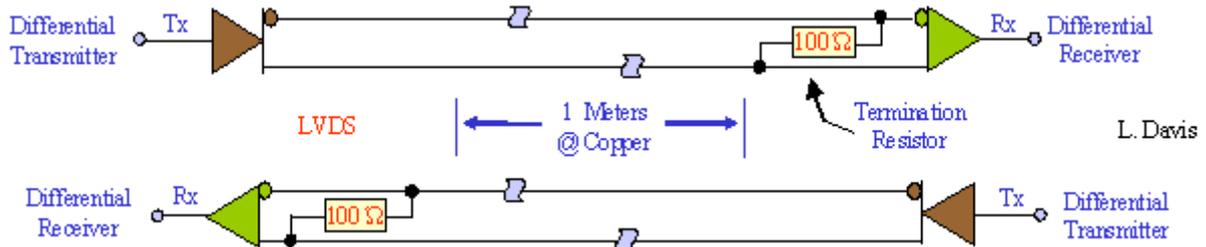


Ilustración 5: Esquema eléctrico un enlace básico PCI-Express, la longitud del cable en general debe ser menor de 50.8cm

Pin	Fila B del conector		Fila A del conector	
	Nombre	Descripción	Nombre	Descripción
01	+12v	+12 V	PRSNT#1	Hot plug presence detect
02	+12v	+12 V	+12v	+12 V
03	RSVD	Reserved	+12v	+12 V
04	GND	Ground	GND	Ground
05	SMCLK	SMBus clock	JTAG2	TCK
06	SMDAT	SMBus data	JTAG3	TDI
07	GND	Ground	JTAG4	TDI
08	+3.3v	+3.3 V	JTAG5	TMS
09	JTAG1	+TRST#	+3.3v	+3.3 V
10	3.3Vaux	3.3v V	+3.3v	+3.3 V
11	WAKE#	Link Reactivation	PWRGD	Power Good
12	RSVD	Reserved	GND	Ground
13	GND	Ground	REFCLK+	Reference Clock Differential pair
14	HSOp (0)	Transmitter Lane 0, Differential pair	REFCLK-	Reference Clock Differential pair
15	HSOn (0)	Transmitter Lane 0, Differential pair	GND	Ground
16	GND	Ground	HSIp (0)	Receiver Lane 0, Differential pair
17	PRSNT#2	Hotplug detect	HSIn (0)	Receiver Lane 0, Differential pair
18	GND	Ground	GND	Ground

Tabla 3: Pines del conector x1 (36 pines)

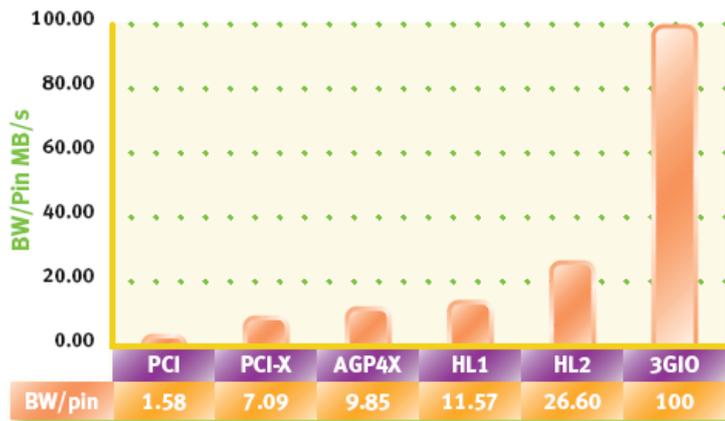


Ilustración 6: Ancho de banda/pin de diferentes buses

La transmisión serie de esta tecnología de bus permite reducir drásticamente el número de pines de los conectores reduciendo la complejidad del cableado y enrutamiento de pistas. En la XXX se puede ver la relación “Ancho de banda/pin” de diferentes tecnologías de bus, calculada como “Velocidad máxima de pico/número total de pines incluyendo alimentación, direccionamiento y control”.

1.3 Arquitectura de red

PCI-Express, a diferencia de PCI, es una red construida mediante enlaces punto a punto y conmutadores de paquetes (Ilustración 7 e Ilustración 8). Las redes de acceso múltiple como PCI son redes difícilmente escalables y sus prestaciones se degradan rápidamente al aumentar el número de usuarios/dispositivos.

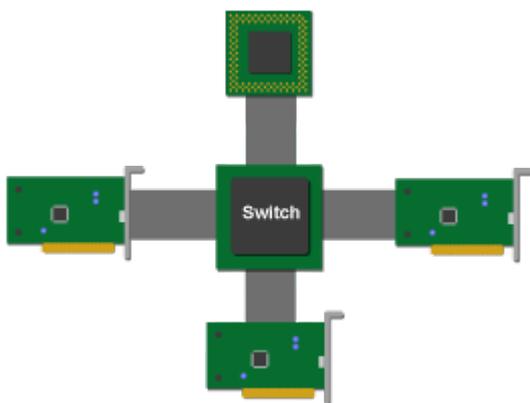


Ilustración 7: Red punto a punto (PCI-Express)

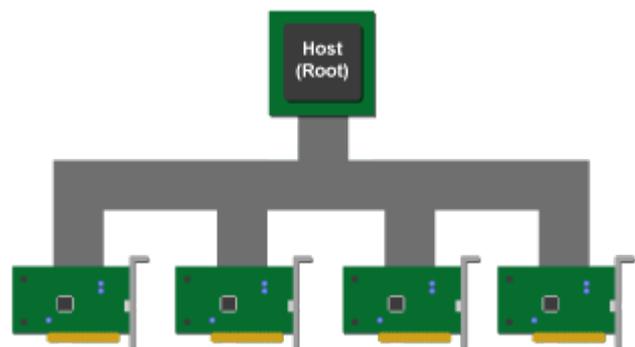


Ilustración 8: Red de acceso múltiple (PCI)

En PCI-Express, en cambio, la escalabilidad del bus es doble ya que, por una parte, podemos añadir más enlaces para conectar más dispositivos a media que se necesite y, por otro lado, podemos

aumentar la velocidad de cada enlace de la red usando más *lanes*⁷ (conectores más anchos). Una arquitectura de red típica PCI-Express se puede ver en la Ilustración 9, el detalle del conmutador de tramas se muestra en la Ilustración 10

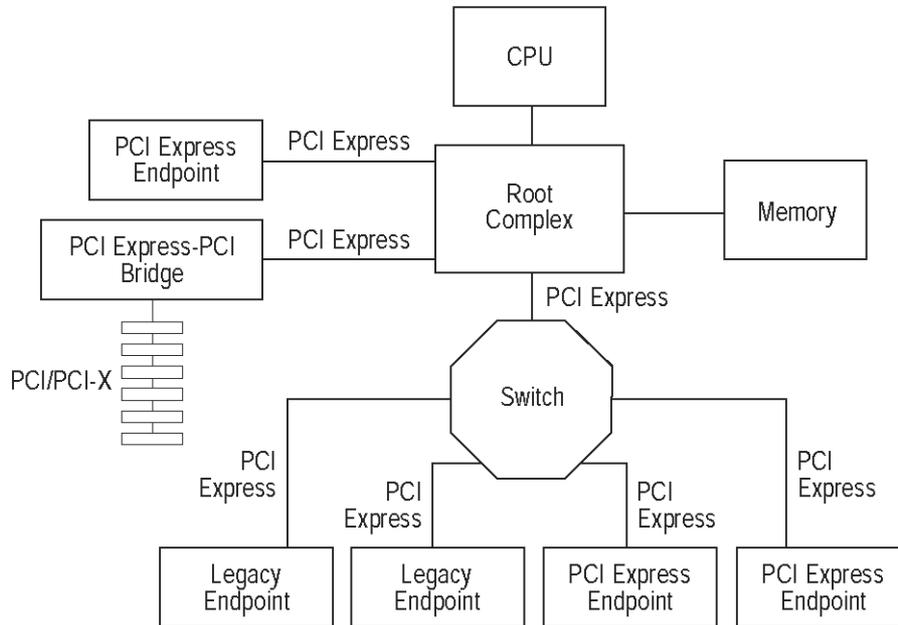


Ilustración 9: Topología típica de un bus PCI-Express conectado al módulo de mayor nivel en la jerarquía de E/S del sistema (Root Complex)

Igual que un router de Internet, el conmutador encamina las tramas entre las distintas tarjetas conectadas al bus, cada una de las cuales tiene su propia dirección de red.

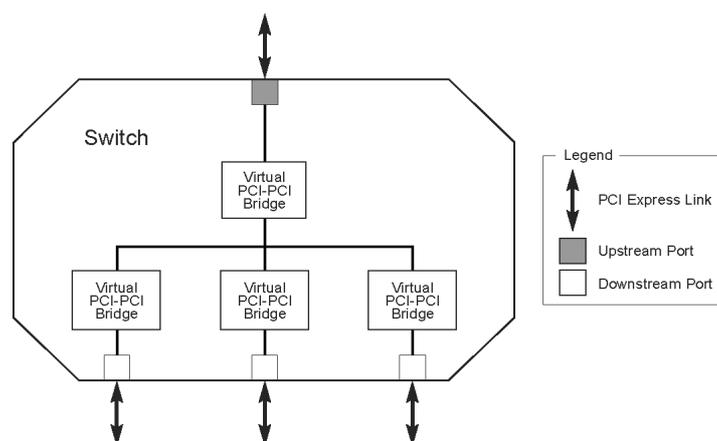


Ilustración 10: Detalle de un conmutador PCI-Express

⁷ En el futuro también se incrementará la frecuencia de reloj, aumentado así la velocidad básica de cada *lane* que podrá llegar, según los expertos, hasta los 10Gbps/sentido que es el límite de las líneas de transmisión de cobre.

La arquitectura OSI de PCI-Express tiene 3 niveles en hardware, como muestra la Ilustración 11, que se explicarán con más detalle en el siguiente apartado

- [Nivel Software]** No está implementado en la tarjeta, representa a todo el software del sistema operativo (los drivers) que utilizan los servicios que ofrece el nivel de transacción para enviar datos/mensajes a las tarjetas conectadas al bus. Como se ha dicho, la interfaz software es exactamente igual que en el caso de PCI, no hay que cambiar el software.

- Nivel de transacción** Ofrece un conjunto de primitivas `read_*` y `write_*` no bloqueantes al nivel SW para que este pueda ordenar la transmisión y recepción de tramas (que en este nivel se llaman TLP - Transaction Layer Packet) desde o hacia la tarjeta.

- Nivel de enlace** Secuenciación de los paquetes y encaminamiento, fiabilidad en el enlace (comprobación de CRC y retransmisión) y control de flujo, entre otras cosas.

- Nivel físico** De/serialización de los datos, codificación con el código 8B/10B, de/modulación de los símbolos, recuperación de sincronismo, blanqueador (scrambler) basado en LFSR, igualación de canal (mediante preénfasis/deénfasis) usando secuencias de entrenamiento, negociación del enlace (¿cuántos lanes se usan?), etc, etc, etc...

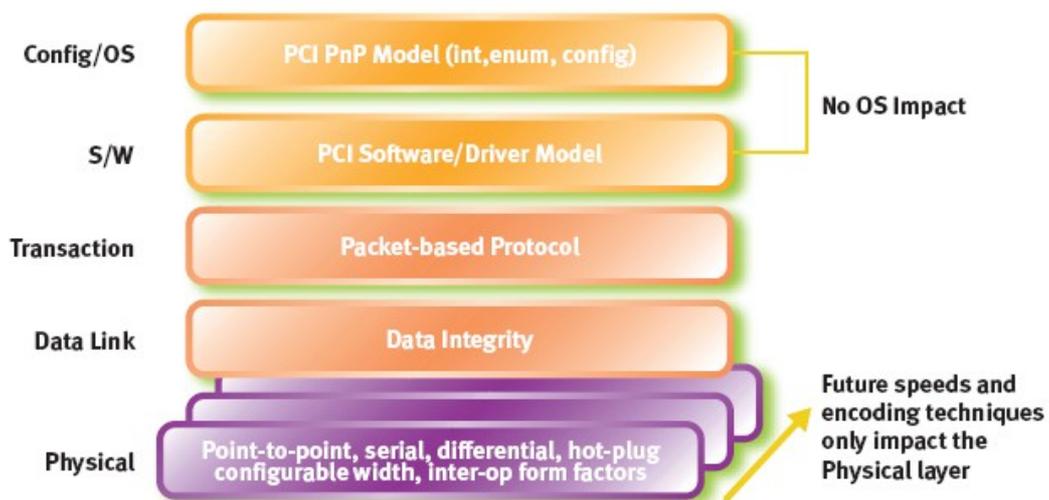


Ilustración 11: Arquitectura en niveles o capas del bus PCI-Express

Igual que ocurre en el modelo OSI o Internet, la trama del nivel i se encapsula dentro de la trama del nivel $i-1$ en un proceso conocido como *tunneling* hasta que, finalmente, el último nivel (el nivel físico) transmite la trama por el medio.

Las cabeceras que se introducen en cada nivel contienen información de control que permite la secuenciación de las tramas, su encaminamiento y la provisión de fiabilidad en el enlace entre otras cosas. Sin embargo, estas cabeceras también añaden redundancia que resta capacidad al canal y aumenta la latencia, algo crítico en un bus de E/S. El formato de trama de PCI-Express es un compromiso entre características, flexibilidad y latencia, como se explica en el Apartado 2.

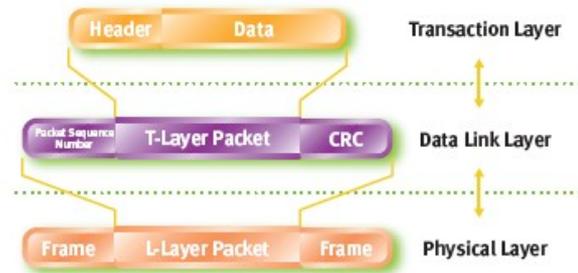


Ilustración 12: Encapsulación (tunneling) de tramas en cada nivel

1.4 Calidad de servicio (QoS)

Las aplicaciones multimedia actuales cada vez son más exigentes con el sistema E/S, ya que cada vez generan más tráfico que tiene restricciones temporales muy concretas.

A modo de ejemplo, una tarjeta capturadora de televisión digital, necesita entregar las tramas MPEG en la memoria principal con un retardo no superior a cierto umbral, si la trama llega a destiempo entonces no sirve para porque el instante de reproducción ya ha pasado y es como si nunca hubiera llegado.

La característica de tiempo real de este tráfico obliga al bus de E/S a diferenciarlo del resto y a darle un tratamiento especial, prioritario (en telemática esto se llama disciplina de cola).

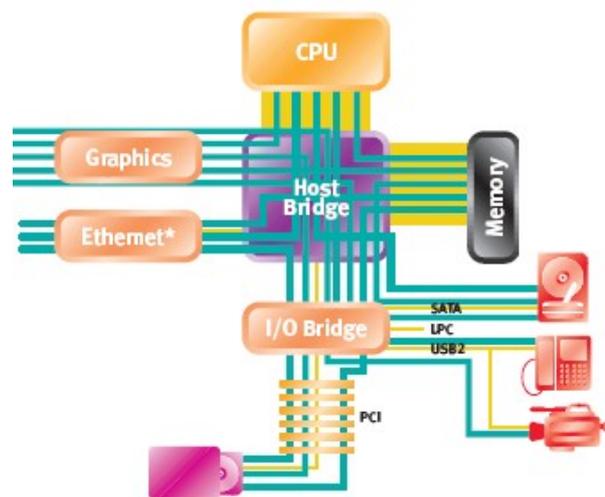


Ilustración 13: Tipos de tráfico de E/S

En PCI-Express existe, como se estudiará en el siguiente apartado, un mecanismo básico de etiquetado del tráfico (TC – Traffic Class) y de disciplina de cola en los buffers (VC – Virtual Channels) proporcionando así calidad de servicio (QoS – Quality of Service), algo de lo que carece PCI 2.2 o PCI-X.

Además, como consecuencia de la arquitectura de red punto a punto, la velocidad del enlace está disponible para cada conector/dispositivo, no está compartida entre todos, como se muestra en la Ilustración 14

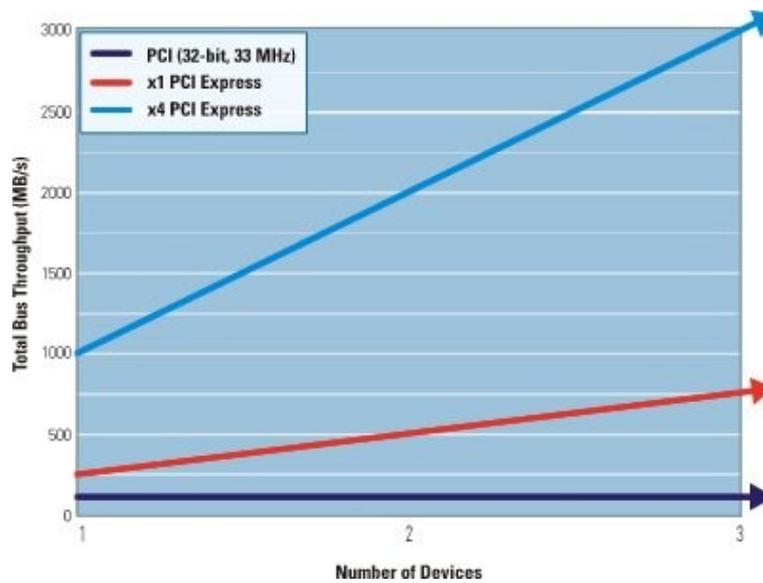


Ilustración 14: En PCI-Express, la velocidad de transmisión no se comparte (a diferencia de lo que ocurre en PCI)

1.5 Comparación con otros buses

PCI-Express no fue diseñada para interconectar procesadores, o memoria principal. Sin embargo, aunque existen otros buses mejores y que pueden sustituir a PCI-Express, tal como [HyperTransport](#) o quizá incluso [InfiniBand](#), su relación no es de competencia entre sí porque tienen objetivos distintos.

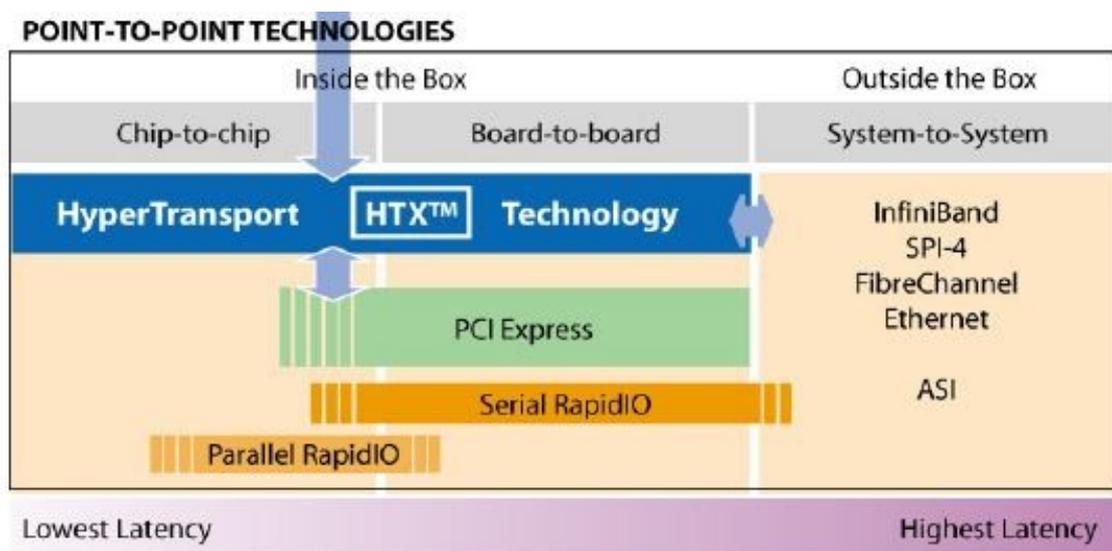


Ilustración 15: Ámbito de aplicación de los buses PCIe, HyperTransport, e InfiniBand

HyperTransport e InfiniBand son buses de altas prestaciones, muy caros y demasiado complejos para ser usadas como un simple bus local para el PC.

En la Ilustración 15 se ve como HyperTransport permite conexiones chip-chip o tarjeta-tarjeta a una velocidad muy alta (actualmente 41.6GB/s) y con una latencia muy baja debido principalmente a que su formato de trama tiene una cabecera más corta, como se muestra en la Ilustración 16.

Así pues, estos buses están dirigidos a segmentos de mercado diferentes (grandes centros de datos y de cálculo) y no compiten con PCI-Express.

PCI-Express surge como una alternativa más sencilla y barata propuesta por Intel (grupo de trabajo "Arapahoe") cuya relación coste/prestaciones es excelente (su coste es similar al de PCI) y su enorme flexibilidad le convierte en una tecnología de bus preparada para resistir el paso del tiempo. En los próximos años se convertirá en el sistema de E/S unificado para equipos de sobremesa y portátiles, es lo que se conoce como bus de E/S de tercera generación (3GIO).

La Ilustración 17 muestra la evolución de los buses del PC a lo largo de la historia comparados con otras redes y la Ilustración 18 los compara en función no sólo de su velocidad sino también de su latencia.

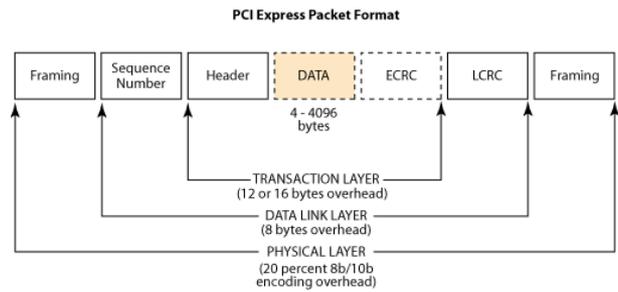
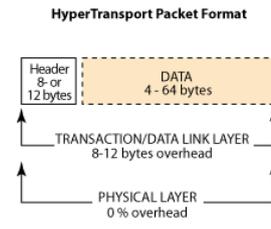


Ilustración 16: Trama HT vs PCI-Express

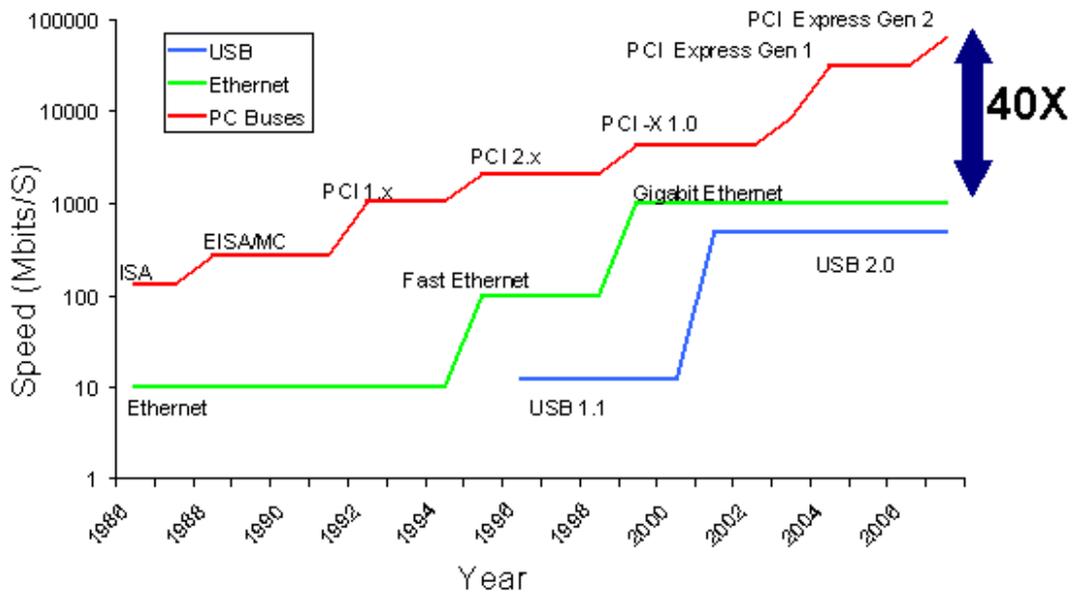


Ilustración 17: Evolución de los buses del PC, de ethernet y USB. PCI-Express se presenta en el mercado como bus PCI para la próxima década

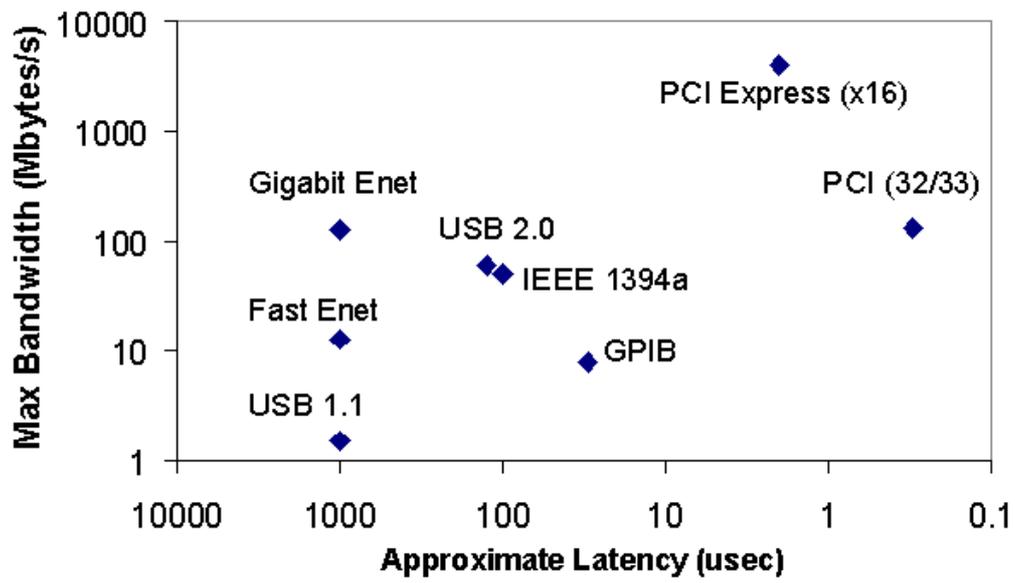


Ilustración 18: Velocidad vs latencia de PCI-Express respecto a otros buses y redes de comunicaciones

2. La norma “PCI-Express Base Specification Revisión 1.1”

Los siguientes niveles se implementan en el hardware de las tarjetas PCIe. La definición de niveles sirve fundamentalmente para facilitar la comprensión de la norma pero no implica ninguna realización/implementación particular del hardware. En cada nivel se distingue entre el receptor y el transmisor de información (Ilustración 19).

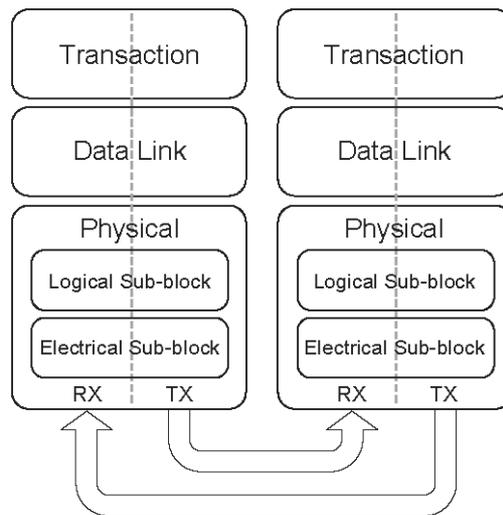


Ilustración 19: Niveles de la arquitectura de red

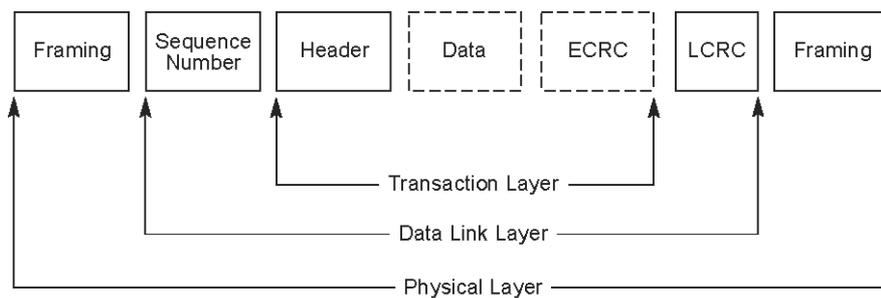


Ilustración 20: Cabeceras añadidas en cada nivel

Las tramas se generan en el nivel de transacción para enviar la información desde el nivel SW a la tarjeta (TX-RX). A medida que la trama en el transmisor baja hacia abajo en la pila niveles, se le van añadiendo cabeceras. En el receptor ocurre el proceso inverso. La Ilustración 20 muestra las cabeceras que se añaden en cada nivel.

2.1 Nivel de transacción

Construye las tramas TLP (Transaction Layer Protocol) a petición del SW. Las tramas TLP sirven para comunicar al extremo transacciones (primitivas u órdenes de control) de tipo read, write no bloqueantes⁸ (pipelined full split-transaction protocol) y ciertas clases de eventos. Todas las tramas tienen un número de secuencia único, de modo que es posible asociar unívocamente una trama de respuesta con la trama que la originó. Hay cuatro clases de transacciones:

- Memory. Pueden ser de tipo read o write y sirven para que la tarjeta lea/escriba datos en una dirección de memoria principal (por ejemplo enviar muestras de voz almacenadas en la memoria a una tarjeta de sonido)
- IO. Pueden ser de tipo read o write y sirven para que la tarjeta lea/escriba datos en una dirección de E/S. Este tipo de transacción está obsoleto y se aconseja no usarla ya que podrían desaparecer de la norma en revisiones futuras
- Configuration. Acceder a los registros d E/S de la tarjeta para configurarla (por ejemplo configurar una tarjeta de sonido a 44100Hz 16-bit)
- Message⁹. Sirven para indicar eventos como interrupciones u órdenes relacionadas con la gestión de energía (in-band signaling). Podemos ver estas transacciones como “hilos virtuales” en el bus para notificar este tipo de cosas, como ocurre en PCI (out band signaling).

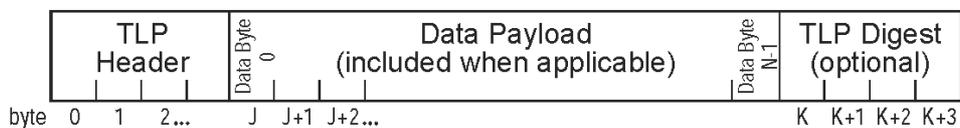


Ilustración 21: Formato genérico de las tramas TLP

La Ilustración 21 muestra el formato genérico de un TLP, su longitud es variable y se indica en la cabecera. Dependiendo del tipo concreto de transacción los campos de la cabecera varían.

El encaminamiento de TLPs se hace viendo ciertos datos de su cabecera que dependen del tipo de TLP. Para los de tipo Memory se usa la dirección de 32 o 64 bits de la tarjeta que figura en la cabecera y para los de tipo Configuration el conmutador examina el Id de bus, de tarjeta, etc

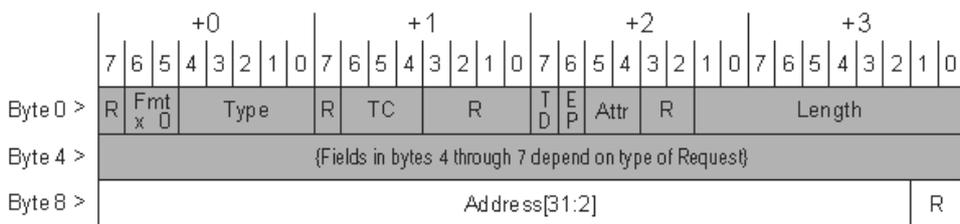
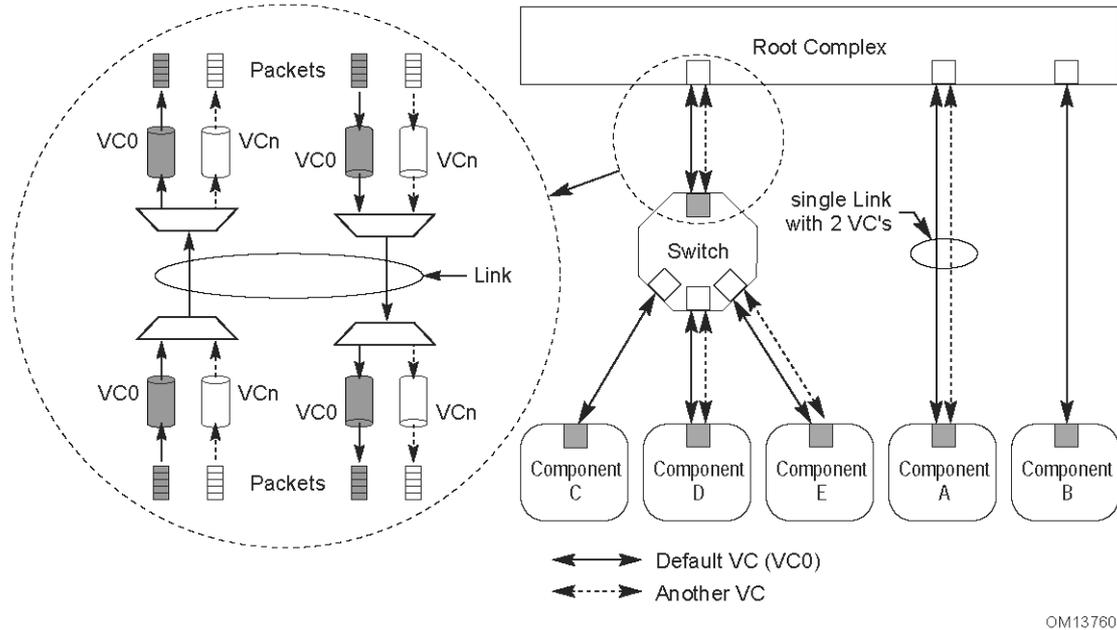


Ilustración 22: Dirección de 32 bits asignada a la tarjeta destino, usada por el conmutador para encaminar TLPs tipo Memory

⁸ Existen unos temporizadores que, al vencer, informan al nivel SW que algo ha ido mal

⁹ Esta transacción es nueva en PCI-Express, PCI no la tenía, en su lugar tenía señalización física en el bus.

El mecanismo de calidad de servicio (QoS) se basa en asignar etiquetas TC (Traffic Class) a los datos contenidos en cada buffer de transmisión, llamado VC (Virtual Channel). Estos buffers se implementan en una memoria de la propia tarjeta/conmutador.



OM13760

Ilustración 23: Concepto de VC (Virtual Channel)

Cada trama TLP está etiquetada con una etiqueta TC, y esta a su vez asociada con un VC concreto, de modo que la asignación de TLPs a VCs es bastante flexible. Cada VC representa un flujo de paquetes en espera de ser transmitidos por el canal. En la disciplina de cola usada para asignar el canal a cada VC es donde reside el mecanismo (básico) de QoS. Se puede asignar una prioridad fija a cada VC o bien usar una estrategia de Round Robin. El usar Round Robin no asegura la equidad en el acceso al canal para los TLPs, ya que puede haber muchos más paquetes asignados a un VC que a otro.

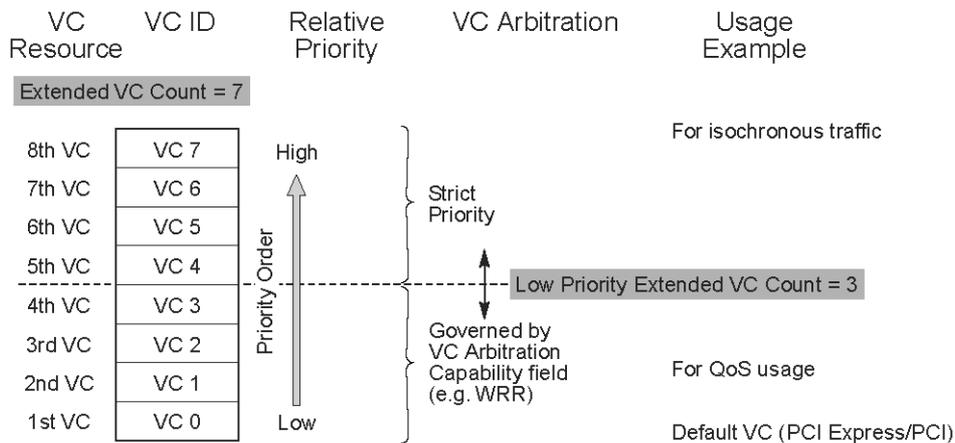


Ilustración 24: Disciplinas de cola para los distintos VCs

El mecanismo de control de flujo¹⁰, por otra parte, se basa en un intrincado sistema de créditos consumidos/otorgados individualmente a cada VC y mensajes de control, cada trama que sale por un VC consume un número de créditos variable que depende del tipo de TLP, cuando se ha agotado el crédito ya no se puede transmitir más hasta volver a recuperarlo. No entraremos en más detalles.

2.2 Nivel de enlace

Este nivel añade un número de secuencia a las tramas TLP y un CRC para hacer el enlace fiable, en caso de error intenta la retransmisión (estrategias ARQ)

Este nivel además de procesar las tramas TLP, también genera sus propias tramas (DLLP - Data Link Layer Packet) que le sirve para monitorizar el estado del enlace en todo momento, de ello se encarga una máquina de estados implementada en el hardware de la tarjeta.

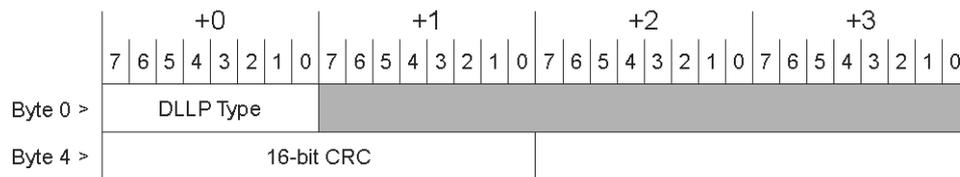


Ilustración 25: Trama DLLP para el control y monitorización del enlace

El polinomio generador del CRC de 32 bits que se añade al TLP expresado en binario es 0x04C11DB7 y se puede generar con un circuito como el de la Ilustración 26

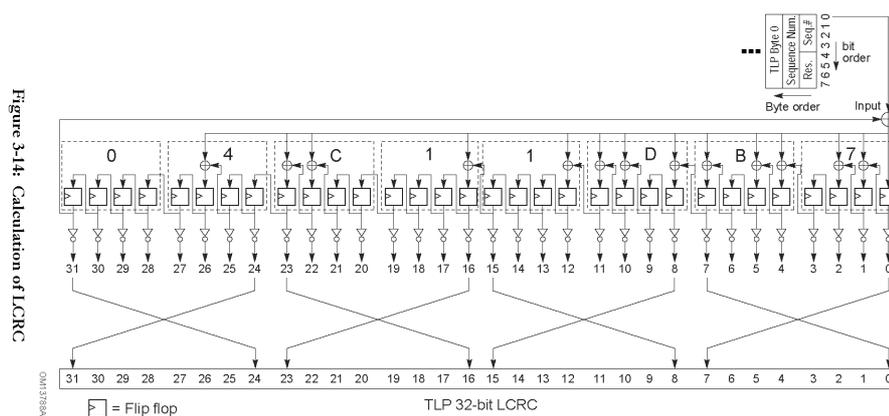


Ilustración 26: Circuito para calcular el CRC-32

10 Control de flujo sirve para que un transmisor rápido no abrume a un receptor lento, no tiene nada que ver con QoS

2.3 Nivel físico

Este nivel está dividido a su en dos subbloques:

- el bloque lógico, circuitos que hace operaciones con bits pero que aun está separado del medio físico y de las señales eléctricas que viajan a su través, de/serialización de las tramas, codificación con
- El bloque eléctrico, circuitos que representan la interfaz con el medio físico, igualación canal, adaptación de impedancia con el medio, código de línea, recuperación del sincronismo, etc

2.3.1 Bloque lógico

La primera operación de este bloque es codificar cada símbolo de entrada de 8 bits en un símbolo de 10 bits usando un código bloque 8b/10b (IEEE 802.3z, 36.2.4). Este código trata los 8 bits de entrada como 3 + 5 bits que son asignados a un código de 4 + 6 bits respectivamente. No todas las secuencias 10b son posibles y hay 12 combinaciones binarias que representan códigos especiales que se usan, entre otras cosas, para marcar el comienzo y fin de trama (alineamiento de trama)

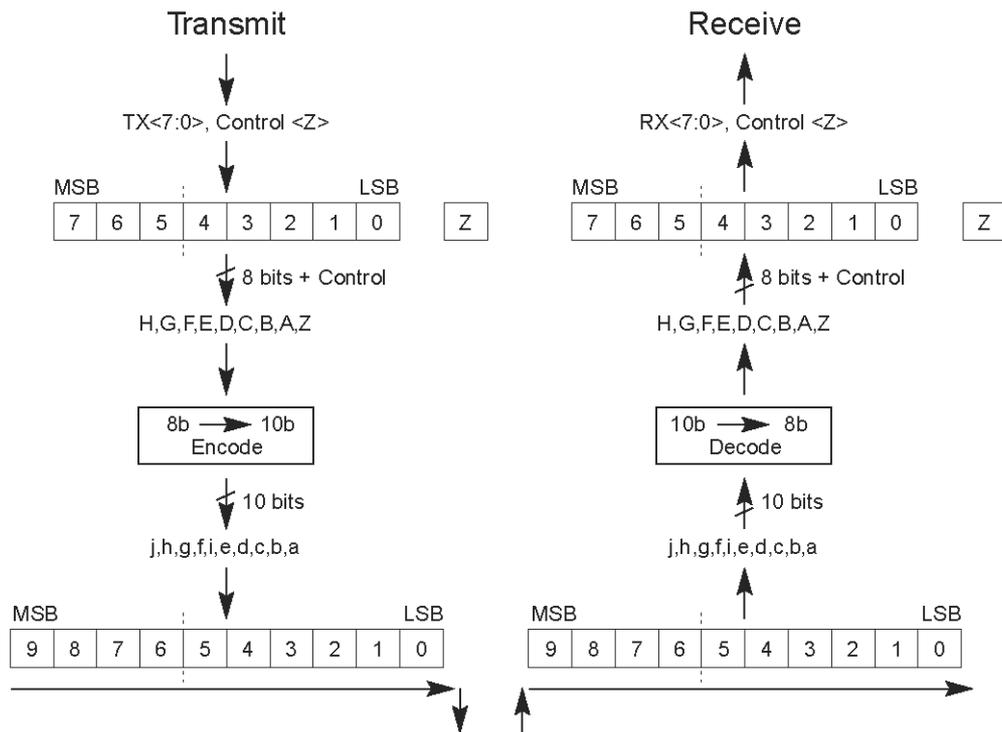


Ilustración 27: Código bloque 8b/10b

OM13793

Lo siguiente es poner en serie para su transmisión los símbolos de 10 bits que salen del codificador, para el caso de 1 lane, la Ilustración 28 muestra el orden de transmisión.

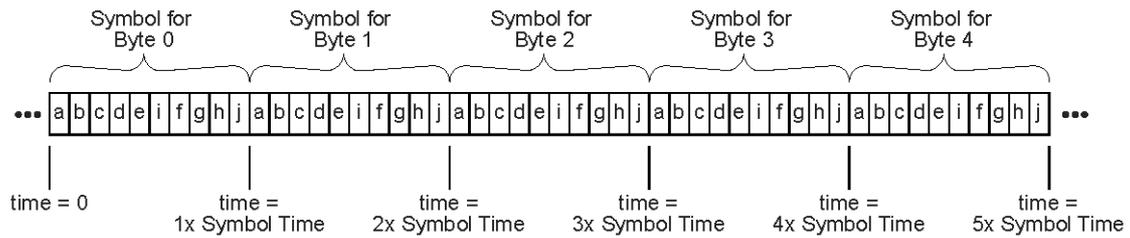


Ilustración 28: Orden de transmisión para un conector x1

Una vez puestos en serie, se les suma una secuencia de bits pseudoaleatoria para blanquear el espectro de la señal modulada y también para facilitar la recuperación del sincronismo. La secuencia de bits pseudoaleatorio se puede generar con un filtro lineal realimentado LFSR (Linear FeedBack Shift Register) cuyo polinomio es $G(x) = x^{16} + x^5 + x^4 + x^3 + 1$. Este polinomio maximiza la longitud¹¹ de la secuencia generada y por tanto genera una distribución uniforme

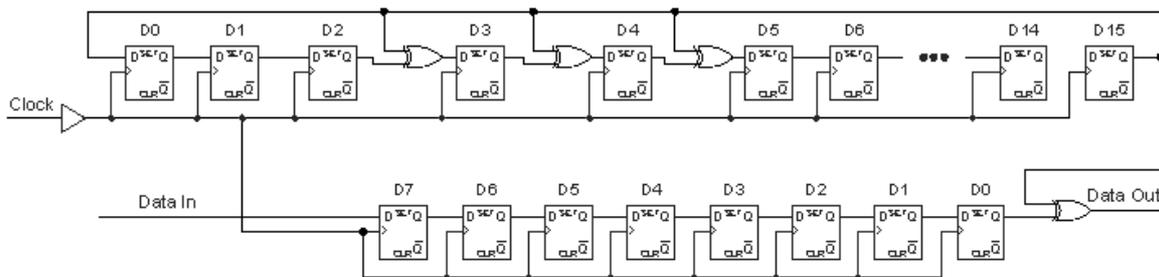


Ilustración 29: Registro LFSR usado como scrambler

Además de estas partes, hay otras encargadas de generar la secuencia de entrenamiento del canal, de negociar el enlace, etc. El receptor haría todas estas operaciones a la inversa.

2.3.2 Bloque físico

La modulación de los bits se salen del bloque lógico genera una señal de tensión diferencial entre los dos conductores del par de la lane. (LVDS - Low Voltage Differential Signaling). La Ilustración 30 y la Ilustración 31 muestran la forma de onda de la señal modulada y el diagrama de ojo mínimo que de cumplir al salir del transmisor (existe otro correspondiente al receptor)

¹¹ La secuencia de salida de un LFSR se repite siempre periódicamente, con un periodo que depende de cuáles de sus biestables estén realimentados

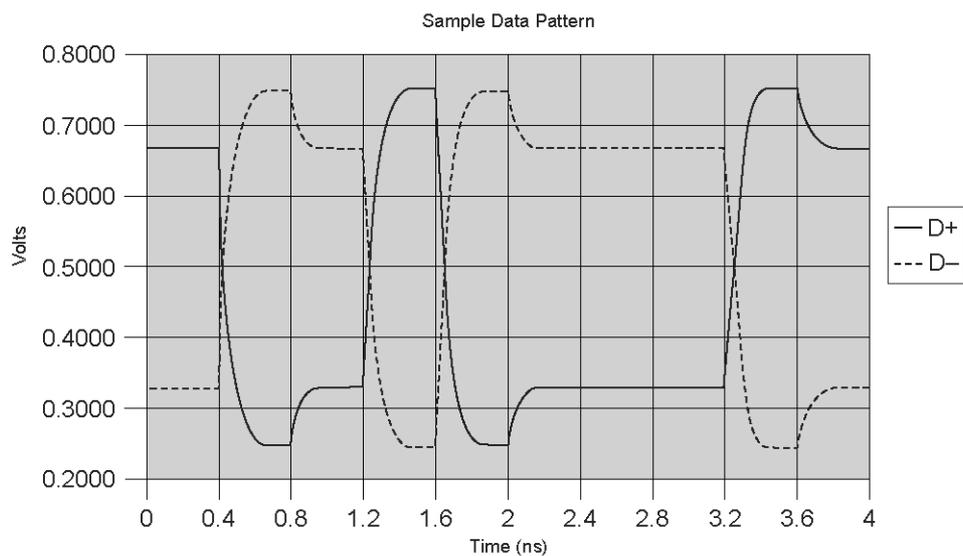


Ilustración 30: Señal LVDS correspondiente a la secuencia 1001000011 de izquierda a derecha (aplicando -3.5dB de deénfasis en el canal)

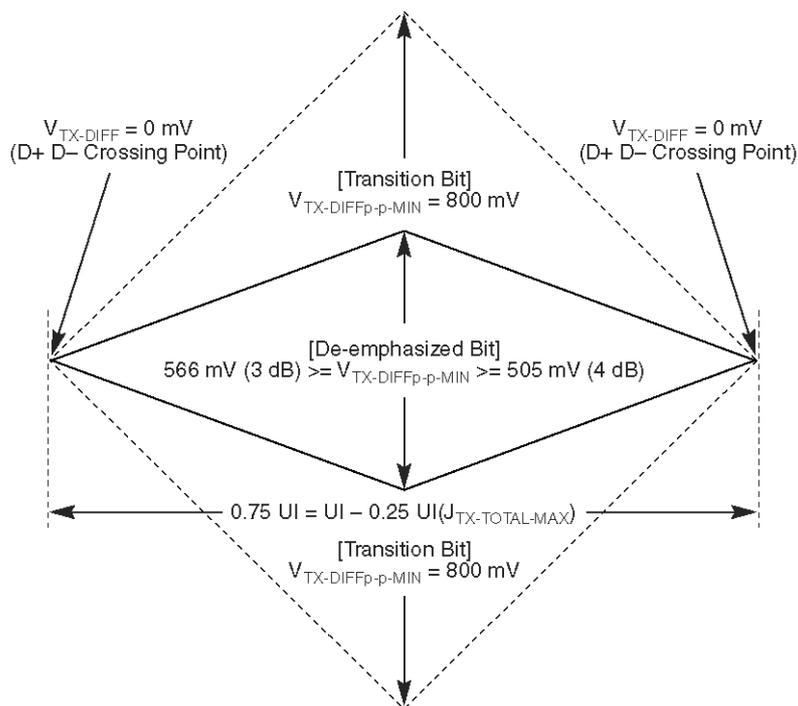


Ilustración 31: Restricciones sobre el diagrama de ojo que tiene que cumplir la señal transmitida (diferencial)

3. PCI-Express en el mercado

Al igual que ocurrió con PCI e ISA, se espera que PCI-Express coexista al principio, debido a la inercia del mercado, con el bus PCI convencional y una migración total y lenta hacia PCI Express.

Examinemos el impacto del bus PCI-Express en 5 áreas diferentes:

1. Estaciones de trabajo (workstations)
2. Equipos portátiles
3. Instrumentación electrónica

3.1 Estaciones de trabajo

PCI-Express está pensado para ser usado sólo como bus local. Debido a que se basa en el bus PCI, las tarjetas actuales pueden ser reconvertidas a PCI-Express cambiando solamente la capa física. La velocidad superior de PCI-Express permite reemplazar casi todos los demás buses, AGP y PCI incluidos.

Acerca de la arquitectura actual de los PCs:

En el PC actual existen dos chips que gestionan las transferencias de datos entre la CPU, la memoria y otros periféricos, tal como se muestra en la Ilustración 32.

1. NorthBridge o MemoryHub¹². Conecta la CPU con la memoria principal, la memoria cache L2 (nivel 2), el puerto AGP y el SouthBridge. Las innovaciones tecnológicas en la CPU y memoria principal obligan a rediseñar este chip casi con cada nuevo procesador que sale al mercado, de ahí de esté físicamente separado del SouthBridge
2. SouthBridge o IOHub. Conecta el NorthBridge con los diferentes buses del PC como ATA, USB, PCI.... No hay diferenciación entre tipos de tráfico, por lo que el nivel de QoS es nulo.

La idea de usar una sola tecnología de bus para conectar todos los periféricos de un ordenador es utópica y jamás podrá ser realidad debido a gran cantidad de periféricos diferentes cada uno de los cuales tiene sus propias particularidades. Aun así, es posible poner cierto orden en el caos de buses existente en el PC actual y simplificar el chipset de la placa base. PCIe será el bus al cual se conectarán todos los demás buses del PC mediante los adaptadores adecuados. Un puente PCI-PCIe (PCI to PCIe bridge) por ejemplo, traduce los paquetes PCIe en la señalización clásica de PCI.

La idea de Intel es tener un solo controlador PCIe comunicándose con todos los dispositivos, en vez de con el actual sistema NorthBridge/SouthBridge como se muestra en la Ilustración 33. En este sistema, los transeptores USB, ATA, PCI... transferirán sus datos a través del nuevo bus.

¹² Es la nueva nomenclatura de Intel para sus nuevos chipsets (Intel Hub Architecture) más rápidos y con cierto grado de diferenciación de tráfico, aunque la idea básica es la misma: separar el acceso a memoria principal del acceso a los periféricos de E/S

Mientras eso no ocurra, el controlador PCI-Express se seguirá conectando al SouthBridge junto a todo lo demás en una arquitectura que podemos llamar “de transición”, como se muestra en la Ilustración 34.

El bus PCI-Express ya está presente en las nuevas placas base para PC en el mercado (año 2006) como por ejemplo la serie 900 de Intel. Los nuevos conectores PCIe sobre placas base en formato BTX¹³ (Balanced Technology eXtended) se muestran en la Ilustración 36

El anuncio en 2005 de las compañías [Nvidia](#) y [ATI](#) —los dos mayores fabricantes de tarjetas gráficas— diciendo que iban fabricar tarjetas gráficas PCIe, a dado el impulso fundamental y definitivo para su aceptación en el mercado. En la Ilustración 35 se muestra una tarjeta gráfica PCIe de nVidia de gama alta para aplicaciones gráficas exigentes, como por ejemplo los videojuegos.

Ya no se desarrollan mejoras en el puerto AGP porque ha quedado obsoleto y está siendo reemplazado por el bus PCI-Express (x16). En este bus podemos conectar más de una tarjeta y hacer procesamiento de vídeo en paralelo, nVidia llama a esta tecnología [SLI](#) y ATI la llama [CrossFire](#).

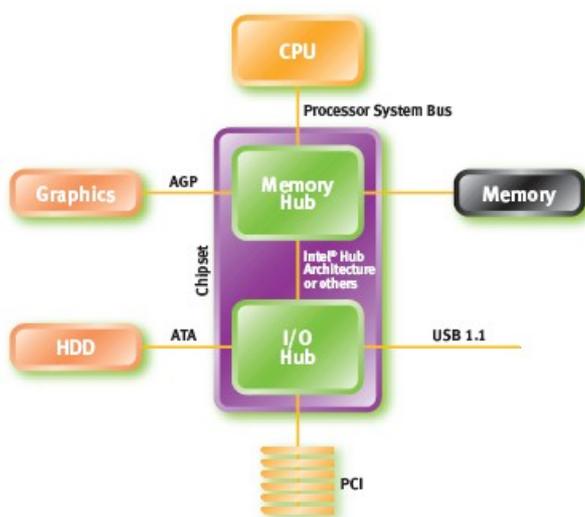


Ilustración 32: Arquitectura del PC actual

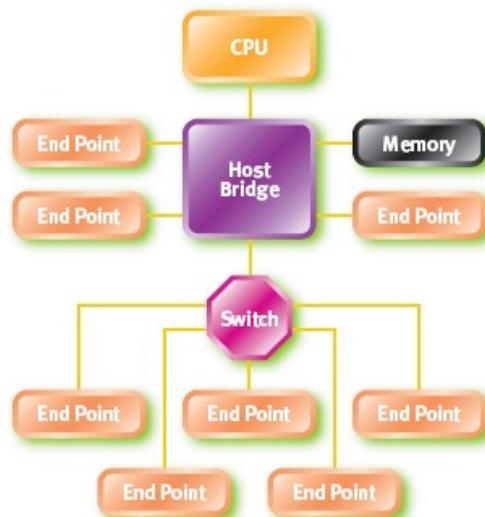


Ilustración 33: Arquitectura del PC en el futuro.
Un conmutador PCI-Express reemplaza al SouthBridge por completo.

13 El formato BTX es el sucesor del formato ATX, es más pequeño que este último e incompatible con el excepto por la fuente de alimentación, que se puede reutilizar.

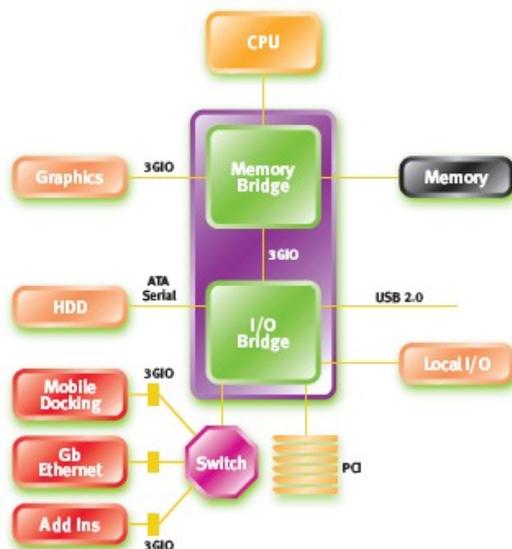


Ilustración 34: Arquitectura de transición del PC para el año 2005 y siguientes. Coexisten PCI-Express, PCI y otras interfaces todas unidas al SouthBridge

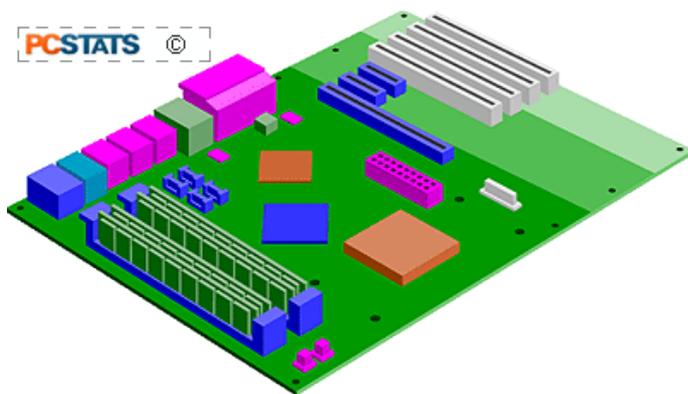


Figure 1. BTX handles a variety of motherboard features and system sizes.

Ilustración 36: Slots PCI-Express en azul junto a los viejos slots PCI en blanco, la placa base es de formato BTX



Ilustración 35: Tarjeta gráfica nVidia PCI-Express (x16) XFX 7900GT PCX DDR3 HDTV, con 256MB de memoria de vídeo y salida digital DVI y HDTV (335€)

3.2 Ordenadores portátiles. ExpressCard

Las ExpressCards son tarjetas compactas (small form factor) para ordenadores y otros dispositivos portátiles y que sustituyen a las antiguas CardBus. A través de su conector (ver Ilustración 37) la tarjeta se puede comunicar con el host mediante dos interfaces distintas:

- PCI-Express, velocidad básica x1, o bien
- USB 2.0

La elección de una interfaz u otra queda a elección de la propia tarjeta en función de cual sea la interfaz más apropiada para abordar su tarea. Cualquier host que sea compatible con tarjetas ExpressCard debe poseer estas dos interfaces asociados al mismo conector de la Ilustración 37. También se pueden des/enchufar en caliente (hot-plugable).

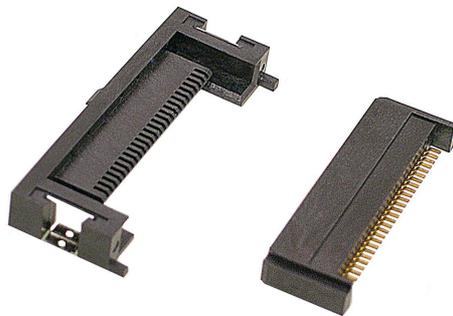


Ilustración 37: Conector hembra y macho de una tarjeta ExpressCard

Los tamaños de las tarjetas son dos, ver la Ilustración 38:

- ExpressCard/34, 34 mm de ancho
- ExpressCard/54, 54 mm de ancho y en forma de “L”

Respecto a su posición en el mercado, esta es de absoluto éxito. Hewlett-Packard, por ejemplo, empezó ya en 2004 a vender portátiles con estas tarjetas, Dell Computer las ha incorporado en sus modelos Inspiron y Latitude. Apple vende ya sus nuevos portátiles MacBook con CPU Intel con estas tarjetas... Es claro que en pocos años sustituirán a las viejas CardBus.

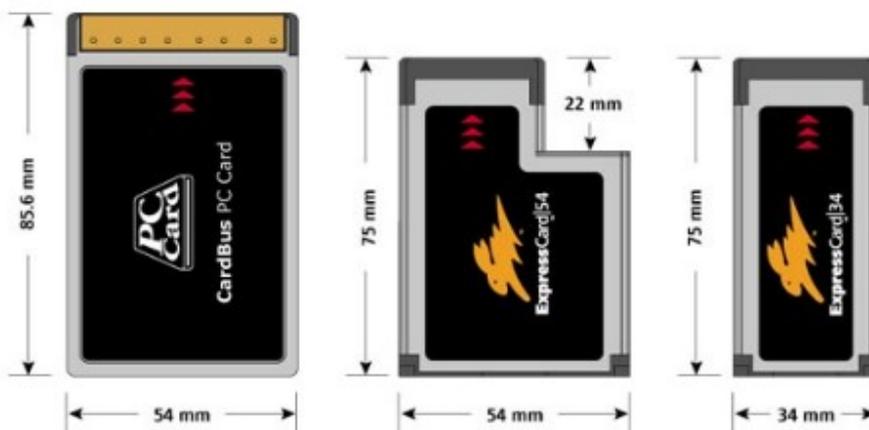


Ilustración 38: Tamaño (form factor) de las tarjetas ExpressCard (segunda y tercera por la izquierda) comparadas con una antigua tarjeta CardBus (la primera por la izquierda)

3.3 Instrumentación PCI-Express y PXI-Express

Acerca de PXI y PXI-Express:

PXI es una norma para instrumentos de medida y control que usa las especificaciones eléctricas de PCI pero con un backplane y con un conector distinto. Al backplane se añaden buses especiales de sincronismo y se usa un conector mucho más robusto y de muchos más pines llamado Eurocard preparado para resistir ambientes industriales. La Ilustración 39 se muestran:

1. Un chasis PXI, que tiene las ranuras donde se insertan las tarjetas (típicamente entre 4 y 18 ranuras) además del backplane y los conectores Eurocard
2. Controlador, es un PC en forma de tarjeta que controla el bus que debe estar situado en la ranura más a la izquierda
3. Las propias tarjetas, que contienen los instrumentos de medida y control



Ilustración 39: Chasis PXI con el controlador (tarjeta más a la izquierda) y las tarjetas (osciloscopios, multímetros, etc)

PXI-Express es el equivalente pero tomando la señalización eléctrica de PCI-Express en lugar de PCI beneficiándose así de la mayor velocidad de PCI-Express. Los chasis PXI-Express son totalmente compatibles con PXI porque lo que han hecho es usar los pines no usados del conector Eurocard para llevar las señales de PCI-Express, como muestra la Ilustración 40



Ilustración 40: Backplane de un chasis preparado para PXI-Express. Se pueden insertar tanto tarjetas PXI como PXI-Express

Como consecuencia de la compatibilidad software entre PCI y PCI-Express, las tarjetas PXI y PXI-Express también son compatibles por software. Desde el punto de vista del software, PCI, PXI, PCI-Express y PXI-Express son la misma tarjeta.



Ilustración 41: Dos tarjetas con distinto conector (PCI-Express y PXI-Express) y un mismo software para las dos

Como ejemplo de productos PCIe en el mercado, en la Ilustración 43 se muestra la tarjeta de adquisición NI PCIe-6251 de National Instruments de 16 entradas y 4 salidas analógicas con una resolución de 16 bits y una frecuencia de muestreo de 1.25MSample/s (entre otras prestaciones) dotada de conector PCI-Express x1 (250MB/s). Cada canal genera un tráfico de $1.25\text{MS/s} \times 16\text{bits} = 2.38\text{ MB/s}$, así pues con un conector x1, podría tener hasta un máximo de 80 canales aproximadamente.

Otro ejemplo de tarjeta que necesita gran velocidad de transferencia es un capturador de imagen (frame grabber), que sirve para capturar la señal de vídeo de las cámaras de alta resolución que se usan en aplicaciones de visión artificial. La Ilustración 42 muestra el capturador de imagen NI PCIe-1429, que puede capturar señales de vídeo de cualquier cámara profesional que cumpla la norma CameraLink® y que tiene un conector PCIe x4 capaz de absorber un tráfico de 680MB/s (el máximo permitido por la norma CameraLink®)

Por último, en la Ilustración 44 se muestra el primer chasis PXIe lanzado al mercado por National Instruments en enero de 2006.



Ilustración 42: Frame grabber NI PCIe-1429 captura la señal de cualquier cámara que cumpla la norma CameraLink (2800€)



Ilustración 43: Tarjeta de adquisición NI PCIe-6251 con conector x1 (250MB/s), 16 entradas y 4 salidas analógicas a 16bits 1.25 MSample/s



Ilustración 44: Chasis PCI-Express NI PXIe-1062Q

[MANOFAC] [PCIX1.1] [RAVI] [PCI3.0] [NI] [PROAKIS] [OJEDA] [PXISA] [CAMERALINK]
[NIPCIe6251] [INFINIBAND] [HYPERTRANSPORT] [INTEL]

Bibliografía

[MANOFAC] <http://www.interfacebus.com/>

[PCIX1.1] PCI-SIG, "PCI-Express Base Specification Revision 1.1", 2005.

[RAVI] Ravi Budruck, Don Anderson, Tom Shanley, "PCI EXPRESS System Architecture", 2003.

[PCI3.0] PCI-SIG, "PCI Local Bus Specification Revision 3,0", 2002.

[NI] <http://www.ni.com/pciexpress/esa/>

[PROAKIS] John G. Proakis, Dimitris G. Manolakis, "Tratamiento digital de señales", 1998.

[OJEDA] Luis Jacobo Álvarez Ruiz de Ojeda, "Diseño digital con lógica programable", 2004.

[PXISA] <http://www.pxisa.org/>

[CAMERALINK] <http://www.electronicstalk.com/guides/cameralink.html>

[NIPCIe6251] <http://sine.ni.com/nips/cds/view/p/lang/es/nid/201813>

[INFINIBAND] <http://www.infinibandta.org/home>

[HYPERTRANSPORT] <http://www.hypertransport.org/>

[INTEL] <http://www.intel.com/technology/pciexpress/>