Manteniendo la Coherencia de Cachés con Tecnología Nanofotónica

José L. Abellán, Eduardo Padierna[1], Alberto Ros y Manuel E. Acacio[2]

Resumen— **Los protocolos de coherencia basados en directorio están considerados como la mejor elección de diseño para proporcionar el máximo rendimiento en el mantenimiento de la coherencia en los multiprocesadores de memoria compartida, a pesar de su alta demanda de memoria y área requerida para la estructura de directorio. En este trabajo proponemos una nueva solución eficiente y escalable al problema del mantenimiento de la coherencia basándonos en un co-diseño de la red de interconexión y el protocolo de coherencia. Nuestra propuesta llamada ECONO, es un protocolo de coherencia para sistemas con un alto número de núcleos que usa PhotoBNoC, una sub-red dedicada basada en tecnología nanofotónica y de muy bajo costo que se utiliza para enviar mensajes *broadcast* atómicos emitidos por ECONO. Considerando un sistema de 256 núcleos, *ECONO+PhotoBNoC* representa una solución eficiente en términos de rendimiento, energía y consumo de área al problema de la coherencia.**

Palabras clave— **Multiprocesadores, Tecnología nanofotónica, Coherencia de caché.**

I. Introducción y Motivación

Desde la pasada década, el número de núcleos (*cores*) en las arquitecturas multicore ha ido creciendo a ritmo constante con el objetivo de obtener un incremento de rendimiento sostenible. Hoy en día, hemos alcanzado productos *manycore* a nivel comercial como el 72-core x86 Knights Landing MIC de Intel [1], o prototipos de incluso mil cores, como el sistema KiloCore [2].

En estos sistemas manycore, los protocolos de coherencia basados en directorio con invalidación ante las escrituras parecen ser la única alternativa viable para ofrecer rendimiento y escalabilidad. Aparte de estas características fundamentales, para su viabilidad, un protocolo también debe de resultar poco complejo y ser eficiente en términos de consumo de área y energía.

Estos requisitos son muchas veces opuestos y, por tanto, es difícil llegar a un diseño que los aúne todos al mismo tiempo. Por ejemplo, consideremos cómo los protocolos Hammer [3] y Directory [4] garantizan la coherencia. En Hammer, al no almacenar información de los compartidores de los bloques de memoria, se necesitan enviar desde la caché de último nivel (Last-Level Cache o LLC) tantos mensajes como número de cachés privadas de último nivel haya en el sistema. Por otro lado, en Directory, se codifica en la LLC información precisa de coherencia acerca de quiénes son los compartidores (p. ej., a través de un full bit-vector), pudiendo así enviar un número de mensajes de coherencia igual al número de compartidores de los bloques. De este modo, Directory es más eficiente en términos de rendimiento y consumo de energía, ya que sólo inyecta a través de la red de interconexión (Network-on-chip o NoC) los mensajes necesarios para garantizar un sistema de memoria coherente. Por otro lado, Hammer es más eficiente en cuanto a consumo de área, ya que no dedica ningún recurso hardware para codificar la lista de compartidores de los bloques de memoria almacenados en las cachés privadas.

La eficiencia del protocolo de coherencia también depende del diseño de la NoC empleada. Por ejemplo, la NoC del procesador MIT RAW [5] consume hasta un 40 % del total de la energía del chip. Aunque muchas propuestas se centran sólo en mejorar el diseño de la NoC, se ha demostrado que un co-diseño de la NoC junto a capas de más alto nivel (p. ej., el protocolo de coherencia) son una mejor opción [6]. En este trabajo también seguimos este modelo de diseño y proponemos un protocolo de coherencia llamado ECONO para el mantenimiento eficiente de la coherencia en sistemas manycore futuros.

ECONO, por lo tanto, es un protocolo de coherencia capaz de reunir los beneficios de Hammer en términos de área, y los de rendimiento y consumo energético de Directory. Esto se consigue utilizando una red dedicada on-chip de muy bajo coste para enviar los mensajes atómicos broadcast que ECONO requiere. ECONO ya fue presentado en [7] utilizando tecnología de G-Lines [8] para implementar esta red. En este trabajo resolvemos los problemas de integración y escalabilidad de las G-Lines, proponiendo una red dedicada fotónica llamada PhotoBNoC.

Las contribuciones principales de este trabajo son las siguientes:

- Se propone una sub-red nanofotónica on-chip llamada PhotoBNoC para el envío atómico eficiente de los mensajes de coherencia broadcast del protocolo ECONO.
- Se diseña PhotoBNoC para garantizar atomicidad en la entrega de los mensajes, alta rapidez y bajo consumo de energía. Esto se consigue mediante un esquema basado en canales Single-Write Broadcast-Reader (SWBR) que son distribuidos en segmentos de distinta longitud dependiendo de la distancia a los nodos receptores.
- Se evalúa la escalabilidad y beneficios de PhotoBNoC para un sistema de 256 cores utilizando aplicaciones paralelas. La evaluación considera tiempo de ejecución, tráfico de red, consumo de energía y área requerida.

[1]Dpto. de Ciencias de la Computación, Universidad Católica de Murcia, e-mail: {jlabellan,epadierna}@ucam.edu.

[2]Dpto. de Ingeniería y Tecnología de Computadores, Universidad de Murcia, e-mail: {aros,meacacio}@ditec.um.es.

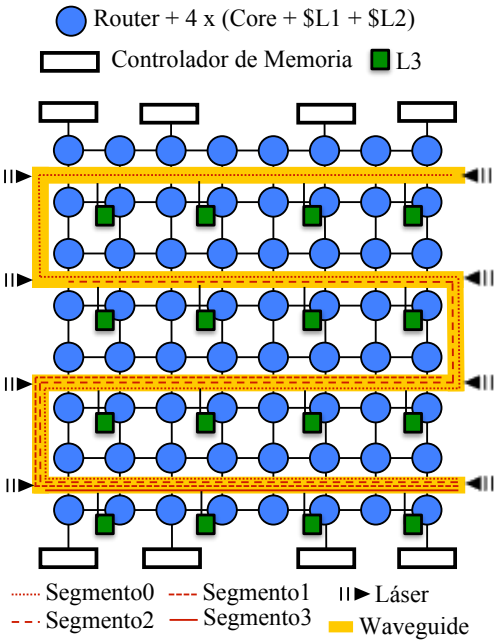


Fig. 1: Sistema de 256 cores bajo estudio. Nótese la disposición de las redes malla 2D eléctrica concentrada y la PhotoBNoC (Laser + Waveguide).

TABLA I: Sistema 256-core (11nm, 1GHz y 0.6V).

CORE	
Pipeline	2 vías, superescalar, fuera de orden
Reorder Buffer	40 entradas
Est. Reserva	36 entradas
Pred. Saltos	2 bit, 128 entradas
Unidades	1 FPU, 2 ALU, 1 MULT
JERARQUÍA DE CACHÉ Y MEMORIA	
Bloque	64 bytes
I/D L1 Privada	4 vías, 32 KB @ 1+4 ns
L2 Privada	8 vías, 256 KB @ 3+8 ns
L3 Compartida	16×[16 vías, 4-MB] Bancos @ 6+16 ns
Memoria	8 controladores, 8 PIDRAM @ 50 ns
RED DE INTERCONEXIÓN	
Topología	Malla 2D concentrada
Enrutamiento	X-Y
Control de Flujo	Virtual-channel (6 VCs) Credit Backpressure (16 Flits/VC)
Latencia Router	2 ciclos
Paquetes	72 bytes (datos); 8 bytes (control)
Flit	72 bytes

II. Plataforma de Estudio

A. Sistema Manycore

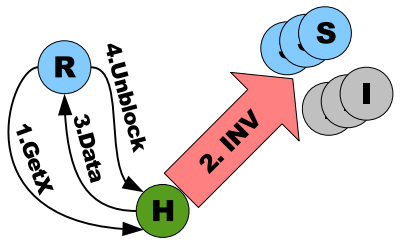
Como se puede ver en la Figura 1, el sistema de estudio consta de 256 cores, una jerarquía de cachés de tres niveles (L1 y L2 privadas, L3 compartida distribuida) basada en la arquitectura del TILE-Gx72 chip [9], con una red de interconexión híbrida que consta de una malla 2D concentrada eléctrica y la PhotoBNoC que será descrita en la Sección IV. El manycore está interconectado a un sistema de memoria de altas prestaciones aprovechando la tecnología PIDRAM [10] para la interfaz manycore-memoria.

B. Tecnología Nanofotónica

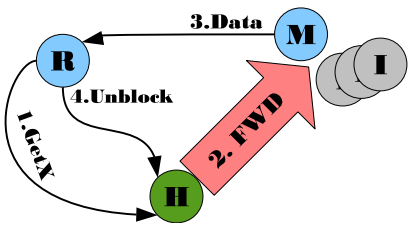
La tecnología nanofotónica ha sido propuesta como reemplazo de la tecnología CMOS eléctrica para implementar NoCs en futuros manycores [11], [12], [13]. La razón es que ofrece mucha más densidad de ancho de banda, mínima latencia de comunicación global y menor consumo de energía dinámica. Es por ello por lo que estudiamos en este trabajo una red dedicada nanofotónica para ECONO. Los parámetros de la tecnología nanofotónica utilizados en este trabajo (ver Tabla II) han sido obtenidos de estudios recientes [14], [15], [16]. Asumimos un sistema de integración monolítico, un ancho de banda de enlace de 8 Gbit/segundo/lambda, 16 lambda/waveguide/dirección, y una latencia convencional de enlace de 3 ciclos (un ciclo para la transmisión, un ciclo para la conversión electro-óptica y otro para la opuesta).

TABLA II: Tec. Nanofotónica. *Dinam.+Estát..

Laser source efficiency	20 %
Coupler loss, Splitter loss	1 dB, 0.2 dB
Modulator insertion loss	1 dB
Waveguide loss	1 dB/cm
Crossing loss	0.05 dB
Filter through loss	1e-3 dB
Filter drop loss	0.5 dB
Photodetector loss	0.1 dB
Non-linearity loss	1 dB
Modulator driver circuit energy*	0.035 pJ/b
Receiver circuit energy*	0.035 pJ/b
Thermal tuning power	16 μW/K
Receiver sensitivity	-17



(a) Invalidar compartidores de bloque.



(b) Recuperar bloque de su propietario.

Fig. 2: Mantenimiento de la coherencia en ECONO. Los mensajes a través de las fechas más gruesas (`2.INV` y `2.FWD`) viajan a través de PhotoBNoC.

III. Protocolo de Coherencia

En esta sección presentamos nuestro protocolo de coherencia, llamado ECONO (*Express COherence NOtification*). En la Figura 2 se muestra cómo ECONO actúa para dos escenarios típicos de mantenimiento de coherencia de cachés: invalidación de compartidores de un bloque y reenvío de datos desde el propietario de un bloque. En cualquier otra situación nuestro protocolo actúa de forma similar a cómo se hace en Directory o Hammer. La explicación se llevará a cabo tomando como base la jerarquía de memoria de nuestro sistema manycore. En particular, entre los niveles de caché L3 y L2.

El primer escenario se muestra en la Figura 2a. En este caso, un bloque de memoria es compartido

por varias cachés L2s (ver estado S) y un core sufre un fallo de escritura en su L2 privada para acceder a este bloque (ver estado R). Tras este evento, la L2 solicitante envía al directorio en la L3 la solicitud para obtener permiso de escritura y la última copia del bloque (1.GetX).

En Hammer, a diferencia de Directory, como el directorio no almacena la lista de compartidores de un bloque, se ha de enviar un número de mensajes de invalidación igual al total de cachés L2s del sistema para invalidar las copias privadas del bloque (ver estado S) y así garantizar el acceso de escritura exclusivo del solicitante. Cada mensaje de invalidación que alcanza una caché L2 privada tiene como respuesta un mensaje de reconocimiento (ACK) que se envía al directorio. Una vez que el directorio recibe todos los ACKs (tantos como cachés L2 privadas), se envía el bloque de datos al solicitante (ver 3.Data). Por último, el solicitante envía un mensaje de tipo Unblock al directorio para que éste pueda continuar manejando peticiones de memoria (el uso de Unblocks facilita el manejo de las condiciones de carrera [17]).

ECONO se distingue de Hammer en que en lugar de enviar tantos mensajes como cachés de L2 privadas haya en el sistema, se envía un único mensaje broadcast a través de nuestra red dedicada PhotoBNoC (ver 2.INV en Figura 2a). Este mensaje se envía de manera atómica para poder determinar con exactitud cuántos ciclos se van a emplear en la transmisión del mismo. Así, el directorio sabrá cuándo todas las cachés L2 han recibido el mensaje para eliminar sus copias del bloque evitando la necesidad de utilizar ACKs.

El segundo caso aparece en la Figura 2b. En este caso existe una sola copia del bloque que pertenece a una caché L2 propietaria (ver estado M). Tanto si la petición de la caché L2 privada solicitante (ver R) necesita permiso de lectura como de escritura sobre este bloque, se envía una petición al directorio (en la figura se muestra el caso de escritura, GetX). En caso de escritura, se tendría que invalidar la copia del bloque (pasaría de estado M a I). En el caso de lectura se tendría que realizar un "downgrade" de la copia privada en el propietario (de M pasaría a estado S).

En Hammer, a diferencia de Directory, la petición al propietario desde el directorio se realiza enviando tantos mensajes como cachés L2 privadas haya en el sistema para que la solicitud alcance al propietario del bloque. Una vez que el propietario recibe la petición del directorio, se cambia el estado del bloque de acuerdo a la solicitud (invalidación o downgrade) y además se reenvía la copia del bloque al solicitante (ver 3.Data). El solicitante, tras recibir el bloque de datos, envía el mensaje Unblock al directorio como en el caso anterior.

En ECONO se utiliza un sólo mensaje broadcast que es enviado atómicamente a través de la PhotoBNoC para alcanzar a todas las cachés L2 privadas (ver 2. FWD). Nótese que nuestros mensajes broadcast atómicos ofrecen también la ventaja de simplificar el diseño y verificación del protocolo de coherencia [25].

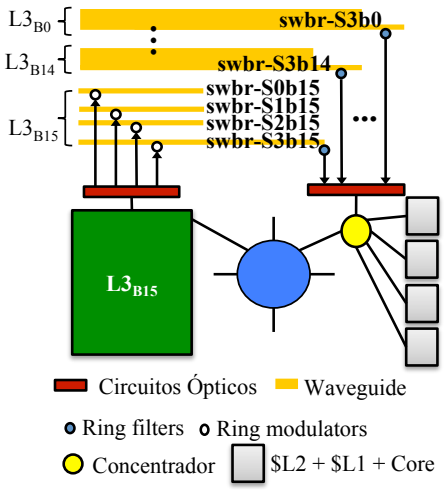


Fig. 3: Detalle de la interconexión de un banco de L3 con el concentrador de un router que conecta cuatro cachés L2 privadas. Se muestran también los canales SWBR necesarios con los cuatro segmentos por canal. Es decir, swbr-SXbY, donde X significa el identificador de segmento (de 0 a 3), e Y se refiere al identificador del banco de L3 (de 0 a 15).

IV. Red de Interconexión Fotónica

ECONO emplea mensajes broadcast atómicos que llamamos ACN (*Atomic Coherence Notification*) y son enviados sobre una red dedicada de bajo costo llamada PhotoBNoC, descrita en esta sección. A diferencia de la red diseñada en nuestro trabajo anterior [7], este trabajo propone el uso de tecnología nanofotónica para mejorar las prestaciones de la red en términos de escalabilidad de integración (nodos inferiores a 65nm), mayor densidad de ancho de banda (lo que reduce el impacto en área en el chip) y menor consumo energético.

Como los ACNs se envían desde el directorio en los bancos de L3 a las cachés L2 privadas, la red PhotoBNoC está compuesta de canales unidireccionales desde L3 a L2. Además, como se han de enviar en broadcast, cada L3 tendrá que estar conectada a todas las L2s. Para implementar esto de manera eficiente reduciendo el consumo energético de los láseres que alimentan los canales SWBR, utilizaremos canales SWBR segmentados y concentración de red, de manera que desde cada L3 se tendrán 4 sub-canales SWBR (o segmentos) que conectan el directorio con los 64 routers (16 routers por segmento). A su vez, cada router difundirá el mensaje de coherencia enviado desde el directorio a sus 4 L2 privadas a través de un concentrador. La Figura 3 muestra la disposición de la PhotoBNoC, mientras que la Figura 4 detalla la parte del emisor y la Figura 5 la parte receptor de los mensajes ACN.

Como los ACNs se envían de forma atómica, para garantizar su atomicidad, es decir, impedir que se pueda enviar algún otro mensaje durante la emi-

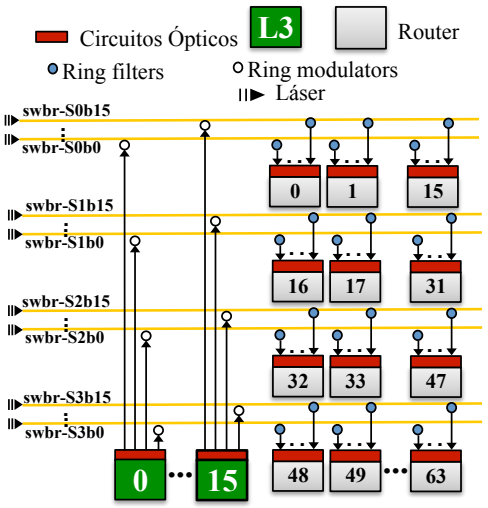


Fig. 4: Los 64 canales nanofotónicos de tipo SWBR requeridos en PhotoBNoC para nuestro sistema de 256 cores. Por simplicidad, en lugar de mostrar las 256 cachés L2 del sistema de estudio, se muestran los 64 routers del sistema, donde cada router interconecta 4 L2s privadas a través de un concentrador (ver Figura 3). Nótese cómo cada canal SWBR conecta un banco de L3 con 16 routers.

sión del mensaje ACN actual retrasando su entrega, la PhotoBNoC se diseña teniendo en cuenta que: (1) se usan canales SWBR haciendo que el escritor en el canal siempre tenga disponible su canal para enviar datos (hay un sólo escritor); (2), para permitir que los receptores siempre acepten el mensaje ACN, se utiliza un buffer en el que se van encolando los mensajes ACN (véase el buffer ACN Buffer Queue, o ABQ, en la Figura 5). Experimentalmente se ha determinado que el tamaño de este buffer debe ser de 16 entradas para evitar desbordamiento y por tanto retraso en la entrega de los ACNs.

Dada la configuración de tecnología nanofotónica asumida en este trabajo (8Gbit/segundo/lambda), la latencia de envío de los mensajes ACN de 9 Bytes (otros tamaños son considerados en [7]) es de 9 ciclos. A esto hace falta sumar 3 ciclos por la latencia de los canales nanofotónicos (ver Sección II-B). Así, el directorio sabe que en 12 ciclos (más uno de encolado en la ABQ), ha llegado el mensaje para invalidar las copias en todas las L2 privadas. Por esta razón, no hace falta utilizar los mensajes ACKs descritos en la Sección III.

V. Entorno de Evaluación

La fase de evaluación se llevó a cabo utilizando el simulador Sniper 6.1 [18] para estudiar el sistema de estudio descrito en la Sección II-A. Sniper fue ampliado para que soportara tanto los protocolos de coherencia de caché ECONO, Hammer y Directory, como la infraestructura PhotoBNoC. Para nuestro estudio sobre potencia disipada del sistema manycore, utilizamos la herramienta McPAT que viene integrada en Sniper para estimar la potencia disipada por los cores y la jerarquía de caché del sistema manycore. Por otro lado, se implementó un interfaz

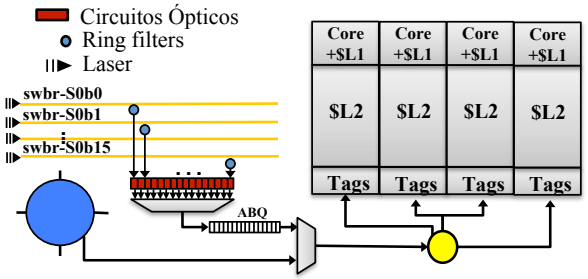


Fig. 5: Principales componentes en el receptor de los mensajes broadcast atómicos. ABQ es el buffer para encolar temporalmente los mensajes ACN transmitidos a través de la PhotoBNoC. Nótese cómo hace falta un multiplexor para que las L2s puedan recibir mensajes de la malla 2D eléctrica o de la PhotoBNoC. Usamos una política imparcial round-robin para transmitir los dos tipos de mensajes al concentrador local.

TABLA III: Aplicaciones evaluadas y configuración.

Suite	Aplicaciones	Entrada
SPLASH-2	cholesky (CH)	tk29.O matrix
	fft	4M complex points
PARSEC	blackscholes (BLK)	sim_large
	fluidanimate (FLU)	sim_large
	swaptions (SW)	sim_large
NPB	cg, bt, is	large
AIB	Kmeans (KM)	100 clusters
MANTEVO	hpccg (HPC)	[100,100,100]
UHPC	mdynamics (MDY)	water_xlarge.tpr
	schock (SCK)	large

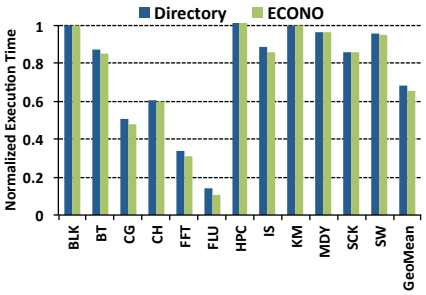
a la herramienta DSENT con el fin de estimar la potencia disipada así como el sobrecoste (overhead) de área de las dos arquitecturas NoC: malla 2D concentrada y PhotoBNoC.

Para la evaluación se utilizaron aplicaciones multihilo de los siguientes conjuntos de benchmarks: NPB [19], Splash2 [20], PARSEC [21], AIB [22], MANTEVO [23] y UHPC [24]. En particular, se seleccionaron las 12 aplicaciones detalladas en la Tabla III que ofrecían mayor escalabilidad para nuestro sistema de 256 cores. Para la evaluación experimental asumimos una simulación con calentamiento de cachés y un número limitado de instrucciones (15×10^7) que constituyen la fase de ejecución paralela.

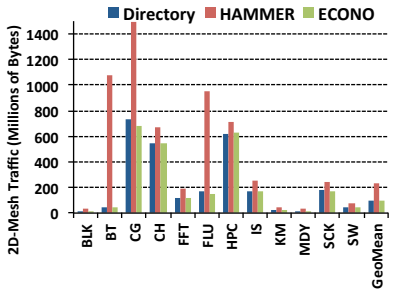
VI. Resultados Experimentales

A. Evaluación de Rendimiento

En términos del tiempo de ejecución (ver Figura 6a), se hace patente que ECONO lo reduce en términos generales con respecto al protocolo Hammer hasta un 34 %, mientras que Directory lo hace en un 31 %. Además, como se puede observar, ECONO mejora ligeramente el rendimiento de Directory en algunos benchmarks (BT, CG, FFT, FLU e IS). En términos del tráfico generado en la NoC eléctrica por los tres protocolos (Figura 6b), el protocolo Hammer genera hasta 2.3 veces más tráfico de media que Directory y ECONO. Esto es debido a que en cada evento de coherencia se han de enviar tantos mensa-



(a) Tiempos de ejecución normalizados respecto a Hammer.



(b) Tráfico total en la NoC malla 2D.

Fig. 6: Análisis preliminar de los protocolos ECONO, Hammer y Directory.

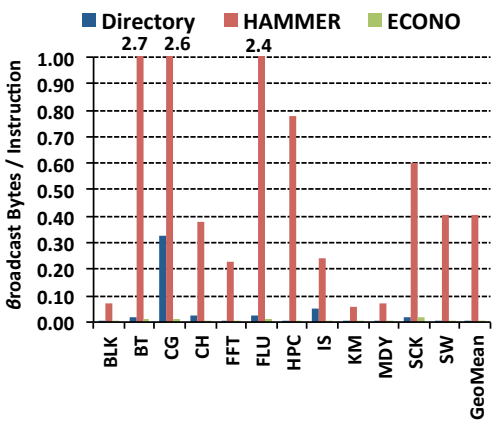


Fig. 7: βroadcast bytes por instrucción.

jes como cachés privadas de último nivel hay. Por el contrario, ECONO muestra un comportamiento similar a Directory, el cual hace un seguimiento de los propietarios y los compartidores de bloque, de forma que se envían los mensajes justos. Aunque ECONO no mantiene información de directorio y precisa de broadcasts, el envío de estos mensajes a través de la PhotoBNoC dedicada reduce en gran manera el tráfico en la red principal.

Para entender las diferencias de rendimiento entre ECONO y los protocolos de referencia explicados anteriormente, examinamos la actividad que realiza cada protocolo para mantener la coherencia. Para ello nos fijamos en dos escenarios: por un lado, el número de mensajes de broadcast, o más específicamente la cantidad de bytes enviados en broadcast por instrucción (βPI), y por el otro, el tipo de acciones de coherencia según los tipos de mensajes de broadcast: Invalidation, Fwd_Write y Fwd_Read.

En el primer escenario, representado en la Figura 7, se observa que ECONO tiene un βPI cercano a cero, ya que usa un único mensaje de broadcast a través del uso de la PhotoBNoC, de forma que se puede llegar a todas las cachés L2 ahorrando una gran cantidad de bytes de broadcast. Hammer en cambio puede llegar a los 2.7 βPI. En cuanto a cómo se ve afectado el rendimiento según los βPI alcanzados por cada protocolo, por un lado, tenemos que con valores de βPI mínimos, el rendimiento apenas varía, como puede verse en el comportamiento de los benchmark BLK y KM (muy bajo βPI, menos de 0.06), que muestra que apenas hay una mejora en el rendimiento de ECONO y Directory frente a Hammer (Figura 6a). Por otro lado, diferencias significativas en los βPI de cada protocolo, es decir, cambios importantes en la dinámica de coherencia, producen un impacto sobre el rendimiento y en la demanda de tráfico de red. ECONO alcanza mejoras de rendimiento y ahorro en el tráfico de la red malla 2D similares a las de Directory. Analizamos a continuación el resto de los benchmarks, y justificamos cómo ECONO puede alcanzar incluso mejor rendimiento que Directory.

BT, CG y FLU muestran los mayores niveles de βPI. En BT, aunque las diferencias entre Hammer y Directory son muy significativas, esto no se traduce en una reducción importante en el tiempo de ejecución, lo cual se debe a cargas de trabajo desbalanceadas que hacen que solo 22 de los 256 cores estén activos el 90 % del tiempo. En CG y FLU, sí se percibe una gran mejora en el rendimiento, con diferencias entre ambos. En CG el mantenimiento de la coherencia es más costoso, pues se envían más mensajes de broadcast debido a que el número de compartidores es superior al de FLU. En FLU, más del 35 % son de tipo FWD_Read y FWD_Write, que son mensajes punto a punto, y que Hammer resuelve mediante broadcast. A este porcentaje hay que añadir los mensajes Invalidation, enviados a un único propietario con copia. En CG en cambio, menos del 3 % pertenecen a las dos primeras categorías, por lo que el efecto negativo de la broadcast en Hammer se hace notar menos que en FLU.

HPC presenta mejoras insignificantes de rendimiento pese a que la diferencia de βPI entre Hammer y Directory es la mayor (200 veces). Este benchmark tiene un rendimiento muy pobre en cuanto a localidad espacial y temporal de su working set y con una tasa muy alta de fallos de LLC por lo que el porcentaje de tráfico de broadcast es cercano a cero, la mayoría de los mensajes son punto a punto tipo Fwd_Read o Fwd_Write (95 %) y el 98 % de las acciones están relacionadas con un solo poseedor. Aun así, el protocolo Hammer está por encima de los demás en tráfico de red debido a su uso de broadcast, que es más costoso.

SCK e IS obtienen ventajas moderadas en rendimiento cuando se usa Directory. En el caso de SCK, el 94 % de los mensajes son de la categoría Fwd_Read, que implica un solo mensaje en Directory y broadcast en Hammer. Sin embargo, esto no

TABLA IV: Número de componentes nanofotónicos en la red PhotoBNoC. Se asume: 8 Gbit/segundo/-lambda; 16 longitudes de onda/waveguide/dirección y 1 lambda/canal SWBR. LD = Lambdas, MD = Modulators, FL = Filters, WG = Waveguides. ADF = Área dispositivo nanofotónico en mm^2 con anillos de radio de $10\mu m$, separación de $410\mu m$ entre waveguides y un área de floorplan de 256 mm^2. Debido a los 4 segmentos de los que está compuesto cada canal PhotoBNoC (Figura 1), se consideran longitudes de waveguide de 15 mm, 33 mm, 51 mm y 69 mm.

LD	MD	FL	WG	ADF (% Área total chip)
64	64	1024	4	2.352 (0.91)

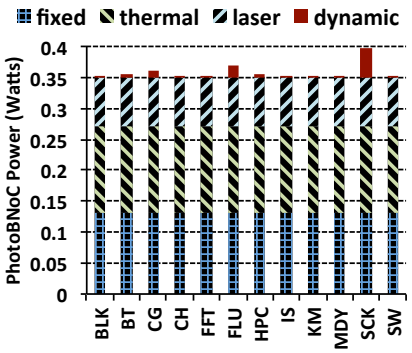
causa mucha degradación del rendimiento, ya que las cachés que no mantienen copia del bloque involucrado en la acción de coherencia, simplemente ignoran el mensaje. En contraste, en IS hay una gran cantidad de mensajes de invalidación (99 %) con una cantidad significativa de compartidores, lo cual impide que Directory obtenga una mejora sustancial de rendimiento respecto a Hammer. De forma similar SW y MDY muestran pobres mejoras de rendimiento cuando se comparan Directory y Hammer. La causa es que son cargas de trabajo de intensivas en cálculo que apenas generan tráfico de red.

Los dos últimos benchmark, CH y FFT, muestran grandes mejoras en el rendimiento cuando se comparan Directory y Hammer. La causa es que se llevan a cabo un gran número de operaciones de sincronización que involucran miles de operaciones de lectura, modificación y escritura de variables globales, y dichas operaciones son manejadas de una forma mucho más eficiente por Directory. En Hammer, estas operaciones implican broadcast, lo que incrementa la latencia de las operaciones de sincronización y da lugar a un bajo IPC. Esto explica las enormes mejoras que se producen al utilizar el protocolo Directory.

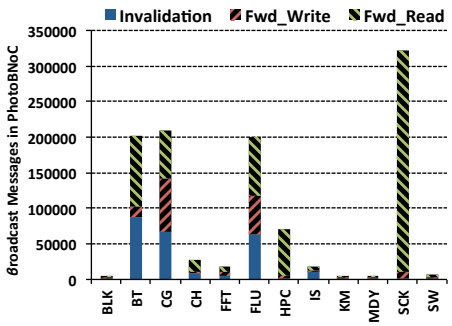
Finalmente, ECONO además de mostrar mejoras de rendimiento similares a las de Directory, en algunos benchmarks lo sobrepasa. En concreto, en BT, CG, FFT, FLU e IS los tiempos de ejecución se reducen hasta un 5 % de media. Hay dos razones para ello: la primera es que, aunque ECONO utiliza broadcast, la transmisión a través de la PhotoBNoC de un único mensaje de broadcast precisa de una latencia fija de 12 ciclos para llegar a todas las cachés L2. Sin embargo, la ruta más larga para Directory es de 46 ciclos, por lo que ECONO optimiza la latencia para todas las caches alejadas más de 12 ciclos (75 % del total). La segunda es que ECONO no precisa de mensajes ACK para las acciones de coherencia con Invalidación, por lo que las LLC no tienen que esperar hasta que reciben todas las confirmaciones y también se ahorra tráfico en la red de interconexión principal (malla 2D), lo que reduce también la contención en los recursos de la red.

B. Análisis de Potencia y Energía

La Figura 8a muestra el análisis de la potencia disipada por la PhotoBNoC y que se corresponde con el



(a) Disipación de potencia en PhotoBNoC.



(b) Tráfico total en la NoC malla 2D.

Fig. 8: Mensajes βroadcast en PhotoBNoC.

número de componentes basados en fotónica de silicio requeridos y listados en la Tabla IV. Este análisis consta de potencia fija (potencia estática y de reloj), térmica (potencia para estabilización térmica), láser (potencia de láser) y potencia dinámica (potencia dependiente del tráfico de datos). Nuestro diseño PhotoBNoC es muy eficiente en términos de potencia, pues no alcanza en ningún caso los 0.4 Watts. Esto se corresponde con una energía media por bit de 625 fJ/b, y se encuentra dentro de los rangos típicos considerados en redes basadas en fotónica de silicio: entre 100 fJ/b y 2 pJ/b, valores correspondientes que van desde los diseños más agresivos a más conservadores [13].

Para entender mejor los resultados de potencia obtenidos, la Figura 8b muestra el número de mensajes de broadcast transmitidos a través de la PhotoBNoC para cada benchmark. La parte de potencia dinámica depende del número de mensajes de broadcast y del tiempo de ejecución. Se puede observar que el benchmark SCK genera la mayor disipación de potencia dinámica, pues demostró tener el mayor número de mensajes de broadcast y el IPC más alto (menor tiempo de ejecución), y su valor es de solo 0.04 Watts.

Sumando la potencia de la interconexión PhotoBNoC al total de la red malla 2D, incluyendo potencia estática, dinámica y de reloj, la potencia de la primera representa tan solo un 6.37 % del total. Además, el diseño de la red principal malla 2D es también muy eficiente, ya que se disipan menos de 9 Watts en el peor caso (IS). Es más, de media, el diseño de red con PhotoBNoC para ECONO, obtiene una potencia de red comparable a la de Hammer, solo que en

TABLA V: Requisitos de espacio de almacenamiento para diferentes protocolos tipo Directory respecto a la PhotoBNoC requerida por ECONO para un sistema de 256 cores.

FM	HC	SCD	ECONO
59.18 %	24.22 %	12.50 %	6.10 %

	FM vs ECONO	SCD vs ECONO
	9.71×	2.04×

algunos benchmark (CH, FFT y FLU), al alcanzar ECONO un IPC mucho más alto, la potencia disipada en general es mayor en ECONO y en Directory que en Hammer.

Respecto al consumo de energía, teniendo en cuenta y sumando todas las fuentes de consumo, ECONO presenta mejoras en la energía similares a Directory respecto a Hammer, y para los benchmark CG y FLU, se obtuvieron resultados de consumo de energía hasta un 4 % menor.

C. Coste de Área On-Chip

Como se ha visto, ECONO obtiene unos resultados ligeramente mejores que Directory en cuanto a rendimiento y consumo de energía. Aparte de estos beneficios, ECONO ahorra en área on-chip, ya que no precisa almacenar información de coherencia acerca de propietarios y compartidores de bloque. Sin embargo, al utilizarse la PhotoBNoC para la realización eficiente de las operaciones de broadcast, es necesario analizar el área on-chip que precisa dicha interconexión. Además es interesante estudiarla en relación a distintas implementaciones del protocolo Directory.

Para la estimación del área, asumimos los mismos componentes de fotónica de silicio que en [26]: waveguides de modo simple, con una separación de 4μm para minimizar la cantidad de cruces entre las waveguides (lo que aumenta las pérdidas de señal teniendo que compensarlo con mayor consumo de potencia de los láseres). El diámetro de los ring resonators (modulators y filters) son de 10 μm. La Tabla IV muestra el número de componentes y el área ocupada por ellos en la PhotoBNoC. Los cálculos realizados para determinar el área de los dispositivos fotónicos de la PhotoBNoC dan como resultado un área de 2.3 mm^2, lo que representa un 0.9 % del total del área del chip.

Para entender la magnitud del coste de área requerido por PhotoBNoC, éste se compara con tres tipos de representaciones de Directory cuyos requisitos de área para nuestro sistema de 256 cores se muestran en la Tabla V, a saber, el esquema tradicional Full-Map con vector de bits (FM), el esquema jerárquico [28] (HC) y el esquema [27] SCD. Este coste de área se da en porcentaje del área requerida al agregar todas las cachés L2 basándonos en el estudio llevado a cabo en [27]. La última columna es para ECONO. En este caso, los porcentajes derivan del coste de área on-chip considerando el sistema PhotoBNoC. La tabla muestra como el protocolo ECONO tiene un coste de área 9.71 veces menor que Full-Map y reduce 2 veces el área respecto al mucho más eficiente y escalable protocolo SCD. De este estudio podemos afirmar que ECONO es también el protocolo más escalable en términos de coste de área.

VII. Trabajo Relacionado

Las bondades de los protocolos orientados a broadcast han sido aprovechadas por propuestas como ATAC [29], que utiliza enlaces con tecnología fotónica para optimizar las transmisiones de broadcast. ATAC opera de la misma manera que un directorio limitado convencional, pero cuando se desborda la lista de compartidores y ante una invalidación se recurre a un broadcast muy eficiente sobre la red óptica. Por otro lado, en Atomic Coherence [25], se simplifican las acciones de mantenimiento de la coherencia a algo muy próximo a las que se tomarían en un protocolo basado en bus. Esta atomicidad evita las condiciones de carrera gracias al uso de un mutex óptico. ECONO, sin embargo, habilita las condiciones de carrera desde los cores solicitantes al nodo home y la serialización solo se emplea para el envío de notificaciones de coherencia de forma atómica, incrementando así la concurrencia. Sin embargo estas notificaciones son pocas de forma que se ahorra tráfico y energía.

Existe una nueva serie de avances tecnológicos que buscan limitar los problemas de energía y sobrecoste de área de las redes eléctricas on-chip. Así, se ha demostrado que los enlaces nanofotónicos de silicio, utilizados en este trabajo para implementar la PhotoBNoC de manera eficiente, son mucho más rápidos en comunicación global, tienen un menor consumo de energía dinámica, y ofrecen una mayor densidad de ancho de banda que los enlaces eléctricos convencionales. Sin embargo, estos enlaces nanofotónicos pueden presentar un gran coste de área on-chip y gran disipación de potencia procedente de los láseres utilizados, así como la proveniente de la estabilización térmica de los ring resonators que utiliza. En este trabajo, se ha diseñado PhotoBNoC teniendo en cuenta estos costos y se ha conseguido obtener un diseño de muy bajo costo en área que al igual que en [30], [31], se basa en un diseño de NoC híbrido donde una subred fotónica muy ligera es únicamente utilizada para la difusión eficiente de mensajes ACN de forma rápida y atómica.

VIII. Conclusiones

Este trabajo propone el diseño de una red broadcast nanofotónica eficiente llamada PhotoBNoC. Dicha red es el substrato para el protocolo de coherencia ECONO. En concreto, PhotoBNoC garantiza transmisiones rápidas y atómicas para la difusión de mensajes de coherencia. Para ello emplea canales nanofotónicos SWBR segmentados según distancia para ahorrar potencia de láser y reducir coste de área en el chip. Las transmisiones atómicas se garantizan a través de los canales SWBR, así como arbitraje local y buffers con colas en los nodos destino.

Los resultados para un multiprocesador de 256 cores muestran que la combinación ECONO+PhotoBNoC representa una solución

de alto rendimiento y más eficiente en energía y área al problema de coherencia de cachés para futuros manycores. En concreto, ECONO+PhotoBNoC puede alcanzar (e incluso mejorar) el rendimiento del protocolo Directory, no escalable y con grandes requisitos de área. Al mismo tiempo, mejora el rendimiento con respecto a Hammer sin perder su ventaja respecto al área. En cuanto a la disipación de potencia, nuestro estudio desvela que PhotoBNoC apenas tiene impacto sobre ella y que ECONO+PhotoBNoC puede conseguir los resultados de Directory en términos de energía.

Agradecimientos

Este trabajo ha sido financiado por el Ministerio de Economía y Competitividad (MINECO) y la Comisión Europea FEDER mediante el proyecto "TIN2016-78799-P", y por la Fundación Séneca-Agencia de Ciencia y Tecnología de la Región de Murcia a través del proyecto "19295/PI/14".

Referencias

[1] A. Sodani, R. Gramunt, J. Corbal, H.-S. Kim, K. Vinod, S. Chinthamani, S. Hutsell, R. Agarwal, Y.-C. Liu, Knights landing: Second-generation intel xeon phi product, IEEE Micro 36 (2) (2016) 34–46.

[2] B. Bohnenstiehl, A. Stillmaker, J. Pimentel, T. Andreas, B. Liu, A. Tran, E. Adeagbo, B. Baas, A 5.8 pj/op 115 billion ops/sec, to 1.78 trillion ops/sec 32nm 1000-processor array, in: Proc. of 2016 Symposium on VLSI Technology and Circuits, 2016, pp. 1–2.

[3] A. Ahmed et al., AMD Opteron Shared Memory MP Systems, in: Proc. HotChips Symposium, 2002.

[4] L. A. Barroso et al., Piranha: A Scalable Architecture Based on Single-Chip Multiprocessing, in: Proc. IEEE International Symposium on Computer Architecture, 2000.

[5] M. B. Taylor et al., The Raw Microprocessor: a Computational Fabric for Software Circuits and General-Purpose Programs, IEEE Micro 22(2) (2002) 25–35.

[6] M. Lodde, J. Flich, M. E. Acacio, Heterogeneous noc design for efficient broadcast-based coherence protocol support, in: Proc. of the 6th IEEE/ACM International Symposium on Networks-on-Chip, 2012, pp. 59–66.

[7] J. L. Abellán, A. Ros, J. Fernández, M. E. Acacio, Econo: Express coherence notifications for efficient cache coherency in many-core cmps, in: Embedded Computer Systems: Architectures, Modeling, and Simulation (SAMOS XIII), 2013 International Conference on, IEEE, 2013, pp. 237–244.

[8] T. Krishna, A. Kumar, L.-S. Peh, J. Postman, P. Chiang, M. Erez, Express virtual channels with capacitively driven global links, IEEE Micro 29 (4) (2009) 48–61.

[9] Mellanox Technologies, Seventy two core processor soc with 8x 10gb ethernet ports, pcie and networking offloads, `http://www.mellanox.com/related-docs/prod_multi_core/PB_TILE-Gx72.pdf` (2016).

[10] S. Beamer, et al., Re-architecting DRAM memory systems with monolithically integrated silicon photonics, in: 37th Annual International Symposium on Computer Architecture, ISCA 2010, 2010, pp. 129–140.

[11] A. Joshi, et al., Silicon-photonic clos networks for global on-chip communication, in: Networks-on-Chip, 2009. NoCS 2009. 3rd ACM/IEEE International Symposium on, 2009, pp. 124–133.

[12] L. Ramini, D. Bertozzi, L. Carloni, Engineering a bandwidth-scalable optical layer for a 3d multi-core processor with awareness of layout constraints, in: Proc. International Symposium on Networks-on-Chip (NOCS), 2012, pp. 185–192.

[13] J. L. Abellán, C. Chao, A. Joshi, Electro-photonic noc designs for kilocore systems, ACM Journal on Emerging Technologies in Computing, 2016, pp. 1–25.

[14] M. Georgas, et al., A Monolithically-Integrated Optical Receiver in Standard 45-nm SOI, IEEE Journal of Solid-State Circuits 47.

[15] B. Moss, et al., A 1.23pj/b 2.5gb/s monolithically integrated optical carrier-injection ring modulator and all-digital driver circuit in commercial 45nm soi, in: ISSCC, 2013, pp. 18–20.

[16] J. S. Orcutt, et al., Nanophotonic integration in state-of-the-art cmos foundries, Opt. Express.

[17] D. J. Sorin et al., A Primer on Memory Consistency and Cache Coherence, Synthesis Lectures on Computer Architecture#16, 2011.

[18] T. E. Carlson, et al., Sniper: Exploring the level of abstraction for scalable and accurate parallel multi-core simulations, in: Proc. SC, 2011, pp. 1–12.

[19] D. Bailey, et al., The NAS parallel benchmarks, Tech. Rep. RNR-94-007 (mar 1994).

[20] S. C. Woo, et al., The SPLASH-2 programs: characterization and methodological considerations, in: Proc. ISCA, 1995, pp. 24–36.

[21] C. Bienia, et al., The PARSEC benchmark suite: Characterization and Architectural Implications, in: Proc. PACT, 2008, pp. 72–81.

[22] W. keng Liao, Parallel k-means data clustering. (2005). URL `http://users.eecs.northwestern.edu/~wkliao/Kmeans/index.html`

[23] M. A. Heroux, et al., Improving performance via mini-applications, Sandia National Laboratories, Tech. Rep.

[24] D. Campbell, et al., Ubiquitous high performance computing: Challenge problems specification, Tech. Rep. HR0011-10-C-0145, Georgia Institute of Technology (2012).

[25] D. Vantrease et al., Atomic Coherence: Leveraging Nanophotonics to Build Race-Free Cache Coherence Protocols, in: Proc. IEEE International Symposium on High-Performance Computer Architecture, 2011.

[26] J. L. Abellán, A. Coskun, A. Gu, W. Jin, A. Joshi, A. B. Kahng, J. Klamkin, C. Morales, J. Recchio, V. Srinivas, T. Zhang, Adaptive tuning of photonic devices in a photonic noc through dynamic workload allocation, Computer- Aided Design of Integrated Circuits and Systems, IEEE Transactions on 36(5) (2017) 801–814. doi:10.1109/TCAD.2016.2600238.

[27] D. Sanchez and C. Kozyrakis., SCD: A Scalable Coherence Directory with Flexible Sharer Set Encoding, in: Proc. IEEE International Symposium on High-Performance Computer Architecture, 2012.

[28] S. Guo et al., Hierarchical Cache Directory for CMP, Journal of Computer Science and Technology 25(2) (2010) 246-256.

[29] G. Kurian et al., ATAC: A 1000-Core Cache-Coherent Processor with On- Chip Optical Network, in: Proc. IEEE International Conference on Parallel Architectures and Compilation Techniques, 2010.

[30] Z. Li, et al., Spectrum: A hybrid nanophotonic-electric on-chip network, in: Proceedings of the 46th Annual Design Automation Conference, DAC '09, 2009, pp. 575–580.

[31] S. Bahirat, S. Pasricha, Meteor: Hybrid photonic ring-mesh network-on- chip for multicore architectures, ACM Trans. Embed. Comput. Syst. 13 (3s) (2014) 116:1–116:33.