



Enzo Rucci

Doctor en Ciencias Informáticas,
Facultad de Informática, UNLP
erucci@lidi.info.unlp.edu.ar

FPGAs: ¿los procesadores del futuro?

Las Field Programmable Gate Arrays (FPGA) son circuitos integrados reconfigurables compuestos de interconexiones programables que combinan bloques lógicos programables, de memoria embebida y de procesamiento de señales digitales, entre otros. Dicho en palabras más simples, una FPGA es "hardware programable". A diferencia de las CPUs y las GPUs, los recursos de una FPGA pueden ser configurados para crear pipelines de instrucciones específicos para el problema a resolver. Si bien operan a frecuencias de reloj más bajas y tienen picos de rendimiento inferiores, la posibilidad de adaptar el hardware para cada aplicación particular le permite a las FPGAs lograr mejores tasas de rendimiento en la mayoría de los casos. En forma adicional, suelen ser más eficientes desde el punto de vista energético ya que no hay desperdicio en los recursos de silicio.

Desafortunadamente no todas son ventajas para las FPGAs. Entre las desventajas se puede mencionar un costo alto de programación y de desarrollo. En general, la programación de FPGAs se realiza a través de lenguajes de descripción de hardware (HDL, por sus siglas en inglés), como Verilog o VHDL. Estos lenguajes suelen ser tediosos, propensos a errores y requieren mantener una noción explícita del paso del tiempo. En forma adicional, los tiempos de compilación y síntesis pueden requerir varias horas, dependiendo del diseño. A estos inconvenientes, debemos sumarle un costo de adquisición más elevado que otros aceleradores como GPUs y Xeon Phi's, aunque es importante resaltar que este costo puede ser amortizado de mejor manera por las FPGAs debido a su mayor eficiencia energética.

Desde sus inicios, las FPGAs han evolucionado continuamente

incrementando sus recursos disponibles e incorporando características como estándares para la interconexión de redes y E/S de alta velocidad. Aunque tradicionalmente fueron usadas para procesamiento de señales digitales, en la actualidad sus dominios de aplicación se encuentran en continua expansión. Existen múltiples ejemplos que reflejan esta tendencia y es posible mencionar algunos de ellos. Recientemente, Microsoft ha revelado que fue capaz de duplicar la productividad de su buscador Bing mediante la integración de FPGAs a sus centros de datos. En forma similar, el buscador chino Baidu está utilizando estos dispositivos para acelerar aplicaciones de aprendizaje automático como reconocimiento de imágenes y de voz al mismo tiempo que planea usarlos para el desarrollo de autos autónomos. Por su parte, la Organización Europea para la Investigación Nuclear (CERN, por sus siglas en francés) ha reportado que evalúa el uso de FPGAs para el cálculo de ecuaciones complejas y de otras aplicaciones para los que actualmente emplea CPUs y GPUs. Por último, grandes empresas prestadoras de servicios de cloud (como Amazon y Alibaba) ofrecen hoy en día la posibilidad de usar FPGAs en sus plataformas.

El crecimiento de las FPGAs no es casual. En sentido opuesto, es posible identificar dos tendencias claras que permiten explicar este fenómeno. La primera comprende la creciente integración de FPGAs con CPUs debido a la consolidación del uso de aceleradores en la comunidad de cómputo de alto rendimiento (HPC, por sus siglas en inglés) como una manera de mejorar las prestaciones al mismo tiempo que se limita el consumo de potencia. En particular, importantes empresas de procesadores han establecido diferentes

acuerdos con los principales fabricantes de FPGAs para desarrollar arquitecturas híbridas CPU-FPGA. IBM ha anunciado una asociación estratégica junto a Xilinx para lograr mayor rendimiento y eficiencia energética en aplicaciones a través de la aceleración con FPGAs sobre sistemas basados en IBM Power. Por su parte, Intel ha adquirido recientemente Altera y planea combinar sus FPGAs con los procesadores Xeon para lograr un dispositivo integrado y altamente personalizado. La segunda tendencia consiste en reducir el tradicional costo elevado de programación y desarrollo. Desde hace varios años, tanto Altera como Xilinx se encuentran trabajando en herramientas de alto nivel que faciliten la programación de estos aceleradores y reduzcan los tiempos de desarrollo; en particular, a través del estándar OpenCL para programación heterogénea. Si bien los expertos probablemente continuarán usando HDLs, la posibilidad de poder usar OpenCL permitirá que el número de programadores que empleen FPGAs crezca más rápidamente.

El mercado de los aceleradores para HPC y centros de datos se encuentra dominado por las GPUs con una competencia cada vez más fuerte de parte de los Xeon Phi. En esta disputa, las FPGAs buscan ganar terreno con sus tasas altas de eficiencia energética como principal fortaleza. Por su parte, Intel ha pronosticado que el 30% de los centros de datos estarán equipados con FPGAs para 2020, siendo utilizadas para tareas como aprendizaje automático, analítica de datos, cifrado y cómputo científico, entre otras. En ese sentido, gigantes como Microsoft y Baidu ya han dado el primer paso. Si Intel acierta en sus predicciones, los próximos años marcarán grandes cambios para HPC y los centros de datos.