1. En los modelos que tiene como procesador una G80. Si ejecutamos XXXXXX cuyos bloques de hilos consumen cada uno 7Kb XXXXX compartida. ¿Cuántos bloques SP se podrían utilizar a la XXXX?
   1. Se ejecutarían 3 bloques(SP) en cada SM
   2. Se ejecutarían 2 bloques (SP) en cada SM
   3. Se ejecuta un único bloque (SP) por SM
   4. Ninguna de las anteriores.
2. Suponiendo que tenemos un kernel que tiene 64 blocks XXXX bloque tiene un medio wraps de hilos. ¿Cuál es la cantidad XXXXX del grid? ¿Qué posición ocuparan el Hilo 12 del bloque 6?
   1. 768,205
   2. 1024,108
   3. 2048,204
   4. Ninguna de las anteriores.
3. ¿Cuál de las siguientes configuraciones no pertenecen a la clasificacion de Hwang-Briggs?
   1. Computadores pipeline.
   2. Computadores Multithreading.
   3. Computadores matriciales.
   4. Sistemas multiprocesador.
4. ¿Cuál es el ancho de banda máximo de la tarjeta GTX 780?

|  |  |
| --- | --- |
| Especificaciones de la GPU |  |
| Nucleos Cuda | 2880 |
| Frecuencia de reloj normal | 875 |
| Frecuencia accelerada | 928 |
| Tasa de relleno de texturas | 210 GigaTexels/s |
| Especificaciones de la memoria |  |
| Frecuencia de la memoria (GBPS) | 7.0 |
| Cantidad de memoria | 3072 MB |
| Interfaz de memoria | 384-bit GDDR5 |

* 1. 672 Gbit/s
  2. 336 Gbit/s
  3. 336 GByte/s
  4. 84 GByte/s

1. Cuando se lanza un núcleo:
   1. Los bloques de hilos se distribuyen de manera aleatoria sobre distintos SM.
   2. Los bloques de hilos se distribuyen de manera secuencial sobre los distintos SM.
   3. Los bloques de hilos se parten en warps que se distribuyen de manera aleatoria sobre los distintos SM.
   4. Ninguna de las anteriores.
2. El ratio entre número de operaciones en coma flotante y de accesos a la memoria global limita el rendimiento ya que
   1. Hay un ratio optimo, valores por encima de ese ratio no mejorarían el rendimiento.
   2. Mejor rendimiento cuanto mas próximo es el ratio a 1.
   3. Menos rendimiento cuanto mayor es el ratio.
   4. Mayor rendimiento cuanto menor es el ratio.
3. ¿Cuál de las siguientes afirmaciones es falsa?
   1. El vertex shader es donde se transforma y se determina la posición final de cada vértice.
   2. Fragment Shadders realiza las transformaciones referentes a los fragmentos tras la operación realizada por el vértice shadder.
   3. La rasterización convierte cada polígono a una serie de fragmentos de tamaño pixel.
   4. El Cuilling es la técnica que eliminan los triángulos cuya cara no esta orientada hacia la cámara.
4. Cual de las siguientes afirmaciones es falsa respecto a las Compute Capabilities?
   1. Los cambios en la primera cifra implican cambios de generación mientras que la segunda implica una revisión.
   2. Las primeras GPUs de CUDA son de compute capabilities 1.0.
   3. La categorización incluye dos números. La primera cifra indica cambios de revisión y la segunda indica la version.
   4. Las GPU’s actuales pertenecen a las compute capabilities 3.0 y 3.5.
5. ¿Cuál de las siguientes afirmaciones es falsa respecto al compilador XXXX?
   1. Cuda compilador (NVCC) utiliza un preprocesador (cudafe) que divide la aplicación en la CPU y la GPU en partes, creando un archivo preprocesador C para la entrada GPU.
   2. El compliador Open64(nvopencc), que emite un lenguaje ensamblador hemos creado llamado GPU assembly.
   3. XXXX asigna los registros y la lista de las instrucciones de acuerdo XXXX chip en particular que se está utilizando (programación).
   4. XXXXXX fichero fuente conteniendo extensiones XXX.
6. Supongamos que un núcleo tiene bloques de 256 hilos, 16 instrucciones independientes de acceso a la memoria global por hilo y que cada hilo usa 8 registros y los accesos a la memoria global consumen 200 ciclos. ¿Cuántos bloques de un SM se pueden ejecutar con estas condiciones?
   1. 2 bloques del SM
   2. 3 bloques del SM
   3. 4 bloques del SM
   4. Ninguna de las anteriores.
7. ¿Cuál de las siguientes afirmaciones es falsa respecto a las diferentes configuraciones entre las G80, G200 y Fremi?
   1. La evolución de Cuda cores de cada una es, 128, 240 y 512.
   2. En la arquitectura de Fremi existe una memoria L2 cache de 768KB mientras que en la G80 y G200 no existe.
   3. La memoria compartida en G80 y G200 es de 32 KB y en la de Fremi puede ser configurable de 32KB o de 48KB.
   4. El ancho de direcciones en la G80 y G200 es de 32 bits y en la de Fremi es de 64 bits.