

DynamicGraphMultiProcessor??

Dynamic Graph Multi Processor ??

??

1. Etched GPU TFLOPS
2. + +
3. Etched H100 800 3.3%
Transformer Sohu FLOPS
90% GPU 30%
4. Scaling Law Scale Up
Scaling Law Scale Down

????

- AI
- Int8 pattern
- 2D =>
- GPU DRAM
- NoC Cache Fork/Join
NoC flow control
- NoC/Cache
- consistency corhenrency fence+sync
- 1. GMP fork/join
- 2. fence
- 3. fence
- IP cpu cache
IP IP
cache_hint
- DSA
 - SOC
 - NOC
-

2. launch
3. launch launch
4. launch sub graph graph group
"signal"
9. delay
1. data hzd
10. stall stall

??

1. Fork Join














1. fork
 - 1.
 2. join
 3. index
 4. launch ID ID IP memory mapping
2. fork load
3. join
 1. launch ID
2. VR L0.5 L0.5
3. VR L0 4
4. TR L0.5, L0.5
5. DTE transpose pad slice deslice
6. PU
- 7.

1.

Width (bits)	4	6	3	3	1	
Meaning	Reuse flags	Wait barrier mask	Read barrier index	Write barrier index	Yield flag	St cycl

2. Reuse flags 4 register cache
3. 6 barrier thread cuda
6
8. async_group
 1. async copy bulk
9. mbarrier fence
10. credit valid/ready
 - 1.
 - 2.
 - 3.

????

1.  die 
 1. 4T int8 
 2.  256MB DDR3 DRAM 1866 16bit  3.7GB/s 
 3. 
2. 
 1.  die 
 2. 
 3. die to die  serdes 

ISA

ctrl

1. 1. Fork(code_index)
 2. launch
 3. Join()

Scalar

1. scalar 

Vector

Tensor