Determinism of GPU solutions for AO real-time computing

l'Observatoire de Paris | LESIA
Laboratoire d'Études Spatiales et d'Instrumentation en Astrophysique

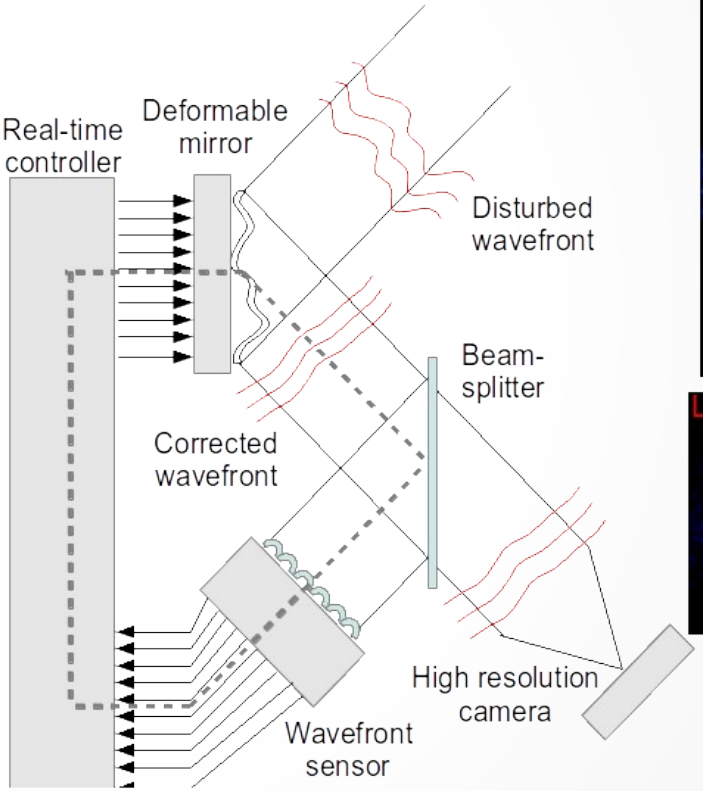
Durham University

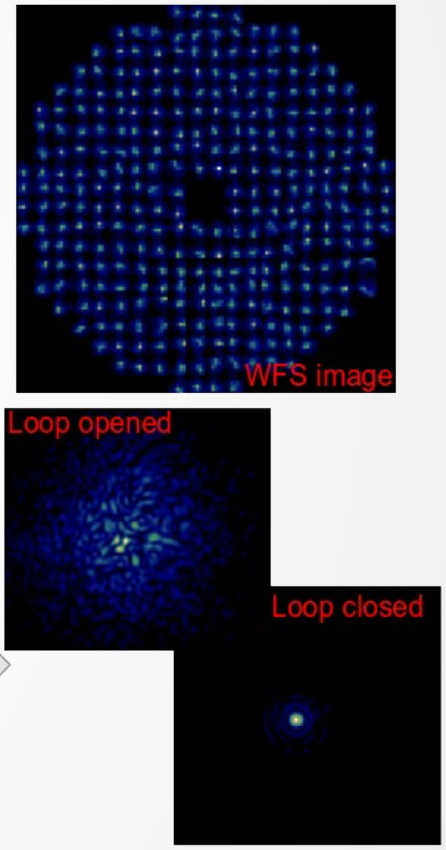
MICROGATE

PLDA

E-ELT

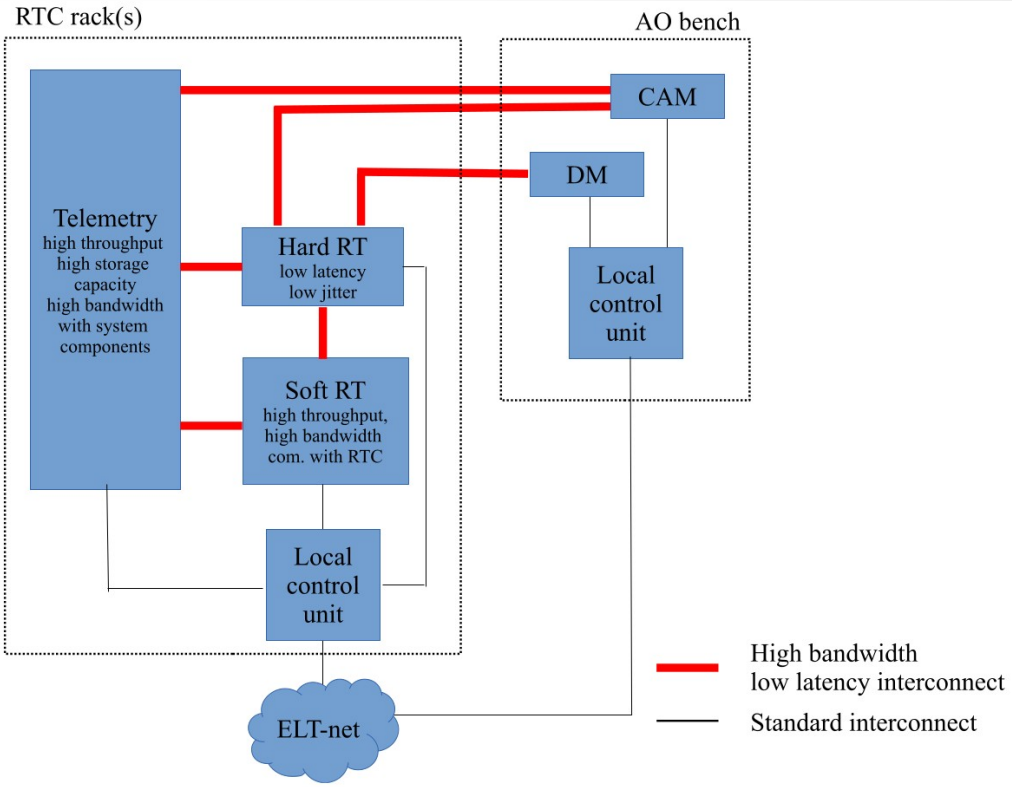
- AO RTC Architecture
  - Hard real time system (~1 kHz)
  - Big computation (5 TFLOPs)
  - Low latency
  - Maximum jitter : ~10%





Jitter

- Where is the jitter ?
  - Data transfer
  - Computation

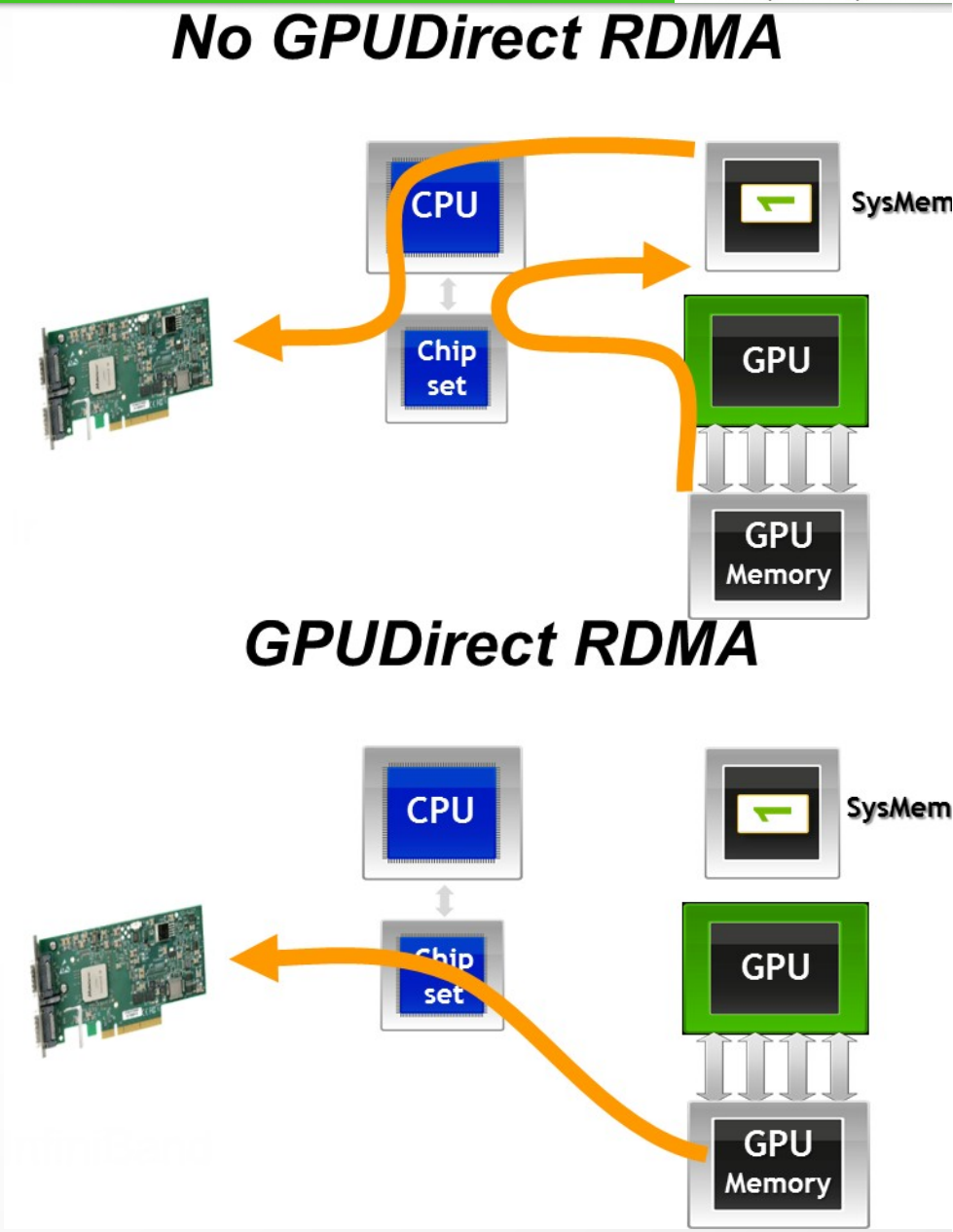


- Jitter with standard transfer and computation

Case	Pipeline	Time (jitter) (ms)
64x64 pixels 8x8 subpupils	copy only	33 (35)
	copy + compute	96 (63)
240x240 pixels 40x40 subpupils	copy only	204 (37)
	copy + compute	576 (57)

Data transfer

- Normal way
  - Main memory is a buffer
  - 2 copies by communication
  - CPU manage the communication
- GPUdirect RDMA (Remote Direct Memory Access)
  - No unnecessary copy
  - CPU only use for launching kernel

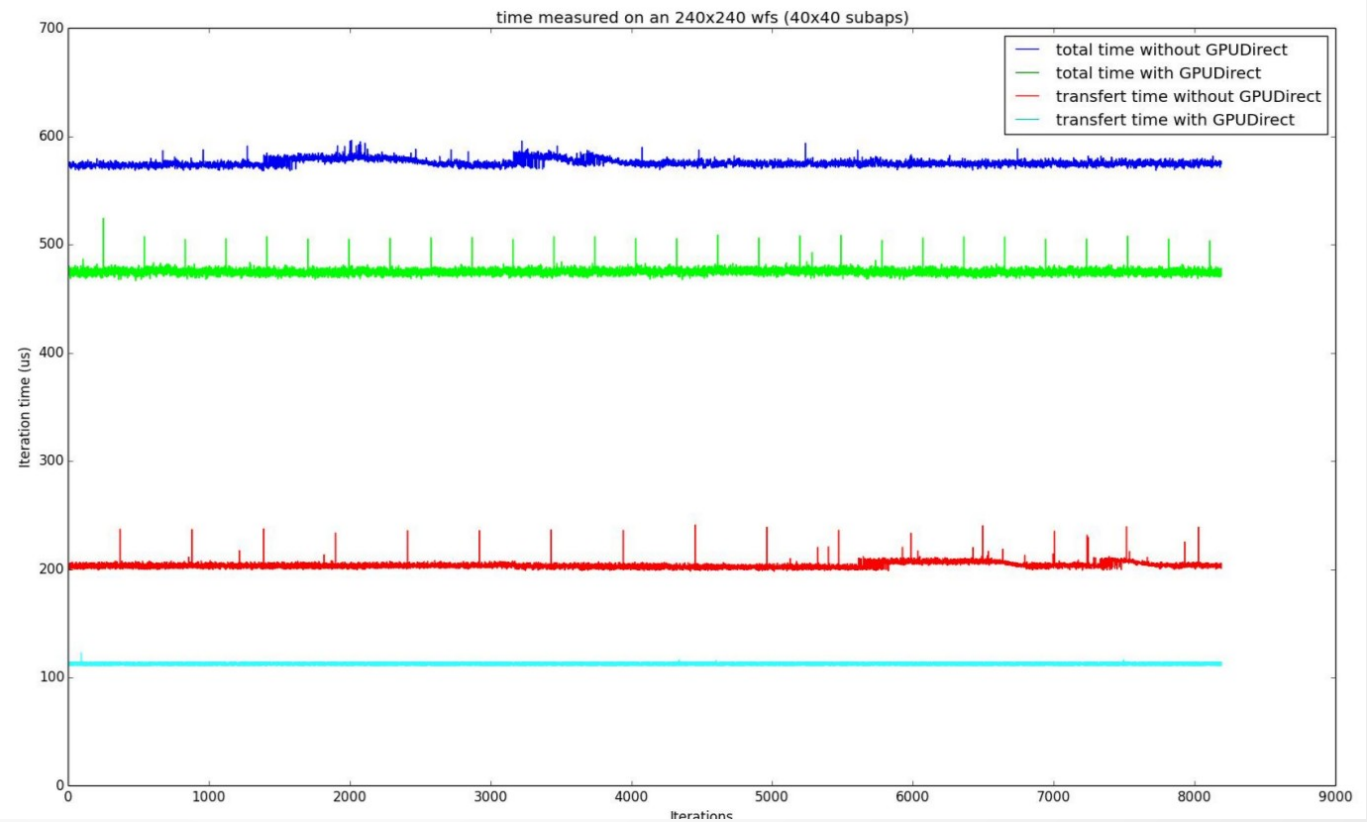


CREDIT : NVIDIA

Transfer result

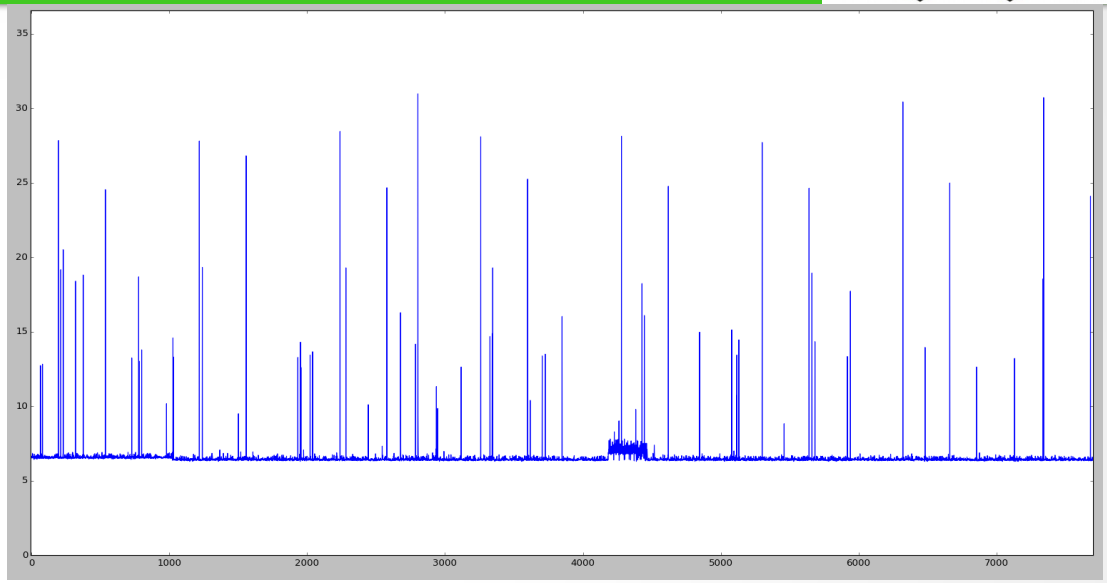
- GPUdirect
  - Reduce jitter during transfer to almost 0
  - Reduces the transfer time by 2

- But jitter still occurs during computations...



Computation

- Normal way
  - High jitter
  - Depends on CPU
  - Need a Real-Time OS



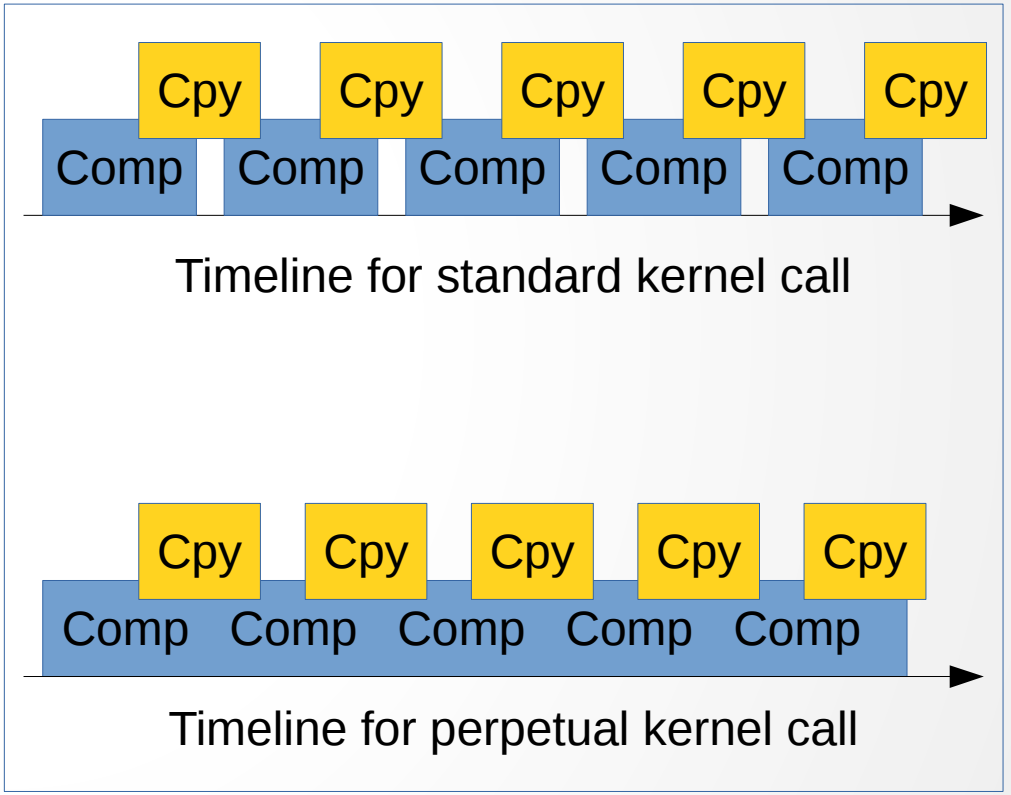
Time (in µs) for 8k empty kernel call
(average : ~6.5µs, peak : ~31µs)

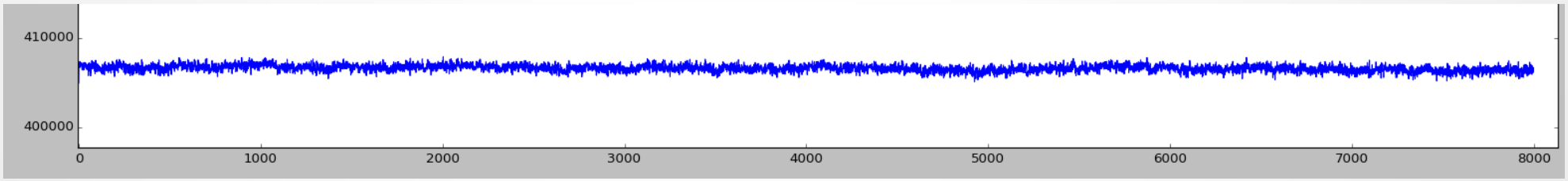
- Jitter with RDMA transfer and standard computation

Case	Pipeline	Time(jitter) (ms)
64x64 pixels 8x8 subpupils	copy only	12 (12)
	copy + compute	69 (59)
240x240 pixels 40x40 subpupils	copy only	112 (10)
	copy + compute	475 (50)

Perpetual kernel

- Pros
  - No scheduler
  - No additional cost
  - New features
    - Reduce computation
    - New synchronization features
- Cons
  - More complex implementation, test and debugging
  - Hardware dependent
  - Can't use any existing library





Clock cycle count for 8k iterations

What's next ?

- Implementation of RTC with perpetual kernel
- Integration with frame grabber
  - Test with pixel generator
  - Integration on the optical bench
  - Full loop profiling
- Study on floating point precision to reduce the number of GPU