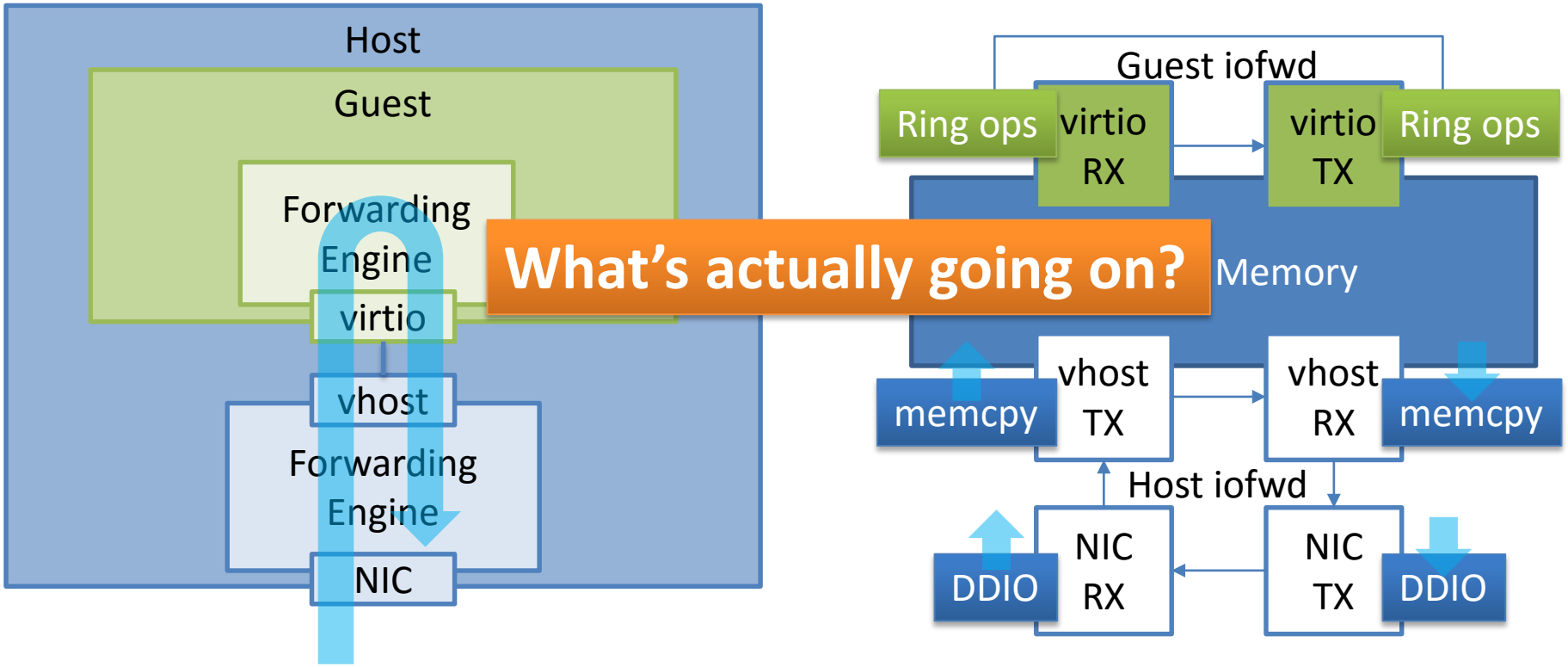
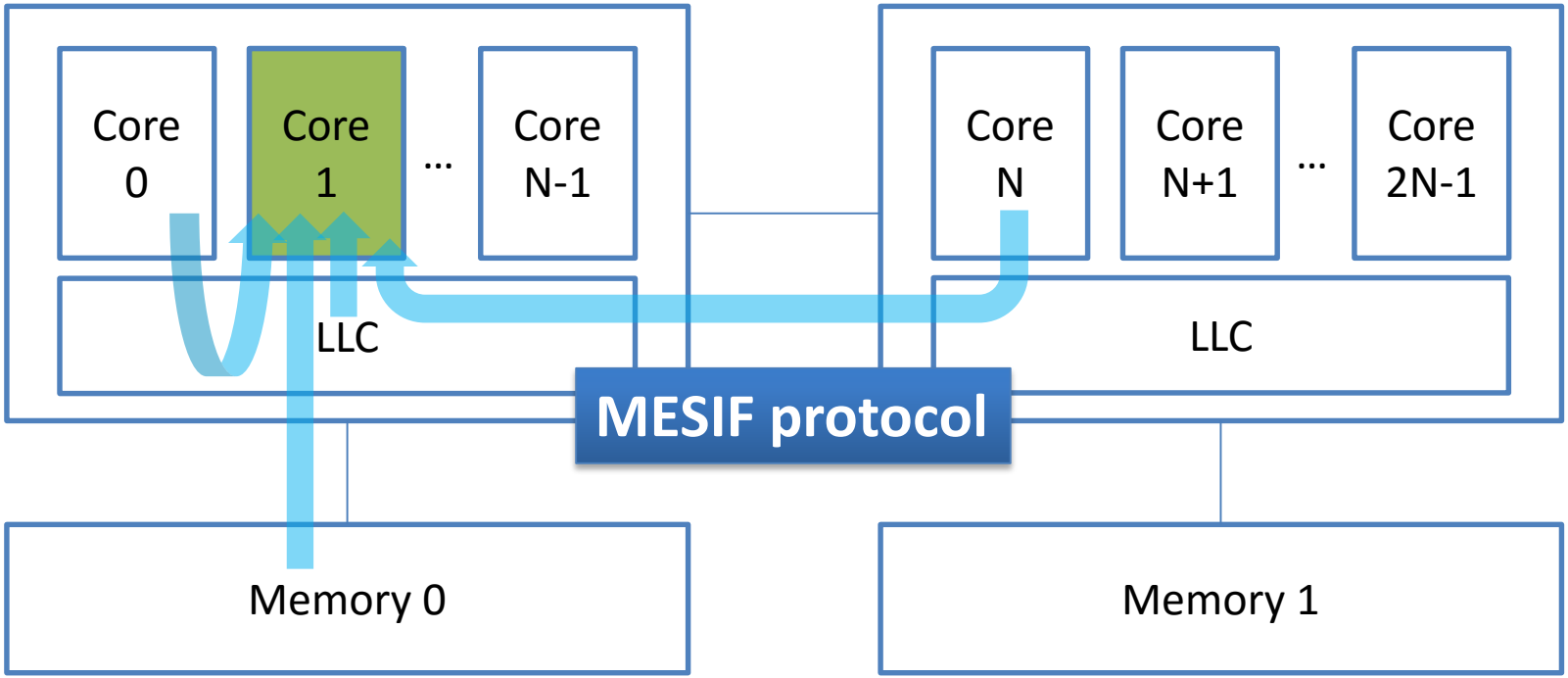
Topic: A Deep Dive into Memory Access

Company: Intel
Title: Software Engineer
Name: Wang, Zhihong

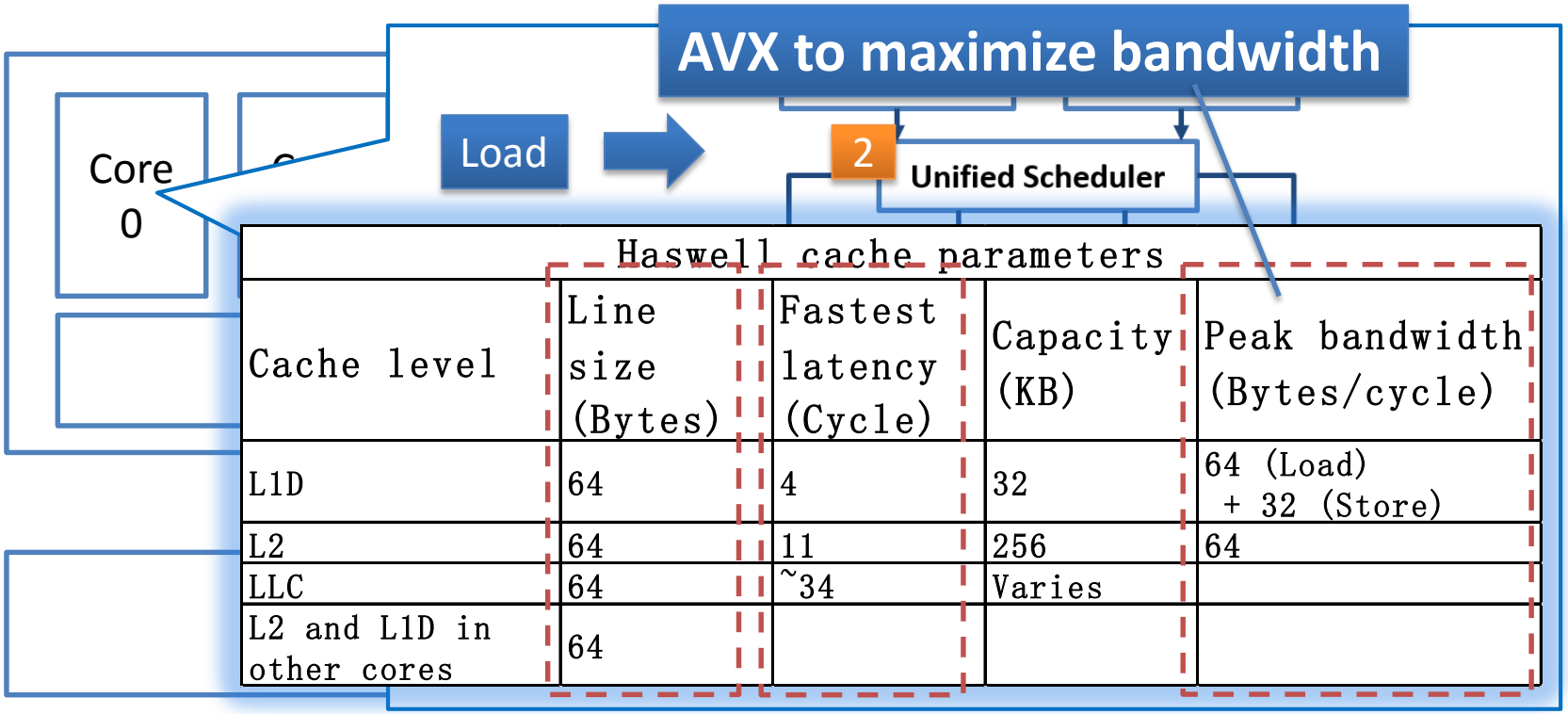
A Typical NFV Scenario: PVP



Overview of Memory System



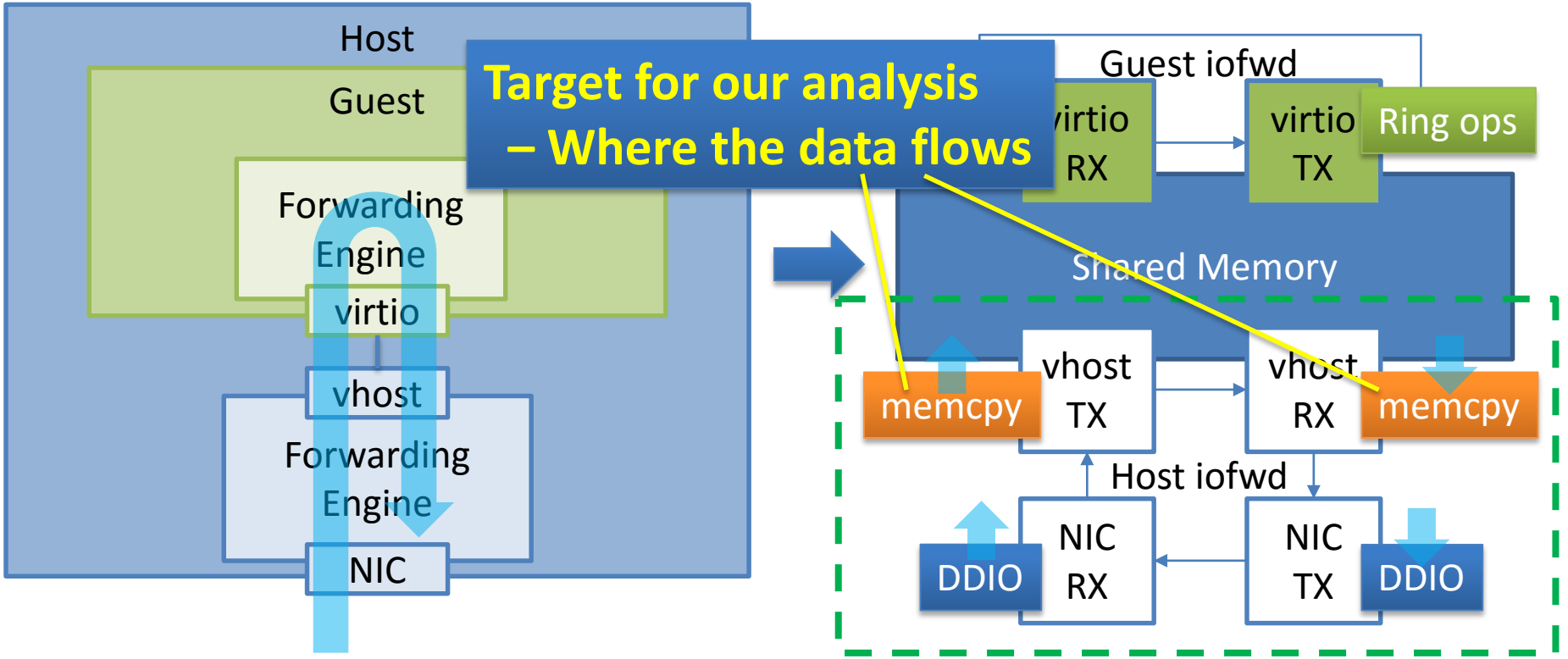
Overview of Memory System (cont'd)



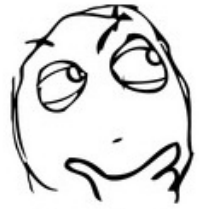
Haswell cache parameters

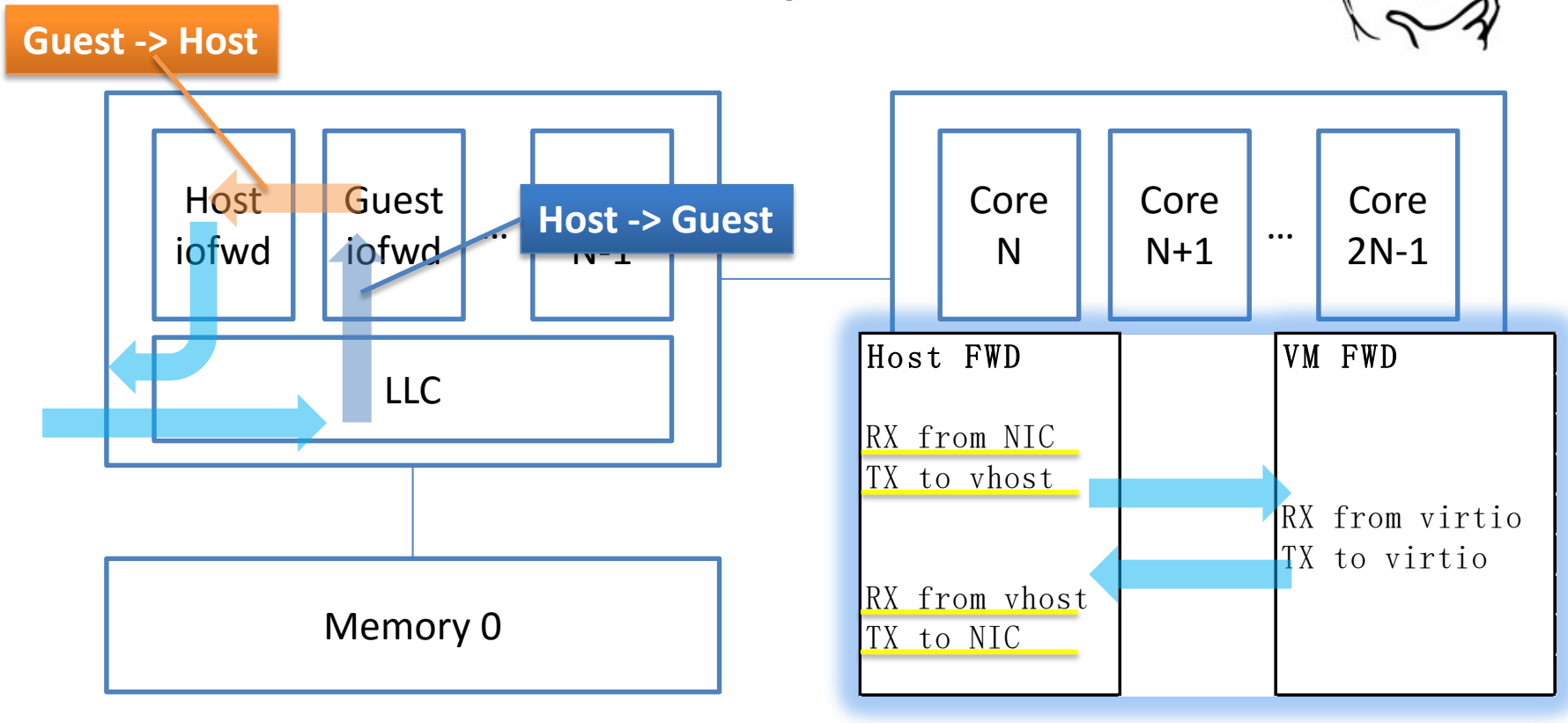
Cache level	Line size (Bytes)	Fastest latency (Cycle)	Capacity (KB)	Peak bandwidth (Bytes/cycle)
L1D	64	4	32	64 (Load) + 32 (Store)
L2	64	11	256	64
LLC	64	~34	Varies	
L2 and L1D in other cores	64			

Let's Do It!

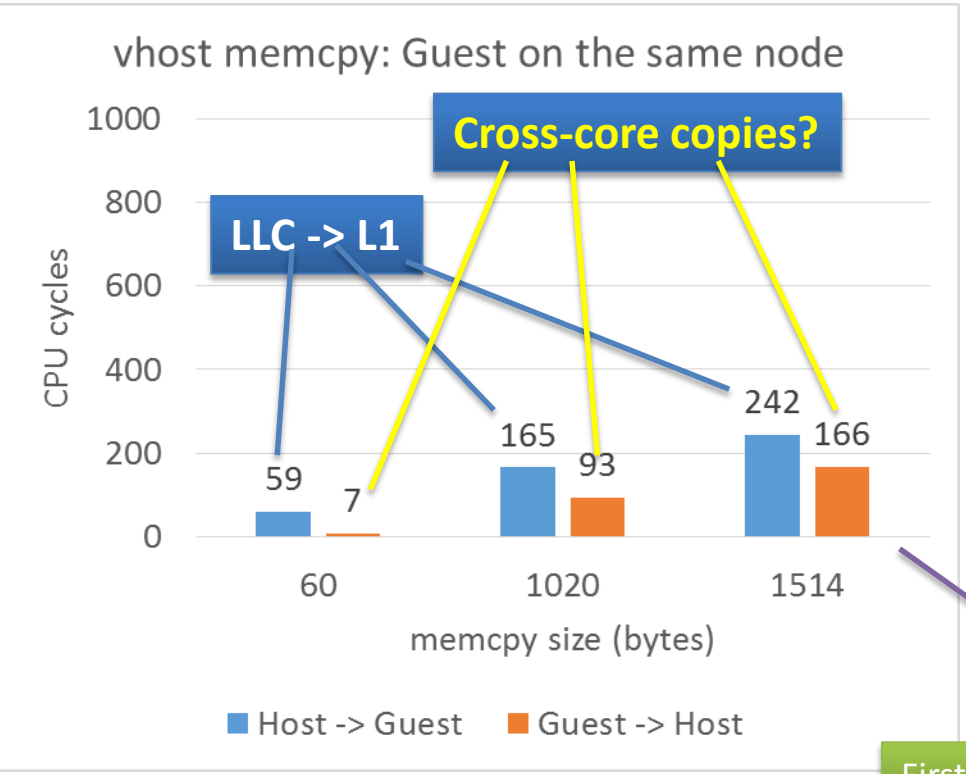

Host
Guest
Forwarding Engine
virtio
vhost
Forwarding Engine
NIC
Target for our analysis
– Where the data flows
Guest iofwd
virtio RX
virtio TX
Ring ops
Shared Memory
memcpy
vhost TX
vhost RX
memcpy
Host iofwd
DDIO
NIC RX
NIC TX
DDIO

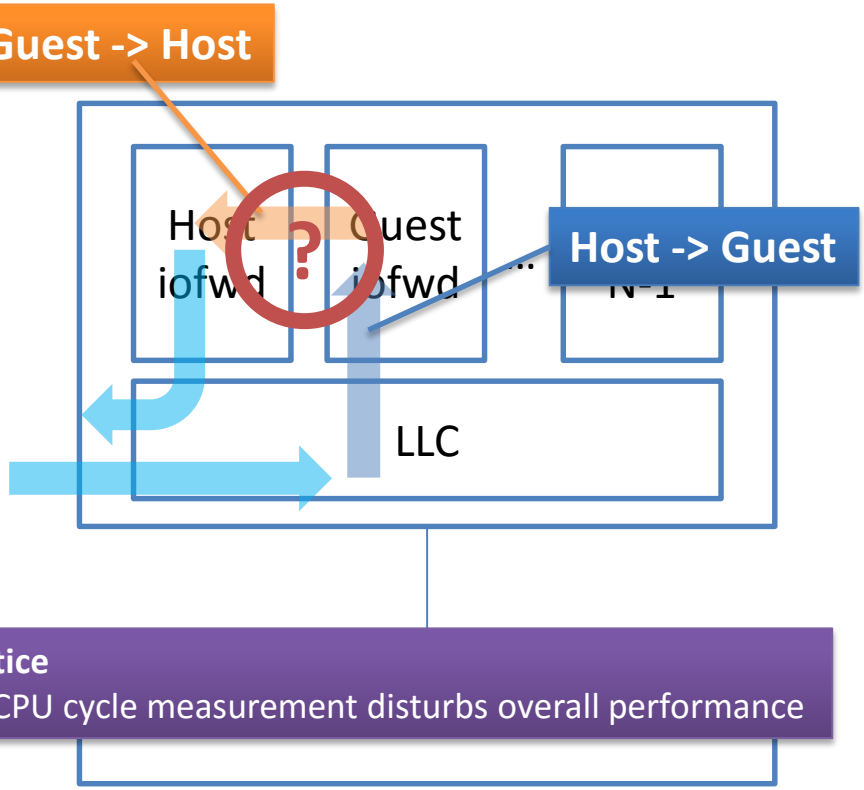
First Impression





Unexpectedly…

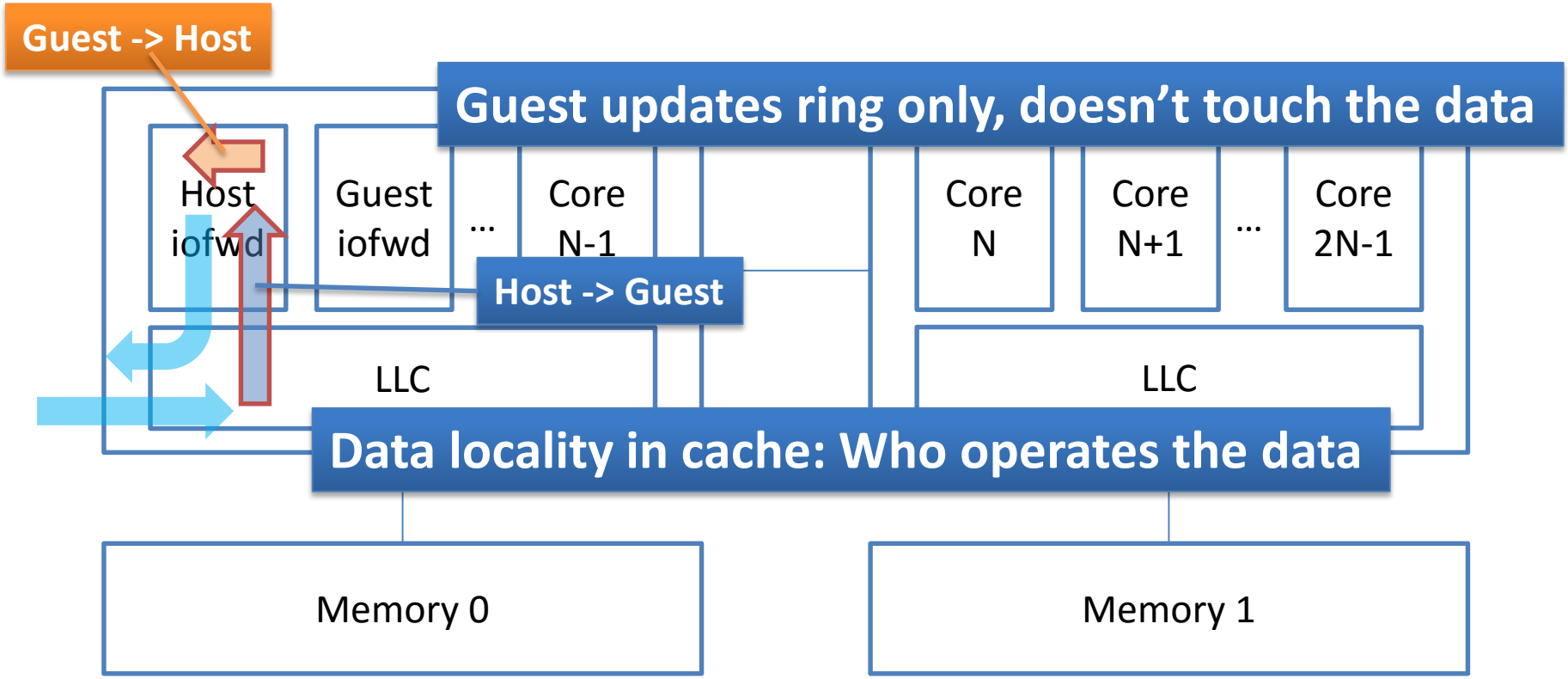




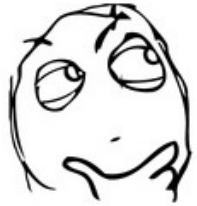
Notice
– CPU cycle measurement disturbs overall performance

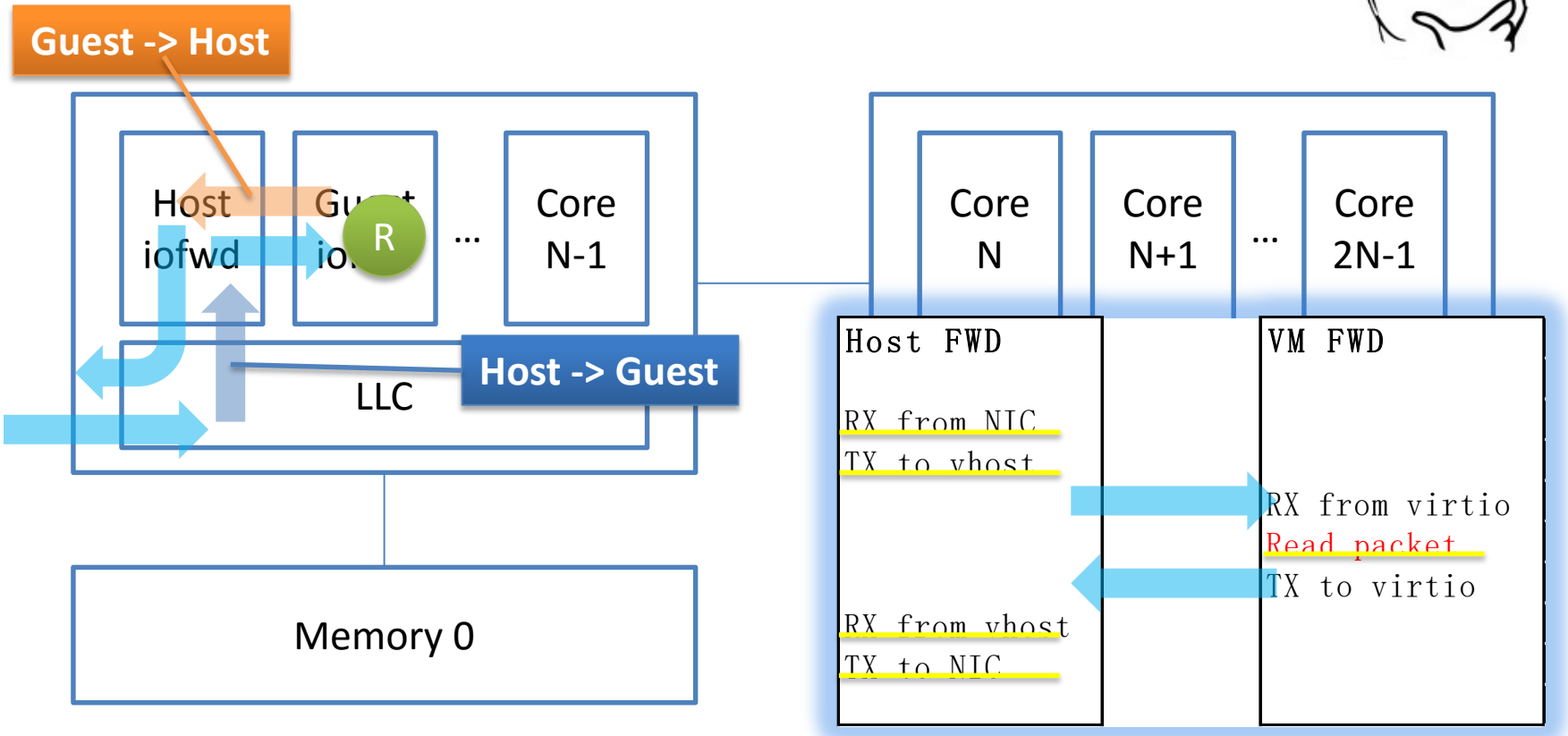
First try

Under The Hood

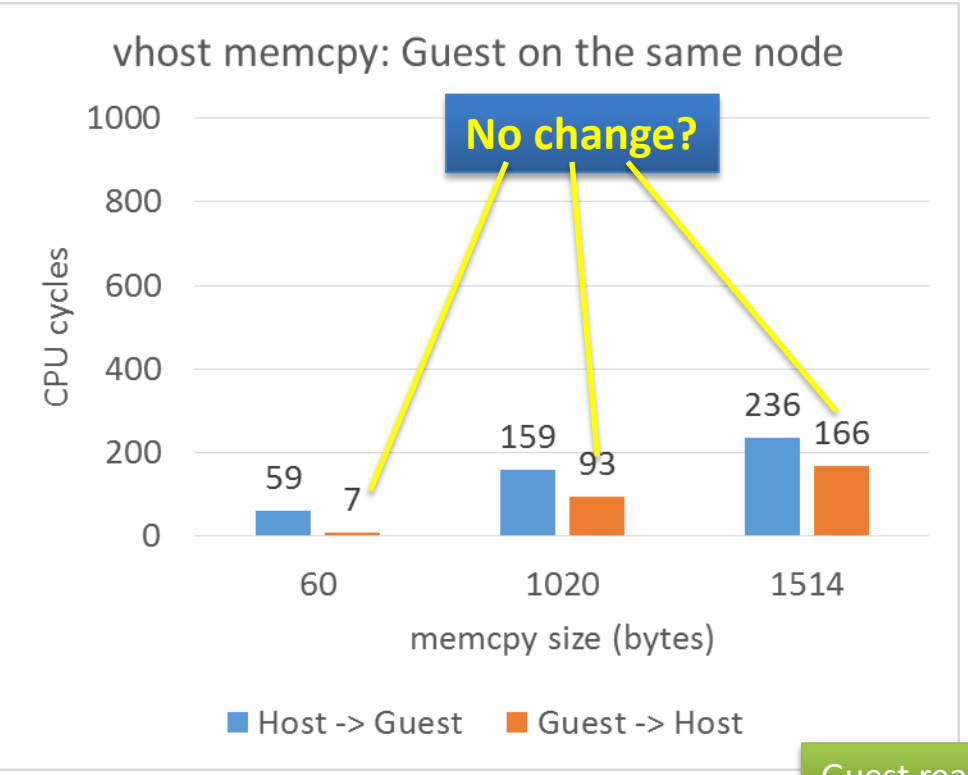


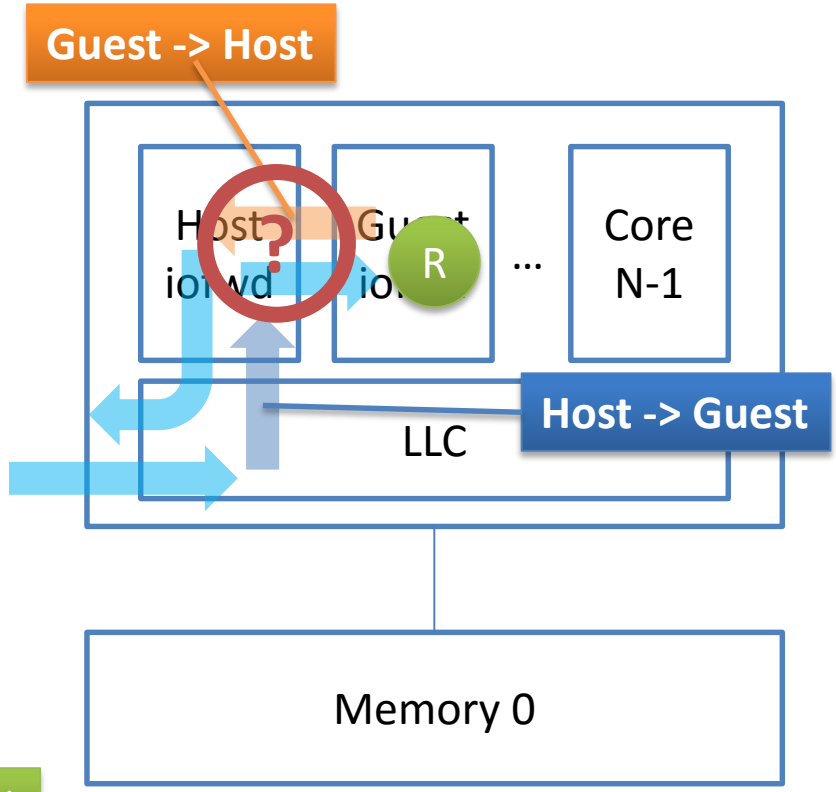
Guest Read The Packet





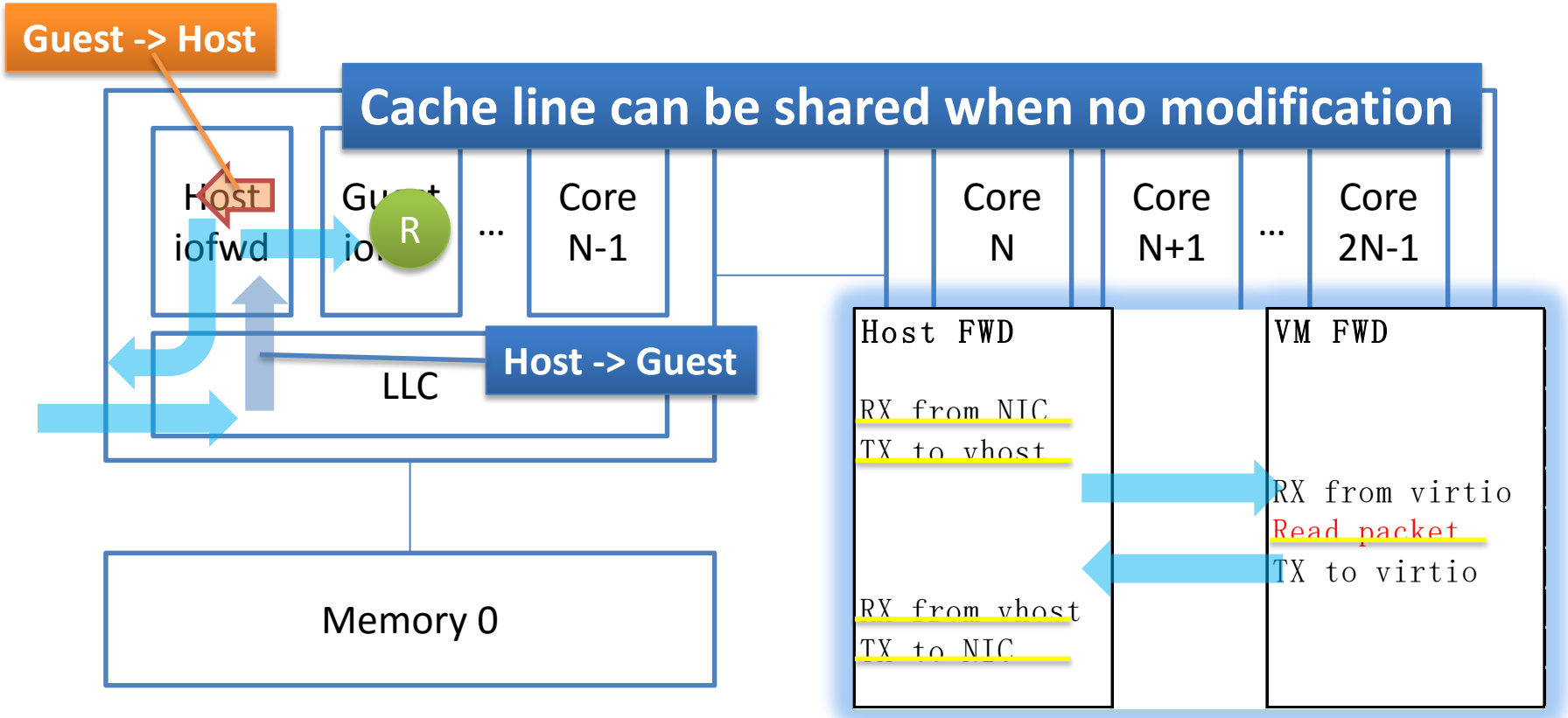
Still Doesn’t Feel Right…



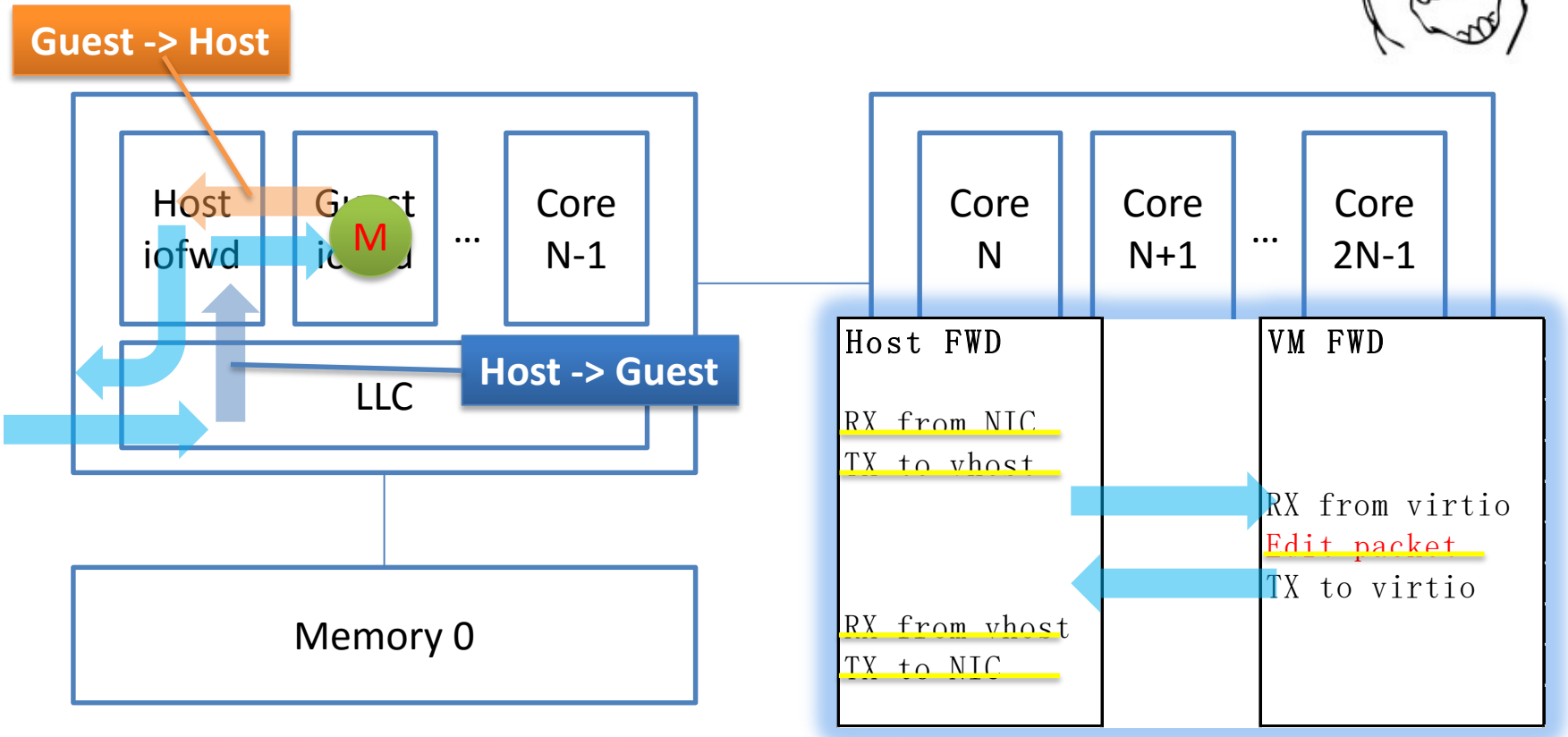


Guest read packet

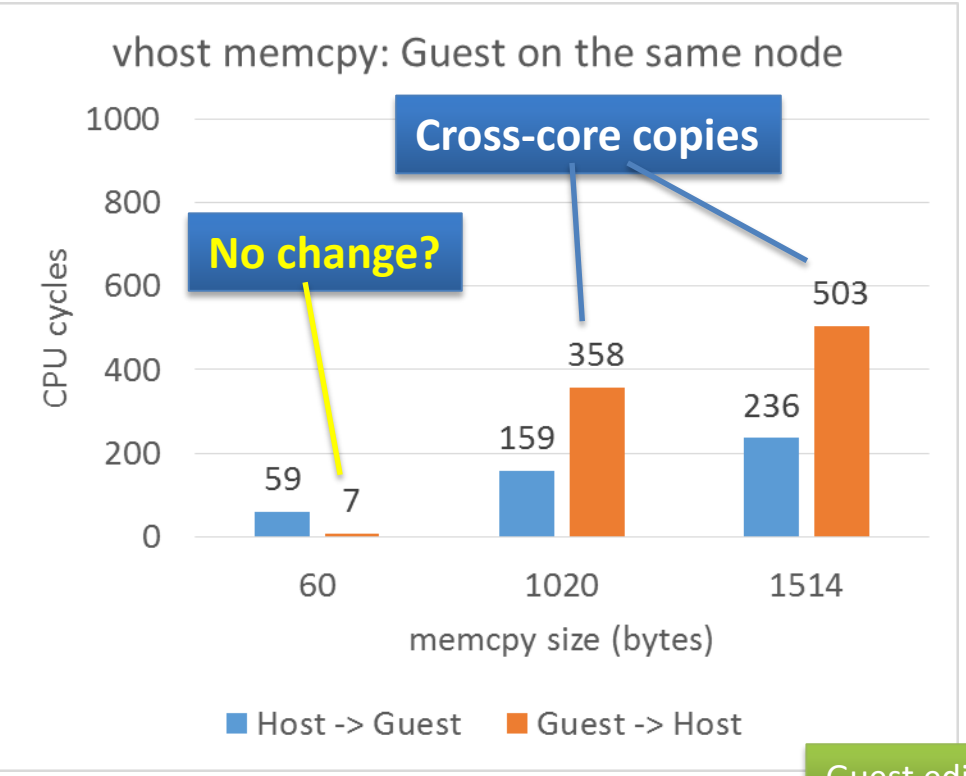
Under The Hood

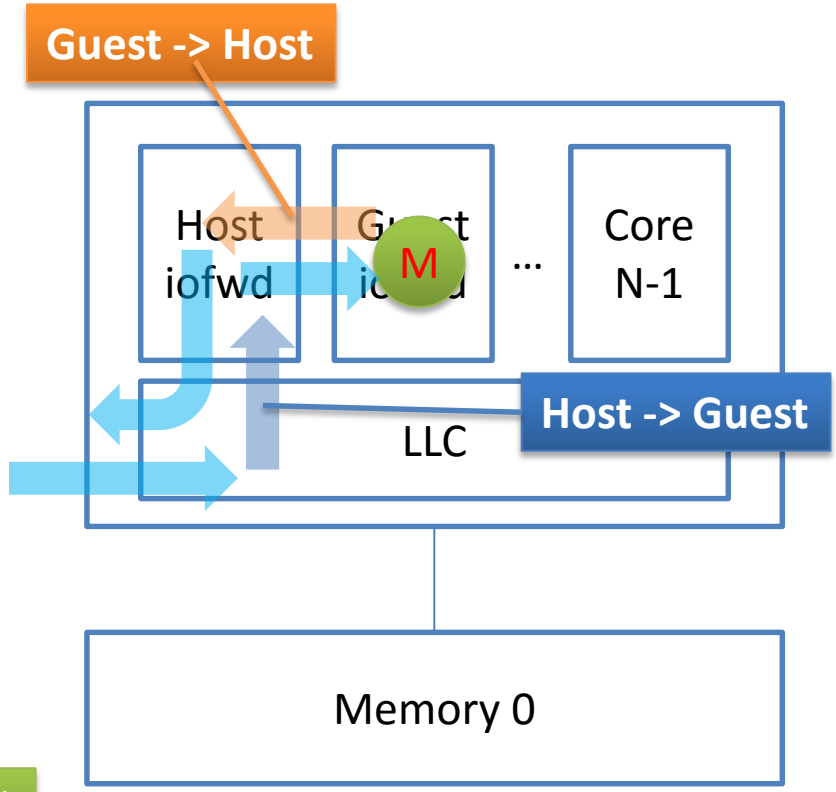


Guest Edit The Packet



Write-back





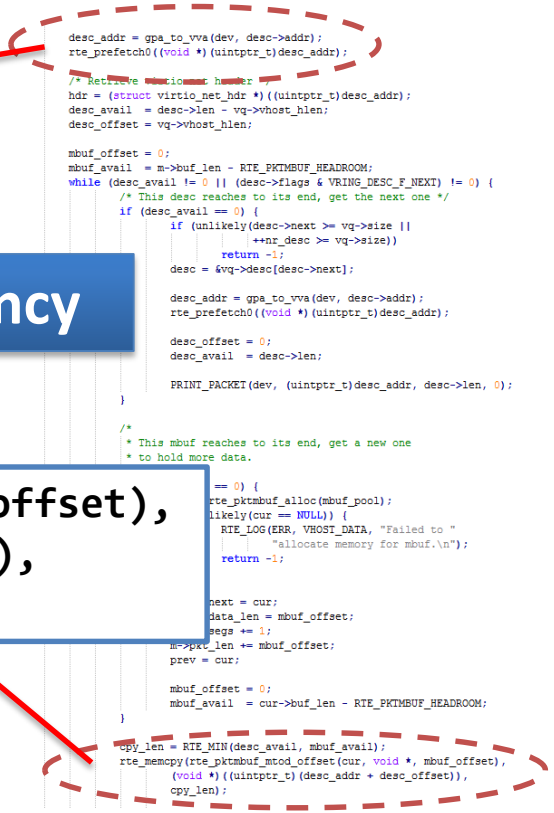
Guest edit packet

Go See Some C Code

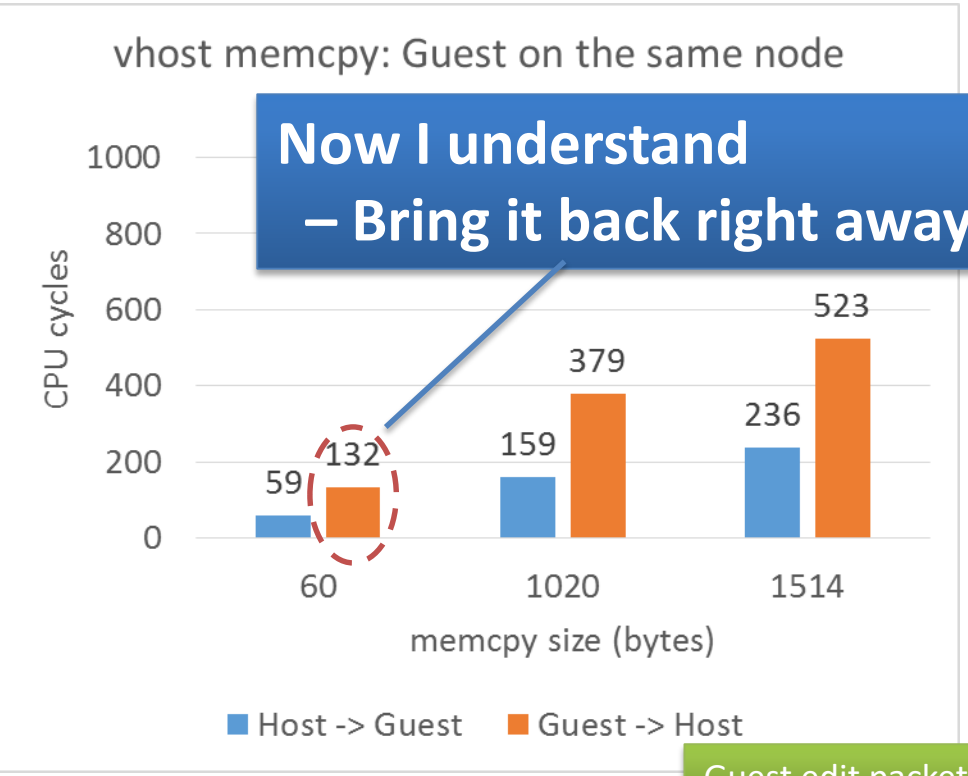
```
desc_addr = gpa_to_vva(dev, desc->addr);
rte_prefetch0((void *)(uintptr_t)desc_addr);
```

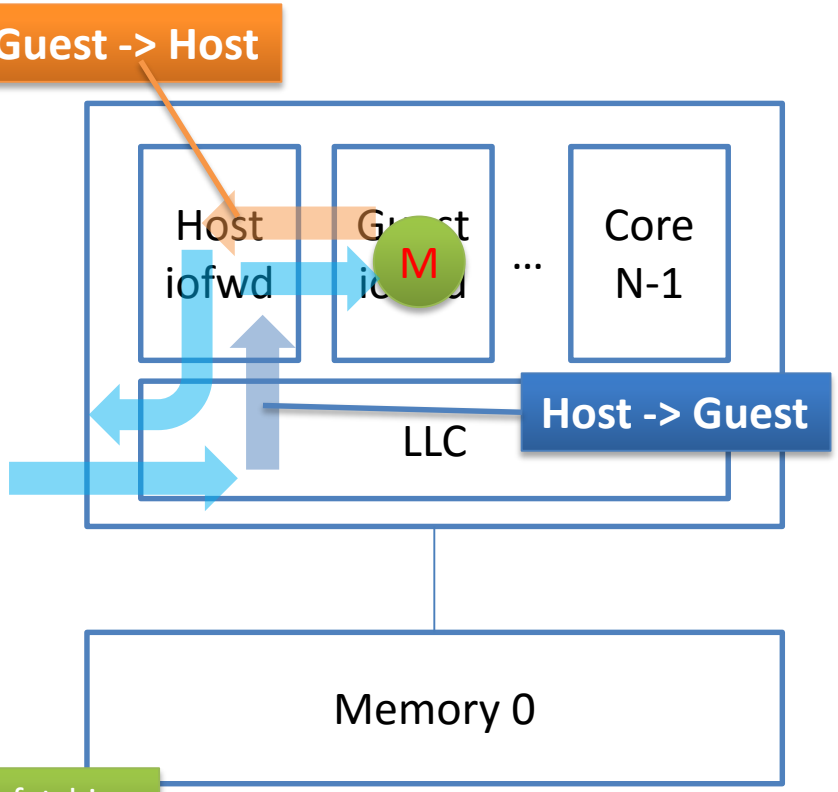
Oh I see – S/W Prefetching to reduce latency

```
rte_memcpy(rte_pktmbuf_mtod_offset(cur, void *, mbuf_offset),
        (void *)((uintptr_t)(desc_addr + desc_offset)),
        cpy_len);
```



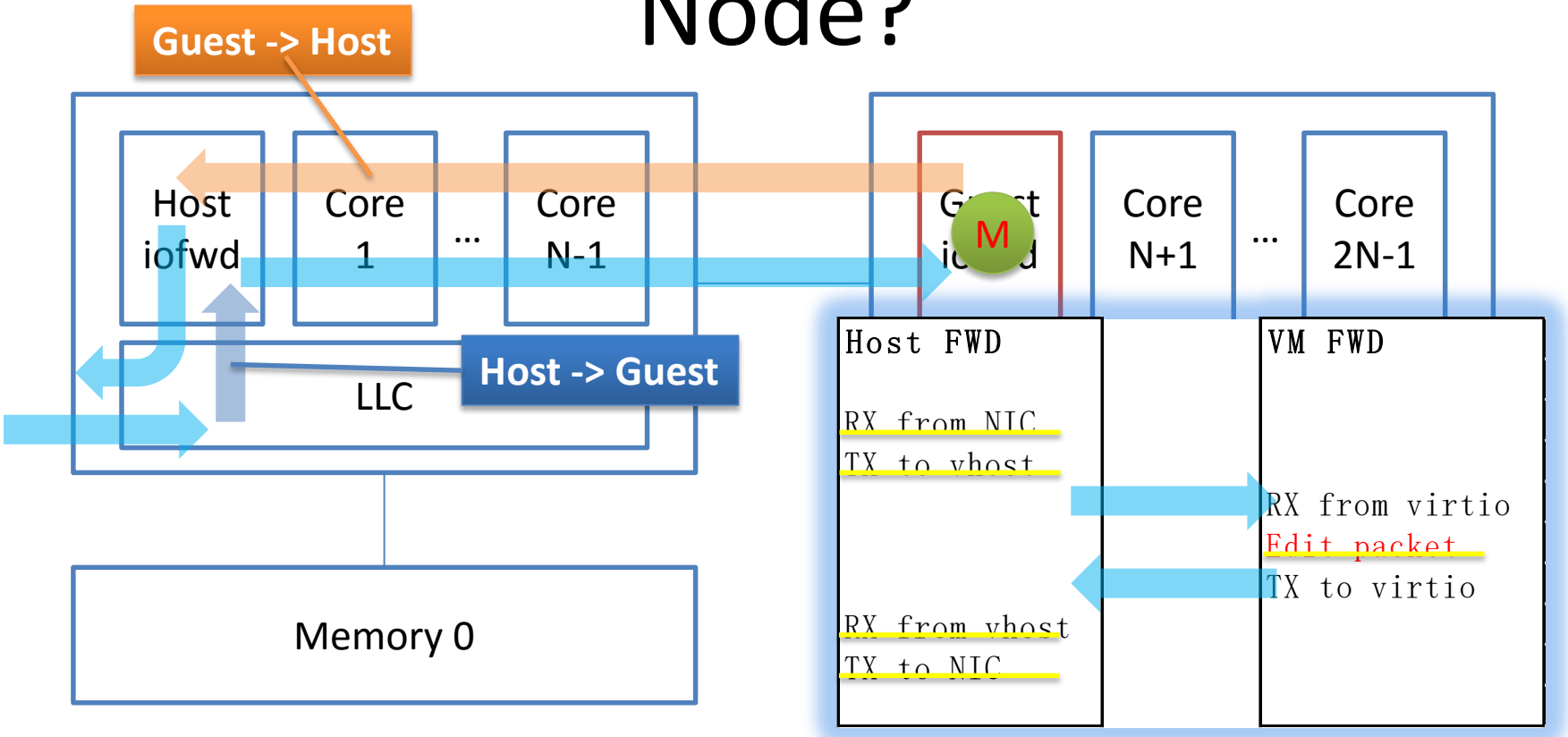
Without S/W Prefetching



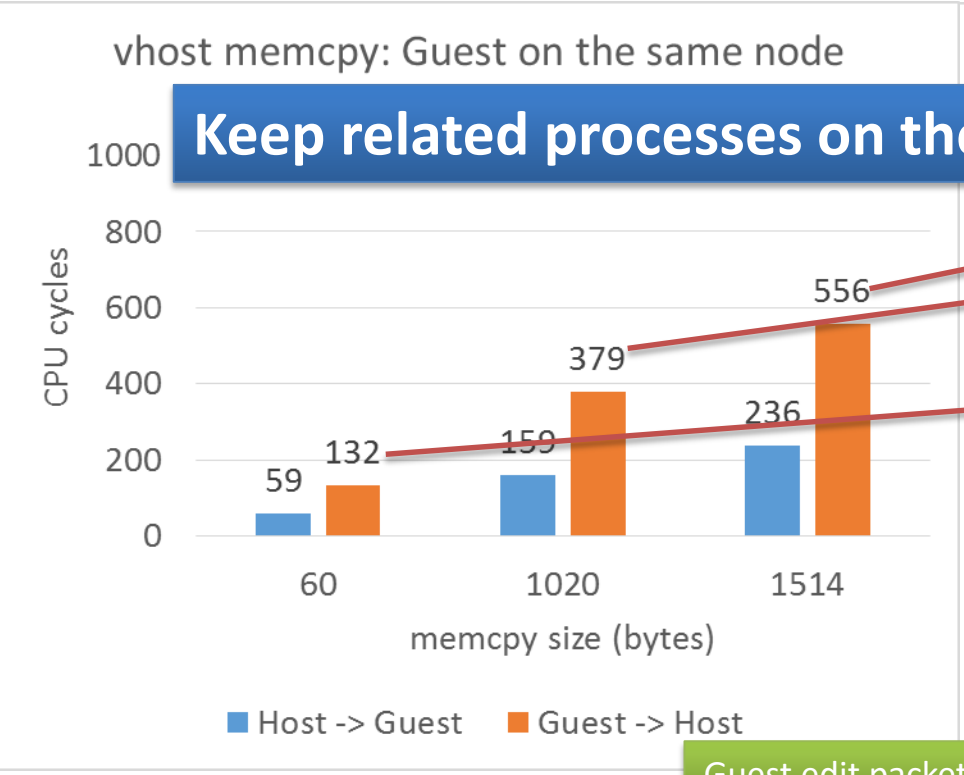


Guest edit packet; No prefetching

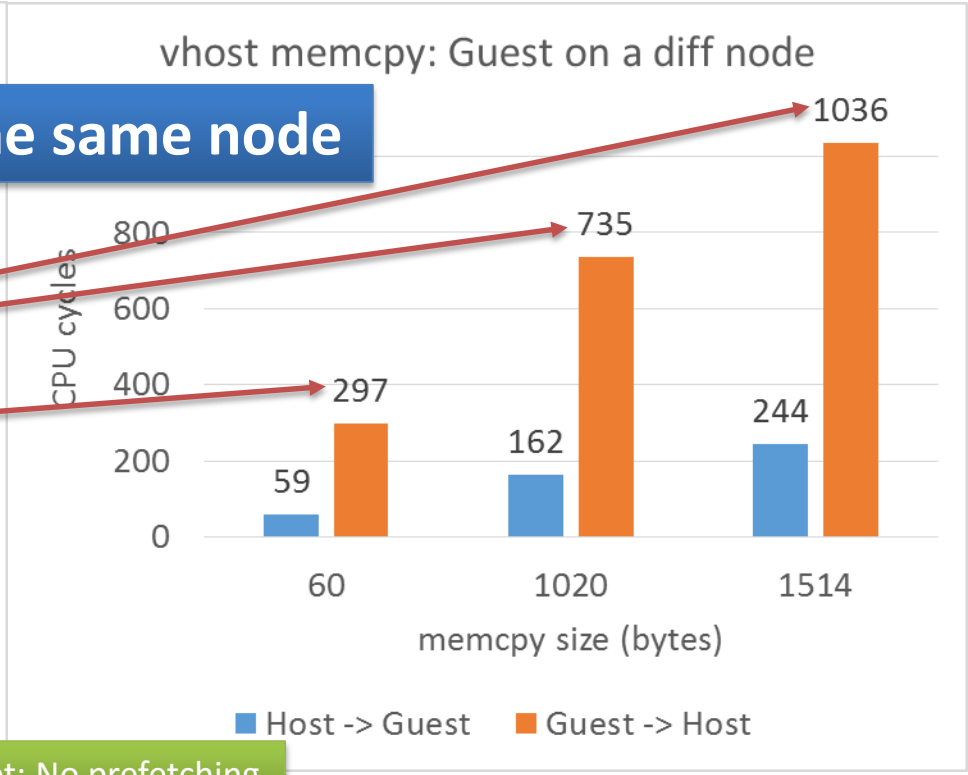
How About Guest In Another Node?



Better NOT…

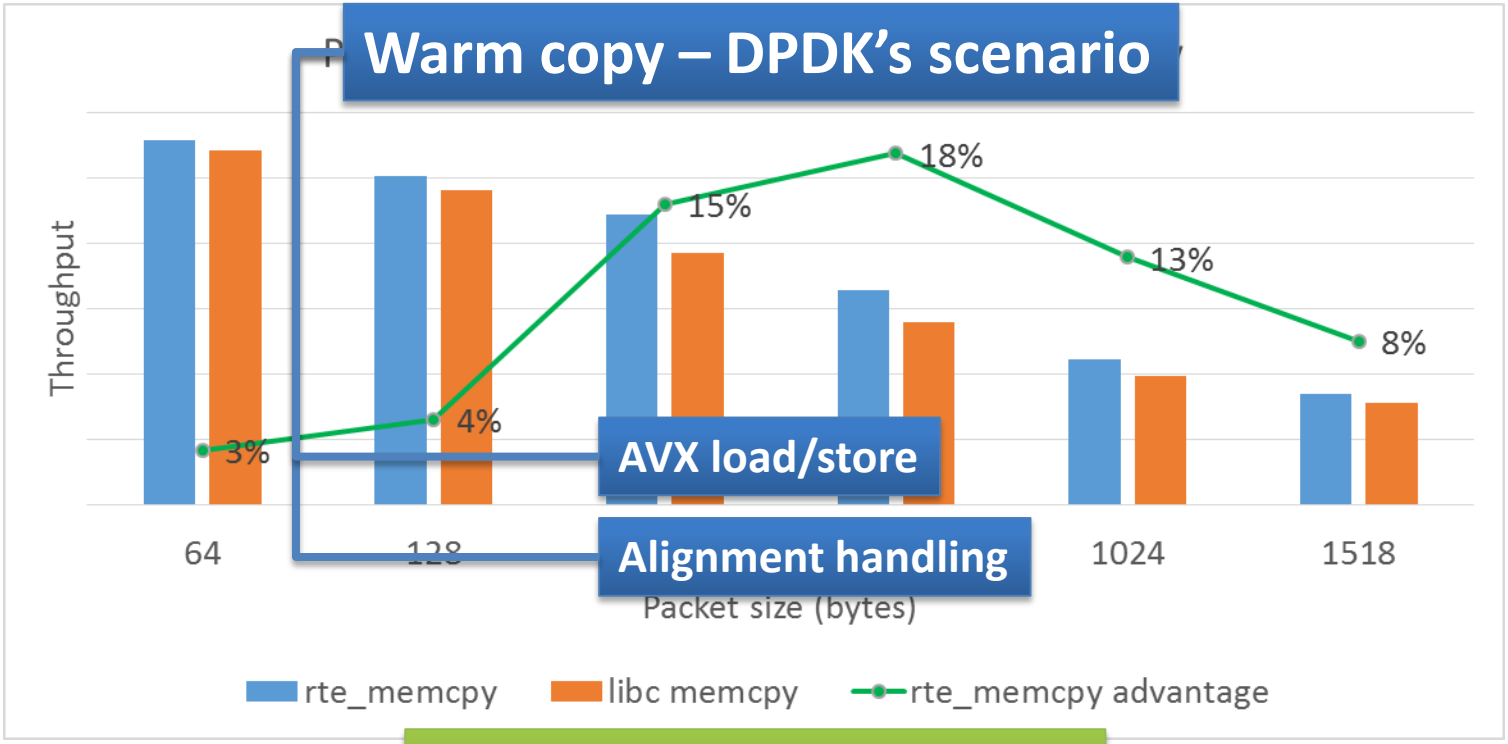


Keep related processes on the same node



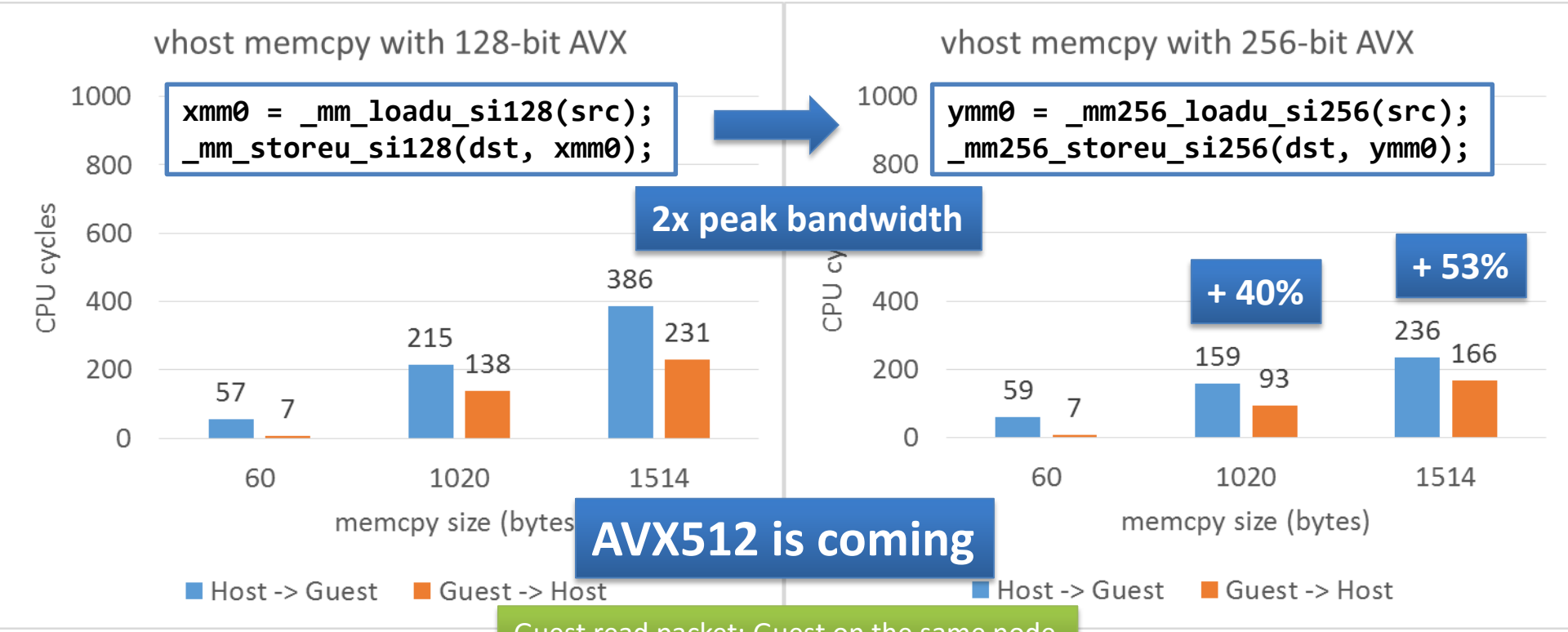
Guest edit packet; No prefetching

rte_memcpy()? Why Even Bother?



Guest edit packet; Guest on the same node

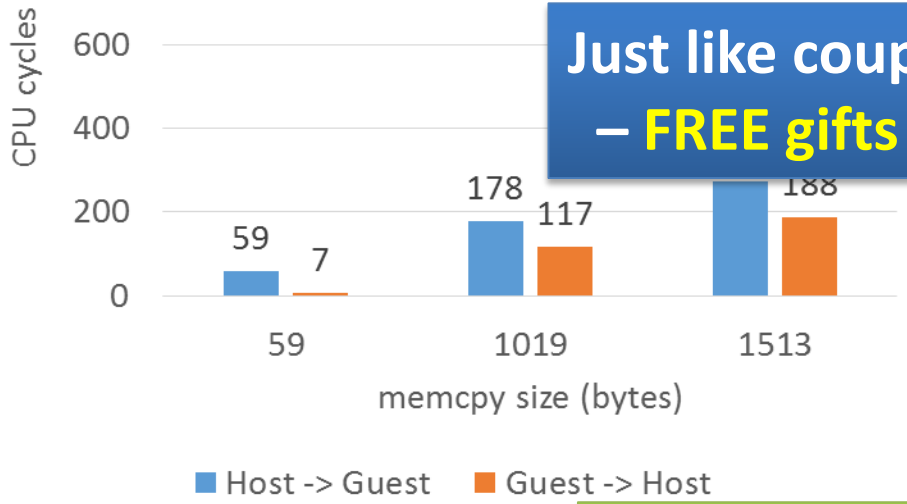
AVX For Bandwidth



Alignment Matters

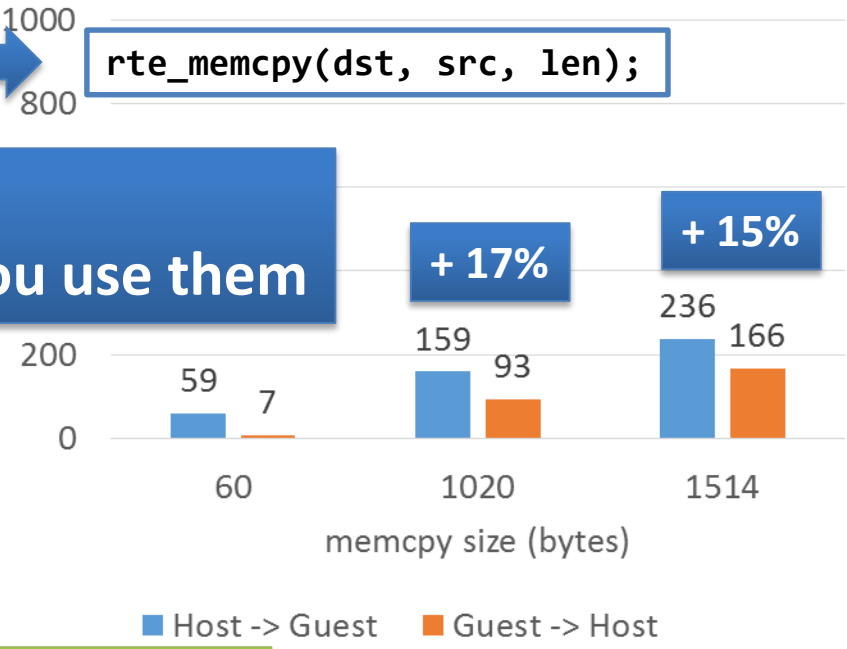
vhost unaligned memcpy

```
rte_memcpy((void *)((uint8_t *)dst + 1),
           src, len - 1);
```



vhost 64 bytes aligned memcpy

```
rte_memcpy(dst, src, len);
```



Just like coupons
– FREE gifts if you use them

Guest read packet; Guest on the same node

Takeaways

- See actual memory behaviors under the hood
  - Intel® 64 and IA-32 Architectures Optimization Reference Manual
- Benefit from new IA technologies
  - AVX, DDIO…


DPDK
DATA PLANE DEVELOPMENT KIT

Cache Allocation Technology