

基于智能网卡的 vhost-user优化接收 方法

阿里云 言枢

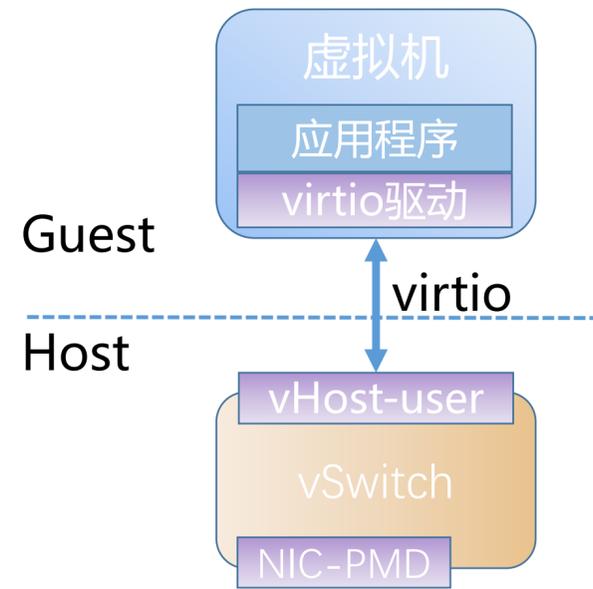
议程

- **为什么需要智能网卡**
- **智能网卡解决方案**
- **接收端的数据通路**
- **零拷贝优化方案**
- **零拷贝的优势**

议程

- **为什么需要智能网卡**
- 智能网卡解决方案
- 接收端的数据通路
- 零拷贝优化方案
- 零拷贝的优势

为什么上智能网卡？



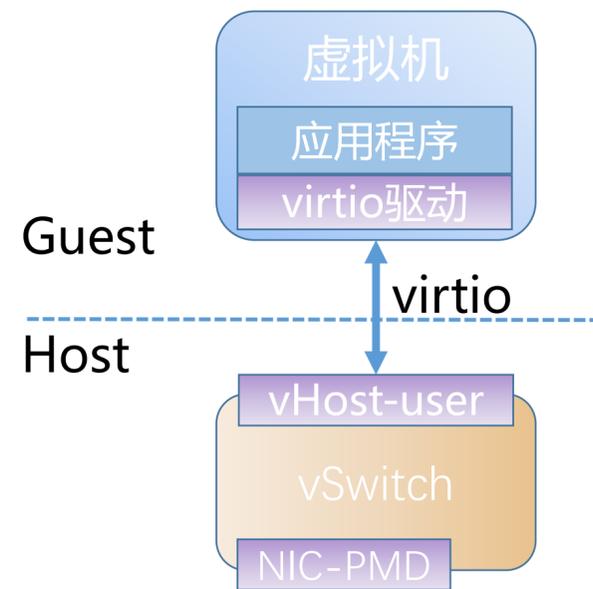
优点

- 接口通用
- 热迁移方便

缺点

- 性能不够好
- 占用主机资源多
(CPU/内存等)

为什么上智能网卡？

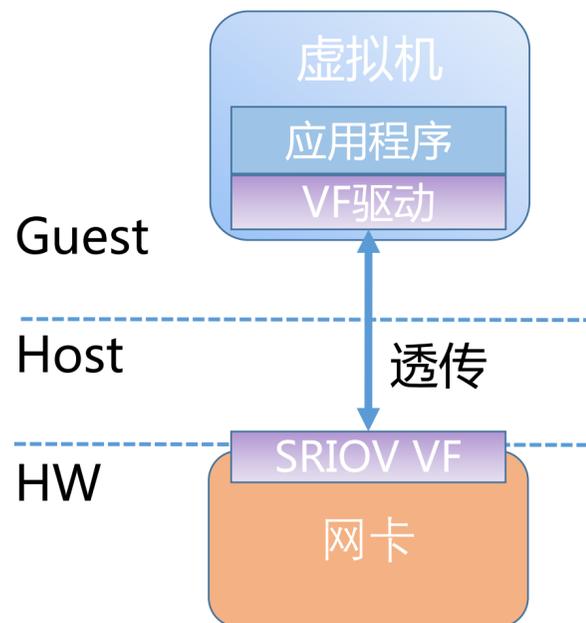


优点

- 接口通用
- 热迁移方便

缺点

- 性能不够好
- 占用主机资源多 (CPU/内存等)



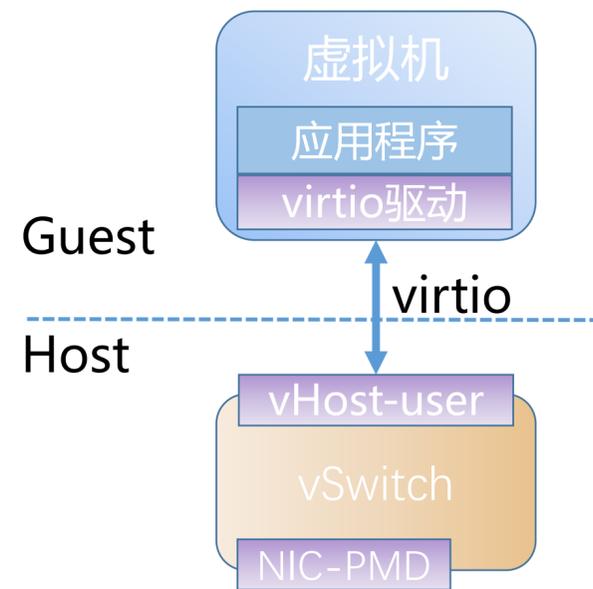
优点

- 性能好
- 不占用主机资源

缺点

- vSwitch不容易部署
- 厂商提供接口，没有统一规范
- 很难做热迁移

为什么上智能网卡？

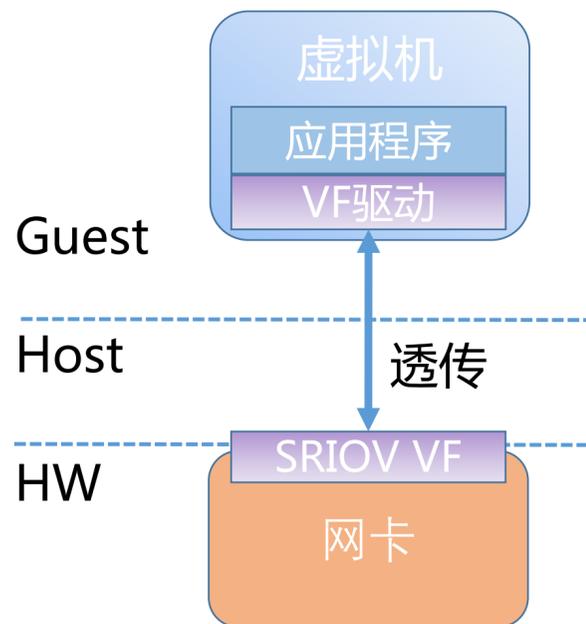
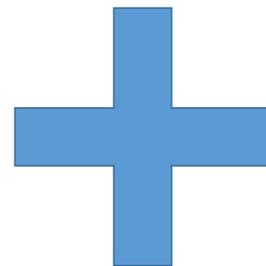


优点

- 接口通用
- 热迁移方便

缺点

- 性能不够好
- 占用主机资源多 (CPU/内存等)

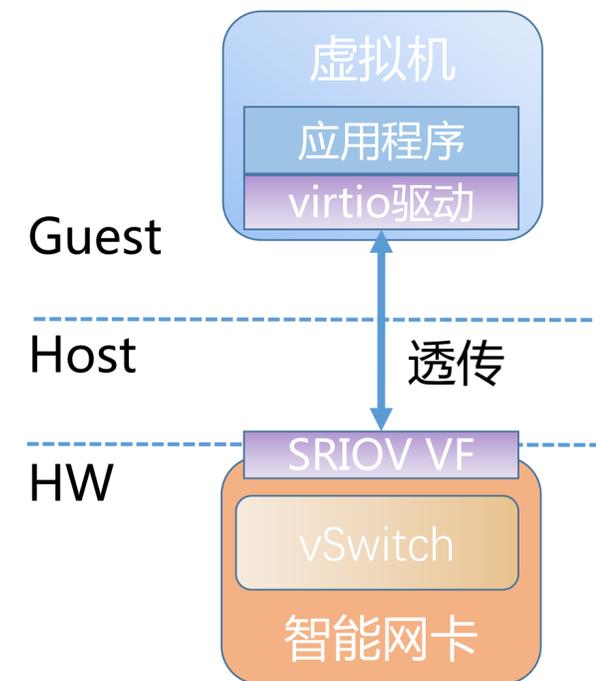


优点

- 性能好
- 不占用主机资源

缺点

- vSwitch不容易部署
- 厂商提供接口，没有统一规范
- 很难做热迁移



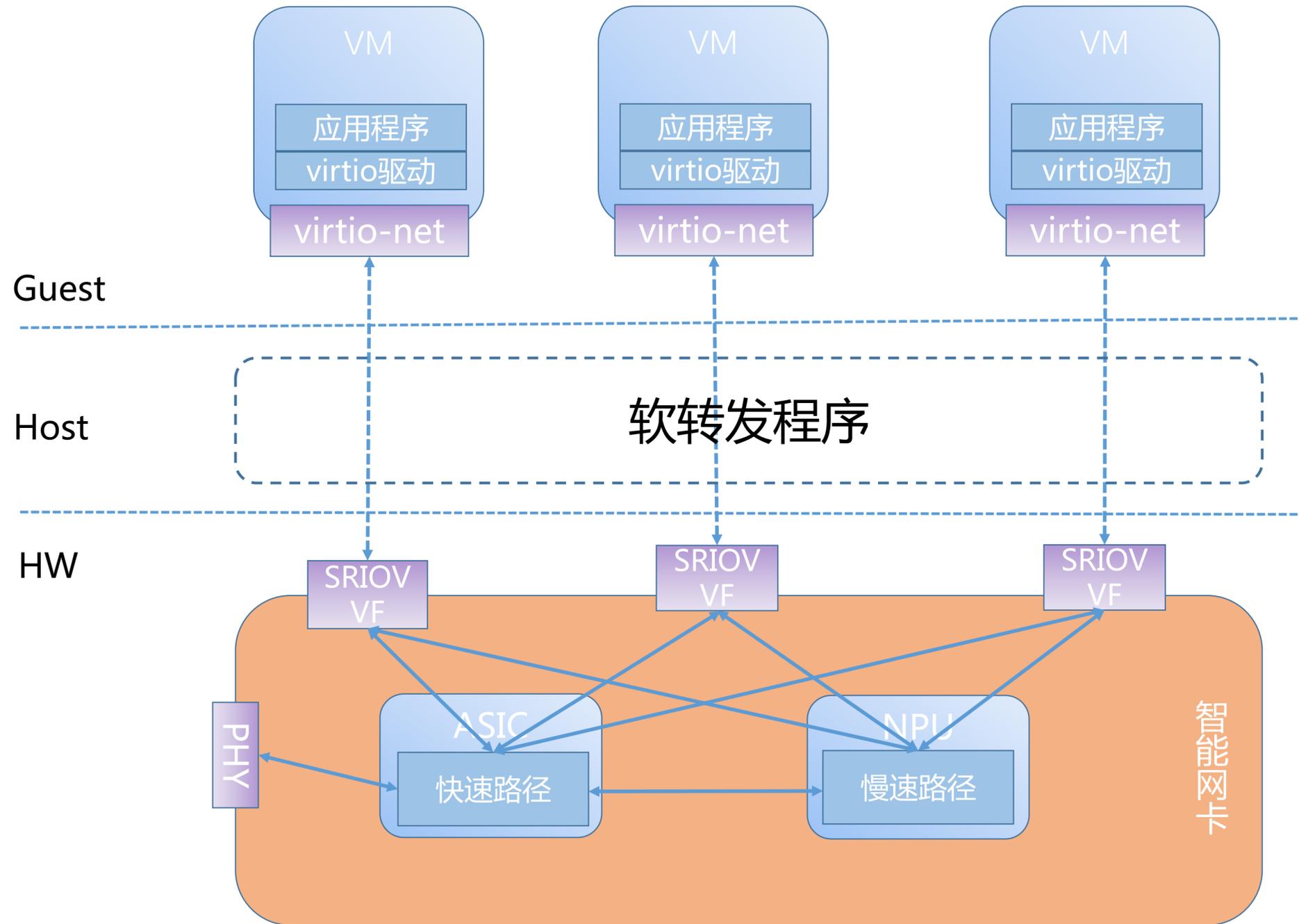
优点

- 性能好
- 不占用主机资源
- 接口通用
- 热迁移方便

议程

- 为什么需要智能网卡
- **智能网卡解决方案**
- 接收端的数据通路
- 零拷贝优化方案
- 零拷贝的优势

智能网卡解决方案

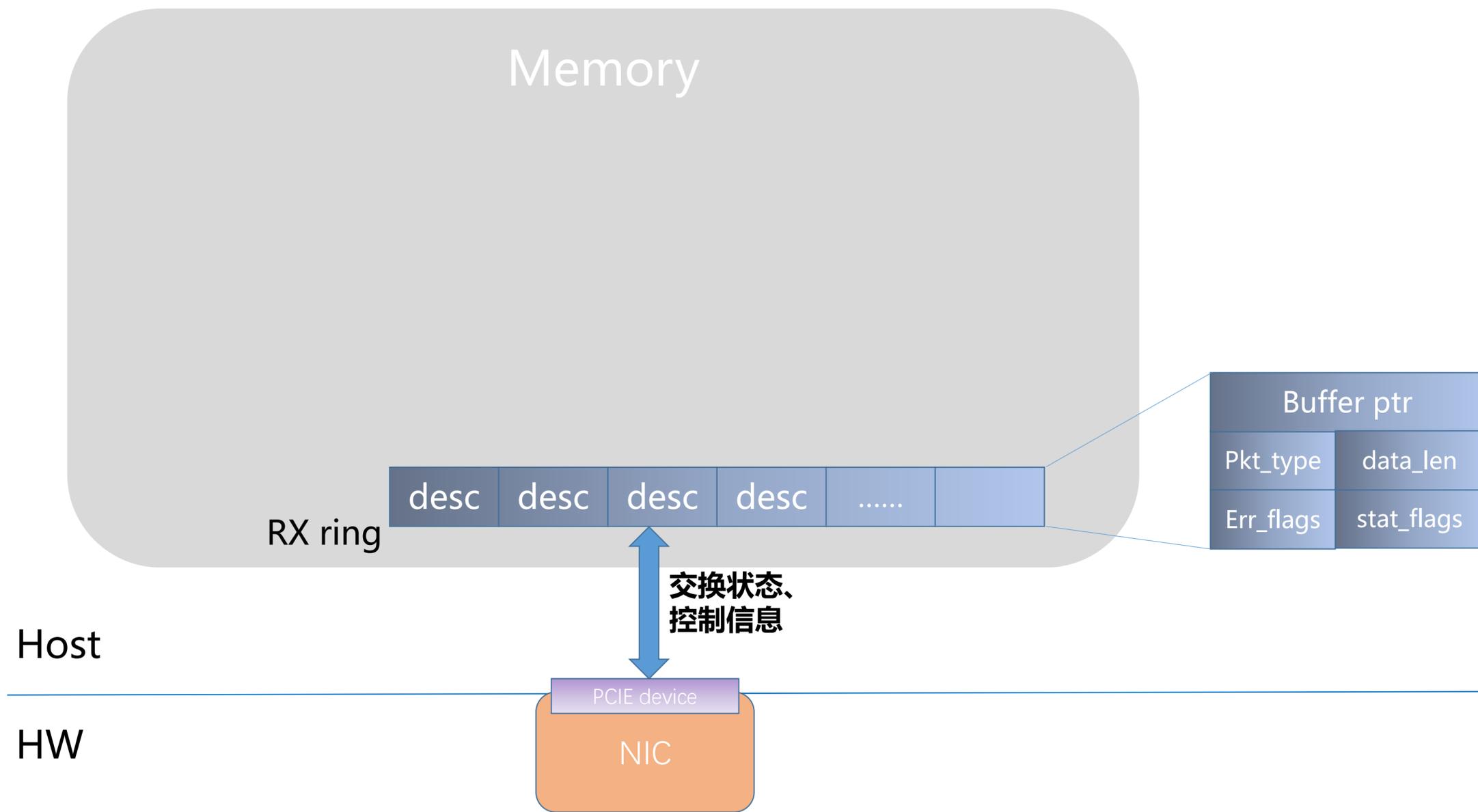


议程

- 为什么需要智能网卡
- 智能网卡解决方案
- **接收端的数据通路**
- 零拷贝优化方案
- 零拷贝的优势

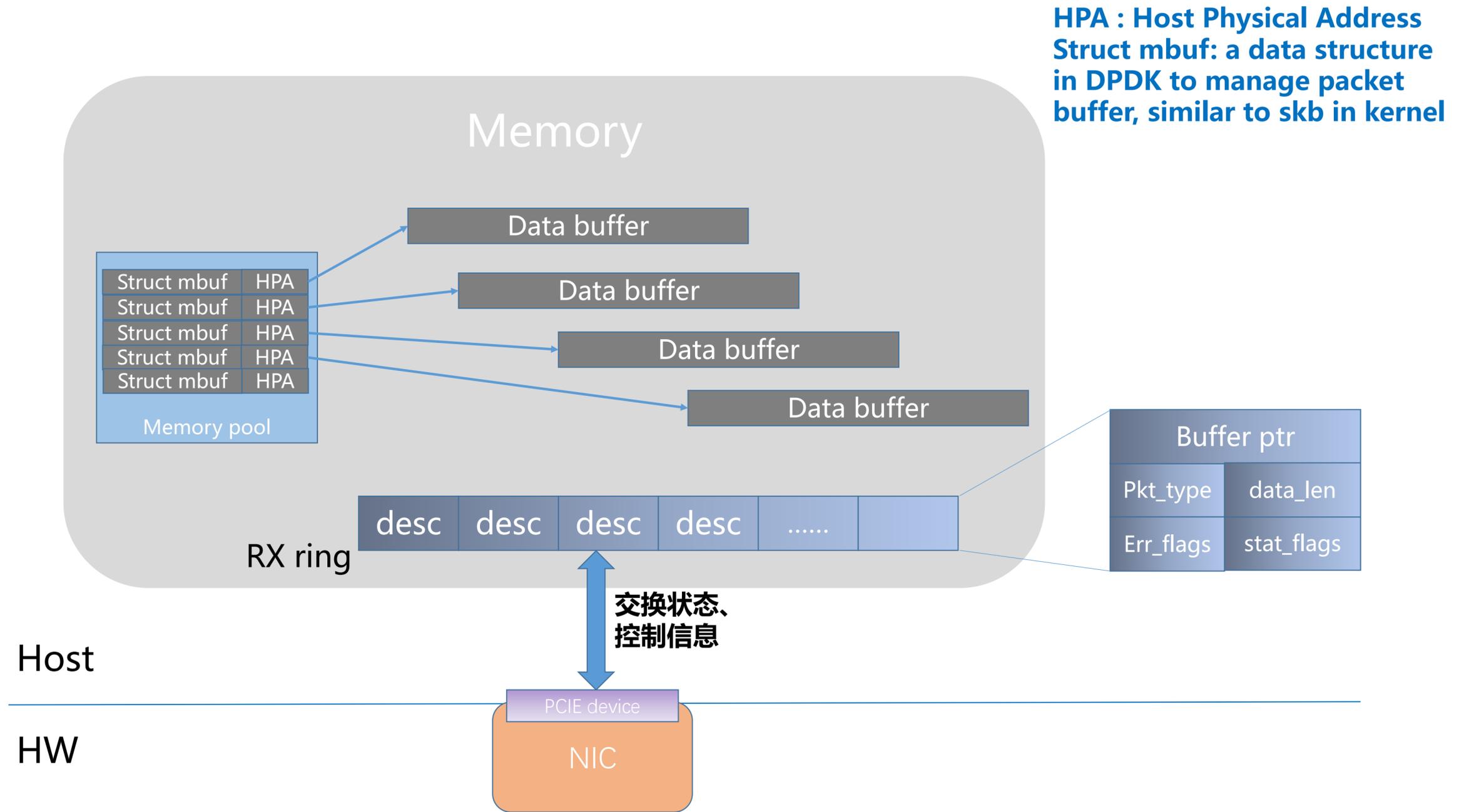
网卡接收队列 – DPDK初始化过程

- 内存中创建一个接收队列和硬件交互



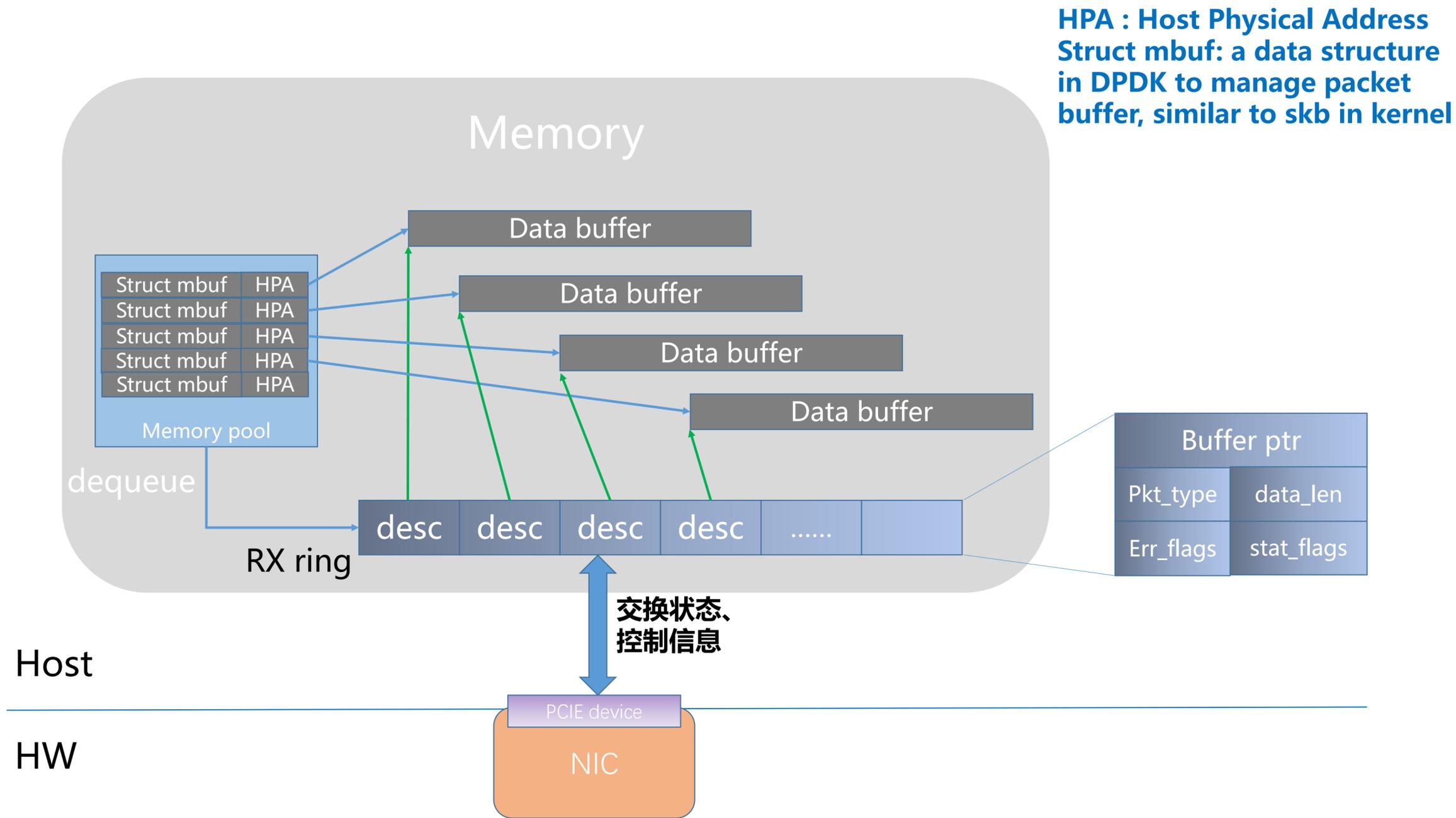
网卡接收队列 – DPDK初始化过程

- 内存中创建一个接收队列和硬件交互
- 内存中创建一个内存池用于存储接收到的报文

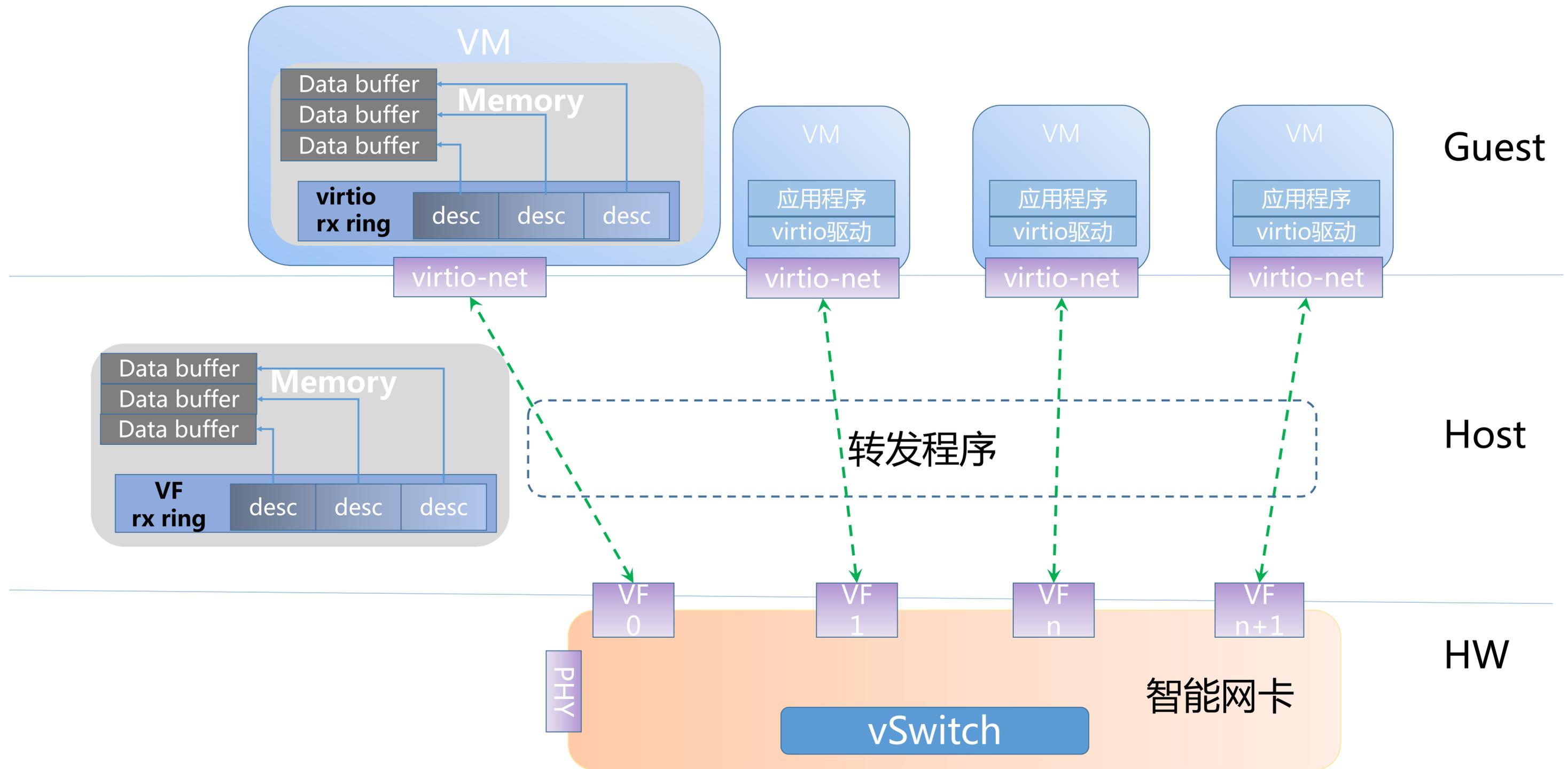


网卡接收队列 – DPDK初始化过程

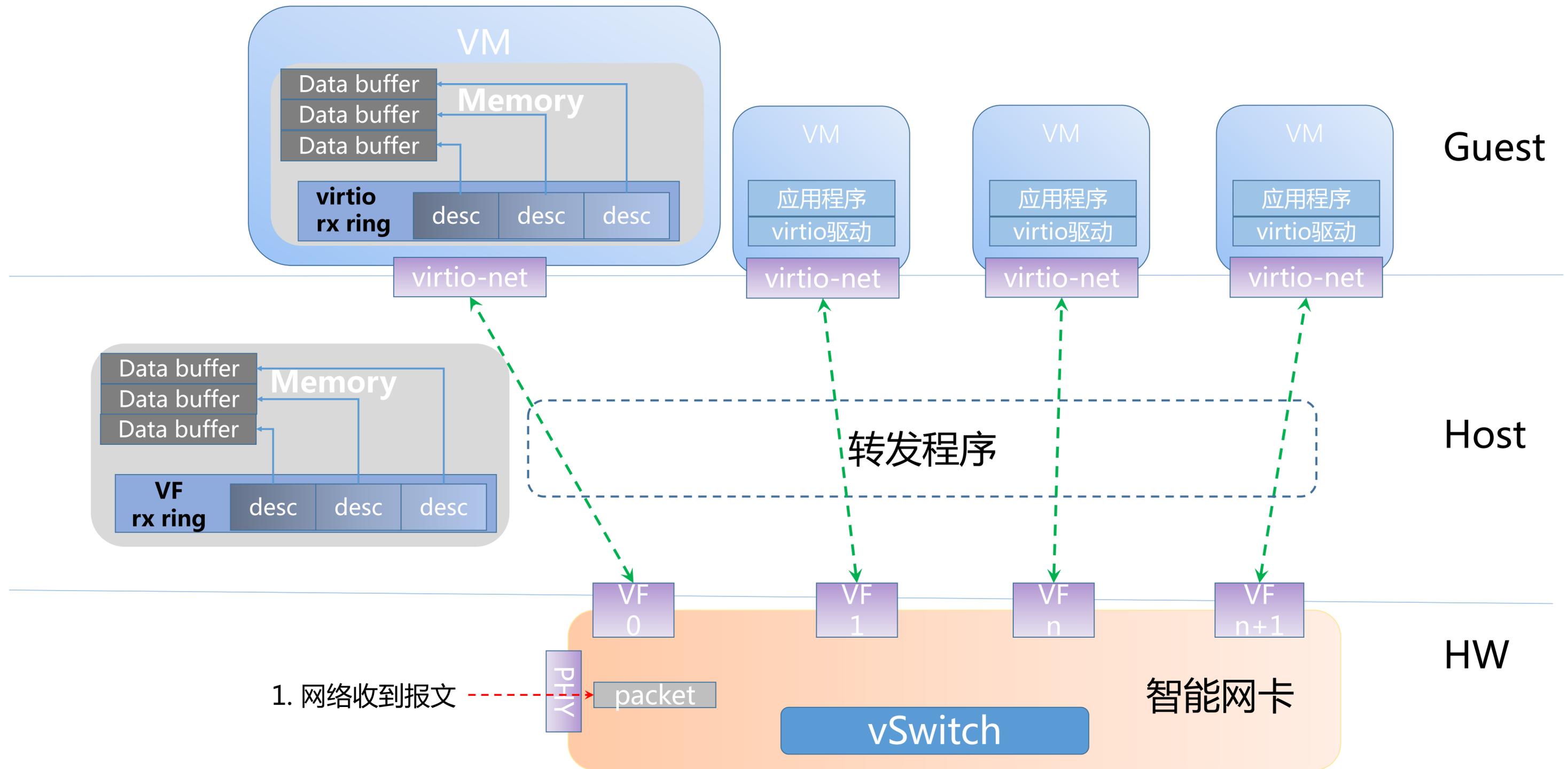
- 内存中创建一个接收队列和硬件交互
- 内存中创建一个内存池用于存储接收到的报文
- 初始化接受队列，从内存池获取缓冲区，填充到描述符



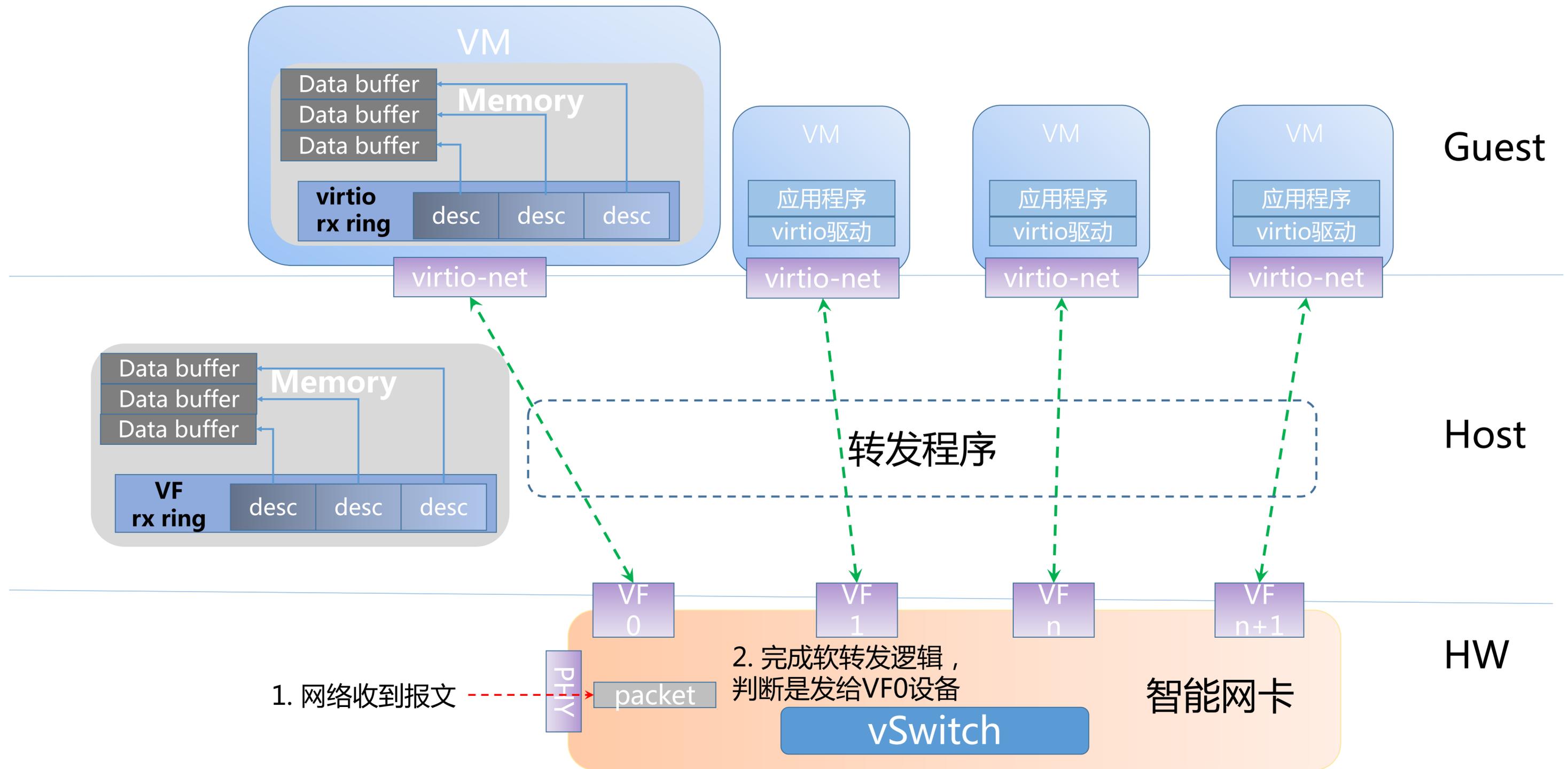
VM接收报文流程



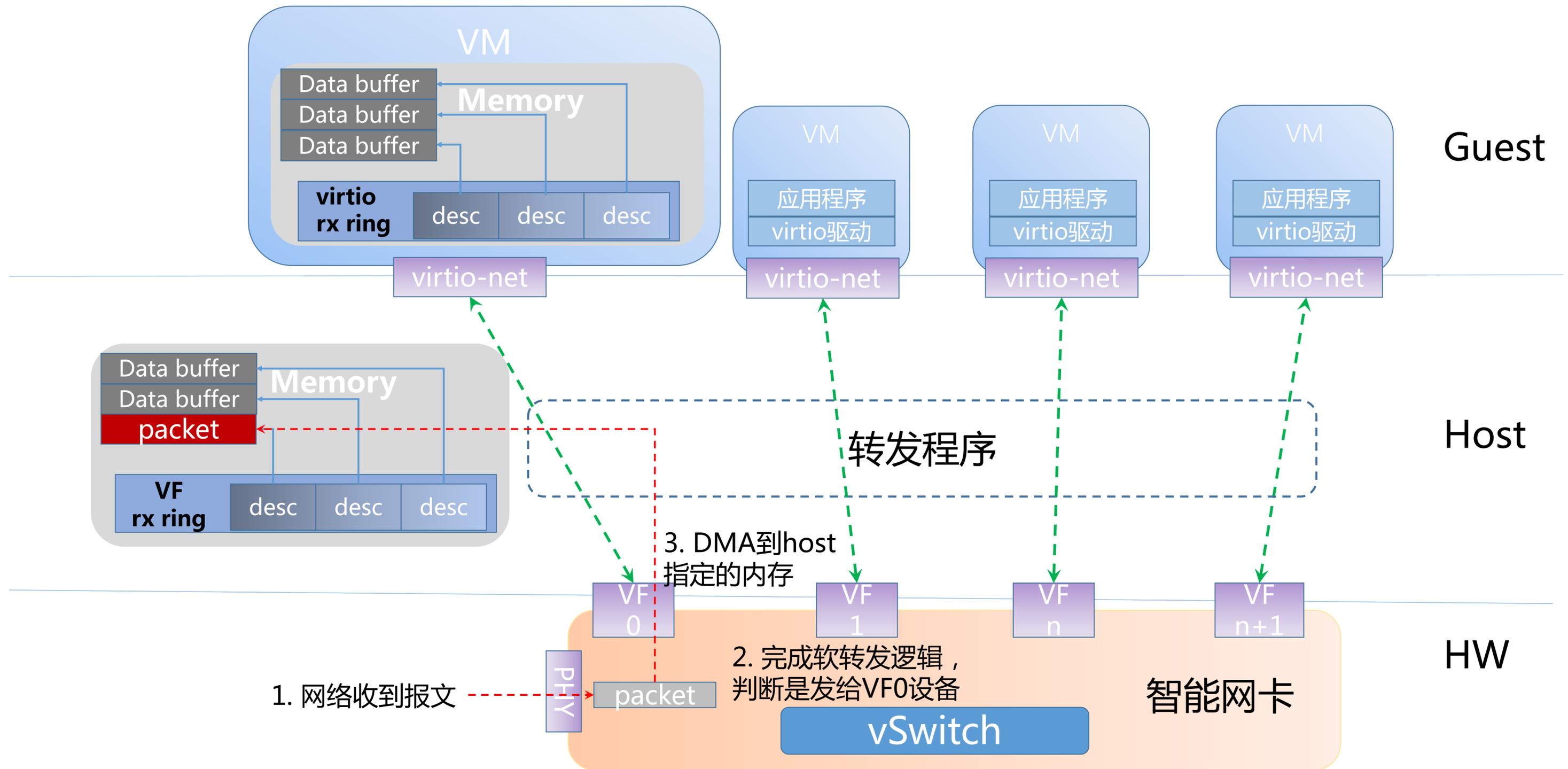
VM接收报文流程



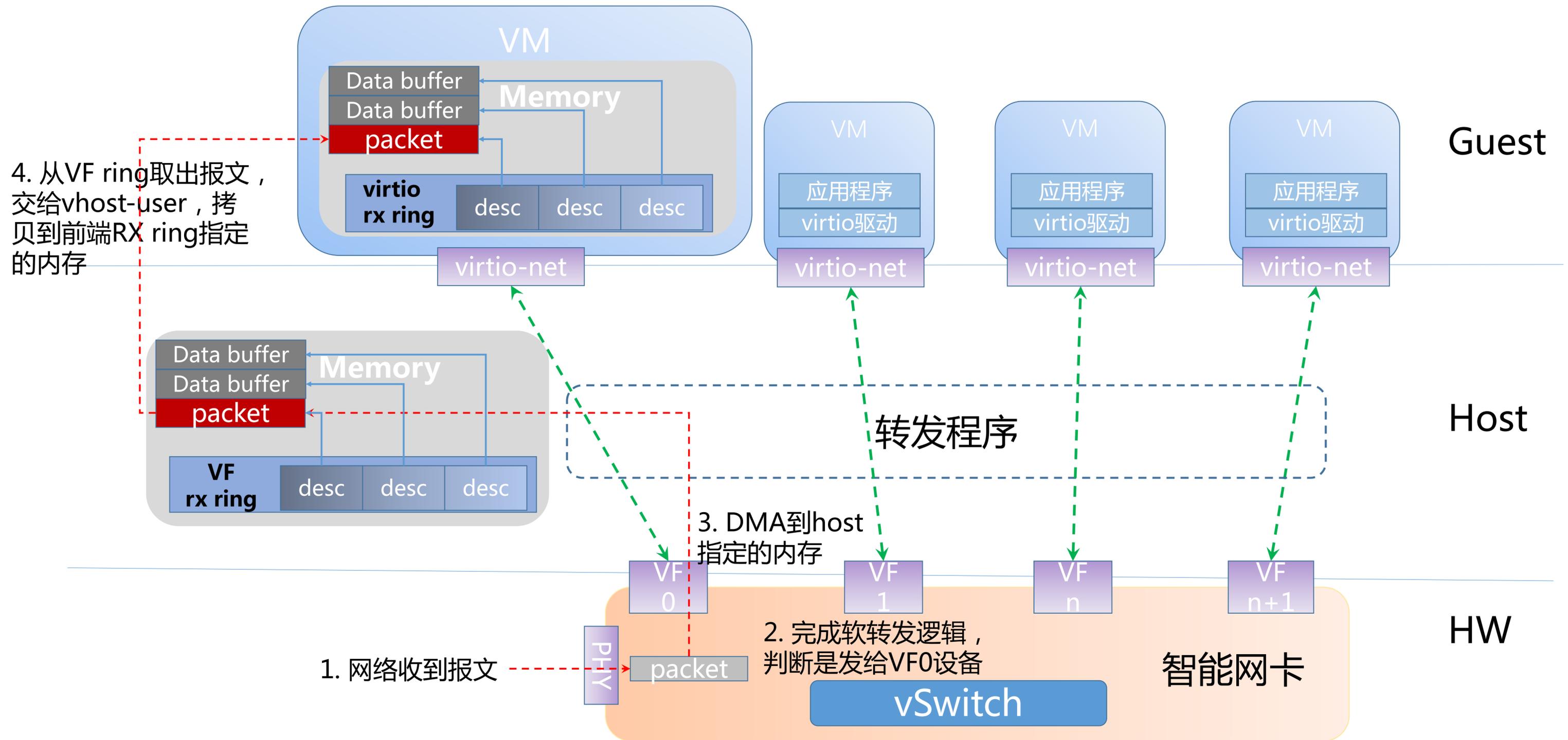
VM接收报文流程



VM接收报文流程



VM接收报文流程

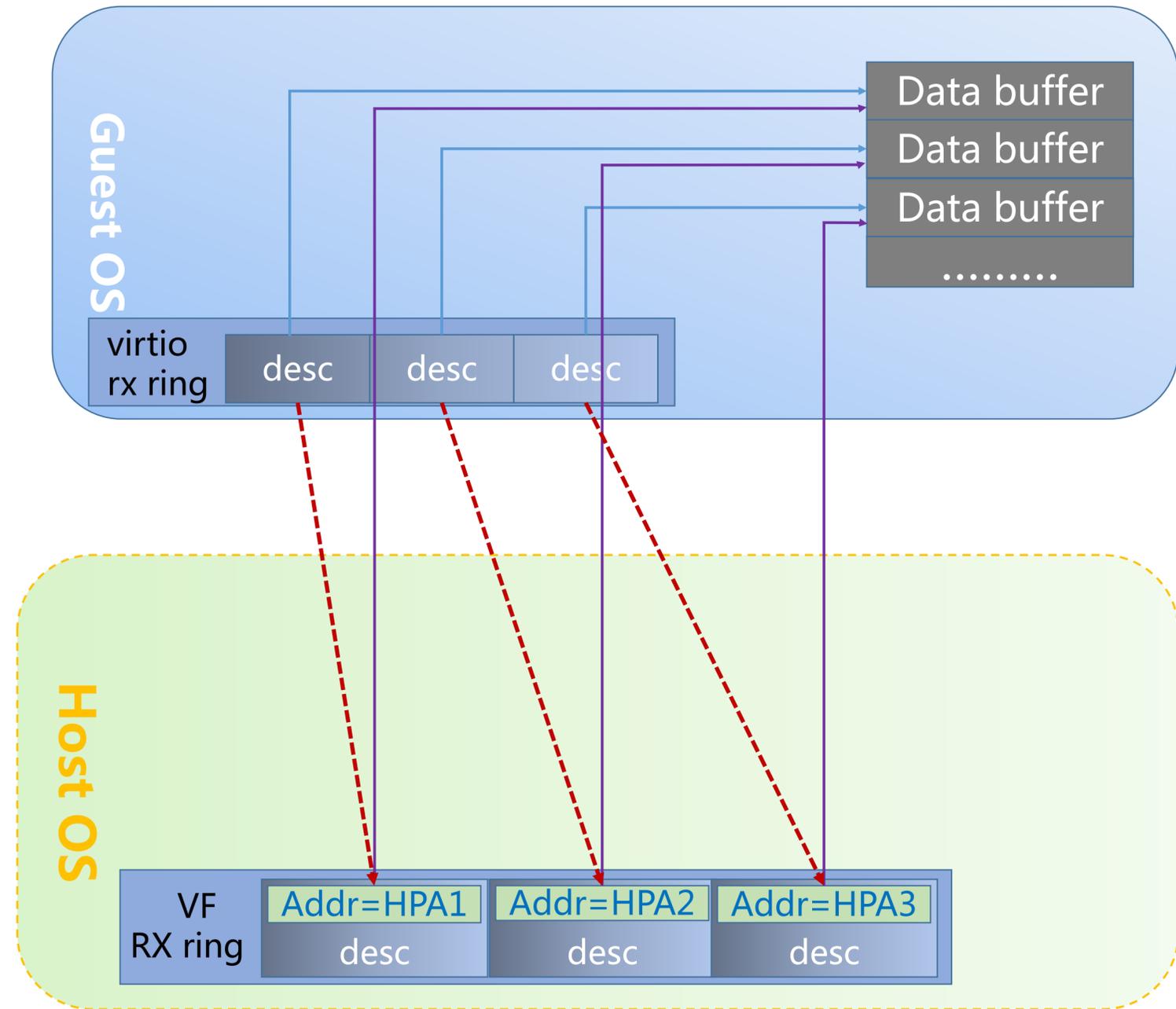


议程

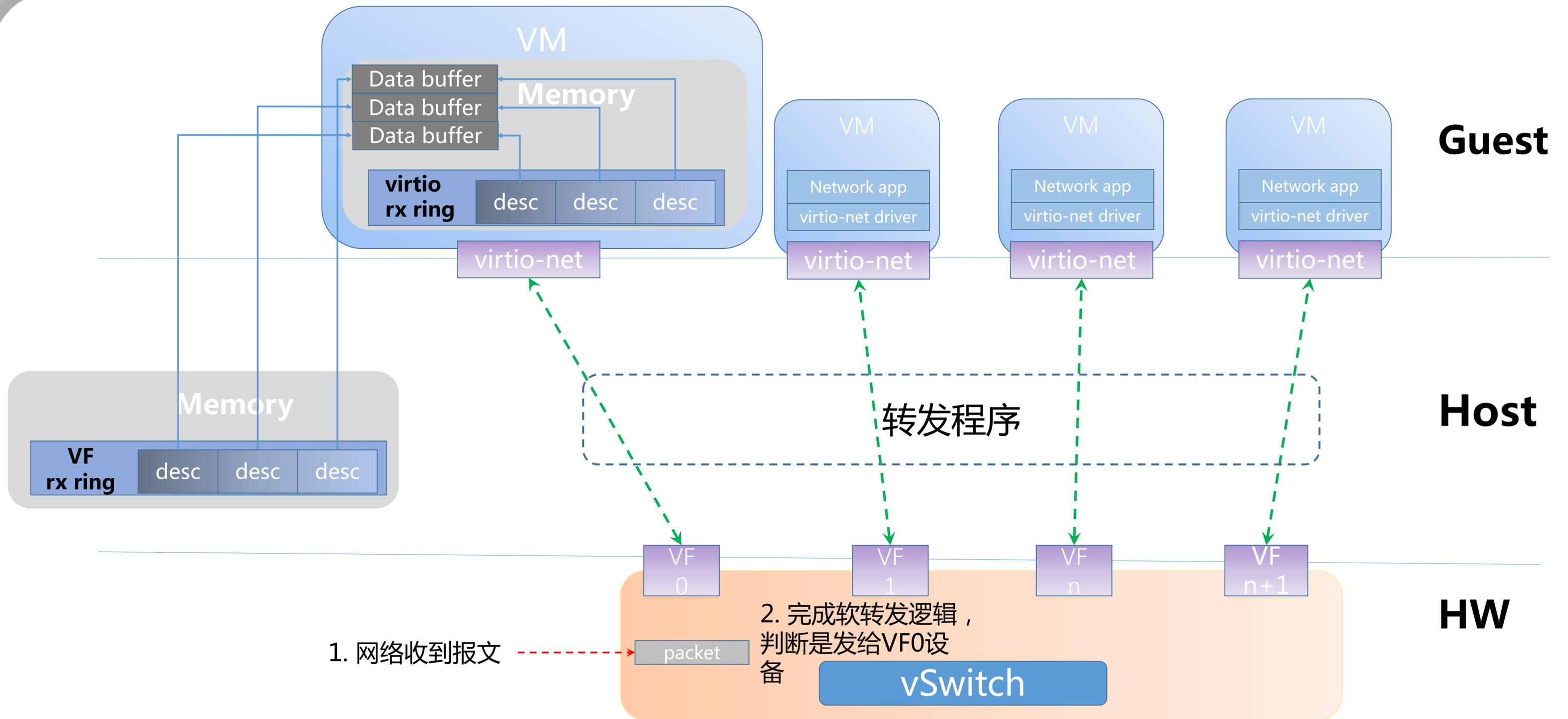
- 为什么需要智能网卡
- 智能网卡解决方案
- 接收端的数据通路
- **零拷贝优化方案**
- 零拷贝的优势

Ring同步模块

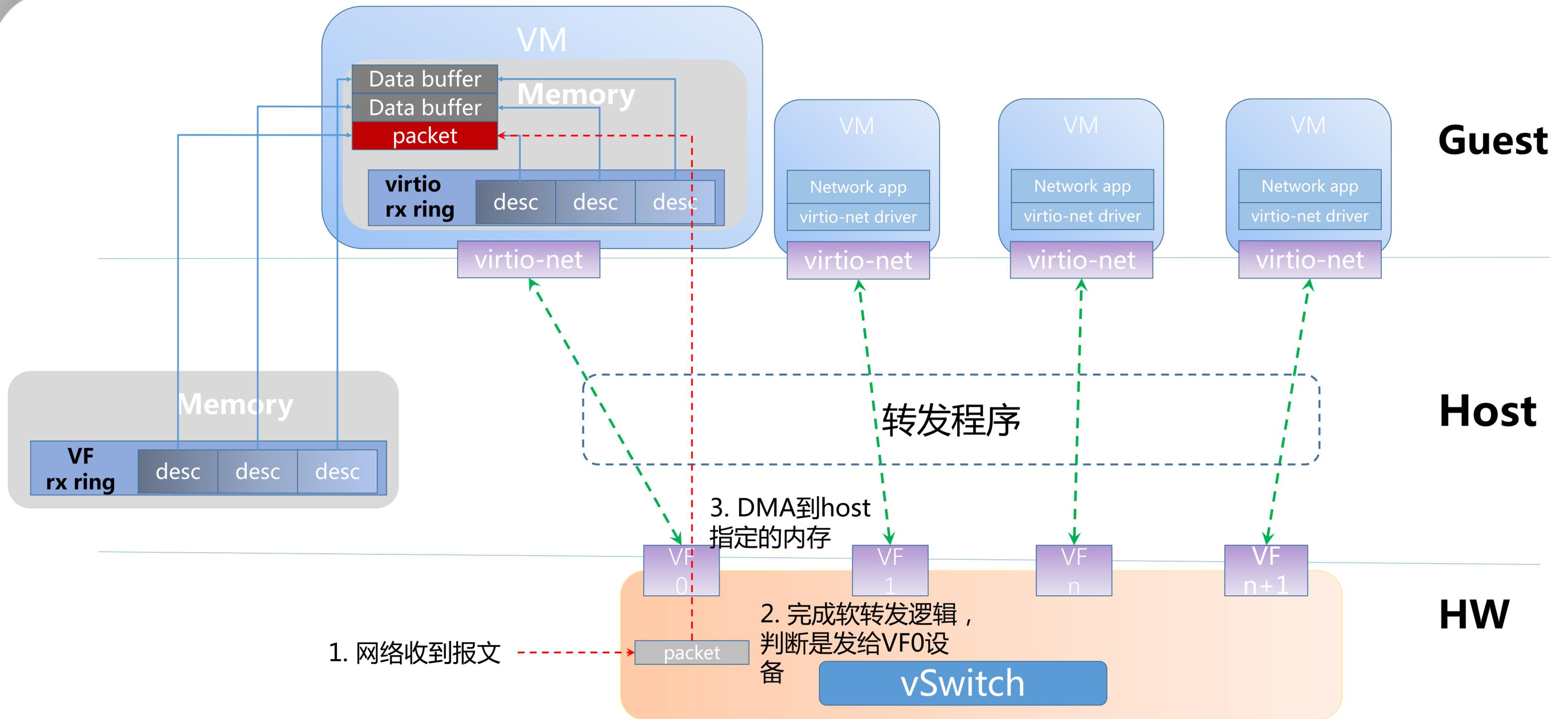
- 监控Guest ring和host ring的描述符更新
- 同步Guest ring和Host ring描述符
- 使Host ring的描述符的数据缓冲区指针指向Guest memory
- GPA to HPA转换



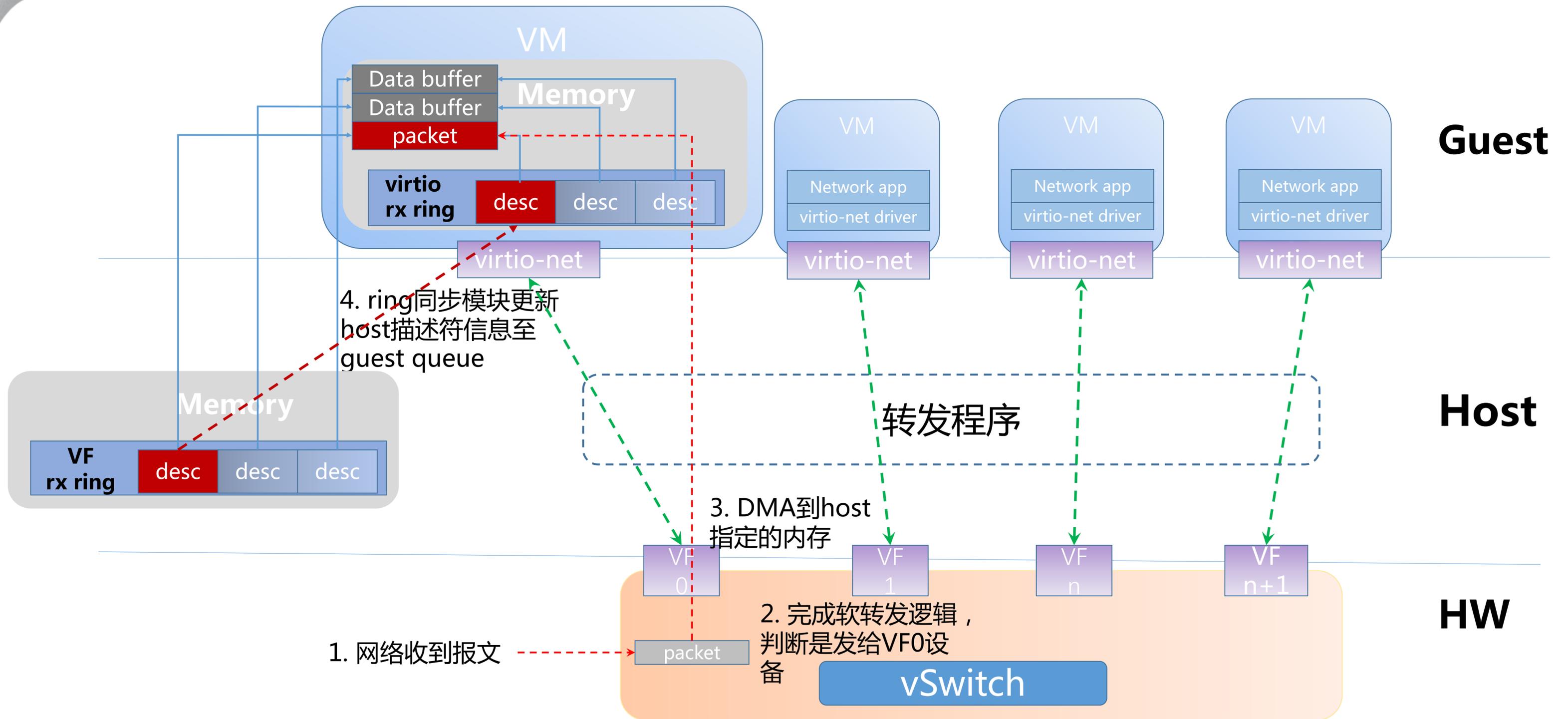
VM接收报文流程 - 零拷贝模式



VM接收报文流程 - 零拷贝模式



VM接收报文流程 - 零拷贝模式



议程

- 为什么需要智能网卡
- 智能网卡解决方案
- 接收端的数据通路
- 零拷贝优化方案
- **零拷贝的优势**

优势

- 性能
 - 接收端减少了内存拷贝
 - Footprint减少，避免cache-coherency
 - 总体性能提高了40%
- 时延
 - 减少拷贝带来的时间开销

为了无法计算的价值 |  阿里云

