

API virtualization

- intercept & interject functionality
add

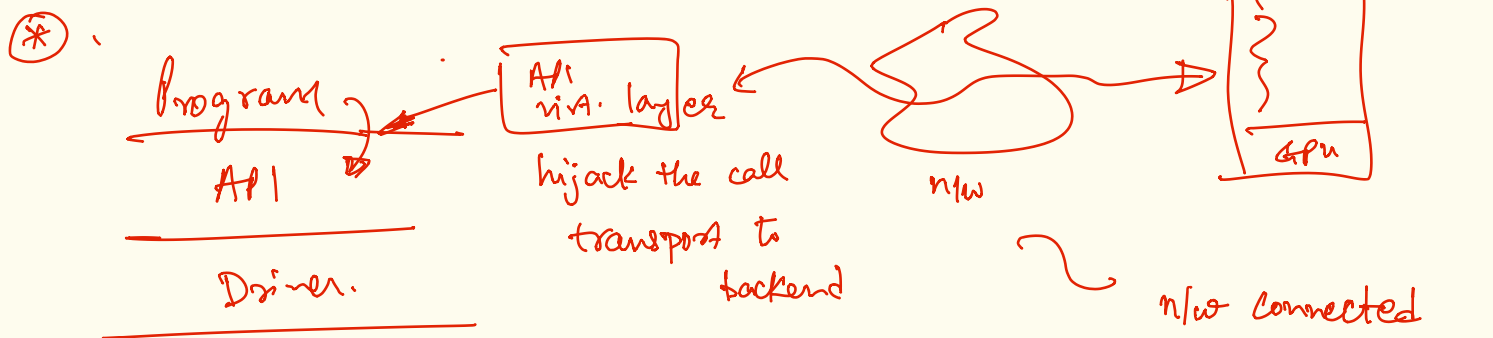
at the API layer
below

- clients / programs use API to build logic & API calls are executed transparently by hypervisor intervention.

- nvidia-smi

- decoupling of access & device.

the API is the interface for interjection.



- (*) cudaMalloc
- cuda copy H to D

system virtualization - VM
m/c

process group - containers
execution/runtime vA.

e.g: GPU access
on the cloud
network.

