# Smart TSC scaling

**Eddie Dong/Kevin Tian**

# Problem

A VM running on top of host A, TSC frequency f0, may be migrated to host B, TSC frequency f1.

# Principle of TSC virtualization

Guest see monotonic increasing value

- Backward of TSC may cause SW to think of huge time elapsed

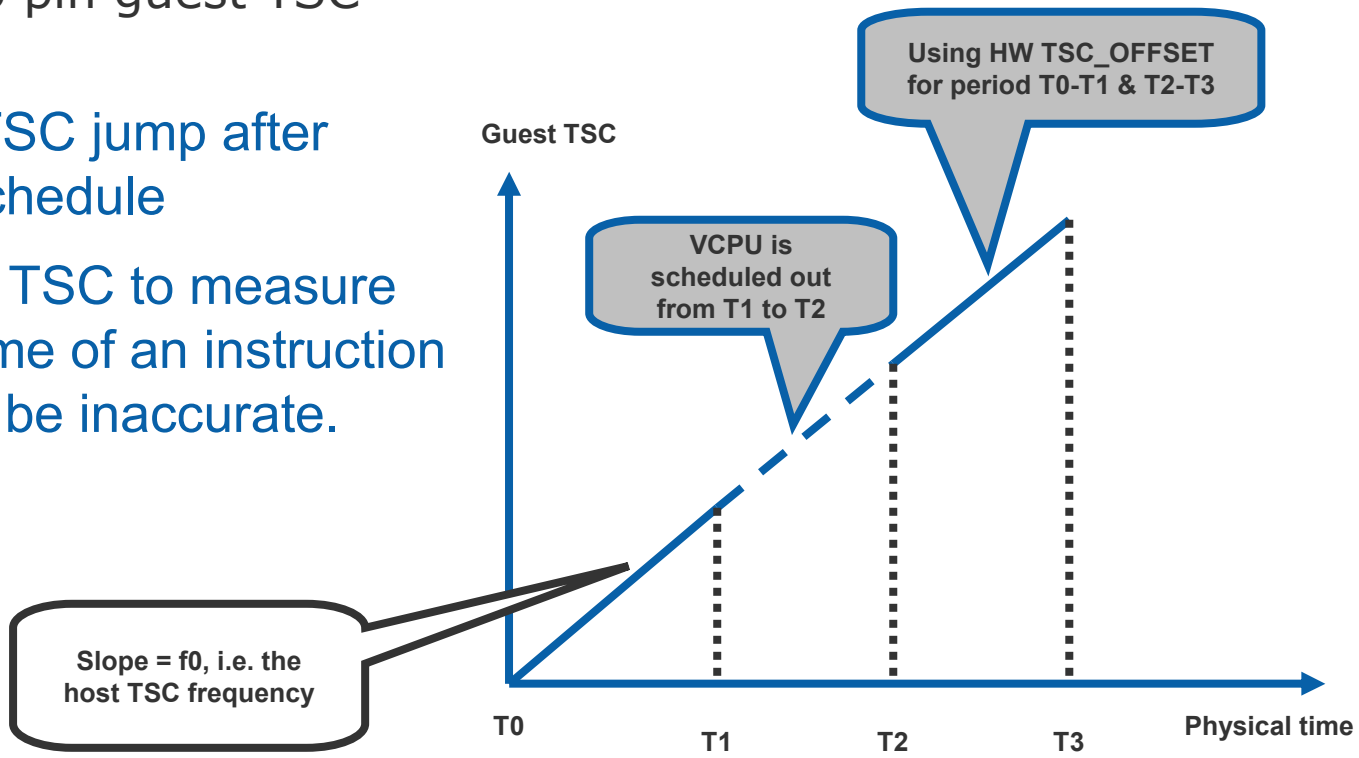
Time elapsed (delta time) from TSC is sync (same) with real time (wall clock) in long time run

- Guest SW can't expect to see same with native delta TSC for an execution of certain instructions.
  - VCPU may be de-scheduled – known issue, for example some micro benchmark won't be trusted in virtual environment.

# Current TSC virtualization

Both Xen & KVM use HW TSC_OFFSET to pin guest TSC with host TSC

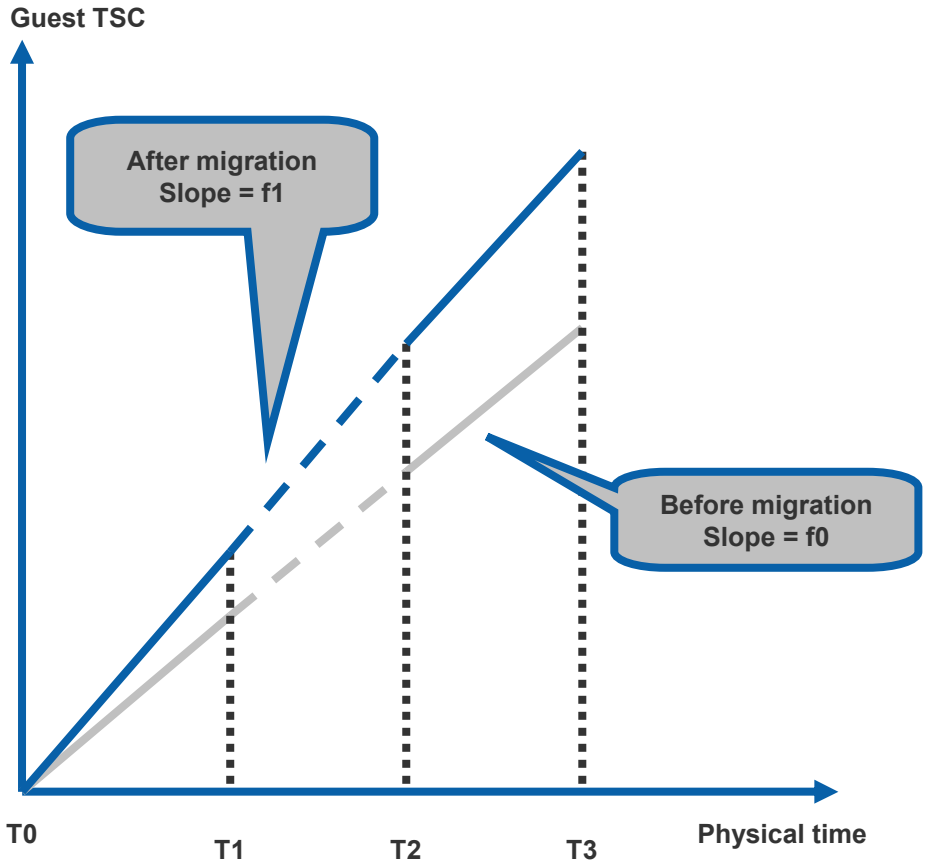
- Guest see TSC jump after VCPU de-schedule
- Using guest TSC to measure execution time of an instruction stream may be inaccurate.



# No scaling (current approach)

Guest delta TSC may be out of sync with wall clock

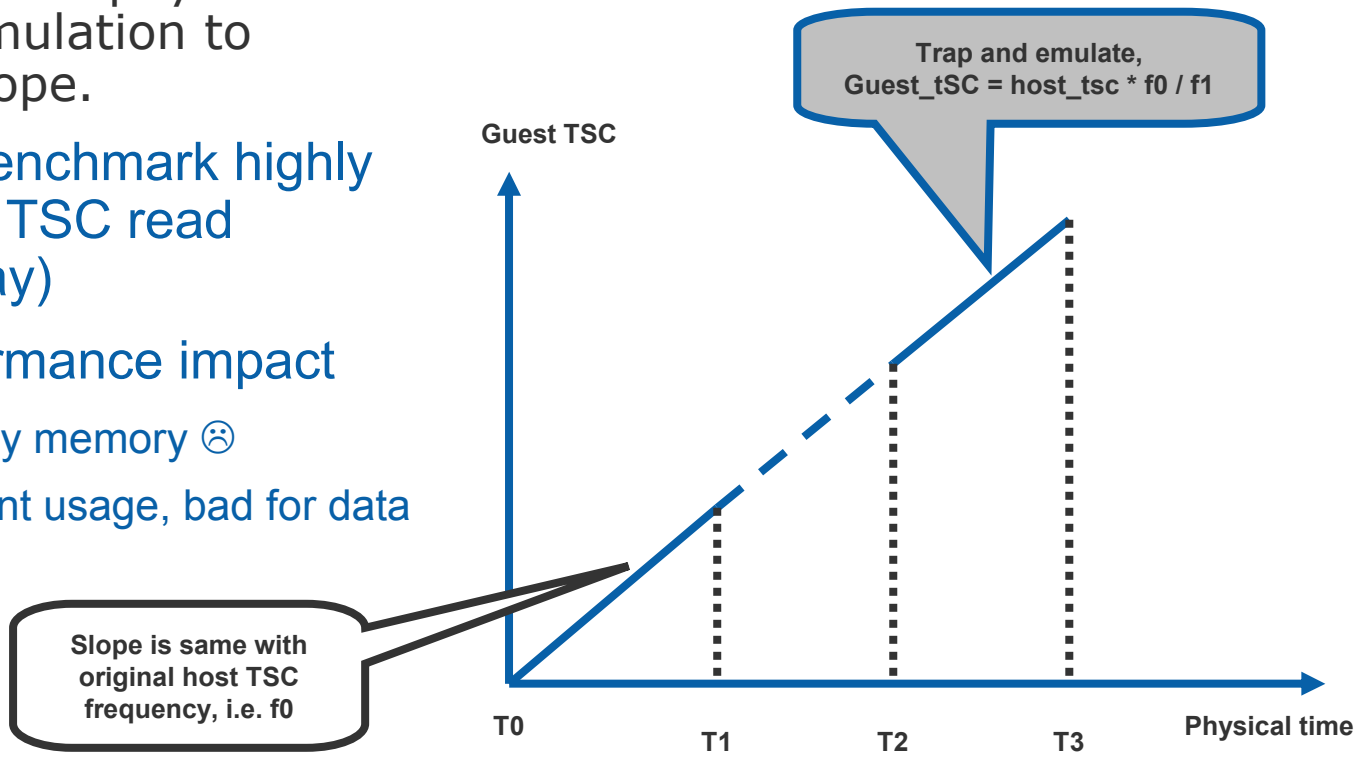
- Any side effect?
- Need more test



# SW re-scaling (or trap and emulate)

Guest see same TSC with pre-migration, with the payment of SW trap and emulation to maintain the slope.

- Database benchmark highly depends on TSC read (gettimeofday)
- ~10% performance impact
  - Base on my memory ☹
  - OK for client usage, bad for data center

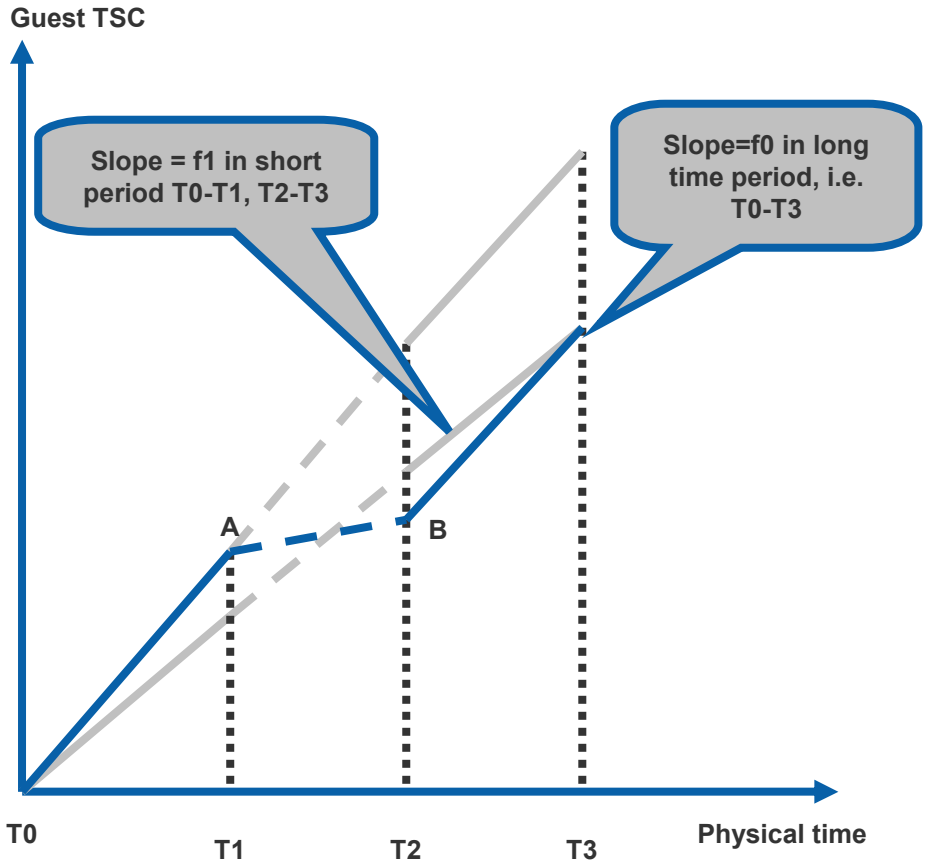


# Smart scaling (alt3)

Using HW TSC_OFFSET when VCPU is running, but re-set TSC_OFFSET when scheduled in to sync guest TSC with wall clock in long time

- From T0 to T3, delta TSC is synced with wall clock.
- TSC_OFFSET is used in T0-T1 and T2-T3.

Fall back to SW rescaling if gTSC can't be maintained monotonically from point A to B



# Smart scaling: Detail consideration

How to choose right rebalancing period, i.e. T0-T3

- Simplest way is to choose 3 scheduler ticks (2 schedule in ticks + 1 schedule out ticks), but then it may have to fall back to SW rescaling when f1 > 150% f0.
  - — Different weight may have different schedule in vs. out ticks.
- Re-adjust per 100ms or 1000ms
- Xen has scheduled out call back API, so easy to implement.
  - — Check with KVM

Any side effect?

SMP consideration