

Describing P2PDMA capabilities in ACPI

Describing the missing information with ACPI HMAT

Leon Romanovsky <leonro@nvidia.com>

Lukas Wunner <lukas@wunner.de>

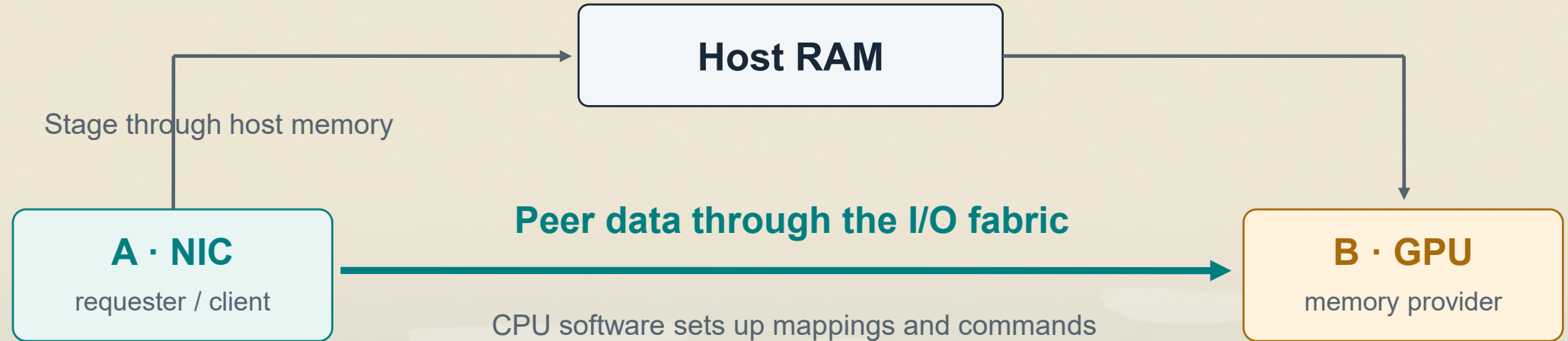
Mahesh Natu <mahesh.natu@intel.com>



LINUX
PLUMBERS
CONFERENCE

PRAGUE, CZECHIA / OCT. 5-7, 2026

P2P moves data between devices



Primary goal: usable P2PDMA between devices assigned to a VM

From peer traffic → routing constraints → missing guest information → HMAT

Describing P2PDMA capabilities in ACPI



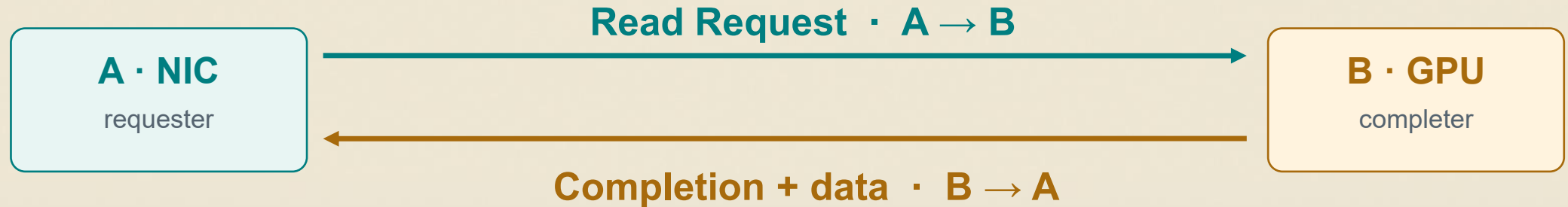
**LINUX
PLUMBERS CONFERENCE**

PRAGUE, CZECHIA / OCT. 5-7, 2026

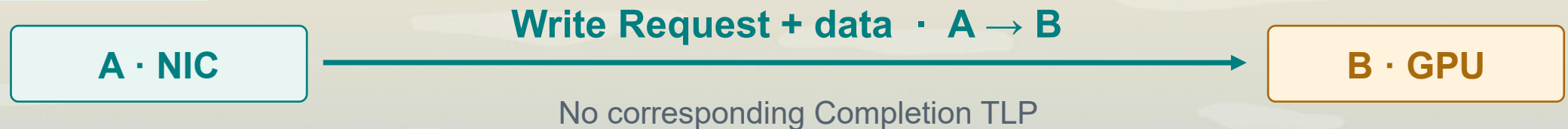


One operation can use two directions

MEMORY READ



POSTED WRITE



An A → B operation includes its return traffic. B → A is a different operation.

Describing P2PDMA capabilities in ACPI



**LINUX
PLUMBERS CONFERENCE**

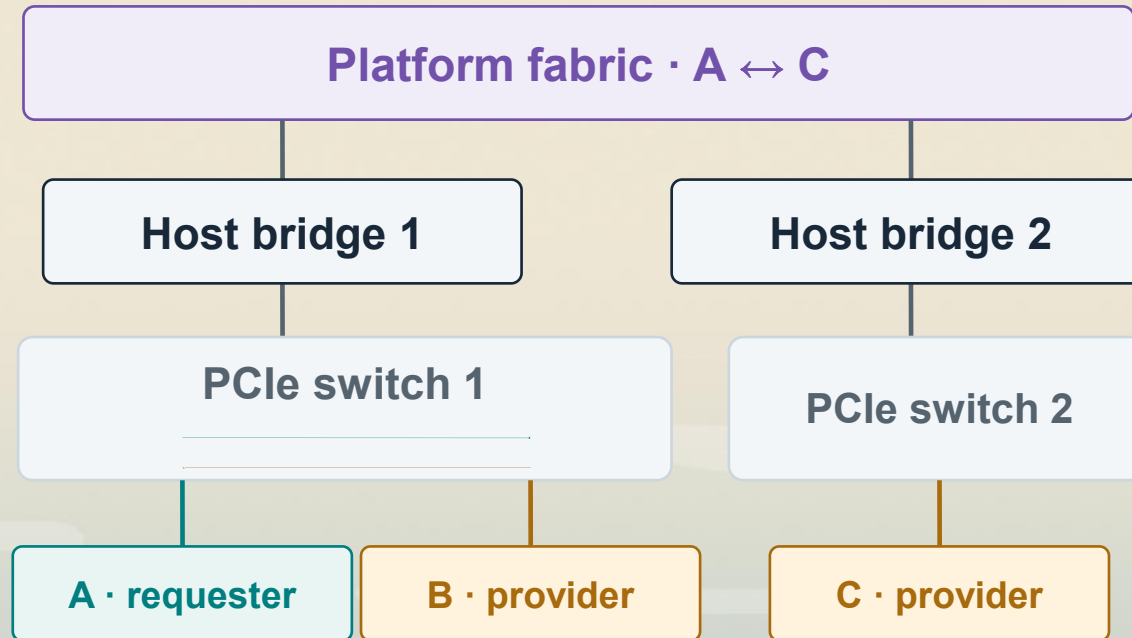
PRAGUE, CZECHIA / OCT. 5-7, 2026



3 / 17

ACS treats the two directions separately

ACS SERIES



Request on A's branch

Redirect / Egress Control
Translated traffic has extra rules

Completion on target's branch

Completion Redirect
Depends on Completion ordering

A ↔ B: shared switch. A ↔ C: separate host bridges; platform support required.

Describing P2PDMA capabilities in ACPI



**LINUX
PLUMBERS CONFERENCE**

PRAGUE, CZECHIA / OCT. 5-7, 2026

Same pair. Four routing answers.

Fixed example: A port = RR + DT · B port = CR · TB clear

Request address	Completion	Request route	Completion route
Untranslated	Strict	Upstream	Upstream
Untranslated	Relaxed	Upstream	Direct
Translated	Strict	Direct	Upstream
Translated	Relaxed	Direct	Direct

Reverse operation, B → A: all four routing combinations are direct in this example.

Routing only. Strict = RO clear; relaxed = RO set. Upstream needs platform support.

RR: Request Redirect · DT: Direct Translated P2P · CR: Completion Redirect · TB: Translation Blocking

Describing P2PDMA capabilities in ACPI



LINUX
PLUMBERS CONFERENCE

PRAGUE, CZECHIA / OCT. 5-7, 2026



A direct route is not the whole answer

ACS SERIES

Device behavior matters

Address translation (ATS) affects Request address type.

A direct route may still lack a usable mapping.

Why this matters in a VM

Virtual PCI topology hides physical path constraints.

QEMU needs to expose the hypervisor's knowledge.

Match actual device traffic with the physical path conditions.

Describing P2PDMA capabilities in ACPI



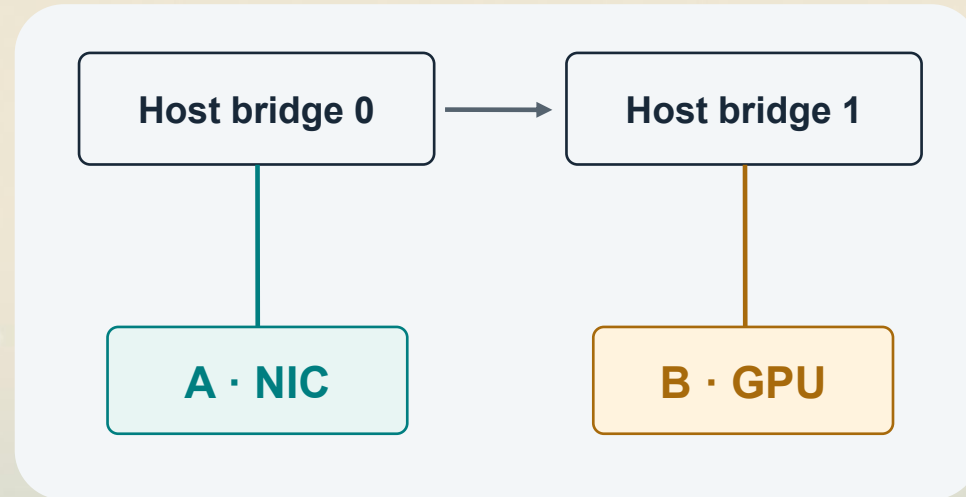
**LINUX
PLUMBERS CONFERENCE**

PRAGUE, CZECHIA / OCT. 5-7, 2026

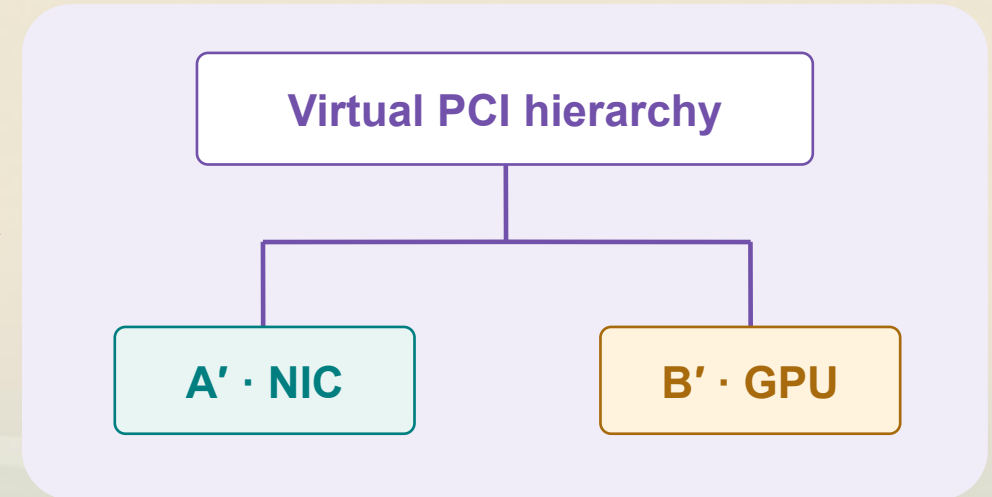


The guest sees a different topology

PHYSICAL HOST



GUEST VIEW



assign

Guest proximity does not reveal fabric rules or performance.

The hypervisor knows relationships that the guest cannot infer from PCI topology.

Describing P2PDMA capabilities in ACPI



LINUX
PLUMBERS CONFERENCE

PRAGUE, CZECHIA / OCT. 5-7, 2026

7 / 17



Describe the transaction the platform supports

01 Reachability

Can A complete an operation against B?

02 Applicability

Which TLP and address conditions apply?

03 Performance

Which usable provider best fits the transfer?

Primary use: QEMU describes hypervisor knowledge so the guest can decide.

Describing P2PDMA capabilities in ACPI



LINUX
PLUMBERS CONFERENCE

PRAGUE, CZECHIA / OCT. 5-7, 2026



HMAT: from host bridges to matrices

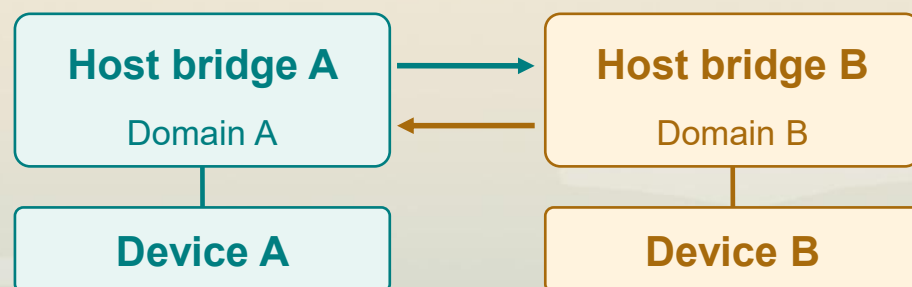
PROPOSED P2P USE

HMAT: Heterogeneous Memory Attribute Table

Memory Proximity Domain

A locality ID. In this P2P proposal, think PCIe host bridge.

Platform fabric



SRAT associates each bridge with its domain ID.

System Locality Latency and Bandwidth Information Structure

A 2D array: initiator rows × target columns.

Separate arrays for latency and bandwidth.

Latency (ns)

From ↓	A	B
A	—	900
B	600	—

Bandwidth (GB/s)

From ↓	A	B
A	—	12
B	20	—

Illustrative values for one transaction class.

QEMU maps host bridges to guest domains and fills the matrices from hypervisor knowledge.

Describing P2PDMA capabilities in ACPI



LINUX
PLUMBERS CONFERENCE

PRAGUE, CZECHIA / OCT. 5-7, 2026

9 / 17



Match the description before choosing a provider

REVISED INTEGRATION

1

Actual traffic

Request address type;
Completion attributes

**2**

Path constraints

Applicable ACS
and translation rules

**3**

Matching HMAT

Direction and
transaction conditions

**4**

Usable mapping

Then compare
matching performance

EARLIER RFC

One non-UIO result for fallback variants

REQUIRED EXTENSION

Keep each transaction's result distinct

A positive description must respect applicable routing and mapping constraints.

Describing P2PDMA capabilities in ACPI



**LINUX
PLUMBERS CONFERENCE**

PRAGUE, CZECHIA / OCT. 5-7, 2026



10 / 17

QEMU translates knowledge into guest tables

PROPOSED

- 1 Obtain validated peer knowledge
- 2 Join host and guest device identities
- 3 Preserve distinctions in guest domains
- 4 Emit guest SRAT + extended HMAT

Identity translation

Host A, B → guest A', B'

Host domains → guest domains

Keep direction and TLP conditions

Guest identities may change. The meaning of each peer relationship must survive.

Describing P2PDMA capabilities in ACPI



LINUX
PLUMBERS CONFERENCE

PRAGUE, CZECHIA / OCT. 5-7, 2026



11 / 17

Agree on the contract; build the virtualized path

ACS series

Routing by TLP class
Mapping constraints

HMAT RFC

Parser, lookup and ranking
Single non-UIO policy

Proposed next step

QEMU-generated HMAT
Guest transaction matching

Questions for a QEMU-to-guest support:

- 1 What peer knowledge does the hypervisor provide?
- 2 Which transaction properties and guarantees must HMAT express?
- 3 How should QEMU and the guest handle incomplete information?

Describing P2PDMA capabilities in ACPI



**LINUX
PLUMBERS CONFERENCE**

PRAGUE, CZECHIA / OCT. 5-7, 2026



12 / 17

Thanks



LINUX
PLUMBERS CONFERENCE

PRAGUE, CZECHIA / OCT. 5-7, 2026



References

Series and specification used in this presentation

TLP series

[\[PATCH v8 00/23\] PCI/P2PDMA:
Route peer-to-peer DMA by TLP class](#)

HMAT first RFC

[\[PATCH RFC 0/7\] PCI/HMAT:
Describe P2PDMA reachability and performance](#)

PCI specification

[PCI Express Base Specification Revision 7.0](#)

Revision used for the ACS examples

Describing P2PDMA capabilities in ACPI



**LINUX
PLUMBERS CONFERENCE**

PRAGUE, CZECHIA / OCT. 5-7, 2026



14 / 17

Backup: the ACS rules behind the example

Request: Translated + TB

Blocked on a traversed port

Request: Translated + DT, no TB

Direct despite applicable RR / EC

Other Requests

RR / EC determine upstream treatment

Completion without RO

CR redirects upstream

Completion with RO

CR does not redirect

Example controls: A = RR + DT, CR clear; B = CR, RR / EC / DT clear; TB clear.

Linux treats enabled EC conservatively here; it does not inspect the Egress Control Vector.

Global ATS can still withhold BUS_ADDR after a direct raw routing result.

Describing P2PDMA capabilities in ACPI



**LINUX
PLUMBERS CONFERENCE**

PRAGUE, CZECHIA / OCT. 5-7, 2026



15 / 17

Backup: semantic requirements for HMAT

Earlier ECR draft

Type 3; Generic Port domain axes
UIO / non-UIO applicability
Six latency / bandwidth selectors

1...0xFFFFE: quantitative entries
0xFFFF: unreachable; 0: reserved

Semantics to agree

Positive-entry operation guarantee
Request and Completion conditions
Unknown versus explicit rejection

Missing / contradictory records
Domain identity and lifecycle rules

UIO, Request address type and Completion RO are distinct applicability dimensions.

Describing P2PDMA capabilities in ACPI



**LINUX
PLUMBERS CONFERENCE**

PRAGUE, CZECHIA / OCT. 5-7, 2026



16 / 17

Backup: tests matrix

Direction	Swap initiator and target; preserve different answers.
TLP applicability	Change address type or Completion RO; match the right record.
Missing data	Distinguish no description from explicit unsupported.
Virtual identities	Keep relationships distinct when assigning guest domains.
Device behavior	Exercise global / per-mapping ATS and cached route answers.
Lifecycle	Validate the declared assignment, hotplug and migration policy.

Keep the guest PCI view fixed; vary the description and verify the decision.

Describing P2PDMA capabilities in ACPI



**LINUX
PLUMBERS CONFERENCE**

PRAGUE, CZECHIA / OCT. 5-7, 2026



17 / 17