

Linuxにおけるパケット処理機構の 性能評価に基づいたNFV導入の一検討

○村松 真†, 川島 龍太†, 中山 裕貴††, 林 經正††, 松尾 啓志†

† 名古屋工業大学大学院

†† 株式会社ボスコ・テクノロジーズ

ICM研究会
2016/03/11

研究目的



アウトライン

◆研究目的

- ✓ パケット処理機構と仮想NICバックエンド機構

◆評価内容

評価(1): 物理サーバ上でのパケット転送

評価(2): VMと仮想スイッチ間のパケット転送

評価(3): VM上でのパケット転送

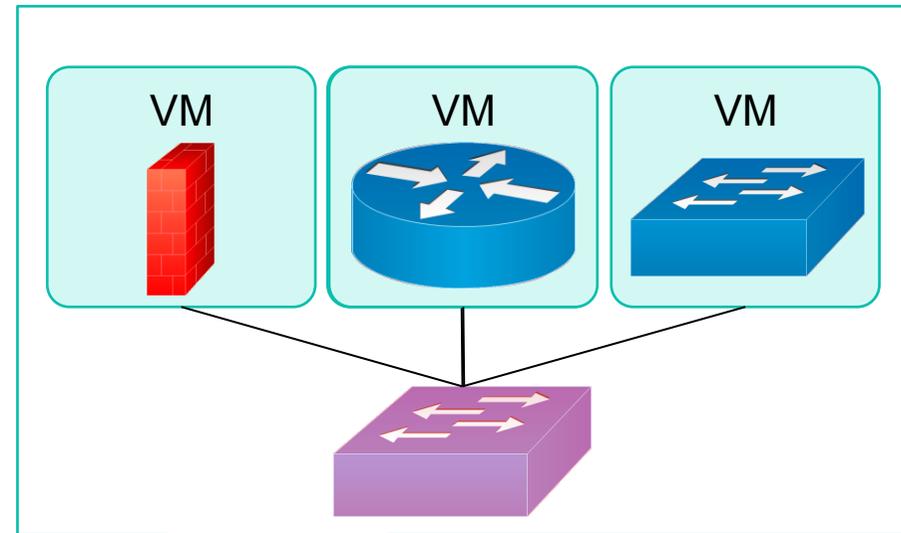
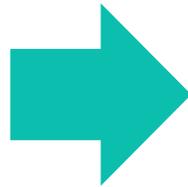
◆NFV環境のサーバ構成の検討

◆まとめと今後の課題

Network Functions Virtualization (NFV)

◆ソフトウェアでネットワーク機能を実現

専用ハードウェア機器

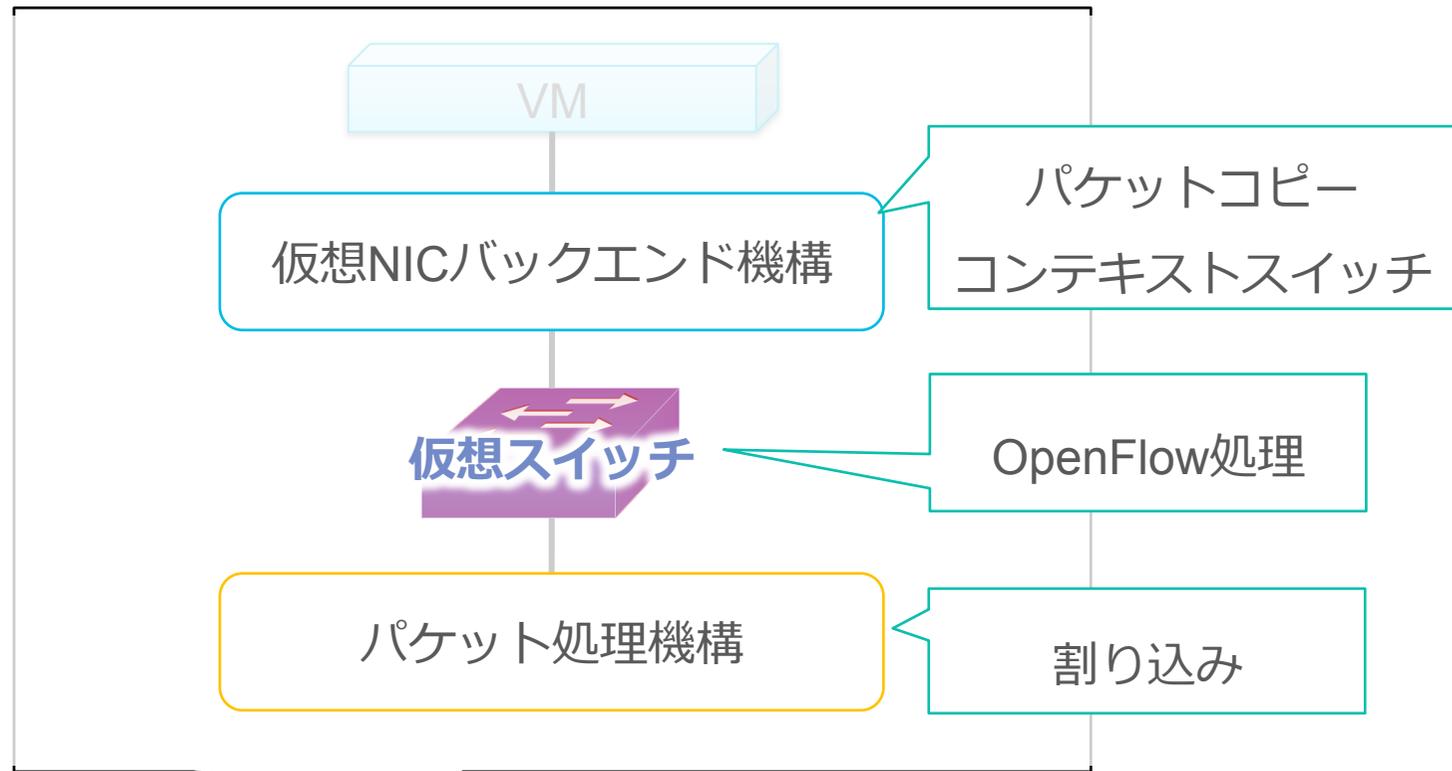


汎用IAサーバ

機器コスト・運用コストを削減

迅速かつ柔軟にサービスを提供可能

汎用IAサーバの課題



新たなパケット処理機構(仮想スイッチ)/仮想NICバックエンド機構が提案

パケット処理機構/仮想NICバックエンド機構

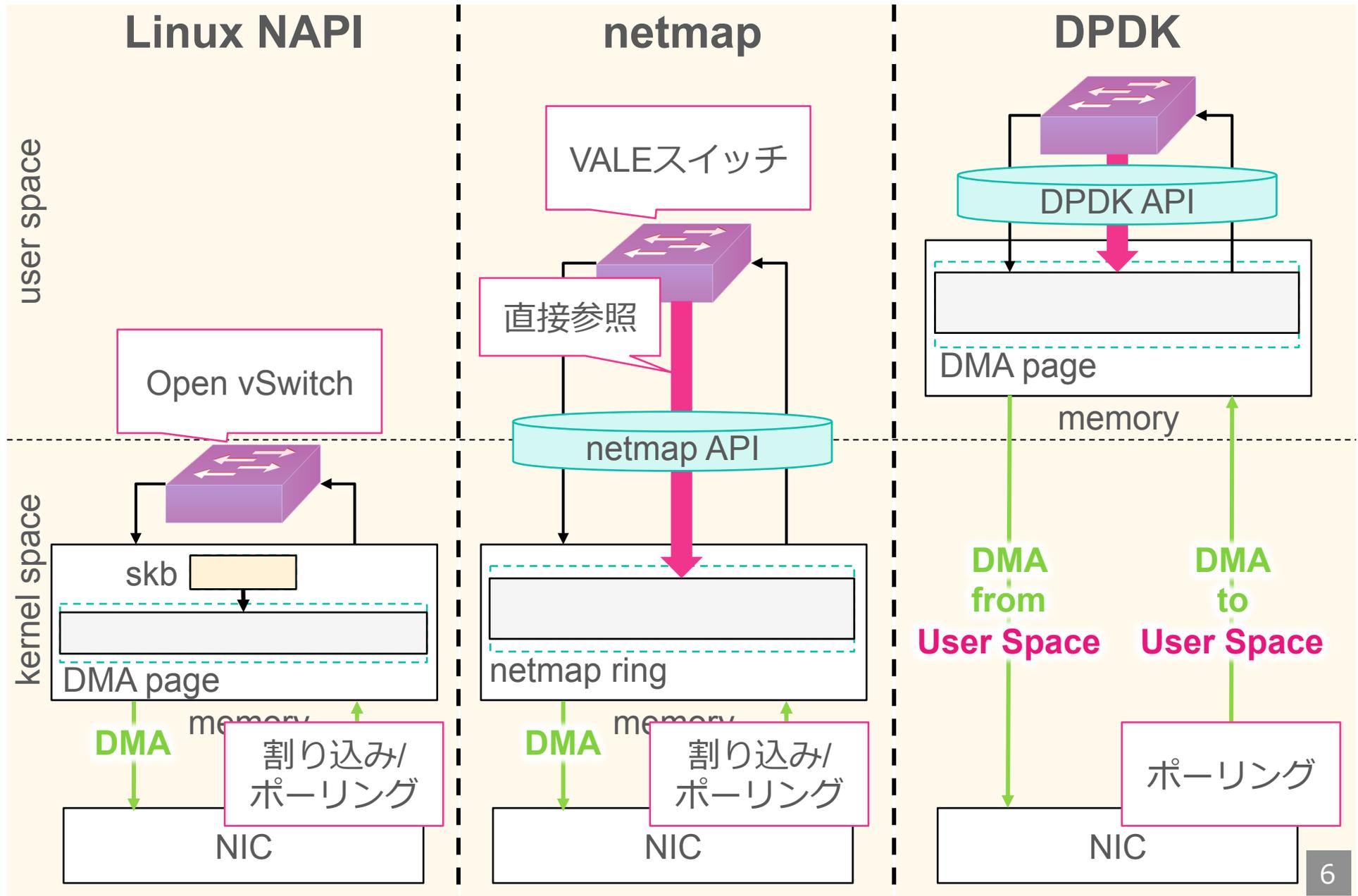
◆パケット処理機構 (仮想スイッチ)

- ✓ NAPI: Linux New API (Open vSwitch)
- ✓ netmap (VALE)
- ✓ Intel DPDK (OVS with DPDK, Lagopus)

◆仮想NICバックエンド機構

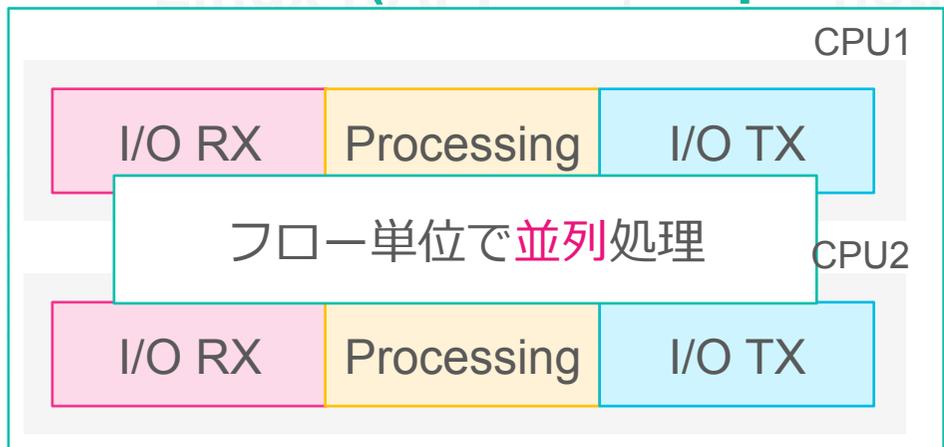
- ✓ vhost-net
- ✓ vhost-user

パケット処理機構と仮想スイッチ

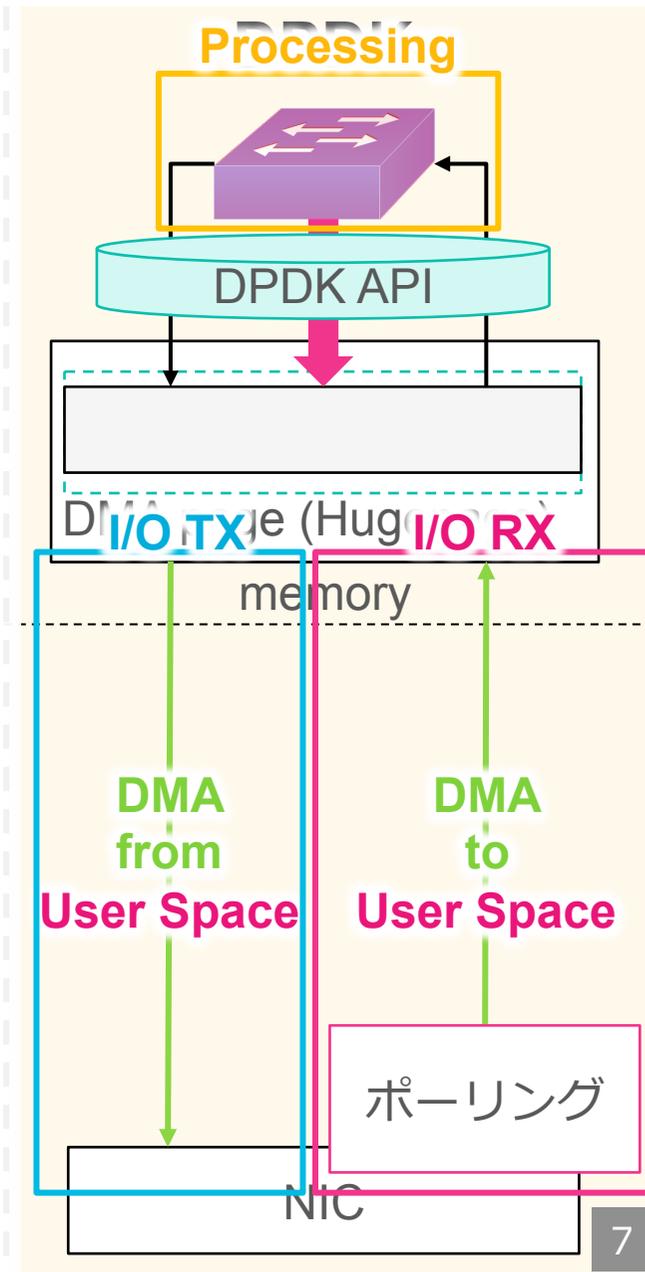
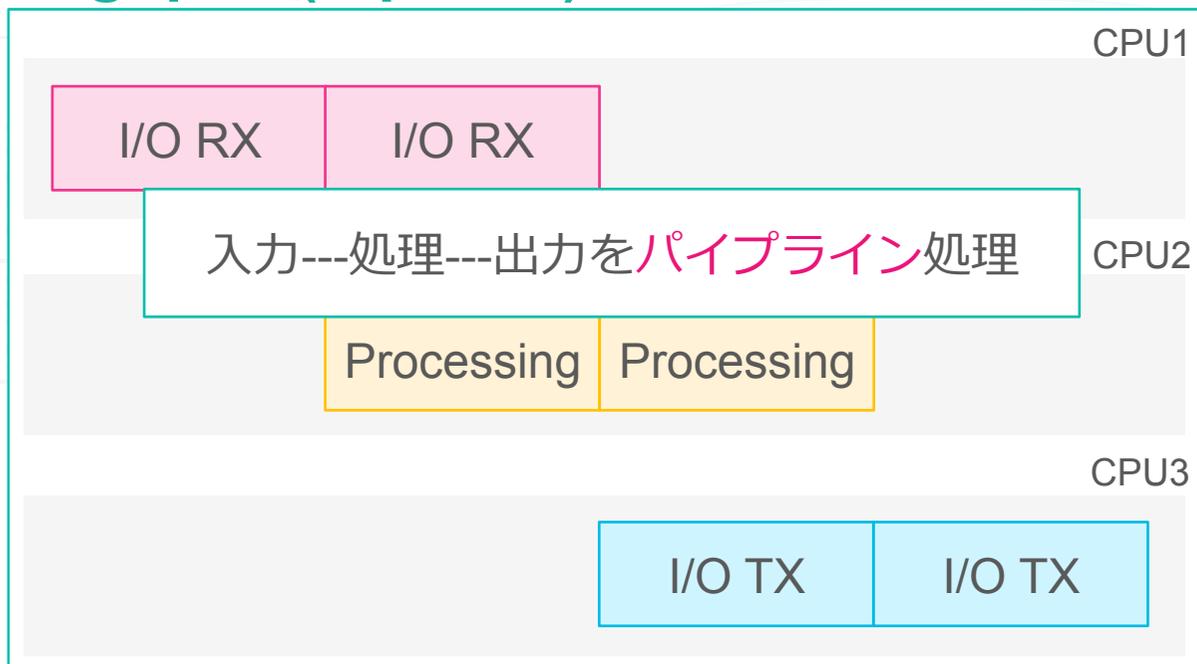


OVS/DPDKとLagopusの違い

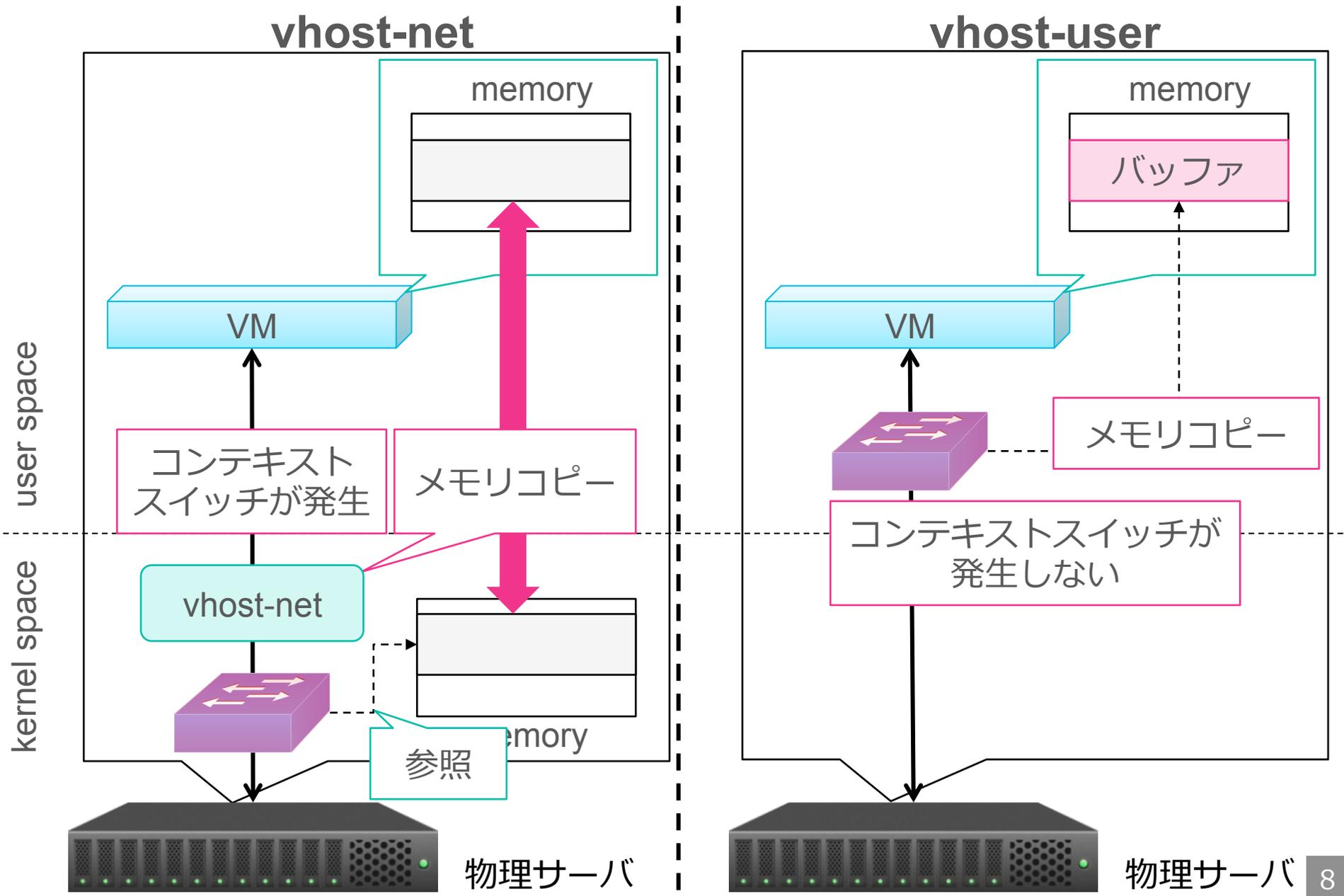
OVS/DPDK (Run to completion)



Lagopus (Pipeline)



仮想NICバックエンド機構



アウトライン

◆研究目的

- ✓ パケット処理機構と仮想NICバックエンド機構

◆評価内容

評価(1): 物理サーバ上でのパケット転送

評価(2): VMと仮想スイッチ間のパケット転送

評価(3): VM上でのパケット転送

◆NFV環境のサーバ構成の検討

◆まとめと今後の課題

本研究における評価対象と評価項目

物理環境の評価対象		仮想化環境の評価対象
パケット処理機構	仮想スイッチ	仮想NICバックエンド機構
Linux NAPI	Open vSwitch	vhost-net
netmap	VALE	vhost-user
Intel DPDK	Lagopus	

評価項目			
フロー数	エントリ数	チェーニング数	VMのパケット処理機構
単一フロー	単一エントリ	単一VM	Linux NAPI
複数フロー	複数エントリ	複数VM	netmap
			Intel DPDK

{
 パケット転送速度
 パケット転送遅延時間・遅延ジッタ

アウトライン

◆研究目的

- ✓ パケット処理機構と仮想NICバックエンド機構

◆評価内容

評価(1): 物理サーバ上でのパケット転送

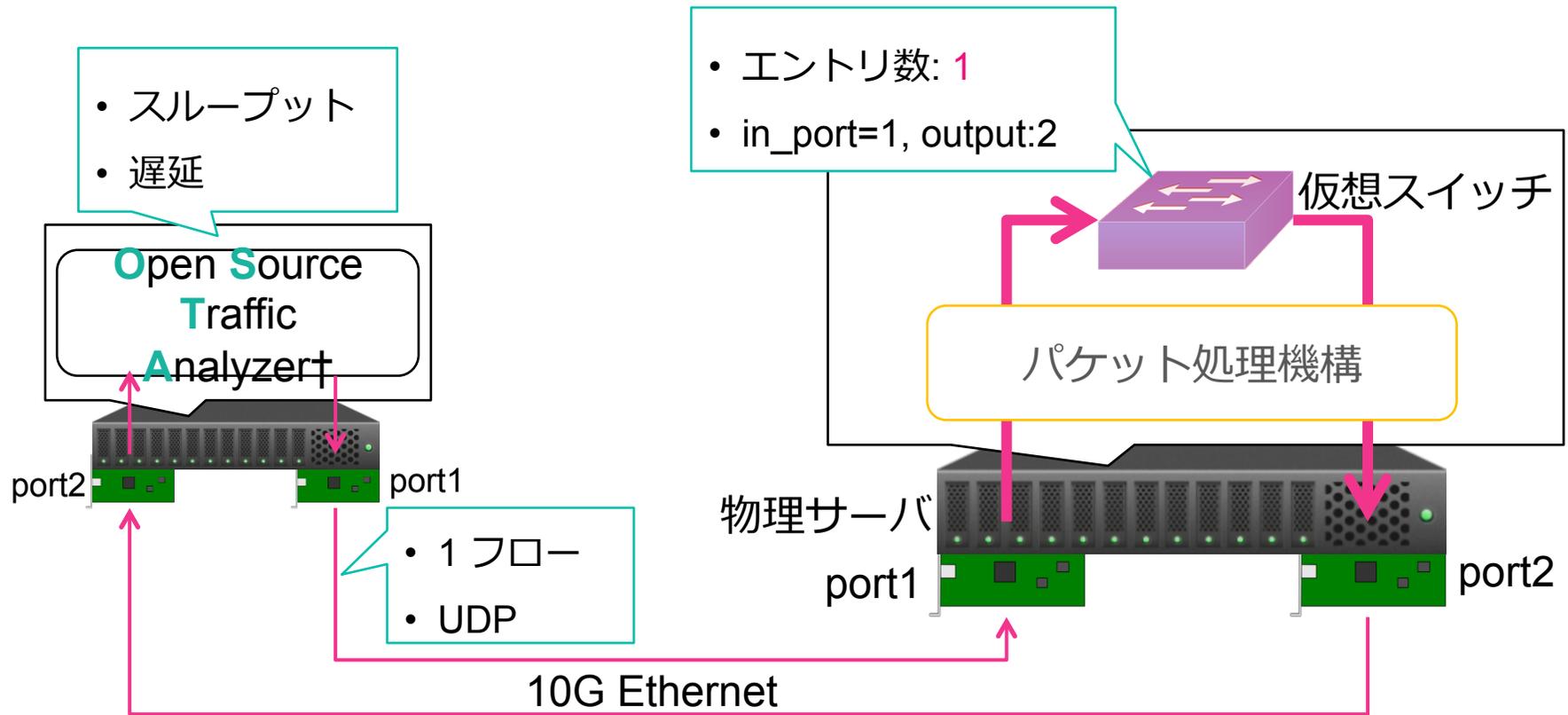
評価(2): VMと仮想スイッチ間のパケット転送

評価(3): VM上でのパケット転送

◆NFV環境のサーバ構成の検討

◆まとめと今後の課題

評価(1): 評価環境



物理サーバの性能

CPU	Intel Core i7 3.4GHz (4core, HT:off)
Memory	16GB
NIC	Intel X540-T2 10Gb Ethernet

ソフトウェアバージョン

OS	Fedora 22 (4.0.4-301-x86 64)
OVS	v2.4.0
netmap	v11.1
Lagopus	v0.2.3
DPDK	v2.1.0

† Open Source Traffic Analyzer , <https://people.kth.se/~danieltt/pktgen/>

評価(1): 評価対象/評価内容

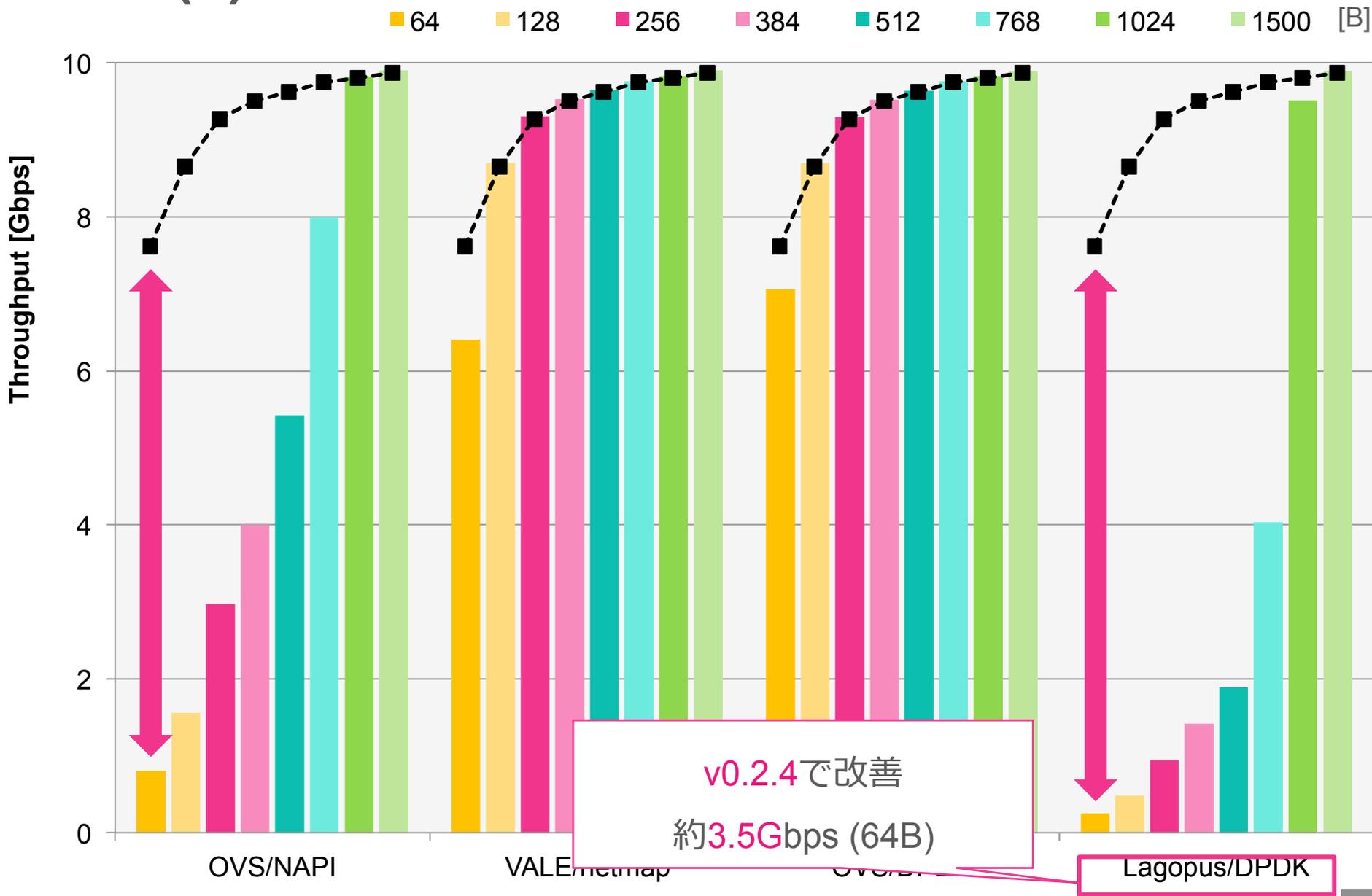
◆ 評価対象

	OVS/NAPI	VALE/netmap	OVS/DPDK	Lagopus/DPDK
仮想スイッチ (動作空間)	OVS (Kernel)	VALE (User)	OVS (User)	Lagopus (User)
パケット処理機構	NAPI	netmap	DPDK	DPDK

◆ 評価内容

- ✓ パケット転送速度
- ✓ パケット転送遅延

評価(1): パケット転送速度



評価(1): 1パケットに必要なCPUサイクル数†

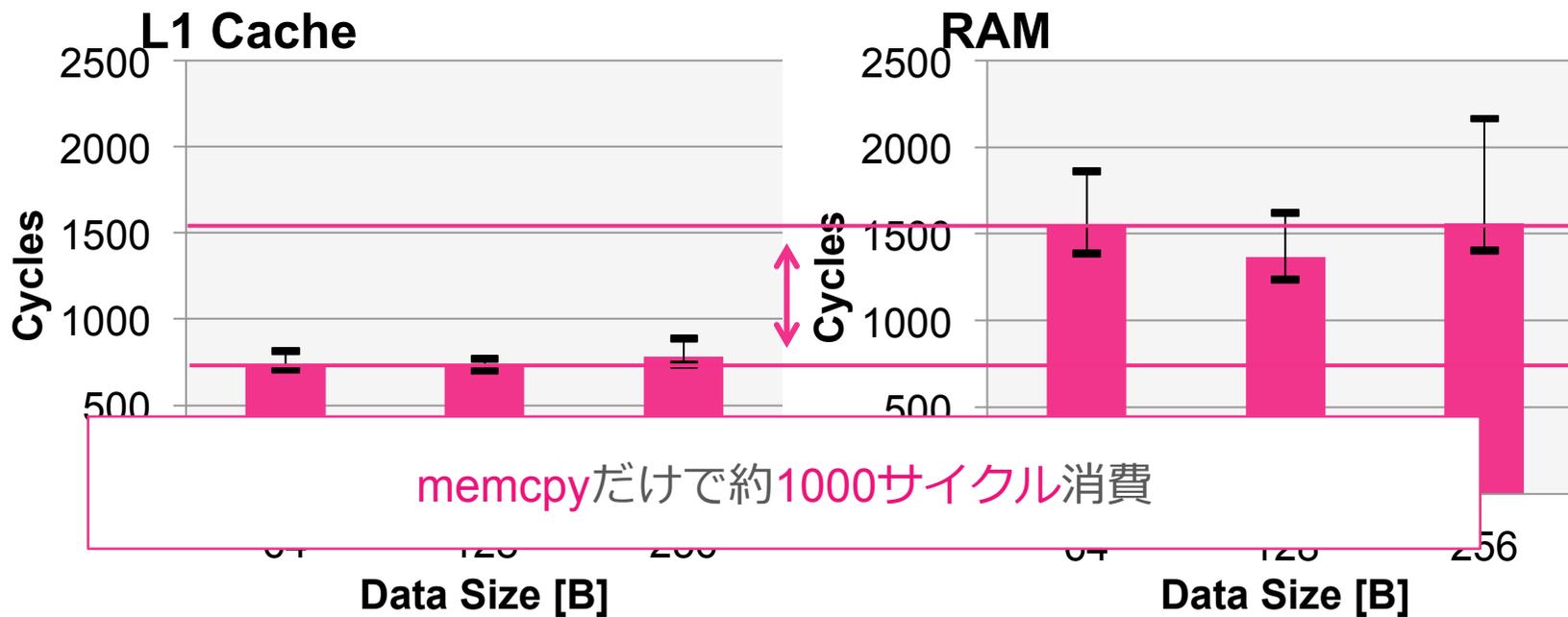
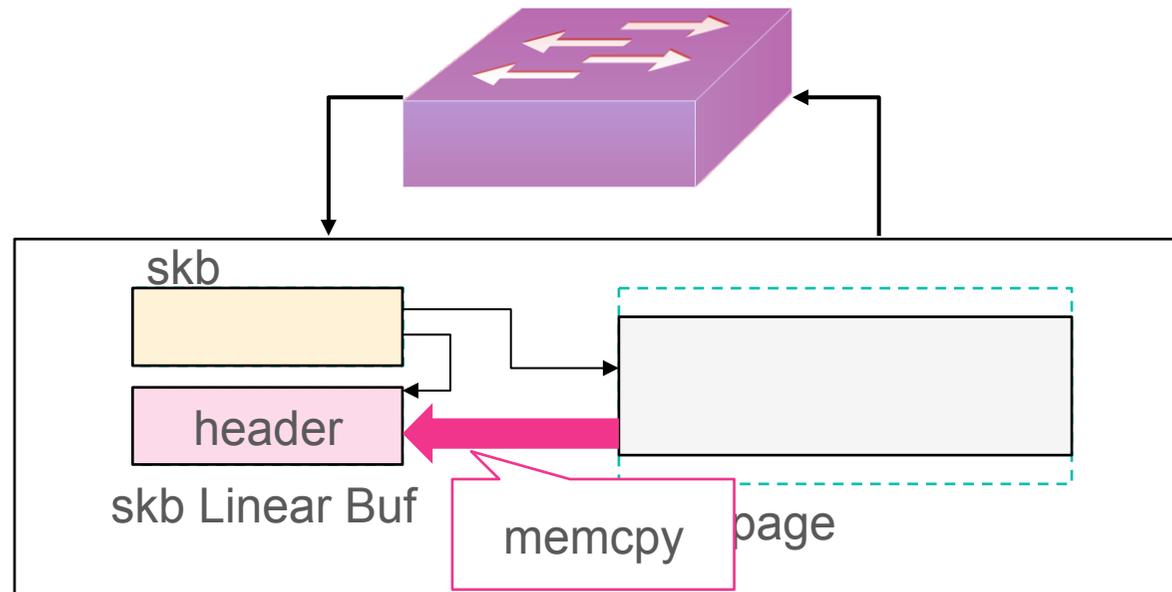
◆ CPU Hz (3.4G) ÷ pps から概算

OVS/NAPI	VALE/netmap	OVS/DPDK	Lagopus/DPDK
2,294	288	262	7,389

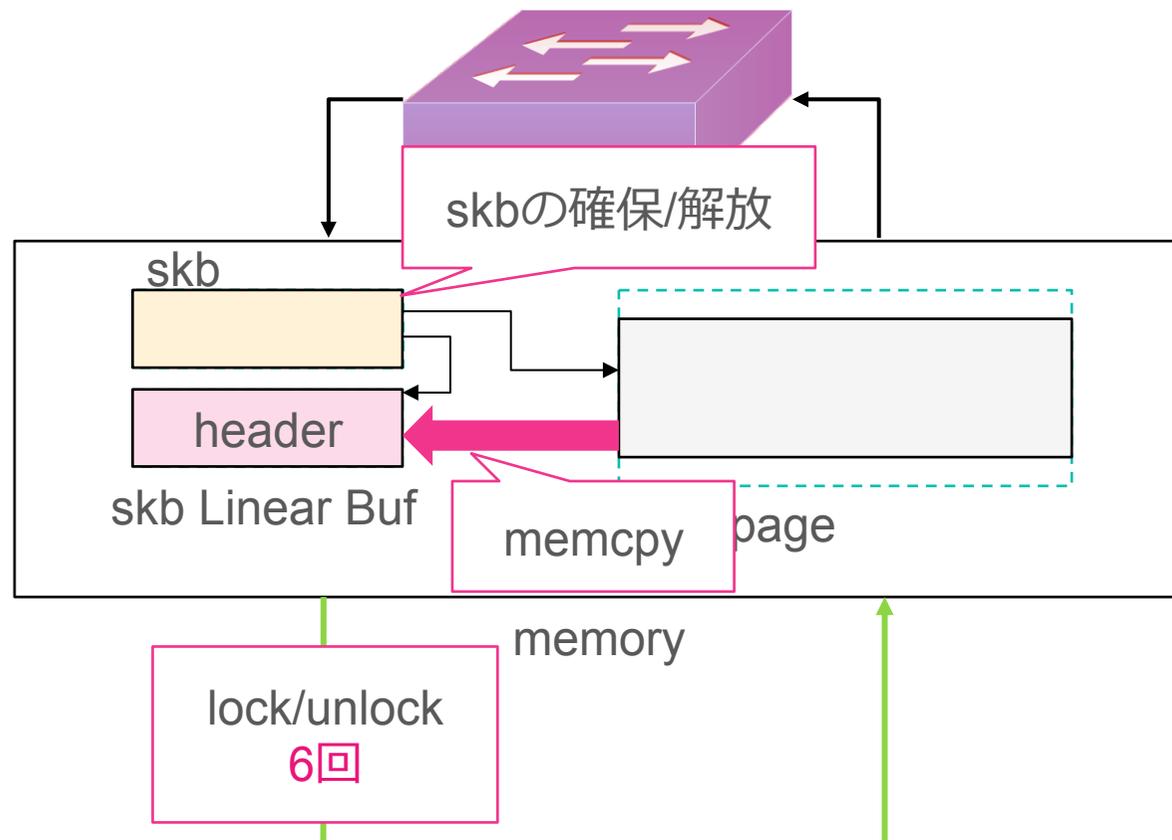
OVS/NAPIに焦点を当てて要因を考察

† Assessing Soft- and Hardware Bottlenecks in PC-based Packet Forwarding Systems, Paul Emmerich et. al, Technische Universität München, ICN 2015

評価(1): OVS/NAPIの性能要因



評価(1): OVS/NAPIの性能要因



根源的な処理に約50%(1000~/2294)のサイクル数を消費

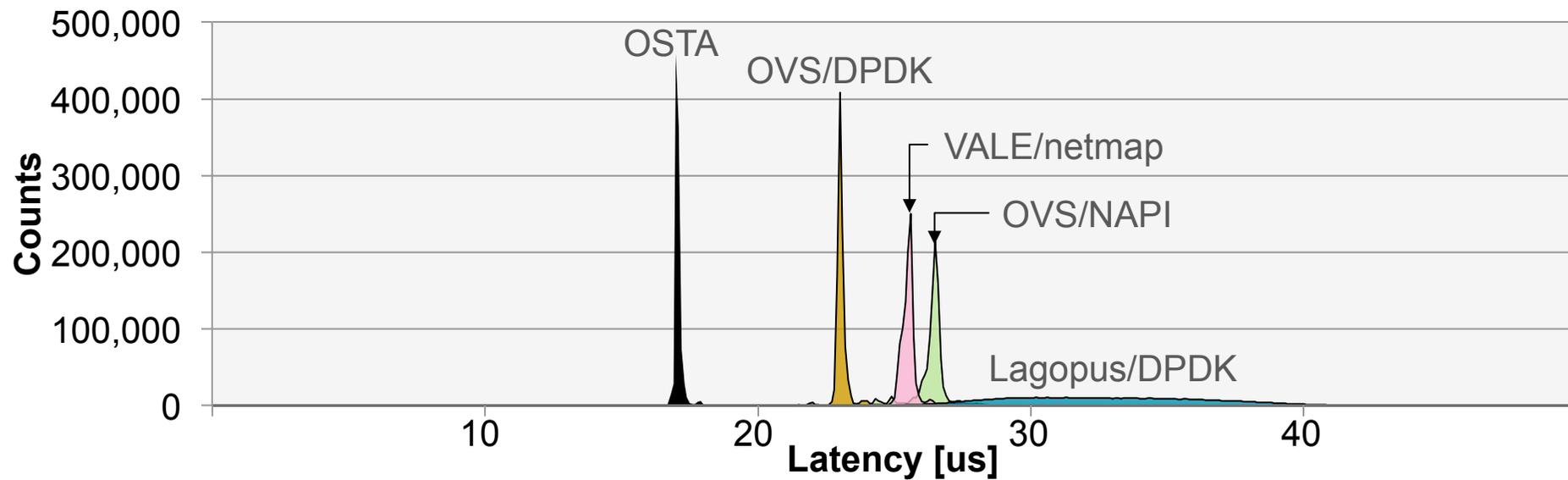


netmap, DPDKはゼロコピー, ロック回数の削減により高速化を実現

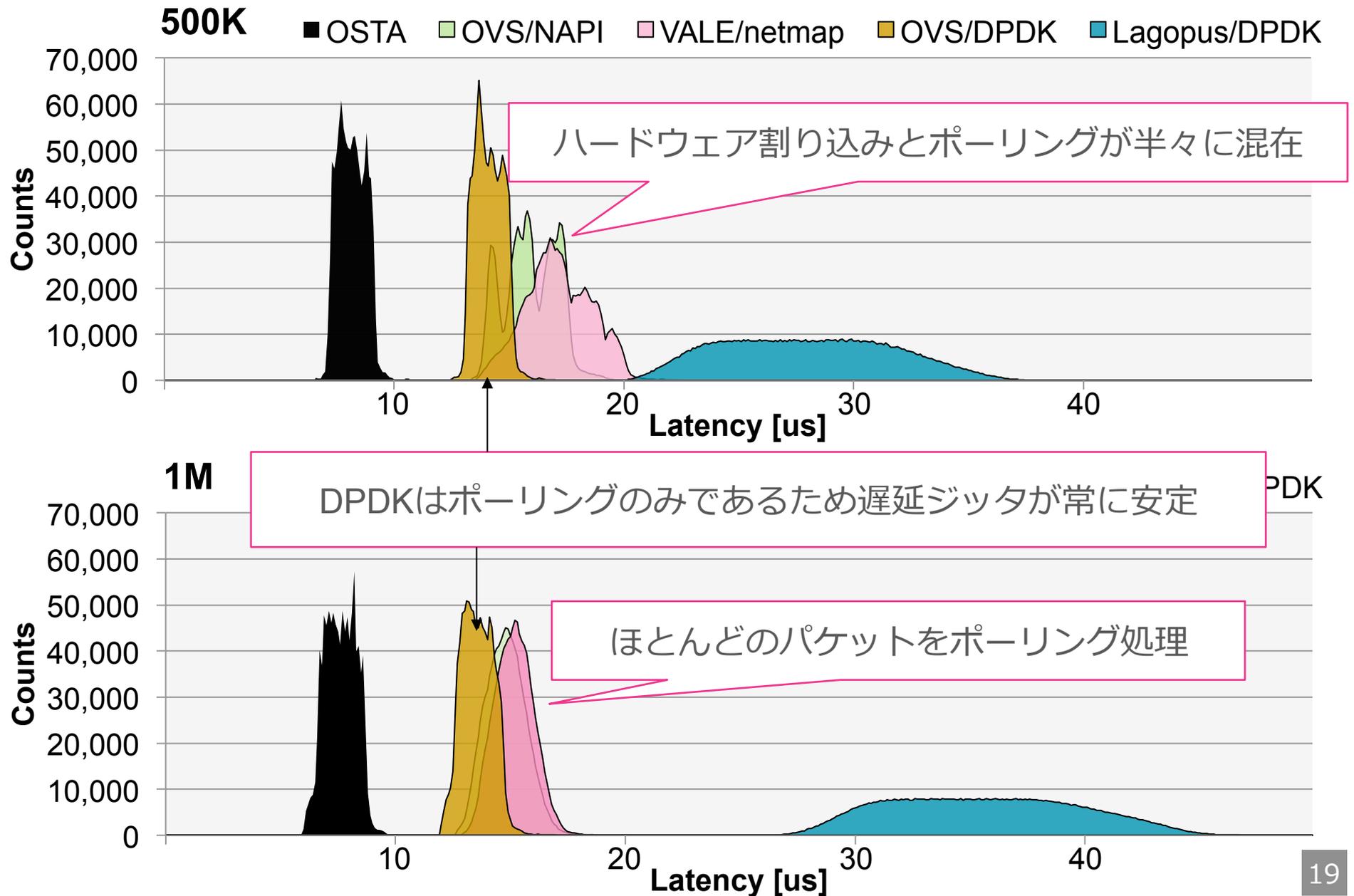
評価(1): パケット転送遅延

◆ 転送レート

✓ 100K pps



評価(1): パケット転送遅延



アウトライン

◆研究目的

- ✓ パケット処理機構と仮想NICバックエンド機構

◆評価内容

評価(1): 物理サーバ上でのパケット転送

評価(2): VMと仮想スイッチ間のパケット転送

評価(3): VM上でのパケット転送

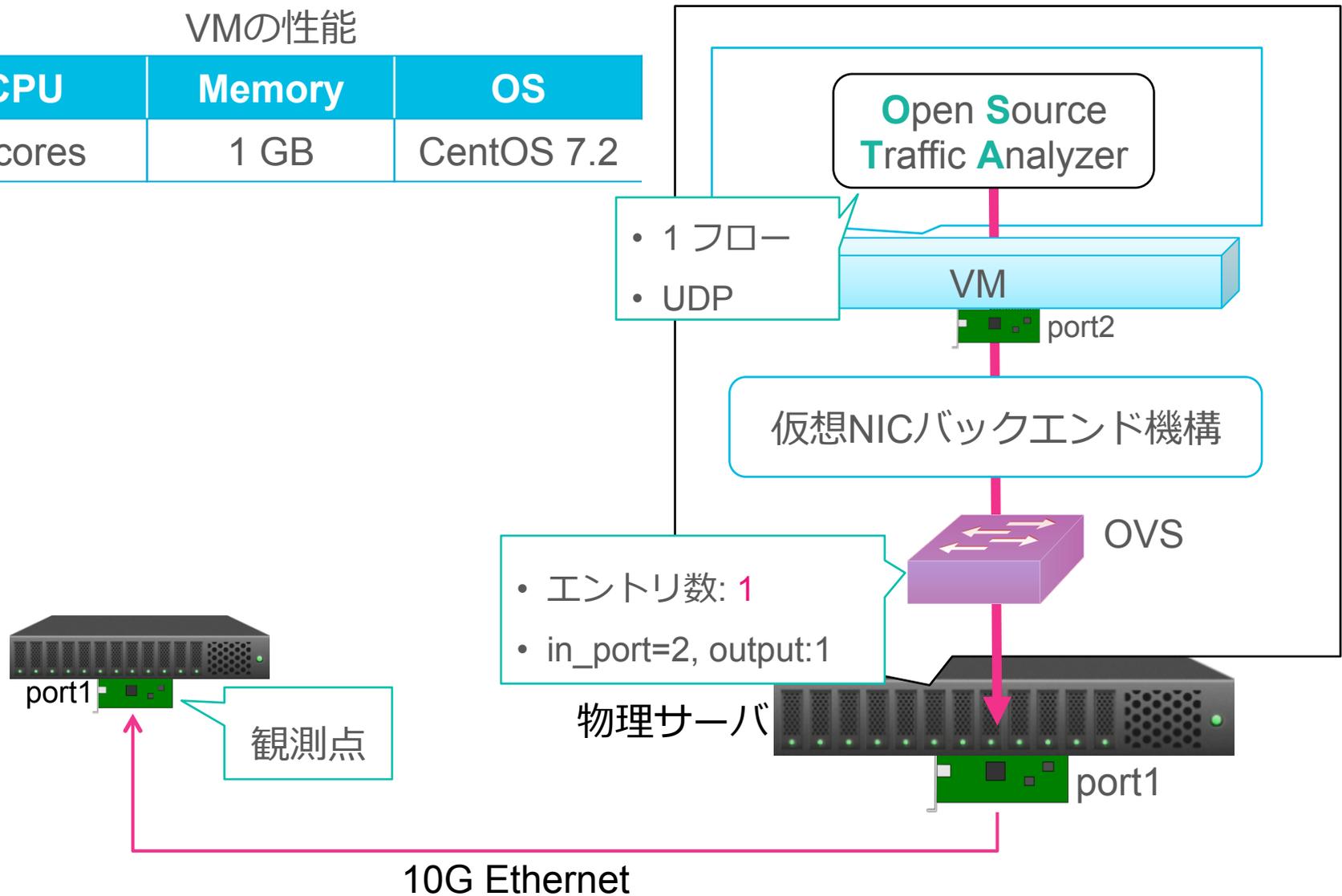
◆NFV環境のサーバ構成の検討

◆まとめと今後の課題

評価(2): 評価環境

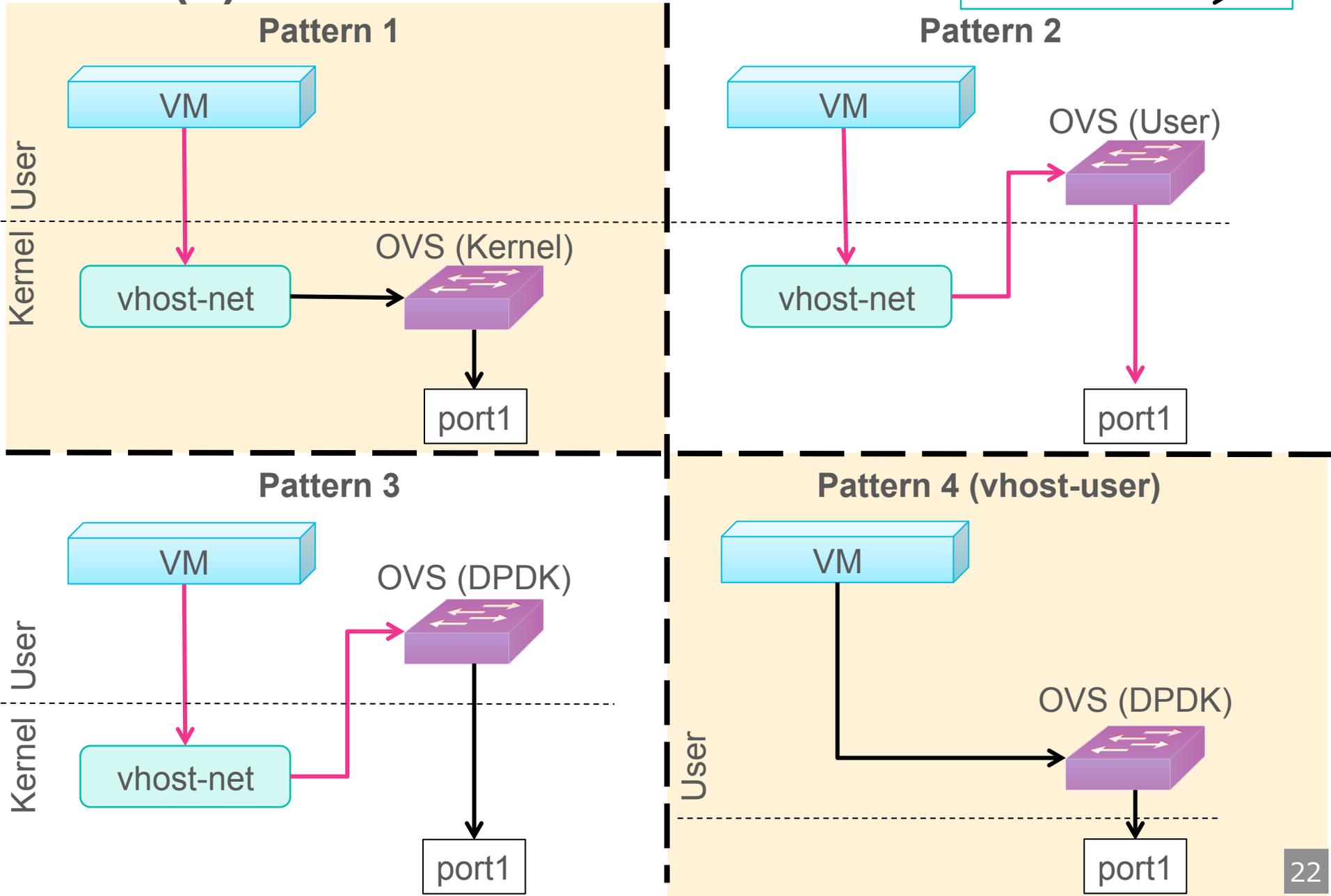
VMの性能

CPU	Memory	OS
2 cores	1 GB	CentOS 7.2

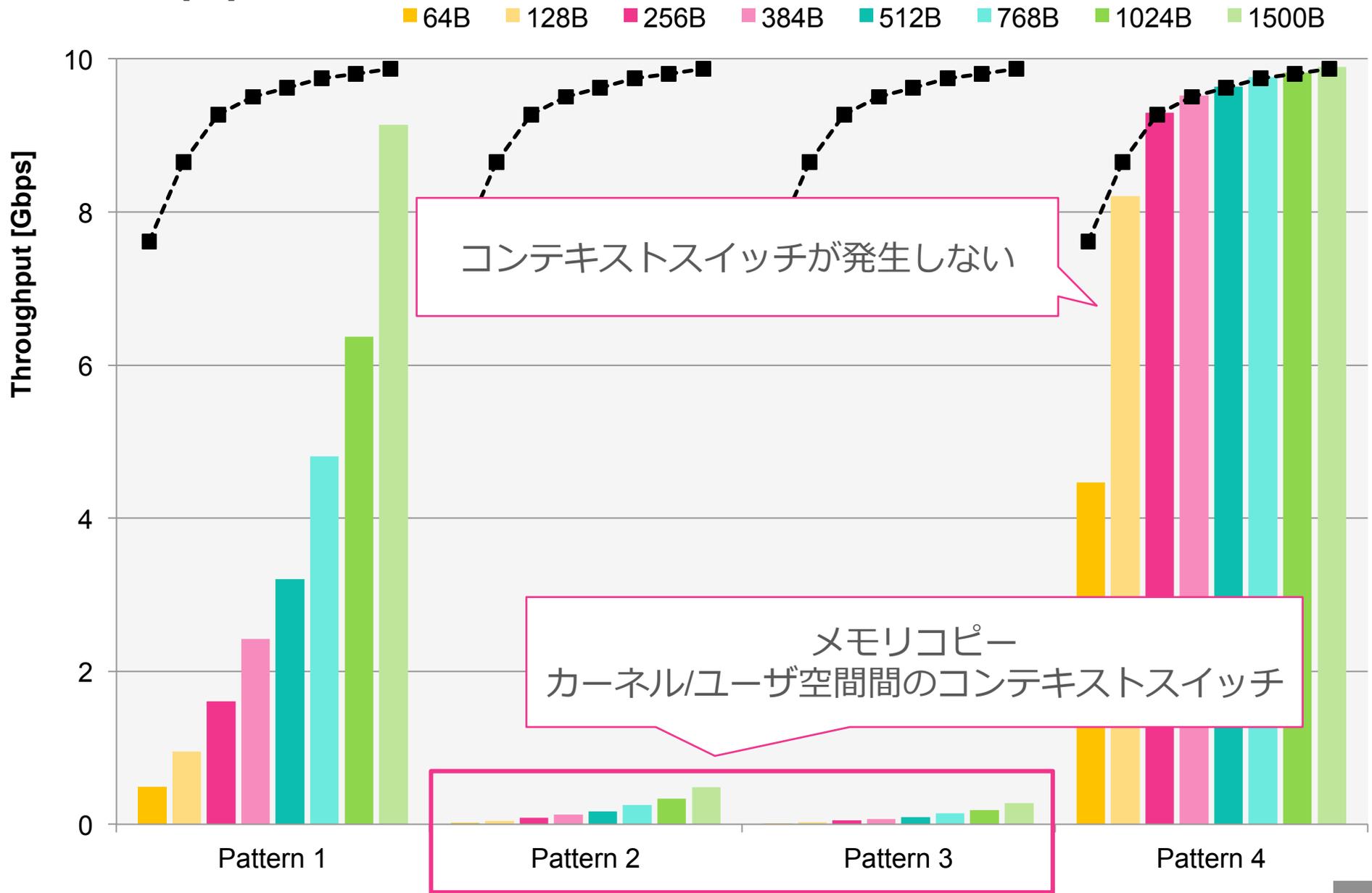


評価(2): 各実装パターンの評価

コンテキストスイッチ
→
コンテキストスイッチ以外
→



評価(2): 評価結果



アウトライン

◆研究目的

- ✓ パケット処理機構と仮想NICバックエンド機構

◆評価内容

評価(1): 物理サーバ上でのパケット転送

評価(2): VMと仮想スイッチ間のパケット転送

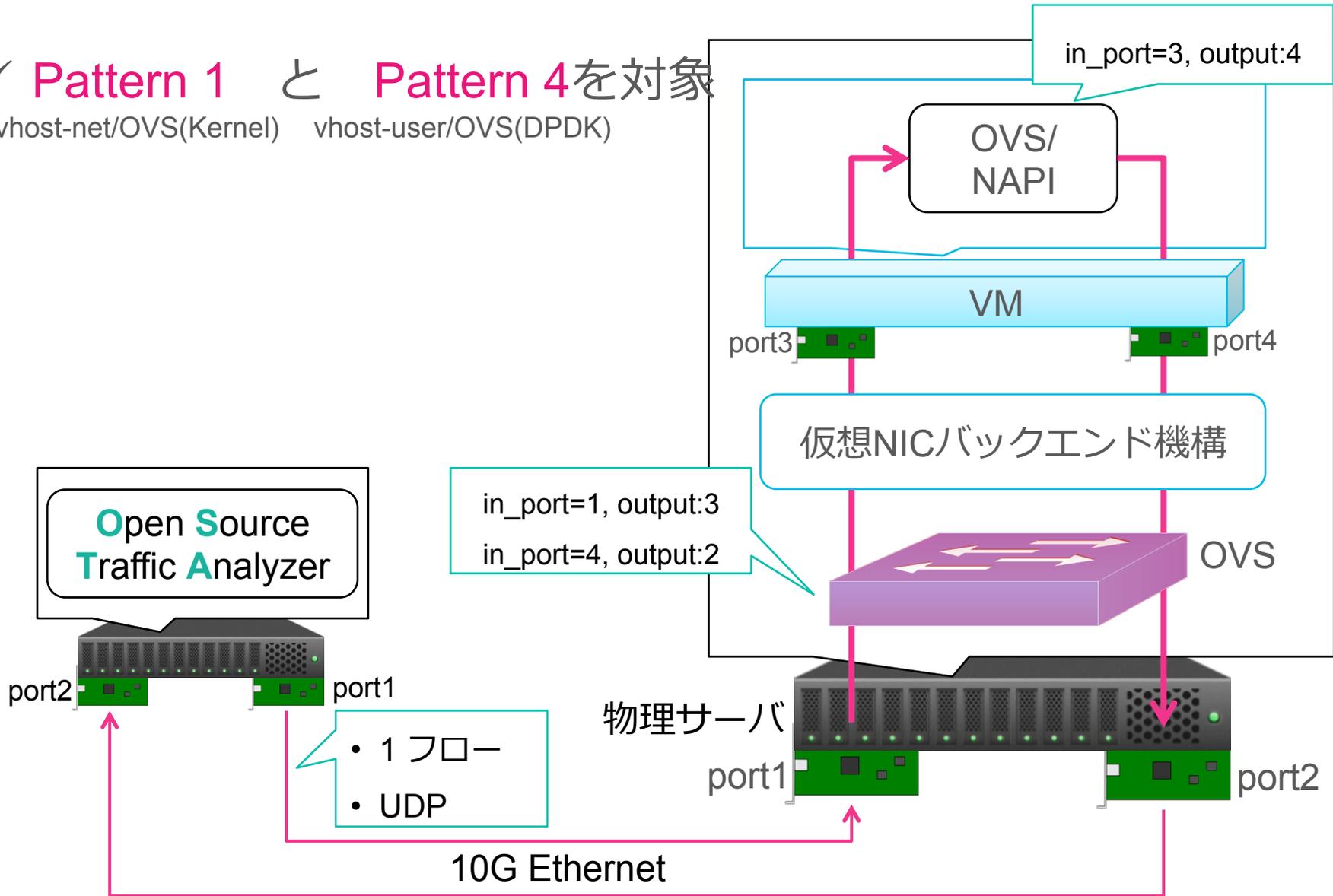
評価(3): VM上でのパケット転送

◆NFV環境のサーバ構成の検討

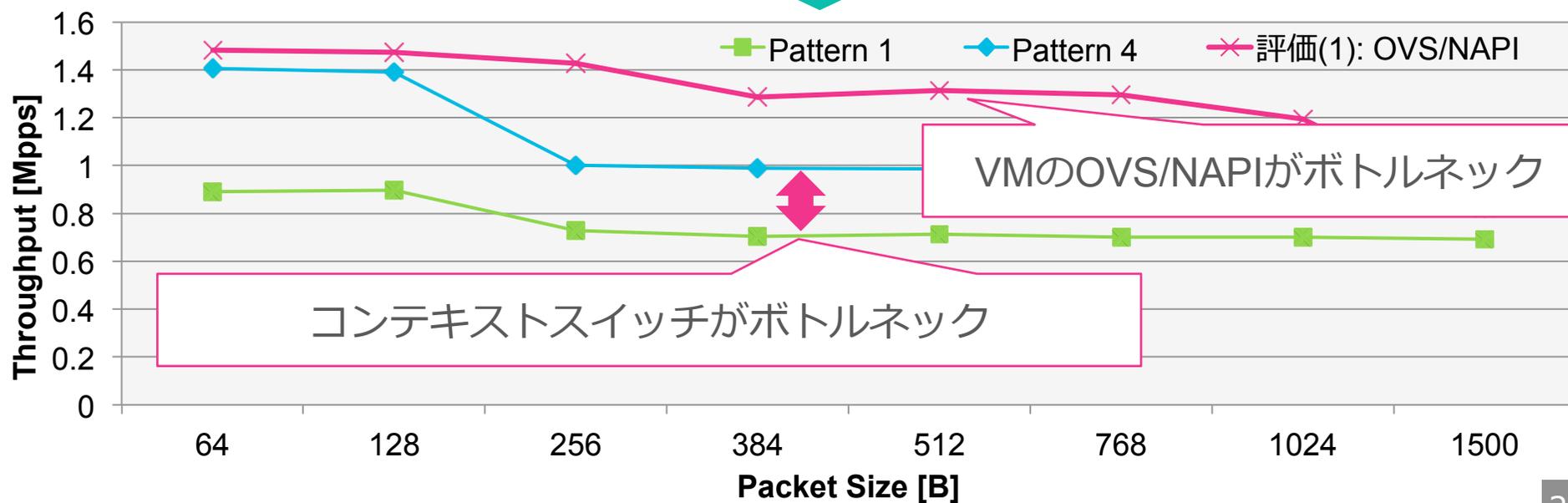
◆まとめと今後の課題

評価(3): 評価環境

✓ **Pattern 1** と **Pattern 4**を対象
vhost-net/OVS(Kernel) vhost-user/OVS(DPDK)



評価(3): パケット転送速度



アウトライン

◆研究目的

- ✓ パケット処理機構と仮想NICバックエンド機構

◆評価内容

評価(1): 物理サーバ上でのパケット転送

評価(2): VMと仮想スイッチ間のパケット転送

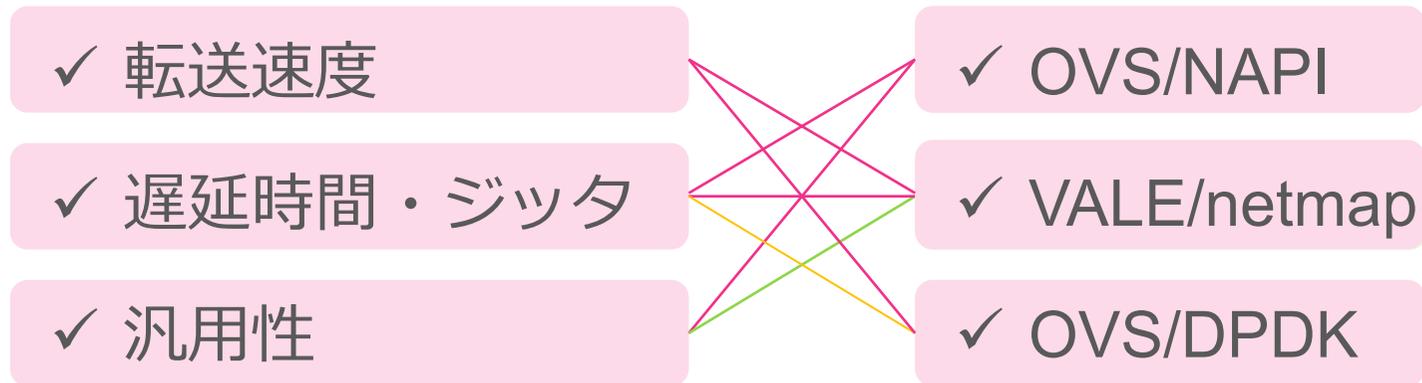
評価(3): VM上でのパケット転送

◆NFV環境のサーバ構成の検討

◆まとめと今後の課題

NFV環境のサーバ構成の検討

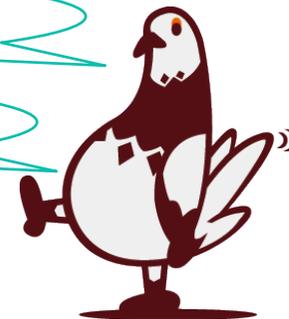
◆評価(1)に基づく検討



状況に応じたパケット処理機構と仮想スイッチを選択

7Mpps per worker thread

遅延ジッタが不安定



まとめと今後の課題

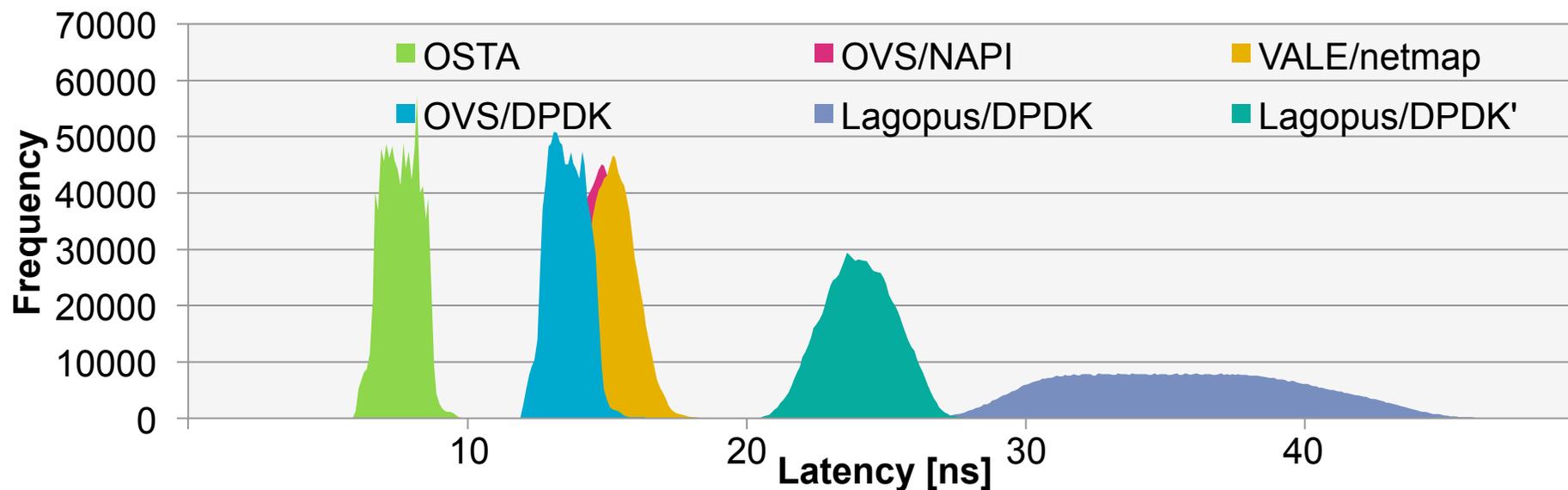
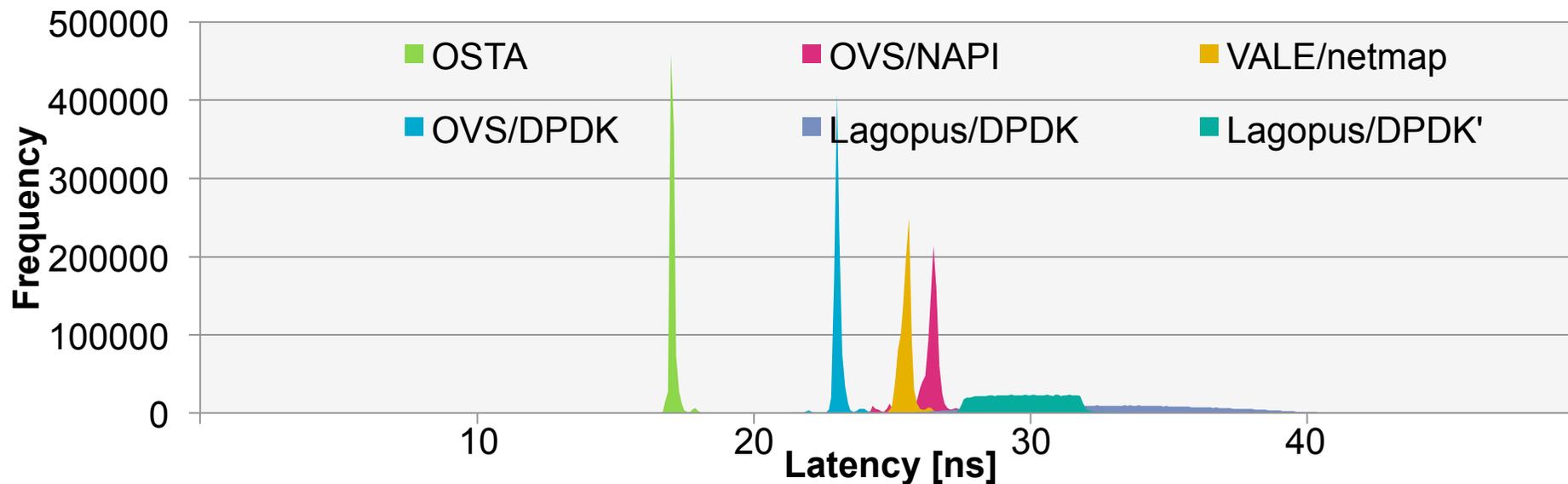
◆まとめ

- ✓ Linux上で様々なパケット処理機構を評価し、性能要因を考察
- ✓ 転送速度を重要視する場合はnetmap, Intel DPDK
 - › New APIではmemcpyやlockなどに処理に大半のサイクル数を消費
- ✓ 遅延ジッタはNew API, netmap, Intel DPDKすべてにおいて抑えられることが確認
 - › 厳密に抑えたい場合はIntel DPDKが適切

◆今後の課題

- ✓ 多数のフローが混在する環境下での評価
- ✓ VM上でDPDKなどのパケット処理機構を用いた際の評価

Lagopus 0.2.4の遅延ジッタ



cycle vs. pps

◆ CPU = Core i7 3770 (3.4GHz, Ivy)
cycles vs. pps (3.4GHz)

