



Същност на Виртуализацията

Виртуализация на ресурси
Виртуализация и Паравиртуализация
Хардуерно ускорена виртуализация



доц. д-р А. Пенев

Съдържание

1. Процесор, памет, диск, мрежа, вход/изход и др.
2. Виртуализация и Паравиртуализация
3. Хардуерно ускорена виртуализация

Виртуализация на Изчисленията

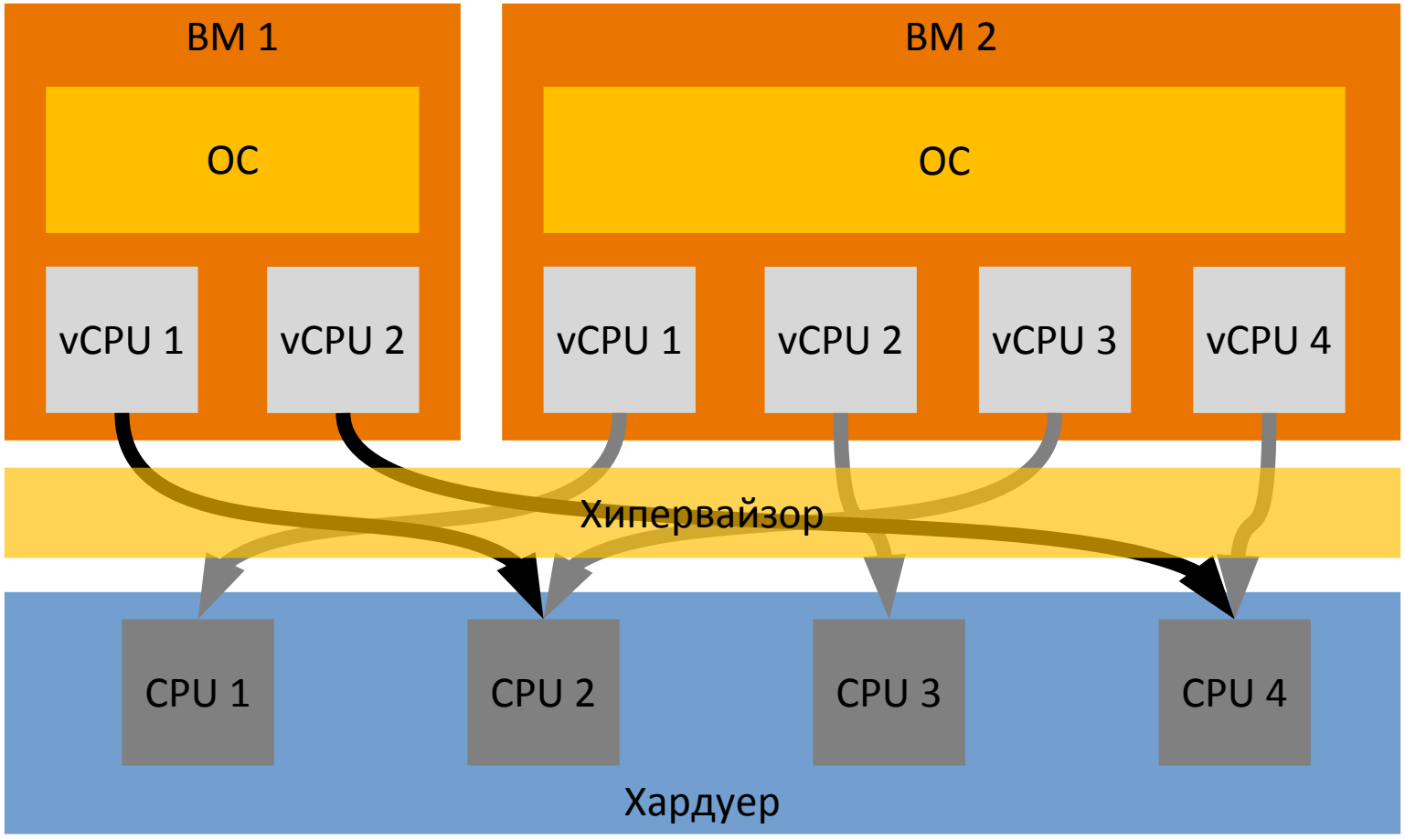
(процесор, GPU)

Време деление и виртуализация на CPU

Предистория

- Първите опити за „виртуализация“ на централния процесор започват през 60-те години на миналия век със създаване на системи работещи с време деление и създаващи илюзията, че отделните потребители работят самостоятелно с машината без да трябва да се изчакват един друг за достъп до нея. Това създава първите многопотребителски ОС, за потребителите на които се създава „виртуалното“ усещане, че машината не е една;
- С появата на многопроцесорните системи възниква необходимост от по-добро и ефективно разпределение на изчислителния ресурс;
- В архитектурата на системите започват да се добавят средства за по-лесна и по-ефективна виртуализация;

Разпределяне на CPU ресурсите. CPU и vCPU



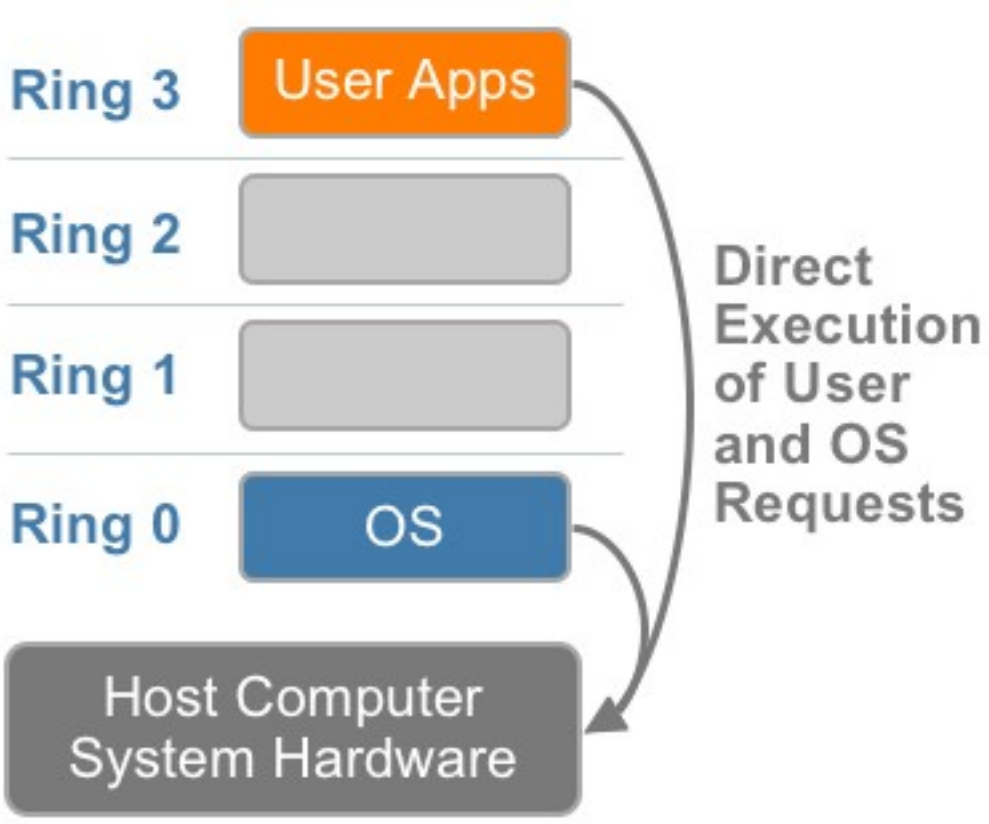
Подходи за CPU виртуализация

- **Емулация** – инструкциите на виртуалните процесори се интерпретират софтуерно;
- **Паравиртуализация** – виртуалните машини съдържат модифицирана ОС, която знае за хипервайзора и комуникира директно с него. Това е един от най-ефективните варианти;
- **Частична виртуализация** – обикновено включва виртуализация на адресното пространство, виртуалната машина симулира множество инстанции на голяма част от хардуерната среда (особено адресни пространства). Обикновено това означава, че цели операционни системи не могат да работят във виртуалната машина;
- **Пълна виртуализация** – виртуалната машина е напълно изолирана от хардуера. Вижда само симулирани от хипервайзора устройства;

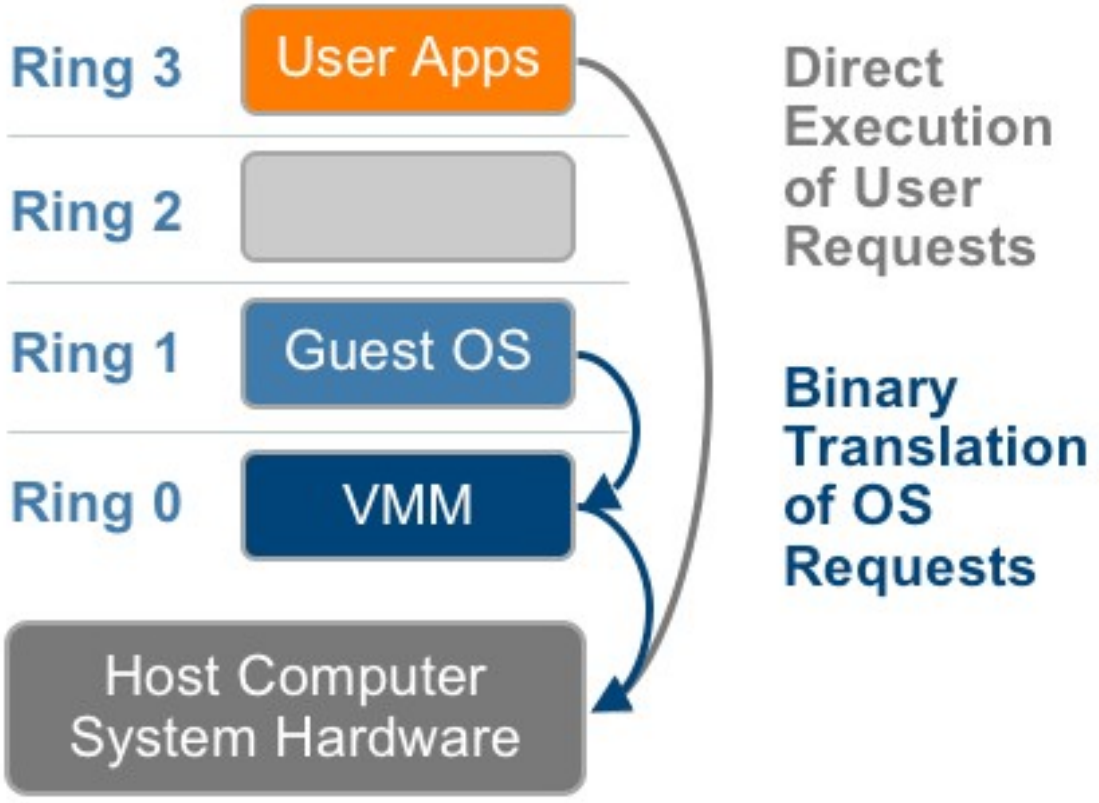
Подходи за CPU виртуализация

- **Бинарна транслация** – хипервайзора променя части от guest ОС, като заменя привилегированите инструкции с код извикващ функции на хипервайзора (VMM). Това не изисква предварително модифициране на цялата ОС;
- **Хардуерно подпомогната виртуализация** – опростява виртуализацията, тъй като пълната или паравиртуализацията са сложни. Intel и AMD добавят допълнителен режим, наречен ниво на привилегирован режим (Ring -1) към процесори x86. Следователно операционните системи все още могат да работят на Ring 0, а хипервайзорът може да работи на Ring „-1“. Всички привилегировани и чувствителни инструкции се улавят автоматично в хипервайзора;

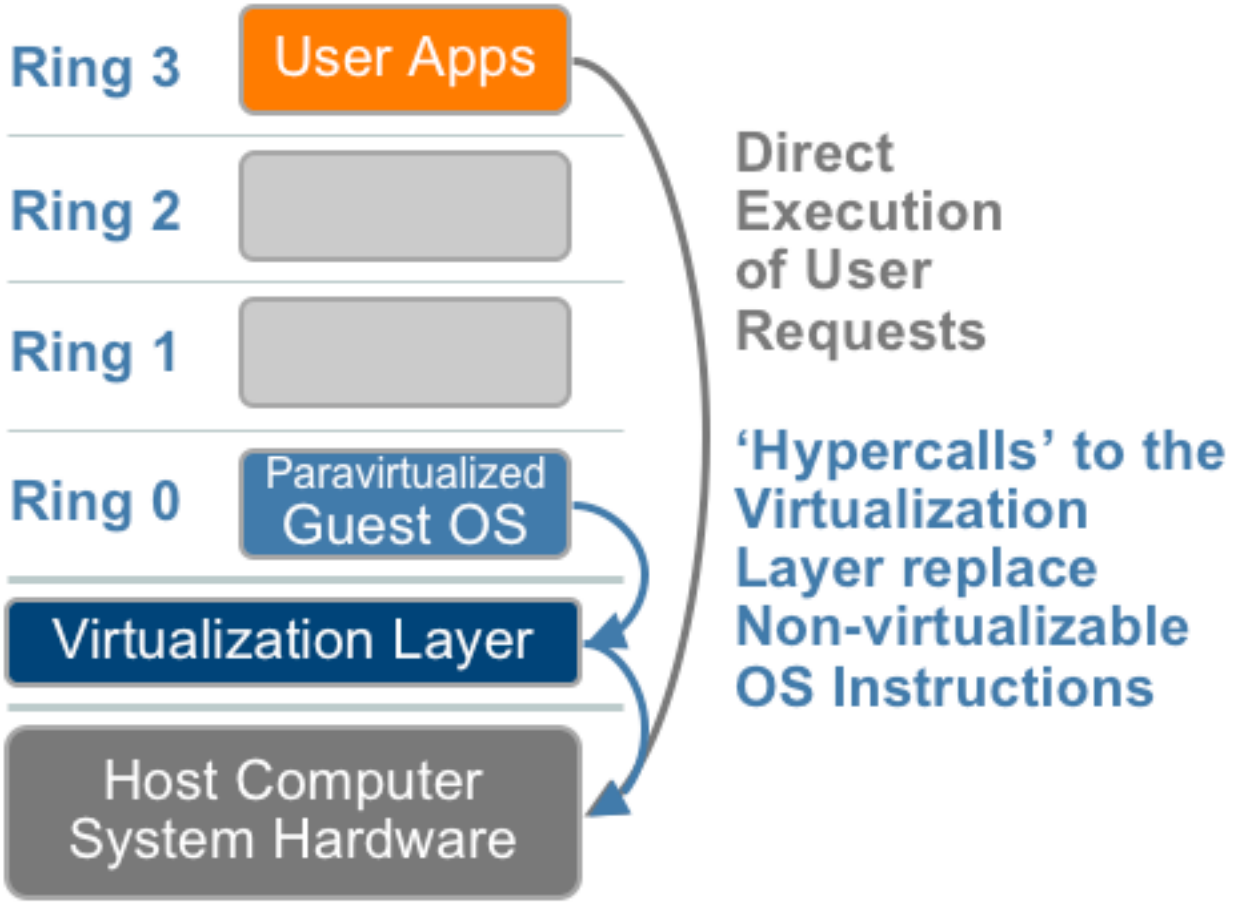
Нормално изпълнение на ОС



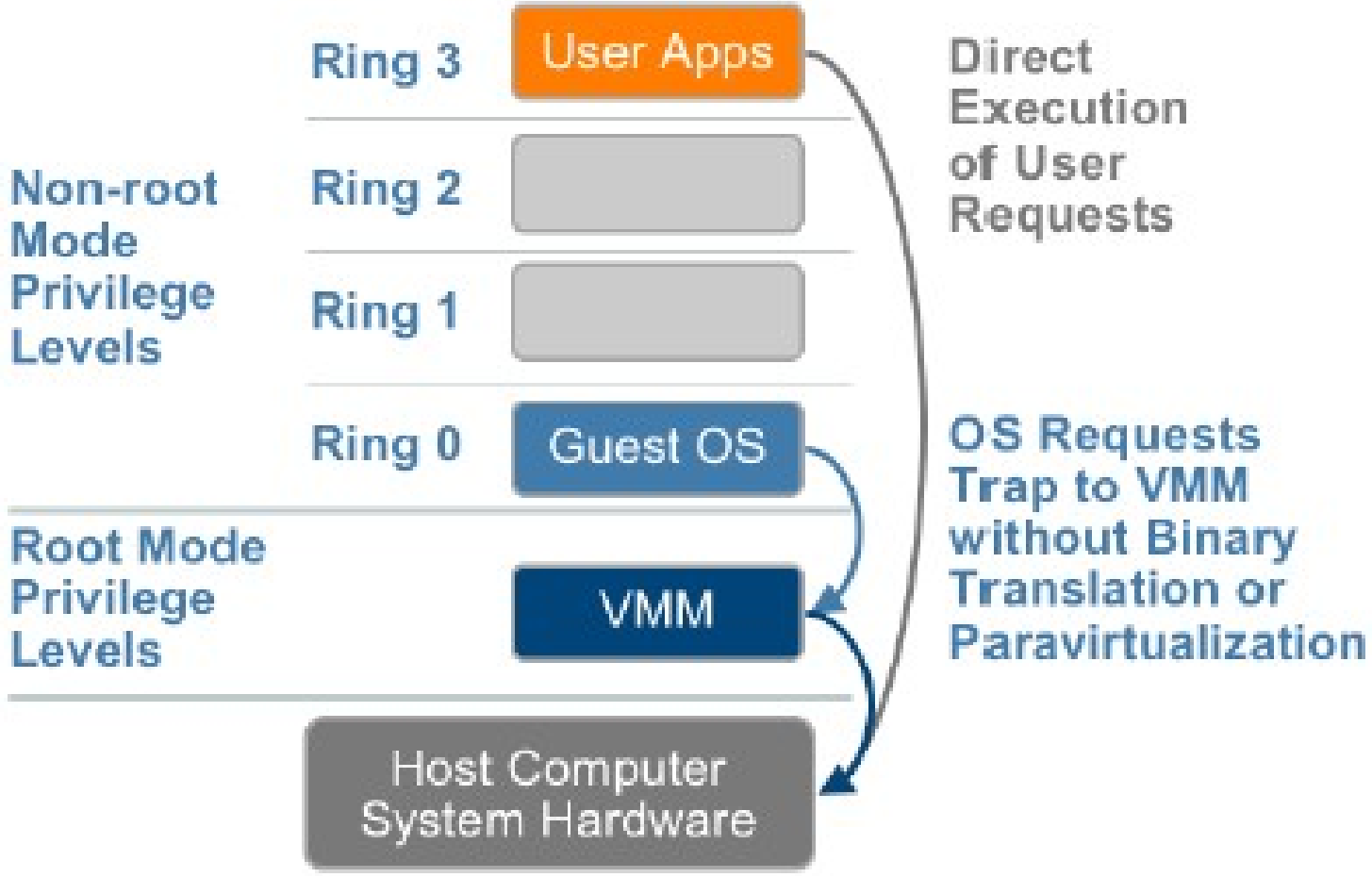
Изпълнение на ОС на Ring 1 след Бинарна транслация



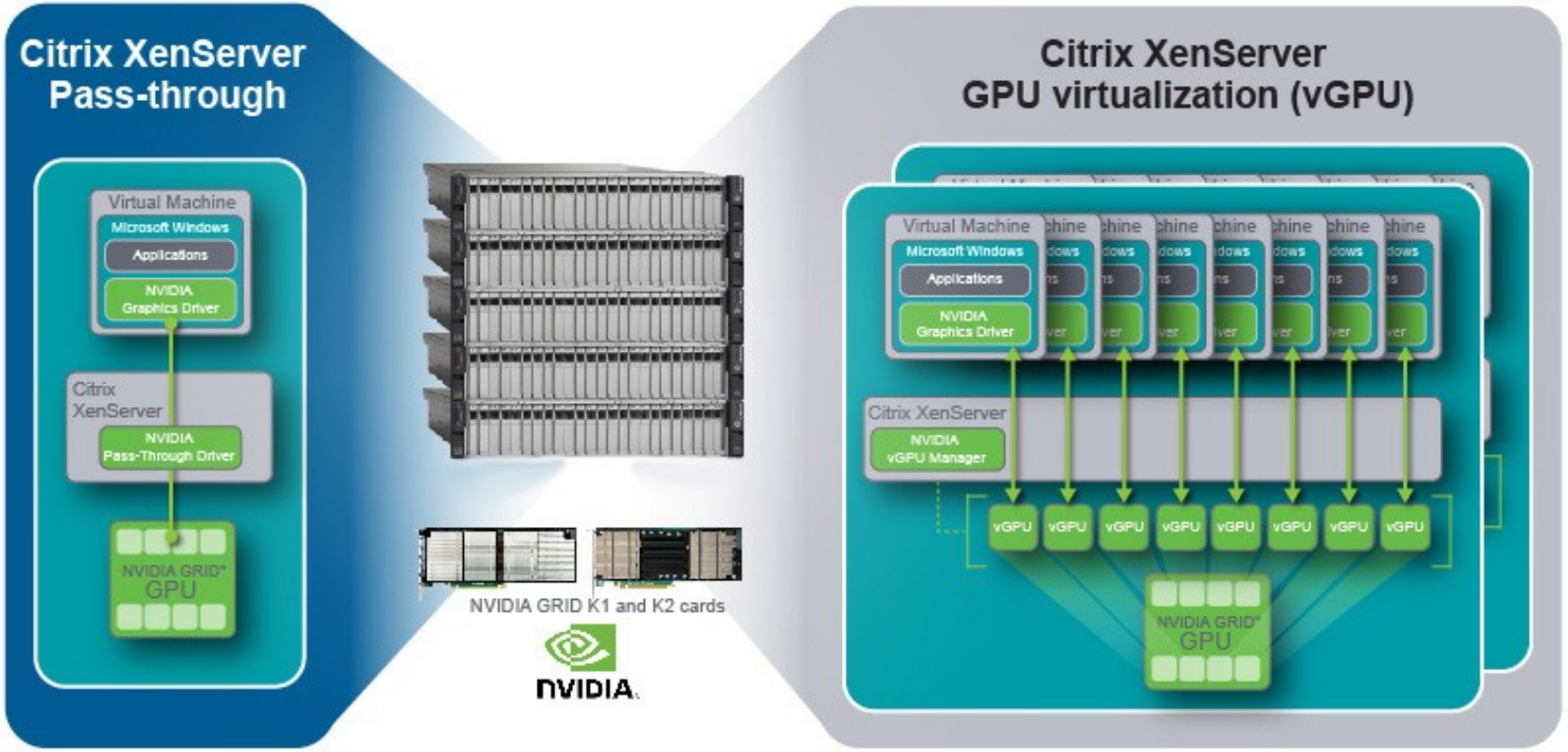
Изпълнение на ОС на Ring 0 при Паравиртуализация



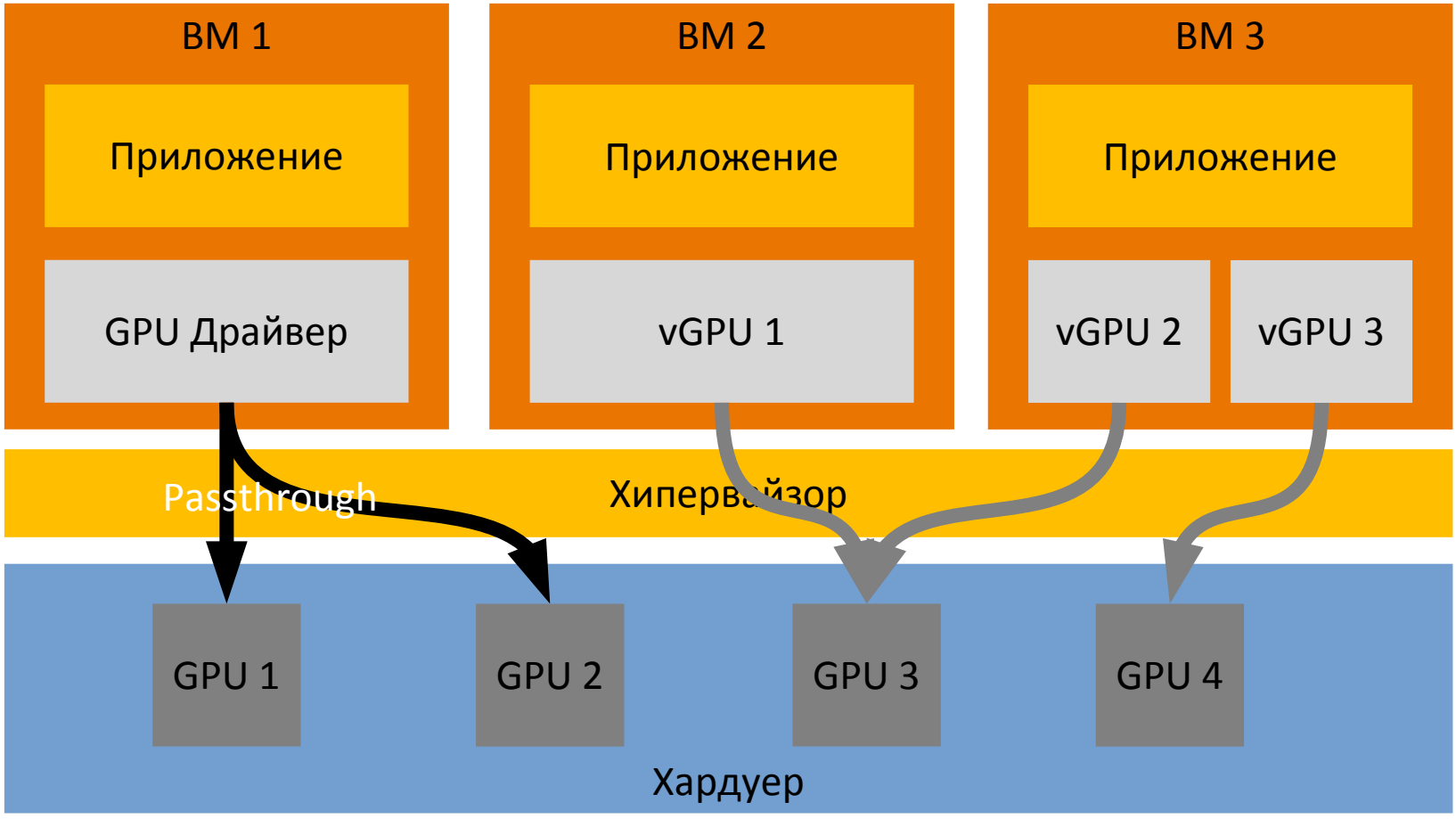
Изпълнение на ОС на при хардуерно ускорена вирт.



GPU и vGPU


Citrix XenServer
Pass-through
Virtual Machine
Microsoft Windows
Applications
NVIDIA
Graphics Driver
Citrix
XenServer
NVIDIA
Pass-Through Driver
NVIDIA GRID
GPU
NVIDIA GRID K1 and K2 cards
NVIDIA
Citrix XenServer
GPU virtualization (vGPU)
Virtual Machine
Microsoft Windows
Applications
NVIDIA
Graphics Driver
Citrix XenServer
NVIDIA
vGPU Manager
vGPU
NVIDIA GRID
GPU

GPU и vGPU


ВМ 1
Приложение
GPU Драйвер
ВМ 2
Приложение
vGPU 1
ВМ 3
Приложение
vGPU 2
vGPU 3
Passthrough
Хипервайзор
GPU 1
GPU 2
GPU 3
GPU 4
Хардуер

Виртуализация на Паметта

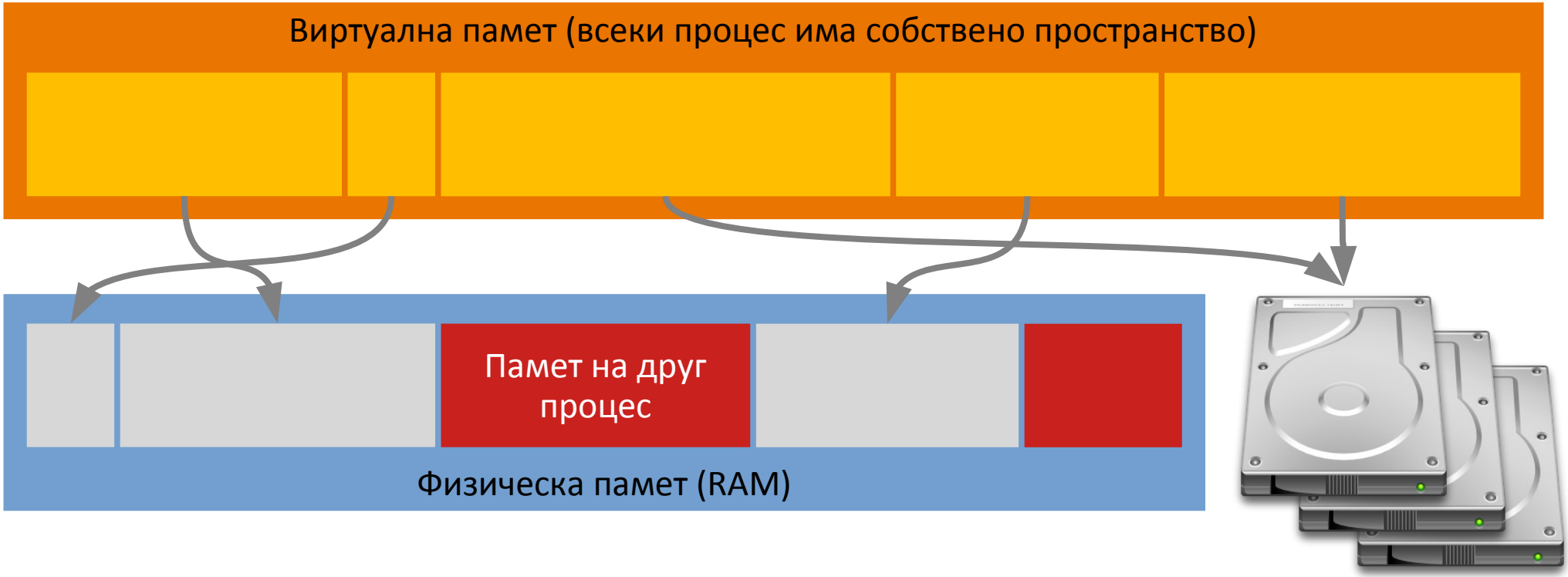
(памет)

Виртуална памет

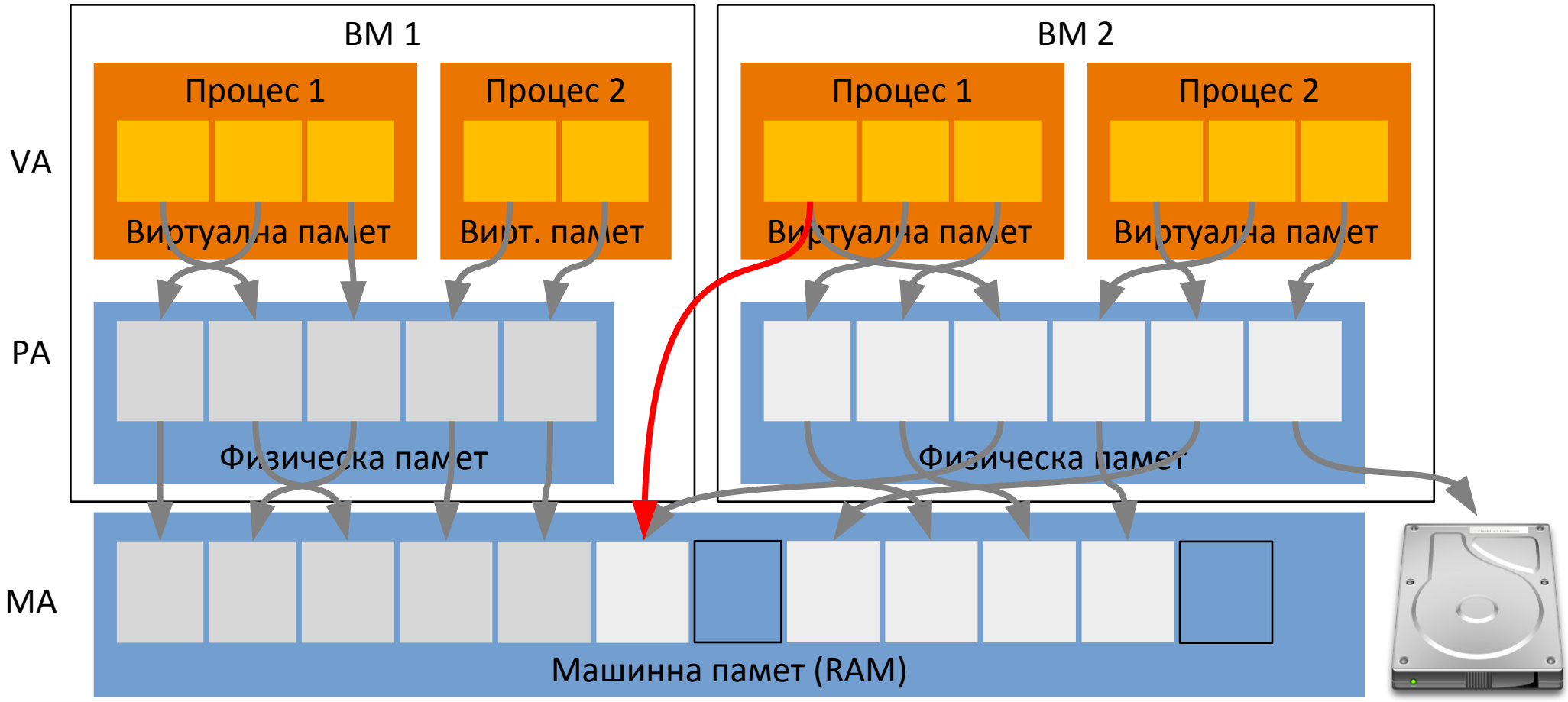
Предистория

- Първите опити за „виртуализация“ на централния процесор започват през 60-те години на миналия век. Наред с това още от 40-те и 50-те години на миналия век се работи за виртуализация и на паметта;
- Компютърните системи изискват все повече оперативна памет. Част от външната памет се използва за разширяване на физическата памет. Виртуализацията на паметта структурира адресното пространство (виртуалното адресно пространство) по различен от класическия начин (последователен). Последователни адреси от виртуалната памет могат да не са последователни във физическата и дори в определен момент да не бъдат в нея;
- Много важна е и изолацията на паметта на отделните процеси/потребители;
- В архитектурата на системите започват да се добавят средства за по-лесна и по-ефективна виртуализация на паметта;

Класическа Виртуална памет



Виртуализация на паметта



Виртуализация на паметта

- Споделяне на физическата системна памет и динамичното ѝ разпределяне за ВМ. Виртуализацията на паметта за ВМ е много подобна на поддръжката на виртуална памет, осигурена от съвременните ОС;
- Всички съвременни процесори x86 включват блок за управление на паметта (MMU) и буфер за превод на търсенето (TLB) за оптимизиране на производителността на виртуалната памет;
- За да стартирате няколко виртуални машини в една система, е необходимо друго ниво на виртуализация на паметта т.е. трябва да се виртуализира MMU;
- Guest ОС продължава да контролира виртуалните си адреси и им съпоставя физически адреси, но няма директен достъп до паметта на Host машината. Хипервайзорът е отговорен за съпоставянето на Guest физическата памет към реалната памет на Host машината (използват се „Shadow Page Tables“ задаващи VA→MA и за избягване на транслацията на две нива);

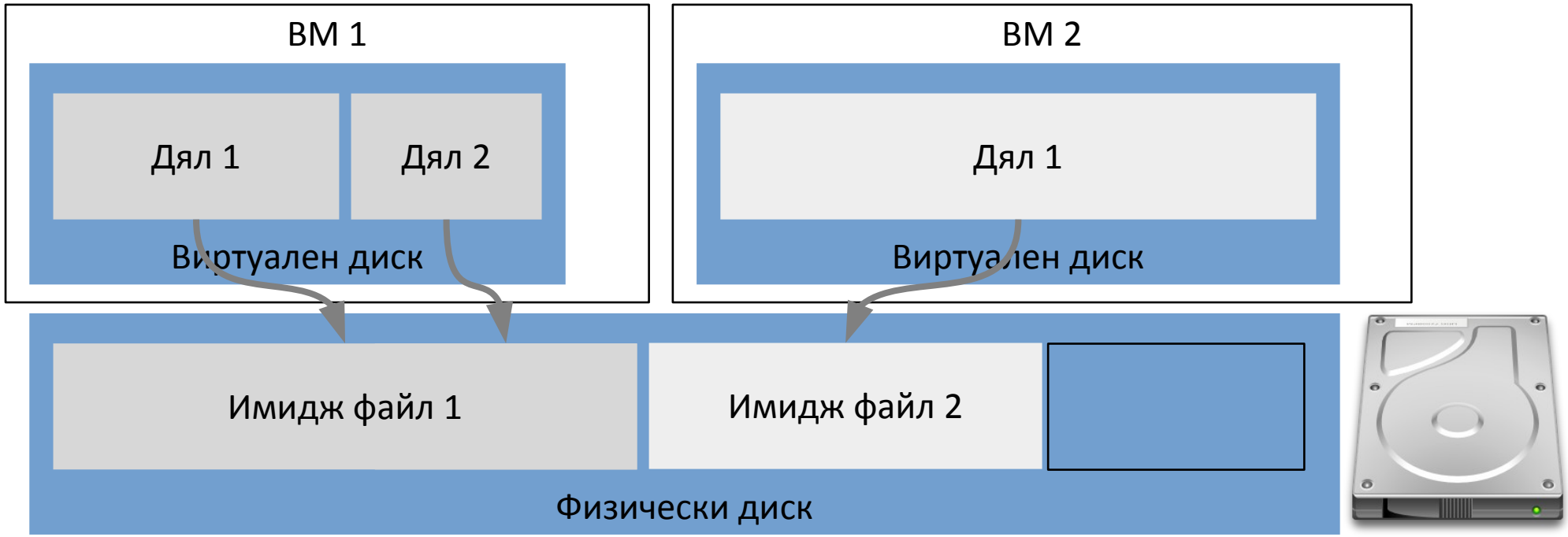
Виртуализация на паметта (техники за ефективност)

- **Shadow Page Table (SPT)** – за избягване на двойното индексиране;
- **Extended Page Table (EPT)** – за разлика от SPT, която е „софтуерна“ техника, EPT е хардуерно подпомогната техника, която може да ускори изпълнението на ВМ от 50% до 500-600%;
- **Hot-Swap** – ползване на диска за увеличаване на размера на виртуалната памет;
- **Kernel Samepage Merging (KSM)** – използване на една и съща страница за различни ВМ, когато тя има едно и също съдържание;
- **Copy on Write (CoW)** – съвместените страници се копират и разделят, ако някой от се опита да запише в тях (ако не са read only);
- **Balooning** – Драйвер работещ в Guest ОС, за да заделя динамично паметта необходима на ОС и при възможност да я върне обратно на хипервайзора за използване от други ВМ;

Виртуализация на Диска/Външната памет

(блокови устройства, файлова система)

Виртуализация на външната памет



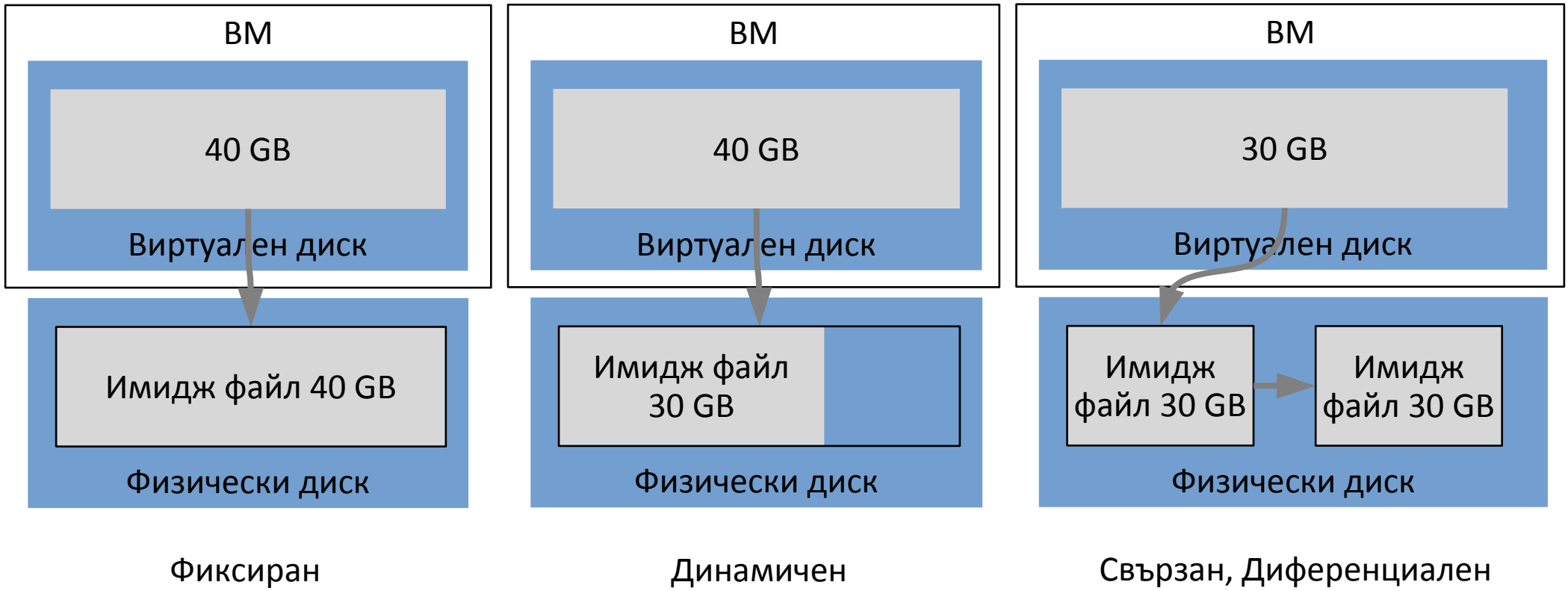
Виртуализация на външната памет

- Обикновено целия диск/външна памет се съхранява във Имидж (Image) файл/файлове;
- Имидж файла на ВМ (VM Image file) съдържа цялото съдържание на диска/външната памет на ВМ;
- Има много файлови формати, но всички те се разделят основно на два вида:
  - **Огледален режим (Mirror)** – Съхраняват всички даннн в диска, включително тези които не се използват т.е. те са пълно копие “сектор-по-сектор”, “байт-по-байт” на диска (RAW формат);
  - **Разреден режим (Sparse)** – Съхраняват всички валидни за потребителя или за системата данни от диска т.е. само наистина необходимите данни (QCOW2, VMDK и др са файлове в такъв формат);

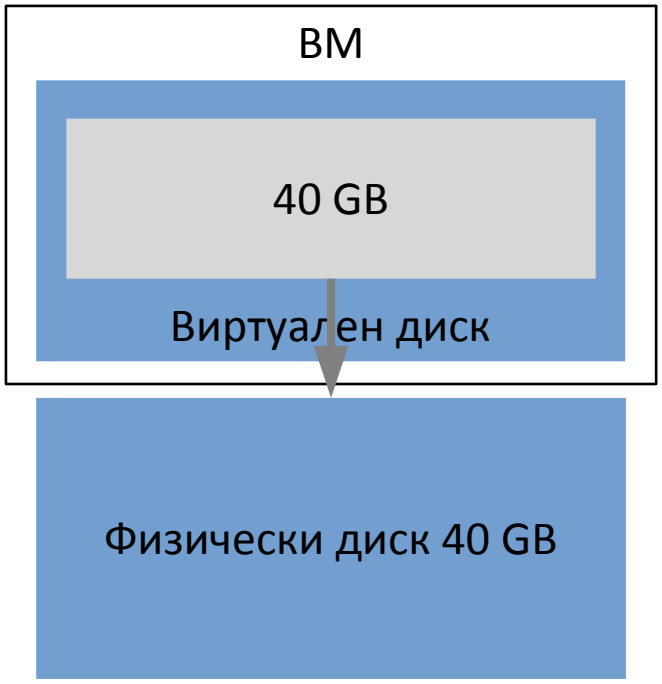
Специфични възможности поддържани от някои имидж файлове

- Дискове със файлови системи, които поддържат „sparse files“ имидж файловите формати могат да съхраняват само действително използваните части от диска;
- Създаване на много Snapshots (съхраняване на състоянието на диска в даден момент), както и бързото му възстановяване към някой от тези моменти;
- Малки имидж файлове дори, когато файловата система не поддържа “sparse files“;
- Диференциални дискове;
- Опционално AES кроптиране;
- Опционална компресия;
- и др.;

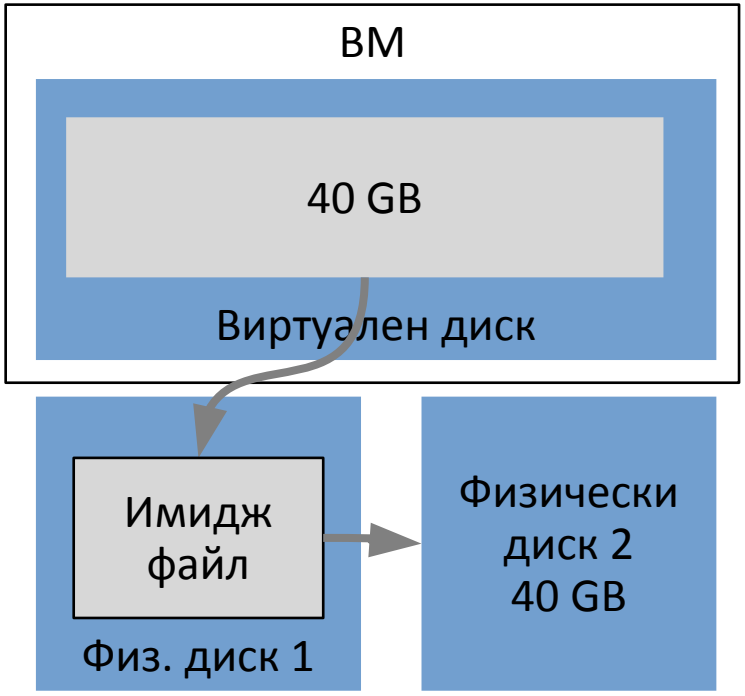
Видове виртуални дискове



Видове виртуални дискове

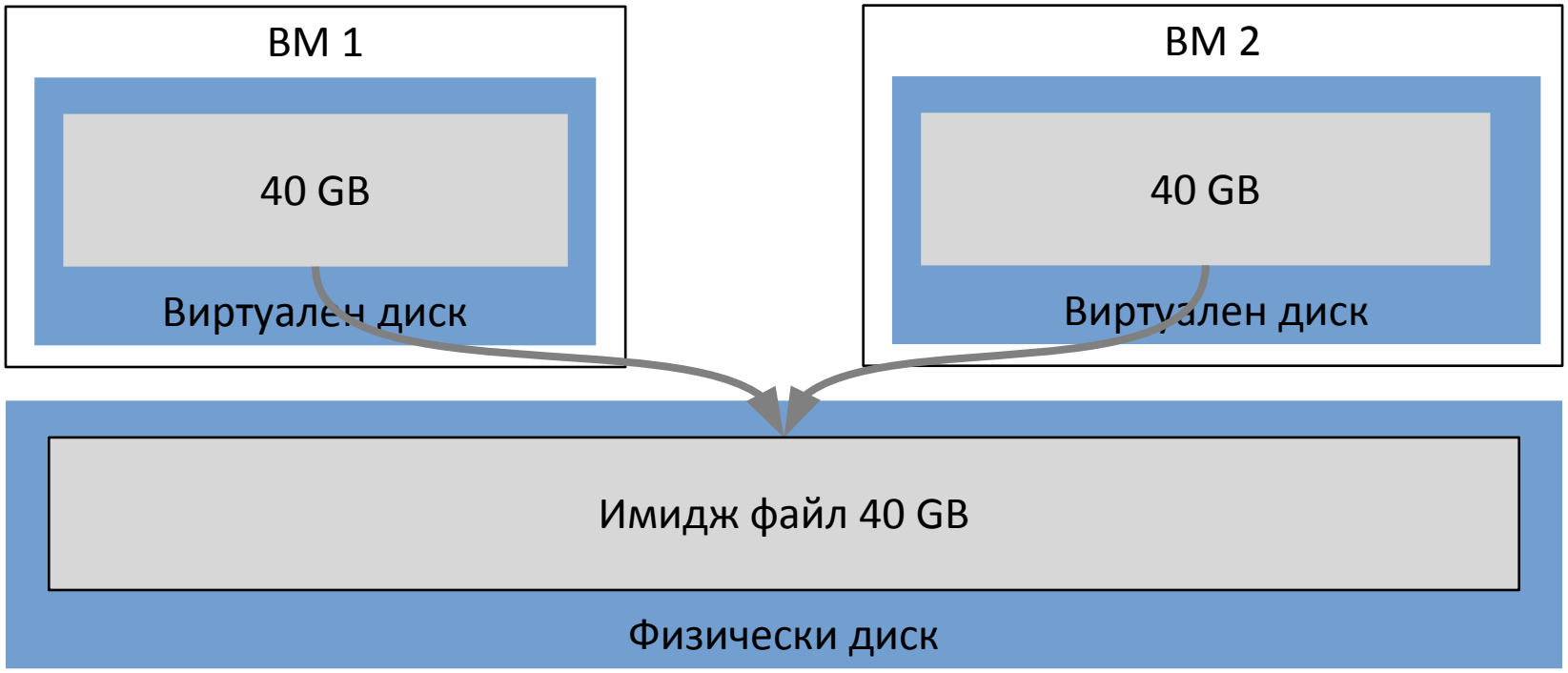


Физически



Свързан диференциален, базиран на физически

Видове виртуални дискове



Споделен

Други

- Имиджи на CD (.iso) – за инсталационни виртуални дискове;
- Имиджи на Флопи дискове;
- Достъп (монтиране) до локални части от файловата система на Host-а – основно чрез виртуален Sharing през (вирт.) мрежата;
- iSCSI – оталечени блокови устройства през TCP/IP;
- и много други;

Виртуализация на Мрежата

(виртуални мрежи)

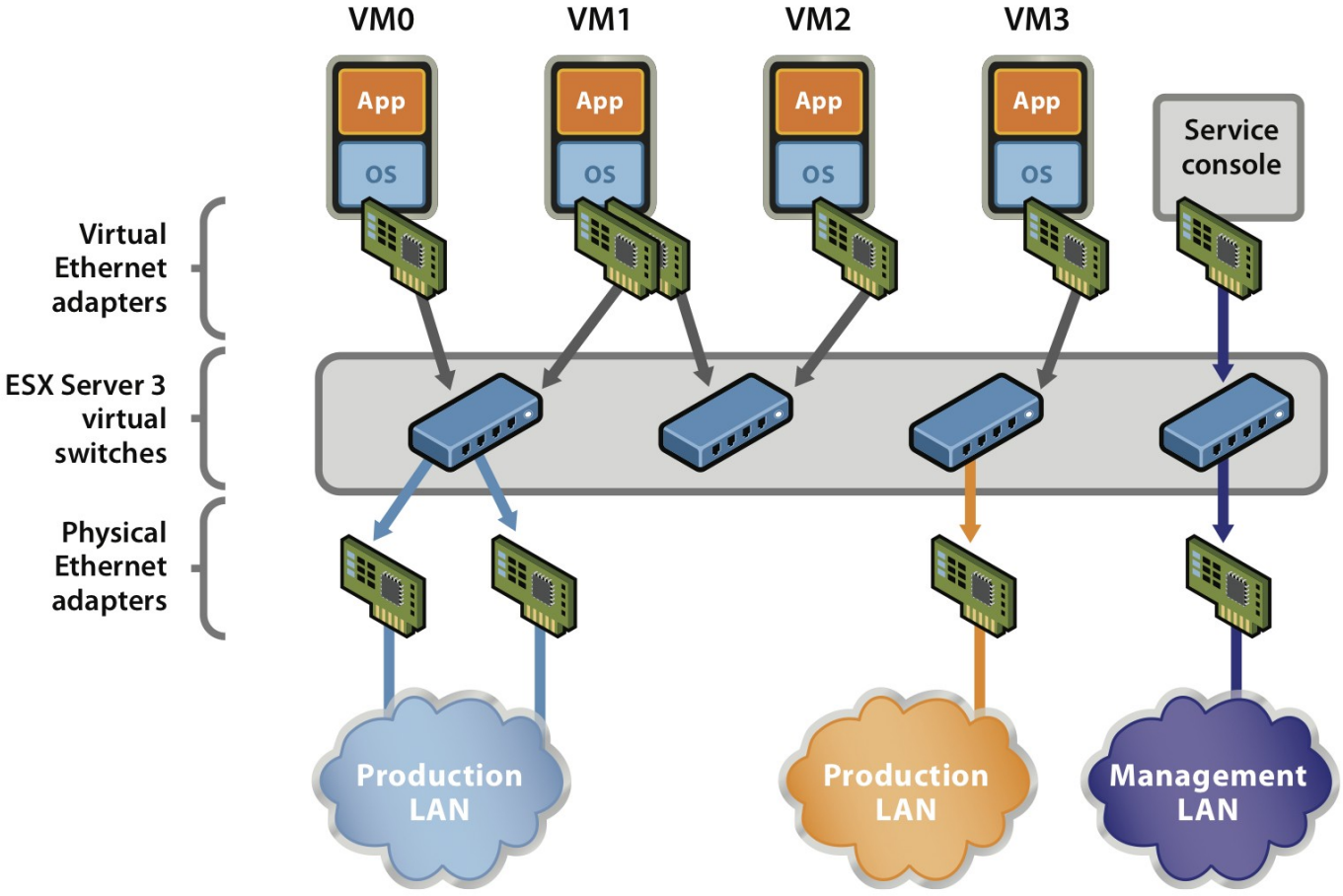
Виртуализация на мрежата

- Виртуалната компютърна мрежа е съвкупност от софтуерни и хардуерни мрежови ресурси, работещи като едно цяло;
- Виртуализацията на мрежата се дели на два вида:
  - **Вътрешна** – осигуряване на подобна на мрежа функционалност на софтуерни компоненти на един мрежов сървър;
  - **Външна** – комбиниране на много мрежи или части от мрежи във една виртуална единица;

Виртуални компоненти на виртуалната мрежа

- Мрежов хардуер, като комутатори (switch) и мрежови интерфейсни карти (NIC);
- Мрежови елементи, като защитни стени (firewall) и балансери (load balancers);
- Виртуални LAN (VLAN) и виртуални машини (VM);
- Мрежови устройства за съхранение (NAS);
- Мрежови елементи за връзка машина-машина, като телекомуникационни устройства и др.;
- Мрежови връзки, като Ethernet и Fiber Channel;
- и др.;

Виртуализация на мрежата



Виртуализация на мрежата

- Прави се виртуализация на мрежовите карти (NIC). Във всяка виртуална машина може да бъдат добавени една или повече vNIC;
- Те се свързват (по различни начини) директно със физическата мрежа (NIC картите на хоста) или през виртуални комутатори (vSwitch);

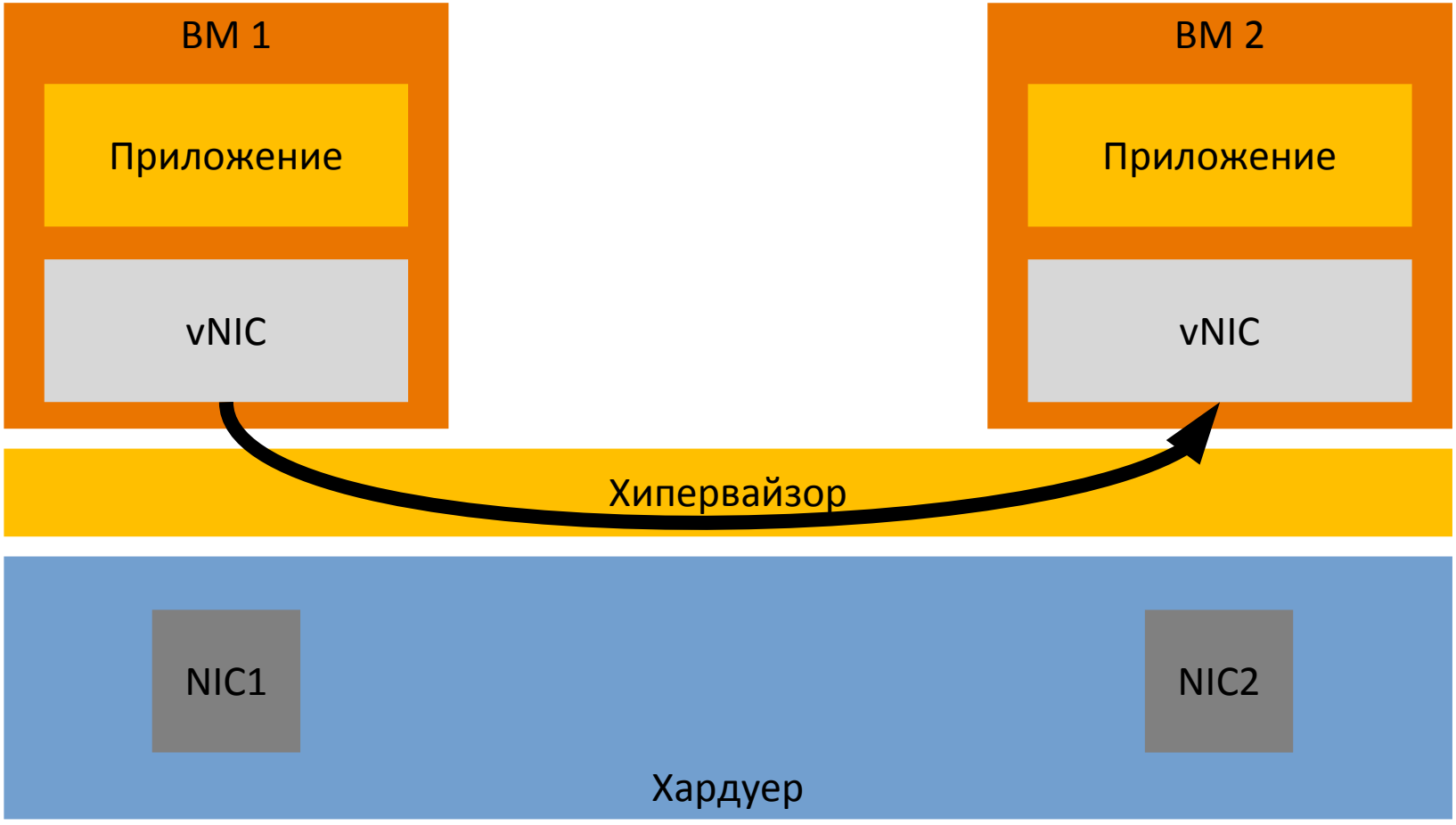
Видове връзки на vNIC с хост машината

- Без връзка;
- NAT;
- NAT мрежа (NAT network);
- Мостов адаптер (Bridge);
- Вътрешна (Internal);
- Само хост (Host-Only);

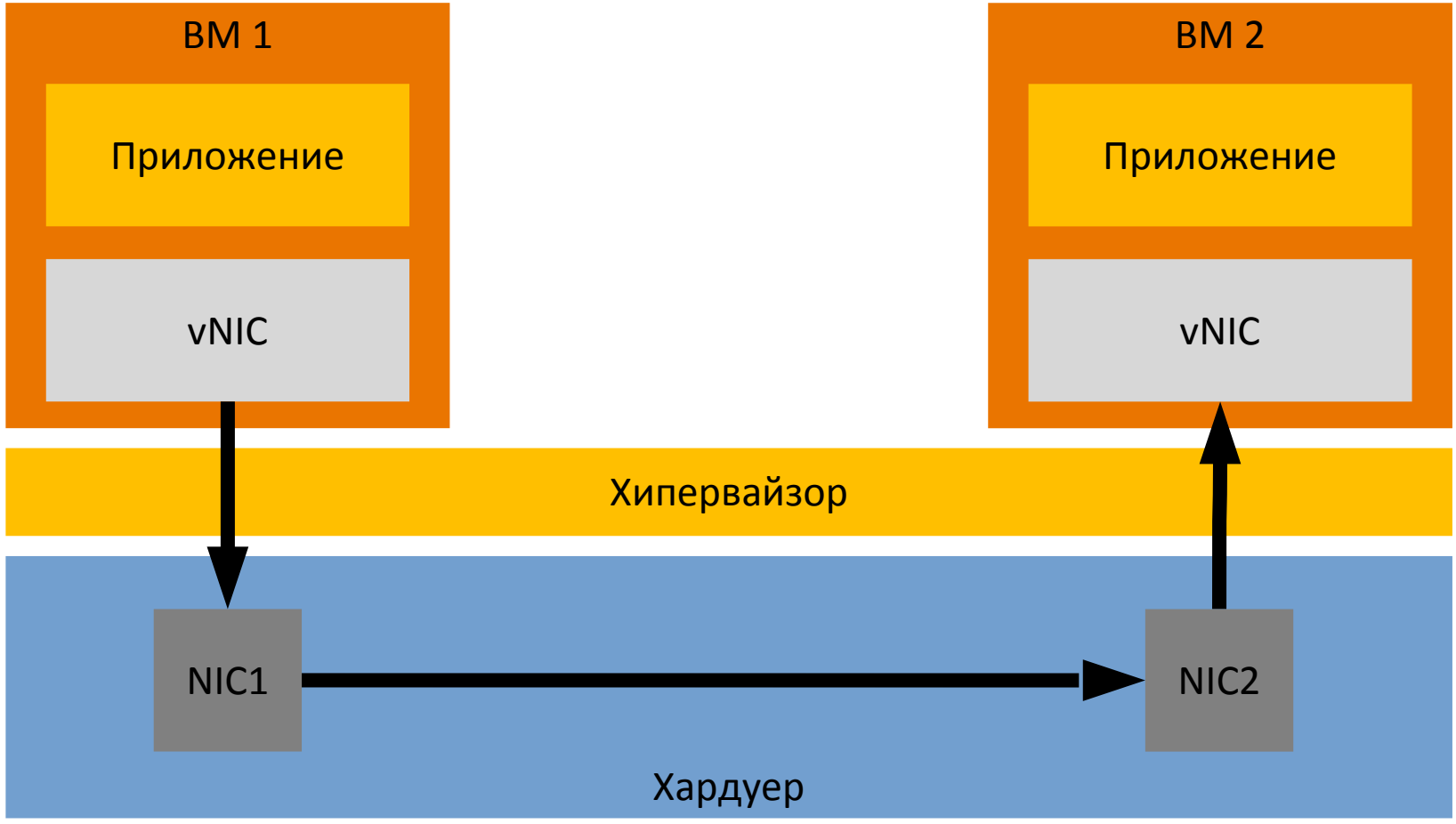
Видове връзки на vNIC с хост машината

Режим	VM → Host	VM ← Host	VM1 ↔ VM2	VM → Net/LAN	VM ← Net/LAN
Disconnect	–	–	–	–	–
Internal	–	–	+	–	–
Host-only	+	+	+	–	–
NAT	+	Port forward	–	+	Port forward
NAT network	+	Port forward	+	+	Port forward
Bridged	+	+	+	+	+

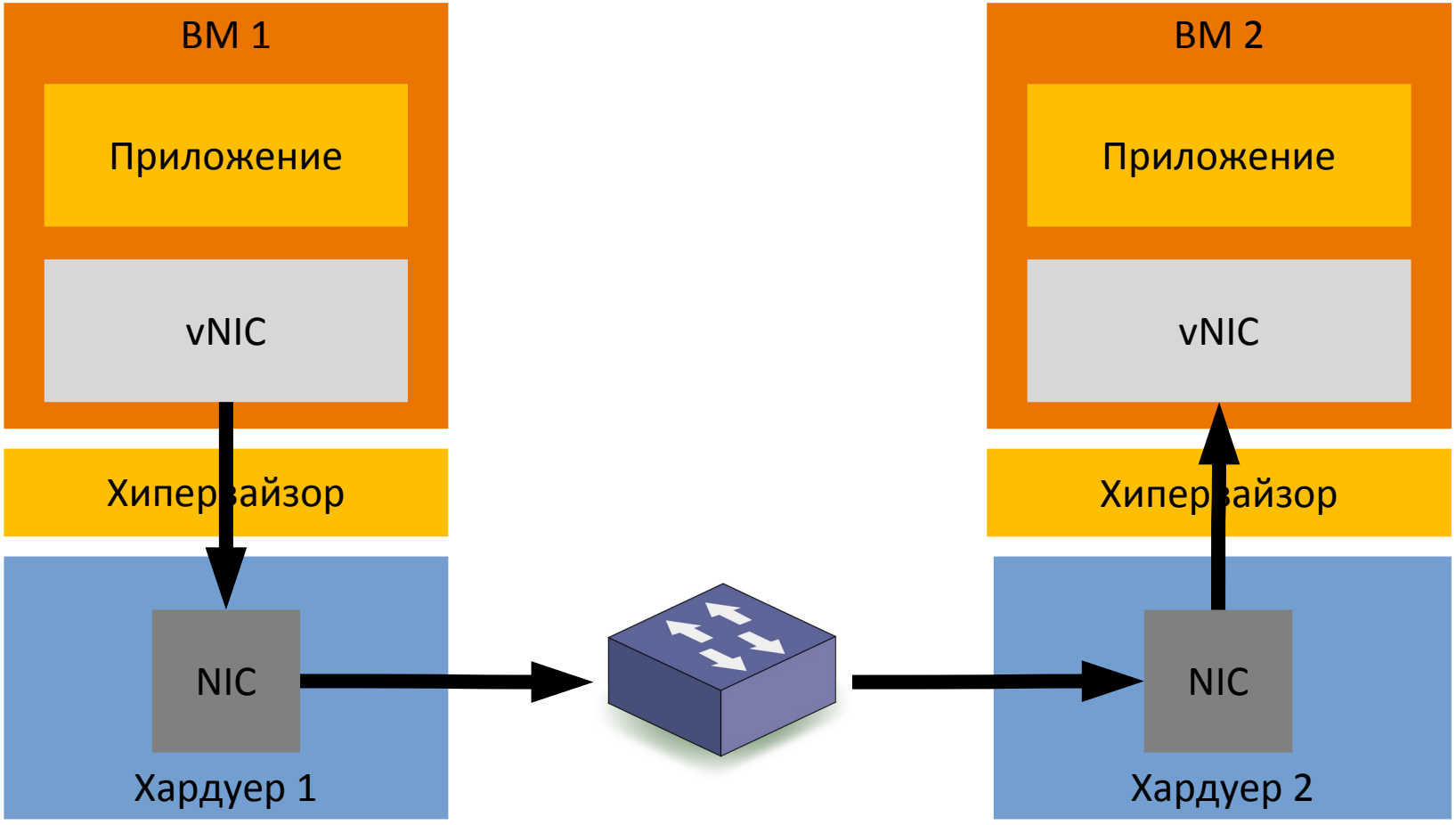
Свързване между виртуални машини



Свързване между виртуални машини



Свързване между виртуални машини



Виртуализация на Вход/Изхода

Виртуализация на В/И

- Виртуализацията на В/И се базира на управление на маршрутизацията на В/И заявки между виртуални устройства и споделения физически хардуер;
- Хипервайзорът виртуализира физическия хардуер и представя на всяка виртуална машина стандартизиран набор от виртуални устройства;
- Тези виртуални устройства ефективно емулират добре познат хардуер (с драйвери на устройства) и превеждат заявките на виртуалната машина към хардуера на системата;
- Тази стандартизация на драйвери на устройства също помага за стандартизиране на ВМ и лесната преносимост на различни платформи, тъй като всички ВМ са конфигурирани да работят на един и същ виртуален хардуер, независимо от реалния физически хардуер в системата;

Предимства

- **Гъвкавост** – виртуализацията на В/И осигурява по-голяма гъвкавост. Виртуалните В/И технологии могат да бъдат динамично разширени и обикновено заменят множество връзки с един кабел. Периодите за изпълнение на общи задачи на центъра за данни - като добавяне на сървъри, съхранение или мрежова свързаност - могат да бъдат намалени от дни до минути;
- **Намалена цена** – понижава разходите и позволява опростено управление на сървъра чрез използване на по-малко карти, кабели и портове на комутатора (switch). Опростява дизайна на мрежовия център за данни, като консолидира и използва по-добре мрежовите комутатори (switch) на LAN и SAN;
- **Намалено окабеляване** – необходим само един кабел за свързване на сървъри както за съхранение, така и за мрежов трафик. Това може да намали връзките в рамките на един сървърен шкаф с 70%;
- **Повишена плътност** – увеличава практическата плътност на В/И, като позволява повече връзки да съществуват в дадено пространство. Това дава възможност за по-голямо използване на 1U сървъри и блейд сървъри;

Работа на периферно устройство

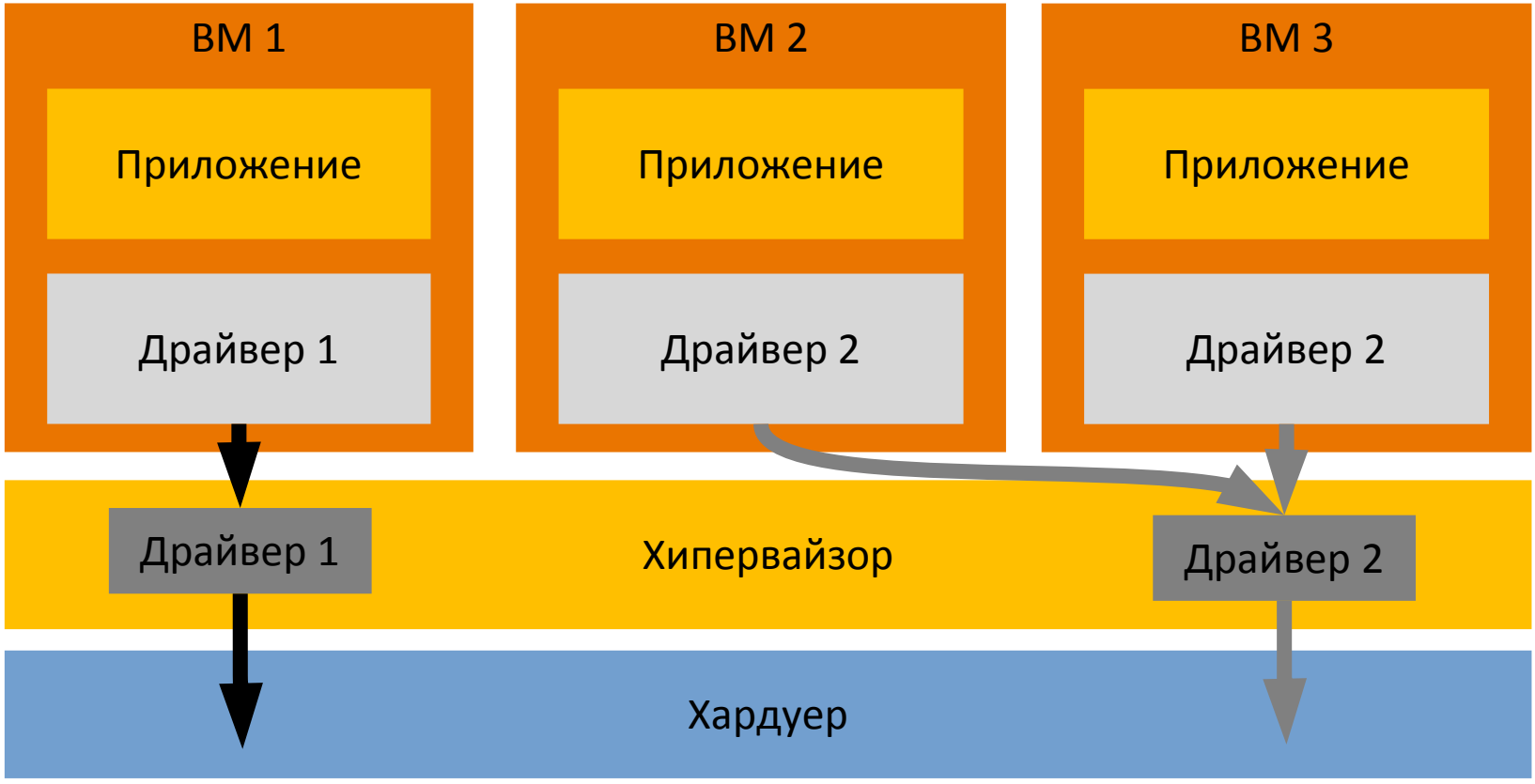


- Драйвера управлява (настройва и получава информация за състоянието) устройството чрез достъп до регистрите му;
- Устройството съобщава на драйвера за събития чрез прекъсване;
- Устройството може да използва DMA за обмен на големи обеми данни с паметта;

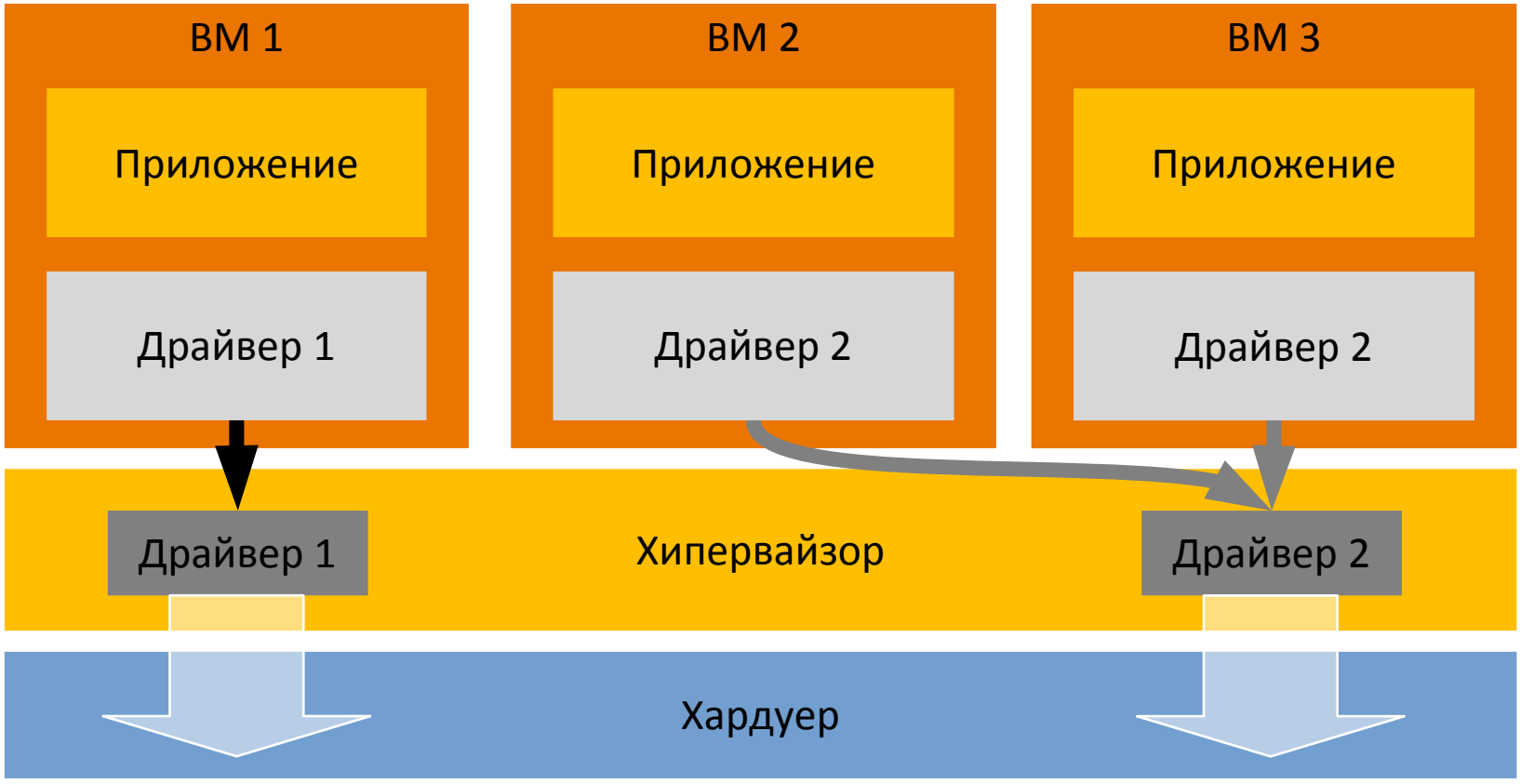
Виртуализацията на В/И изисква хипервайзора да предостави на guest ОС пълен аналог на интерфейса на устройството:

- Същия интерфейс: Софтуерна емулация, Директен достъп;
- Напълно нов интерфейс: Паравиртуализация;

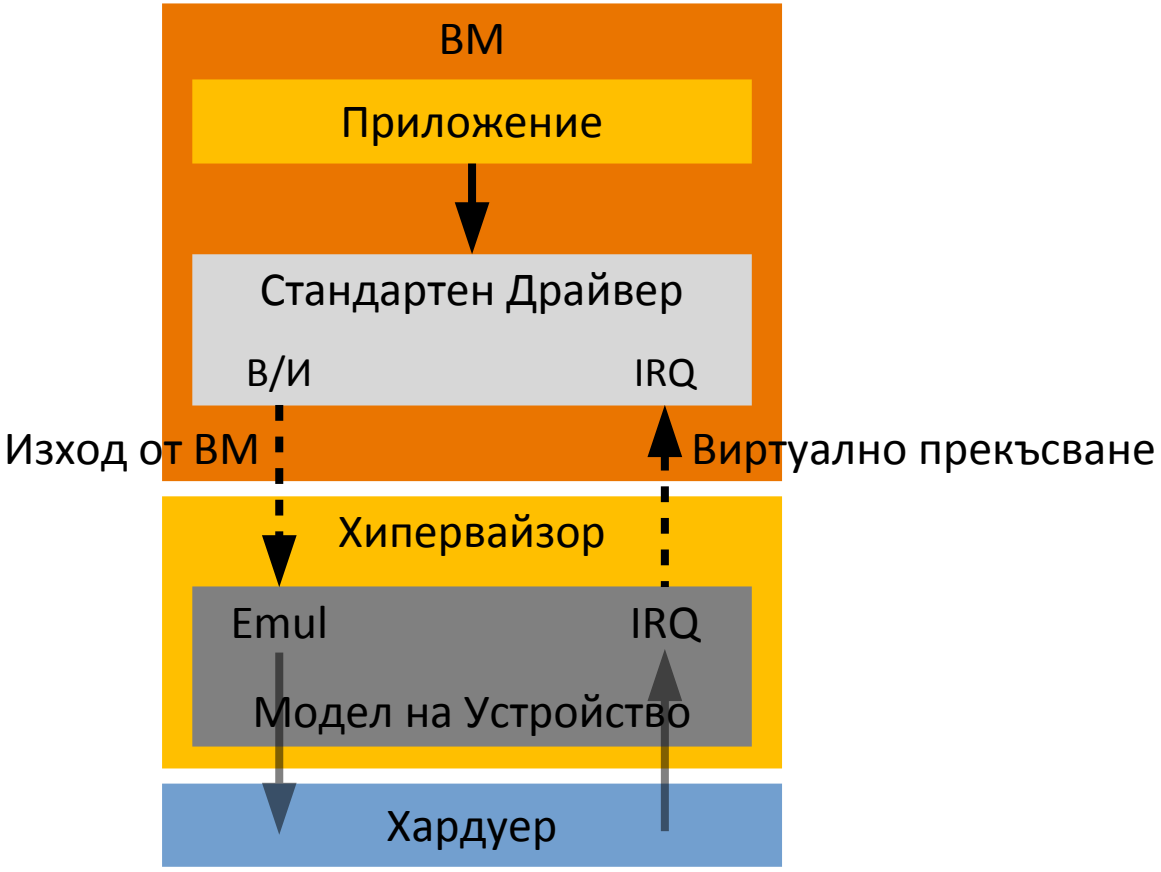
Два модела на работа на драйверите (Xen indirect driver)



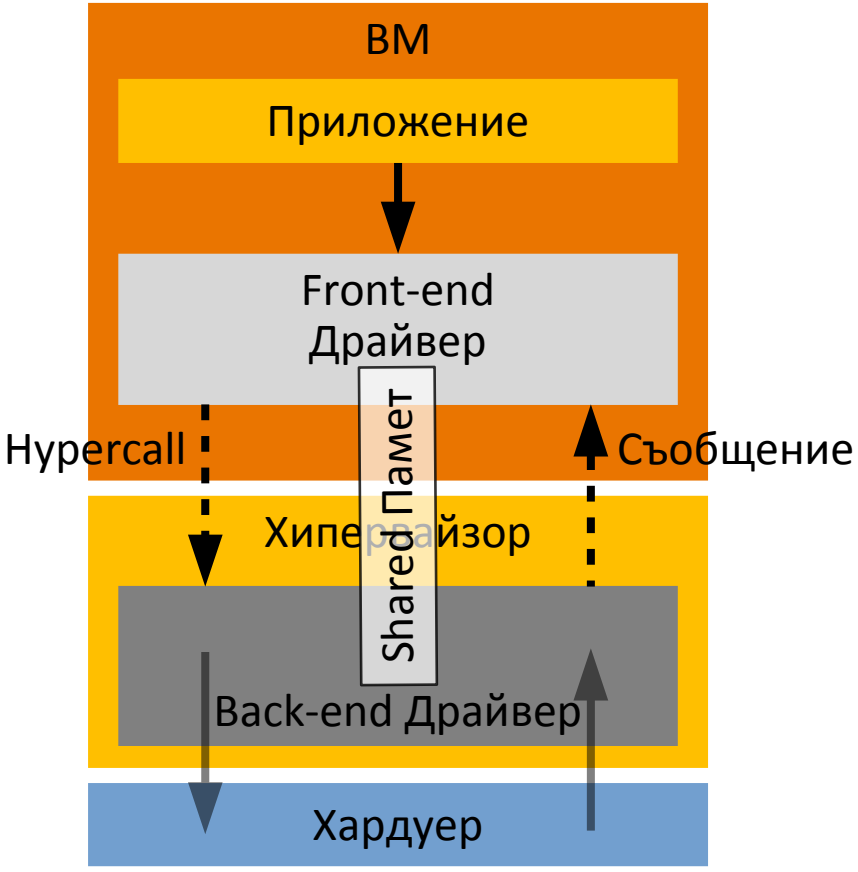
Два модела на работа на драйверите (ESXi direct driver)



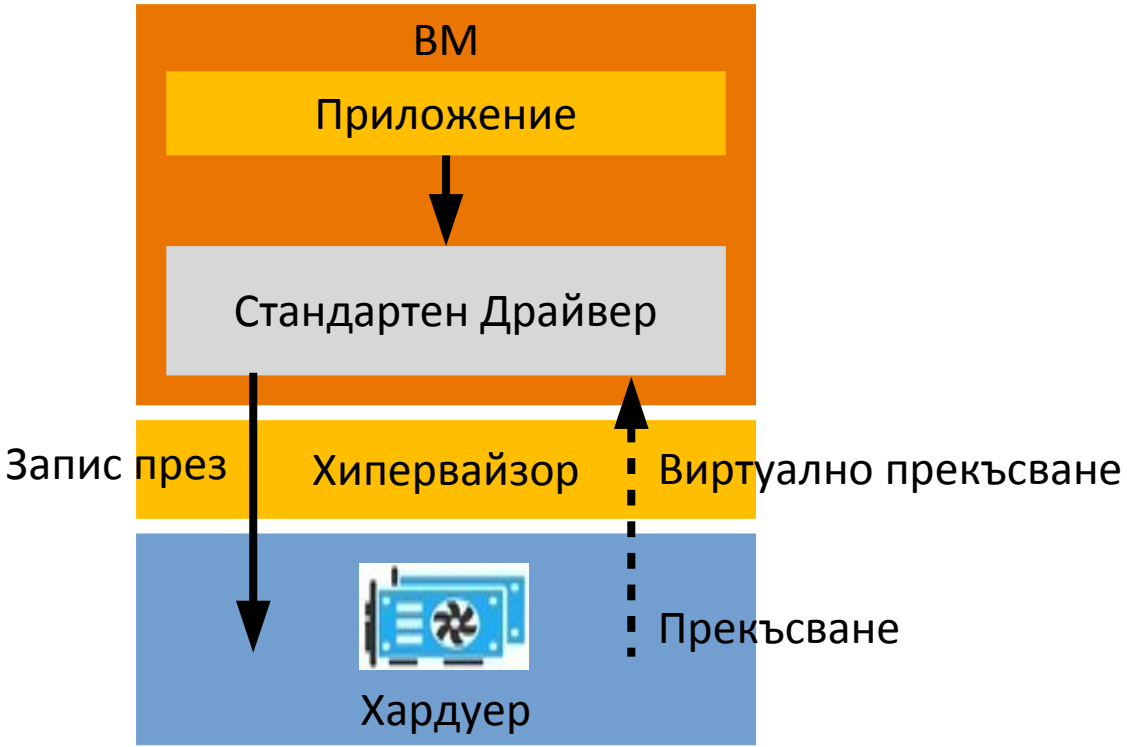
В/И – Софтуерна Емулация



В/И – Паравиртуализация



В/И – Директен достъп



Хардуерно Ускорена Виртуализация

Съвременни разширения на някои процесори за VM

- **VT-x**, **VT-i** и **AMD-V** – Набори инструкции хардуерно подпомагащи виртуализацията на Intel, Intel Itanium и AMD процесорите;
- **AMD-Vi** и Intel **VT-d** (I/O MMU virtualization) – Модула за управление на В/И и паметта (IOMMU) позволява на виртуалните машини директно да използват периферни устройства, като Ethernet, ускорени графични карти и контролери на твърдия диск, чрез DMA и ремапинг на прекъсванията. Това се нарича PCI passthrough. IOMMU също така позволява да се елиминират някои буфери, необходими за комуникацията с периферни устройства;
- **AMD AVIC** и **Intel APICv** – През 2012, AMD въвежда техния Advanced Virtual Interrupt Controller за подобряване на производителността на прекъсванията при виртуализация;
- **Intel GVT-d**, **GVT-g** и **GVT-s** – Технологии за виртуализация на GPU. Интегрираният графичен процесор може да бъде: резервиран за една ВМ (GVT-d), споделен между множество ВМ на базата на време деление, с нормалния графичен драйвер (GVT-g), или споделен между множество ВМ с виртуален графичен драйвер (GVT-s);
- **VT-c** – виртуализация на мрежата и комуникациите;

Въпроси?

apenev@uni-plovdiv.bg