

به نام خدا



ارائه‌ی درس مدارهای واسط

موضوع:

Nonvolatile Memory Express

استاد:

دکتر فصحتی

شمیم رحیمی ۴۰۱۱۰۵۹۵۶

نیم‌سال اول ۱۴۰۳-۱۴۰۴

پروتکل **NVMe** یک پروتکل ارتباطی است که به طور خاص برای حافظه‌های فلش مانند SSDها طراحی شده است. این پروتکل با هدف افزایش سرعت انتقال داده و کاهش تأخیر در مقایسه با پروتکل‌های قدیمی مانند **SATA** و **AHCI** توسعه یافته است. NVMe از گذرگاه **PCIe** (Peripheral Component Interconnect Express) استفاده می‌کند که امکان ارتباط مستقیم با CPU را فراهم می‌کند و به این ترتیب، عملکرد ذخیره‌سازی را به طور چشمگیری بهبود می‌بخشد. در این سند، به بررسی جامع و دقیق‌تر پروتکل NVMe، اجزای آن، کاربردها، مزایا و چالش‌های آن می‌پردازیم.

## 1. کاربرد و چرایی توسعه NVMe

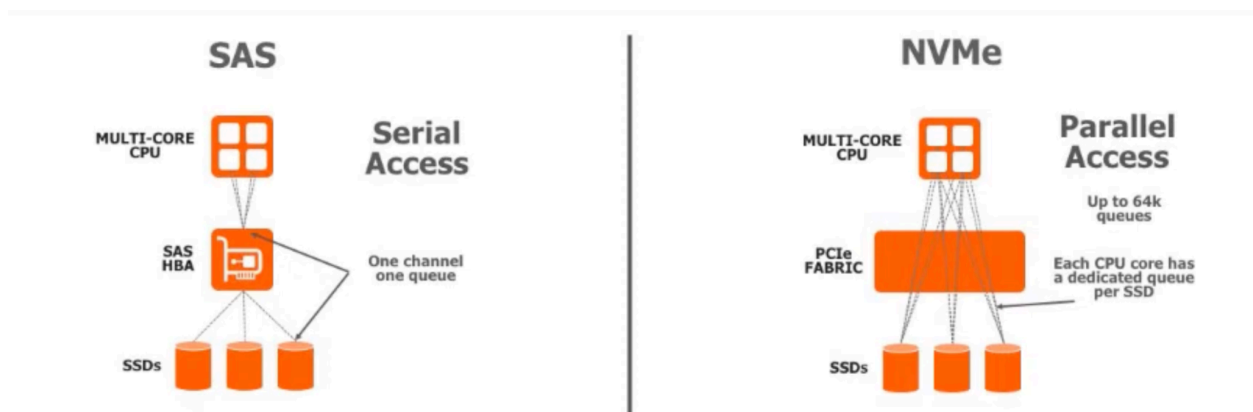
### کاربردهای NVMe

- پروتکل NVMe در سیستم‌هایی که نیاز به دسترسی سریع به داده‌ها دارند، به کار می‌رود. برخی از کاربردهای اصلی NVMe عبارتند از:
- **سرورها و مراکز داده:** در محیط‌هایی که نیاز به پردازش سریع داده‌ها و کاهش تأخیر وجود دارد، NVMe به عنوان یک راه‌حل ایده‌آل استفاده می‌شود.
  - **سیستم‌های گیمینگ و ویرایش ویدئو:** در این سیستم‌ها، سرعت خواندن و نوشتن داده‌ها بسیار مهم است و NVMe می‌تواند عملکرد بهتری را ارائه دهد.
  - **یادگیری ماشین و هوش مصنوعی:** در این حوزه‌ها، حجم داده‌ها بسیار زیاد است و نیاز به دسترسی سریع به داده‌ها وجود دارد. NVMe می‌تواند به بهبود عملکرد این سیستم‌ها کمک کند.

## چرایی توسعه NVMe

پروتکل‌های قدیمی مانند **SATA** و **AHCI** برای هارد دیسک‌های مکانیکی (HDD) طراحی شده بودند و محدودیت‌های زیادی در زمینه سرعت و تأخیر داشتند. با ظهور حافظه‌های فلش مانند SSD، نیاز به یک پروتکل جدید که بتواند از پتانسیل کامل این حافظه‌ها استفاده کند، احساس شد. NVMe با استفاده از گذرگاه PCIe و ساختار پیشرفته‌تر، توانست این محدودیت‌ها را برطرف کند.

NVMe از عمق صف و تعداد دستورات بسیار بیشتری نسبت به SATA و SAS پشتیبانی می‌کند که این امر باعث افزایش کارایی و کاهش تأخیر می‌شود.



## 2. معماری و اجزای NVMe

معماری NVMe شامل چندین بخش اصلی است که در ادامه بررسی می‌شوند:

### Frontend (قسمت جلویی)

- **PCIe Bandwidth: NVMe** از پهنای باند بالای PCIe برای انتقال داده‌ها استفاده می‌کند.
- **Host CPU: NVMe** به طور مستقیم با CPU ارتباط برقرار می‌کند که این امر باعث کاهش تأخیر و افزایش سرعت می‌شود.

## Backend (قسمت پشتی)

- **Flash Channel Controller**: این بخش مسئول مدیریت کانال‌های فلش و انتقال داده‌ها به NAND Flash است.
- **NAND Flash**: این بخش شامل حافظه‌های فلش است که داده‌ها در آن‌ها ذخیره می‌شوند.

## Firmware (نرم‌افزار داخلی)

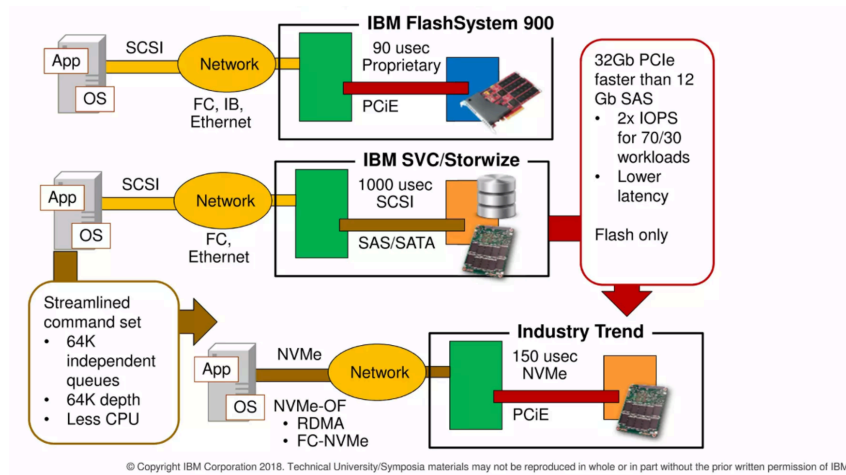
- **GC Manager**: مدیریت Garbage Collection (جمع‌آوری داده‌های بی‌استفاده) را بر عهده دارد.
- **Meta Manager**: مدیریت متاداده‌ها (Metadata) را انجام می‌دهد.
- **NVMe Manager**: مدیریت پروتکل NVMe و ارتباط با Host را بر عهده دارد.

## 3. لایه فیزیکی، اتصالات و مدارات

### لایه فیزیکی

NVMe از سیگنالینگ تفاضلی (Differential Signaling) برای انتقال داده‌ها استفاده می‌کند. این روش شامل دو سیم برای هر سیگنال است (یک سیم داده مثبت و یک سیم داده منفی). این روش مزایای زیر را دارد:

- **کاهش نویز الکترومغناطیسی**: سیگنالینگ تفاضلی باعث کاهش نویز و افزایش پایداری سیگنال‌ها می‌شود.
- **افزایش سرعت انتقال داده**: این روش امکان انتقال داده‌ها با سرعت بالاتر را فراهم می‌کند.
- **کاهش تداخل بین سیگنال‌ها**: سیگنالینگ تفاضلی باعث کاهش تداخل بین سیگنال‌ها می‌شود.



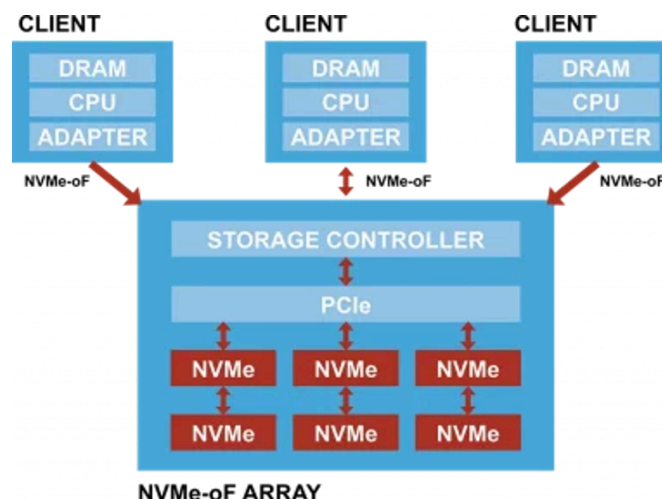
## اتصالات ضروری و اختیاری در NVMe

### • اتصالات ضروری:

- **PCIe Lane(s):** برای انتقال داده‌ها استفاده می‌شود.
- **Clock Signal:** برای همگام‌سازی داده‌ها استفاده می‌شود.
- **Power Supply:** برای تأمین برق دستگاه استفاده می‌شود.

### • اتصالات اختیاری:

- **GPIOs:** برخی از SSDهای NVMe از GPIOs برای کنترل بهتر و عملکرد پیشرفته استفاده می‌کنند.



#### 4. نوع ارتباط و انکودینگ

##### نوع ارتباط

NVMe از ارتباط **سریال** استفاده می‌کند که برخلاف روش‌های قدیمی مانند **SATA** و **PATA** که از ارتباط موازی استفاده می‌کردند، کارایی بهتری دارد. در ارتباط سریال، داده‌ها به صورت پشت سر هم منتقل می‌شوند که این امر باعث کاهش **Overhead** و افزایش سرعت انتقال داده می‌شود.

##### • مزایای ارتباط سریال:

- **کاهش تأخیر:** در ارتباط سریال، داده‌ها به صورت مستقیم و بدون نیاز به هم‌زمان‌سازی بین چندین خط انتقال، منتقل می‌شوند که این امر باعث کاهش تأخیر می‌شود.
- **افزایش پهنای باند:** ارتباط سریال امکان استفاده از پهنای باند بالاتر را فراهم می‌کند.
- **کاهش نویز و تداخل:** در ارتباط سریال، تعداد خطوط انتقال کمتر است که این امر باعث کاهش نویز و تداخل بین سیگنال‌ها می‌شود.

##### انکودینگ

NVMe از انکودینگ **128b/130b** استفاده می‌کند که نسبت به انکودینگ‌های قدیمی مانند **8b/10b** (که در PCIe 2.0 استفاده می‌شد) کارایی بهتری دارد.

##### • مزایای انکودینگ 128b/130b:

- **کاهش Overhead:** در انکودینگ 8b/10b، به ازای هر 8 بیت داده، 2 بیت اضافه برای کنترل خطا اضافه می‌شد که این امر باعث افزایش Overhead می‌شد. در انکودینگ 128b/130b، تنها 2 بیت اضافه برای هر 128 بیت

داده اضافه می‌شود که این امر باعث کاهش Overhead و افزایش پهنای باند می‌شود.

- **افزایش سرعت انتقال داده:** با کاهش Overhead، سرعت انتقال داده افزایش می‌یابد.

## 5. اتصالات بین چندین دستگاه

NVMe از طریق **PCIe Bus** می‌تواند از چندین دستگاه پشتیبانی کند. روش‌های اتصال چندین دستگاه در NVMe عبارتند از:

### 1. استفاده از چندین اسلات M.2 یا U.2

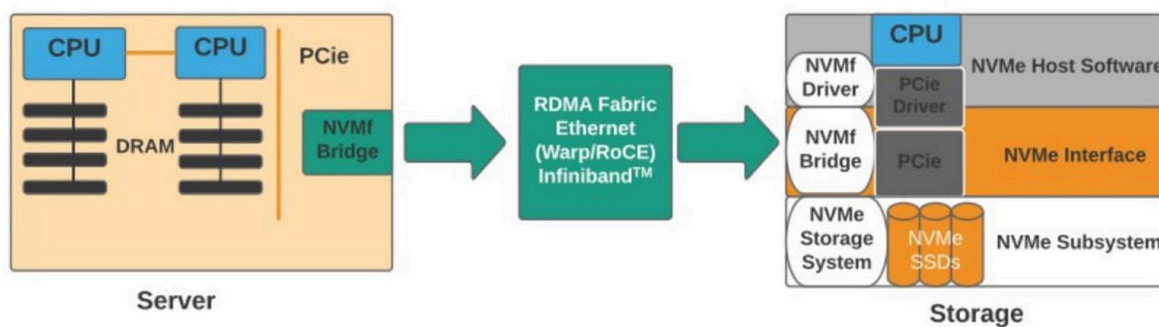
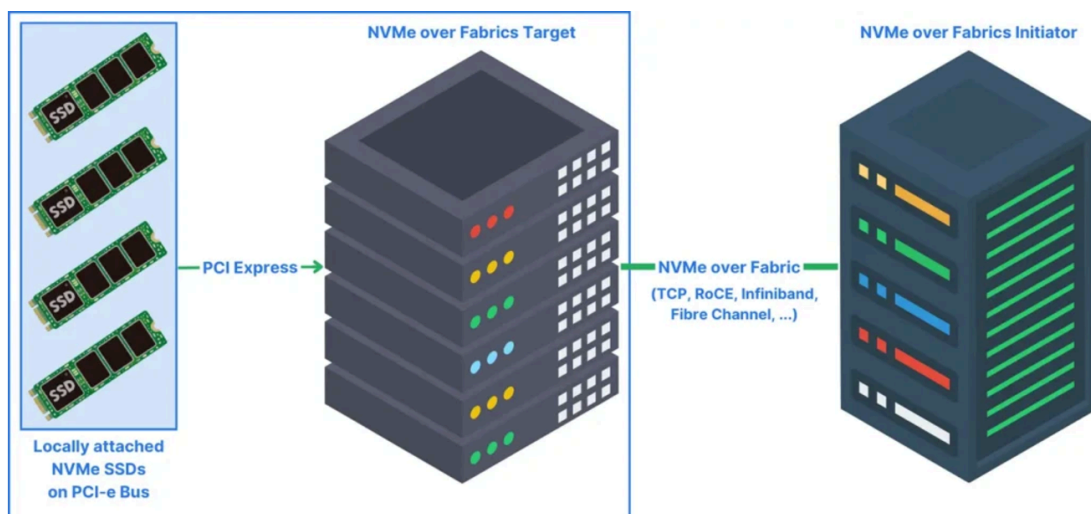
در مادربردهای مدرن، می‌توان از چندین اسلات **M.2** یا **U.2** برای اتصال چندین دستگاه NVMe استفاده کرد. این اسلات‌ها به طور مستقیم به PCIe Bus متصل هستند و امکان اتصال چندین SSD NVMe را فراهم می‌کنند.

### 2. NVMe over Fabrics (NVMe-oF)

این روش امکان اتصال SSDها به چندین سرور از طریق شبکه را فراهم می‌کند. NVMe-oF از پروتکل‌های شبکه مانند **RDMA**، **TCP/IP** و **InfiniBand** استفاده می‌کند تا امکان انتقال داده‌ها بین چندین سرور و دستگاه‌های ذخیره‌سازی را فراهم کند.

#### • مزایای NVMe-oF:

- **اتصال چندین دستگاه:** NVMe-oF امکان اتصال چندین دستگاه ذخیره‌سازی به چندین سرور را فراهم می‌کند.
- **کاهش تأخیر:** با استفاده از پروتکل‌های شبکه سریع مانند RDMA، تأخیر انتقال داده‌ها کاهش می‌یابد.
- **مقیاس‌پذیری:** NVMe-oF امکان افزایش تعداد دستگاه‌های ذخیره‌سازی و سرورها را بدون کاهش عملکرد فراهم می‌کند.



### 3. PCIe Switches

این روش امکان استفاده از چندین SSD بر روی یک گذرگاه PCIe را فراهم می‌کند. با استفاده از **PCIe Switches**، می‌توان چندین دستگاه NVMe را به یک PCIe Bus متصل کرد و از طریق آن‌ها داده‌ها را انتقال داد.

آدرس‌دهی در NVMe روی PCIe



NVMe روی PCIe از آدرس‌دهی ساده‌ای استفاده می‌کند که شامل **Bus Number**، **Device Number**، **Function Number** و **Namespace ID (NSID)** است. هر SSD NVMe می‌تواند چندین Namespace داشته باشد که فضای ذخیره‌سازی را جدا می‌کند.

- **مثال آدرس‌دهی:** یک SSD NVMe ممکن است آدرس زیر را داشته باشد:

○ **Bus Number:** 01

○ **Device Number:** 00

○ **Function Number:** 0

○ **Namespace ID (NSID):** 1

آدرس‌دهی در NVMe-oF

در NVMe-oF، آدرس‌دهی پیچیده‌تر است و از **Transport Address** و **Qualified Name (QN)** برای شناسایی دستگاه‌ها استفاده می‌شود.

- **Transport Address:** این آدرس شامل اطلاعاتی مانند آدرس IP و پورت است که برای شناسایی دستگاه‌ها در شبکه استفاده می‌شود.
- **Qualified Name (QN):** این آدرس شامل اطلاعاتی مانند نام دستگاه و Namespace است که برای شناسایی دقیق‌تر دستگاه‌ها استفاده می‌شود.
- **مثال آدرس‌دهی در NVMe-oF:** یک دستگاه NVMe over TCP ممکن است آدرسی مانند 192.168.1.10:4420 داشته باشد.

7. مدیریت جریان داده

NVMe از روش‌های مختلفی برای مدیریت جریان داده استفاده می‌کند تا عملکرد بهینه و تأخیر کم داشته باشد. برخی از این روش‌ها عبارتند از:

1. صف‌های چندگانه (Multiple Queues)

NVMe از صف‌های ارسال (SQ) و دریافت (CQ) چندگانه استفاده می‌کند که به پردازش هم‌زمان چندین درخواست کمک می‌کند. هر دستگاه NVMe می‌تواند تا 65,000 صف داشته باشد که این امر باعث افزایش کارایی و کاهش تأخیر می‌شود.

## 2. مدیریت جریان داده با محدودیت QoS

برخی از SSDهای NVMe از **Rate Limiting** برای اعمال محدودیت سرعت خواندن/نوشتن استفاده می‌کنند. این روش به مدیریت جریان داده و جلوگیری از ترافیک بیش از حد کمک می‌کند.

## 3. جریان داده در NVMe-oF

در NVMe-oF، جریان داده‌ها با استفاده از پروتکل‌هایی مانند **TCP**، **RDMA** یا **InfiniBand** مدیریت می‌شود. این پروتکل‌ها امکان انتقال داده‌ها با سرعت بالا و تأخیر کم را فراهم می‌کنند.

## 8. تشخیص و تصحیح خطا

NVMe دارای مکانیزم‌های پیشرفته‌ای برای تشخیص و تصحیح خطا است. این مکانیزم‌ها در سطوح مختلفی عمل می‌کنند:

### 1. تصحیح خطا در حافظه

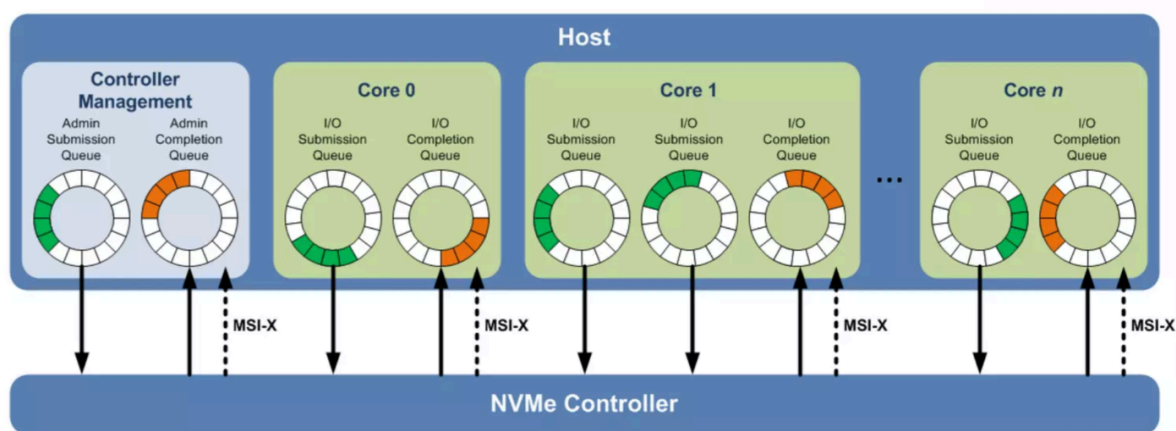
از **LDPC** (Low-Density Parity-Check Code) برای تصحیح خطای سلول‌های حافظه استفاده می‌شود. این کدها امکان تشخیص و تصحیح خطاهای چند بیتی را فراهم می‌کنند.

### 2. تصحیح خطا در انتقال داده

در PCIe، از **Replay Buffers** برای ارسال مجدد داده‌ها در صورت خطا استفاده می‌شود. این روش باعث افزایش قابلیت اطمینان انتقال داده‌ها می‌شود.

3. تصحیح خطا در فرمان‌ها

در سطح نرم‌افزار، از **Error Recovery Mechanisms** برای شناسایی و تصحیح فرمان‌های نامعتبر استفاده می‌شود. این مکانیزم‌ها امکان بازیابی از خطاهای نرم‌افزاری را فراهم می‌کنند.



9. انواع پیام‌ها در NVMe

پیام‌های NVMe به سه دسته اصلی تقسیم می‌شوند:

1. دستورات (Commands)

این دستورات توسط Host به کنترلر ارسال می‌شوند و شامل دستورات خواندن، نوشتن و مدیریتی هستند. هر دستور شامل فیلدهایی مانند **Opcode**، **Namespace ID**، **Logical Block Address (LBA)**، **(NSID)** و **Data Length** است.

فیلد	توضیحات
Opcode	نوع دستور (Read, Write, Identify)
Namespace ID (NSID)	شماره namespace مورد استفاده
Logical Block Address (LBA)	آدرس بلوک مورد نظر برای خواندن/نوشتن
Data Length	مقدار داده انتقالی
Metadata	اطلاعات جانبی برای محافظت از داده‌ها
Command Identifier	شماره یکتای دستور

## 2. پاسخ‌ها (Responses)

این پیام‌ها توسط کنترلر به Host ارسال می‌شوند و شامل اطلاعاتی مانند وضعیت دستور و داده‌های خوانده‌شده هستند. هر پاسخ شامل فیلدهایی مانند **Status Code**، **Data Buffer** و **Command Identifier** است.

فیلد	توضیحات
Status Code	وضعیت اجرای دستور (موفق/ناموفق)
Data Buffer	در پاسخ به خواندن، شامل داده‌های خوانده‌شده
Command Identifier	شناسه دستور مربوطه
Completion Queue Entry (CQE)	نشانه پایان پردازش دستور

## 3. پیام‌های خاص (Special Messages)

این پیام‌ها برای هماهنگی بین کنترلر و Host استفاده می‌شوند و شامل **Completion Queue Entries (CQEs)** و **Asynchronous Event Notifications (AENs)** هستند. CQEs زمانی ارسال می‌شوند که یک دستور تکمیل شود و AENs برای اطلاع‌رسانی رویدادهای مهم مانند پر شدن حافظه یا خرابی دستگاه استفاده می‌شوند.

- SlideShare, "NVMe Overview," Available:  
[https://www.slideshare.net/slideshow/nvme-overview/88249387?from\\_search=0](https://www.slideshare.net/slideshow/nvme-overview/88249387?from_search=0)
- SlideShare, "NVMe Revolution," Available:  
[https://www.slideshare.net/slideshow/s104878-nvmerevolutionjburgv1809b/115662865?from\\_search=1](https://www.slideshare.net/slideshow/s104878-nvmerevolutionjburgv1809b/115662865?from_search=1)
- StorageReview, "NVMe & NVMe-of Background & Overview," Available:  
<https://www.storagereview.com/review/nvme-nvme-of-background-overview>
- NetApp, "What is NVMe?" Available:  
[https://www.netapp.com/data-storage/nvme/what-is-nvme/#:~:text=NvMe%20\(nonvolatile%20memory%20express\)%20is.all%20types%20of%20enterprise%20workloads](https://www.netapp.com/data-storage/nvme/what-is-nvme/#:~:text=NvMe%20(nonvolatile%20memory%20express)%20is.all%20types%20of%20enterprise%20workloads)
- Wikipedia, "NVM Express," Available:  
[https://en.wikipedia.org/wiki/NVM\\_Express](https://en.wikipedia.org/wiki/NVM_Express)