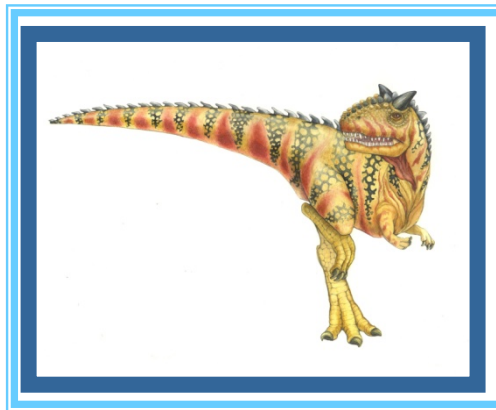


Bölüm 12:

İkincil Depolama Yapısı





Bölüm 12: İkincil Depolama Yapısı

- Yığın Depolama Yapısına (Mass Storage Structure) Genel Bakış
- Disk Yapısı
- Disk Ekleme (Disk Attachment)
- Disk İş Planlaması
- Disk Yönetimi
- Değiş tokuş Alan yönetimi (Swap-Space Management)
- RAID Yapısı
- Stabil Depolama Uygulaması (Stable-Storage Implementation)
- Üçüncü Seviye Depolama Aygıtları (Tertiary Storage Devices)
- İşletim Sistemi Sorunları
- Performans Sorunları





Yığın Depolama Yapısına Genel Bakış

- Manyetik Teyp (Magnetic tape)
 - önceleri ikincil depolama ortamıydı
 - Göreceli olarak kalıcıdır ve büyük miktarda veri tutabilir
 - erişim zamanı yavaştır
 - rastgele erişim diske göre ~1000 kere daha yavaştır
 - genelde yedekleme, düzenli kullanılmayan verilerin depolanması ve sistemler arası transfer ortamı olarak kullanılır
 - Sırada (spool) tutulur ve önceki okuma-yazma kafasını ileri ve geri sarar
 - Veri kafanın altında ise transfer oranı diske yakındır
 - 20-200GB depolama
 - Yaygın teknolojiler 4mm, 8mm, 19mm, LTO-2 ve SDLT





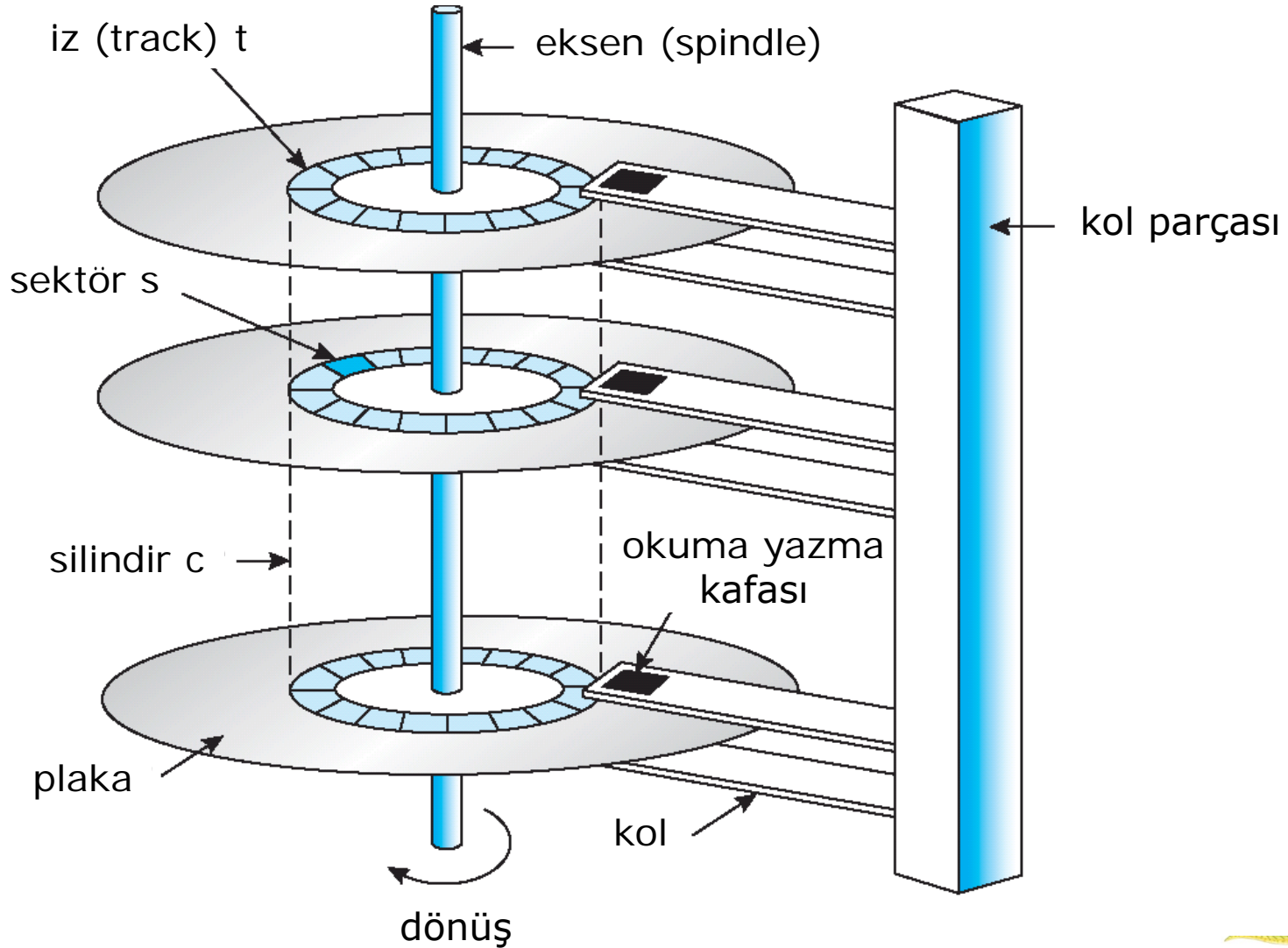
Yığın Depolama Yapısına Genel Bakış

- Modern bilgisayarlarda ikincil depolama manyetik diskler tarafından sağlanmaktadır.
 - Sürücüler saniyede 60 – 200 kere dönerler
 - **Transfer oranı (*Transfer rate*)** sürücü ile bilgisayar arasındaki veri akış oranıdır.
 - **Konumlanma zamanı (rastgele erişim zamanı) (*Positioning time (random-access time)*** disk kolunun istenen silindire gidiş zamanı (**arama zamanı**) (seek time) ve istenen sektörün disk kafası altında dönme zamanının (**dönüş gecikmesi**) (rotational latency) toplamıdır.
 - **Kafa çarpışması (Head crash)** disk kafasının disk alanı ile teması durumunda ortaya çıkar ve kötü sonuçlar doğurabilir
- Diskler çıkarılabilir
- Sürücü bilgisayara **G/Ç veriyolu (*I/O Bus*)** aracılığıyla bağlanır
 - Farklı veriyolları mevcuttur - **EIDE, ATA, SATA, USB, Fibre Channel, SCSI**
 - Bilgisayardaki **Sunucu kontrolörü (Host controller)** diskteki veya depolama dizisindeki **disk kontrolörü** ile konuşmak için veriyolunu kullanır





Hareketli Kafa Disk Mekanizması (Moving-head Disk Mechanism)





Disk Yapısı

- Disk sürücüler mantıksal bloklardan oluşan tek boyutlu diziler şeklinde adreslenirler. Mantıksal blok en küçük transfer birimidir.
- Mantıksal blokların tek boyutlu dizisi diskteki sektörler içerisine sıralı olarak eşlenir.
 - Sektör 0 en dıştaki **silindirin** ilk **izinin** ilk **sektörüdür**.
 - Eşleme önce o iz, daha sonra bu silindirdeki diğer izler, daha sonra da en dıştakinden en içtekine doğru kalan silindirler sırası olmak üzere devam eder.





Disk Ekleme (Disk Attachment)

Bilgisayarlar disk depolama alanlarına iki şekilde erişir:

- G/Ç portları aracılığıyla
- Uzaktaki bir sunucudaki dağıtık dosya sistemi aracılığıyla

1- Sunucuya eklenmiş depolamaya yerel G/Ç portları aracılığıyla erişilir

- Tipik masaüstü PC'lerde **IDE**, **ATA** veya **SATA** G/Ç veriyolu mimarisi kullanılır.
- Daha üst seviye bilgisayar ve sunucular ise **SCSI** veriyolunu kullanır.
- **SCSI**'nin kendisi bir veriyoludur ve bir kabloda 16 ağıta kadar yer alabilir. Sunucuda bir **SCSI başlatıcısı (SCSI initiator)** ve 15 adete kadar **SCSI hedefinden (SCSI targets)** (SCSI diskler) oluşur.
 - Her hedef 8 adete kadar **mantıksal birim** (aygıt kontrolörüne bağlı diskler) barındırabilir
- **FC** yüksek hızlı seri bir mimaridir ve optik fiber veya dört iletkenli bakır kabloda çalışır. İki tipi vardır:
 - 24 bit'lik adres alanı kullanarak anahtarlanan – sunucuların birçok depolama birimine bağlandığı **depolama alan ağları'nın (storage area networks (SAN))** temelidir.
 - 126 ağıt için kullanılabilecek **Aracılık Yapan Döngü (Arbitrated Loop (FC-AL))** olabilir

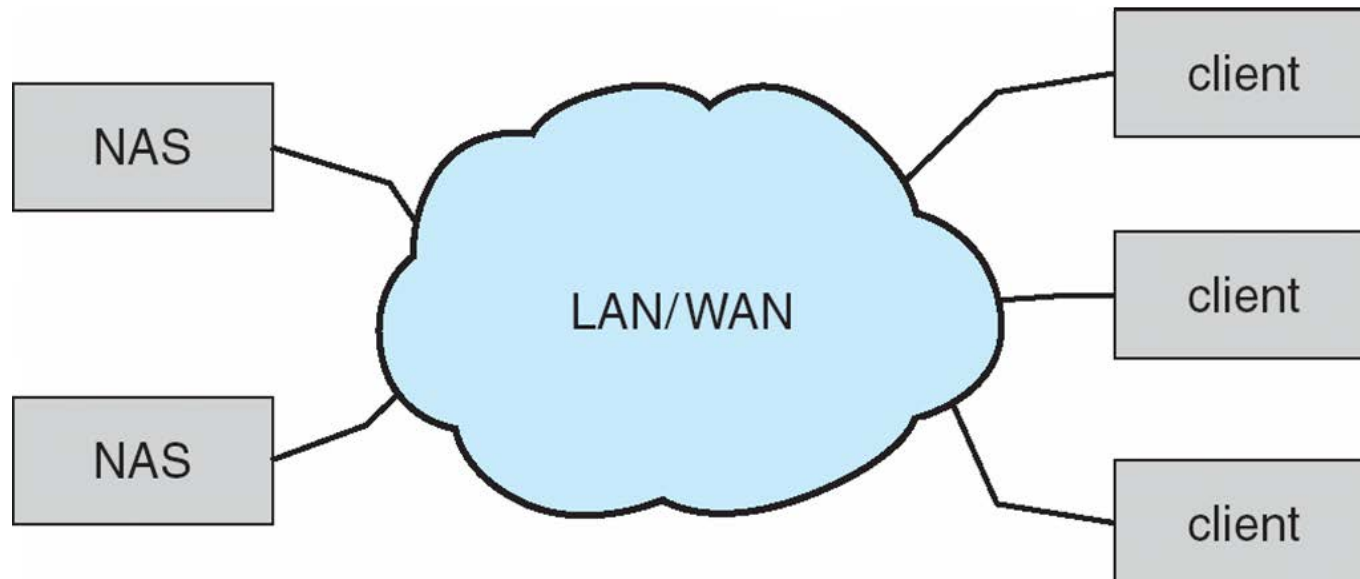




Disk Ekleme (Disk Attachment)

2. Ağa Bağlı Depolama (Network-Attached Storage)

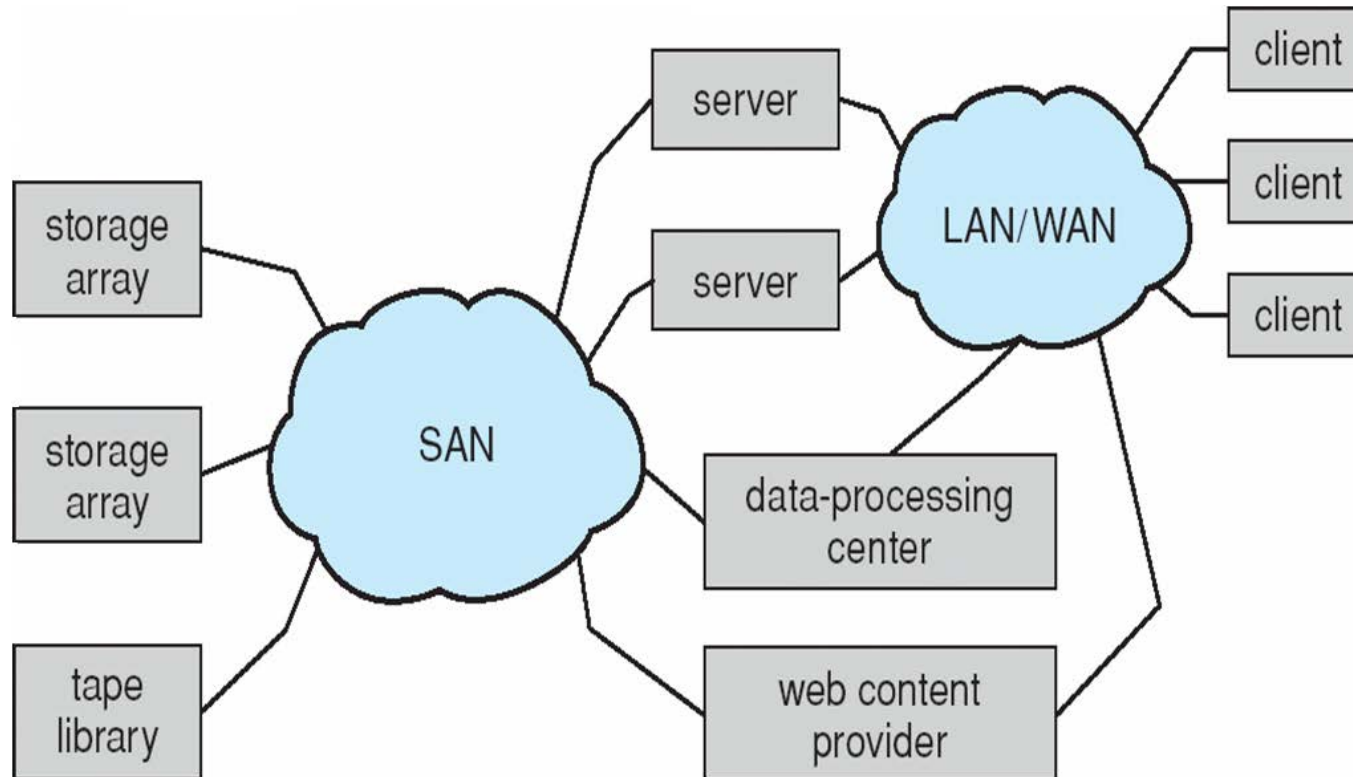
- **Ağa Bağlı Depolama (*Network-attached storage (NAS)*)** yerel bir bağlantı (veriyolu gibi) yerine ağ üzerinde yapılan depolamadır
- NFS ve CIFS yaygın protokollerdir
- Sunucu ve depolama arasında **Uzak Prosedür Çağrıları (*Remote Procedure Calls (RPC)*)** kullanılarak gerçekleştirilir
- Yeni **iSCSI** protokolü SCSI protokolünü taşımak için IP protokolünü kullanır





Depolama Alan Ağı (Storage Area Network)

- Büyük depolama ortamlarında yaygındır
- Çoklu sunucular çoklu depolama dizilerine bağlıdır - esnek





Disk İş Planlaması

- İşletim sistemi donanımı verimli kullanmaktan sorumludur – disk sürücüler için bu hızlı bir erişim zamanı ve disk bant genişliğine (bandwidth) sahip olmak anlamına gelir.
- Erişim zamanının iki ana bileşeni mevcuttur:
 - **Arama zamanı (*Seek time*)** diskin kafaları istenen sektörün bulunduğu silindire götürmesi için gereken zamandır
 - **Dönüşsel gecikme (*Rotational latency*)** ise diskin istenen sektörü disk kafasına döndürmesi için beklenen ek süredir
- Arama zamanı minimize edilmelidir
- Arama zamanı \approx Arama uzaklığı
- **Disk bant genişliği** ise transfer edilen toplam byte sayısının ilk servis isteği ile son transferin bitmesi arasında geçen toplam zamana bölünmesidir.





Disk İş Planlaması

- Disk G/Ç isteklerini gerçekleştirmek için farklı algoritmalar mevcuttur
- Bunları örnek bir istek kuyruğu üzerinde inceleyelim (0-199 silindirlerdeki blok sayıları).

98, 183, 37, 122, 14, 124, 65, 67

Kafa işaretçisi (Head pointer) pozisyonu : 53

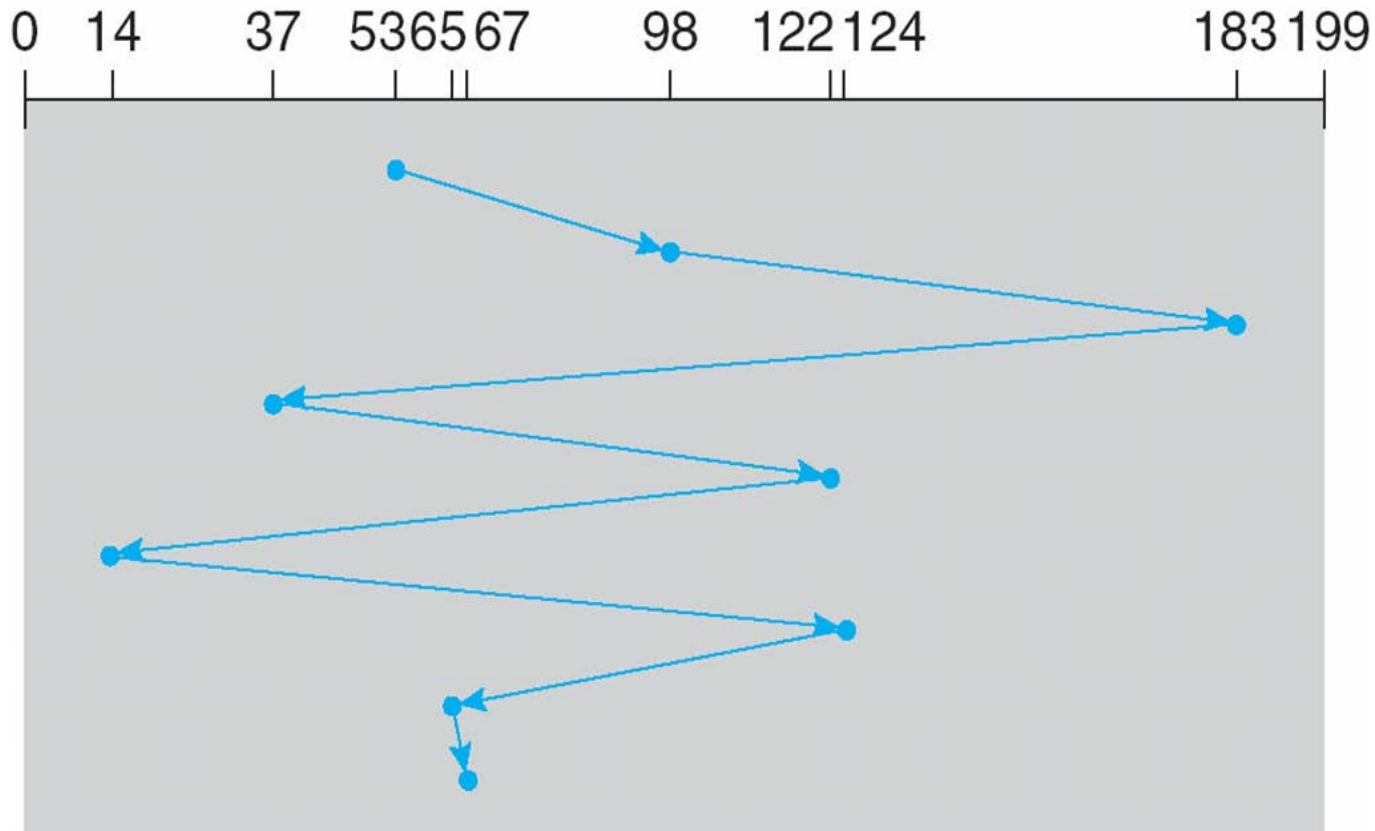




FCFS (First-Come First-Served)

queue = 98, 183, 37, 122, 14, 124, 65, 67

head starts at 53



Gösterimde 640 toplam kafa hareketi gösterilmektedir





SSTF (Shortest Seek Time First)

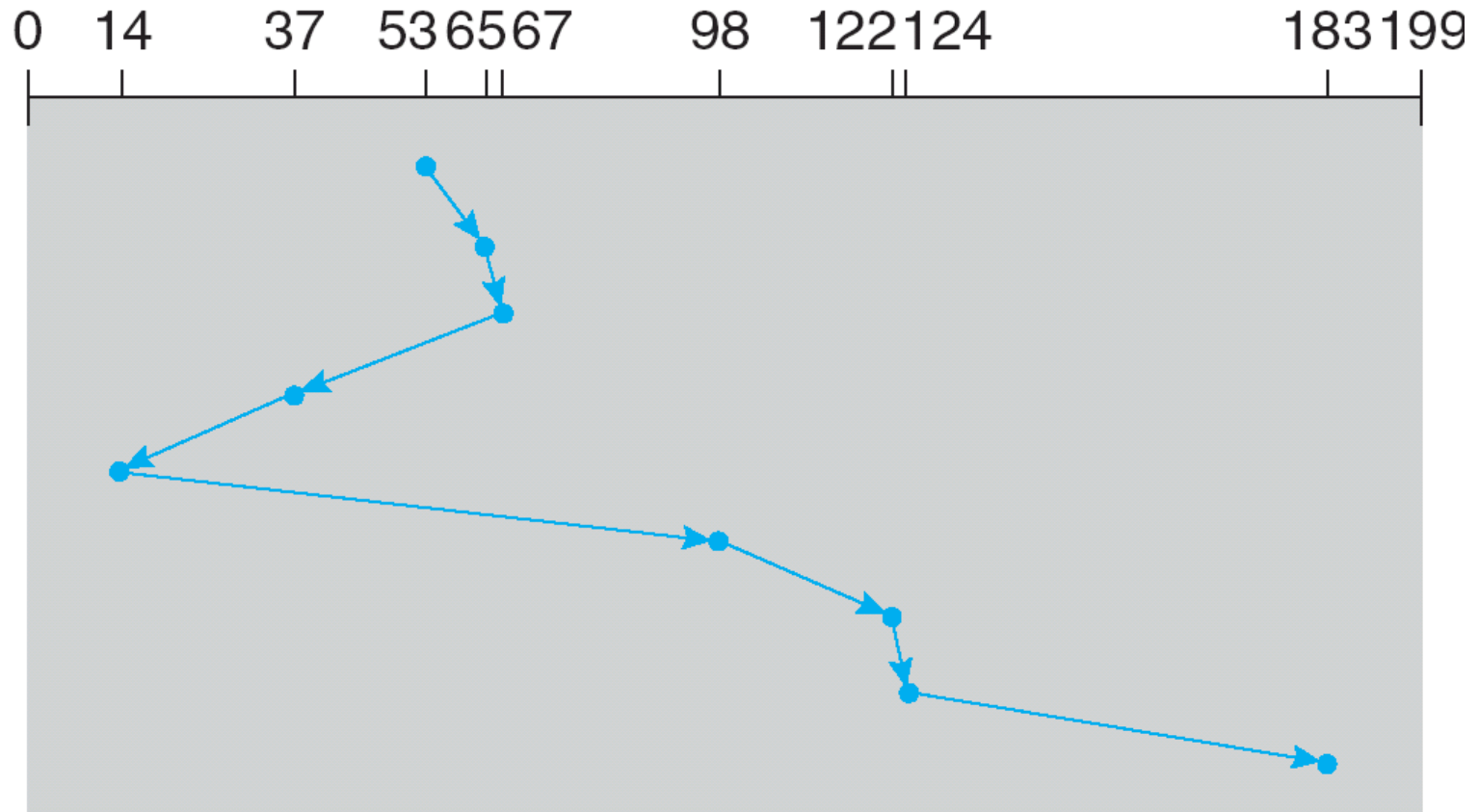
- Mevcut kafa pozisyonu için minimum arama zamanına sahip isteği seçer.
- SSTF iş planlaması SJF iş planlama algoritmasının bir biçimidir. Bazı istekler için açlığa (starvation) sebep olabilir.





SSTF

queue = 98, 183, 37, 122, 14, 124, 65, 67
head starts at 53



Gösterimde toplam kafa hareketleri 236'dır.





SCAN

- Disk kolu diskin başı ya da sonunda bulunarak başlar ve servis vermeye diskin diğer ucuna gidinceye kadar devam eder ve buraya ulaşınca diğer yöne doğru çalışmaya devam eder.
- Bazen ***asansör algoritması (elevator algorithm)*** olarak da adlandırılır.

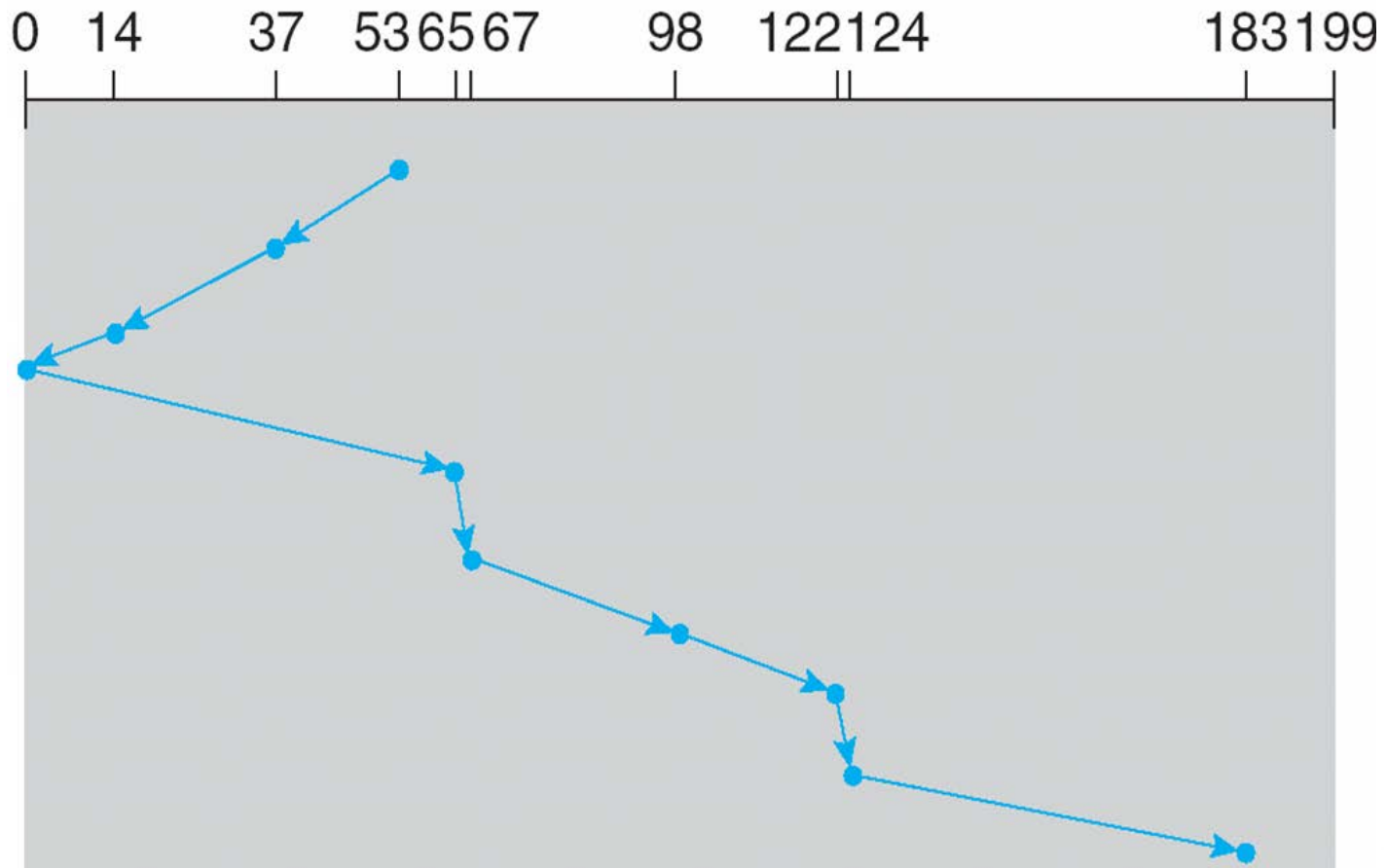




SCAN

queue = 98, 183, 37, 122, 14, 124, 65, 67

head starts at 53



Gösterimde 236 toplam kafa hareketi görülmektedir.





C-SCAN (Circular SCAN)

- SCAN'e göre daha değişmeyen (uniform) bekleme zamanı sunar.
- Kafa diskin bir ucundan diğer ucuna doğru hareket eder. Hareket ettikçe servis verir. Diğer uca ulaştığında dönüş yolunda hiç servis vermeden hemen diskin başına döner.
- Silindirlere son silindirden ilk silindire doğru bir dairesel liste gibi davranır.

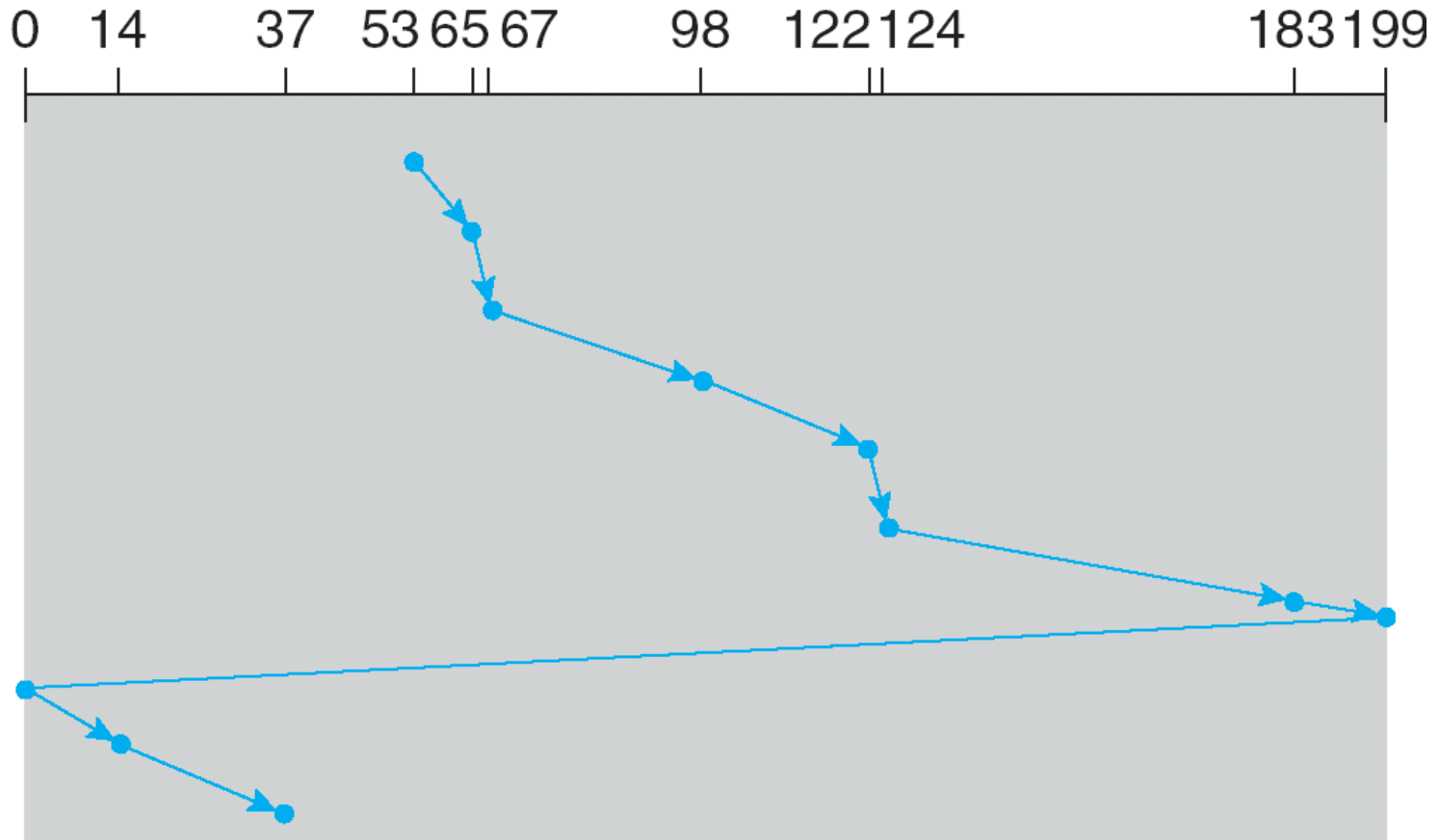




C-SCAN

queue = 98, 183, 37, 122, 14, 124, 65, 67

head starts at 53



Gösterimde 383 toplam kafa hareketi görülmektedir.



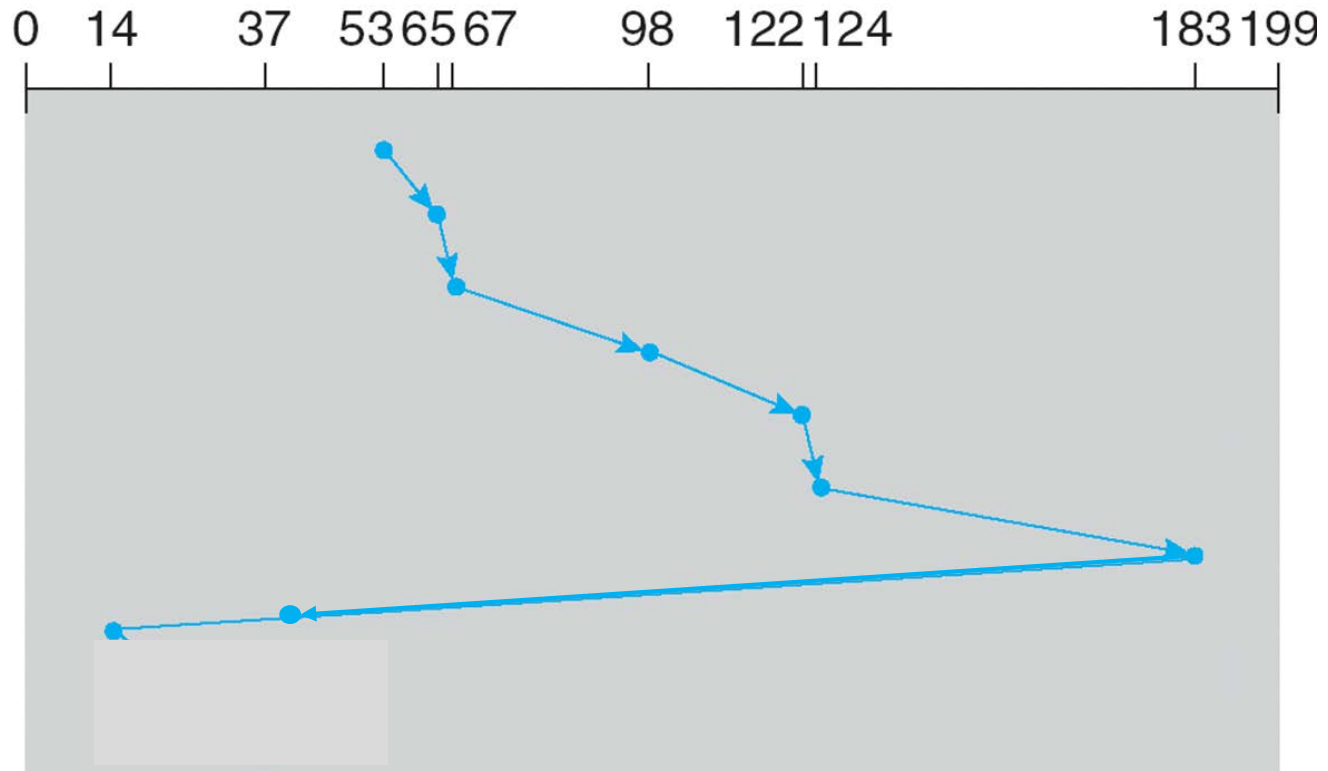


LOOK

SCAN'ın bir versiyonudur.

Kol yalnızca her yöndeki son isteğe kadar gider, sonra hemen geri döner. Önc

queue 98, 183, 37, 122, 14, 124, 65, 67
head starts at 53





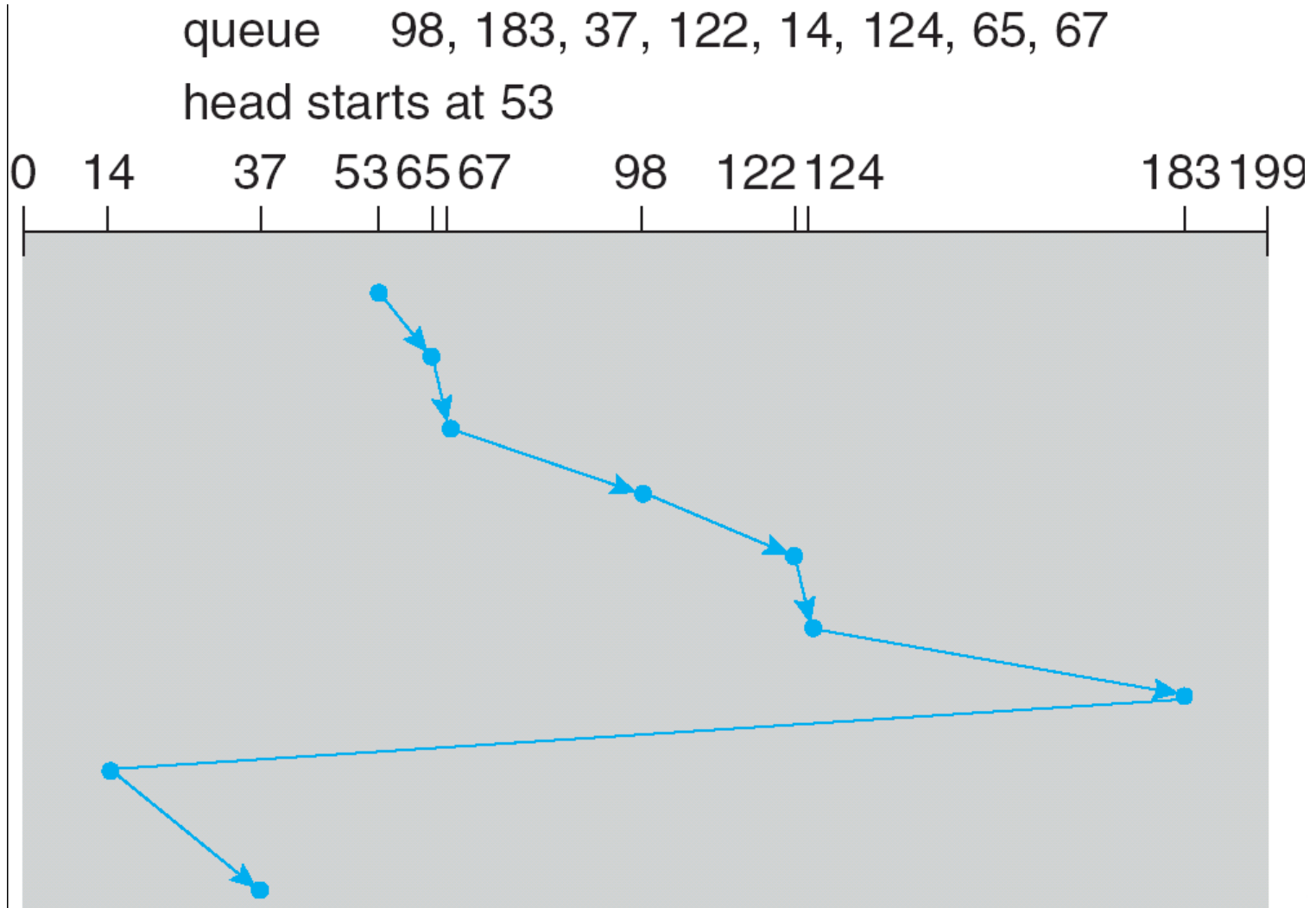
C-LOOK

- C-SCAN'ın bir versiyonudur.
- Kol yalnızca her yöndeki son isteğe kadar gider, sonra hemen geri döner. Önce diskin sonuna kadar gitmesi gerekmez.





C-LOOK



Gösterimde 322 toplam kafa hareketi görülmektedir.





Disk İş Planlaması Algoritması Seçimi

- SSTF yaygındır ve FCFS'ye göre daha yüksek performansa sahiptir
- SCAN ve C-SCAN disk üzerinde ağır yük olan sistemlerde daha iyi çalışır
- Performans, isteklerin sayısı ve tipine göre değişir
- Disk servis istekleri, dosya tahsis metodundan (ardışık veya bağlı liste vb.) etkilenmektedir
 - Ardışık dosya tahsisinde kafa hareketi sayısı daha az olur
 - Bağlı listede ise daha fazla kafa hareketi olur
- Disk iş planlaması algoritması, gerektiğinde farklı bir algoritma kullanılabilmeye izin verecek şekilde işletim sisteminin ayrı bir modülü olarak yazılmalıdır
- SSTF veya LOOK varsayılan algoritma olarak iyi bir seçimdir





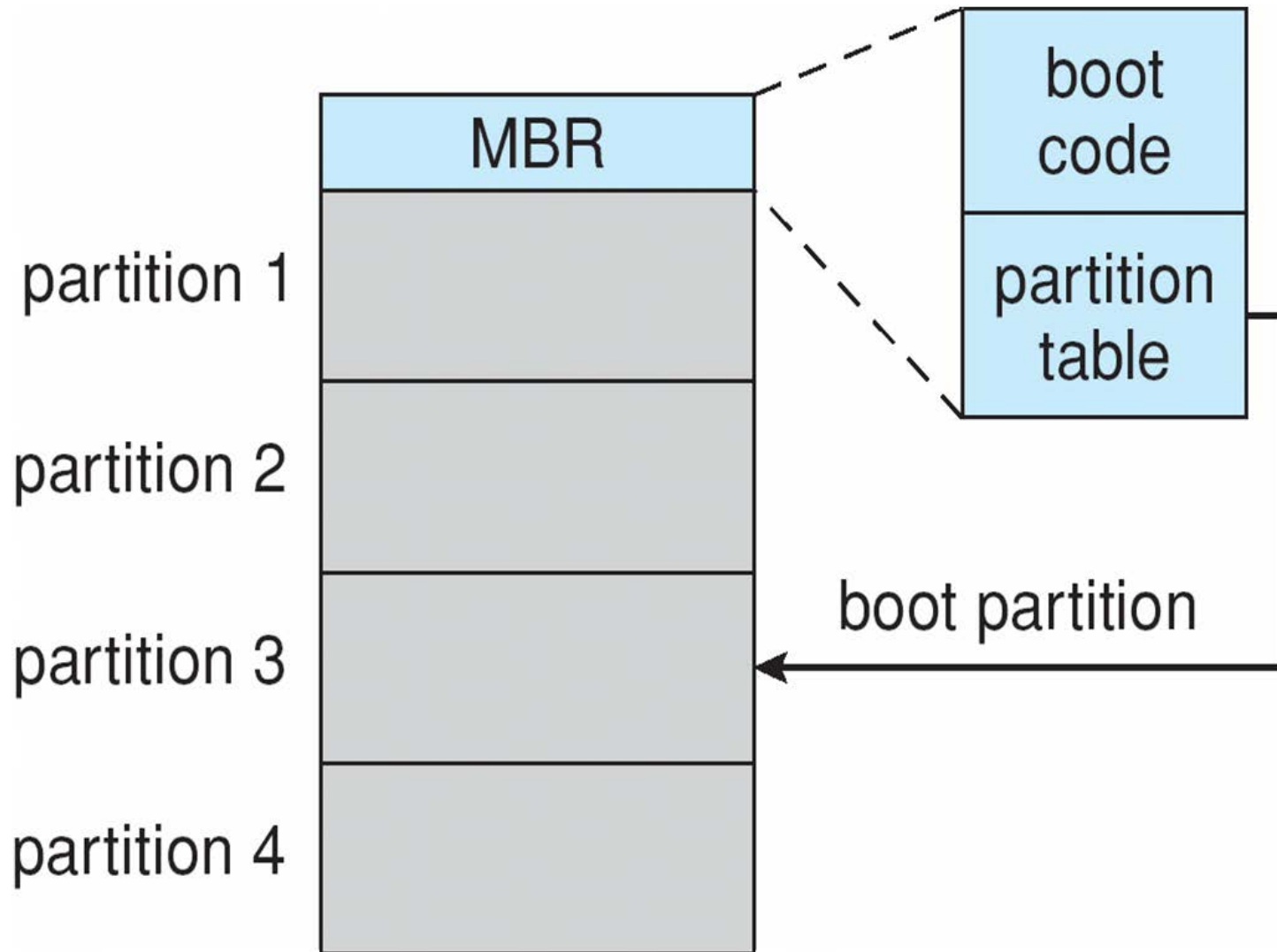
Disk Yönetimi

- **Alt Seviye Biçimlendirme (Low-level formatting)** veya **Fiziksel Biçimlendirme (physical formatting)** — Diski disk kontrolörünün okuyup yazabileceği şekilde sektörlere bölmek
- Bir disk dosyaları tutacak şekilde kullanmak için, işletim sistemi kendi veriyapılarını diske kaydetmelidir.
 - **Bölüm (Partition)** - diskin bir veya daha fazla silindir grubuna bölünmesi
 - **Mantıksal Biçimlendirme (Logical formatting)** veya "bir dosya sistemi yaratmak"
- **Başlatma Bloğu (Boot block)** sistemi başlatır
 - **Önyükleyici (bootstrap)** ROM içerisinde kayıtlıdır
 - **Önyükleyici yükleme (Bootstrap loader)** programı
- **Sektör yedekleme (sector sparing)** gibi metotlar bozuk blokları idare etmek için kullanılır.





Windows 2000'de Diskten Başlatma





Değiş Tokuş Alanı Yönetimi (Swap-Space Management)

- **Değiş Tokuş Alanı (Swap-space)** — Sanal bellek disk alanını ana belleğin bir uzantısı gibi kullanır
- Değiş tokuş alanı normal dosya sisteminden bölünür ya da (daha yaygın olarak) ayrı bir disk bölümü (disk partition) olabilir
- Değiş tokuş alanı Yönetimi
 - 4.3 BSD, süreç başlarken değiş tokuş alanını tahsis eder. Metin segmenti (program) ve veri segmentini tutar.
 - ▶ Çekirdek (Kernel) değiş tokuş alanı kullanımını izlemek için **değiş tokuş haritalarını (swap map)** kullanır.
 - Solaris 2, sanal bellek sayfası ilk yaratıldığında değil, sayfa fiziksel bellekten dışarı çıkartılırken değiş tokuş alanını tahsis eder.





RAID Yapısı

- **RAID** – çoklu disk sürücüler **tekrarlama** (*redundancy*) yoluyla **güvenilirlik** (*reliability*) sağlar
- RAID (Redundant Arrays of Independent Disks) altı farklı seviyeye ayrılmıştır
- Disk güvenliğinin ve performansının arttırmak için ortaya çıkartılmış disk organizasyon biçimleridir.
- Temel fikir birden fazla diski bilgisayara eklemek ve bu disklerin tek bir disk gibi davranmasını sağlamaktır.
- RAID işlemlerini RAID kontrolcü gerçekleştirir.
- Veriler RAID disklerde paylaştırılırlar bu yüzden paralel işlem gerçekleştirmek mümkündür.
- Verilerin diskler üzerinde dağılımlarında farklı şemalar bulunur.





RAID

- Disk kullanım tekniklerindeki gelişmeler ile çoklu disklerin birlikte çalışması sağlanmıştır.
- **Disk şeritleme (*Disk striping*)** bir grup disk tek bir depolama birimi olarak kullanır.
 - **Bit level striping** her byte'taki bit'leri farklı disklere böler
 - **Block level striping** bir dosyanın bloklarını farklı disklere böler
- RAID şemaları performansı ve tekrarlayan veri kaydederek depolama sisteminin güvenilirliğini artırır.
 - **Aynalama (*Mirroring*)** veya **Gölgeleme (*shadowing*)** her diskin kopyasını tutar.
 - **Blok karıştıran parite (*Block interleaved parity*)** daha az tekrarlama kullanır





RAID

- **RAID Level 0:** Sanal disk her birinde k adet sektör bulunan şeritlere (strip) bölünür. $(0, k-1)$ sektörleri Strip0 da, $(k, 2k-1)$ sektörleri Strip1 de,... bulunur.
- Veri striplere bölünerek farklı striplere yazılır.
- Veri aynı anda farklı disklerle yazılır ve okunur. Veri okuma ve yazma hızı yüksektir.
- Disklerden bir tanesi arızalanırsa, veriler kaybolur.





RAID

- **RAID Level 1:** Aynalanmış (Yansımali) diskler.
- Her diskin bir kopya diskli bulunur. Kopya diski birinci diskin verilerini birebir yedekler.
- Yazma işlemi iki diske birden yapılır. Yazma performansı iyi değildir. Okuma performansı iyidir.
- Bir disk arızalandığı zaman diğer disk hemen kullanılır.
- Onarma işleminde arızalı diskin yerine yenisi konulur ve sağlam olan disk birebir bu diske kopyalanır.





RAID Seviyeleri



(a) RAID 0: tekrarlamayan şeritleme



(b) RAID 1: yansımali (mirrored) diskler



(c) RAID 2: bellek tipi hata düzeltici kodlar



(d) RAID 3: bit-karıştırılmış parite



(e) RAID 4: blok-karıştırılmış parite



(f) RAID 5: blok-karıştırılmış dağıtık parite



(g) RAID 6: P + Q tekrarlama

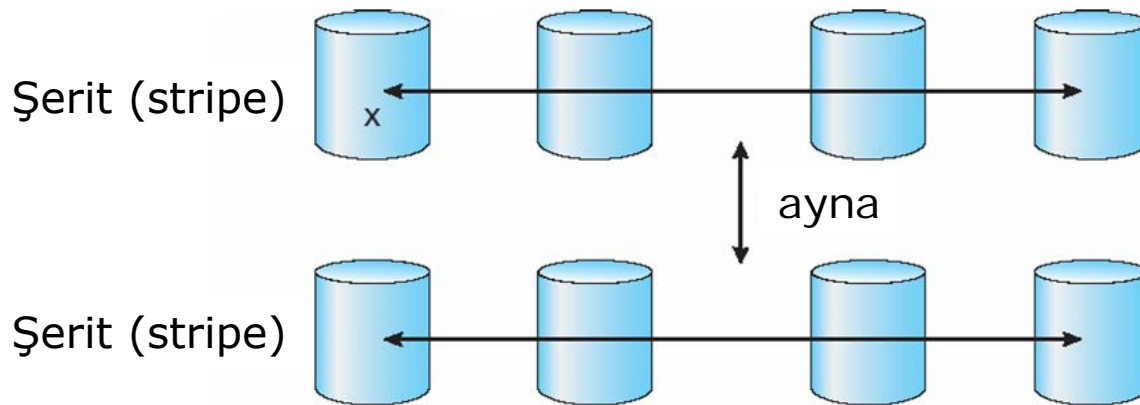
C: Kopya

P: Hata düzeltme bitleri
(Parite)

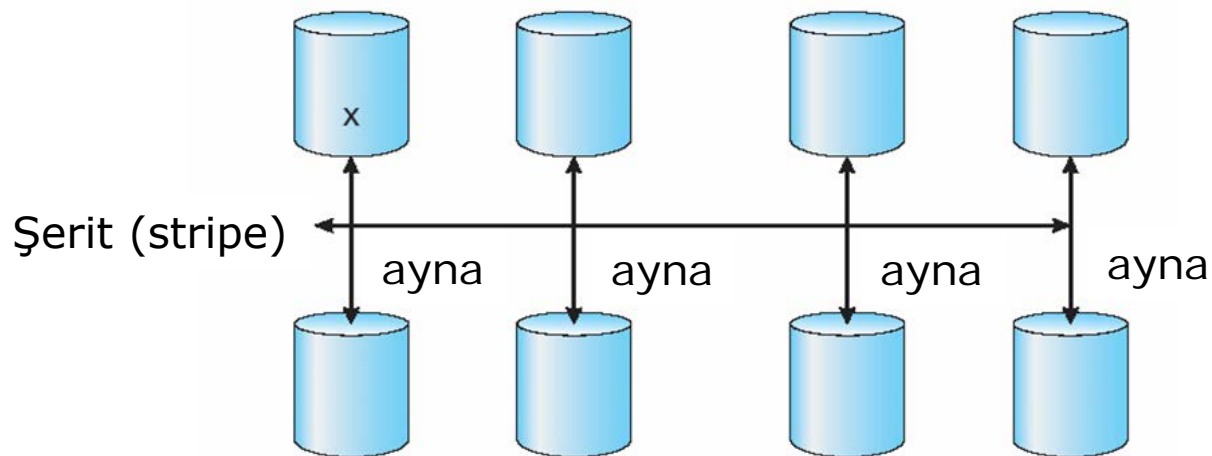




RAID (0 + 1) ve (1 + 0)



a) RAID 0 + 1 tek bir disk hatası ile



b) RAID 1 + 0 tek bir disk hatası ile





Üçüncü Seviye Depolama Aygıtları

- Genellikle üçüncü seviye depolama, taşınabilir ortamlar kullanılarak gerçekleştirilir ve maliyet önemlidir
- Yaygın örnekleri, CD-ROM, DVD-ROM, Flash bellek ve disketlerdir.





Taşınabilir Diskler

- **Floppy disk** — manyetik materyalle kaplanmış ince esnek bir diskir ve plastik bir kap içindedir.
- Çoğu floppy disk 1.44 MB civarında veri saklar ve bozulma riski yüksektir





Taşınabilir Diskler

- **Manyeto-optik diskler** manyetik bir materyalle kaplanmış bir plaka üzerinde veri saklarlar.
 - Bit kaydetmek için zayıf bir manyetik alan lazer ısısı kullanılarak yükseltilir.
 - Lazer ışığı veri okumak için de kullanılır (Kerr effect).

- **Optik diskler** ise manyetizma kullanmaz; lazer ışığı ile değiştirilebilen özel materyaller barındırırlar.





WORM Diskler

- Okuma-yazma disklerindeki veri tekrar tekrar değiştirilebilir.
- WORM (“Write Once, Read Many Times”) (Bir kere yaz, birçok kere oku) disklerine ise sadece bir kere yazılabilir.
- İnce alüminyum film iki cam veya plastik plakanın arasında sıkıştırılmıştır.
- Bir bit yazmak için sürücü lazer ışını kullanarak alüminyum üzerinde küçük bir delik açar; bilgi silinemez veya değiştirilemez.
- Çok dayanıklı ve güvenilirdir.
- ***Salt Okunur (Read Only)*** diskler (CD-ROM ve DVD) fabrikadan önceden kaydedilmiş veri ile gelirler.





Teypler

- Diskle karşılaştırıldığında teyp daha ucuzdur ve daha fazla veri tutar; ancak rastgele erişim daha yavaştır.
- Teypler hızlı rastgele erişime ihtiyaç duyulmayan durumlarda ekonomik bir ortamdır. (Ör., disk verisini yedeklenmesi, büyük miktarda veri tutulması)
- Büyük teyp kurulumları genelde robotik teyp değiştiriciler kullanarak teyplerin teyp sürücüleri arasında ve depolama slotlarında hareketini sağlar.
 - **Yığınlayıcı (*stacker*)** – az miktarda teyp barındıran kütüphane
 - **Silo** – binlerce teyp barındıran kütüphane





Üçüncü Seviye Depolama Aygıtları

İşletim Sistemi Konuları

- Temel İşletim Sistemi işleri fiziksel sürücülerin yönetilmesi ve uygulamalar için bir sanal makine soyutlaması sunmaktır
- Sabit diskler için İşletim Sistemi iki soyutlama sunar:
 - **Ham aygıt (*Raw device*)** – veri blokları dizisi
 - **Dosya Sistemi (*File system*)** – İşletim sistemi farklı uygulamalardan gelen karışık istekleri planlar ve sıralar





Üçüncü Seviye Depolama Aygıtları

Uygulama Arayüzü (Application Interface)

- Birçok İşletim sistemi taşınabilir diskleri sabit disk olarak kabul ederler – yeni bir disk önce biçimlendirilir ve disk üzerinde boş bir dosya sistemi yaratılır.
- Teypler ham depolama ortamı olarak sunulmuşlardır. Uygulama teypdeki bir dosyayı açmaz, tüm teyp sürücüsü ham bir sürücü olarak açılır.





Üçüncü Seviye Depolama Aygıtları

Teyp Sürücüler

- Bir teyp sürücüsündeki temel işlemler disk sürücülerden farklılık gösterir.
- **locate** teybi spesifik bir mantıksal bloğa götürür, bütün ize değil (**seek** 'e denk gelir)
- **read position** işlemi teyp kafasının bulunduğu mantıksal blok numarasını döndürür.
- **space** işlemi harekete izin verir.
- Teyp sürücüler "sadece sonuna eklenebilir" ("append-only") aygıtlardır; teybin ortasındaki bir bloğu güncellemek bu bloğun arkasındaki herşeyi siler.
- Blok bittikten sonra bir EOT (End Of Track (İz Sonu)) işareti konulur.





Dosya Adlandırma (File Naming)

- Taşınabilir ortamlardaki dosyaların adlandırılması bir bilgisayardaki bir taşınabilir ortama yazıp, sonra bu ortamı başka bir bilgisayarda kullanmak istediğimizde sorun olabilir.
- Güncel İşletim Sistemleri genellikle taşınabilir ortamlardaki isim alanı (name space) problemini çözümsüz bırakırlar ve bunun çözümünü uygulamalar ve kullanıcılara bırakırlar.
- Bazı taşınabilir ortamlar (ör., CD'ler) iyi standardize edilmişlerdir, böylece tüm bilgisayarlar bunları aynı şekilde kullanırlar.





Hiyerarşik Depolama Yönetimi (HDY) (Hierarchical Storage Management (HSM))

- Hiyerarşik bir depolama sistemi depolama hiyerarşisini genişleterek üçüncül depolamayı da dahil eder.
- Genellikle bunu dosya sistemini genişleterek gerçekleştirirler.
 - Küçük ve sık kullanılan dosyalar diskte kalır.
 - Büyük, eski, aktif olmayan dosyalar ise arşivlenirler.
- HDYler genellikle süperbilgisayar merkezleri ve diğer büyük kurulumlar gibi çok büyük miktarda veri bulunduran sistemlerde bulunurlar.





Hız

- Üçüncül depolamada iki hız bileşeni **bant genişliği (*bandwidth*)** ve **gecikmedir (*latency*)**.
- Bant genişliği saniye başına byte ile ölçülür.
 - **Sürekli Bant Genişliği (*Sustained bandwidth*)** – büyük bir transfer sırasında ortalama veri oranı; byte sayısı / transfer zamanı.
 - ▶ Sistemin akışının aktığı durumdaki veri oranıdır
 - **Efektif Bant Genişliği (*Effective bandwidth*)** – arama veya konumlama ve kartuş değiştirme dahil tüm G/Ç zamanı boyunca gerçekleşen ortalama hız.
 - ▶ Sürücünün genel veri oranıdır.





Hız

- **Erişim Gecikmesi (*Access latency*)** – verinin yerinin bulunması için gerekli zaman miktarı.
 - Bir disk için erişim zamanı – kolu seçilen silindire götürüp dönüşsel gecikmenin beklenmesi; < 35 milliseconds.
 - Teypdeki erişimde ise teyp kafası istenen bloğa gelinceye kadar teyp tekerleklerinin sarılması; onlarca veya yüzlerce saniye.
 - Genellikle bir teypdeki rastgele erişim bir diskteki rastgele erişime göre bin kere yavaştır.
- Üçüncül depolamanın düşük maliyeti sonucunda birçok ucuz medya tek bir pahalı sürücünün yerine kullanılır.





Güvenilirlik

- Bir sabit disk sürücüsü taşınabilir veya bir teyp sürücüsüne göre daha güvenilirdir.
- Bir optik ortam manyetik veya teybe göre daha güvenilirdir.
- Sabit diskte oluşacak bir kafa çarpması genellikle veriyi yok ederken, optik disk sürücüsü veya teyp sürücüsünde veri hasar görmez.





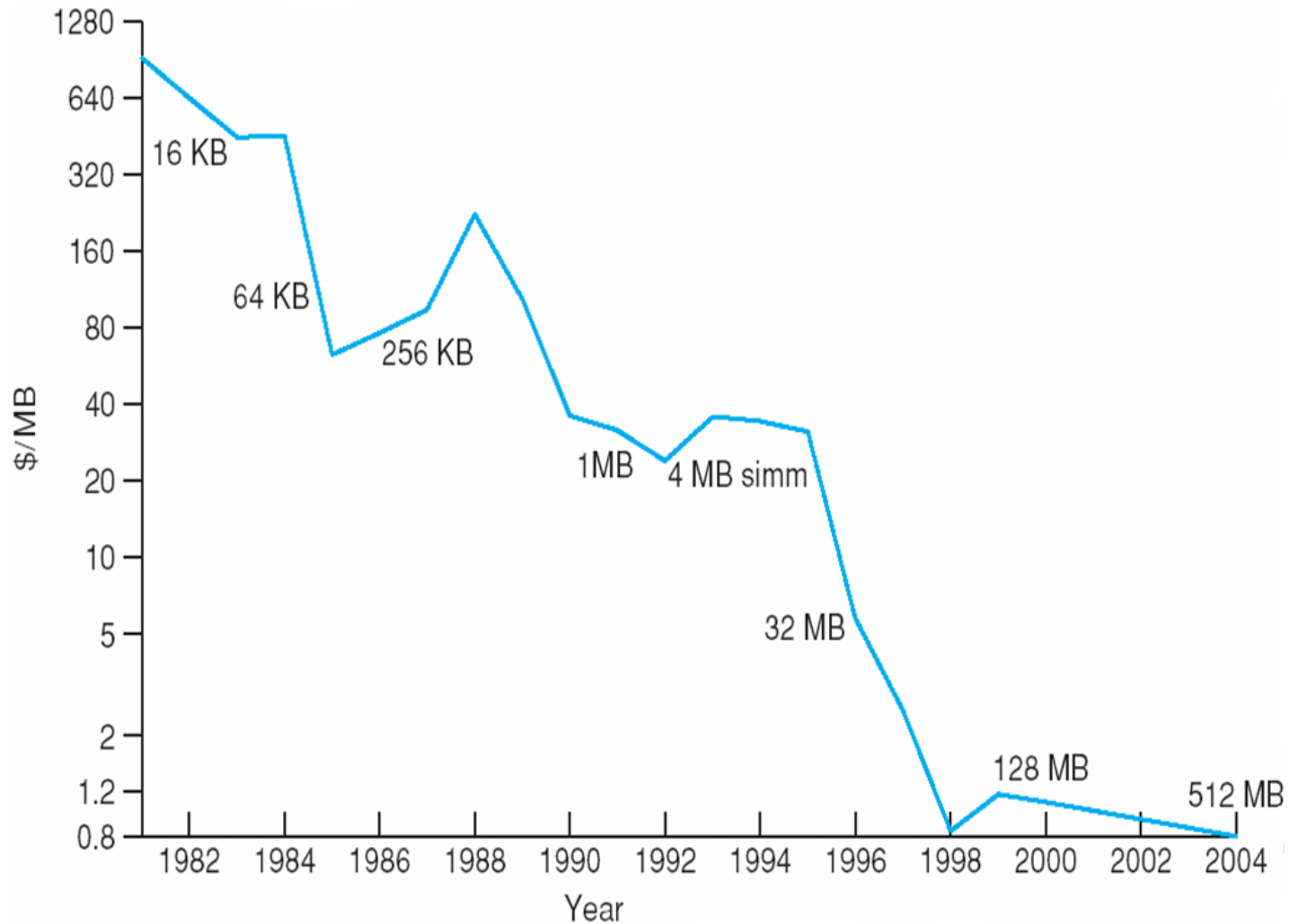
Maliyet

- Anabellek, disk depolamadan çok daha pahalıdır
- Her sürücü için bir teyp kullanılıyorsa sabitdisk depolamanın megabyte başına maliyeti manyetik teyp ile yakındır
- En ucuz teyp sürücüler ile en ucuz disk sürücüler yıllar geçtikçe aynı depolama kapasitesine sahip oldular



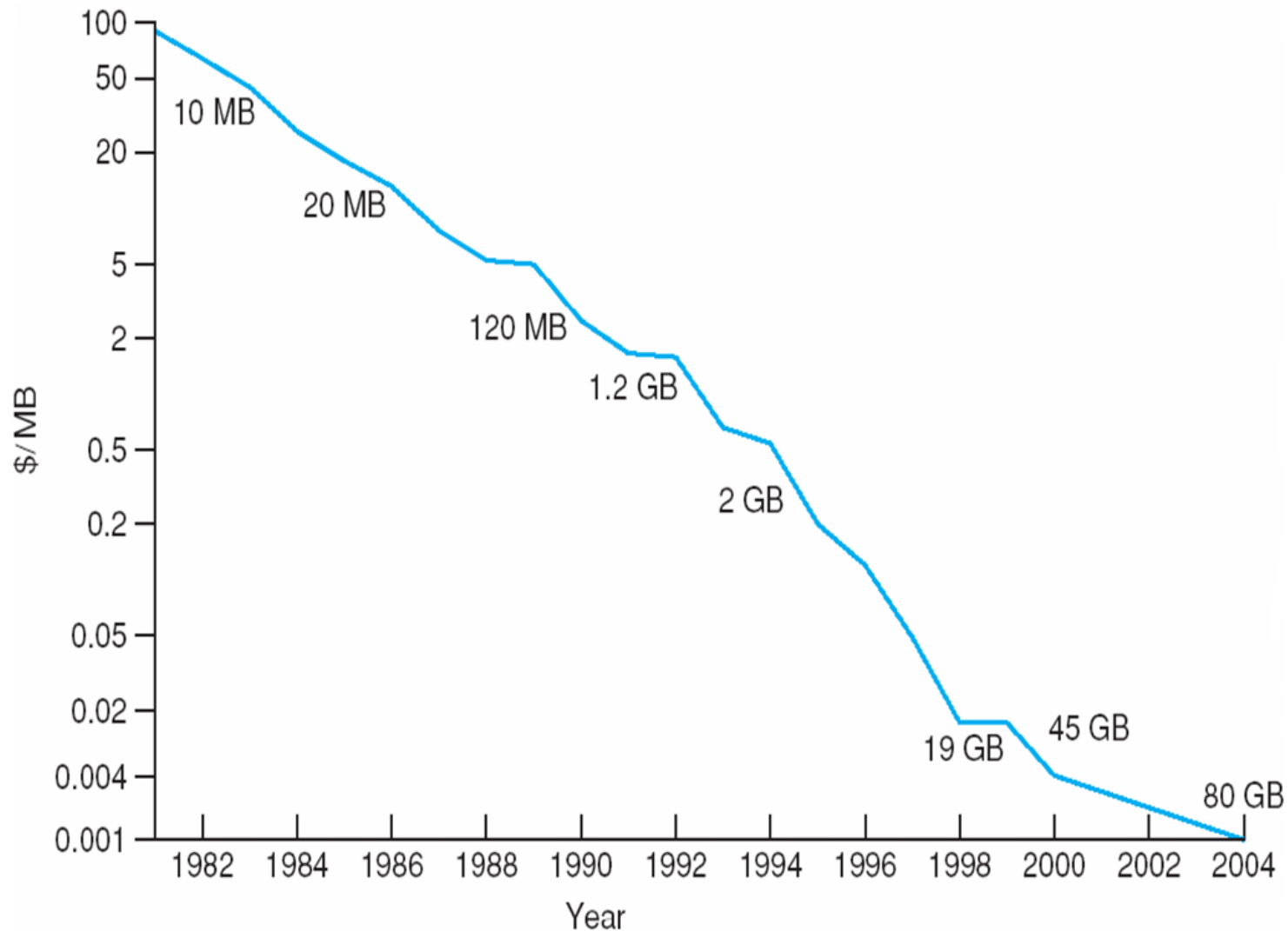


DRAM için Megabyte Başına Fiyat , 1981 - 2004



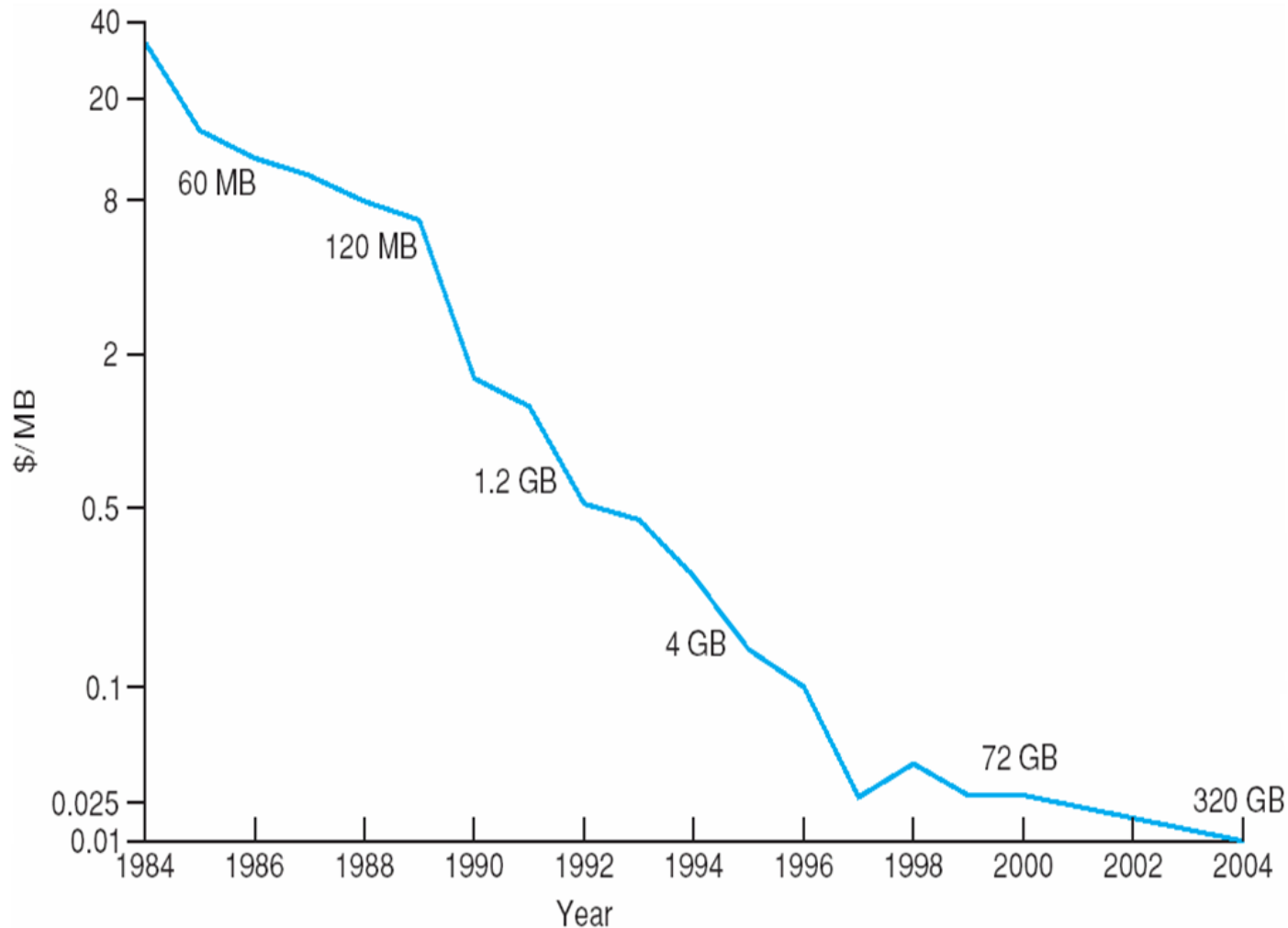


Manyetik Sabitdisk İçin Megabyte Başına Fiyat, 1981 - 2004





Teyp Sürücüsü İçin Megabyte Başına Fiyat, 1984-2000



Bölüm 12 Son

