**4. Запаметяващи устройства**

**Йерархия на паметите – фиг. 4.3.**

**Организация на йерархичната памет**

Регистри и Cache.

управление от компилатора

скоростта се определя от ЦПУ - обикновено в рамките на един такт

Cache

управлява се от MMU (memory management unit) – транслация ВА/ФА

Основна памет.

DRAM чипове ~2^10 Mb.

Управлява се от MMU и ОС

Понякога има йерархична структура от поднива, разширения.

Външни носители.

Управляват се от ОС и потребителя.

Енергонезависими.

Магнитни и оптични дискове (дискови пакети) те са най-голямата по размер

памет за работа в реално време (on-line) - за ОС, системен и потребителски

SW и данни

Третична памет

работят автономно (off-line)

за дублиране, резервиране и архив (backup)

ленти, дискове, наследени устройства

**Магнитни дискови пакети HDD**

Организация – 4.6.1.

синхронни двулицеви магнитни дискове (d = **2.5”, 3.5”** или 5”) и подвижни маг. глави

(от магнитната повърхност се ползва приблизително само външния инч)

по-малките дискове са с по-бързо време за достъп на главите и по-висока скорост на

въртене (заради по-доброто охлаждане)

цилиндри, писти/пътеки (tracks), сектори и празни полета (gaps)

блокове за еднократен достъп (логическо понятие) = 1+ сектора

типични конструктивни параметри (4.6.2.):

7200RPM / 8.3mS – преход между два последователни сектора (**latency**, **rotation time**)

2 диска / 4 глави

214 (=16384) цилиндри

27 (=128) сектора в писта

212 (=4096) байта в сектор

време на главите за позициониране (**seek time**)

старт+стоп 1mS и

преход между два съседни цилиндъра 1μS

**Паралелни HDD интерфейси**

остарели, с наследено приложение

паралелен ATA (Advanced Technology Attachment), IDE (Integrated Drive

Electronics)

къс паралелен плосък кабел

директен паралелен интерфейс към 16-битова магистрала или

посредством драйвер – към 32-битова

28b адрес на сектор т.е. 28220=256Мсектора\*512В=128 GB (Enhanced

IDE)

SCSI (Small Computer System Interface, “Scuzzy”)

универсален паралелен интерфейс ~60 линии (сега със серийна версия

SAS) за свързване на до 16 периферни устройства по специален

команден поротокол

включва арбитриращ прототкол за достъп + TCP/IP

приложение – **RAID сървери** (Redundant Arrays of Independent Disks –

дискови масиви за надеждни репликирани данни или паралелен достъп

с висока производителност

**Серийни HDD интерфейси**

**сериен АТА (SATA)**

сериен буфериран обмен – висока скорост, дистанция и надеждност 3Gb/S

освен физическото ниво дефинира потребителски ориентирани протоколи

поддържа конкурентни транзакции с възстановяване след грешка

**Fiber Channel**

оптичен интерфейс предимно към HDD – SAN (Storage Area Network)

сериен обмен – протокола се поддържа и върху усукана двойка проводници

модели на свързване (4.10)

FC-P2P – за 2 устройства

FC-AL (Arbitrated Loop) – с прекъсване на пръстена или с концентратор (hub);

token-ring протокол, което го различава от P2P на 2 устройства

FC-SF (Switched Fabric) – арбитрищ концентратор за множествен достъп (аналог

на Ethernet)

**SAS (Serial Attached SCSI)**

сериен интерфейс с асоцииран Initiator/Target (HDD/RAID) протокол

без арбитриране – Р2Р

високи скорости без фазово отместване (skewing)

наборът команди поддържа не само HDD/CD/DVD но и скенери, принтери...

по-скъпи устройства от SATA – за сърверни приложения

други стандартни серийни интерфейси като **USB** и **IEEE1394** (FireWire) се свързват към

HDD SAS или SATA посредством буфериращ мост (bridge=layer2switch; repeater/bridge/gateway)

**Ускорен достъп до HDD**

при обработката в ИС част от контекста е във вторичната или дори

третичната памет

скоростта на обработка се опр. не толкова от алгоритъма в

основната памет, а от операциите за пренос на данни в йерархията

т.е. от броя операции за запис и четене на блокове

необх. от ефективна орагнизация на достъпа

размер на блока (ниво ОС/СУБД)

организация на данните – многомерните таблици се записват във

файлове и блокове със последователен достъп

приложенията се организират с оглед минимизиране на операциите за

достъп до вторичната памет – обичайно две фази:

разделна обаработка на подобластите от контекста в основната

памет и запис на резултатите във вторичната памет (обикновено

алокация в нови блокове)

интегриране на резултата за отделните подобласти

възможен многофазен йерархичен вариант

напр. алгоритми за сортиране или търсене

**Размер на блока**

цялата писта на един цилиндър е от порядъка на 512-1024KB и е

възможно 1 сектор ≡ 1 писта, както и блока да се състои от

повече от 1 сектор

плътността на записа нараства с промяна на магнитната глава и

вертикалния запис (Nobel Prize Physics 2007) – 4.13.

фактори за по-голям блок

пре-позиционирането на главите (seek time + rotation time) е 10ки

пъти по-бавно от операцията на обмен (Ч/З)

напр. обмен на блок при BS=16KB е ~0.1mS; пре-позициониране на

главите е средно 5-10mS

фактори за по-малък блок

размера на информация в 1 писта/сектор

блок с повече от 1 сектор не е ефективен по време за достъп

големи части от блоковете може да остават незаети (контекстна

зависимост)

**Методи за ускорен достъп до HDD**

разполагане на свързаните данни в блокове на един и същ цилиндър:

премахва времето за радиално позициониране (seek time) и възможно

редуциране на времето за аксиално позициониране (rotation time a.k.a.

latency)

къс ход (short stroking) – неизползване на пълния капацитет за по-бърз

достъп – 4.14. – в ОС чрез дялове (partitioning)

разпределяне на данните между няколко диска с паралелен достъп

репликиране/дублиране (mirroring) на данните на един и същи диск или на

няколко диска – за защита освен за ускорен достъп

планиране на операциите – ниво ОС или СУБД – определя реда на няколко

конкурентни заявки за достъп до блокове, така че закъсненията между

последователните заявки се минимизират – само при възможност на

пренареждане на последователността

“асансьорен алгоритъм” (elevator algorithm) – заявките се пренареждат

по реда на цилиндрите

предварително зареждане на блокове в ОП – напр. съседни на заявения

**Репликиране на данните**

поддържане на няколко копия на данните в различни

дискове (mirrors)

ускорени операции и отказоустойчивост

при паралелен достъп до различни блокове данни

планирането на операциите може да е динамично и

да се използва диска с най-близко текущо

разположение на главите

**ускорява четенето** на ниво блок ~n пъти (по боря

копия)

**не ускорява записа,** но не го и забавя

**Надеждност на информацията в HDD**

контролни и коригиращи кодове

контролно кодиране с информационен излишък – напр. контролни суми

(checksum)

проверка по четност (parity check) на битовете от сектора

еднобитова сума → 50% вероятност за откриване на грешка

повече контролни битове за “дълги грешки”; напр. 8 контролни бита – по един за

всеки от последователните битове в байтовете; вероятност за откриване = 28/(28-1)

(т.е. 1-1/256)

репликиране на данните – mirroring (обикновено се прилага дублиране)

групова защита в дисков масив за паралелен достъп до различни

блокове “RAID Level 4” – фиг. 4.16

при **запис** на блок в един диск от масива е необходимо преизчисляване на

контролния бит чрез предварително контролно четене на старата стойност и

корекция в контролния диск

освен откриване на грешки – възможност за корекция при повреда на диск

чрез възстановяване на информацията

RAID Level 5 и 6 се отнасят до повреда на два и повече диска

**Логическо разделяне на HDD – дялове**

дялът (partition) е логическа структура от последователни сектори; всеки сектор

принадлежи най-много на един дял

различните ОС се разполагат в различни дялове и достъпът до сетори от друг дял

е невъзможен чрез INT 13h, но е възможен чрез адресиране на логическите

дискове, обхващащи тези сектори

наследено, но незадължително изискване е дяловете обхващат цели цилиндри,

освен първия (най-външен) от който е отделен само нулевия сектор C/H/S=0/0/1 -

резервиран за MBR (Master Boot Record – 512 В сектор в началото на многоделен

HDD)

така първия дял стартира от C/H/S=0/1/1 (вместо от нов цилиндър 1/0/1)

MBD

съдържа таблицата с дяловете на диска

съдържа изпълним код при начално зареждане на ОС от диск

32-битов идентификатор на HDD

дялов сектор (partition sector) – описва всеки дял: начален и краен сектор, тип

(първичен, вторичен, с ОС инсталация – bootable) и др.

ОС-сектор (boot sector) – в началото на дял, който се използва за зареждане на

ОС

съдържа програма, която стартира зареждането на ядрото на съответната ОС

в MS-OS само primary partition може да е bootable – т.е. съдържа ОС-сектор; останалата

част от диска се обобсобява като един вторичен (secondary/extended) дял, който може

да бъде разделен на няколко логически дискове (drive, volume)

**Оптични носители на данни**

лазерен запис и светлинно четене (фотодиод) върху алуминиев

диск – първоначално аналогов звуков формат, впоследствие

универсален цифров запис на данни, структурирани във файлове

пистата за **едностранен сериен** запис и четене на диска

съдържа вдлъбнатини от лазерно облъчване с размери (данни за

CD-стандарт!):

дълбочина \* ширина \* дължина : ~100 \* 500 \* 850 nm

дистанция между съседни писти 1500 nm

дължина λ на вълната на лазерната емисия –780 nm (IR-сектор)

дълбочината на вдлъбнатините е ~0.2\*λ

запис на данни - последователно:

CIRC (Cross-interleaved Reed-Solomon coding) – за откриване и

корекция на грешка (1 конторлен байт на 3 байта данни)

8/14 модулация

NRZI-кодиране (non-return to zero, inverted) – 2 нива

възстановяване при четене – в обратен ред

**Стандартни цифрови оптични носители**

CD-ROM (780 nm лазер – IR) – само четене

CD-R (от recordable) – еднократен запис

CD-RW – многократен запис

DVD-ROM (650 nm лазер – червен) – само четене

DVD-R и DVD+R – еднократен запис

DVD-RAM, DVD-RW и DVD+RW – многократен запис

BD, BD-R, BD-RE (Blue-ray Disc – 405 nm лазер – син) –

четене, запис, презапис

HD DVD [HD DVD-ROM, HD DVD-R, HD DVD-RW (High-

Definition DVD – 405 nm лазер )

**Оптична глава**

състои се от:

лазерен източник

в режим четене ~5mW

[опция] в записващ режим 100÷225mW

записващият лазер е в импулсен режим и енергията му варира

приблизително двойно по време на запис

високите скорости на запис и форматите с по-висока плътност

изискват по-високоенергийни лазери поради краткото време на

въздействие

фокусираща леща

фотодиод

електромагнитен серво-контрол на дистанцията от

повърхността

електродвигател за радиалното отместване

**Оптичен запис**

на -ROM (т.е. фабричен запис) – механично щамповане на пистата върху гладка повърхност –

едновременно се записва цял диск

на ±R (т.е. еднократен запис) – лазерно (т.е. в случая топлинно) необратимо последователно

прогаряне на непрозрачни области в прозрачен органичен слой върху рефлектиращата

Al-повърхност

на ±RW, -RAM (т.е. с презапис) – прогарянето е в прозрачна кристална сплав, която минава

обратимо в аморфна непрозрачна фаза

еднослоен запис (масовият стандарт) – спираловидна прозрачна писта се гравивира

фабрично в поликарбонатното покрите над рефлектиращата повърхност за запис; чрез

нея се коригира и синхронизира треакторията на главата в режим запис

двустранен запис – възможен, но непрактичен поради необходимост от обръщане на

диска в устройството при четене

двуслоен запис – на различна дълбочина чрез промяна на фокусното разстояние на

лещата

горната рефлекторна повърхност е полупрозрачна

и двата слоя имат своя спираловидна водеща писта над съответната рефлекторна

повърхност за запис; за горния слой пистата стартира отвътре навън, за долния –

отвън навътре

**Скорости на оптичен запис**

маркировка за скоростта на запис е спрямо стандартната скорост на

четене от CD

12x/10x/32x означава

запис CD-R 12\* (1.76 MB/s)

запис CD-RW 10\* (1.46 MB/s)

четене от СD 32\* (4.69 MB/s)

оптичинят запис на данни има характериситката на изосинхронен

режим поради постоянната скорост и посока на въртене - без

връщане назад

за целта денните за запис се подготвят в буфери в основната памет

по-нископроизводителни компютри могат да изпаднат в състояние на

празен записващ буфер

загуба на носител или

рестартиране на целия запис

**Оптично четене**

електромагнитния сервомеханизъм поддържа

съответната фокусна дистанция от рефлекторната

повърхност

радиално-въздействащия елдвигател поддържа

главата върху записаната спирална писта

четене

на механични вдлъбнатини от ROМ: базира се на гасящата

интерференция на падащия и отразения от вдлъбнатината

лъч (дълбочината на вдлъбнатината е кратна на дължината

на вълната)

при прогорените непрозрачни петна в R/RAM падащият лъч

изобщо не се отразява

наличието или отсъствието на отразен лъч се детектира от

четящият фотодиод

**Данни върху CD-ROM/RW**

една спирална писта (отвтре навън - така се допускат различни диаметри и форми на CD) с ширина

500 nm и междупистова дистанция 1.5 um

кодирането поддържа EFM (eight-fourteen modulation, конверсия от 8- към 14-битови думи), в която

се гарантират “1” при дълги последователности от “0” за да не може четящият лазер да загуби

писата

защита: коригиращ код срещу еднократна (1 бит) грешка

[при audio CD]:

допълнителна информация за абсолютното разположение на озаглавени части от записа, както и текущото

положение на главата, се добавят в кода (sub-coding), което позоволява бърз преход по заглавия

срещу групова грешка (напр. одраскване) се прилага размесване на данните (interleaving) - непоследователен

запис, т.е. данните, прочетени при последователни обороти, се преподреждат - загубите се отразяват на QoS

CD-R[OM] и CD-RW

поддържат по-сложни коригиращи кодове за възстановяване на грешка 3. ниво CIRC (Cross-interleaved Reed-

Solomon coding) – за откриване и корекция на грешка (1 конторлен байт на 3 байта данни)

поддържат файлова система и планиране на файловете върху пистата, които обаче са неподходящи за

презапис на малки файлове (при -RW)

формат на данните:

кадри (frames) по 24 В

сектори от 94 кадъра т.е. 2352 В, от които

Mode-1 (за данни) 2048 В данни (останалите за коригиращ код)

Mode-2 (аудио-видео файлове) 2336 В данни (къса корекция)

333000 сектора (или блока) в единствената писта; Mode-1 CD-ROM = 682 MВ

**Терабайтови оптични носители**