



Univerzita Hradec Králové
Fakulta informatiky a managementu



evropský
sociální
fond v ČR



EVROPSKÁ UNIE



MINISTERSTVO ŠKOLSTVÍ,
MLÁDEŽE A TĚLOVÝCHOVY



OP Vzdělávání
pro konkurenceschopnost

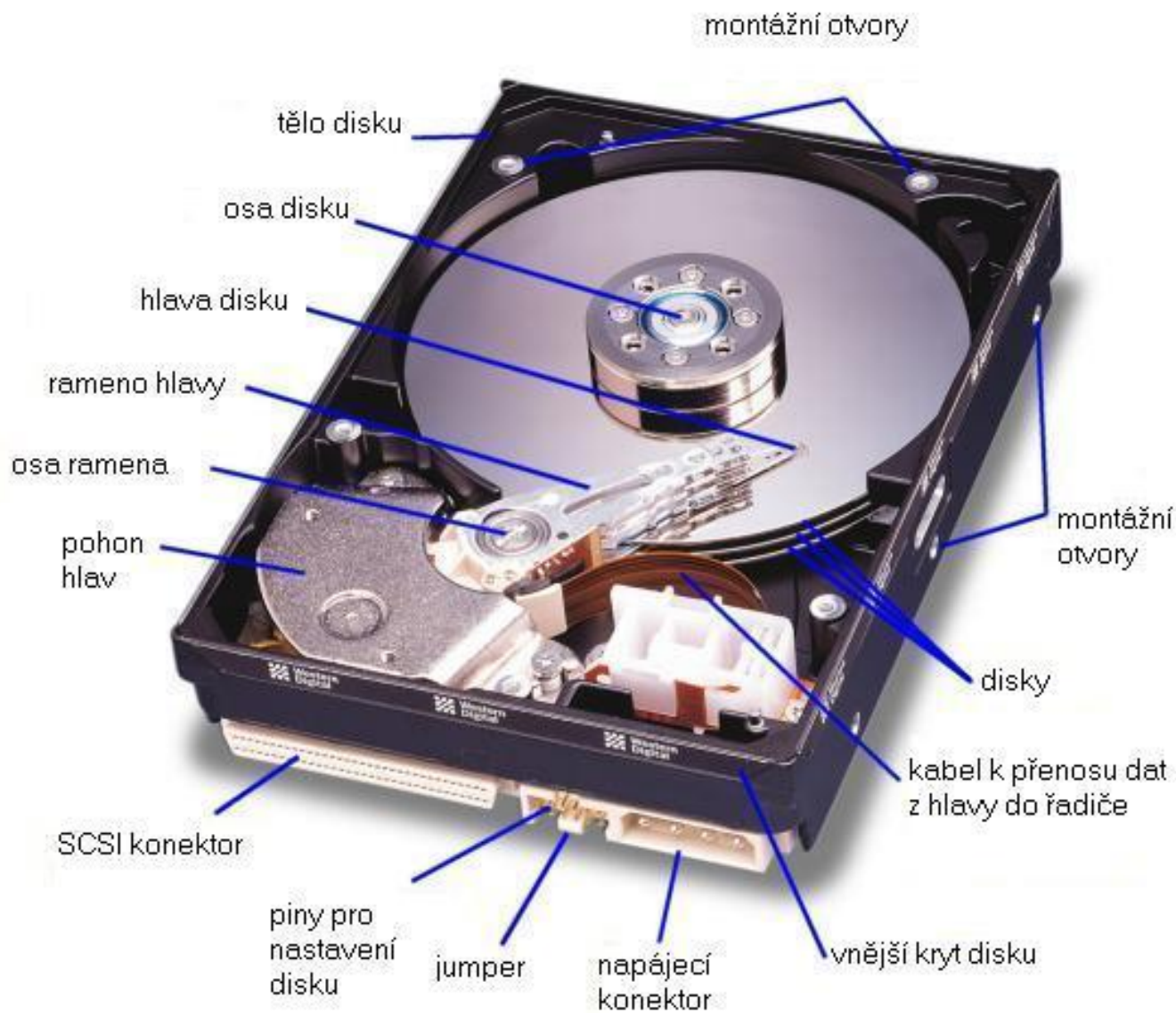
INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Pevný disk

Fyzická struktura

Josef Horálek
upravil Peter Mikulecký

- = Pevný disk je paměť pracující na magnetickém principu
- = Skládá se z několika částí:
 - = médium, na němž jsou uložena data ($\varnothing 3,5''$ a $2,5''$)
 - = magnetické hlavy pro zápis a čtení
 - = mechanika pohybující hlavami
 - = motorek točící diskem
 - = elektronika disku řídící práci disku
 - = desku rozhraní, zajišťující připojení HDD z základní desce
- = Dnes ho lze realizovat i jako SSD (Solid State Disc) bez mechanických částí



- = Datové médium HDD je složeno
 - = z tuhých kotoučů – ploten, umístěných v několika patrech nad sebou
 - = data se zapisují do magnetické vrstvy, která je nanesena na každý kotouč
 - = s tímto povrchem pracují magnetické čtecí/zapisovací hlavy, které se vznášejí nad povrchem (μm)
- = Při vypnutí disku zajistí mechanika magnetických hlav jejich „přistání“ do vyhrazené parkovací oblasti

= Pro orientaci operačního systému na rozsáhlém prostoru disku si OS rozdělí jeho plochu:

- = na stopy (soustředné kružnice)
- = každou stopu na sektory, do kterých zapisuje data.

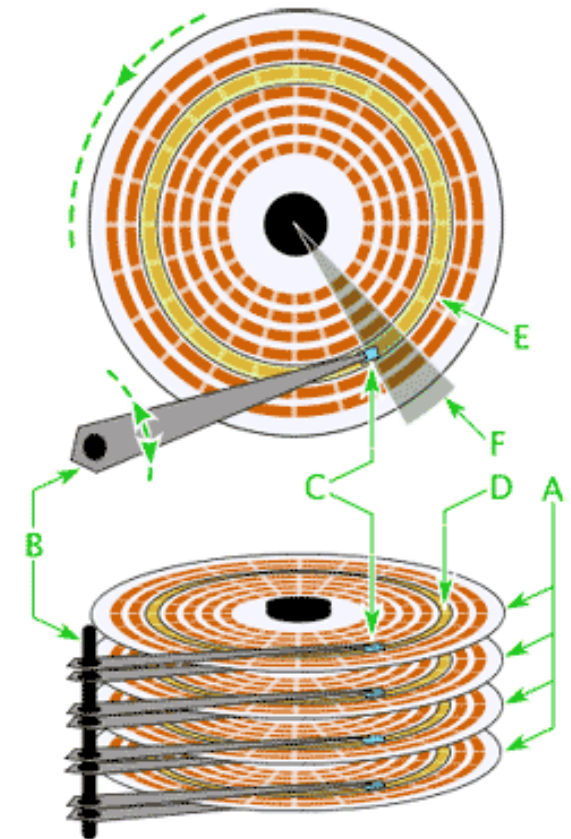
A – plotna (kotouč s magnetickou vrstvou)

B – otočné rameno nesoucí všechny hlavy

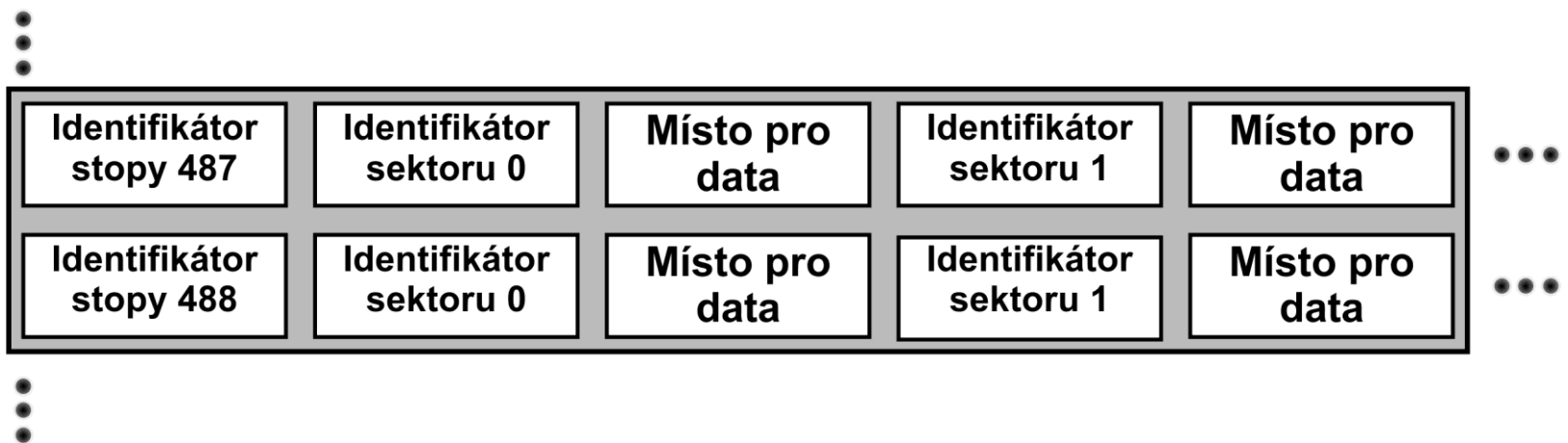
C – čtecí a zapisovací hlava (head)

D,E – cylindry (stopa, přechází všemi plotnami)

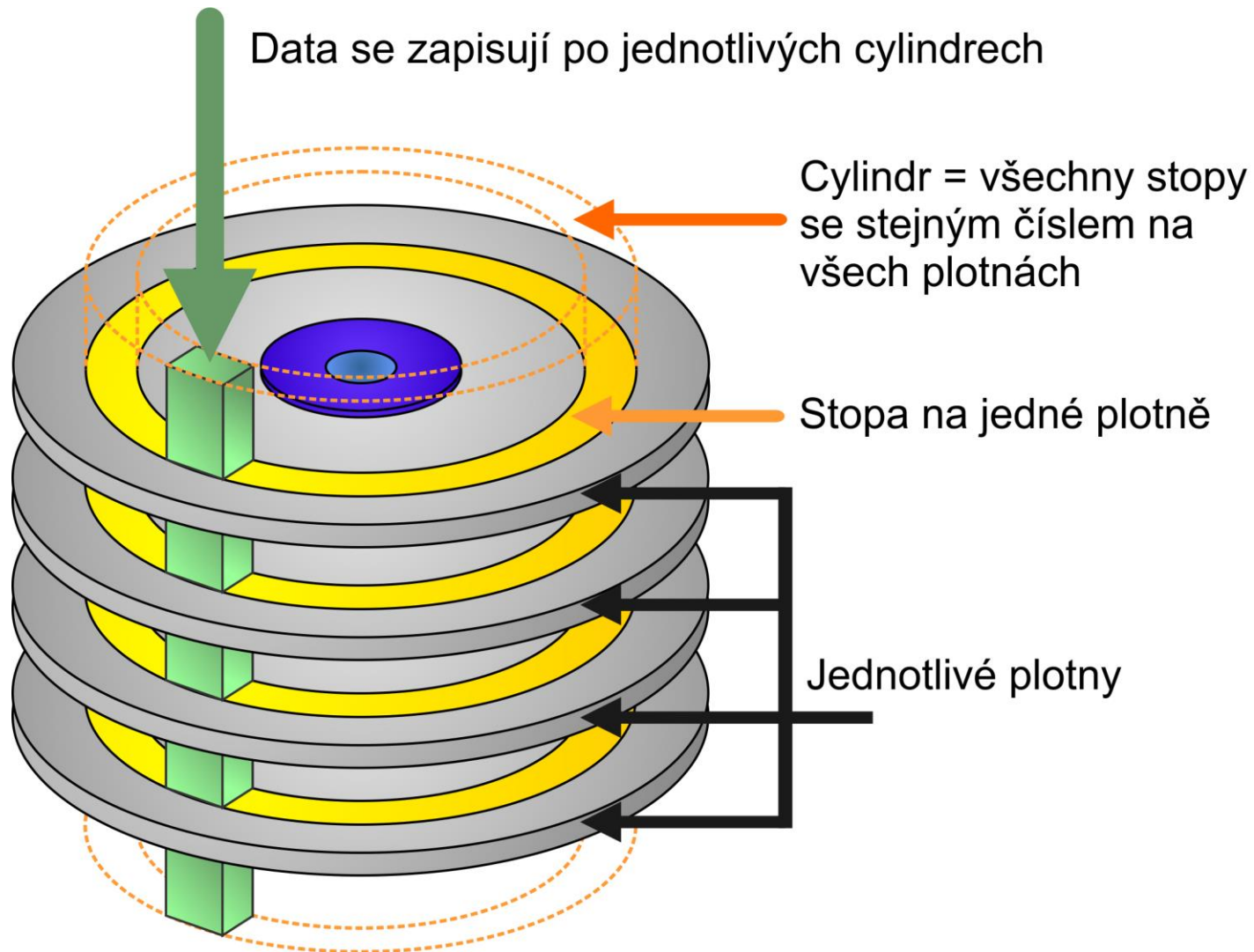
F – sektor (úhlový výsek se stopami)



- = Řadič si musí rozdělit disk na stopy a sektory, které si očísluje
- = fyzické formátování (low format)
- = na začátek každé stopy a sektoru umístí řadič magnetickou značku (identifikátor)
- = tento druh formátování provádí výrobce.



- = Nad každým povrchem „létá“ jedna hlava
 - = zapisuje/čte data
 - = má-li pevný disk 5 kotoučů, může mít až 10 hlav (každý kotouč má 2 povrchy)
- = Všechny hlavy jsou umístěny na společném rameni
 - = pokud řadič posune hlavu číslo 3 (patřící třetímu povrchu) nad stopu 134, posunou se i hlavy nad ostatními kotouči nad stopu 134 svého povrchu
 - = díky společnému rameni se tedy hlavy vždy vznášejí nad stejnou stopou všech povrchů
 - = stejným stopám na různých površích se říká cylindr, řidčeji válec



- = Vyjadřuje rychlost, s níž disk vyhledává data
- = je součtem dvou časů: doby vystavení + doby čekání. Její hodnota se pohybuje pod 10 ms

- = Je čas nutný k pohybu hlav nad určitou stopu
 - = hlavy většinou „přelétávají“ pouze několik stop, a tak je doba vystavení definována jako jedna třetina času potřebného pro pohyb přes celý disk
 - = u moderních disků se pohybuje mezi 2 a 4 ms
 - = výrobci pevných disků se přirozeně snaží dobu vystavení minimalizovat - proto zápis a následné čtení probíhají po cylindrech, nikoli po stopách

- = Pokud hlava „doletí“ nad správnou stopu (je vystavena), nemůže ještě začít se čtením
 - = musí totiž počkat, až se pod ni dotočí ten sektor, v němž se má začít se čtením dat
 - = doba čekání záleží na náhodě, ale jako technická hodnota se uvažuje jedna polovina otáčky disku
 - = dobu čekání chtějí výrobci také snižovat - cesta k tomu je zvýšit otáčky disku

= Metoda pro zkrácení doby čekání

- = při čtení se přečtou data z jednoho sektoru, musí se odeslat přes řadič a BIOS operačnímu systému, který je dále předá aplikačnímu programu
- = ten informace zpracuje a požádá operační systém o nové údaje
- = OS se obrátí na BIOS a řadič, který zorganizuje načtení dalšího sektoru
- = mezitím se však disk pod hlavou pootočí - nestihne začátek následujícího sektoru
- = musí počkat (téměř celou otáčku), až se pod ni sektor opět dostane
- = proto bylo zavedeno prokládání, které ukládá data přes sektory.

Fyzický sektor	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Logický sektor	0	9	1	10	2	11	3	12	4	13	5	14	6	15	7	16	8

- = Stejně jako mikroprocesory i pevné disky používají vyrovnávací paměť
 - = do ní se načítají data z disku a odtud se pak přenášejí na sběrnici
 - = cache podstatně zrychluje práci
 - = dnes o velikosti 2 MB až 64 MB

- = Je nejdůležitějším kritériem disku
 - = první počítače IBM-XT vůbec pevné disky neměly
 - = postupně se začaly objevovat disky s kapacitou 10 MB
 - = horní hranice kapacity dnešních disků je asi 20000 GB, tj. 20 TB (např. Seagate Iron Wolf Pro), ale zítra to už nemusí být pravda
 - = pokud budete chtít připojit velký disk do staršího PC, ujistěte se, že s ním bude umět pracovat BIOS vaší desky (horní kapacitu disku najdete v manuálu). Kapacity disků rostou velmi rychle a starší základní desky nemusí zvládat práci s velkými disky!

- = Každý bit je představován miniaturním **dipólem** zapsaným do magnetického povrchu disku
- = úkolem je miniaturizovat dipóly
- = vytvářet stále jemnější magnetické struktury s možností vyšší hustoty zápisu dat
 - = dříve používaná technologie, kdy se na povrch kotoučů nanášela vrstva oxidů, byla nahrazena vrstvou tenkého filmu
 - = dokonalejší povrch filmu umožňuje menší „výšku letu“ hlavy nad diskem, což znamená potřebu menšího magnetického pole, ta dovoluje použití menších dipólů a větší hustotu stop

- = Data se na magnetická média ukládají pomocí změn magnetického toku.
- = Změna může nastat z kladného toku na záporný, nebo naopak, ze záporného na kladný.
- = Každá takováto změna se při čtení projeví jako **impulz**. K reprezentaci dat na magnetickém médiu se tedy používá přítomnost nebo nepřítomnost impulsu (mezera).
- = Příklad (P značí impuls, N mezeru):

Posloupnost bitů **10101110** se zakóduje jako

PNPNPPPN

- = Teoreticky by bylo možné data zaznamenávat tak, že bit 1 by byl reprezentován impulzem, a 0 mezerou.
- = **To by však nepracovalo!**
- = V případě, že by následovala delší posloupnost nul, která by byla zaznamenána jako dlouhá posloupnost mezer bez jakýchkoliv impulzů, došlo by ke ztrátě synchronizace řadiče pevného disku a nedalo by se přesně určit, kolik mezer (tedy nul) bylo přečteno.
- = Impulzy pomáhají synchronizovat čtená data a řadič disku.
- = Nesmí tedy nikdy dojít k dlouhé posloupnosti mezer. Je nutné zvolit vhodný kompromis mezi ztrátou synchronizace a zbytečně velkým počtem impulzů.

- = Při čtení dipóly rotují pod magnetickou hlavou a vyvolávají v ní elektrické napětí
 - = pokud po sobě následuje několik stejných bitů, například 10000111, stojí řadič (který napětí z hlavy „odebírá“) před problémem, jak od sebe stejné bity oddělit (musí vědět, kolik 0 jde za sebou)
- = Vyvinuty algoritmy úspornějšího zápisu na disk:
 - = MFM (Modified Frequency Modulation)
 - = RLL (Run Length Limited)
 - = PRML (Partial Response Maximum Likelihood)

= V případě FM (Frequency Modulation) se jednotlivé bity kódují následovně:

Bit	Zakódování
0	PN
1	PP

= Takže 101101101 by se kódovalo jako
PPNPPPPNPPPPNPP

= Tedy jednička znamená dva impulsy, nula impuls následovaný mezerou. Zde je zaručené, že nikdy nenastane příliš dlouhá posloupnost mezer, je zde ale vysoký počet impulzů, což omezuje kapacitu média.

= Dnes už zastaralý způsob, nepoužívá se.

= MFM

= Je snaha zmenšit počet impulzů, bity jsou kódovány takhle:

Bit	Zakódování
0	PN jestliže je v řetězci 00 NN jestliže je v řetězci 10
1	NP

= Příklad: 101100 kódování MFM: NPNNNPNPNPN (4 impulsy)

kódování FM: PPPNPPPPPNPN (9 impulsů)

= Zde celkový počet po sobě jdoucích mezer nebude nikdy příliš vysoký, mohou po sobě následovat nejvýše 3 mezery.

= MFM modulace je asi o20% úspornější, než FM.

= U pevných disků se už nepoužívá, pouze u pružných disků (disket), což je už dnes rovněž zastaralé.

= RLL

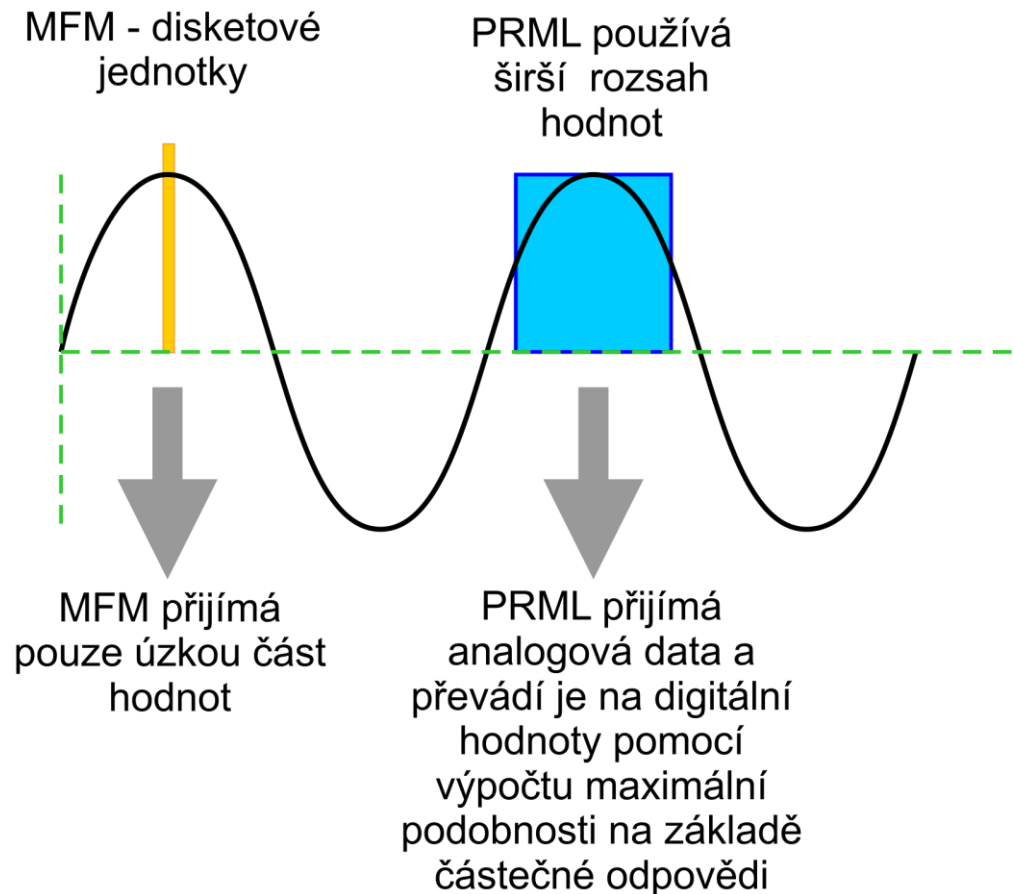
- = často používaná metoda
- = řadič si přepočítá ukládanou posloupnost na novou kombinaci impulsů a mezer dle schématu:

Vzorek	RLL	Počet impulsů	MFM	Počet impulsů
00	PNNN	1	PNPN	2
01	NPNN	1	PNNP	2
100	NNPNNN	1	NPNNPN	2
101	PNNPNN	2	NPNNNP	2
1100	NNNNPNNN	1	NPNPNNPN	3
1101	NNPNNPNN	2	NPNPNNNP	3
111	NNNPNN	1	NPNPNP	3

- = Kódování je asi o 50% úspornější, než u MFM
- = Použito u starších HDD

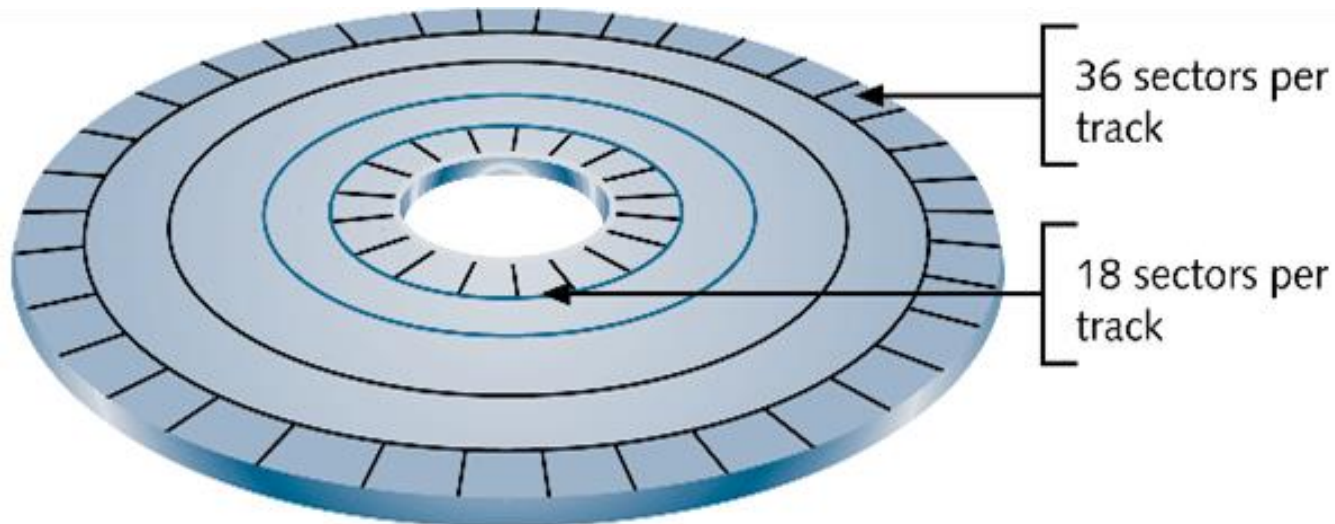
= PRML

- = přináší další zvýšení hustoty ukládaných dat
- = čtené impulsy se zpracovávají digitálním signálovým procesorem (DSP)
- = ten přesně ví, jak má vypadat sled signálů vyvolaný hustě ležícími dipóly, dokáže dopočítat i chybějící údaj
- = rozezná tedy více dipólů na malé ploše, což vede ke zvýšení kapacity disku



- = Díky geometrii disku jsou vnější stopy delší než vnitřní
 - = sektor stopy 0 (první vnější stopa) je delší než stejný sektor poslední stopy, přesto oba sektory nesou stejné množství dat
- = Dipóly jsou malými magnety
 - = mají své severní a jižní póly
 - = stejné póly se odpuzují, opačné přitahují
 - = na vnitřních stopách, kde jsou dipóly blízko u sebe
 - = hrozí nebezpečí, že při jisté kombinaci kladných a záporných impulsů, např. 1100..., mohou v důsledku přitažlivých a odpudivých sil vzniknout bity do sebe, čímž se informace naruší a data budou nečitelná
- = Obrana je prekompenzace
 - = řadič počítá s pohybem dipólů a posouvá zapisované impulsy proti směru předpokládaných přitažlivých sil
 - = bity jsou úmyslně ukládány na geometricky špatné místo, ale působením vzájemných magnetických sil se srovnají
 - = pro prekompenzaci používá zkratka CPZ

- = Při ZBR je plocha povrchu disku rozdělena na zóny
 - = každá zóna má jiný počet sektorů
 - = vnitřní (v níž jsou stopy nejkratší) 35, druhá 36 a poslední, úplně vnější (a nejdelší), 54
 - = optimalizuje se tak počet sektorů - v dlouhých stopách jich je více, v kratších méně
 - = ve srovnání s CPZ zvyšuje ZBR kapacitu
 - = klade vyšší nároky na elektroniku řadiče



- = Spolehlivost je ovlivněna mnoha faktory, z praktického hlediska nás zajímají dva parametry disku:
 - = střední doba mezi chybami – MTBF
 - = Podpora S.M.A.R.T.

- = Střední doba mezi chybami - MTBF (Mean Time Between Failures)
 - = snaží se vystihnout poruchovost disku
 - = je výsledkem simulovaného umělého stárnutí a statistických pravděpodobnostních výpočtů
 - = výsledné hodnoty vycházejí ve statisících hodin.
- = MTBF je jistě indikátorem spolehlivosti

= S.M.A.R.T. (Self-Monitoring Analysis and Reporting Technology)

- = je technologie umožňující předvídat možné selhání pevného disku.
- = založena na sledování přesně definovaných vlastností disků
- = každá vlastnost má přesně definovaný rozsah hodnot
- = pokud se některá z hodnot dostane mimo vymezené hranice, je indikována možnost vzniku chyby
 - = řadič disku, který vlastnosti disku sleduje, uloží zprávu o pravděpodobném vzniku chyby do paměti EEPROM
 - = odtud údaje přečte a vyhodnotí speciální software

- = Ke správné funkci SMART potřebujete:
 - = disk i řadič, které „umí“ SMART
 - = BIOS podporující SMART
 - = speciální software pro zpracování varovného signálu předpovídajícího blížící se chybu disku

- = Raw Read Error Rate
 - = počet chybných čtení dat z plotny
- = Spin Up Time
 - = čas potřebný k roztočení ploten
- = Start/Stop Count
 - = počet start/stop cyklů disku
- = Power On Hours Count
 - = počet odpracovaných provozních hodin
- = Pro správnou funkci S.M.A.R.T. je nutné ho v SETUP aktivovat
 - = Sledování hodnot SMART provádí každý BIOS, při startu počítače načte aktuální stav hodnot SMART
 - = dosáhne-li některá ze sledovaných hodnot kritického stavu je inicializace disku pozastavena, dokud ji uživatel nepovolí

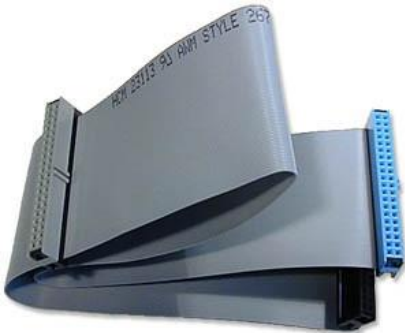
- = Jsou řídicím centrem diskové jednotky
- = jejich úkoly lze shrnout do následujících bodů:
 - = zodpovídají za správné vystavení hlav
 - = při čtení musí co nejrychleji poslat hlavu nad to místo disku, kde jsou uložena hledaná data
 - = organizují vlastní zápis a čtení dat prostřednictvím kódování (při zápisu) nebo dekódování (čtení)
 - = ve spolupráci se sběrnici zajišťují přenos dat mezi diskem a mikroprocesorem
 - = typ řadiče je jedním z nejdůležitějších ukazatelů kvality disku, obvykle se celý pevný disk po typu řadiče pojmenovává.

= EIDE

= SATA

= SCSI

= EIDE datový kabel



= SATA datový kabel



= SCSI datový kabel



- = EIDE vzniklo inovací standardu IDE (Integrated Drive Electronics)
 - = plošný spoj řadiče je součástí mechaniky disku a spojení řadiče se základní deskou je provedeno přes rozhraní
 - = na základní desce jsou standardně dva řadiče EIDE
 - = na každý je možno připojit 2 zařízení
 - = kabel má 40 nebo 80 žil (závisí na přenosové rychlosti) a je připojen přímo na rozhraní základní desky
- = napájení 5V

- = Pevný disk musí určit přesnou fyzickou polohu dat
- = Řadič disku musí umět definovat požadavek, s kterými daty chce pracovat
 - = obě zařízení tedy musí používat stejnou metodu adresace dat
 - = použitá metoda se musí definovat v BIOSu přes Setup
- = CHS – Cylinder/Head/Sector
 - = nejstarší metoda adresování
 - = poloha dat je definována adresou hlavy, sektoru a cylindru
 - = pro přístup používal BIOS rozhraní Int13h
 - = umožňuje adresovat 10 bitů pro adresu cylindru, 8 bitů pro adresu hlavy a 6 bitů pro adresu sektoru
 - = IDE využívalo 16 b pro cylindr, 4 b pro hlavu a 8 b pro na číslo
 - = max. velikost kapacity disku byla 512MB

= XCHS - Large Int13h

- = dokáže využít všech bitů rozhraní Int13h - kapacita disku je maximálně 7,88 GB.
- = dnes se již nepoužívá.

= LBA – Logical Block Addressing

- = adresování převzato od řadičů SCSI
- = sektory jsou číslovány od 0 do max. hodnoty
- = každá sektor má 28 bitovou adresu
- = je možné adresovat 2²⁸ sektorů - 128GB.
- = pro větší kapacity vznikl standard ATA/ATAPI-6 který má 48 bitovou adresu - adresuje 2⁴⁸ tedy 114 milionů GB.
- = LBA dnes standard

- = Jde o sériové rozhraní
- = Oproti EIDE má několik výhod a změn:
 - = k jednomu zařízení jeden kabel – odpadá časová prodleva nutná k přepínání mezi dvěma EIDE disky
 - = maximální délka kabelu 1m
 - = přenosová rychlost SATA je 150 nebo 300MB/s
 - = je možné připojení a odpojení disku za chodu
 - = možnost vytvoření RAID polí
 - = napájení SATA 500 -600mV
 - = jelikož jde o sériový přenos nedochází k přeslechům - zvýšení frekvence sběrnice

- = Zvýšení přenosové rychlosti na 300MB/s
- = NCQ – Native Command Queuing
- = Staggered Spin-Up (střídavé roztáčení)
- = Port Multiplier

- = Technologie umožňující řazení příkazů do front a vykonávat je optimalizovanou cestou
 - = pokud mikroprocesor potřebuje více dat, zadá řadiči více příkazů o žádosti na data, která se nejspíše nebudou nacházet ve stejné geometrické oblasti
- = NCQ zajistí, že se požadavky nezpracují sekvenčně, ale v optimalizovaném pořadí tak, aby hlavičky nemuseli přeskakovat na vzdálené plochy disku

- = Systém zajišťují postupné roztáčení motorků disků při startu počítače
 - = pokud by docházelo k okamžitému startu všech disků, šlo by o energeticky náročnou operaci, jelikož pro start disk potřebuje až 2 A

= Port Multiplier,

- = je jakýsi „přepínač“ připojený k řadiči disku, který zprostředkuje komunikaci s dalšími pevnými disky
- = lze připojit až 15 zařízení
- = aby nedošlo k citelným ztrátám doporučuje se zhruba 5 zařízení

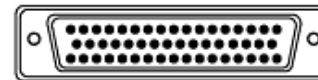
- = Komplexnější rozhraní než EIDE
- = Umožňuje řetězení příkazů



Centronics 50 Male



Centronics 50 Female



DB50 Male



Burndy 68 Male



Micro D 50 Male



Micro D 50 Female



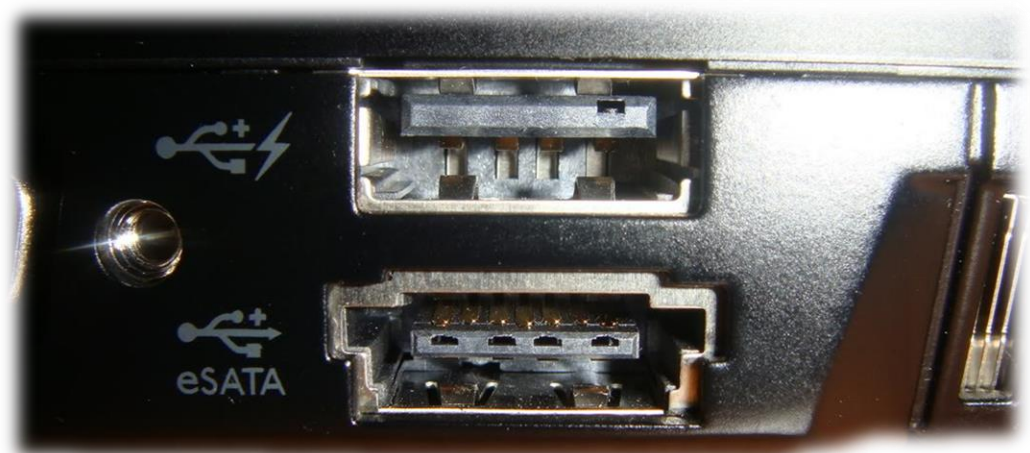
Micro D 68 Male



VHDCI 68 (0.8-mm CHAMP) Male

- = Fyzicky je SCSI realizováno tak, že se do sběrnice slotu na základní desce zasune karta – hostitelský adaptér
- = Od host adaptéru vede datový kabel, který podporuje jednotlivé periferie SCSI
- = Každá z jednotek má vlastní řadič - funkčně samostatná
- = Provoz řídí hostitelský adaptér
- = Přenos řízen hostitelským adaptérem, tedy nezatěžuje procesor
- = Každé zařízení SCSI má vlastní ID0 – 6, hostitelský adaptér ID7
- = Sběrnice musí být zakončena odporem (terminátorem)
- = Datová šířka 8 nebo 16 s frekvencí 5Hz až 40 Hz

- = V dnešní době se pro připojení externích disků používají rozhraní USB2.0, USB3.0, případně eSATA
- = Rozdílem v použité komunikační sběrnici je rychlost přenosu dat a možnosti napájení pevného disku
 - = eSATA konektor neobsahuje napájecí vodiče, nabízí však vyšší přenosovou rychlost proti USB 2.0
 - = S nástupem USB 3.0 se od eSATA upouští
 - = Konektor eSATA je často sdílený s některým z USB portů
kdy kontakty eSATA
jsou umístěny na
horní straně
a USB na spodní





Univerzita Hradec Králové
Fakulta informatiky a managementu

Děkuji za pozornost...

