

# Ekstern sortering



# Ekstern sortering

(sortering på eksterne lagermedier)



Særlige hensyn:

- (1) Det er tidsmæssigt dyrt at tilgå et dataelement
- (2) Der er begrænsninger på tilgangen,  
f.eks. kan et magnetbånd kun læses sekventielt

# Fordel-og-flet



## **Fordeling:**

Opdel filen i blokke på størrelse med det indre lager  
Sorter hver af disse blokke og fordel dem på 2 eller flere bånd

## **Fletning:**

Flet de sorterede blokke til længere sorterede blokke  
Fortsæt på denne måde, indtil hele filen er én sorteret blok

# Balanceret flervejsfletning

Eksempel: 3-vejsfletning af 81 poster



Sorterede blokke (længden målt i antal poster)

1	9 (3)	0	1 (27)	0
2	9 (3)	0	1 (27)	0
3	9 (3)	0	1 (27)	0
<hr/>				
4	0	3 (9)	0	1 (81)
5	0	3 (9)	0	
6	0	3 (9)	0	

3 passager for at sortere 81 poster

Fletningen kan foretages ved hjælp af en prioritetskø

# Balanceret flervejsfletning



$N$ : antal poster

$M$ : størrelse af indre lager (målt i antal poster)

Benyt halvdelen af båndene som indbånd, resten til udbånd

Passage 0: fordel filen i sorterede blokke af størrelse  $M$   
på bånd  $1, 2, \dots, k$

Passage 1:  $k$ -flet blokkene fra bånd  $1, 2, \dots, k$  til blokke af  
størrelse  $kM$  ud på bånd  $k+1, k+2, \dots, 2k$

Passage 2:  $k$ -flet blokkene fra bånd  $k+1, k+2, \dots, 2k$  til  
blokke af størrelse  $k^2M$  ud på bånd  $1, 2, \dots, k$

...

Passage  $p$ :  $k$ -flet blokkene fra indbåndene til en blok af  
størrelse  $k^pM$  ud på et af på udbåndene.



Filen er sorteret, når

$$k^p M \geq N$$

d.v.s. efter

$$p = \log_k(N/M) \text{ passager}$$

Eksempel:

filstørrelse  $10^9$  poster

lagerstørrelse  $10^6$  poster

antal båndstationer 4

antal passager  $\log_2 10^3 = \mathbf{10}$

antal båndstationer 20

antal passager  $\log_{10} 10^3 = \mathbf{3}$

Filen kan sorteres på 3-10 gange den tid, det vil tage at læse eller skrive den

# Polyfasesortering



Reducerer antallet af bånd til cirka det halve af antallet af bånd ved balanceret flervejsfletning

Princip:

Benyt hele tiden  $k-1$  indbånd og 1 udbånd

Algoritme:

Foretag fletning fra de  $k-1$  indbånd ud på udbåndet, indtil et af indbåndene bliver tomt

Det tomme bånd tilbagespoles og benyttes som nyt udbånd for fletning fra de øvrige  $k-1$  bånd

Således fortsættes, indtil filen er sorteret

# Blokfordeling ved polyfasesortering



Fordeling:

Fordel blokkene på  $k-1$  bånd, således at den sidste fletning gør alle indbånd tomme samtidigt

<b>1</b>	<b>21</b>	<b>8</b>	<b>0</b>	<b>5</b>	<b>3</b>	<b>1</b>	<b>0</b>	<b>1</b>
<b>2</b>	<b>13</b>	<b>0</b>	<b>8</b>	<b>3</b>	<b>0</b>	<b>2</b>	<b>1</b>	<b>0</b>
<b>3</b>	<b>0</b>	<b>13</b>	<b>5</b>	<b>0</b>	<b>2</b>	<b>0</b>	<b>1</b>	<b>0</b>

**3 bånd, 34 blokke, 7 passager**

Fordelingsmønsteret kan bestemmes ved brug af generaliserede Fibonacci

$$F^k(N) = F^k(N-1) + F^k(N-2) + \dots + F^k(N-k)$$
$$F^k(0 \leq N \leq k-2) = 0, F^k(k-1) = 1$$



# Afløserteknikken

(engelsk: replacement selection)



En teknik, der muliggør at blokkene i fordelingsfasen kan blive af længde  $2M$

Organiser posterne i det indre lager som en hob

Når lageret er fyldt, og der ankommer en ny post, skrives den mindste post fra hoben ud på båndet og afløses af den nye post

Der er nu 2 tilfælde:

- (1) Den nye post er større end eller lig med den sidst udskrevne post.  
Posten indsættes i hoben
- (2) Den nye post er mindre end den sidst udskrevne post.  
Posten tilbageholdes indtil videre

En blok er udskrevet, når alle poster i det indre lager er tilbageholdt (hoben er tom)  
Derefter organiseres de tilbageholdte poster som en hob, og der fortsættes som hidtil