UNIVERSITÀ PARTHENOPE
DiST
Dipartimento di Scienze e Tecnologie

Intelligent Signal Processing

File Placement

Angelo Ciaramella

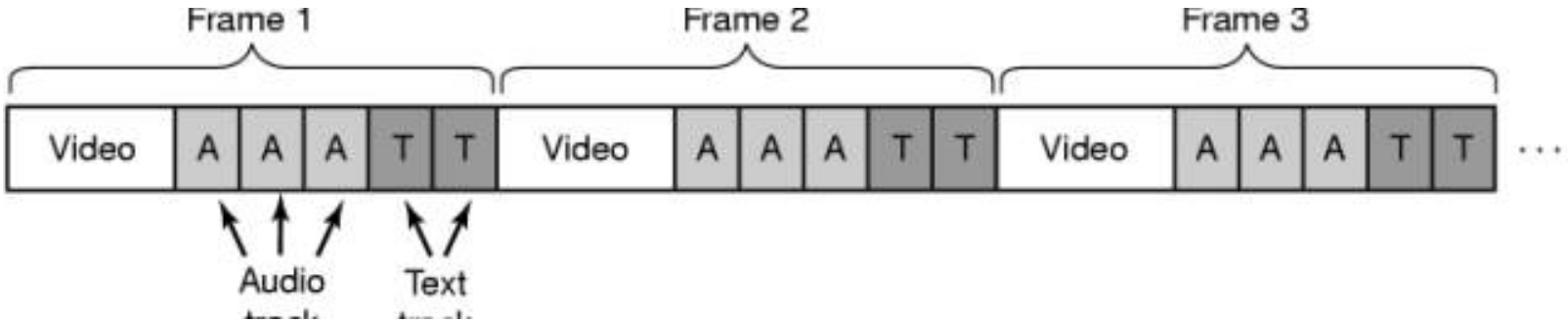
Introduction

- Multimedia files are
  - very large
  - often written only once but read many times
  - tend to be accessed sequentially
  - playback must also meet strict Quality of Service criteria

Single disk

- Multimedia files
  - data can be streamed to the network or output device at the requisite speed and without jitter
  - seeks during a frame is highly undesirable
- One complication
  - presence of video, audio, and text
    - each stored as separate contiguous files

Interleaving



Interleaving video, audio and text in a single contiguous file per movie.

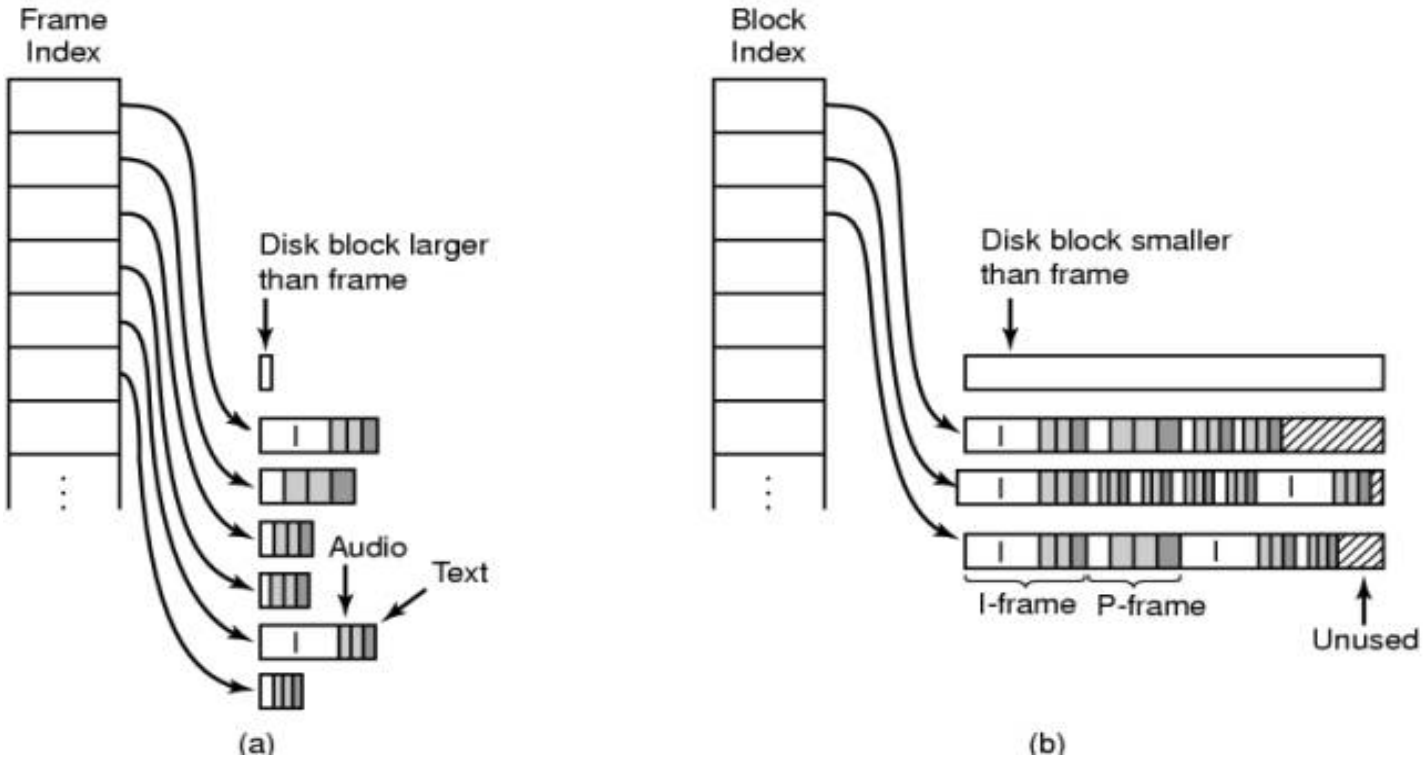
Interleaving

- Interleaving
  - eliminates all seeks (on a single-user system)
  - does not require any overhead for keeping track
  - random access is impossible
  - *fast forward* and *fast backward* are impossible without additional data structures and complexity

File oraganization strategies

- Two alternative file placement organizations
  - *Small disk blocks*
    - disk block size is chosen to be considerably smaller than the average frame size
  - *Large disk blocks*
    - multiple frames in each block

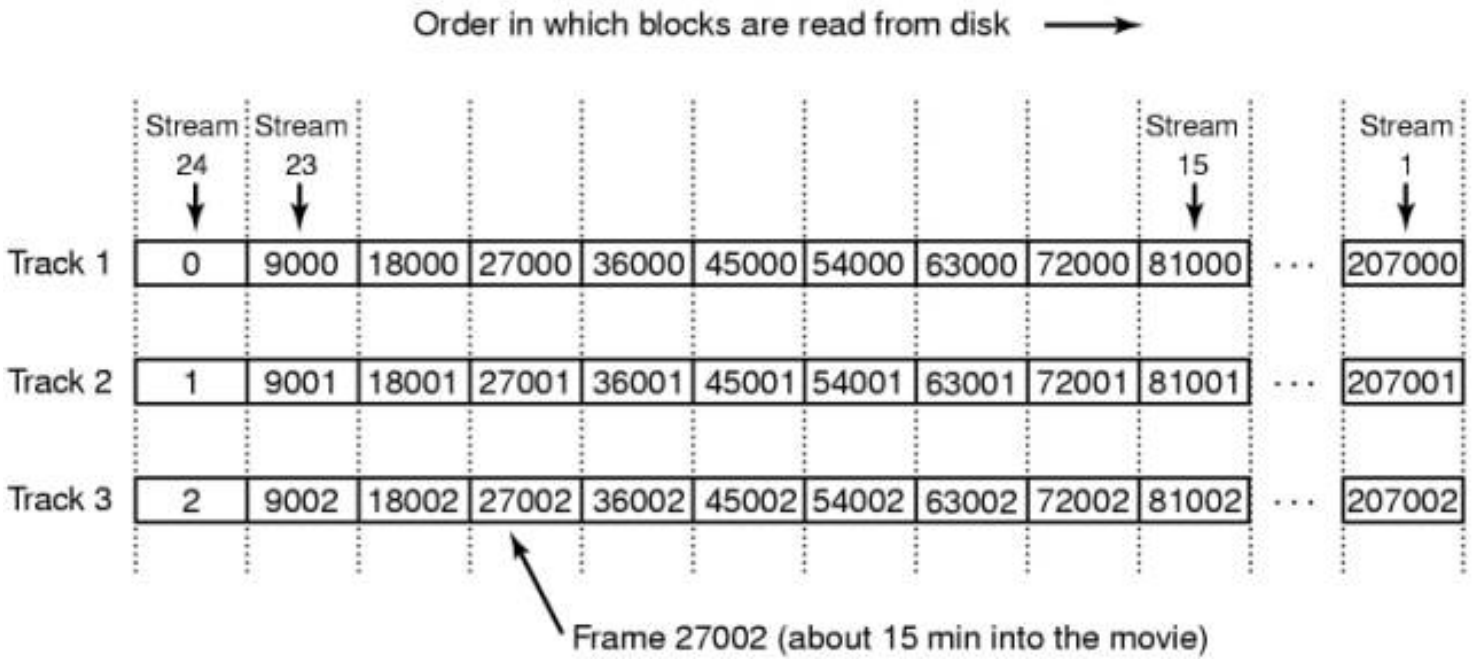
File oraganization strategies



Noncontiguous movie storage: (a) small disk blocks; (b) large disk blocks

Near Video on Demand

30 frames/sec – new stream starting every 5 min



Optimal frame placement for near video on demand (Chen and Thapar, 1997)

Near Video on Demand

- A simple buffering strategy is to use double buffering
  - while one buffer is being played out onto 24 streams, another buffer is being loaded in advance
  - when the current one finishes, the two buffers are swapped

- how large to make the buffer
  - e.g., largest track in the movie is known in advance, so a buffer of precisely that size can be chosen
  - trade-off here between memory used for buffers and *quality of the movies*

Multiple files

- On a Video Server
  - time will be wasted moving the disk head from movie to movie when multiple movies are being viewed simultaneously by different customers

- Observation
  - some movies are more popular than others
  - taking popularity into account when placing movies on the disk

Zipf's law

- Zipf's law
  - George Zipf, Harvard professor of linguistics
  - if the movies, books, Web pages, or words are ranked on their popularity, the probability that the next customer will choose the item ranked *k*-th in the list is C/k
  - If there are N movies, C is computed such that

$$C + C/2 + C/3 + \ldots + C/N = 1$$

Zipf's law

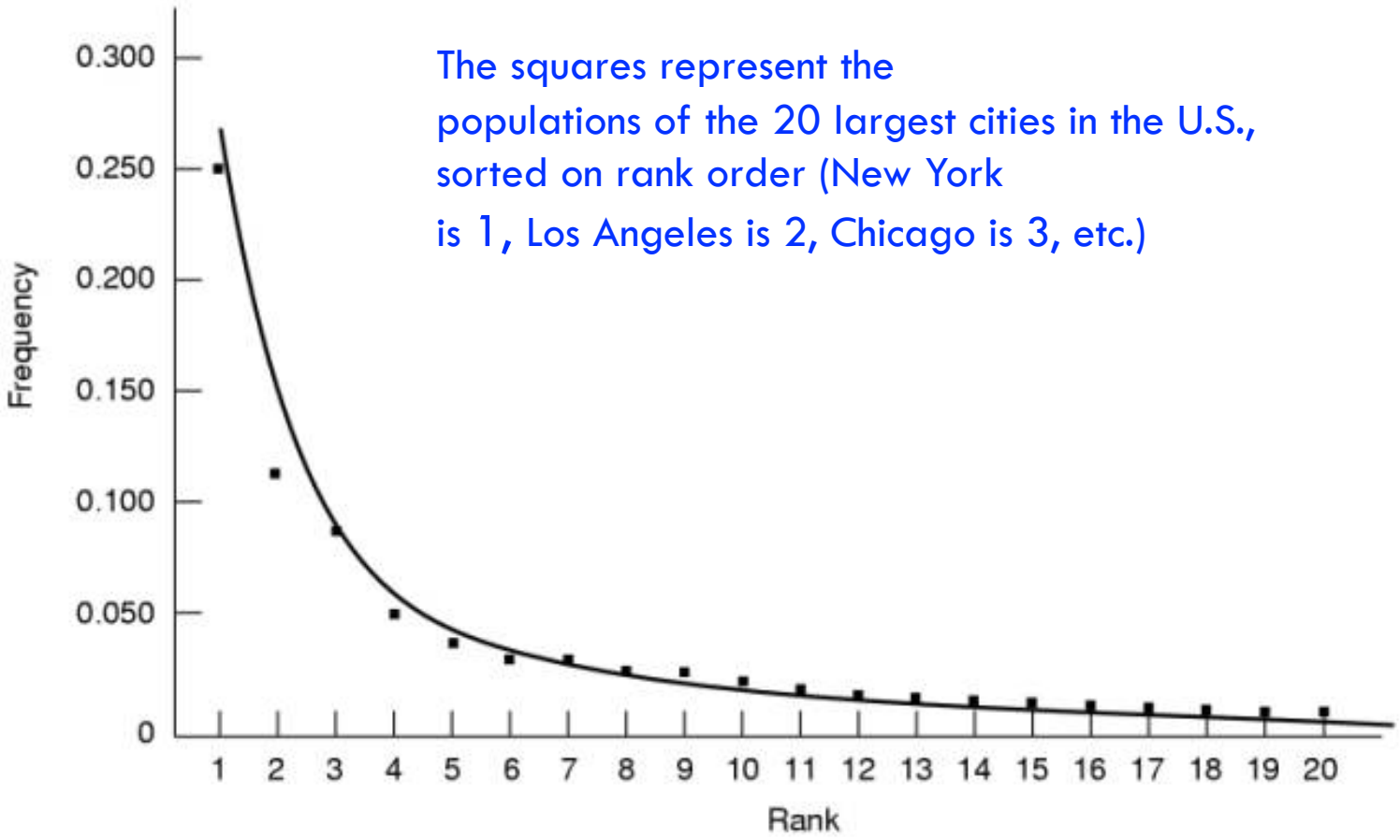
N. population	C
10	0.341
100	0.193
1000	0.134
10000	0.102

C values varying *N*

N. movies	Probabilities
1	0.134
2	0.067
3	0.045
4	0.034
5	0.027

Probabilities
for the top five movies
with *N*=1000

Zipf's law



The squares represent the populations of the 20 largest cities in the U.S., sorted on rank order (New York is 1, Los Angeles is 2, Chicago is 3, etc.)

Zipf's law predicts that the second largest city should have a population half of the largest city and the third largest city should be one third of the largest city, and so on.

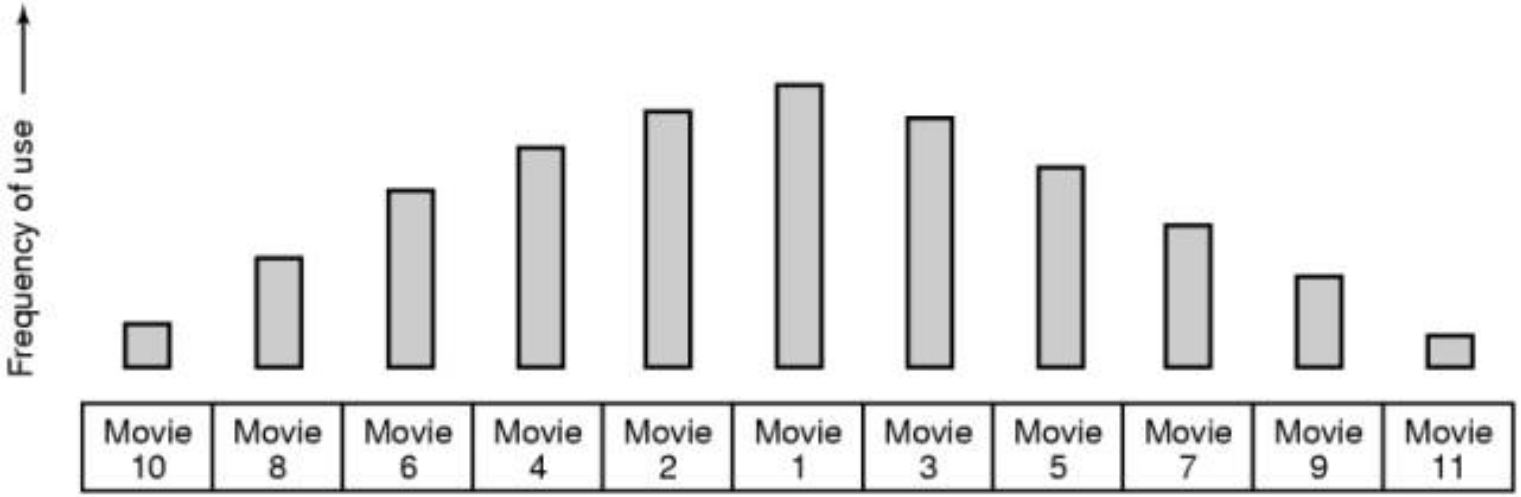
Zipf's law

- For movies on a video server
  - Zipf's law states that the most popular movie is chosen twice as often as the second most popular movie, three times as often as the third most popular movie, and so on
  - e.g., movie 50 has a popularity of C/50 and movie 51 has a popularity of C/51 , so movie 51 is 50/5 1as popular as movie 50, only about a 2% difference

Organ-Pipe algorithm

- Organ-Pipe algorithm
  - Grossman and Silverman (1973) and Wong (1983)
  - Studies have shown that the best strategy is surprisingly simple and distribution independent
  - placing the most popular movie in the middle of the disk, with the second and third most popular movies on either side of it

Organ-Pipe algorithm

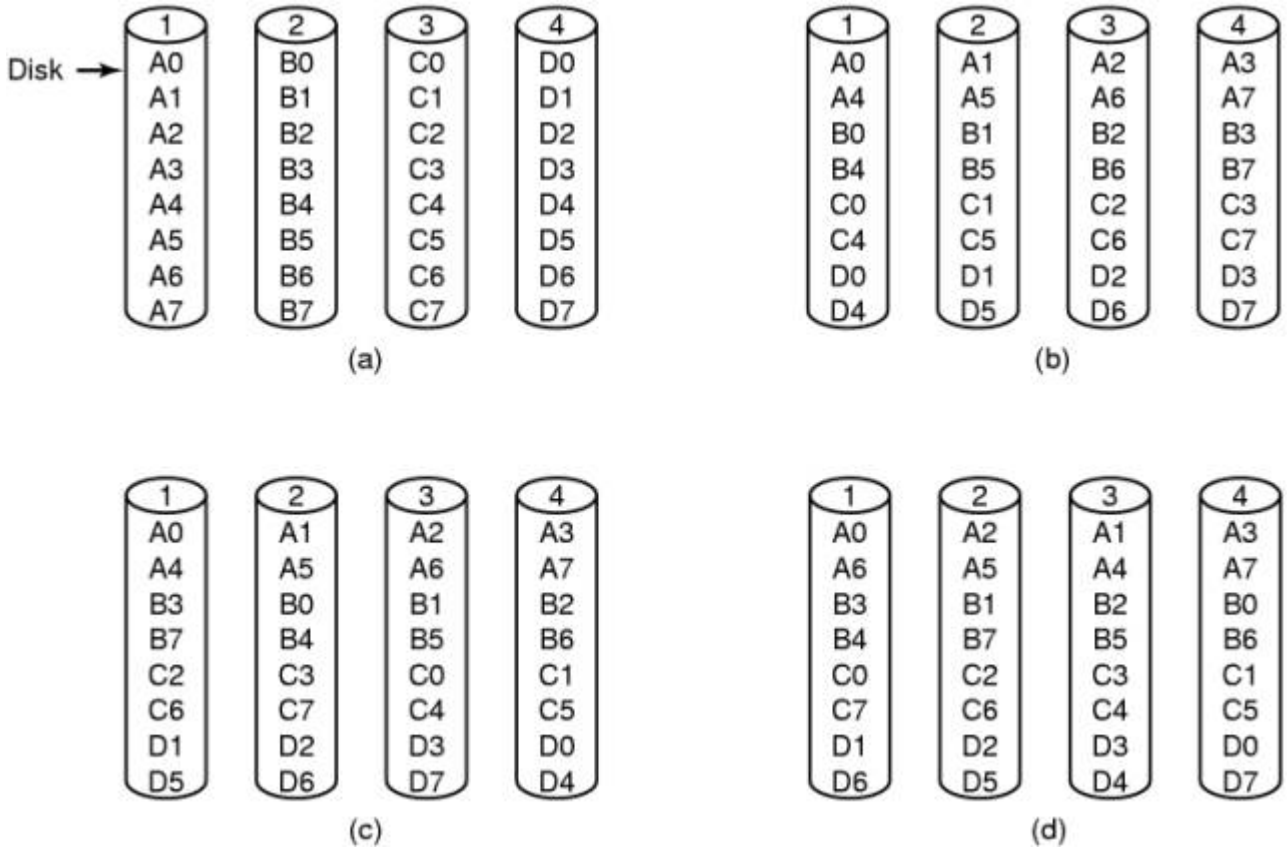


The organ-pipe distribution of files on a video server

Organ-Pipe algorithm

- Organ-Pipe algorithm
  - With 1000 movies and a Zipf's law distribution
    - the top five movies represent a total probability of 0.307
    - the disk head will stay in the cylinders allocated to the top five movies about 30% of the time

Multiple Disks – Disk Farm



Four ways of organizing multimedia files over multiple disks. (a) No striping. (b) Same striping pattern for all files. (c) Staggered striping. (d) Random striping.