Continuous Media Storage Servers

Leana Golubchik

University of Maryland

John C.S. Lui

CUHK

Richard R. Muntz

UCLA

Multimedia Storage Systems Performance Analysis

Dick Muntz

Leana Golubchik - University of Maryland
John C.S. Lui - CUHK
Richard R. Muntz - UCLA

Outline:

- Introduction
  - Why is it important? Why is it different?
  - Types of issues
  - Metrics and workload characterization
- Video on demand systems
- 3D interactive systems
- A few things that I did not cover

Leana Golubchik - University of Maryland
John C.S. Lui - CUHK
Richard R. Muntz - UCLA

Some things I did not cover:

- processor scheduling
- networking
- multimedia presentations
- teriary storage
- distributed multimedia systems

Leana Golubchik - University of Maryland
John C.S. Lui - CUHK
Richard R. Muntz - UCLA

Why is multimedia important?

It is estimated that the percentage of storage devoted to multimedia will grow from 10% in 1995 to over 50% by 2005.

Why is multimedia "different" ?

- large size of objects
  - compression is particularly important
    - constant bit rate (CBR)
    - variable bit rate (VBR)
    - scalable compression (multilayer)
- realtime/quality of service requirements
  - delay bounds
  - jitter (single stream, multiple streams)
  - frame rate, quality of image/audio (PSNR)
  - synchronization

Leana Golubchik - University of Maryland
John C.S. Lui - CUHK
Richard R. Muntz - UCLA

Workload characterization

- stream = one delivery stream,
  may have multiple clients or single client

  two delivery streams of the same object which are offset in time are different streams

  not all streams to the same object have exactly the same characteristics, e.g., scalable compression
- arrival rate of requests for objects with a certain QoS
- characteristics of access streams:
  - strictly continuous, e.g., scheduled video
  - VCR functions
  - frequent jumps, e.g., news on demand
  - CBR, VBR
  - frame rate

Leana Golubchik - University of Maryland
John C.S. Lui - CUHK
Richard R. Muntz - UCLA

Workload characterization (cont'd.)

- relative frequency of access to objects,
  e.g., Zipf distribution for relative frequency of access to videos

- mixed workloads
  - realtime and non-realtime, necessary but often ignored
  - CBR for each object but multiple rates
  - video/audio and 3D interactive

Leana Golubchik - University of Maryland
John C.S. Lui - CUHK
Richard R. Muntz - UCLA

Disk systems

disk performance issues

- seek times: convex function of number of tracks traversed. Worst case for N seeks is when all cylinders traversed and requests are equally spaced.
- rotational latency: with no relation to track boundaries latency is uniformly distributed. If whole tracks are fetched, rotational latency can be made insignificant ... but you have to program the controller.
- track settle time
- controller cache ineffective
- disk zones: most disks have constant density of bits per inch on track. Outer tracks can have several times the capacity of inner tracks and, consequently, several times the transfer rate.
- thermal recalibration: disk controller goes on vacation every so often, e.g., 50 msec.

Leana Golubchik - University of Maryland
John C.S. Lui - CUHK
Richard R. Muntz - UCLA

Disk systems (cont'd.)

- T(n) <= t-seek(n) + n(t-rot + B/r-min)

  where:

T(n)	=	time to serve n requests
t-seek(n)	=	max seek time for n requests
t-rot	=	maximum rotation time
B	=	disk block size (assume const. here)
r-min	=	minimum transfer rate

- can also add thermal recalibration time to right hand side
- can treat seek times, rotational latencies, etc. as variables and estimate probability that T(n) <= x; actually usually want x such that P(T(n) > x) <= a
- if the particular n requests are given, then can take this information into account, e.g., seek times are known more precisely.

Leana Golubchik - University of Maryland
John C.S. Lui - CUHK
Richard R. Muntz - UCLA

Disk systems (cont'd.)

- disk seek time is too high to ignore and pay the price of a random seek per request
- all proposed layout and scheduling schemes propose to schedule requests in "rounds" i.e., a set of requests are scheduled using some form of SCAN algorithm to reduce seek time per request.
- A taxonomy of cases ... for audio/video:

	single disk	multiple disks
CBR one obj. type	Fellini (AT&T)	simple striping, streaming RAID
CBR mult. obj. type	Fellini	staggered striping
VBR	E. Chang (CDL, CTL)	Vin, et.al.

- RIO, at UCLA ... more later

Leana Golubchik - University of Maryland
John C.S. Lui - CUHK
Richard R. Muntz - UCLA

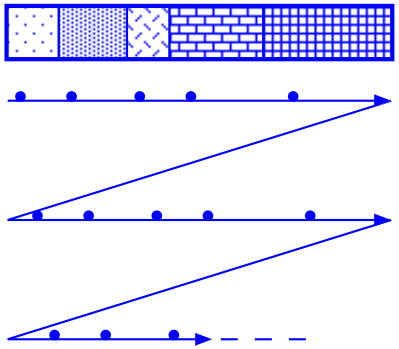
Disk systems (cont'd.)

- Single disk, CBR, all objects with the same rate constant disk blocks

Define S as a "service round"

Each stream gets served once per service round

Constraints:



If N = number of streams, then T(N) <= S

If p = playout rate of objects, then B/p >= S

(B = disk block size)

Let T(N) = aN + NB/r

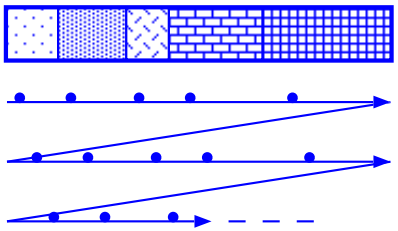
then: aN + NB/r <= S <= B/p

or N <= B/[p(a + B/r)]

Leana Golubchik - University of Maryland
John C.S. Lui - CUHK
Richard R. Muntz - UCLA

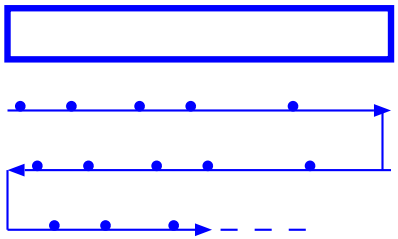
Disk systems (cont'd.)

- **Single disk, CBR, all objects with the same rate constant disk blocks**



(sequential allocation)

- **buffer requirements are Sp per stream**



(random allocation)

- **buffer requirements are 2Sp per stream**

Leana Golubchik - University of Maryland
John C.S. Lui - CUHK
Richard R. Muntz - UCLA

Disk systems (cont'd.)

- Single disk, CBR encoding, multiple object types
  - for given object i, B_i= block size read

    simple change to the previous formula for admission control

- Single disk, VBR encoding
  - several choices

    CDL = constant data length disk block (1 read per cycle)

    CTL = constant time length disk block (1 read per cycle)

    Hybrid = constant data length disk block but multiple read during a service cycle

Leana Golubchik - University of Maryland
John C.S. Lui - CUHK
Richard R. Muntz - UCLA

Disk systems (cont'd.)

- **Single disk, VBR encoding**

  CTL = constant time length disk block (1 read per cycle)

  now the read time per cycle is a variable

  two types of admission control

  deterministic

  (a) use worst case seek times, etc.

  (b) use knowledge of the block sizes required on all future cycles

  stochastic

  use estimates of tail of distribution based on first few moments

  Hybrid = constant data length disk block but multiple read during a service cycle

  CDL = constant data length disk block (1 read per cycle)

Leana Golubchik - University of Maryland
John C.S. Lui - CUHK
Richard R. Muntz - UCLA

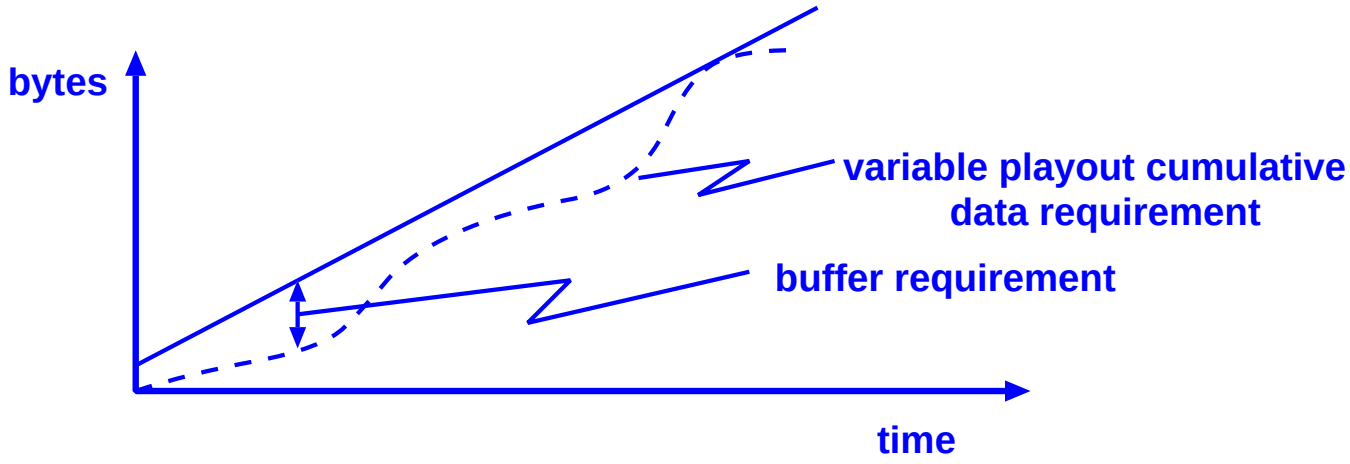
Disk systems (cont'd.)

- Single disk, VBR encoding

  CDL = constant data length disk block (1 read per cycle)

  hardly worth discussing

  in this case there is a constant data rate off the disk and one has the problem of ensuring there is no starvation in the playout ... large buffer requirement



  Hybrid = constant data length disk block but multiple read during a service cycle

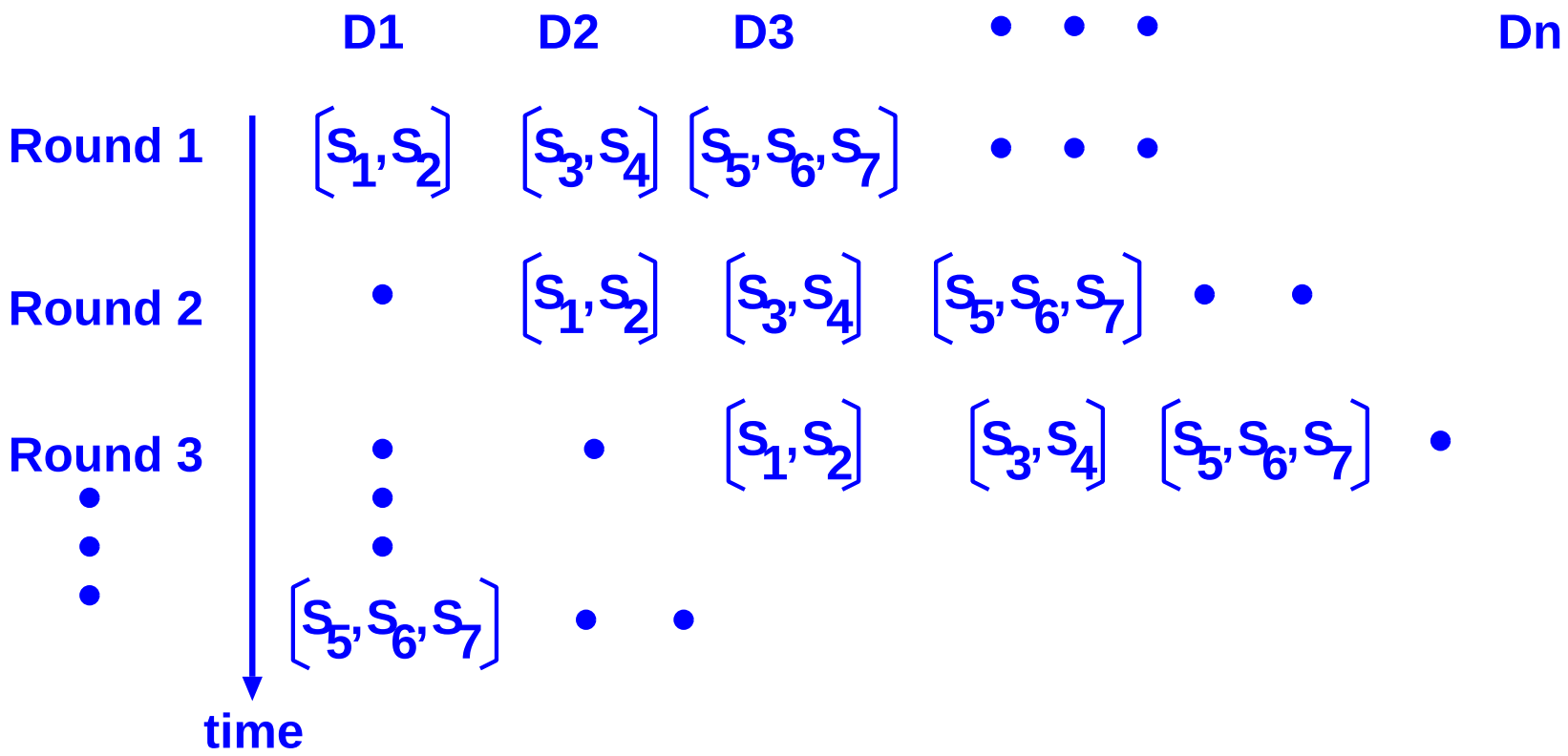
  basically the same admission control scheme alternatives as with CTL

Leana Golubchik - University of Maryland
John C.S. Lui - CUHK
Richard R. Muntz - UCLA

Disk systems (cont'd.)

Multiple disks, CBR, one type of object

data striped in fixed size blocks over disks

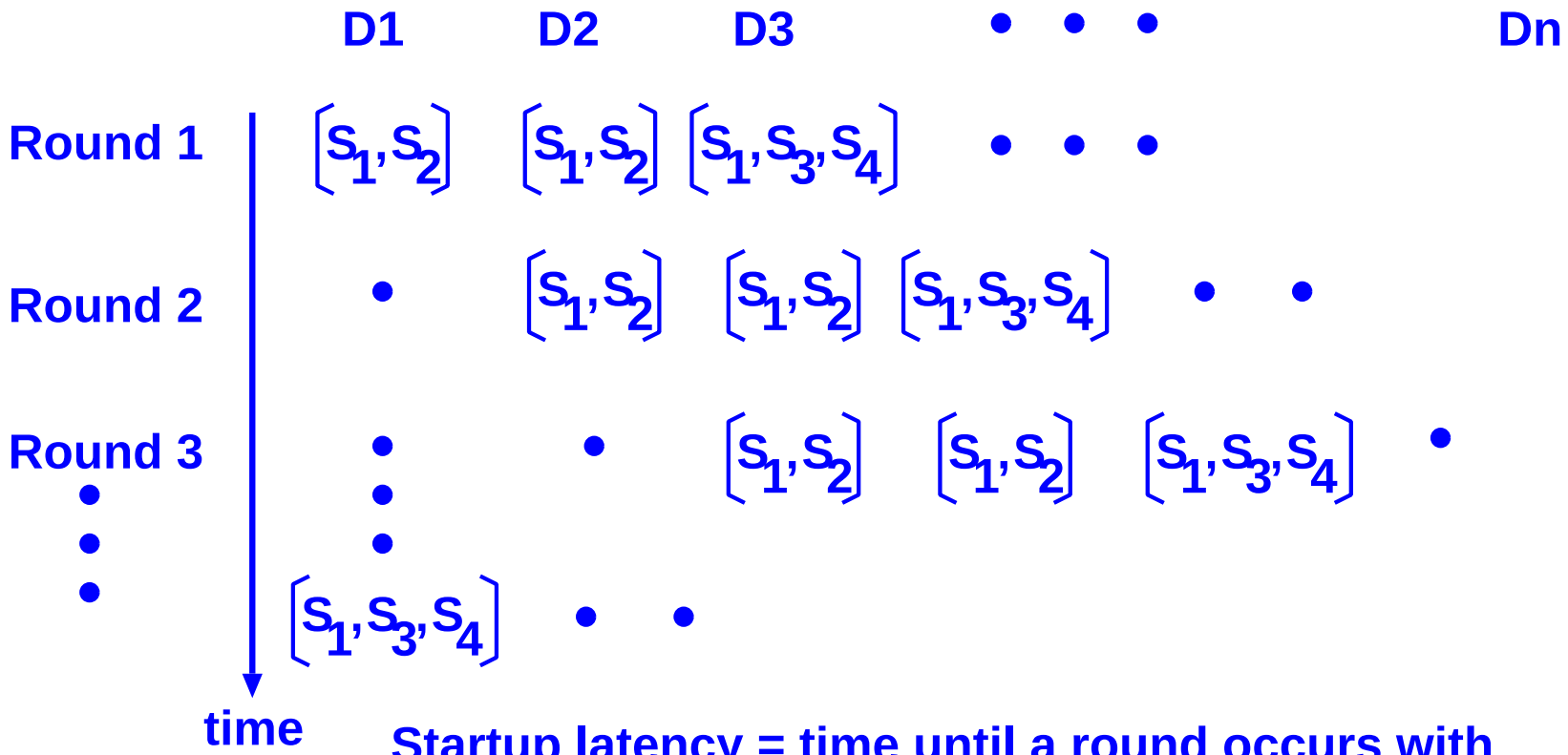


Startup latency = time until a round occurs with an "available slot" on the disk where the object starts

Leana Golubchik - University of Maryland
John C.S. Lui - CUHK
Richard R. Muntz - UCLA

Disk systems (cont'd.)

Multiple disks, CBR, multiple rate objects
data striped in fixed size blocks over disks
(staggered striping)



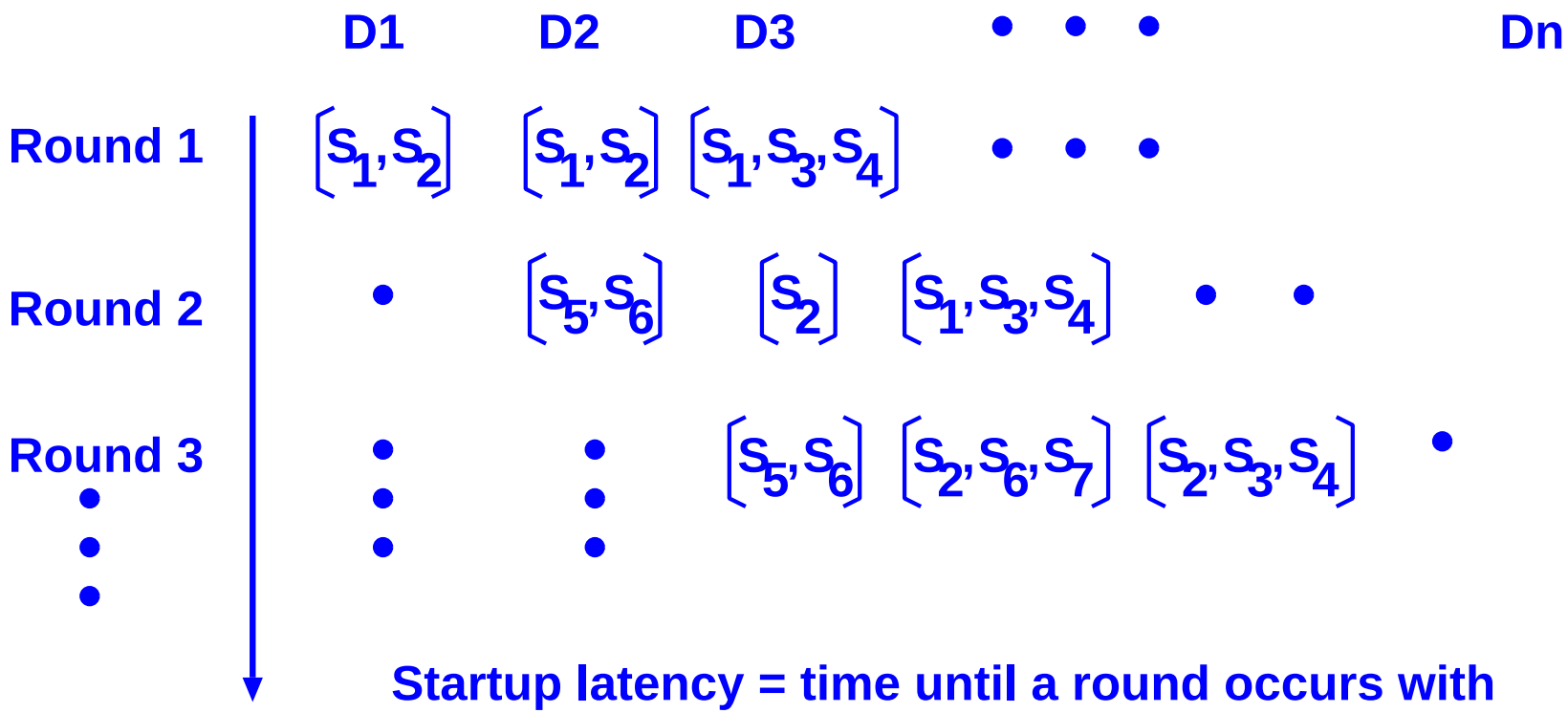
Startup latency = time until a round occurs with the requisite number of available slots in consecutive positions that cover the initial set of blocks for the object

Leana Golubchik - University of Maryland
John C.S. Lui - CUHK
Richard R. Muntz - UCLA

Disk systems (cont'd.)

Multiple disks, VBR

data striped in fixed size blocks over disks



Startup latency = time until a round occurs with the requisite number of available slots in consecutive positions that cover the initial set of blocks for the object

Leana Golubchik - University of Maryland
John C.S. Lui - CUHK
Richard R. Muntz - UCLA

Disk systems (cont'd.)

Scalable compression (layered compression)

Temporal or spatial:

Temporal:

Ex. pattern for a GOP in MPEG encoding

$I_1 B_1 B_2 P_1 B_3 B_4 P_2 B_5 B_6 P_3 B_7 B_8$

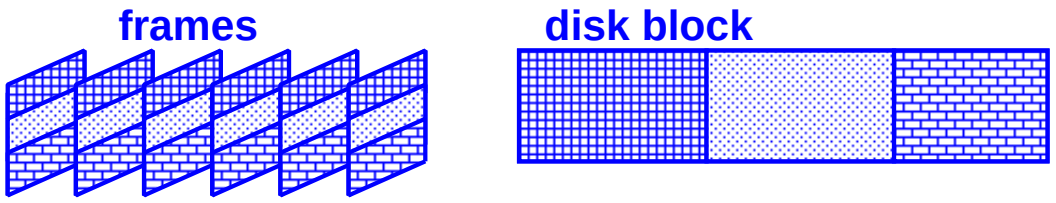
layout in a disk block

$I_1 P_1 P_2 P_3 B_1 B_2 B_3 B_4 B_5 B_6 B_7 B_8$

two GOP in same disk block

$I_1 I_2 P_{11} P_{12} P_{13} P_{21} P_{22} P_{23} B_{11} B_{12} B_{13} \bullet \bullet \bullet$

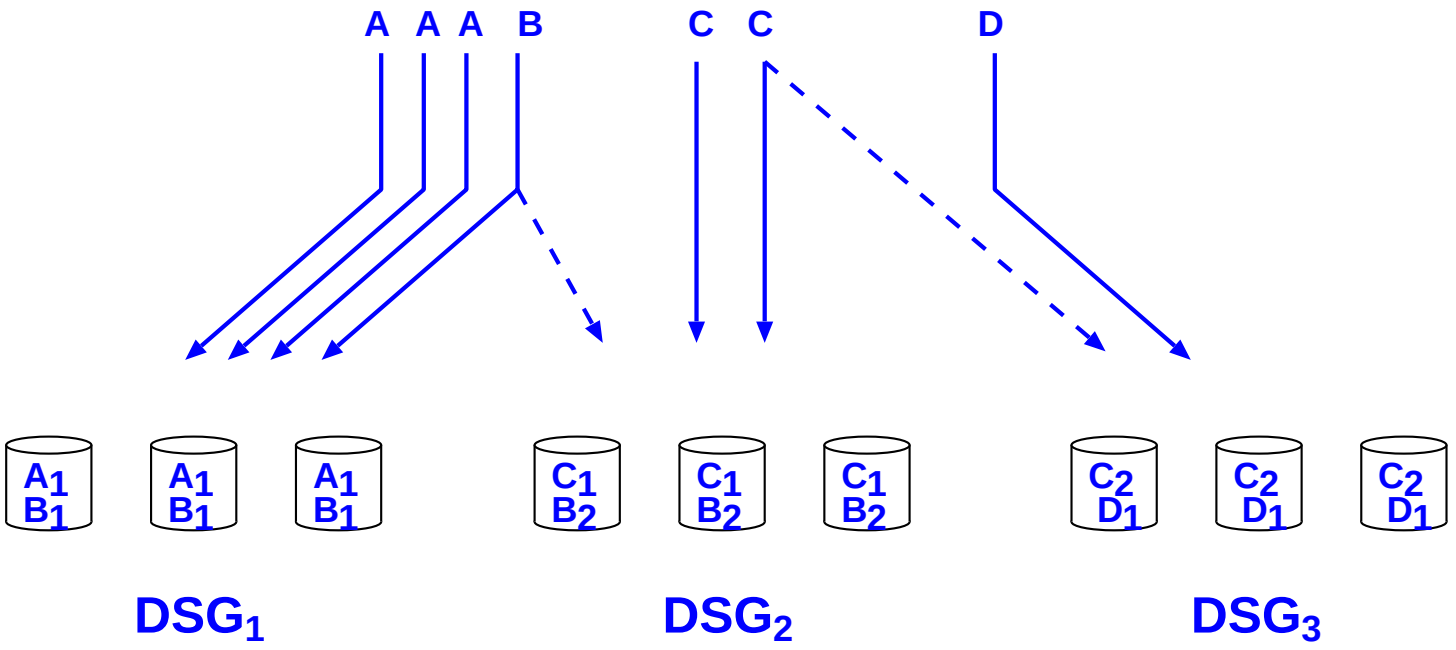
Spatial:



Leana Golubchik - University of Maryland
John C.S. Lui - CUHK
Richard R. Muntz - UCLA

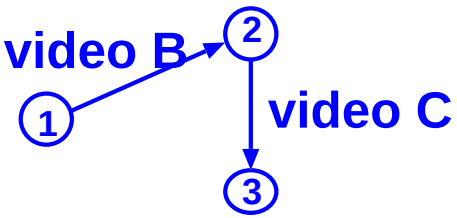
Disk systems (cont'd.)

DASD dancing [Wolfe et.al.]



Dynamic Algorithm

Ex. DSG1 overloaded
DSG3 underloaded
w.r.t opt.



Leana Golubchik - University of Maryland
John C.S. Lui - CUHK
Richard R. Muntz - UCLA

Disk systems (cont'd.)

Some of the weirdnesses of disks

- disk zones

  track pairing schemes [Birk]

- thermal recalibration
  - can be scheduled via the disk controller
  - add additional buffering to mask delay
  - buy more expensive disks which mask calibration
  - disk arrays: one disk calibrated at a time, use parity to mask unavailability of that disk

- heterogeneous disks

  DASD dancing

  [+ A. Dan, et.al.]

Leana Golubchik - University of Maryland
John C.S. Lui - CUHK
Richard R. Muntz - UCLA