DBMS Fall 2001, Exercise 1, Monday, Nov 12, MT2
(Group 1 at 2–4 pm, Group 2 at 4–6 pm)

1. (Ex. 2.2.1 in Textbook) The *Megatron 777 (M777)* disk has the following characteristics: There are 10 surfaces, with 10,000 tracks each. Tracks hold an average of 1000 sectors of 512 bytes each. 20% of each track is used for gaps. The disk rotates at 10,000 RPM. The time to move the head n tracks is 1 + 0.001n milliseconds.

   Answer the following questions about the M777 disk:
   
   (a) What is the capacity of the disk?
   (b) What is the maximum seek time?
   (c) What is the maximum rotational latency?
   (d) If a block is 16 KB, what is the transfer time of a block?
   (e) What is the average seek time?
   (f) What is the average rotational latency?

2. (Ex. 2.3.1 in Textbook) Using Two-Phase, Multiway Merge-Sort, how long would it take to sort the sample relation whose sorting was considered at the lecture, using the M777 disk described above? Consider the case where the blocks are stored in random locations on the disk. Assume that other characteristics of the machine and the data remain the same (except for the disk used to store the data).

3. Consider again performing the above sorting task with the same hardware as above. How fast could you make the sorting by applying the tricks that were discussed at the lecture: arranging blocks sequentially along cylinders, using buffers of the size of a track or a cylinder, and double buffering?

4. (a) Present a reasonable physical representation for patient records of the following kind: Patients are identified by their birth date, a social security number, and a 10-digit patient number, which all are represented as fixed-length fields. Additionally, the records contain variable-length fields for the name, the address and the case history (*Fin.* hoitokertomus) of the patient. (Notice that the
DBMS must be able to recognize and interpret the fields of such records.)

(b) How much storage space would an average patient record take? Assume the following: The lengths of patient names vary uniformly between 10–50 characters, the lengths of their addresses similarly btw 20–80 characters, and the lengths of case histories btw 0–3000 characters.

5. Consider implementing a file of 100,000 patient records, each of which is 1000 bytes long. Assume that the size of blocks is 2 KB (2048 bytes), out of which 40 bytes are used for the block header. There is a sparse primary index on the file based on 10 character patient numbers, and a two-level secondary index based on 20 character family names of patients. A family name occurs in the file 10 times on the average.

(a) How much storage is required by the primary index and by the secondary index of the file? Assume that each DB address requires 4 bytes and that all records are unspanned.

(b) How many disk accesses are needed to fetch a record by its patient number if the primary index is held in main memory?

(c) How many disk accesses are needed to fetch the 10 (on the average) patient records by a family name if the second level of the secondary index is held in main memory?