1 DRAM Refresh [80 points]

A memory system is composed of eight banks, and each bank contains 32K rows. The row size is 8 KB.

Every DRAM row refresh is initiated by a command from the memory controller, and it refreshes a single row. Each refresh command keeps the command bus busy for 5 ns.

We define *command bus utilization* as a fraction of the total time during which the command bus is busy due to refresh.

The retention time of each row depends on the temperature (T). The rows have different retention times, as shown in the following Table 1:

Retention Time	Number of rows
$(128 - T)$ ms, $0°C \leq T \leq 128°C$	2^8 rows
$2 * (128 - T)$ ms, $0°C \leq T \leq 128°C$	2^{16} rows
$4 * (128 - T)$ ms, $0°C \leq T \leq 128°C$	all other rows
$8 * (128 - T)$ ms, $0°C \leq T \leq 128°C$	2^8 rows

Table 1: Retention time

1.1 Refresh Policy A [20 points]

Assume that the memory controller implements a refresh policy where all rows are refreshed with a fixed refresh interval, which covers the worst-case retention time (Table 1).

(a) [5 points] What is the maximum temperature at which the DRAM can operate reliably with a refresh interval of 32 ms?

T=96°C.

Explanation. Looking at the first column of Table 1, it is obvious that the DRAM rows with the least retention time (i.e., the first row of the table) can operate at temperatures up to 96°C (128 - 96) when refreshed at fixed 32 ms refresh period.

(b) [15 points] What command bus utilization is directly caused by DRAM refreshes (with refresh interval of 32 ms)?

4.096%.

Explanation. The time that the command bus spends on refresh commands in 32ms is 2^{18}*5ns, where 2^{18} is the number of rows (8 banks, 32K rows per banks), and 5ns is the time that the bus is busy for each refresh command.

Utilization = $\frac{2^{18} * 5ns}{32ms} = 4.096\%$

1.2 Refresh Policy B [15 points]

Now assume that the memory controller implements a refresh policy where all rows are refreshed only *as frequently as required* to correctly maintain their data (Table 1).

(a) [15 points] How many refreshes are performed by the memory controller during a 1.024 second period? (with T=64°C)

The number of refreshes in 1.024 seconds is 1313280.

Explanation. 2^8 rows are refreshed every 64ms $\implies$ $(1024/64) * 2^8 = 2^{12}$ refresh commands in 1.024 seconds.
2^{16} rows are refreshed every 128ms $\implies$ $(1024/128) * 2^{16} = 2^{19}$ refresh commands in 1.024 seconds.
2^8 rows are refreshed every 512ms $\implies$ $(1024/512) * 2^8 = 2^9$ refresh commands in 1.024 seconds.

The remaining rows are $2^{18} - (2^8 + 2^{16} + 2^8) = 2^{18} - 2^{16} - 2^9$, so: $2^{18} - 2^{16} - 2^9$ rows are refreshed every 256ms $\implies$ $(1024/256) * (2^{18} - 2^{16} - 2^9) = 2^{20} - 2^{18} - 2^{11}$ refresh commands in 1.024 second.

Total: $2^{12} + 2^{19} + 2^9 + 2^{20} - 2^{18} - 2^{11} = 1313280$ refresh commands in 1.024 second.

1.3 Refresh Policy C [25 points]

Assume that the memory controller implements an even smarter policy to refresh the memory. In this policy, the refresh interval is fixed, and it covers the worst-case retention time (64ms), as the refresh policy in part 1.1. However, as an optimization, a row is refreshed only if it has *not* been **accessed** during the past refresh interval. For maintaining correctness, if a cell reaches its maximum retention time without being refreshed, the memory controller issues a refresh command.

(a) [5 points] Why does a row *not* need to be refreshed if it was accessed in the past refresh interval?

The charge of a DRAM cell is restored when it is accessed.

(b) [20 points] A program accesses all the rows repeatedly in the DRAM. The following table shows the access interval of the rows, the number of rows accessed with the corresponding access interval, and the retention times of the rows that are accessed with the corresponding interval.

Access interval	**Number of rows**	**Retention times**
1ms	2^{16} rows	64ms, 128ms or 256ms
60ms	2^{16} rows	64ms, 128ms or 256ms
128ms	all other rows	128ms or 256ms

What command bus utilization is directly caused by DRAM refreshes?

Command bus utilization: 0.512%.

Explanation. The rows that are accessed every 1 ms and every 60 ms do *not* need to be refreshed.

The remaining rows (2^{17}) are accessed once every two refresh intervals (128 ms). So, the command bus utilization is $\frac{2^{17} * 5ns * 100}{128ms} = 0.512\%$

1.4 Refresh Policy D [20 points]

Assume that the memory controller implements an approximate mechanism to reduce refresh rate using Bloom filters, as we discussed in class. For this question we assume the retention times in Table 1 with a constant temperature of $64°C$.

One Bloom filter is used to represent the set of all rows that require a 64 ms refresh rate.

The memory controller's refresh logic is modified so that on every potential refresh of a row (every 64 ms), the refresh logic probes the Bloom filter. If the Bloom filter probe results in a "hit" for the row address, then the row is refreshed. Any row that does *not* hit in the Bloom filter is refreshed at the default rate of once per 128 ms.

(a) [20 points] The memory controller performs 2107384 refreshes in total across the channel over a time interval of 1.024 seconds. What is the false positive rate of the Bloom filter? (NOTE: False positive rate = Total number of false positives / Total number of accesses).

- Hint: $2107384 = 2^3 * (2^{18} + 2^{11} - 2^{10} + 2^9 - 2^8 - 1)$

False Positive rate = $(2^{10} - 1)/2^{18}$.

Explanation. First, there are 2^8 rows that need to be refreshed every 64ms. All the other rows ($2^{18} - 2^8$) will be refreshed every 128ms.

At 64 ms, we have 2^8 rows which actually require this rate, plus P out of the $2^{18} - 2^8$ other rows which will be false-positives. In a period of 1.024s, this bin is refreshed 16 times. The number of refreshes for this bin is $16 * (2^8 + P * (2^{18} - 2^8))$.

At 128 ms we have the rest of the rows which are $2^{18} - 2^8$ minus the false positives refreshed in the 64ms bin. So, the number of rows are $2^{18} - 2^8 - P * (2^{18} - 2^8)$. In a period of 1.024s, this bin is refreshed 8 times. The number of refreshes for this bin is $8 * ((2^{18} - 2^8) - (P * (2^{18} - 2^8))$

Thus, in total we have $16 * (2^8 + P * (2^{18} - 2^8)) + 8 * ((2^{18} - 2^8) - P * (2^{18} - 2^8)) =$ 2107384 . Solving the equation, P=2^{-8}.

Finally, False positive rate = Total number of false positives / Total number of accesses = $P * (2^{18} - 2^8))/2^{18} = 2^{-8} * (2^{18} - 2^8)/2^{18} = (2^{10} - 1)/2^{18}$