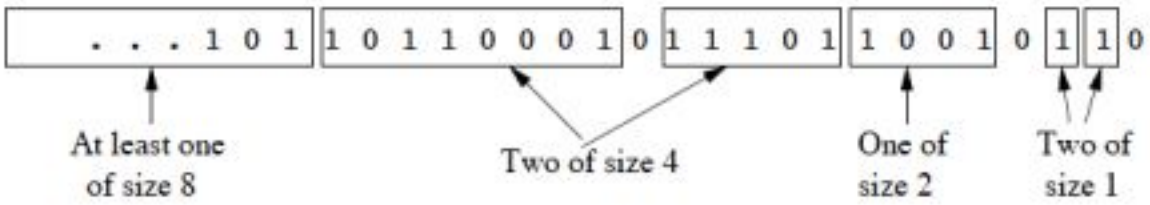
Tuesday Quiz

1. [1+1 points] In DGIM algorithm, what is the upper bound on the error rate? Explain Why.
   The upper bound on the error rate in DGIM algorithm is 50%.

   Assume that the last bucket is of size 2^r. Then, by assuming that 2^r-1 of its 1s are within the last bucket, the error is at most 2^r-1.

   There is at least 1 bucket of each of the sizes smaller than 2^r, the sum of theirs sizes is -
   1 + 2 + 4 + ... + 2^r-1 = (2^r)-1.
   Hence, the error is at most 50%

2. [1+1 points] What are the two types of queries in the data streaming model? Explain each with an example.
   Standing queries: Executed whenever a new tuple arrives. E.g: report each new maximum value ever seen in the stream
   Ad-hoc queries: Normal queries asked one time for a particular purpose. E.g., what is the maximum value so far?
3. [4 points] Figure shows a bit stream divided into buckets in a way that satisfies the DGIM rules.
   According to the DGIM algorithm, estimate the number of 1's the last k positions (where the rightmost position is the last) for k = (a) 8 (b) 19. In each case, how far off the correct value is your estimate?



   a) [2 points] k=8
   b) [2 points] k=19

   Ans: a) Estimated number of 1's in the last k=8 position: 4
   Actual number of 1's: 4
   Estimated is the same as actual. So it is far off by 0%
   b) Estimated number of 1's in the last k=19 position: 10
   Actual number of 1's: 10
   Estimated is the same as actual. So it is far off by 0%

   For part a, the estimated number of buckets is not 3. Since for k=8, all the buckets are completely covered, hence we do not divide by 2. There is no partially covered bucket in this case.

4. [2 points] Stream: a, b, c, b, d, a, c, d, a, b, d, c, a, a, b. Use AMS to calculate the surprise number for this stream.
   n = stream size = 15

   E[f(X)] = (1/15) {[15(1+3+5+7+9)] + [15(1+3+5+7)] + [15(1+3+5)] + [15(1+3+5)]}

   = (1/15) * 885

   = 59