Topic: External memory sorting with polyphase merging, Reynolds's paper.[1]

1. Fill in the blank entries in the following tables, indicating the number of runs on each of the five tapes used in a *polyphase merge-sort of order 4*. Row n of each table summarizes the distribution of runs on the tapes immediately following the nth merge, with the 0th row summarizing the initial distribution of runs (before any merges). Leave space in each cell to answer Question 2.

	# runs on tape				
m	1	2	3	4	5
0	8 (1)	8 (1)	7 (1)	7 (1)	0
1	1 (1)	1 (1)	0	0	7 (4)
2	0	0	1 (6)	0	6 (4)
3	1 (10)	0	0	0	5 (4)
4	0	1 (14)	0	0	4 (4)
5	1 (18)	0	0	0	3 (4)
6	0	1 (22)	0	0	2 (4)
7	1 (26)	0	0	0	1 (4)
8	0	1 (30)	0	0	0

	# runs on tape				
m	1	2	3	4	5
0	10 (1)	9 (1)	5 (1)	6 (1)	0
1	5 (1)	4 (1)	0	1 (1)	5 (4)
2	4 (1)	3 (1)	1 (7)	0	4 (4)
3	3 (1)	2 (1)	0	1 (13)	3 (4)
4	2 (1)	1 (1)	1 (19)	0	2 (4)
5	1	0	0	1 (25)	1 (4)
6	0	1 (30)	0	0	0

2. Augment the entries in the tables of Question 1 by adding (parenthesized) the sizes of the sorted runs in each cell, assuming all initial runs have 1000 records.

   ⓐ *The parenthesized numbers in the tables are in units of 1000 records.*

3. Using the definition in Reynolds's paper, list the first 20 k-generalized Fibonacci numbers for $k = 2, 3, 4, 5$.

   ⓐ

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
0	1	1	2	3	5	8	13	21	34	55	89	144	233	377	610	987	1597	2584	4181
0	0	1	1	2	4	7	13	24	44	81	149	274	504	927	1705	3136	5768	10609	19513
0	0	0	1	1	2	4	8	15	29	56	108	208	401	773	1490	2872	5536	10671	20569
0	0	0	0	1	1	2	4	8	16	31	61	120	236	464	912	1793	3525	6930	13624

4. Using the method suggested by Reynolds's paper, determine the initial distribution of 82 runs on 6 tapes for a 5-way polyphase merge. That is, indicate the number of runs initially written to each of the tapes, numbered 1 through 6. Show the intermediate steps used in arriving at the final distribution. Then indicate the result of each merge step in tabular form, as in Question 1, until only one run remains.

[1] Samuel W. Reynolds, "A Generalized Polyphase Merge Algorithm," *Communications of the ACM* 4/8 (1961).

ⓐ *Determination of perfect run distributions for a 5-way polyphase merge using the 5th-order Fibonacci numbers (left), and run distribution (right):*

$n:$	$[0,3]$	4	5	6	7	8	9
$F^{(5)}(n):$	0	1	1	2	4	8	16
$j:$		5	6	7	8	9	10
$c_5:$		1	1	2	4	8	16
$c_4:$		1	2	3	6	12	24
$c_3:$		1	2	4	7	14	28
$c_2:$		1	2	4	8	15	30
$c_1:$		1	2	4	8	16	31
$\sum_i c_i:$		5	9	17	33	65	129

	runs on tape						*total runs*	
step	1	2	3	4	5	6	*writ.*	*rem.*
1	1	1	1	1	1	0	5	77
2	2	2	2	2	1	0	9	73
3	4	4	4	3	2	0	17	65
4	8	8	7	6	4	0	33	49
5	16	15	14	12	8	0	65	17
6	31	17	14	12	8	0	82	0

Using Reynolds's notation, we have $c_1 = 16$, $E = 2$, $c_E = 15$,*and* $x = 17-15 = 2$, *with* $c_1 - c_E = 16 - 15 = 1$. *Thus we have* $2 \le E \le k-1$ *and* $c_1 - c_E < x < c_1 + c_{E+1} - c_E$.

These values lead to the case summarized by the box in the bottom-right corner of the flowchart in the paper. Proceeding as directed by the flowchart in that case, we merge $c_1 = 16$ *runs from tapes 1 and 2 on to tape 6 giving the following distribution of runs on tapes 1 through 6:* $(15, 1, 14, 12, 8, 16)$.

We pass $c_E + x - c_1 = 15+2-16 = 1$ *run from tape 2 to tape 6, giving:* $(15, 0, 14, 12, 8, 17)$.

We then merge $c_E + x - c_1 = 1$ *runs from tapes 6 and 3 (2-way merge) to tape 2:* $(15, 1, 13, 12, 8, 16)$.

Finally, we pass $c_{E+1} - c_E - x + c_1 = 14 - 15 - 2 + 16 = 13$ *runs from tape 3 to tape 2:* $(15, 14, 0, 12, 8, 16)$.

We are left with a good (k-Fibonacci-based) distribution of runs depicted to the right. (The sequence of runs on tapes 6, 1, 2, 4, and 5 is identical to the sequence in step 5 of the earlier distribution table.) Normal polyphase merging can now begin as a five-way merge with results initially written to tape 3. Further, each merging stage uses all available tapes (5-way merge) unlike what happens when the initial distribution of runs is selected naively.

1	2	3	4	5	6
15	14	0	12	8	16
7	6	8	4	0	8
3	2	4	0	4	4
1	0	2	2	2	2
0	1	1	1	1	1
1	0	0	0	0	0