Final In-Class Exam  
(4:05 PM – 5:20 PM : 75 Minutes)

- This exam will account for either 15% or 30% of your overall grade depending on your relative performance in the midterm and the final. The higher of the two scores (midterm and final) will be worth 30% of your grade, and the lower one 15%.

- There are three (3) questions, worth 75 points in total. Please answer all of them in the spaces provided.

- There are 14 pages including four (4) blank pages and one (1) page of appendices. Please use the blank pages if you need additional space for your answers.

- The exam is open slides and open notes.

**Good Luck!**

<table>
<thead>
<tr>
<th>Question</th>
<th>Pages</th>
<th>Score</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Parallel DFT</td>
<td>2–5</td>
<td></td>
<td>30</td>
</tr>
<tr>
<td>2. Trapping the Median</td>
<td>7–9</td>
<td></td>
<td>30</td>
</tr>
<tr>
<td>3. Files on Compact Discs</td>
<td>12</td>
<td></td>
<td>15</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td></td>
<td>75</td>
</tr>
</tbody>
</table>
**Question 1. [30 Points] Parallel DFT.** Given the coefficient vector \(\langle a_0, a_1, \ldots, a_{n-1}\rangle\) of a polynomial \(P(x) = a_0 + a_1 x + a_2 x^2 + \ldots + a_{n-1} x^{n-1}\), the PAR-REC-DFT function shown below (in Figure 1) computes another vector \(\langle y_0, y_1, \ldots, y_{n-1}\rangle\), where \(y_i = P((\omega_n)^i)\) and \(\omega_n\) is the primitive \(n\)-th root of unity. The output vector \(\langle y_0, y_1, \ldots, y_{n-1}\rangle\) is called the Discrete Fourier Transform (DFT) of the input vector \(\langle a_0, a_1, \ldots, a_{n-1}\rangle\). We assume for simplicity that \(n\) is a power of 2.

```
PAR-REC-DFT( \langle a_0, a_1, \ldots, a_{n-1}\rangle )

(Input is the coefficient vector \(\langle a_0, a_1, \ldots, a_{n-1}\rangle\) of a polynomial \(P(x) = a_0 + a_1 x + a_2 x^2 + \ldots + a_{n-1} x^{n-1}\). The output is another vector \(\langle y_0, y_1, \ldots, y_{n-1}\rangle\), where \(y_i = P((\omega_n)^i)\) and \(\omega_n\) is the primitive \(n\)-th root of unity. We assume for simplicity that \(n\) is a power of 2.)

1. if \(n = 1\) then return \(\langle a_0\rangle\)
2. else
3. \(\langle y_{0\text{even}}, y_{1\text{even}}, \ldots, y_{\frac{n}{2}-1\text{even}}\rangle \leftarrow \text{spawn} \text{PAR-REC-DFT}( \langle a_0, a_2, \ldots, a_2n-2\rangle )\) \{even numbered coefficients\}
4. \(\langle y_{0\text{odd}}, y_{1\text{odd}}, \ldots, y_{\frac{n}{2}-1\text{odd}}\rangle \leftarrow \text{PAR-REC-DFT}( \langle a_1, a_3, \ldots, a_{2n-1}\rangle )\) \{odd numbered coefficients\}
5. sync
6. \(w_0 \leftarrow 1\)
7. parallel for \(j \leftarrow 1\) to \(\frac{n}{2} - 1\) do
8. \(w_j \leftarrow n\)-th primitive root of unity \{i.e., \(w_j \leftarrow e^{\frac{2\pi i}{n}}\), where \(i = \sqrt{-1}\)\}
9. \(\langle s_0, s_1, \ldots, s_{\frac{n}{2}-1}\rangle \leftarrow \text{PREFIX-SUM}( \langle w_0, w_1, \ldots, w_{\frac{n}{2}-1}\rangle, \times \rangle\) \{prefix sum using the product operator\}
10. parallel for \(i \leftarrow 0\) to \(\frac{n}{2} - 1\) do
11. \(y_j \leftarrow y_{j\text{even}} + s_j y_{j\text{odd}}\)
12. \(y_{\frac{n}{2}+j} \leftarrow y_{j\text{even}} - s_j y_{j\text{odd}}\)
13. return \(\langle y_0, y_1, \ldots, y_{n-1}\rangle\)
```

Figure 1: A parallel recursive divide-and-conquer algorithm for computing the Discrete Fourier Transform (DFT) of a 1D array (vector).
1(a) [10 Points] Write down a recurrence relation describing the work done (i.e., $T_1$) by PAR-REC-DFT, and solve it.
1(b) [10 Points] Write down a recurrence relation describing the span (i.e., $T_\infty$) of PAR-REC-DFT, and solve it. Please assume that the span of a parallel for loop with $n$ iterations is $\mathcal{O}(\log n + k)$, where $k$ is the maximum span of a single iteration.
1(c) [10 Points] Find the parallel running time (i.e., $T_p$) and parallelism of PAR-REC-DFT.
**Question 2. [30 Points] Trapping the Median.** Given an array \( A[1 : n] \) of \( n \) distinct numbers as input, the function TRAP-MEDIAN shown below (in Figure 2) returns another array \( A'[1 : n'] \) containing \( n' \) distinct numbers from \( A \) such that w.h.p. in \( n, n' = \mathcal{O}\left(\frac{n^2}{4}\right) \) and \( A' \) still includes the median of \( A \). We assume for simplicity that \( n \) is an odd positive integer.

```
TRAP-MEDIAN(A, n)
(Input is an array \( A[1 : n] \) of \( n \) distinct numbers, where \( n \) is an odd positive integer. Output is an array \( A'[1 : n'] \) containing \( n' \) distinct numbers from \( A \) such that w.h.p. in \( n, n' = \mathcal{O}\left(\frac{n^2}{4}\right) \) and \( A' \) contains the median of \( A \).)

1. choose each entry of \( A \) with probability \( \frac{n - 1}{4} \) independent of others, and collect them in an array \( B \)
2. \( m \leftarrow |B| \)
3. if \( \left\lceil \frac{m}{2} - \sqrt{n} \right\rceil > 0 \) and \( \left\lceil \frac{m}{2} + \sqrt{n} \right\rceil \leq m \) then
4. sort \( B \) using an optimal sorting algorithm
5. \( x \leftarrow B \left\lfloor \frac{m}{2} - \sqrt{n} \right\rfloor, \ y \leftarrow B \left\lceil \frac{m}{2} + \sqrt{n} \right\rceil \)
6. \( r_x \leftarrow \)number of items in \( A \) with value \( \leq x \)
7. \( r_y \leftarrow \)number of items in \( A \) with value \( \leq y \)
8. if \( r_x < \frac{n + 1}{2} < r_y \) then \( \{ \text{if } x \text{ is smaller than the median of } A, \text{ and } y \text{ is larger than the median} \} \)
9. \( n' \leftarrow \)number of items in \( A \) with value between \( x \) and \( y \) \( \{ \text{count each } z \text{ in } A \text{ with } x < z < y \} \)
10. allocate an array \( A'[1 : n'] \)
11. scan \( A \) again, and copy each number \( z \in (x, y) \) from \( A \) to \( A' \)
12. return \( A' \)
13. else return nil
14. else return nil
```

Figure 2: Trap the median of \( n \) numbers in a set of size asymptotically smaller than \( n \).

2(a) [12 Points] Prove that \( n^\frac{3}{4} - n^\frac{7}{16} < m < n^\frac{3}{2} + n^\frac{7}{16} \) holds w.h.p. in \( n \) (in Step 2).
2(b) [12 Points] Show that \( r_x < \frac{n+1}{2} < r_y \) holds w.h.p. in \( n \) (in Step 8). You may assume that 
\( m = \Theta \left( n^{\frac{3}{4}} \right) \) holds w.h.p. in \( n \) (from part 2(a)).
2(c) [6 Points] Show that the running time of Trap-Median is $\mathcal{O}(n)$ w.h.p. in $n$. You may use the results you proved in parts 2(a) and 2(b), if needed.
Use this page if you need additional space for your answers.
Use this page if you need additional space for your answers.
Question 3. [15 Points] Files on Compact Discs. I have $m > 0$ files and a set $S$ of $n > 1$ compact discs (CDs). I have copied each file to exactly two of the CDs in $S$. Different files may be copied to different CD pairs. Now given that for each file I know the two CDs I copied them to, I want to find a subset $S' \subseteq S$ such that each file is contained in at least one CD of $S'$, and $|S'|$ is as small as possible.

3(a) [15 Points] Give a polynomial-time 2-approximation algorithm for solving this problem. In other words, the size of the subset returned by your algorithm must not be more than 2 times larger than the size of the subset returned by an optimal algorithm.
Use this page if you need additional space for your answers.
APPENDIX I: USEFUL TAIL BOUNDS

Markov’s Inequality. Let $X$ be a random variable that assumes only nonnegative values. Then for all $\delta > 0$, $Pr[X \geq \delta] \leq \frac{E[X]}{\delta}$.

Chebyshev’s Inequality. Let $X$ be a random variable with a finite mean $E[X]$ and a finite variance $Var[X]$. Then for any $\delta > 0$, $Pr[|X - E[X]| \geq \delta] \leq \frac{Var[X]}{\delta^2}$.

Chernoff Bounds. Let $X_1, \ldots, X_n$ be independent Poisson trials, that is, each $X_i$ is a 0-1 random variable with $Pr[X_i = 1] = p_i$ for some $p_i$. Let $X = \sum_{i=1}^{n} X_i$ and $\mu = E[X]$. Following bounds hold:

Lower Tail:
- for $0 < \delta < 1$, $Pr[X \leq (1 - \delta)\mu] \leq \left(\frac{e^{\delta}}{1-\delta}(1-\delta)^{\delta}\right)^{\mu}$
- for $0 < \delta < 1$, $Pr[X \leq (1 - \delta)\mu] \leq e^{-\frac{\mu^2}{2}}$
- for $0 < \gamma < \mu$, $Pr[X \leq \mu - \gamma] \leq e^{-\frac{\gamma^2}{2\mu}}$

Upper Tail:
- for any $\delta > 0$, $Pr[X \geq (1 + \delta)\mu] \leq \left(\frac{e^{\delta}}{1+\delta}(1+\delta)^{\delta}\right)^{\mu}$
- for $0 < \delta < 1$, $Pr[X \geq (1 + \delta)\mu] \leq e^{-\frac{\mu^2}{4}}$
- for $0 < \gamma < \mu$, $Pr[X \geq \mu + \gamma] \leq e^{-\frac{\gamma^2}{2\mu}}$

APPENDIX II: THE MASTER THEOREM

Let $a \geq 1$ and $b > 1$ be constants, let $f(n)$ be a function, and let $T(n)$ be defined on the nonnegative integers by the recurrence

$$T(n) = \begin{cases} \Theta(1), & \text{if } n \leq 1, \\ aT\left(\frac{n}{b}\right) + f(n), & \text{otherwise,} \end{cases}$$

where, $\left[\frac{n}{b}\right]$ is interpreted to mean either $\left\lfloor\frac{n}{b}\right\rfloor$ or $\left\lceil\frac{n}{b}\right\rceil$. Then $T(n)$ has the following bounds:

**Case 1:** If $f(n) = O(n^{\log_b a - \epsilon})$ for some constant $\epsilon > 0$, then $T(n) = \Theta(n^{\log_b a})$.

**Case 2:** If $f(n) = \Theta(n^{\log_b a \log^k n})$ for some constant $k \geq 0$, then $T(n) = \Theta(n^{\log_b a \log^{k+1} n})$.

**Case 3:** If $f(n) = \Omega(n^{\log_b a + \epsilon})$ for some constant $\epsilon > 0$, and $af\left(\frac{n}{b}\right) \leq cf(n)$ for some constant $c < 1$ and all sufficiently large $n$, then $T(n) = \Theta(f(n))$. 

14