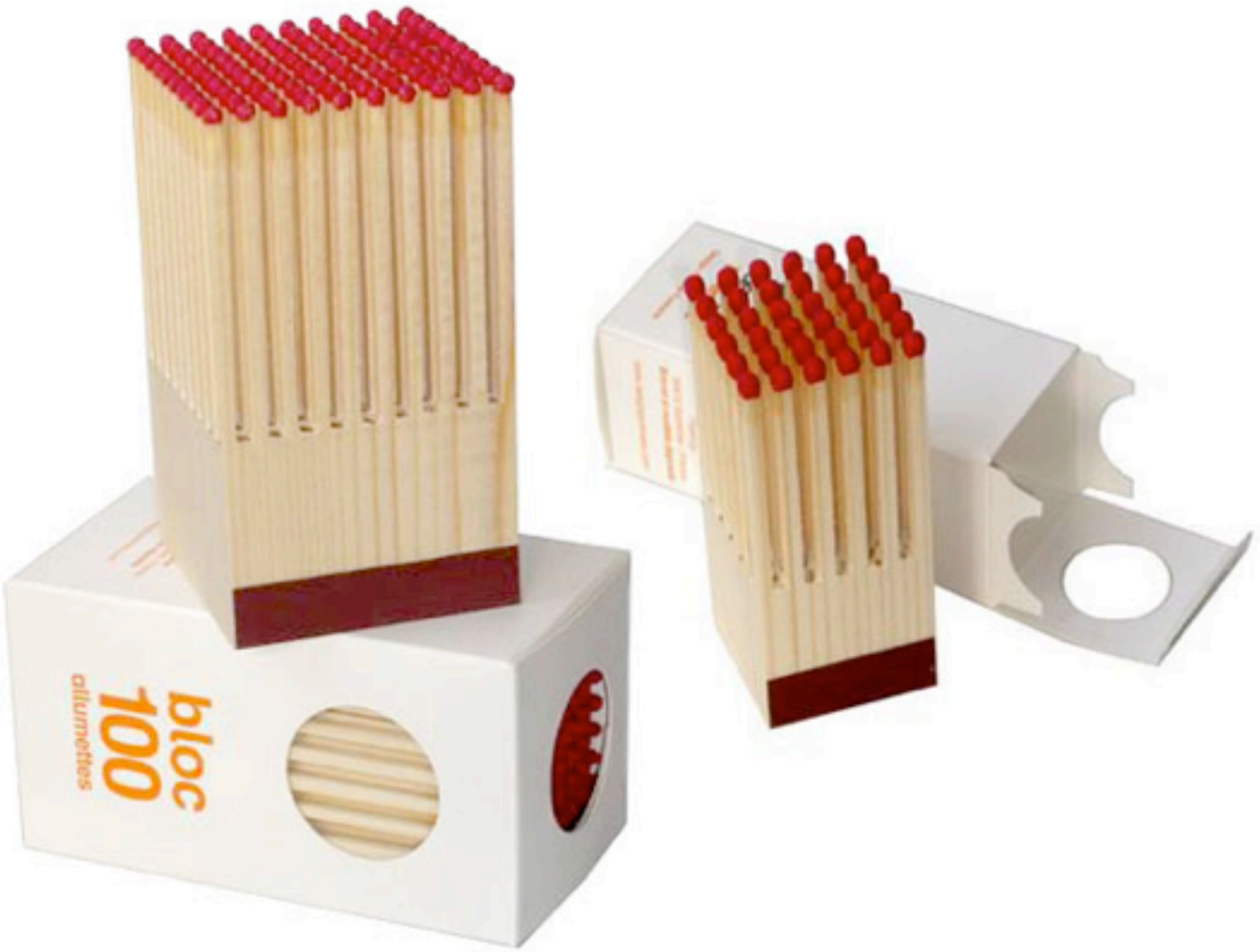
L27

4102

12.02.2013

abhi

randomization



http://kitsunenoir.com/blogimages/bloc-matches.jpg

What is your best strategy to survive & become King ??

CHECK PROCEDURE:

RANDOMLY PICK 50 MATCHES AND LIGHT THEM

IF ONE FAILS, REJECT THE BOX.

IF ALL SUCCEED, ACCEPT IT.

PR THAT TEST FAILS

THREE CASES TO CONSIDER:

Spse the uncle tampers with

< 50 matches ⇒ You camp + live. or you don't camp + live.

> 50 matches ⇒ Kill the uncle!! b/c you catch him

50 matches ⇒ one case of failure ⇒ you test pick the 50 good matches & camp w/ 50 bad ones.

What is the pr of failure ??

100 matches, 50 good ones. You pick those 50 good ones!!

$$\left(\frac{50}{100}\right)\left(\frac{49}{99}\right)\left(\frac{48}{98}\right)\cdots\left(\frac{1}{50}\right)$$

$$\frac{1}{\binom{100}{50}} \rightarrow$$

$$\boxed{\binom{2n}{n} > \frac{2^{2n-1}}{\sqrt{n}}}$$ via Stirling identity

$$\frac{1}{\binom{100}{50}} < 2^{-98}$$

PR OF DEATH:

9.911653021418339067376749696883601495412210270643283767892785256889073029997327393587632943101698342E-30

10^{-31}

PR OF DEATH:

9.91165302141833906737674969688360149
5412210270643283767892785256889073029
997327393587632943101698342E-30

PR OF ROYAL FLUSH:

1.53908E-6

10^{-5}

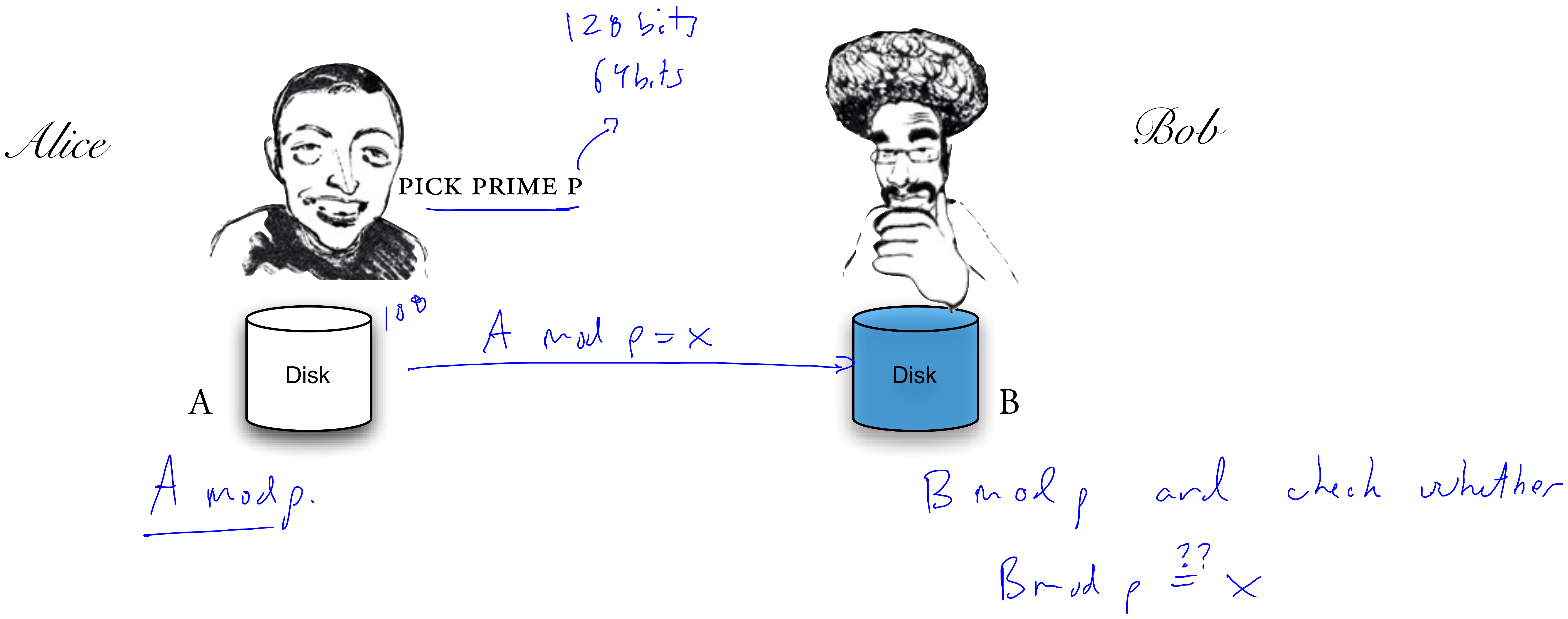
PR THAT YOU...

Age in 1990	Total U.S.	White Male	White Female	Black Male	Black Female
20	0.102%	0.128%	0.045%	0.307%	0.074%

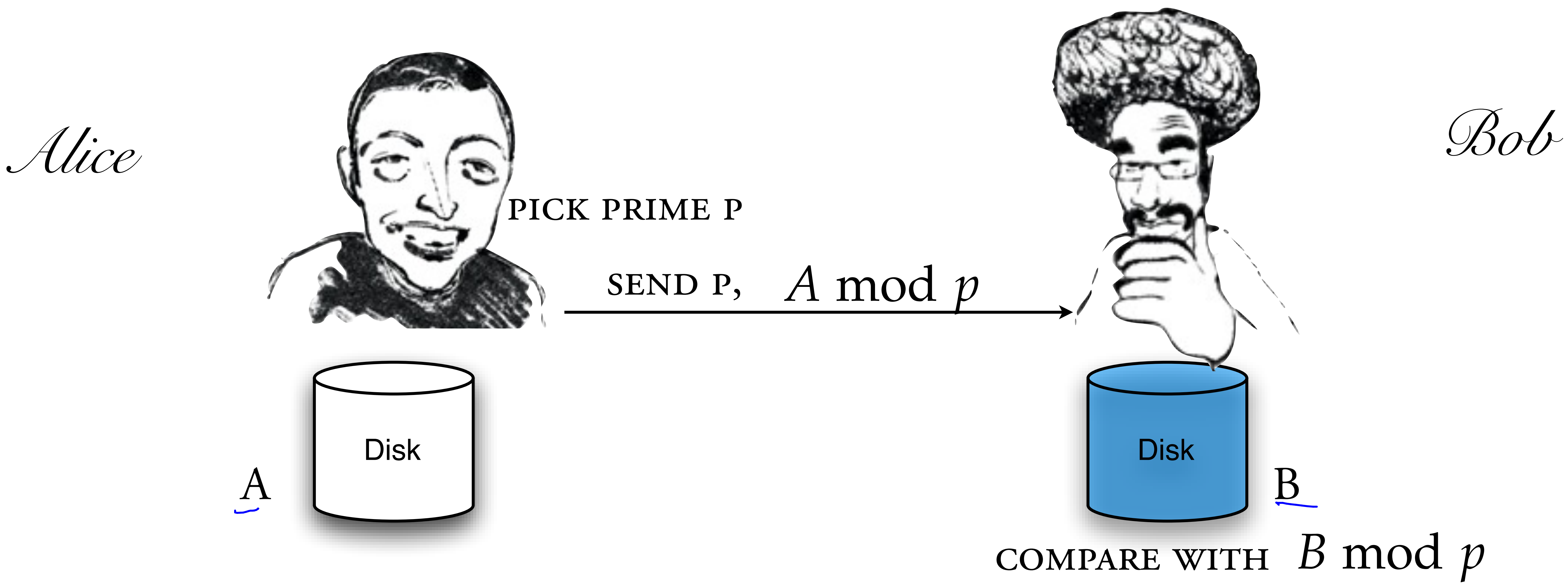
10^{-3}

FINGERPRINTING

FINGERPRINTING



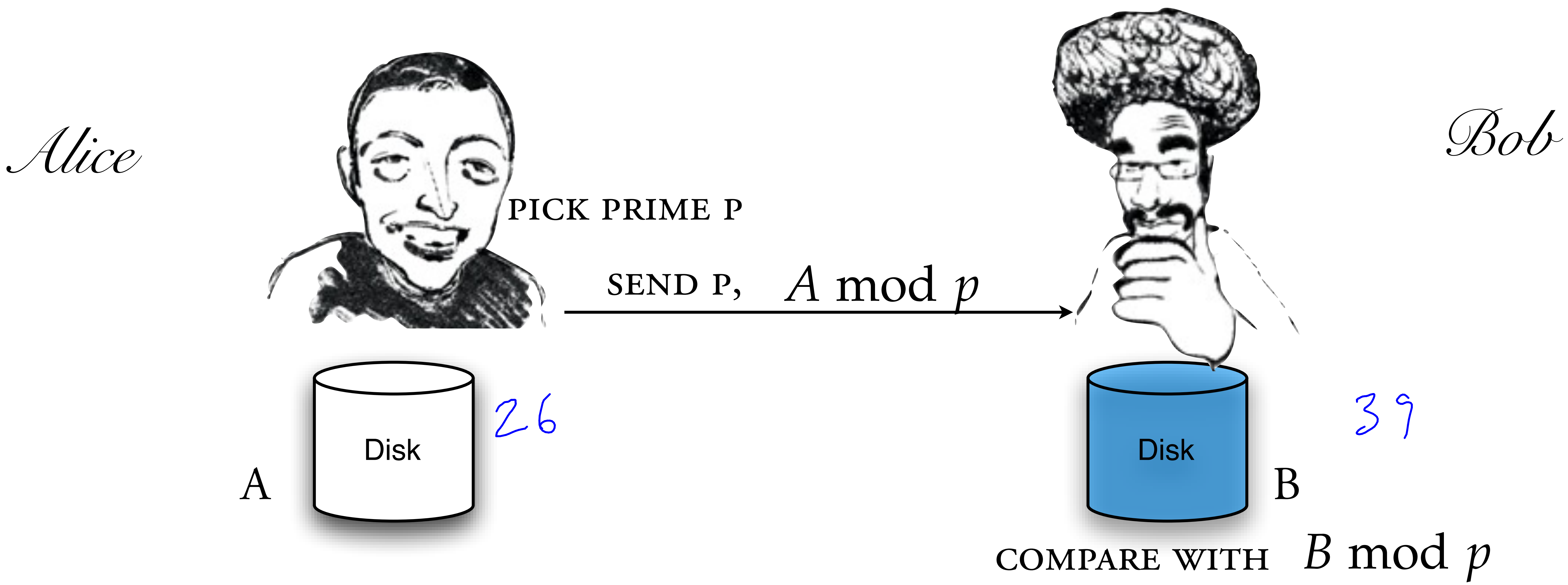
FINGERPRINTING



IF A=B, THEN this protocol certainly succeeds!!

FINGERPRINTING

$p = 13.$



IF $A \neq B$ THEN there is some small chance that the protocol makes an error and convinces Alice Bob that their disks are equal.

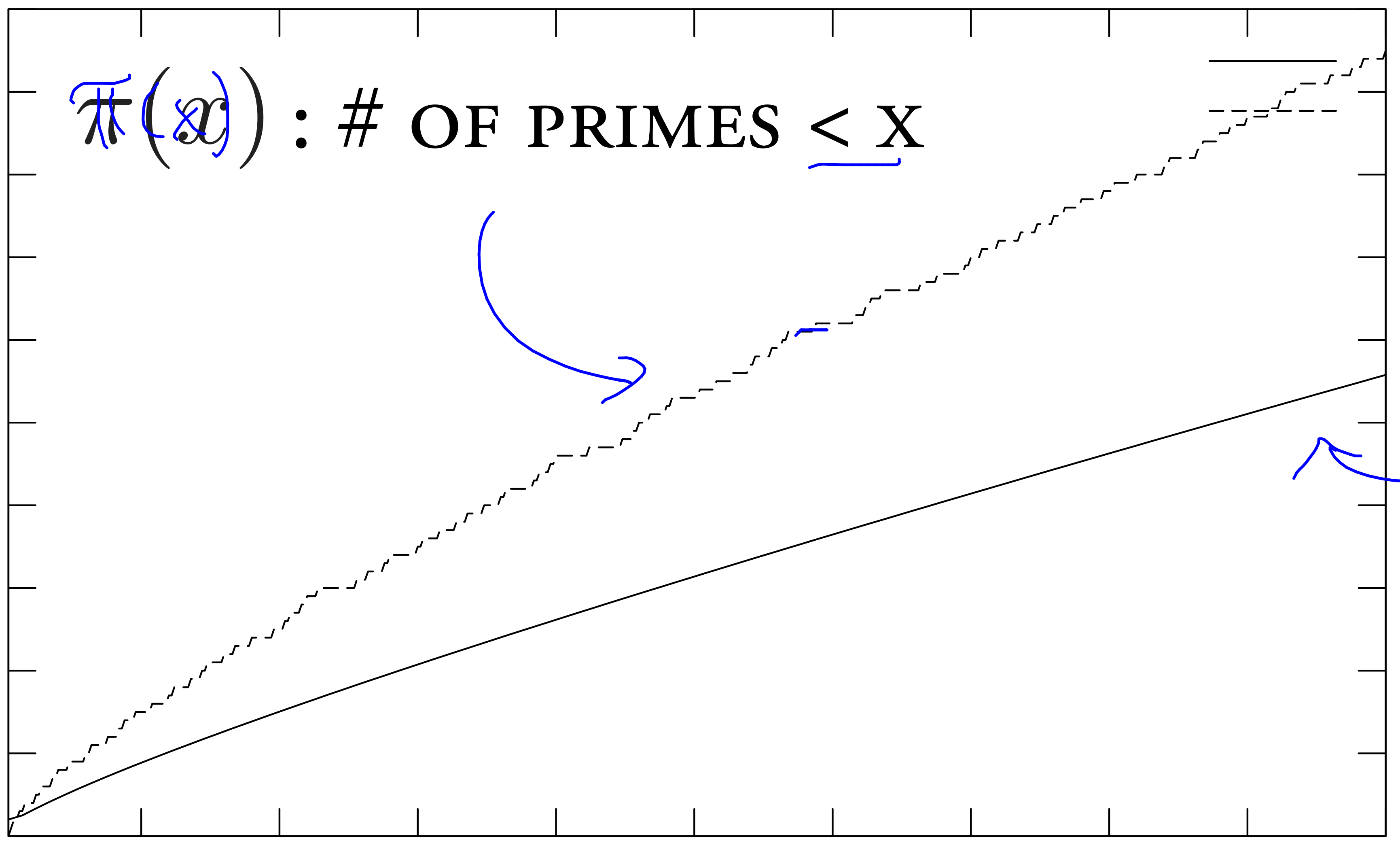
NUMBER OF PRIMES

∞ EUCLID.

how many primes are there that are $< 2^{64}$

NUMBER OF PRIMES

THERE ARE CERTAINLY INFINITELY MANY



$\pi(x)$: #

$$\pi(x) > \frac{x}{\log x}$$

$\frac{x}{\log x}$

$x \sim 2^{64}$

$$\pi(2^{64}) > \frac{2^{64}}{64} > 2^{58}$$

$\sim 2^{6}$

$$\pi(2^{128}) > \frac{2^{128}}{128} \sim 2^{121}$$

LEMMA: # OF PRIME DIVISORS OF X < LOG(X)

$\log_2(x)$

2 is the smallest prime.

if x has t prime divisors than

$$x \geq 2^t . \qquad \square$$

PR OF FALSE MATCH:

Spse that $A \neq B$, but the protocol concludes "match".

$A \bmod p = B \bmod p$

$\Rightarrow (A-B) = 0 \bmod p \Rightarrow p$ divides $(A-B)$

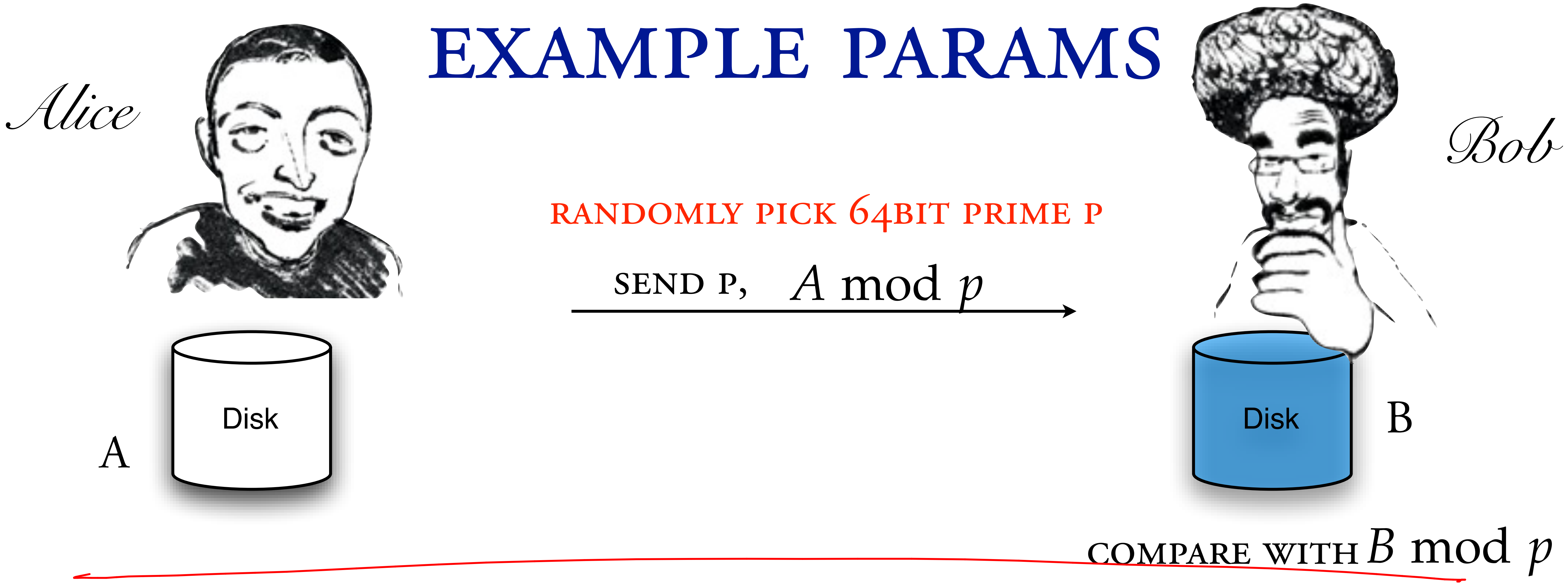
$\Rightarrow$ how many prime divisors can $(A-B)$ have.?

$\log(A-B) \sim \underline{\log(2^{2^{40}})} \sim \underline{2^{40}}$

$\Rightarrow$ how many 64 bit primes are there?? $\underline{2^{58}}$.

Pr failure is $< \frac{2^{40}}{2^{58}} \sim \underline{2^{-18}}$

EXAMPLE PARAMS



STRING MATCHING

PATTERN

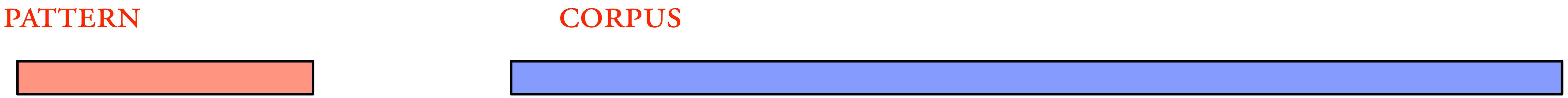
Netflix

CORPUS

A squabble between a group fighting spam and a Dutch company that hosts Web sites said to be sending spam has escalated into one of the largest computer attacks on the Internet, causing widespread congestion and jamming crucial infrastructure around the world. Millions of ordinary Internet users have experienced delays in services like Netflix or could not reach a particular Web site for a short time.

However, for the Internet engineers who run the global network the problem is more worrisome. The attacks are

STRING MATCHING

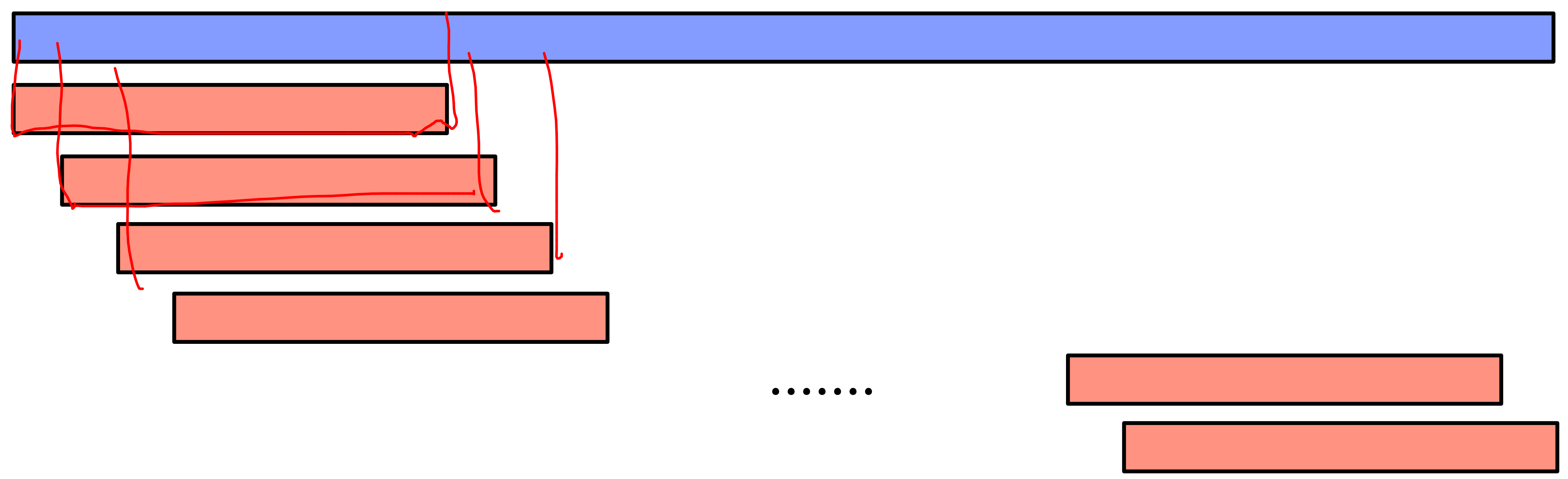


STRING MATCHING

O(m)

O(n)

BRUTE FORCE:



```
for (int i = 0, j=0; i < n-m; i++) {
   while (j < m && t[i+j] == p[j]) { j++; }
   if (j == m) return i;
}
return -1;
```

Running time $O(n \cdot m)$

SIMPLE ALGORITHM

AAAAAAAAAAAAAAAAAAAAAAAAAA

AAAAAAB

BRUTE FORCE WORST CASE:

KMP ALGORITHM

ABCDABCDABCDEFH

ABCDABHI

KMP ALGORITHM

ABCDABCDABCDEFH

ABCDABHI

SLIDING RULE

given that P[1....q] matches T[m+1...m+q], but a mismatch occurs at q+1, then:

Text

SLIDING RULE

Given that P[1....q] matches T[m+1...m+q], but a mismatch occurs at q+1, then:

Find the longest prefix of P[1...q] that is also a suffix of P[1...q]

Slide so that P[1...p] matches T[1-p+1,...]

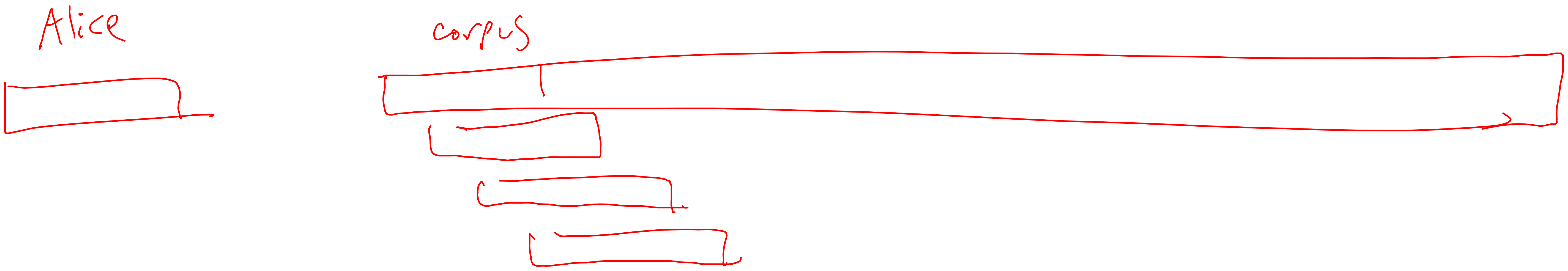
$\Theta(n+m)$

1	2	3	4	5	6	7	8	9	10	11
X	**Y**	**X**	**Y**	**Y**	**X**	**Y**	**X**	**Y**	**X**	**X**

Text

1	2	3	4	5	6	7	8	9	10	11
X	**Y**	**X**	**Y**	**Y**	**X**	**Y**	**X**	**Y**	**X**	**X**
0	0	1	2	0	1	2	3	4	3	1

NEW IDEA FOR STRING MATCH



STRING MATCHING

PICK RANDOM T-BIT PRIME

COMPUTE PATTERN MOD PRIME (Alice operation)

FOR I=1...N

COMPUTE NEXT CORPUS MOD PRIME

COMPARE, OUTPUT MATCH IF SAME

Bob's operation over sliding windows of size m

PICK AN 80-BIT PRIME P

WHAT IS THE PROBABILITY OF A MISMATCH AT THE FIRST POSITION?

128 bit primes.

if the pattern is of size m bits

pattern has $\leq m$ prime divisors.

$\Pr[\text{mismatch @ position } 0] \leq \frac{m}{2^{121}}$ → # of 128-bit primes.

$\Pr[\text{" " " } 1] \leq \frac{m}{2^{121}}$ ....

(small)

There are n positions to match, so $\Pr[\text{Alg fails}] \leq \frac{n \cdot m}{2^{121}}$

PR OF ANY MISMATCH:

STRING MATCHING

PICK RANDOM T-BIT PRIME

COMPUTE PATTERN MOD PRIME

FOR I=1...N

COMPUTE NEXT CORPUS MOD PRIME

COMPARE, OUTPUT MATCH IF SAME

STRING MATCHING EXAMPLE

PATTERN

26535

TEXT

31415926535897931231273 98

mod 13

= 7

p = 13

26535 mod 13 =

2

STRING MATCHING EXAMPLE

PATTERN

26535

TEXT

3141592653589793123127398

Given that 31415 mod 13 = 7,
How can I compute 14159 mod 13?

Hint: 10000 mod 13 = 3

STRING MATCHING EXAMPLE

PATTERN

26535

TEXT

314159265358979312312 7398

14159 =

STRING MATCHING

PICK RANDOM T-BIT PRIME

COMPUTE PATTERN MOD PRIME

FOR I=1...N

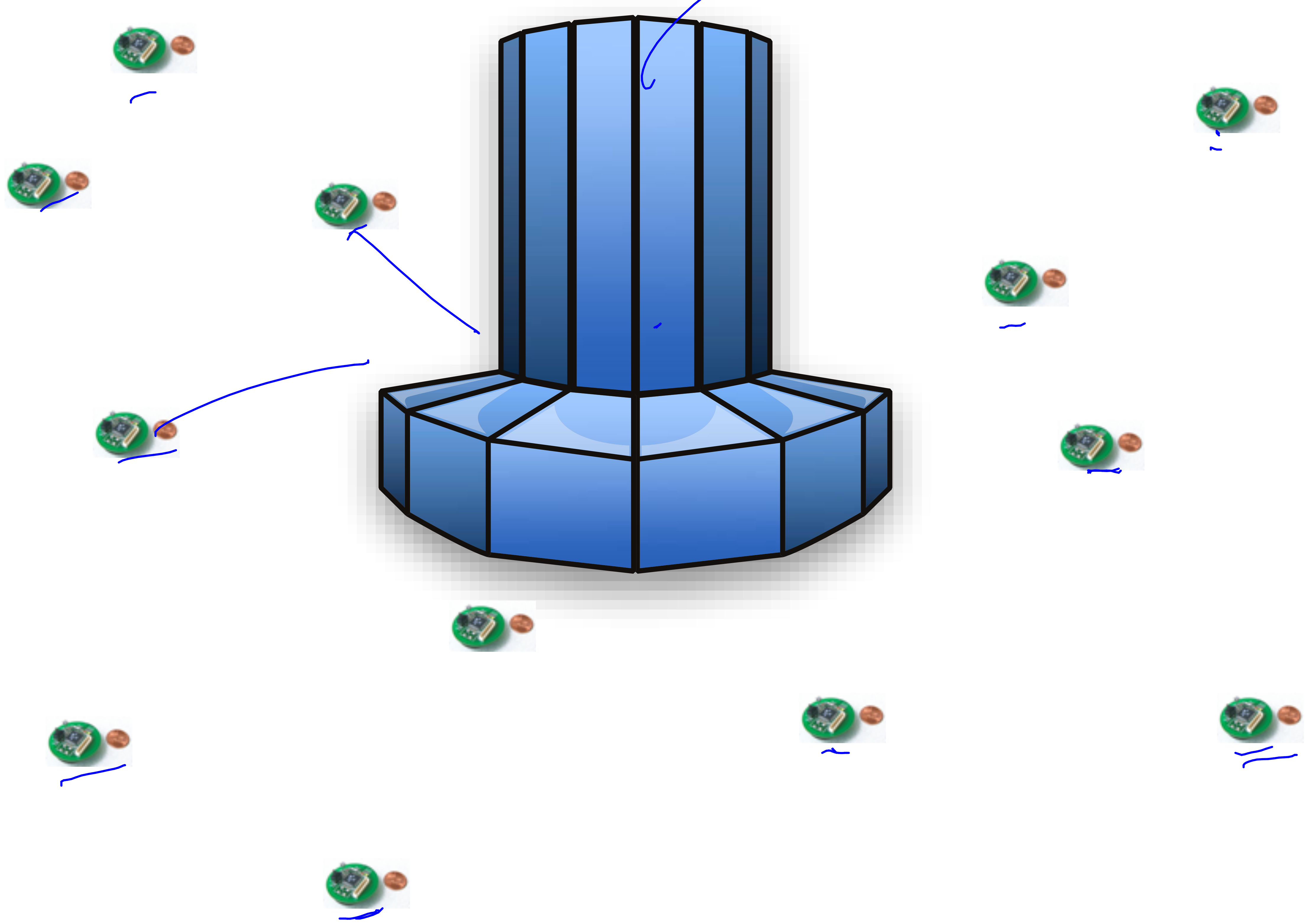
COMPUTE NEXT CORPUS MOD PRIME

COMPARE, OUTPUT MATCH IF SAME

} $\Theta(1)$ mod p operations

$\Theta(n+m)$ mod p operations.

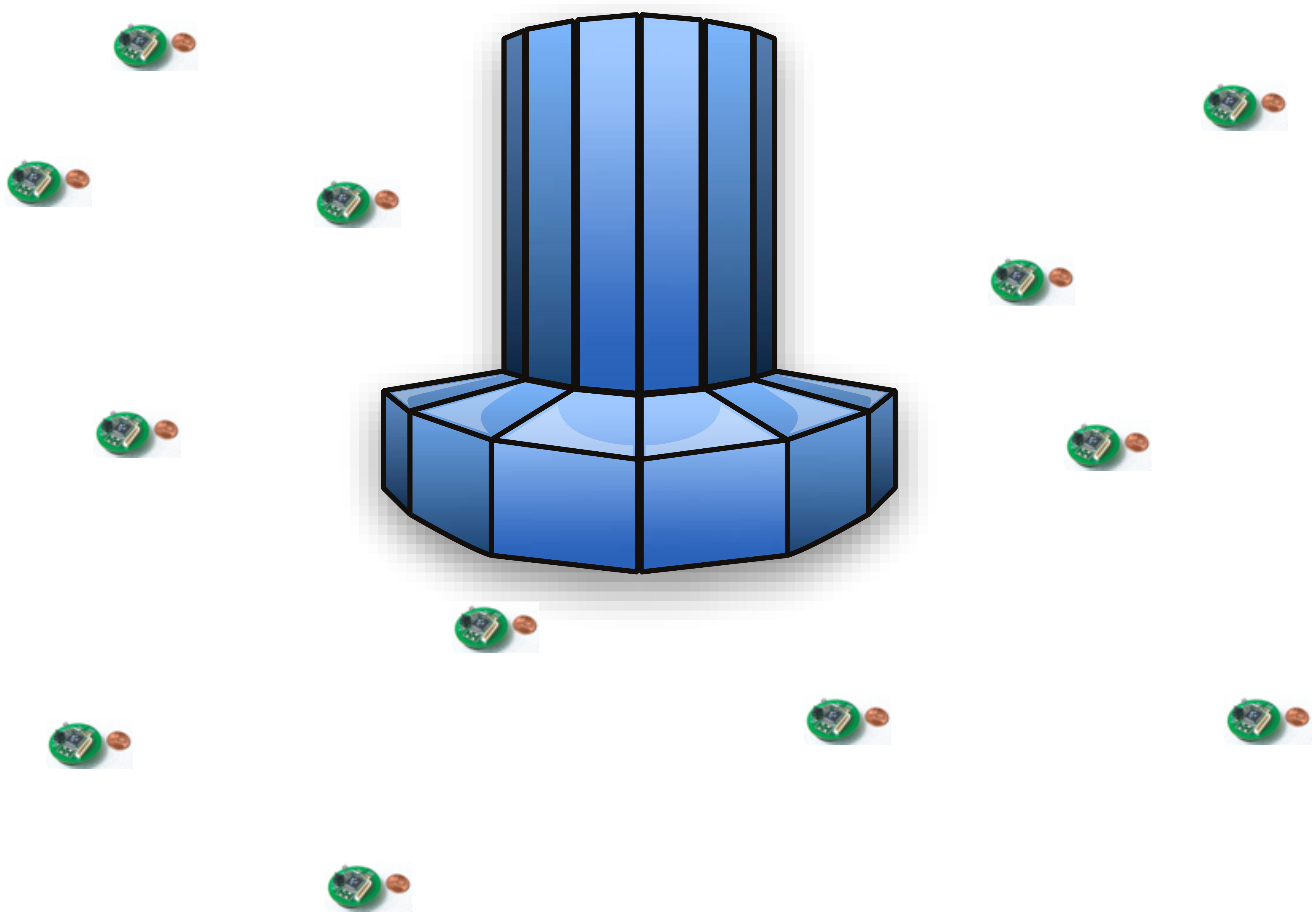
FIRST EXAMPLE



GOAL:

DEVISE A RELIABLE METHOD FOR NODES TO SEND MESSAGE TO THE SERVER WITH AS LITTLE COORDINATION AS POSSIBLE.

FIRST EXAMPLE



SIMPLE ALGORITHM

AT TIME T, FLIP A COIN THAT IS HEADS WITH PR $\frac{1}{n}$

IF HEADS, THEN BROADCAST. IF SUCCESS, THEN STOP.

ELSE WAIT AND TRY AGAIN.

REPEAT $cn \log n$ TIMES

ANALYZE THE SIMPLE ALGORITHM

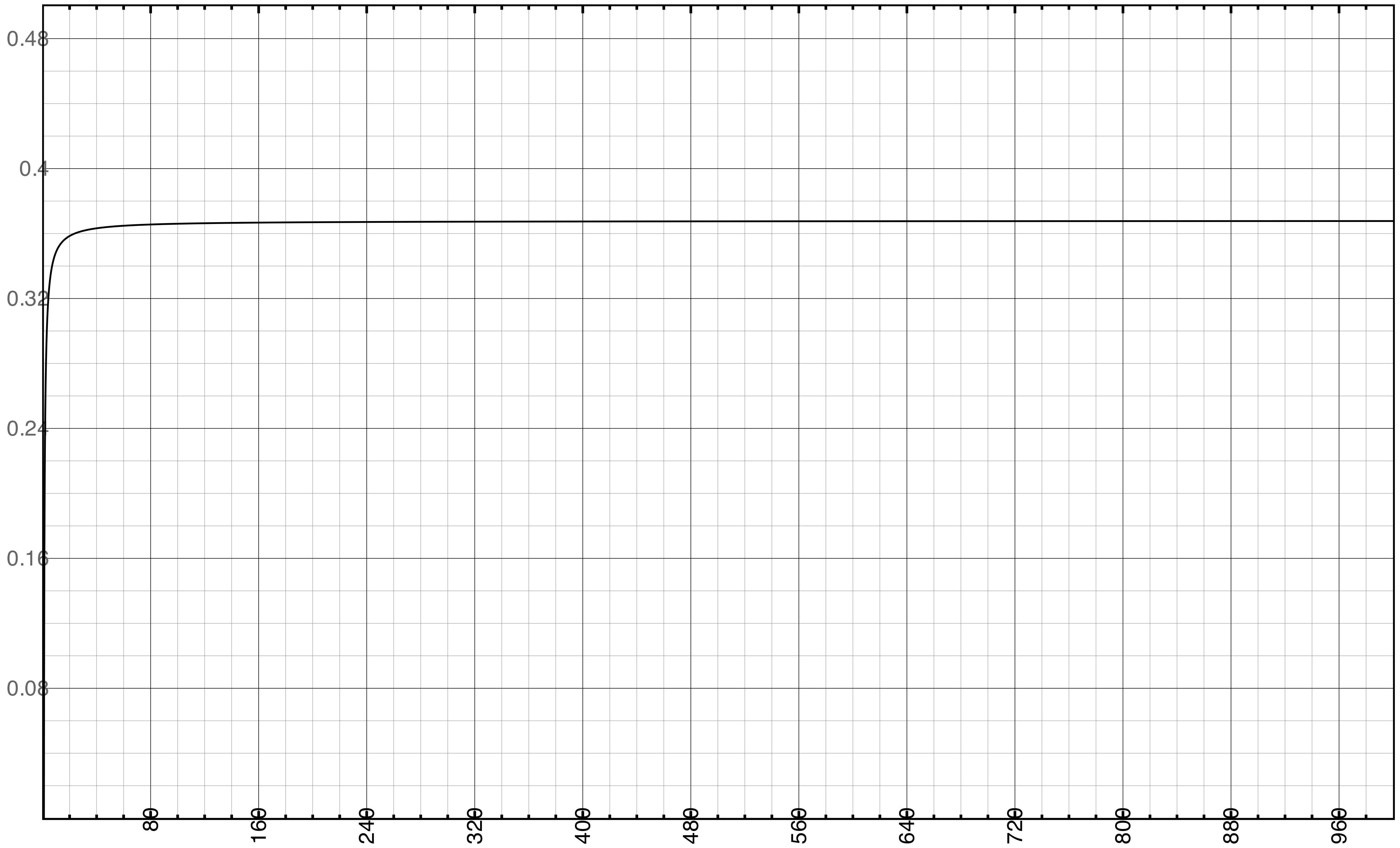
$$S_{i,t} =$$

$$\Pr[S_{i,t} = 1] =$$

$$\Pr[S_{i,t} = 1] = \frac{1}{n}\left(1 - \frac{1}{n}\right)^{n-1}$$

FACT: IF $f(n) = \left(1 - \frac{1}{n}\right)^n$ THEN

FACT: IF $f(n) = \left(1 - \frac{1}{n}\right)^n$ THEN



$$S_{i,t} = \text{node } i \text{ succeeds in sending at time } t$$

$$\frac{1}{en} \leq \Pr[S_{i,t} = 1] \leq \frac{1}{2n}$$

FAILURE

$$F_{i,t} =$$

FAILURE

$F_{i,t} =$ NODE i FAILS TO SEND AT TIMES 1,2,...,t

$$\Pr[F_{i,t}] = \bigwedge_{j=1}^{t} \Pr[\overline{S_{i,j}}]$$

FAILURE

$F_{i,t} =$ NODE i FAILS TO SEND AT TIMES 1,2,...,t

$$\Pr[F_{i,t}] = \bigwedge_{j=1}^{t} \Pr[\overline{S_{i,j}}] = \prod_{j=1}^{t} \Pr[\overline{S_{i,j}}]$$

$$\Pr[F_{i,t}] = \bigwedge_{j=1}^{t} \Pr[\overline{S_{i,j}}] = \prod_{j=1}^{t} \Pr[\overline{S_{i,j}}]$$

FAILURE

$$F_{i,t} = \text{NODE } i \text{ FAILS TO SEND AT TIMES } 1,2,\ldots,t$$

$$\Pr[F_{i,t}] = \bigwedge_{j=1}^{t} \Pr[\overline{S_{i,j}}] = \prod_{j=1}^{t} \Pr[\overline{S_{i,j}}]$$

FOR $t = O(n \ln n)$

$$\Pr[F_{i,t}] = n^{-c}$$

ALL FAIL

$$F_t =$$

$$\Pr[F_t] =$$

ALL FAIL

$F_t =$ SOME NODE i FAILS TO SEND AT TIMES 1,2,...,t

$$\Pr[F_t] = \bigvee_{i=1}^{n} \Pr[F_{i,t}]$$

ALL FAIL

F_t = SOME NODE i FAILS TO SEND AT TIMES 1,2,...,t

$$\Pr[F_t] = \bigvee_{i=1}^{n} \Pr[F_{i,t}] \leq \sum_{i=1}^{n} \Pr[F_{i,t}] \leq \sum_{i=1}^{n} n^{-c}$$

SUMMARY

AT TIME T, FLIP A COIN THAT IS HEADS WITH PR $\frac{1}{n}$

IF HEADS, THEN BROADCAST. IF SUCCESS, THEN STOP.

ELSE WAIT AND TRY AGAIN.

REPEAT $O(n \ln n)$ TIMES

WITH PROBABILITY

EVERY NODE SUCCEEDS IN SENDING MESSAGE.

TOOLS WE USED

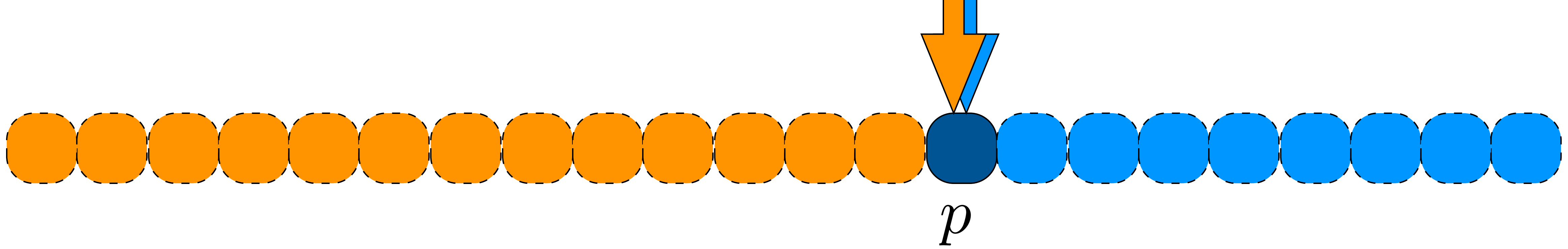
ANALYSIS OF $\left(1-\frac{1}{n}\right)^n$

PROBABILITY THAT MANY INDEPENDENT EVENTS ALL OCCUR:

PROBABILITY THAT ONE OUT OF N EVENTS OCCURS:

SECOND EXAMPLE:

MEDIAN



SELECT $(i, A[1, \ldots, n])$

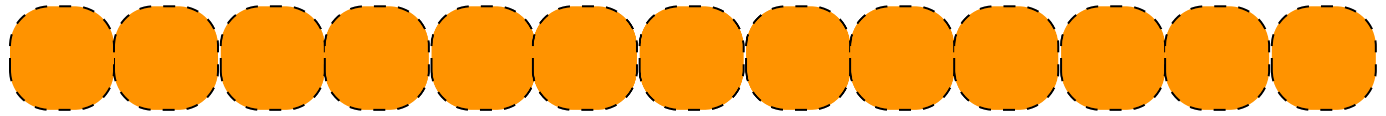
PICK FIRST ELEMENT

PARTITION LIST ABOUT THIS ONE

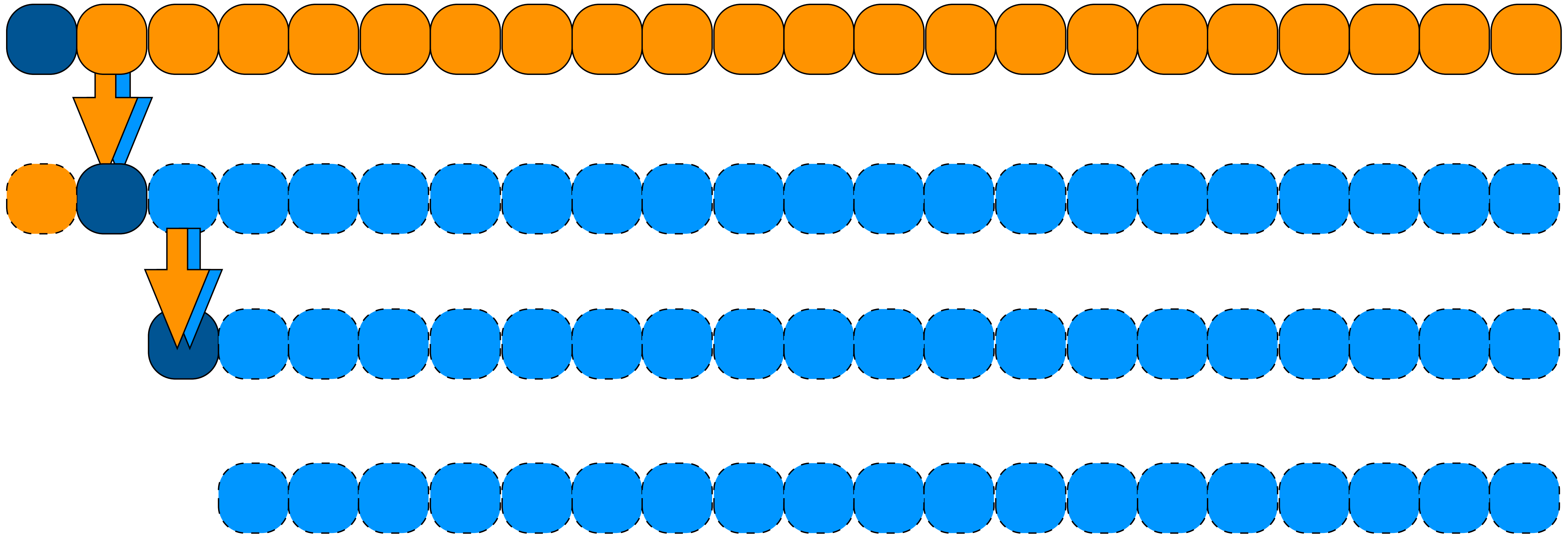
IF PIVOT IS POSITION i, RETURN PIVOT

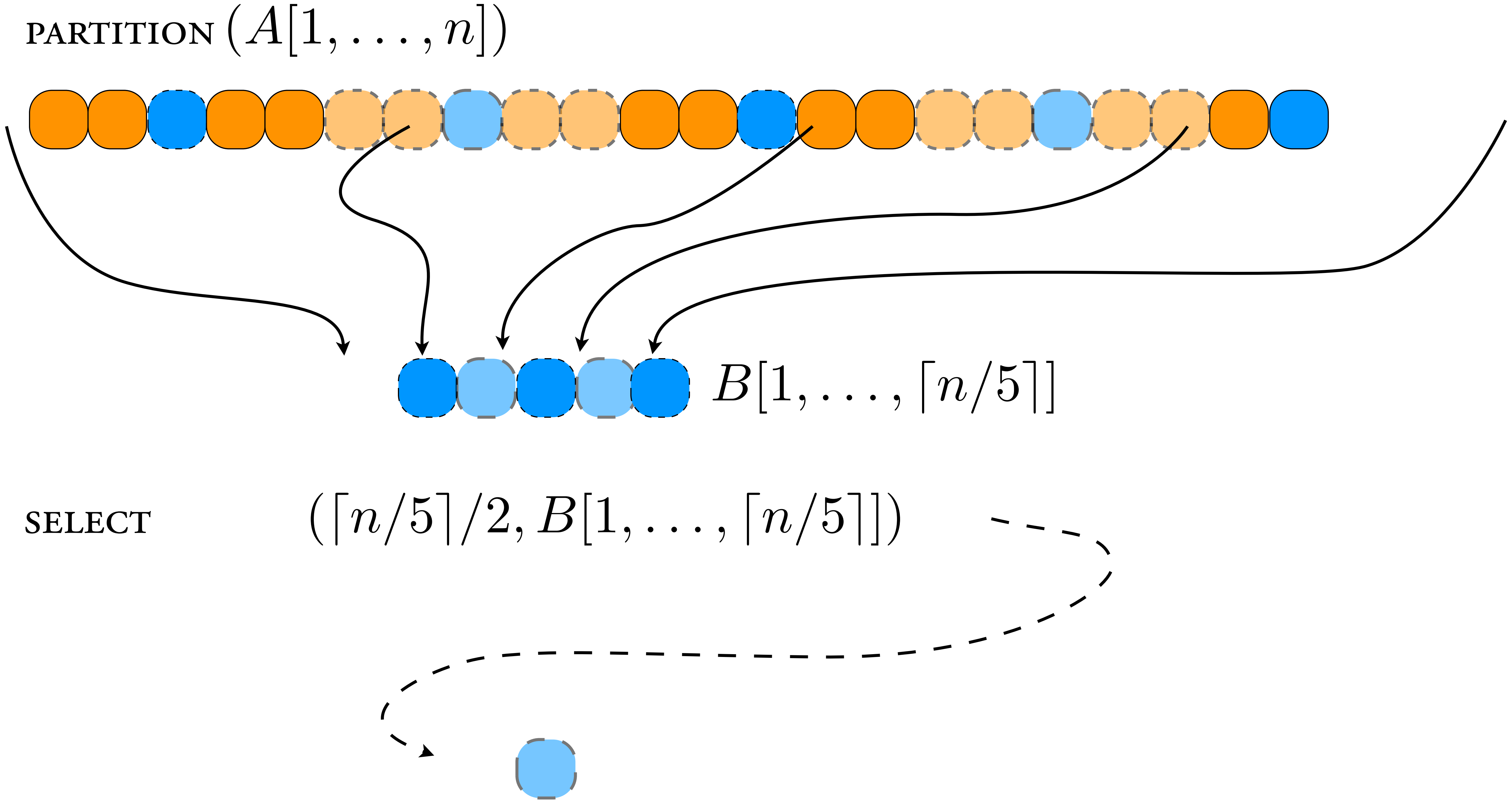
ELSE IF PIVOT IS IN POSITION > i SELECT $(i, A[1, \ldots, p-1])$

ELSE SELECT $((i-p-1), A[p+1, \ldots, n])$



PROBLEM: WHAT IF WE ALWAYS PICK BAD PARTITIONS?

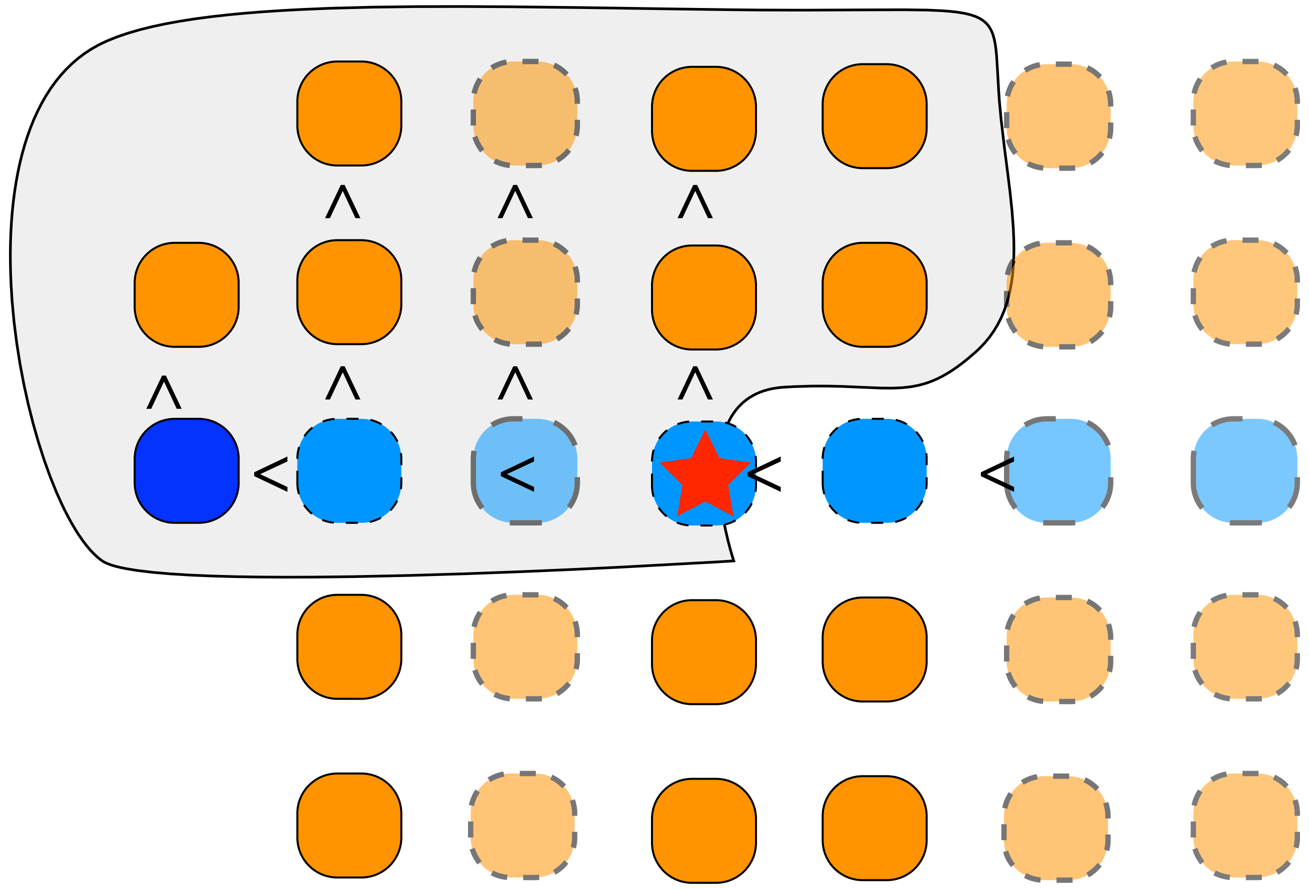



PARTITION $(A[1, \ldots, n])$
$B[1, \ldots, \lceil n/5 \rceil]$
SELECT $(\lceil n/5 \rceil / 2, B[1, \ldots, \lceil n/5 \rceil])$

A NICE PROPERTY OF OUR PARTITION

$$3\left(\left\lceil \frac{1}{2}\lceil n/5\rceil \right\rceil - 2\right)$$

$$\geq \frac{3n}{10} - 6$$

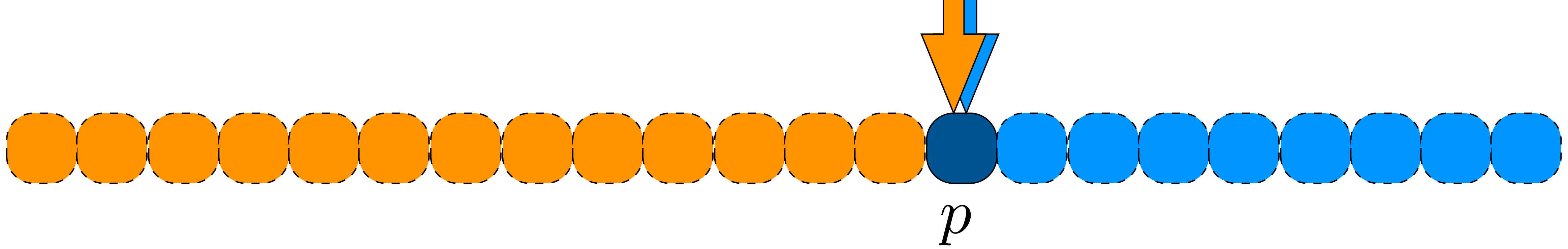


THIS IMPLIES THERE ARE

AT MOST $\dfrac{7n}{10} + 6$ NUMBERS

LARGER THAN ★

/SMALLER



SELECT $(i, A[1, \ldots, n])$

~~PICK FIRST ELEMENT~~

PIVOT = PARTITION $(A[1, \ldots, n])$

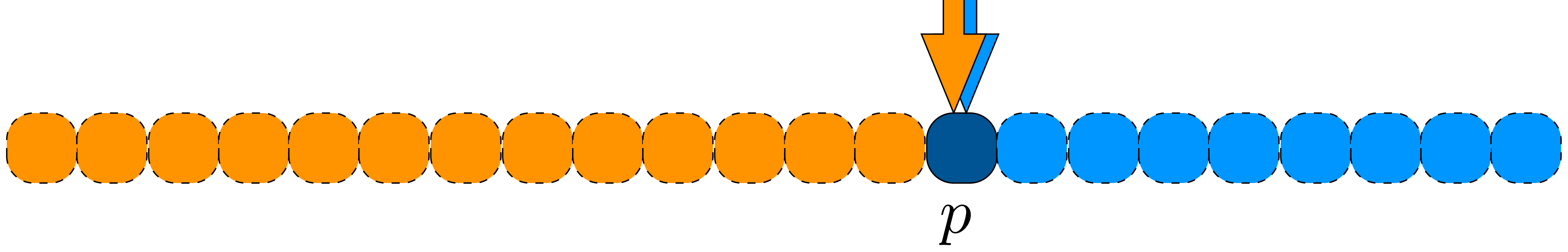
IF PIVOT IS POSITION i, RETURN PIVOT

ELSE IF PIVOT IS IN POSITION > i SELECT $(i, A[1, \ldots, p-1])$

ELSE SELECT $((i-p-1), A[p+1, \ldots, n])$

$$S(n) = S(\lceil n/5 \rceil) + O(n) + S(7n/10 + 6)$$

$$\Theta(n)$$



RANDOMIZEDSELECT $(i, A[1, \ldots, n])$

PICK RANDOM PARTITION ELEMENT

PARTITION LIST ABOUT THIS ONE

IF PIVOT IS POSITION i, RETURN PIVOT

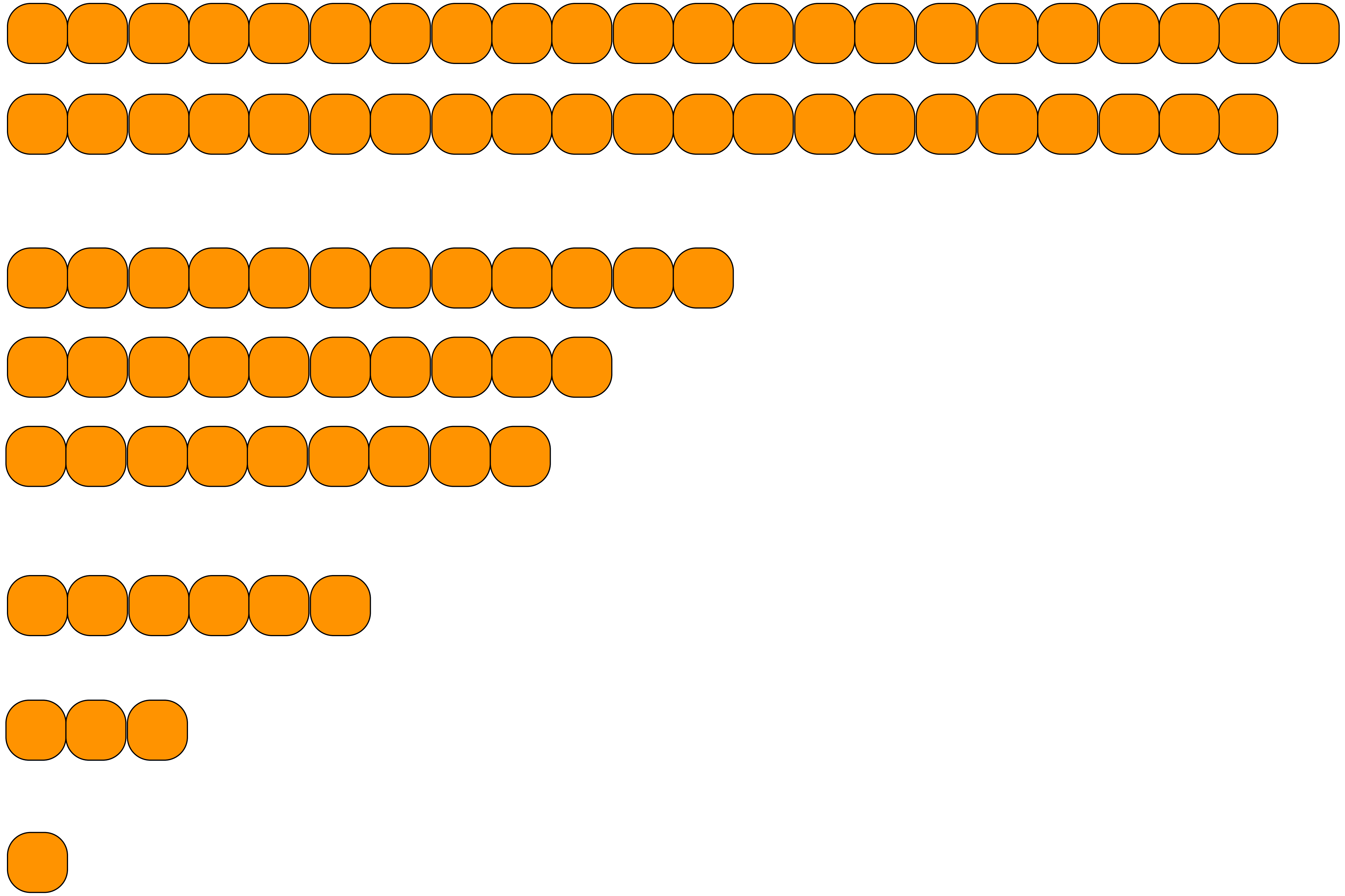
ELSE IF PIVOT IS IN POSITION > i SELECT $(i, A[1, \ldots, p-1])$

ELSE SELECT $((i-p-1), A[p+1, \ldots, n])$

RUNNING TIME ANALYSIS

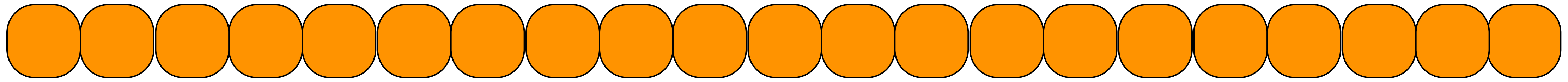
RECURSIVE CALLS

PHASES



PHASES

ALGORITHM IS IN PHASE J IF



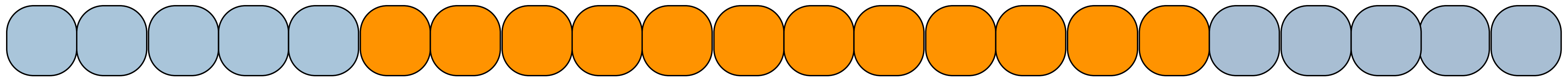
SIZE OF INPUT LIST IS $< \left(\frac{3}{4}\right)^{j} n$

RandomizedSelect $(i, A[1, \ldots, n])$

PICK RANDOM PARTITION ELEMENT

PARTITION LIST ABOUT THIS ONE

....



X_j = NUMBER OF

STEPS IN PHASE J

$E[X_j] =$

$X_j =$ **NUMBER OF**

STEPS IN PHASE J

$$E[X_j] = \sum_{j=0}^{\infty} j \cdot \Pr[X_j = j]$$

$\Pr[X_j = 1] =$

$\Pr[X_j = 2] =$

$\Pr[X_j = j] =$

LINEARITY OF EXPECTATION

$$\forall X, Y, \quad E[X + Y] = E[X] + E[Y]$$

EXPECTED RUNNING TIME

$$E[X] =$$