Goal
Consider a set

\[ S = \{x_1 < x_2 < \ldots < x_n\} \]

from a totally ordered universe. This set can dynamically change by adding or removing elements. Our goal is to search \( S \) for an element \( k \).
Searching

Goal
Consider a set

\[ S = \{x_1 < x_2 < \ldots < x_n\} \]

from a totally ordered universe. This set can dynamically change by adding or removing elements. Our goal is to search \( S \) for an element \( k \).

Bottom and Top Elements
We add a bottom element \( -\infty \) and top element \( +\infty \) to the universe such that

\[ -\infty < x_1 < x_2 < \ldots < x_n < +\infty. \]

These elements can simplify the implementation of the search.
Linked List Representation

Implementation
We can represent the set $S$ by an ordered linked list. The problem is that we cannot index into this list, so the search is slow.

Search Trees
A search tree can speed up the search, but can be a bit awkward to maintain under insert and delete operations.

Idea Behind Skip Lists
We want to obtain the speed of a binary search tree but combine it with the ease of maintaining a sorted linked list.
Skip lists were invented by Bill Pugh in 1990.

They offer an expected search time of $O(\log n)$.

They generalize linked lists and are easy to implement.
A **descending filtration** is a sequence $S_i$ of subsets of $S$ such that

$$\emptyset = S_r \subseteq S_{r-1} \subseteq \cdots \subseteq S_1 = S.$$ 

In computer science, the $S_i$ are called levels. The idea is that $S_k$ for a large $k$ is easy to search, since it has fewer elements than $S_1$.

The idea is that we implement each $S_i$ by a sorted linked list. Each element $x$ in $S_i$ is also linked to the element $x$ in the finer level $S_{i-1}$. 
Searching in a Skip List

Search

When we search for $k$:
- If $k = \text{key}$, done!
- If $k < \text{next key}$, then $k$ is not in this list, so go down a level
- If $k \geq \text{next key}$, then go right
Example: Search for 61
Example: Search for 29
Deterministic Construction

Construction
If the set $S_1 = S$ is fixed, then we could choose to include every other element into $S_2$. Next, put every other element of $S_2$ into $S_3$, and so forth.

Problem
We want to be able to insert and delete elements. These operations destroy the nice structure!
Randomized Construction

Construction
Let $S_1 = S$. For every element $x$ in $S_k$, include $x$ in $S_{k+1}$ with probability $1/2$.

Expected Number of Elements

\[
\begin{align*}
E[|S_1|] &= n, \\
E[|S_2|] &= n/2, \\
E[|S_3|] &= n/4, \\
\vdots 
\end{align*}
\]
We say that an element $x_k$ has **height** $\ell$ if and only if

$$x_k \in S_\ell, \quad \text{but} \quad x_k \notin S_{\ell+1}.$$ 

Let $X_k$ be the random variable that gives the height of the element $x_k$. We have

$$\Pr[X_k = \ell] = p(1 - p)^{\ell-1}.$$ 

So for $p = 1/2$, we have

$$\Pr[X_k = \ell] = (1/2)^\ell = 2^{-\ell}.$$
Interlude: Jensen’s Inequality
Jensen’s Inequality for Convex Functions

Proposition (Jensen’s Inequality)

Let $X$ be a random variable with $E[X] < \infty$. If $f : \mathbb{R} \to \mathbb{R}$ is a convex function, so $\supset$, then

$$f(E[X]) \leq E[f(X)].$$
Proposition (Jensen’s Inequality)

Let $X$ be a random variable with $E[X] < \infty$. If $f: \mathbb{R} \to \mathbb{R}$ is a convex function, then

$$f(E[X]) \leq E[f(X)].$$

Proof.

Since $f$ is convex, we can find a linear function $g(x) = ax + b$ which lies entirely below the graph of $f$, but touches $f$ at $E[X]$. In other words, we can choose real numbers $a$ and $b$ such that

$$f(E[X]) = g(E[X])$$

and $g(x) \leq f(x)$ for all $x \in \mathbb{R}$. 
Proof. (Continued).

Since \( f(x) \geq g(x) \) for all \( x \in \mathbb{R} \), it follows that

\[
E[f(X)] \geq E[g(X)]
\]

\[
= E[aX + b] = aE[X] + b
\]

\[
= g(E[X]) = f(E[X]).
\]
Jensen’s Inequality for Concave Functions

Proposition (Jensen’s Inequality)

Let $X$ be a random variable with $E[X] < \infty$. If $f : \mathbb{R} \rightarrow \mathbb{R}$ is a concave function, so $\preceq$, then

$$E[f(X)] \leq f(E[X]).$$

Proof.

If $f$ is concave, then $-f$ is convex. So

$$-f(E[X]) \leq E[-f(X)] = -E[f(X)]$$

by Jensen’s inequality for convex functions. Thus,

$$f(E[X]) \geq E[f(X)].$$
Proposition

The expected maximum height of a skip list with \( n \) elements is given by

\[
E \left[ \max_{1 \leq k \leq n} X_k \right] \in O(\log n).
\]
Proof.

Let $\alpha$ be a real number in the range $1 < \alpha < 2$. Then

$$E \left[ \max_{1 \leq k \leq n} X_k \right] \leq \log_\alpha E \left[ \alpha^{\max_{1 \leq k \leq n} X_k} \right]$$

$$= \log_\alpha E \left[ \max_{1 \leq k \leq n} \alpha^{X_k} \right].$$

Since $\alpha^{X_k} \geq 1$, we can estimate the right-hand side by the sum

$$E \left[ \max_{1 \leq k \leq n} X_k \right] \leq \log_\alpha E \left[ \sum_{k=1}^{n} \alpha^{X_k} \right].$$
Continued.

\[
E \left[ \max_{1 \leq k \leq n} X_k \right] \leq \log_\alpha E \left[ \sum_{k=1}^{n} \alpha^X_k \right] = \log_\alpha \left( \sum_{k=1}^{n} \sum_{k \geq 1} \alpha^k 2^{-k} \right) \\
= \log_\alpha \left( \sum_{k=1}^{n} \frac{1}{1 - \alpha/2} \right) \\
= \log_\alpha n + \log_\alpha \frac{1}{1 - \alpha/2} = O(\log n),
\]

which is what we wanted to show.
Proposition

The number of levels of a skip list of a set with \( n \) elements satisfies \( O(\log n) \) with high probability.

Proof.

Let \( X_k \) denote the random variable giving the number of levels of the \( k \)-the element of \( S \). Then

\[
Pr[X_k > t] \leq (1 - p)^t.
\]

So

\[
Pr[\max_k X_k > t] \leq n(1 - p)^t = \frac{n}{2^t}
\]

for \( p = 1/2 \). Choosing \( t = a \log n \) and \( r = \max_k X_k \), we can conclude that

\[
Pr[r > a \log n] \leq \frac{1}{n^{a-1}}
\]

for any \( a > 1 \). \( \square \)