Randomized Algorithms

The power of two choices: suppressing extreme loads

Alex Gittens
Department of Computer Science, Rensselaer Polytechnic Institute

Fall 2026

Learning Objectives

By the end of this lecture, you should be able to:

1. **Compare** one-choice and two-choice load balancing.
2. **Explain** why reaching the next load level requires both probes to hit heavily loaded bins.
3. **Bound** the number of high placements even though successive placements are dependent.
4. **Explain** why repeated squaring produces a $\log \log n$ maximum load.

A load-balancing problem

There are n initially empty bins and n balls arriving one at a time.

For each ball, we may inspect only a few randomly sampled bins before choosing where to place it. The average load after all balls arrive is exactly 1, but the busiest bin may have a much larger load.

$$L_{\max} = \max_{v \in [n]} L_v(n).$$

Our question is:

How much can one additional random probe reduce the worst congestion?

One random choice

If each ball chooses one uniformly random bin, then the load of a fixed bin is

$$L_v(n) \sim \mathrm{Bin}\left(n, \frac{1}{n}\right).$$

Using a Chernoff bound for a sum of independent Bernoulli trials, for $k \geq 1$,

$$\Pr(L_v(n) \geq k) \leq \left(\frac{e}{k}\right)^k.$$

A union bound over all n bins gives

$$\Pr(L_{\max} \geq k) \leq n\left(\frac{e}{k}\right)^k.$$

The transition occurs near

$$k \log k \approx \log n,$$

which gives

$$L_{\max} = O\left(\frac{\log n}{\log \log n}\right) \qquad \text{with high probability.}$$

The matching lower bound requires a complementary argument.

Two random choices

For each arriving ball:

1. sample two bins independently and uniformly from $[n]$;
2. inspect their current loads;
3. place the ball in the less loaded sampled bin.

Ties may be broken arbitrarily.

A heavily loaded bin can receive another ball only when the other probe is also heavily loaded.
The maximum load now satisfies

$$L_{\max} = \log_2 \log n + O(1) \qquad \text{with high probability.}$$

We will prove the upper bound

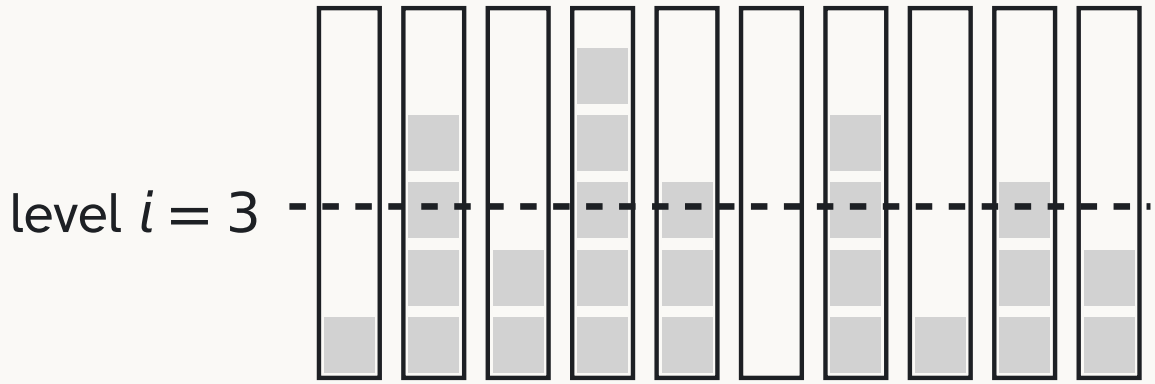
$$L_{\max} \leq \log_2 \log n + O(1) \qquad \text{with high probability.}$$

The matching lower bound requires a separate argument.

Heavy bins at level i

After t balls have arrived, define

$$H_i(t) = \{v : L_v(t) \geq i\}, \qquad B_i(t) = |H_i(t)|, \qquad b_i(t) = \frac{B_i(t)}{n}.$$



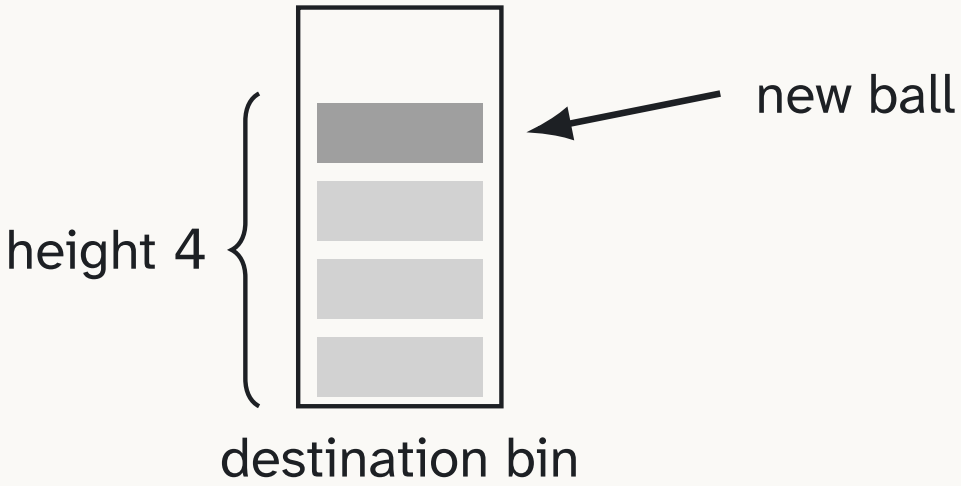
How many bins have load at least 3?

$$B_3(t) = 5, \qquad b_3(t) = \frac{5}{10} = \frac{1}{2}.$$

$B_i(t)$ counts the bins that have reached level i. The fraction $b_i(t)$ will determine how likely the process is to create level $i + 1$.

Ball height

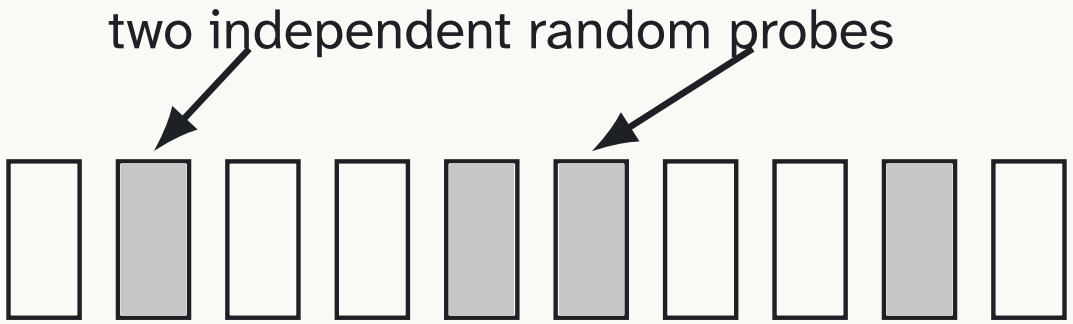
The *height* of a ball is one plus the load of its destination immediately before placement.



A ball has height at least $i + 1$ exactly when it enters a bin whose current load is at least i.
Thus every bin that finishes with load at least $i + 1$ must have received a ball of height at least $i + 1$.

Two heavy probes create the next layer

Suppose the shaded bins currently have load at least i.



$b_i(t) = 4/10$ of the bins are heavy

The chosen bin is the lighter of the two probes. Therefore a ball can have height at least $i + 1$ only if *both* probes land in the level-i heavy set. Conditioned on everything that has happened through time t, the heavy set is fixed and the next two probes are fresh. Hence

$$\Pr(\text{height} \geq i + 1 \mid \text{history through time } t) = b_i(t)^2.$$

High balls count the bins in the next layer

For a fixed level i, define

$$X_t^{(i)} = \mathbf{1}\{\text{ball } t \text{ has height at least } i+1\}, \qquad S_i = \sum_{t=1}^{n} X_t^{(i)}.$$

Every bin that reaches level $i+1$ needs at least one ball of height $i+1$. Therefore

$$\boxed{B_{i+1}(n) \le S_i.}$$

If the level-i heavy fraction always stayed below some number β, then each ball would have conditional probability at most β^2 of contributing to S_i.

This suggests that S_i should behave, for upper-tail purposes, like a binomial random variable with mean $n\beta^2$. The indicators $X_t^{(i)}$, however, are not independent.

History determines the next success probability

The placement of ball t changes the bin loads. This can change $b_i(t)$ and therefore the probability that ball $t+1$ reaches height $i+1$.

$$\Pr(X_{t+1}^{(i)} = 1 \mid \mathcal{F}_t) = b_i(t)^2,$$

where

$$\mathcal{F}_t = \text{all random choices and placements through ball } t.$$

The notation $\mathcal{F}_t$ simply gives a short name to the information already revealed by time t. Once we condition on $\mathcal{F}_t$, the current loads are fixed, while the next two probes remain independent and uniform.

Ordinary binomial independence is unavailable, but the conditional success probabilities are still enough for an MGF argument.

Chernoff bound from conditional probabilities

Consider any sequence of indicators $Z_1, \ldots, Z_n$. The choice of Z_t may depend on the outcomes before time t.
Suppose that, for every possible history through time $t-1$,

$$\Pr(Z_t = 1 \mid \mathcal{F}_{t-1}) \le p.$$

Set $\mu = np$. Then, for every $\delta > 0$,

$$\boxed{\Pr\left(\sum_{t=1}^{n} Z_t \ge (1+\delta)\mu\right) \le \exp\left(-\frac{\delta^2}{2+\delta}\mu\right).}$$

In particular, taking $\delta = 1$ gives

$$\boxed{\Pr\left(\sum_{t=1}^{n} Z_t \ge 2\mu\right) \le e^{-\mu/3}.}$$

The number $\mu = np$ is the mean of a comparison binomial random variable. It need not equal $\mathbb{E}\sum_t Z_t$. The proof appears in the appendix.

From level i to level $i+1$

For each level, define

$$E_i = \{B_i(n) \le \beta_i n\}.$$

The transition from level i to level $i+1$ fails when

$$E_i \cap E_{i+1}^c$$

occurs: level i meets its target, but level $i+1$ does not.
Because $B_i(t)$ never decreases, E_i implies

$$B_i(t) \le \beta_i n \qquad \text{for every } t \le n.$$

Thus, when E_i holds, the level-i heavy set stays small throughout the process.
The tempting next step is not justified:

$$\Pr(X_{t+1}^{(i)} = 1 \mid \mathcal{F}_t, E_i) \le \beta_i^2.$$

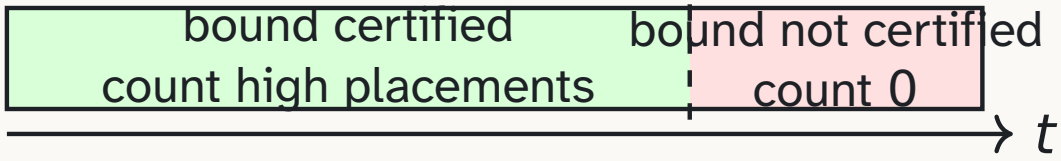
Conditioning on the final event E_i reveals information about future probes and can change the distribution of the next two choices.
We instead need a count whose conditional probability is bounded using only information already revealed at time t.

Truncation enables the Chernoff bound

Define

$$\widetilde{X}_{t+1}^{(i)} = X_{t+1}^{(i)} \mathbf{1}\{B_i(t) \le \beta_i n\}.$$

This indicator counts a high placement only while the observed history certifies that the level-i heavy set has at most $\beta_i n$ bins.



Before the threshold is exceeded, the conditional success probability is at most β_i^2. Afterward, the truncated indicator is zero. Thus, for every possible history,

$$\Pr(\widetilde{X}_{t+1}^{(i)} = 1 \mid \mathcal{F}_t) \le \beta_i^2.$$

This is exactly the hypothesis of the conditional Chernoff bound, with $p = \beta_i^2$.

On E_i, truncation changes nothing

On E_i, the inequality $B_i(t) \le \beta_i n$ holds throughout the process. Therefore every safe indicator equals its original indicator:

$$E_i \quad \Longrightarrow \quad \widetilde{X}_t^{(i)} = X_t^{(i)} \text{ for all } t.$$

Since $B_{i+1}(n) \le S_i$, for every threshold a,

$$\boxed{\Pr\big(E_i \cap \{B_{i+1}(n) > a\}\big) \le \Pr(\widetilde{S}_i > a).}$$

The Chernoff bound for $\widetilde{S}_i$ holds without assuming E_i. The equality $\widetilde{S}_i = S_i$ is needed only on the event E_i appearing on the left.

A deterministic starting level

There are exactly n balls. If $B_4(n)$ bins each contain at least four balls, those bins contain at least $4B_4(n)$ balls. Hence

$$4B_4(n) \le n, \qquad \text{so} \qquad B_4(n) \le \frac{n}{4}.$$

Set

$$\boxed{\beta_4 = \frac{1}{4}, \qquad E_4 = \{B_4(n) \le \beta_4 n\}.}$$

The event E_4 holds deterministically. It gives the recursion a starting level whose heavy-bin population is already known to be small.

The transition estimate

For the truncated indicators, take

$$p = \beta_i^2, \qquad \mu_i = np = n\beta_i^2.$$

Here μ_i is the comparison parameter in the conditional Chernoff bound. It need not equal $\mathbb{E}\widetilde{S}_i$.
Choose

$$\boxed{\beta_{i+1} = 2\beta_i^2.}$$

Then the next threshold is

$$\beta_{i+1} n = 2n\beta_i^2 = 2\mu_i.$$

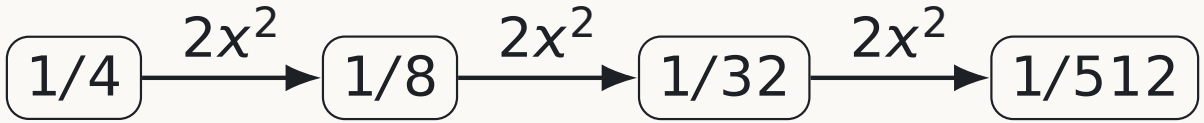
The conditional Chernoff bound and the agreement of the true and truncated counts on E_i give

$$\boxed{\Pr(E_i \cap E_{i+1}^c) \le \Pr(\widetilde{S}_i \ge 2\mu_i) \le e^{-\mu_i/3} = e^{-n\beta_i^2/3}.}$$

Repeated squaring

Starting from $\beta_4 = 1/4$,

$$\beta_5 = \frac{1}{8}, \qquad \beta_6 = \frac{1}{32}, \qquad \beta_7 = \frac{1}{512}.$$



Set $z_i = 2\beta_i$. Then

$$z_{i+1} = z_i^2, \qquad z_4 = \frac{1}{2}.$$

Consequently,

$$z_{4+k} = 2^{-2^k}, \qquad \boxed{\beta_{4+k} = 2^{-1-2^k}.}$$

Each new level doubles the exponent. This is the source of the $\log\log n$ scale.

$C \log n$ gives polynomially small failure

For the transition from level i to level $i+1$,

$$\Pr(E_i \cap E_{i+1}^c) \le e^{-\mu_i/3}, \qquad \mu_i = n\beta_i^2.$$

We want every recursive step to succeed at once. If

$$\mu_i \ge C \log n,$$

then

$$e^{-\mu_i/3} \le e^{-(C/3)\log n} = n^{-C/3}.$$

This is polynomially small in n. There are only $O(\log \log n)$ recursive levels, so a union bound gives

$$O(\log \log n)\, n^{-C/3} = o(1).$$

The Chernoff bound does not suddenly stop working below $C \log n$. We choose this cutoff because it gives a simple, uniform probability budget for the entire recursive phase.

The first failure identifies a valid transition

Let j be the largest level for which

$$n\beta_j^2 \geq C \log n,$$

where $C > 3$ is fixed.
Define the event that every recursive threshold holds:

$$\mathcal{G} = \bigcap_{i=4}^{j+1} E_i.$$

The event E_4 holds deterministically. If the induction fails, let E_k be its first failed event:

$$E_4 \checkmark \quad \cdots \quad E_{k-1} \checkmark \quad E_k \times \quad E_{k+1}, \ldots \text{ irrelevant.}$$

Then E_{k-1} still holds, so the bound for the transition from level $k-1$ to level k applies.

The total failure probability

Every outcome in $\mathcal{G}^c$ belongs to at least one event

$$E_i \cap E_{i+1}^c, \qquad i = 4, \ldots, j.$$

Therefore

$$\begin{aligned}
\Pr(\mathcal{G}^c) &\le \sum_{i=4}^{j} \Pr(E_i \cap E_{i+1}^c) \\
&\le \sum_{i=4}^{j} n^{-C/3} \\
&\le O(\log\log n) n^{-C/3} = o(1).
\end{aligned}$$

No independence between the events E_i is required. The union bound uses only the transition estimates already proved.

$$\boxed{\Pr\left(\bigcap_{i=4}^{j+1} E_i\right) = 1 - o(1).}$$

One more level leaves only $O(\log n)$ bins

With high probability, E_{j+1} holds:

$$B_{j+1}(n) \le \beta_{j+1} n.$$

Since j is the largest level satisfying $n\beta_j^2 \ge C \log n$,

$$n\beta_{j+1}^2 < C \log n.$$

Apply the safe-count MGF argument once more, now from level $j+1$ to level $j+2$. The comparison mean is smaller than $C \log n$.
Choose a sufficiently large constant $K > C$. A Chernoff bound at the larger threshold $K \log n$ gives

$$\Pr\big(E_{j+1} \cap \{B_{j+2}(n) > K \log n\}\big) \le e^{-c_K \log n} = n^{-c_K}$$

for a constant $c_K > 0$ that grows with K.
Therefore

$$\boxed{B_{j+2}(n) \le K \log n \qquad \text{with high probability.}}$$

The next load layer disappears

Let

$$G = \{B_{j+2}(n) \le K \log n\}.$$

On G, monotonicity gives $B_{j+2}(t) \le K \log n$ for every t. Define a safe indicator that records a height-$(j+3)$ ball only while the observed heavy set has at most $K \log n$ bins.
At every step, its conditional success probability is at most

$$q = \left(\frac{K \log n}{n}\right)^2.$$

A union bound over the n arrivals gives

$$\Pr(G \text{ and some ball reaches height } j+3) \le nq = \frac{K^2 \log^2 n}{n} = o(1).$$

Since $\Pr(G^c) = o(1)$, with high probability no ball reaches height $j+3$.
Therefore

$$\boxed{L_{\max} \le j+2 \qquad \text{with high probability.}}$$

Repeated squaring needs $\log \log n$ levels

Recall

$$\beta_{4+k} = 2^{-1-2^k}$$

and the cutoff relation

$$n\beta_j^2 \asymp \log n.$$

Writing $k = j - 4$ gives

$$n\, 2^{-2-2^{k+1}} \asymp \log n.$$

Taking logarithms once,

$$2^{k+1} = \Theta(\log n).$$

Taking logarithms again,

$$k = \log_2 \log n + O(1).$$

Since $j = k + 4$,

$$\boxed{j = \log_2 \log n + O(1).}$$

Combining this with $L_{\max} \le j + 2$ proves the upper bound.

Two choices square the heavy-bin fraction

For one choice, the extreme load is governed by the factorial tail of a single bin and a maximum over n bins:

$$L_{\max} \asymp \frac{\log n}{\log \log n}.$$

For two choices, a ball can create level $i+1$ only when both probes lie in the current level-i heavy set:

$$\beta_i \longmapsto 2\beta_i^2.$$

After normalization, the recurrence becomes

$$z_i \longmapsto z_i^2.$$

Thus the exponent doubles at every load level. It reaches order $\log n$ after only order $\log \log n$ levels.

$$\boxed{L_{\max} \le \log_2 \log n + O(1) \qquad \text{with high probability.}}$$

d choices

Suppose each ball samples $d \geq 2$ bins and joins the least loaded one. A ball reaches height at least $i + 1$ only if all d probes lie in the level-i heavy set.
Thus the one-step probability becomes

$$b_i(t)^d,$$

and the heavy-set recurrence has the form

$$\beta_{i+1} = C_d \beta_i^d$$

for a suitable constant C_d.
After normalization, taking negative logarithms changes the recurrence into multiplication by d. The exponent therefore grows like d^i. Reaching a polynomially small heavy fraction takes

$$i = \frac{\log \log n}{\log d} + O(1).$$

The maximum-load scale becomes

$$\frac{\log \log n}{\log d} + O(1).$$

Exercises for reconstruction

1. **Conditional Chernoff bound.** Let $Z_1, \dots, Z_n \in \{0, 1\}$ satisfy

$$\Pr(Z_t = 1 \mid \mathcal{F}_{t-1}) \le p.$$

   Reproduce the tower-property argument and derive the multiplicative Chernoff bound for $\sum_t Z_t$.

2. d **choices.** Derive the level-to-level recurrence when each ball samples d bins and joins the least loaded one. Use it to recover the scale $\log \log n / \log d$ up to an additive constant.

3. **Sampling without replacement.** Suppose the two probes must be distinct. If $B_i(t) = r$, compute the probability that both probes lie in the heavy set, upper-bound it by $(r/n)^2$, and explain why the proof still goes through.

Questions?

The conditional MGF of one indicator

Because Z_t takes values in $\{0, 1\}$,

$$e^{\theta Z_t} = 1 + (e^\theta - 1)Z_t.$$

Conditioning on the history gives

$$\begin{aligned}\mathbb{E}[\, e^{\theta Z_t} \mid \mathcal{F}_{t-1}] &= 1 + (e^\theta - 1)\mathbb{E}[Z_t \mid \mathcal{F}_{t-1}] \\ &= 1 + (e^\theta - 1)\Pr(Z_t = 1 \mid \mathcal{F}_{t-1}) \\ &\le 1 + p(e^\theta - 1).\end{aligned}$$

This bound holds after every possible history. That uniform conditional bound replaces independence in the Chernoff argument.

The tower property gives the full MGF bound

Let $T_t = Z_1 + \cdots + Z_t$. Since T_{t-1} is determined by the history through time $t-1$,

$$\begin{aligned}\mathbb{E}[\,e^{\theta T_t}] &= \mathbb{E}\big[e^{\theta T_{t-1}}\mathbb{E}(e^{\theta Z_t} \mid \mathcal{F}_{t-1})\big] \\ &\le \big(1 + p(e^{\theta}-1)\big)\mathbb{E}[\,e^{\theta T_{t-1}}].\end{aligned}$$

Repeating this step gives

$$\mathbb{E}[\,e^{\theta T_n}] \le \big(1 + p(e^{\theta}-1)\big)^n \le \exp\big(\mu(e^{\theta}-1)\big), \qquad \mu = np.$$

This is the same MGF upper bound used for a binomial random variable with mean μ.

Markov's inequality gives the Chernoff bound

For $\theta > 0$, Markov's inequality gives

$$\begin{aligned}\Pr(T_n \geq (1+\delta)\mu) &\leq e^{-\theta(1+\delta)\mu}\mathbb{E}[\,e^{\theta T_n}\,] \\ &\leq \exp\big(\mu(e^{\theta}-1)-\theta(1+\delta)\mu\big).\end{aligned}$$

Choose $\theta = \log(1+\delta)$. Then

$$\Pr(T_n \geq (1+\delta)\mu) \leq \left(\frac{e^{\delta}}{(1+\delta)^{1+\delta}}\right)^{\mu} \leq \exp\left(-\frac{\delta^2}{2+\delta}\mu\right).$$

For $\delta = 1$, this becomes

$$\Pr(T_n \geq 2\mu) \leq e^{-\mu/3}.$$