Theorem

Let $X_1, \ldots, X_n$ be independent, identically distributed, $0 - 1$ random variables with $Pr(X_i = 1) = E[X_i] = p$. Let $\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$, then for any $\delta \in [0, 1]$ we have

$$Prob(\bar{X}_n \geq (1 + \delta)p) \leq e^{-np\delta^2/3}$$

and

$$Prob(\bar{X}_n \leq (1 - \delta)p) \leq e^{-np\delta^2/2}.$$
Let $X_1, \ldots, X_n$ be independent, $0 - 1$ random variables with $\Pr(X_i = 1) = E[X_i] = p_i$. Let $\mu = \sum_{i=1}^{n} p_i$, then for any $\delta \in [0, 1]$ we have

$$\Pr\left(\sum_{i=1}^{n} X_i \geq (1 + \delta)\mu\right) \leq e^{-\mu\delta^2/3}$$

and

$$\Pr\left(\sum_{i=1}^{n} X_i \leq (1 - \delta)\mu\right) \leq e^{-\mu\delta^2/2}.$$
Packet Routing on Parallel Computer

Communication network:
Packet Routing on Parallel Computer

Communication network:

- nodes - processors, switching nodes;
- edges - communication links.
The $n$-cube

The 3-cube:

The 4-cube:
The $n$-cube

The $n$-cube:
$N = 2^n$ nodes: $0, 1, 2, \ldots, 2^n - 1$.
Let $\bar{x} = (x_1, \ldots, x_n)$ be the number of node $x$ in binary.
Nodes $x$ and $y$ are connected by an edge iff their binary representations differ in exactly one bit.
Bit-wise routing: correct bit $i$ in the $i$-th transition - route has length $\leq n$. 
The $n$-cube

The 3-cube:

The 4-cube:
Model and Computational problem

- Up to one packet can cross an edge per step, each packet can cross up to one edge per step.
- A permutation communication request: each node is the source and destination of exactly one packet.
- What is the time to route an arbitrary permutation on the $n$-cube?
Two phase routing algorithm:

1. Send packet to a randomly chosen destination.
2. Send packet from randomly chosen destination to real destination.

Path: Correct the bits, starting at $x_1$ to $x_n$.

Any greedy queuing method - if some packet can traverse an edge one does.
Theorem

The two phase routing algorithm routes an arbitrary permutation on the \(n\)-cube in \(O(\log N) = O(n)\) parallel steps with high probability.

- We focus first on phase 1. We bound the routing time of a given packet \(M\).
- Let \(e_1, ..., e_m\) be the \(m \leq n\) edges traversed by a given packet \(M\) is phase 1.
- Let \(X(e)\) be the total number of packets that traverse edge \(e\) at that phase.
- Let \(T(M)\) be the number of steps till \(M\) finished phase 1.
Lemma

\[ T(M) \leq \sum_{i=1}^{m} X(e_i). \]

• We call any path \( P = (e_1, e_2, \ldots, e_m) \) of \( m \leq n \) edges that follows the bit fixing algorithm a possible packet path.
• We denote the corresponding nodes \( v_0, v_1, \ldots, v_m \), with \( e_i = (v_{i-1}, v_i) \).
• For any possible packet path \( P \), let \( T(P) = \sum_{i=1}^{m} X(e_i) \).
• If phase I takes more than $T$ steps then for some possible packet path $P$,

$$T(P) \geq T$$

• There are at most $2^n \cdot 2^n = 2^{2n}$ possible packet paths.

• Assume that $e_k$ connects $(a_1, \ldots, a_i, \ldots, a_n)$ to $(a_1, \ldots, \bar{a}_i, \ldots, a_n)$.

• Only packets that started in address

$$(*, \ldots, *, a_i, \ldots, a_n)$$

can traverse edge $e_k$, and only if their destination addresses are

$$(a_1, \ldots, a_i-1, \bar{a}_i, *, \ldots, *)$$

• There are no more than $2^{i-1}$ possible packets, each has probability $2^{-i}$ to traverse $e_i$. 
• There are no more than $2^{i-1}$ possible packets, each has probability $2^{-i}$ to traverse $e_i$.

• $\mathbb{E}[X(e_i)] \leq 2^{i-1} \cdot 2^{-i} = \frac{1}{2}$.

• $\mathbb{E}[T(P)] \leq \sum_{i=1}^{m} \mathbb{E}[X(e_i)] \leq \frac{1}{2} \cdot m \leq n$.

• **Problem**: The $X(e_i)$’s are not independent.
• A packet is active with respect to possible packet path $P$ if it ever use an edge of $P$.
• For $k = 1, \ldots, N$, let $H_k = 1$ if the packet starting at node $k$ is active, and $H_k = 0$ otherwise.
• The $H_k$ are independent, since each $H_k$ depends only on the choice of the intermediate destination of the packet starting at node $k$, and these choices are independent for all packets.
• Let $H = \sum_{k=1}^{N} H_k$ be the total number of active packets.
• 
  $$\mathbb{E}[H] \leq \mathbb{E}[T(P)] \leq n$$

• Since $H$ is the sum of independent $0-1$ random variables we can apply the Chernoff bound

  $$\Pr(H \geq 6n) \leq \Pr(H \geq 6\mathbb{E}[H]) \leq 2^{-6n}.$$
For a given possible packet path \( P \),

\[
\Pr( T(P) \geq 30n) \leq \Pr( T(P) \geq 30n \mid H \geq 6n) \Pr( H \geq 6n) + \Pr( T(P) \geq 30n \mid H < 6n) \Pr( H < 6n) \\
\leq \Pr( H \geq 6n) + \Pr( T(P) \geq 30n \mid H < 6n) \\
\leq 2^{-6n} + \Pr( T(P) \geq 30n \mid H < 6n).
\]
Lemma

*If a packet leaves a path (of another packet) it cannot return to that path in the same phase.*

Proof.

Leaving a path at the \(i\)-th transition implies different \(i\)-th bit, this bit cannot be changed again in that phase.

Lemma

*The number of transitions that a packet takes on a given path is distributed \(G\left(\frac{1}{2}\right)\).*

Proof.

The packet has probability \(1/2\) of leaving the path in each transition.
The probability that the active packets cross edges of $P$ more than $30n$ times is less than the probability that a fair coin flipped $36n$ times comes up heads less than $6n$ times.

Letting $Z$ be the number of heads in $36n$ fair coin flips, we now apply the Chernoff bound:

$$
\Pr(T(P) \geq 30n \mid H \leq 6n) \leq \Pr(Z \leq 6n) \\
\leq e^{-18n(2/3)^2/2} = e^{-4n} \leq 2^{-3n-1}.
$$

$$
\Pr(T(P) \geq 30n) \leq \Pr(H \geq 6n) + \Pr(T(P) \geq 30n \mid H \leq 6n) \\
\leq 2^{-6n} + 2^{-3n-1} \leq 2^{-3n}
$$
As there are at most $2^{2n}$ possible packet paths in the hypercube, the probability that there is any possible packet path for which $T(P) \geq 30n$ is bounded by

$$2^{2n}2^{-3n} = 2^{-n} = O(N^{-1})$$.
• The proof of phase 2 is by symmetry:
• The proof of phase 1 argued about the number of packets crossing a given path, no “timing” considerations.
• The path from “one packet per node” to random locations is similar to random locations to “one packet per node” in reverse order.
• Thus, the distribution of the number of packets that crosses a path of a given packet is the same.
### Definition
A routing algorithm is **oblivious** if the path taken by one packet is independent of the source and destinations of any other packets in the system.

### Theorem
Given an *N*-node network with maximum degree *d* the routing time of any deterministic oblivious routing scheme is

\[ \Omega \left( \sqrt{\frac{N}{d^3}} \right) \].
Set Balancing

Given an $n \times n$ matrix $A$ with entries in $\{0, 1\}$, let

$$
\begin{pmatrix}
  a_{11} & a_{12} & \ldots & a_{1n} \\
  a_{21} & a_{22} & \ldots & a_{2n} \\
  \vdots & \vdots & \ddots & \vdots \\
  a_{n1} & a_{n2} & \ldots & a_{nn}
\end{pmatrix}
\begin{pmatrix}
  b_1 \\
  b_2 \\
  \vdots \\
  b_n
\end{pmatrix}
=
\begin{pmatrix}
  c_1 \\
  c_2 \\
  \vdots \\
  c_n
\end{pmatrix}.
$$

Find a vector $\bar{b}$ with entries in $\{-1, 1\}$ that minimizes

$$
\|A\bar{b}\|_\infty = \max_{i=1,\ldots,n} |c_i|.
$$
Theorem

For a random vector $\bar{b}$, with entries chosen independently and with equal probability from the set \{-1, 1\},

$$\Pr(||A\bar{b}||_\infty \geq \sqrt{4n \ln n}) \leq \frac{2}{n}.$$ 

The $\sum_{i=1}^{n} a_{j,i} b_i$ (excluding the zero terms) is a sum of independent $-1, 1$ random variable. We need a bound on such sum.
Chernoff Bound for Sum of \([-1, +1]\) Random Variables

**Theorem**

Let \(X_1, \ldots, X_n\) be independent random variables with

\[
Pr(X_i = 1) = Pr(X_i = -1) = \frac{1}{2}.
\]

Let \(X = \sum_{1}^{n} X_i\). For any \(a > 0\),

\[
Pr(X \geq a) \leq e^{-\frac{a^2}{2n}}.
\]

de Moivre – Laplace approximation: For any \(k\), such that \(|k - np| \leq a\)

\[
\binom{n}{k} p^k (1 - p)^{n-k} \approx \frac{1}{\sqrt{2\pi np(1 - p)}} e^{-\frac{a^2}{2np(1-p)}}
\]
For any $t > 0$, 

$$
\mathbb{E}[e^{tX_i}] = \frac{1}{2}e^t + \frac{1}{2}e^{-t}.
$$

$$
e^t = 1 + t + \frac{t^2}{2!} + \cdots + \frac{t^i}{i!} + \cdots
$$

and

$$
e^{-t} = 1 - t + \frac{t^2}{2!} + \cdots + (-1)^i\frac{t^i}{i!} + \cdots
$$

Thus,

$$
\mathbb{E}[e^{tX_i}] = \frac{1}{2}e^t + \frac{1}{2}e^{-t} = \sum_{i \geq 0} \frac{t^{2i}}{(2i)!}
$$

$$
\leq \sum_{i \geq 0} \frac{(\frac{t^2}{2})^i}{i!} = e^{t^2/2}
$$
\[ E[e^{tX}] = \prod_{i=1}^{n} E[e^{tX_i}] \leq e^{nt^2/2}, \]

\[ Pr(X \geq a) = Pr(e^{tX} > e^{ta}) \leq \frac{E[e^{tX}]}{e^{ta}} \leq e^{t^2 n/2 - ta}. \]

Setting \( t = a/n \) yields

\[ Pr(X \geq a) \leq e^{-\frac{a^2}{2n}}. \]
By symmetry we also have

**Corollary**

Let $X_1, \ldots, X_n$ be independent random variables with

$$Pr(X_i = 1) = Pr(X_i = -1) = \frac{1}{2}.$$

Let $X = \sum_{i=1}^{n} X_i$. Then for any $a > 0$,

$$Pr(|X| > a) \leq 2e^{-\frac{a^2}{2n}}.$$
Application: Set Balancing

**Theorem**

For a random vector $\vec{b}$, with entries chosen independently and with equal probability from the set $\{-1, 1\}$,

$$Pr(||A\vec{b}||_\infty \geq \sqrt{4n \ln n}) \leq \frac{2}{n} \quad (1)$$

- Consider the $i$-th row $\vec{a}_i = a_{i,1}, \ldots, a_{i,n}$.
- Let $k$ be the number of 1’s in that row.
- $Z_i = \sum_{j=1}^{k} a_{i,j} b_{i,j}$.
- If $k \leq \sqrt{4n \ln n}$ then clearly $Z_i \leq \sqrt{4n \ln n}$. 
If $k > \sqrt{4n \log n}$, the $k$ non-zero terms in the sum $Z_i$ are independent random variables, each with probability $1/2$ of being either $+1$ or $-1$.

Using the Chernoff bound:

$$Pr \left\{ |Z_i| > \sqrt{4n \log n} \right\} \leq 2e^{-4n \log n/(2k)} \leq \frac{2}{n^2},$$

where we use the fact that $n \geq k$.

The result follows by union bound ($n$ rows).
Hoeffding’s Inequality

Large deviation bound for more general random variables:

**Theorem (Hoeffding’s Inequality)**

Let $X_1, \ldots, X_n$ be independent random variables such that for all $1 \leq i \leq n$, $E[X_i] = \mu$ and $Pr(a \leq X_i \leq b) = 1$. Then

$$Pr(\left| \frac{1}{n} \sum_{i=1}^{n} X_i - \mu \right| \geq \epsilon) \leq 2e^{-2n\epsilon^2/(b-a)^2}$$

**Lemma**

*(Hoeffding’s Lemma)* Let $X$ be a random variable such that $Pr(X \in [a, b]) = 1$ and $E[X] = 0$. Then for every $\lambda > 0$,

$$E[e^{\lambda X}] \leq e^{\lambda^2(a-b)^2/8}.$$
Proof of the Lemma

Since $f(x) = e^{\lambda x}$ is a convex function, for any $\alpha \in (0, 1)$ and $x = \alpha a + (1 - \alpha) b$,

$$f(X) \leq \alpha f(a) + (1 - \alpha) f(b).$$

For $\alpha = \frac{b-x}{b-a} \in (0, 1)$, $x = \alpha a + (1 - \alpha) b$, thus,

$$e^{\lambda x} \leq \frac{b-x}{b-a} e^{\lambda a} + \frac{x-a}{b-a} e^{\lambda b}.$$

Taking expectation, and using $\mathbf{E}[X] = 0$, we have

$$\mathbf{E}[e^{\lambda X}] \leq \frac{b}{b-a} e^{\lambda a} - \frac{a}{b-a} e^{\lambda b} \leq e^{\lambda^2 (b-a)^2 / 8}.$$

[Proof of the second inequality in the next slide.]
\[ E[e^{\lambda X}] \leq \frac{b}{b-a} e^{\lambda a} - \frac{a}{b-a} e^{\lambda b} = e^{\phi(t)}, \]

for

- \( \phi(t) = -\theta t + \log(1 - \theta + \theta e^t) \),
- \( \theta = \frac{-a}{b-a} \)
- \( t = \lambda(b-a) \)

\( \phi(0) = \phi'(0) = 0 \), and \( \phi''(t) \leq 1/4 \) for all \( t \).

By Taylor’s theorem, for any \( t > 0 \) there is a \( t' \in [0, t] \) such that

\[ \phi(t) = \phi(0) + t\phi'(0) + \frac{1}{2} t^2 \phi''(t') \leq \frac{1}{8} t^2. \]

Thus,

\[ E[e^{\lambda X}] \leq e^{\phi(t)} \leq e^{t^2/8} = e^{\lambda^2(b-a)^2/8}. \]
Proof of the Bound

Let $Z_i = X_i - \mathbb{E}[X_i] = X_i - \mu$ and $Z = \frac{1}{n} \sum_{i=1}^{n} Z_i$.

$$
Pr(Z \geq \epsilon) \leq e^{-\lambda \epsilon} \mathbb{E}[e^{\lambda Z}] \leq e^{-\lambda \epsilon} \prod_{i=1}^{n} \mathbb{E}[e^{\lambda Z_i/n}] \leq e^{-\lambda \epsilon + \frac{\lambda^2 (b-a)^2}{8n}}
$$

Set $\lambda = \frac{4n \epsilon}{(b-a)^2}$ gives

$$
Pr(|\frac{1}{n} \sum_{i=1}^{n} X_i - \mu| \geq \epsilon) = Pr(Z \geq \epsilon) \leq 2e^{-2n \epsilon^2/(b-a)^2}
$$
A More General Version

Theorem

Let $X_1, \ldots, X_n$ be independent random variables with $\mathbb{E}[X_i] = \mu_i$ and $\Pr(B_i \leq X_i \leq B_i + c_i) = 1$, then

$$\Pr(|\sum_{i=1}^{n} X_i - \sum_{i=1}^{n} \mu_i| \geq \epsilon) \leq 2e^{-\frac{2\epsilon^2}{\sum_{i=1}^{n} c_i^2}}$$

Proof.

Homework!
Application: Job Completion

We have $n$ jobs, job $i$ has expected run-time $\mu_i$. We terminate job $i$ if it runs $\beta \mu_i$ time. When will the machine will be free of jobs? $X_i =$ execution time of job $i$. $0 \leq X_i \leq \beta \mu_i$.

$$
Pr(\sum_{i=1}^{n} X_i - \sum_{i=1}^{n} \mu_i \geq \epsilon \sum_{i=1}^{n} \mu_i) \leq 2e^{-\frac{2\epsilon^2 (\sum_{i=1}^{n} \mu_i)^2}{\sum_{i=1}^{n} \beta^2 \mu_i^2}}
$$

Assume all $\mu_i = \mu$

$$
Pr(\sum_{i=1}^{n} X_i - n \mu \geq \epsilon n \mu) \leq 2e^{-\frac{2\epsilon^2 n^2 \mu^2}{n\beta^2 \mu^2}} = 2e^{-2\epsilon^2 n / \beta^2}
$$

Let $\epsilon = \beta \sqrt{\frac{\log n}{n}}$, then

$$
Pr(\sum_{i=1}^{n} X_i - n \mu \geq \beta \mu \sqrt{n \log n}) \leq 2e^{-\frac{2\beta^2 \mu^2 n \log n}{n\beta^2 \mu^2}} = \frac{2}{n^2}
$$