Recall:

**Def[typical sequence]:**

$x^n \in$-typical if $|-1/n \log(p(x^n)) - H(X)| \leq \varepsilon$

It means $2^{-n(H(X) + \varepsilon)} \leq p(x^n) \leq 2^{-n(H(X) - \varepsilon)}$.

**Def[Jointly typical sequence]:**

$x^n y^n \in$-jointly-typical if

(a) $|-1/n \log(p(x^n)) - H(X)| \leq \varepsilon$

(b) $|-1/n \log(p(y^n)) - H(Y)| \leq \varepsilon$

(c) $|-1/n \log(p(x^n y^n)) - H(XY)| \leq \varepsilon$

where $p(x^n y^n) = \prod_{i=1}^{n} p(x_i, y_i)$.

[The strong typicality equivalence of (c) implies those of (a,b).]

**Def[Jointly-typical set]:** $A_{n,\varepsilon} = \{x^n y^n \in$-jointly typical}$
Joint asymptotic equipartition (Joint AEP) theorem:

Let \((X^n,Y^n)\) be sequences of length \(n\)
drawn iid according to \(p(x^n y^n) = \prod_{i=1}^{n} p(x_i y_i)\). Then:

1. \(\forall \delta > 0, \exists n_0 \text{ s.t. } \forall n \geq n_0, \Pr(X^nY^n \in A_{n,\epsilon}) > 1-\delta\)
2. \((1-\delta) 2^n [H(XY)-\epsilon] \leq |A_{n,\epsilon}| \leq 2^n [H(XY)+\epsilon]\)
3. Let \(W^n,Z^n\) be rv's (same sample space as \(X^n,Y^n\)) w/ dist^n \(q(x^n y^n) = p(x^n) p(y^n)\).
   i.e. \(q\) is a dist^n that has the same marginal as \(p\),
   but outcomes \(x^n, y^n\) are independent.
   Then, \(\Pr_q (W^n Z^n \in A_{n,\epsilon}) \leq 2^{-n[I(X:Y)-3\epsilon]}\)
   Also, for large \(n\),
   \((1-\delta) 2^{-n[I(X:Y)+3\epsilon]} \leq \Pr_q (W^n Z^n \in A_{n,\epsilon})\)
Joint asymptotic equipartition (Joint AEP) theorem:

Proof:

[1] Given \( \varepsilon, \delta \), we can apply AEP on \( X^n, Y^n \), and \((XY)^n\).

thus, \( \exists n_0 \) s.t. \( \forall n \geq n_0 \),

the \( \varepsilon \)-typical sets \( T_{n,\varepsilon}^X, T_{n,\varepsilon}^Y, T_{n,\varepsilon}^{XY} \)

all have prob \( \geq 1-\delta/3 \).

\[
A_{n,\varepsilon} = T_{n,\varepsilon}^X \cap T_{n,\varepsilon}^Y \cap T_{n,\varepsilon}^{XY}
\]

\[
A_{n,\varepsilon}^c = T_{n,\varepsilon}^X \cup T_{n,\varepsilon}^Y \cup T_{n,\varepsilon}^{XY}
\]

By the union bound,

\[
\Pr(X^nY^n \in A_{n,\varepsilon}^c) \leq \Pr(X^nY^n \in T_{n,\varepsilon}^X) + \Pr(X^nY^n \in T_{n,\varepsilon}^Y) + \Pr(X^nY^n \in T_{n,\varepsilon}^{XY}) \leq \delta
\]

\[
\Pr(X^nY^n \in A_{n,\varepsilon}) \geq 1-\delta.
\]
Joint asymptotic equipartition (Joint AEP) theorem:

Proof:
[2] Using the same proof as in AEP, condition (c) implies

\[ \forall x^n y^n \in A_{n,\varepsilon}, \]
\[ (1-\delta) 2^{-n(H(XY)+\varepsilon)} \leq p(x^n y^n) \leq 2^{-n(H(XY)-\varepsilon)} \]
Joint asymptotic equipartition (Joint AEP) theorem:

Proof:
[3] Let $W^n, Z^n$ be rv's (same sample space as $X^n, Y^n$) w/ dist $q(x^n, y^n) = p(x^n) p(y^n)$.

\[
\Pr_q (x^n, y^n \in A_{n,\epsilon}) = \frac{\sum_{x^n, y^n \in A_{n,\epsilon}} p(x^n) p(y^n)}{2^n[H(XY)] - \epsilon} \times 2^{-n[H(X)] + \epsilon} \times 2^{-n[H(Y)] + \epsilon} \leq 2^{-n[I(X:Y) + 3\epsilon]}
\]

lower bound on $|A_{n,\epsilon}|$ lower bounds on $p(x^n)$ and $p(y^n)$

\[
(1-\delta) 2^n[H(XY) - \epsilon] \times 2^{-n[H(X)] + \epsilon} \times 2^{-n[H(Y)] + \epsilon} = 2^{-n[I(X:Y) + 3\epsilon]}
\]

upper bound on $|A_{n,\epsilon}|$ upper bounds on $p(x^n)$ and $p(y^n)$
More observations:

Given $y^n \in T_{n,\varepsilon}^Y$, how many $x^n \in T_{n,\varepsilon}^X$ is s.t. $x^n y^n \in A_{n,\varepsilon}$?

Call this set $S_{y^n}$.

$$p(x^n|y^n) = \frac{p(x^n y^n)}{p(y^n)} \approx 2^{-n[H(XY)-H(Y)]} = 2^{-n[H(X|Y)]}$$

$\uparrow$ since $x^n y^n \in A_{n,\varepsilon}$,

$$1 = \sum_{x^n \in S} p(x^n|y^n) \approx |S_{y^n}| 2^{-n[H(X|Y)]}$$

Hence, $|S_{y^n}| \approx 2^{nH(X|Y)}$. Fraction of such $x^n \approx 2^{-nI(X:Y)}$.

Similarly, given $x^n \in T_{n,\varepsilon}^X$, $\approx 2^{nH(Y|X)} y^n$'s are jointly typical with it, and the fraction of such $y^n \approx 2^{-nI(X:Y)}$. 
What's going on?

We're comparing 2 distributions, p and q, on $x^n y^n$. We can list $x^n$'s along a column, $y^n$'s along a row. Can focus only on $x^n$'s, $y^n$'s typical wrt to the common marginal dist'n's. Put $p(x^n y^n), q(x^n y^n)$ in each box.

<table>
<thead>
<tr>
<th>$p(x^n y^n)$</th>
<th>$y^n(1)$</th>
<th>...</th>
<th>$y^n(_)$</th>
<th>$q(x^n y^n)$</th>
<th>$y^n(1)$</th>
<th>...</th>
<th>$y^n(_)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x^n(1)$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$x^n(2)$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$2^n (H(X) + \varepsilon)$

$2^n (H(Y) + \varepsilon)$
What's going on?

1. Mostly $\approx 0$'s except for $2^n[H(XY)+\varepsilon]$ (\(\approx\) equiprobable) entries.
2. Fix a $y^n$ (column). $\approx 2^n[H(X|Y)\pm 2\varepsilon]$ "nonzero" (\(\approx\)equiprobable) entries. A random entry (row) $x^n y^n$ is nonzero with prob $\approx 2^n[H(X|Y)\pm 2\varepsilon] / 2^n[H(X)+\varepsilon] = 2^n[I(X:Y)\pm 3\varepsilon]$. Similarly for fix $x^n$ (row).

So, LHS $\propto$ 0/1 matrix with $\approx$ equal row & column sums. AEP[3] holds row/column-wise.

$2^n(H(Y)+\varepsilon)$

<table>
<thead>
<tr>
<th>$x^n(1)$</th>
<th>...</th>
<th>$x^n(n)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$y^n(1)$</td>
<td>...</td>
<td>$y^n(n)$</td>
</tr>
</tbody>
</table>

basically uniform @ entry $\approx 2^{-n[H(X)+H(Y)\pm 2\varepsilon]}$
Now ready for Shannon's noisy coding theorem.

The rate $R$ is called achievable if, $\forall \ n,$

$\exists \ \eta_n, \ \zeta_n \to 0, \ E_n, \ D_n$ encoder & decoder s.t.

$max_M \ Pr(D_n \circ E_n(M) \neq M) \leq \zeta_n, \ M \in \{1, \ldots , k=2^{n(R-\eta_n)}\}.$

With rules still TBD: Note notation recycling.

$E_n(M) = \hat{x}_M$ (labeled by $M$ with length $n$) $= \begin{bmatrix} x_{M1} & x_{M2} & \ldots & x_{Mn} \end{bmatrix}$

$D_n$ takes $y^n$ to some $W.$

Channel capacity for $N := \sup$ over all achievable rates

$= \sup_{p(x)} I(X:Y) = \sup_{p(x)} I(X:N(X))$
Proof structure:

1. Direct coding theorem:
   a. Show ∀ p(X), I(X:Y) is an achievable rate by analyzing the prob of failure of a random code and random message. That it vanishes ⇒ ∃ at least one code with vanishing average prob of error.

   b. Choose a subset of better codewords that gives vanishing worse case prob of error.

2. Converse: At any higher rate, prob of error → 0.
Part 1a. Let $R = I(X;Y) - \eta$ (will find $\eta$). Need $E_n, D_n$ with prob error $\leq \zeta_n$

* Fix any $p(x)$.

* Write down $A_{\epsilon,n}$ for $XY$ with $\Pr(Y=y|X=x)$ given by $N$.

* $\forall n$ (fixed from now on) let $k = 2^{n(R-\eta_n)}$. (Will find $\eta_n$.)

$E_n$: Pick $k$ codewords (each $x_{Mj}$ chosen iid $\sim p(x)$). Call it $C_n$. Fixed & known to Alice & Bob once choosen.

$x_1 = x_{11}, x_{12}, \ldots, x_{1n}$

$x_2 = x_{21}, x_{22}, \ldots, x_{2n}$

$\ldots$

$x_K = x_{k1}, x_{k2}, \ldots, x_{kn}$

Everything refers to this particular code $C_n$ from now on.
Part 1a. Let $R = I(X:Y) - \eta$ (will find $\eta$). Need $E_n, D_n$ with prob error $\leq \zeta_n$.

* Fix any $p(x)$.
* Write down $A_{\varepsilon,n}$ for $XY$ with $\Pr(Y=y|X=x)$ given by $N$.
* \forall $n$ (fixed from now on) let $k = 2^{n(R-\eta_n)}$. (Will find $\eta_n$.)

$E_n$: Pick $k$ codewords (each $x_{Mj}$ chosen iid $\sim p(x)$). Call it $C_n$. Fixed & known to Alice & Bob once choosen.

\begin{align*}
X_1 &= x_{11}, x_{12}, \ldots, x_{1n} \\
X_2 &= x_{21}, x_{22}, \ldots, x_{2n} \\
&\quad \vdots \\
X_K &= x_{k1}, x_{k2}, \ldots, x_{kn} \\
\end{align*}

Say, $M=2$.

$E_n(2) = x_2 = x_{21} x_{22} \ldots x_{2n}$

Let $y^n = y_1 y_2 \ldots y_n$ be received.

$\Pr(y^n|x_M) = \prod_{i=1}^{n} \Pr(y_i|x_{Mi})$
**Dₙ**: typical set decoding

Given \( y^n \), let \( S_y^n = \{ x^n | x^n y^n \in A_{\varepsilon,n} \} \).

If there is a unique \( x^n \in S_y \), output \( W \) s.t. \( E_n(W) = x^n \).

Else, output \( W = k + 1 \) (representing an error).

In what ways will this fail?

Either - no such \( x^n \) \( \text{Err}_0 \)
- or \( \exists M' \neq M \) with \( E_n(M')y^n \in A_{\varepsilon,n} \) \( \text{Err}_{M'} \)

Prob of error for a given message \( M \) for code \( C_n \):

\[
\lambda_M(C_n) = \Pr(W \neq M | MC_n) = \Pr(\text{Err}_0 \cup_{M' \neq M} \text{Err}_{M'} | MC_n)
\]

Worse case prob of error:

\[
P_{e}^{\max}(C_n) = \max_M \lambda_M(C_n)
\]

Ave (arithmetic) prob of error:

\[
P_{e}^{\text{ave}}(C_n) = 1/k \sum_M \lambda_M(C_n)
\]
Now, upper bound, for this $n$:

$$\Pr_{C_n} \left[ P_{e^{\text{ave}}} (C_n) \right]$$

* just many iid \(\text{wrt a particular } C_n\)
draws to \(X \sim p(x)\) \(\text{but averaged over } M\).

$$= \Pr_{C_n} \left[ 1/k \sum_M \lambda_M (C_n) \right]$$

each \(M\) chosen similarly

thus \(\lambda_M\) independent of \(M\)

$$= \Pr_{C_n} \lambda_1 (C_n)$$

$$= \Pr_{C_n} (W \neq 1 | M=1) = \Pr_{C_n} (\text{Err}_0 \cup_{M' \neq 1} \text{Err}_{M'} | M=1)$$

\(\bigcup\) \(\text{bdd}\)

$$\leq \Pr_{C_n} (\text{Err}_0 | M=1) + (k-1) \Pr_{C_n} (\text{Err}_{M' \neq 1} | M=1)$$
Bounding $\Pr_{C_n}(\text{Err}_0|M=1)$:

By joint AEP [1], $\forall \delta > 0$, $\exists n_0$ s.t. $\forall n \geq n_0$,

$$\Pr(X^nY^n \in A_{n,\varepsilon}) > 1 - \delta$$

Given $n$, $\exists \delta_n, \varepsilon_n$ for which $\Pr(X^nY^n \in A_{n,\varepsilon_n}) > 1 - \delta_n$.

[And $\delta_n, \varepsilon_n \to 0$.]

Here:

$x_{M=1} = x_{11} \ldots x_{1n}$ drawn iid $\sim p(x)$, and

$y^n = y_1 \ldots y_n$ drawn $\sim p(y|x_{1i})$

Thus, $x_{1i}y_i$ iid $\sim p(xy)$ and $\Pr(x_{M=1} y^n \in A_{n,\varepsilon_n}) > 1 - \delta_n$.

$\Pr_{C_n}(\text{Err}_0|M=1) \leq \delta_n$.

BACK 1 SLIDE.
Bounding $\Pr_{\mathcal{C}_n}(\text{Err}_{M' \neq 1} | M=1) = \Pr_{\mathcal{C}_n}(x_{M'} y^n \in A_{n,\varepsilon_n})$:

By joint AEP [3], $\forall \delta > 0$, $\exists n_0$ s.t. $\forall n \geq n_0$,

\[ W^n, Z^n \sim q(x^n y^n) = p(x^n) p(y^n). \]

\[ (1-\delta) 2^{-n[I(X:Y)+3\varepsilon]} \leq \Pr_q(W^n Z^n \in A_{n,\varepsilon}) \leq 2^{-n[I(X:Y)-3\varepsilon]} \]

Given $n$, $\exists \delta_n, \varepsilon_n$ for which

\[ (1-\delta_n) 2^{-n[I(X:Y)+3\varepsilon_n]} \leq \Pr_q(W^n Z^n \in A_{n,\varepsilon_n}) \leq 2^{-n[I(X:Y)-3\varepsilon_n]} \]

[And $\delta_n, \varepsilon_n \to 0$.]

Here:

$x_{M'} = x_{M'1} \ldots x_{M'n}$ drawn independent of $x_1$ and

$y^n = y_1 \ldots y_n$ iid $\sim p(y|x_{1i})$, independent of $x_{M'}$.

Thus, $\Pr_{\mathcal{C}_n}(\text{Err}_{M' \neq 1} | M=1) \leq 2^{-n[I(X:Y)-3\varepsilon_n]}$. 
Now, upper bound, for this $n$:

$$\Pr_{C_n} \left[ P_{e \text{ ave}} (C_n) \right]$$

$$\leq \Pr_{C_n} (\text{Err}_0 | M=1) + (k-1) \Pr_{C_n} (\text{Err}_{M' \neq 1} | M=1)$$

$$\leq \delta_n + (k-1) \ 2^{-n[I(X:Y)-3\varepsilon_n]}$$

but $k=2^n(R-\eta_n)$, $R=I(X:Y)-\eta$

$$\leq \delta_n + 2^n[I(X:Y)-\eta-\eta_n] \ 2^{-n[I(X:Y)-3\varepsilon_n]}$$

$$\leq \delta_n + 2^n[-\eta-\eta_n+3\varepsilon_n]$$

choose $\eta = \text{small constant}$

$$\leq \delta_n + 2^{-n\eta} =: \zeta_{ \text{ave} } \n_n.$$  

Thus, $\exists \ C_n (E_n,D_n)$ with $P_{e \text{ ave}} (C_n) \leq \zeta_{ \text{ave} } \n_n$. 
Part 1b.

Worse case prob of error: \( P_e^{\text{max}}(C_n) = \max_M \lambda_M(C_n) \)

Ave (arithmetic) prob of error: \( P_e^{\text{ave}}(C_n) = \frac{1}{k} \sum_M \lambda_M(C_n) \)

For the code \( C_n \) obtained in 1a, order \( M \) in ascending order of \( \lambda_M(C_n) \). Keep the first half. Call this new code \( C'_n \).

\[ P_e^{\text{ave}}(C_n) = \frac{1}{k} \sum_M \lambda_M(C_n) \geq \frac{1}{k} \left[ \sum_{M \not\in C'_n} P_e^{\text{max}}(C'_n) + \sum_{M \in C'_n} \lambda_M(C_n) \right] \geq \frac{1}{2} P_e^{\text{max}}(C'_n). \]

Thus, \( C'_n \) has worse case error prob \( \leq \zeta_n^{\text{ave}}/2 =: \zeta_n \to 0. \)

[rate for \( C'_n = \text{rate for } C_n - 1/n. \)]

Thus \( R=I(X:Y)-\eta \) achievable on \( C'_n \) for any \( \eta>0. \)

"Sup over \( R\)" gives capacity \( \geq \max_{p(x)} I(X:Y). \)