"If $\tilde{R} > G(C)$: $\exists \delta > 0$ \, $\forall N \, \exists n > N$ : $\delta$ code with $\frac{K}{N} \geq \tilde{R} \& \, P_e \leq \delta$"  

Tools:  
1. **Data Processing Inequality (DPI)** for $A \rightarrow B \rightarrow C$ Markov chain:  
   \[ I(B: C) \geq I(A: C) \, \& \, H(A|B) \leq H(A|C) \]  
2. If $X^N$ arbitrary and $Y^N$ channel output: 
   \[ I(X^N: Y^N) \leq \sum_{i=1}^{N} I(X_i: Y_i) \leq N \cdot G(C) \]  
3. **Fano's inequality** for $S \rightarrow T \rightarrow \hat{S}$ Markov chain, $p = P(C(S \neq \hat{S}))$ 
   \[ H([P(1-P(S))] + p \cdot \log \#AS \geq H(C(S|\hat{S})) \geq H(S|T) \]  

Proof of the converse: Consider $(N, K)$-code with $\frac{K}{N} \geq \tilde{R} > G$.  
Let $S \in \{1, \ldots, 2^K\}$ uniform. Recall: $S \rightarrow X^N \rightarrow Y^N \rightarrow \hat{S}$.  

Then:  
1. $H(S|Y^N) = H(S) - I(S: Y^N) \geq H(S) - I(X^N; Y^N)$  
2. $S \rightarrow X^N \rightarrow Y^N$ Markov chain \[ H(S|Y^N) \leq 1 + P(C(S \neq \hat{S})) \cdot \log \#AS = 1 + P_e \cdot K \] \[ \rightarrow \, K - N \cdot G \leq 1 + P_e \cdot K \] \[ \rightarrow \, P_e \geq \frac{1}{K} (K - N \cdot G - 1) = 1 - \frac{N \cdot G}{K} - \frac{1}{K} \geq \frac{1 - \frac{G}{K}}{ \frac{1}{NK} } \rightarrow \] \[ \frac{1 - \frac{G}{K}}{ \frac{1}{NK} } \] Can never go below this for large enough $N$ \[ \text{Are we happy? What questions does Shannon's theorem leave unaddressed?} \] \[ \text{Algorithmics, large $N$, ... how to even compute $G$?} \]
Shannon’s Theorem vs. Practice

Need large block size $N$ for joint typicality vs. fixed packet size
Codebook $X^N(1),\ldots,X^N(2^K)$ exponentially large in $N$ (if $P > 0$)
Random codes vs predictable performance

A family of codes is "very good" if $\frac{k}{N} \to 0$ & $P_C \to 0$
"good" if $\frac{k}{N} \geq R$ & $P_C \to 0$ for some $\tilde{R} > 0$
"bad" otherwise

... and "practical" if efficient encode + decode

In practice:
* most codes are linear (e.g., linear function of $S^k$)
* "easy" to come up with "plausible" encoders — but optimal decoding is in general (NP) hard! unlike for compression!

$$\sigma_{opt}(y^N) = \arg\max_S P(S|y^N)$$

Why? If $P(s)$ arbitrary prior, want to choose $S$ to maximize $P(S=y^N)$

$$= \sum_{y^N} P(S=s(y^N), T^N=y^N)$$

Choose $S=s(y^N)$ that maximizes $P(s(y^N)=P(s|y^N)$

For erasure channel:

$S_1 \oplus S_2 \oplus S_3 \oplus t_1 = 0$
$S_2 \oplus \ldots \oplus S_k \oplus t_2 = 0$

* types of decoders: "algebraic" vs. "iterative"

Types of codes:
* block codes: e.g. Hamming, Reed-Salomon, LDPC codes
  Storage, bar codes, Sat comm
* convolutional: e.g. turbo codes
  $3G/4G/LTE$, Sat comm.
* linear streaming codes