1. Let R be an achievable rate, i.e., for any $\epsilon > 0$, there exists for sufficiently large n and (n, M) code such that

$$\frac{1}{n} \log M > R - \epsilon \quad \text{and} \quad \lambda_{max} < \epsilon.$$

2. Consider

$$\begin{aligned}
\log M &= H(W) \\
&= H(W|\hat{W}) + I(W;\hat{W}) \\
&\leq H(W|\hat{W}) + I(\mathbf{X};\mathbf{Y}) \\
&= H(W|\hat{W}) + h(\mathbf{Y}) - h(\mathbf{Y}|\mathbf{X}) \\
&\leq H(W|\hat{W}) + \sum_{i=1}^{n} h(Y_i) - h(\mathbf{Y}|\mathbf{X}) \\
&= H(W|\hat{W}) + \sum_{i=1}^{n} h(Y_i) - \sum_{i=1}^{n} h(Y_i|X_i) \\
&= H(W|\hat{W}) + \sum_{i=1}^{n} I(X_i;Y_i).
\end{aligned}$$

3. Let V be a mixing random variable distributed uniformly on $\{1, 2, \cdots, n\}$ which is independent of X_i, $1 \leq i \leq n$.

4. Let $X = X_V$ and Y be the output of the channel with X being the input.

5. Then

$$\begin{aligned}
E\kappa(X) &= EE[\kappa(X)|V] \\
&= \sum_{i=1}^{n} \Pr\{V = i\} E[\kappa(X)|V = i] \\
&= \sum_{i=1}^{n} \Pr\{V = i\} E[\kappa(X_i)|V = i] \\
&= \sum_{i=1}^{n} \frac{1}{n} E\kappa(X_i) \\
&= E\left[\frac{1}{n}\sum_{i=1}^{n} \kappa(X_i)\right] \\
&\leq P.
\end{aligned}$$

6. By the concavity of mutual information with respect to the input distribution,

$$\frac{1}{n}\sum_{i=1}^{n} I(X_i;Y_i) \leq I(X;Y) \leq C(P).$$

7. It follows that

$$n(R - \epsilon) < \log M \leq H(W|\hat{W}) + nC(P).$$

8. Invoke Fano's inequality to conclude that $R \leq C(P)$.