Why is $S(\rho) \leq H(\{p_i\})$?

We know that when Alice sends a state from the ensemble $\{p_i, |\psi_i\rangle\}$, such that $\rho = \sum_i p_i |\psi_i\rangle\langle\psi_i|$ the optimal number of qubits needed is $S(\rho)$.

Now, for any state $\rho = \sum_i p_i \rho_i$, where ρ_i has an orthogonal support i.e., ρ_i's are mutually orthogonal or

$$\mathrm{Tr}(\rho_i \rho_j) = 0, \quad \forall\, i \neq j,$$

we know that;

$$S(\rho) = H(\{p_i\}) + \sum_i p_i S(\rho_i)$$

For any general $\{\rho_i\}$, this relation becomes an inequality i.e., (see Nielsen and Chuang; Section 11.3.6).

$$S(\rho) \leq H(\{p_i\}) + \sum_i p_i S(\rho_i)$$

If $\rho_i = |\psi_i\rangle\langle\psi_i|$ such that $\rho = \sum_i p_i |\psi_i\rangle\langle\psi_i|$

then we have;

$$S(\rho) \leq H(\{p_i\})$$

as $S(\rho_i) = S(|\psi_i\rangle\langle\psi_i|) = 0$