The DCN is composed of two atomic operations $\mathcal{S}_\theta$ and $\mathcal{M}_\phi$.

- Split block $\mathcal{S}_\theta$ is a *Set2Set* model that splits an input set X into a binary partition $X = X_0 \sqcup X_1$. It does so by sampling from element-wise binary probabilities $p_\theta(z_m \mid X)$. It is applied recursively until scale J. This gives a probability distribution over hierarchical partitions of X of depth J:

$$\mathcal{P}(X) = \{X_{j,k}\ ; 0 \leq j < J ; 0 \leq k < n_j\}\ ,\ \text{with } X_{j,k} = X_{j+1,2k} \sqcup X_{j+1,2k+1}$$

$$\{X_{j+1,2k}, X_{j+1,2k+1}\} = \mathcal{S}_\theta(X_{j,k})\ ,\ \text{with } X_{j,k} = X_{j+1,2k} \sqcup X_{j+1,2k+1}\ ,\ (j < J, k \leq 2^j)$$

$$\mathcal{P}(X) \sim \mathbf{S}_\theta(X)$$

- Merge block $\mathcal{M}_\phi$ is a PtrNet that takes two sequences Y_0, Y_1 and outputs a permutation of a subsequence of the union of both. PtrNet output is a stochastic matrix giving the probability distributions over indexes. Binarize along rows through argmax to find the permutation. The merge block is applied recursively traversing the hierarchical partition $\mathcal{P}(X)$ upwards.

$$Y_{j,k} = \mathcal{M}_\phi(Y_{j+1,2k}, Y_{j+1,2k+1})\ \ ,\ (1 \leq k \leq 2^j, j < J)\ , \text{and } \hat{Y} = \mathcal{M}_\phi(Y_{1,0}, Y_{1,1})\ .$$