Idealized equations:

§1. Control Unit

$$q_i = U_i^{[d\times 2d]} q + b_i^{[d]} \tag{c0}$$

$$cq_i = W_{\mathsf{cq}}^{[d\times d]} c_{i-1} + q_i \tag{c1}$$
$$ca_{is} = W_{\mathsf{ca}}^{[1\times d]} (cq_i \odot \mathbf{cw}_s) \tag{c2.1}$$
$$cv_{is} = \mathrm{softmax}(ca_{is}) \tag{c2.2}$$
$$\mathbf{c}_i = \sum_s cv_{is}\, \mathbf{cw}_s \tag{c2.3}$$

Only eqn. (c0) above is *position dependent*

§2. Read Unit

$$\hat{\mathbf{k}}_{hw} = U_{\mathsf{proj}}^{[d\times d]} \mathbf{k}_{hw} + b_{\mathsf{proj}}^{[d]} \tag{r0}$$

$$I_{ihw} = \mathbf{m}_{i-1} \odot \hat{\mathbf{k}}_{hw} \tag{r1}$$
$$I'_{ihw} = \hat{\mathbf{k}}_{hw} + W_{\mathsf{I}}^{[d\times d]} I_{ihw} + b_{\mathsf{I}}^{[d]} \tag{r2}$$
$$ra_{ihw} = W_{\mathsf{ra}}^{[1\times d]} (\mathbf{c}_i \odot I'_{ihw}) \tag{r3.1}$$
$$rv_{ihw} = \mathrm{softmax}(ra_{ihw}) \tag{r3.2}$$
$$\mathbf{r}_i = \sum_s rv_{ihw}\, \hat{\mathbf{k}}_{hw} \tag{r3.3}$$

Note that $\hat{\mathbf{k}}_{hw}$ can be *precomputed* before the reasoning steps.

§3. Write Unit

$$\mathbf{m}_i = W_{\mathsf{m}}^{[d\times d]} \mathbf{r}_i + b_{\mathsf{m}}^{[d]} \tag{w1}$$