Question 1:

Given: $q(X_{1-T}|X_0) = \prod_{t=1}^{T} q(X_t|X_{t-1})$

Show: $q(X_{1-T}|X_0) = q(X_T|X_0)\ \prod_{t=T}^{2} q(X_{t-1}|X_t, X_0)$

Proof :

$$q(X_T|X_0) \prod_{t=T}^{2} q(X_{t-1}|X_t, X_0)$$

$$= q(X_T|X_0) \prod_{t=T}^{2} \frac{q(X_{t-1}, X_t, X_0)}{q(X_t, X_0)}$$

$$= q(X_T|X_0) \prod_{t=T}^{2} \frac{q(X_t|X_{t-1})q(X_{t-1}|X_0)q(0)}{q(X_t|X_0)q(X_0)}$$

$$= q(X_T|X_0) \prod_{t=T}^{2} \frac{q(X_t|X_{t-1})q(X_{t-1}|X_0)}{q(X_t|X_0)}$$

$$= \cancel{q(X_T|X_0)}\ \frac{q(X_T|X_{T-1})\cancel{q(X_{T-1}|X_0)}}{\cancel{q(X_T|X_0)}}\ \frac{q(X_{T-1}|X_{T-2})\cancel{q(X_{T-2}|X_0)}}{\cancel{q(X_{T-1}|X_0)}} \dots$$

$$\frac{q(X_3|X_2)\cancel{q(X_2|X_0)}}{\cancel{q(X_3|X_0)}}\ \frac{q(X_2|X_1)q(X_1|X_0)}{\cancel{q(X_2|X_0)}}$$

$$= q(X_T|X_{T-1})q(X_{T-1}|X_{T-2})q(X_{T-2}|X_{T-3}) \dots q(X_2|X_1)q(X_1|X_0)$$

$$= \prod_{t=1}^{T} q(X_t|X_{t-1}) = q(X_{1-T}|X_0)$$

Question 2:

(4) $q(X_t|X_0) = N(X_t; \sqrt{\overline{a}_t}X_0, (1-\overline{a}_t)I)$

$$q(X_t|X_{t-1}) := N\left(X_t; \sqrt{1-\beta_t}X_{t-1}, \beta_t I\right)\ ,\ \overline{a}_t := \prod_{s=1}^{T} a_s\ ,\ a_t := 1 - \beta_t$$

$$X_t = \sqrt{1-\beta_t}X_{t-1} + \beta_t I$$

$$X_{t-1} = \sqrt{1-\beta_{t-1}}X_{t-2} + \beta_{t-1} I$$

$$\therefore X_t = \sqrt{1-\beta_t}(\sqrt{1-\beta_{t-1}}X_{t-2} + \beta_{t-1} I) + \beta_t I$$

$$\therefore X_t = \sqrt{\overline{a}_t}X_0 + (1-\overline{a}_t)I$$

$$\therefore q(X_t|X_0) = N(X_t; \sqrt{\overline{a}_t}X_0, (1-\overline{a}_t)I)$$

(6) $q(X_{t-1}|X_t, X_0) = N\left(X_{t-1}; \tilde{u}_t(X_t, X_0), \tilde{\beta}_t I\right)$

$where\ \tilde{u}_t(X_t, X_0) = \frac{\sqrt{\bar{a}_{t-1}}\beta_t}{1-\bar{a}_t} X_0 + \frac{\sqrt{\bar{a}_t}(1-\bar{a}_{t-1})}{1-\bar{a}_t} X_t$ and $\tilde{\beta}_t \coloneqq \frac{1-\bar{a}_{t-1}}{1-\bar{a}_t}\beta_t$

$$q(X_{t-1}|X_t, X_0) = \frac{q(X_t|X_{t-1}, X_0)q(X_{t-1}|X_0)}{q(X_t|X_0)}$$

$$= \frac{N\left(X_t; \sqrt{a_t}X_{t-1}, (1-a_t)I\right)N\left(X_{t-1}; \sqrt{\bar{a}_{t-1}}X_0, (1-\bar{a}_{t-1})I\right)}{N\left(X_t; \sqrt{\bar{a}_t}X_0, (1-\bar{a}_t)I\right)}$$

$$\propto \exp\left\{-\frac{1}{2}\left[\frac{\left(-2\sqrt{a_t}X_tX_{t-1} + a_tX_{t-1}^2\right)}{1-a_t} + \frac{\left(X_{t-1}^2 - 2\sqrt{\bar{a}_{t-1}}X_{t-1}X_0\right)}{1-\bar{a}_{t-1}} + C(X_t, X_0)\right]\right\}$$

$$\propto \exp\left\{-\frac{1}{2}\left[\left(\frac{a_t}{1-a_t} + \frac{1}{1-a_{t-1}}\right)X_{t-1}^2 - 2\left(\frac{\sqrt{a_t}X_t}{1-a_t} + \frac{\sqrt{\bar{a}_{t-1}}X_0}{1-\bar{a}_{t-1}}\right)X_{t-1}\right]\right\}$$

$$= \exp\left\{-\frac{1}{2}\left(\frac{1-\bar{a}_t}{(1-a_t)(1-a_{t-1})}\right)\left[X_{t-1}^2 - 2\frac{\frac{\sqrt{a_t}X_t}{1-a_t} + \frac{\sqrt{\bar{a}_{t-1}}X_0}{1-\bar{a}_{t-1}}}{\frac{1-\bar{a}_t}{(1-a_t)(1-a_{t-1})}}X_{t-1}\right]\right\}$$

$$= \exp\left\{-\frac{1}{2}\left(\frac{1}{\frac{(1-a_t)(1-a_{t-1})}{1-\bar{a}_{t-1}}}\right)\left[X_{t-1}^2 - 2\frac{\left(\frac{\sqrt{a_t}X_t}{1-a_t} + \frac{\sqrt{\bar{a}_{t-1}}X_0}{1-\bar{a}_{t-1}}\right)(1-a_t)(1-a_{t-1})}{1-\bar{a}_t}X_{t-1}\right]\right\}$$

$$= \exp\left\{-\frac{1}{2}\left(\frac{1}{\frac{(1-a_t)(1-a_{t-1})}{1-\bar{a}_{t-1}}}\right)\left[X_{t-1}^2 - 2\frac{\sqrt{a_t}X_t(1-a_{t-1}) + \sqrt{\bar{a}_{t-1}}X_0(1-a_t)}{1-\bar{a}_t}X_{t-1}\right]\right\}$$

$$\propto N\left(X_{t-1}; \frac{\sqrt{a_t}X_t(1-a_{t-1}) + \sqrt{\bar{a}_{t-1}}X_0(1-a_t)}{1-\bar{a}_t}, \frac{(1-a_t)(1-a_{t-1})}{1-\bar{a}_{t-1}}I\right)$$

$$= N\left(X_{t-1}; \tilde{u}_t(X_t, X_0), \tilde{\beta}_t I\right)$$

(8) $L_{t-1} = E_q\left[\frac{1}{2\sigma_t^2}\left|\left|\tilde{u}_t(X_t, X_0) - \tilde{u}_\theta(X_t, t)\right|\right|^2\right] + C$

$D_{KL}(q(X_{t-1}|X_t, X_0)||p_\theta(X_{t-1}|X_t))$, $p_\theta(X_{t-1}|X_t) = N(X_{t-1}; \mu_\theta(X_t, \mathrm{t}), \sigma_t^2 I)$

$$D_{KL}(N(X_{t-1}; \tilde{\mu}_t(X_t, X_0), \sigma_t^2 I)||N(X_{t-1}; \mu_\theta(X_t, \mathrm{t}), \sigma_t^2 I))$$

$$= \frac{1}{2}\left[\left(\mu_\theta(X_t, \mathrm{t}) - \tilde{\mu}_t(X_t, X_0)\right)^T \Sigma_q(t)^{-1}\left(\mu_\theta(X_t, \mathrm{t}) - \tilde{\mu}_t(X_t, X_0)\right)\right]$$

$$= \frac{1}{2}\left[\left(\mu_\theta(X_t, \mathrm{t}) - \tilde{\mu}_t(X_t, X_0)\right)^T (\sigma_t^2(t)I)^{-1}\left(\mu_\theta(X_t, \mathrm{t}) - \tilde{\mu}_t(X_t, X_0)\right)\right]$$

$$= \frac{1}{2\sigma_t^2}[||\mu_\theta(X_t, \mathrm{t}) - \mu_q(X_t, X_0)||_2^2]$$

From Equation(5), $L_{t-1} = E_q(D_{KL}\left(q(X_{t-1}|X_t, X_0)\middle|\middle|p_\theta(X_{t-1}|X_t)\right))$

$$= E_q\left[\frac{1}{2\sigma_t^2}\left|\left|\tilde{u}_t(X_t, X_0) - \tilde{u}_\theta(X_t, t)\right|\right|^2\right] + C$$