Since S(t) = Q(t) + B(t) at any time t, and by linearity of integrals,

$$\int_{0}^{T} S(t)dt = \int_{0}^{T} [Q(t) + B(t)]dt = \int_{0}^{T} Q(t)dt + \int_{0}^{T} B(t)dt,$$

where T is the fixed ending time of the simulation (T = 20 for this example). Dividing this relation through by T, we see that the time-average level of S(t) can be obtained by adding the time-average level of Q(t) to the time-average level of S(t) could have been obtained from the existing output without adding any state variables.

In general, to get the maximum level of the pointwise sum of two curves, we *cannot* just add the maximum values of the two curves together since these maxima might not have occurred at the same time, so it would generally be necessary to keep the new state variable representing the new curve to get the maximum level. In this particular example, though, the logic implies that the maximum of the queue length Q(t) would have to be obtained when the server is busy (if that ever happens), i.e., when B(t) = 1, its maximum, so that the maxima of the two curves *do* occur at the same time. So, *in this example*, it is valid to add the maximum of Q(t) to the maximum of B(t) to get the maximum of S(t), so that the new state variable tracking S(t) would not have been necessary.