1 The mathematics of two-way ranging

1.1 Single-sided two-way ranging error due to clock offset

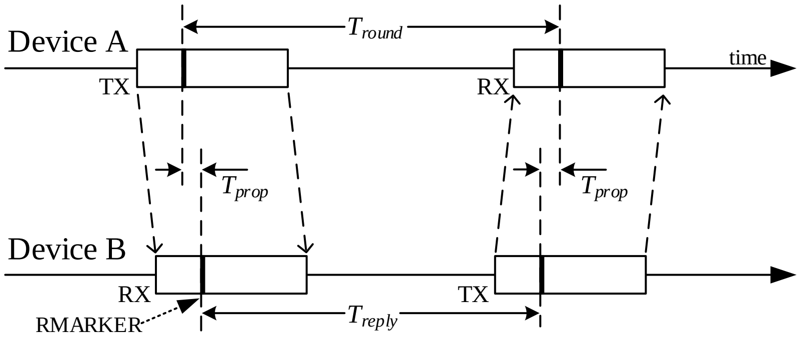

Figure 1 – single-sided two-way ranging

The operation of SS-TWR is as shown in Figure 1, where device A initiates the exchange and device B responds to complete the exchange. Each device precisely timestamps the transmission and reception times of the message frames, and so can calculate times T_{round} and T_{reply} by simple subtraction. Hence, the resultant time-of-flight, T_{prop}, may be estimated by the equation:

$$\hat{T}_{prop} = \frac{1}{2}\left(T_{round} - T_{reply}\right)$$

With reference to Figure 1, the times T_{round} and T_{reply} are measured independently by device A and B using their respective the local clocks, which have some clock offset error e_A and e_B from their nominal frequency. As a result T_{prop} is then actually:

$$T_{prop} = \frac{1}{2}\left(T_{round} \times (1 + e_A) - T_{reply} \times (1 + e_A)\right)$$

And, the error in the range estimate equation (given above) is:

$$\hat{T}_{prop} - T_{prop} = \frac{1}{2}\left(T_{round} - T_{reply}\right) - \frac{1}{2}\left(T_{round} \times (1 + e_A) - T_{reply} \times (1 + e_B)\right)$$

Or:

$$\hat{T}_{prop} - T_{prop} = \frac{1}{2}\left(-T_{round} \times e_A + T_{reply} \times e_B\right)$$

Or, since T_{reply} is 500 to 5000 times bigger than T_{prop}, for the purpose of simplifying the error estimate we can take T_{reply} as (almost) the same as T_{round} and can be use T_{reply} in place of T_{round} to give an error estimate:

$$error = \hat{T}_{prop} - T_{prop} \approx \frac{1}{2}(e_B - e_A) \times T_{reply}$$

Based on this equation, Table 1 presents the typical errors in SS-TWR time-of-flight estimation depending on the reply time, T_{reply}, and the total clock offset error.

Table 1 – typical clock induced error in SS-TWR time-of-flight estimation

T_{reply} \ clock error	2 ppm	5 ppm	10 ppm	20 ppm	40 ppm
100 µs	0.1 ns	0.25 ns	0.5 ns	1 ns	2 ns
200 µs	0.2 ns	0.5 ns	1 ns	2 ns	4 ns
500 µs	0.5 ns	1.25 ns	2.5 ns	5 ns	10 ns
1 ms	1 ns	2.5 ns	5 ns	10 ns	20 ns
2 ms	2 ns	5 ns	10 ns	20 ns	40 ns
5 ms	5 ns	12.5 ns	25 ns	50 ns	100 ns

Table 1 shows that quite accurate results are achievable when short messages are used (e.g. with shorter preamble lengths and higher data rates) to give small reply time and more accurate clock sources are employed.

Note: the reply time T_{reply} is not just the RX-to-TX turnaround time but also includes the message length.

An error of 1 ns in time-of-flight is equivalent to a 30 cm error in distance.

1.2 Derivation of (asymmetric) double-sided two-way ranging formula

Figure 2 shows the two round trips of double sider ranging (with the optimal 3 message exchange) and defines the terms used for the round trip times, R_a and R_b, and the reply times, D_a and D_b, for the pair of devices, A and B, participating in the two-way ranging exchange to measure the time-of-flight, T_f.

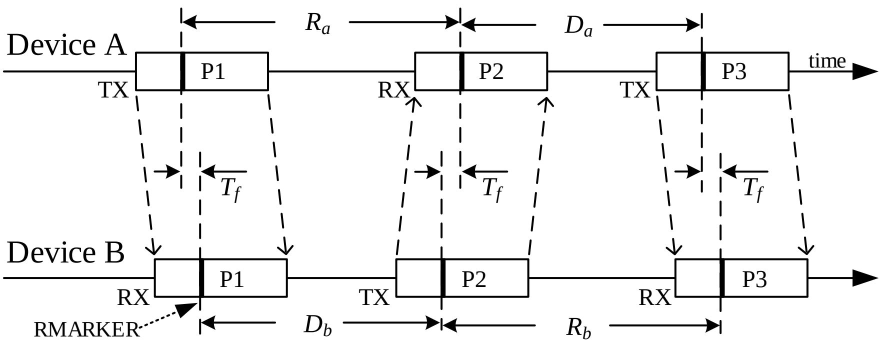

Figure 2 – terms used in deriving the double-sided two-way ranging formula

In Figure 2, device A transmits a message, P1, to device B. Device B receives this message a short time later, D_b, it transmits a message, P2, back to device A. Message P2 arrives at device A at a time R_a after it transmitted message P1. These then have the relationship:

$$R_a = 2T_f + D_b \qquad (1)$$

So

$$T_f = \frac{1}{2}(R_a - D_b) \qquad (2)$$

In practice, the times are measured by real clocks in A and B, which will run independently either faster or slower than an ideal clock, synchronized to their local reference frequency generator which can be assumed to be a constant frequency over the duration of the ranging exchange. Let us say that Clocks A and B run respectively at k_a and k_b times the frequency of an ideal, true, clock. Any time measurements will be multiplied by these constants, k_a or k_b. Denoting the actual time estimates for R_a and D_a as $\hat{R}_a$ and $\hat{D}_a$ respectively, and similarly $\hat{R}_b$ and $\hat{D}_b$ as the estimates of R_b and D_b. Then, since R_a is measured at A by A's clock:

$$\hat{R}_a = k_a R_a \tag{3}$$

And, similarly

$$\hat{D}_a = k_a D_a \tag{4}$$

$$\hat{R}_b = k_b R_b \tag{5}$$

$$\hat{D}_b = k_b D_b \tag{6}$$

Using $\hat{R}_a$, $\hat{D}_a$ etc. as estimates for R_a, D_a etc. we introduce a measurement error, so for example if $\hat{T}_f$ is an estimate of T_f we can say for a single round trip exchange

$$\hat{T}_{f1} = \frac{1}{2}(\hat{R}_a - \hat{D}_b) = \frac{1}{2}(k_a R_a - k_b D_b) \tag{7}$$

And, the error in the estimation is

$$\hat{T}_{f1} - T_f = \frac{1}{2}((k_a - 1)R_a - (k_b - 1)D_b) \tag{8}$$

For the UWB physical layer the values of the expressions $(k_a - 1)$ and $(k_b - 1)$ may be up to 20 ppm, i.e. $20\text{x}10^{-6}$, and for accurate ranging it is important to keep the error below 100ps ($1\text{x}10^{-10}$) which means the delays e.g. $\hat{D}_b$ must be kept below about 5 µs which is not possible where even short UWB frames are typically > 100 µs long. The solution is to use two round trip delays.

We know from (4) that

$$D_a = \frac{\hat{D}_a}{k_a} \tag{10}$$

And similarly

$$D_b = \frac{\hat{D}_b}{k_b} \tag{11}$$

Then from (10) and (1) we can say

$$R_a = 2T_f + \frac{\hat{D}_b}{k_b} \tag{12}$$

And from (12) and (3)

$$\hat{R}_a = 2k_a T_f + \frac{k_a \hat{D}_b}{k_b} \tag{13a}$$

And similarly

$$\hat{R}_b = 2k_b T_f + \frac{k_b \hat{D}_a}{k_a} \tag{13b}$$

In (12) and (13) we have quantities that we can measure, $\hat{R}_a$, $\hat{R}_b$, $\hat{D}_a$ and $\hat{D}_b$. But we have no way of measuring k_a or k_b. And the errors in these quantities swamp the value of T_f. There is however one thing we can do. If we multiply $\hat{R}_a$ by $\hat{R}_b$, the bulk of the value of the product will be the product of $\hat{D}_a$ and $\hat{D}_b$. For this term in the product, the k_a and k_b constants cancel each other out.

Then, from (12) & (13)

$$\hat{R}_a\hat{R}_b = 4k_ak_bT_f^{\,2} + \frac{k_a\hat{D}_b}{k_b}\frac{k_b\hat{D}_a}{k_a} + 2k_bT_f\frac{k_a\hat{D}_b}{k_b} + 2k_aT_f\frac{k_b\hat{D}_a}{k_a} \tag{14}$$

$$\hat{R}_a\hat{R}_b = 4k_ak_bT_f^{\,2} + \hat{D}_a\hat{D}_b + 2T_f(k_a\hat{D}_b + k_b\hat{D}_a) \tag{15}$$

$$\hat{R}_a\hat{R}_b - \hat{D}_a\hat{D}_b = 4k_ak_bT_f^{\,2} + 2T_f(k_a\hat{D}_b + k_b\hat{D}_a) \tag{16}$$

And, from (16) and (13)

$$\frac{\hat{R}_a\hat{R}_b-\hat{D}_a\hat{D}_b}{\hat{R}_a+\hat{D}_a} = \frac{4k_ak_bT_f^{\,2}+2T_f(k_a\hat{D}_b+k_b\hat{D}_a)}{2k_aT_f+\frac{k_a\hat{D}_b}{k_b}+\hat{D}_a} \tag{17}$$

On left hand side, taking out $2T_f$ and multiplying above and below by k_b

$$\frac{\hat{R}_a\hat{R}_b-\hat{D}_a\hat{D}_b}{\hat{R}_a+\hat{D}_a} = 2T_fk_b\,\frac{2k_ak_bT_f+(k_a\hat{D}_b+k_b\hat{D}_a)}{2k_ak_bT_f+k_a\hat{D}_b+k_b\hat{D}_a} \tag{18}$$

So finally

$$\frac{\hat{R}_a\hat{R}_b-\hat{D}_a\hat{D}_b}{\hat{R}_a+\hat{D}_a} = 2T_fk_b \approx 2T_f \tag{19a}$$

And similarly

$$\frac{\hat{R}_a\hat{R}_b-\hat{D}_a\hat{D}_b}{\hat{R}_b+\hat{D}_b} = 2T_fk_a \approx 2T_f \tag{19b}$$

We now have two possible estimates for T_f, and since k_a and k_b are very close to unity, i.e. $0.99998 < k_a, k_b < 1.00002$, we can estimate T_f as follows

$$\hat{T}_{fa} = \frac{1}{2}\frac{\hat{R}_a\hat{R}_b-\hat{D}_a\hat{D}_b}{\hat{R}_b+D_b} \tag{20}$$

$$\hat{T}_{fb} = \frac{1}{2}\frac{\hat{R}_a\hat{R}_b-\hat{D}_a\hat{D}_b}{\hat{R}_a+\hat{D}_a} \tag{21}$$

These estimates are very close to the actual T_f because k_a and k_b are very close to one and, crucially, their accuracy is independent of the response delays employed at A and at B.

Whether a system should use formula (20) or formula (21) would depend on whether it expects which clock it expects to be more accurate. For example, if the system knows B has a higher accuracy clocks then it should use formula 21. If it expects neither to be more accurate than the other and it is more accurate to use the average result from (20) and (21) since this will always be as good as, or better than, the worst of the two (21) and (20).

This average can be approximated by the following formula that combines (20) and (21):

$$\hat{T}_{fab} = \frac{\hat{R}_a\hat{R}_b-\hat{D}_a\hat{D}_b}{\hat{R}_a+\hat{D}_a+\hat{R}_b+\hat{D}_b} \tag{22}$$

For double-sided two-way ranging it is generally suggested that the delay times $\hat{D}_a$ and $\hat{D}_b$ have to be nearly equal for the overall error to be acceptably small, but this is not a restriction when employing equations (20), (21) and (22), that burden is removed, so it is NOT required the use of the same response time at each end. This gives much more implementation flexibility within the individual nodes participating in a ranging exchange and also facilitates more efficient implementations when a group of N nodes want to find the ½·N·(N-1) distances between each other. Here, instead of separated ranging exchanges, messages can be combined, (e.g. a response from node B to Node A could also act as a ranging initiation message other nodes), to do the distance measurements with a much reduced over the air traffic, perhaps as few as 2·N messages.

1.3 Asymmetric double-sided two-way ranging error due to clock offset

Formulas (19a) and (19b) give the time of flight results (TOF) results including the clock offset errors. For the average formula then we can similarly say

$$\hat{T}_{fab} = T_f \left(\frac{k_a + k_b}{2}\right)$$

Thus, the error in the estimated $\hat{T}_{fab}$ is:

$$Error = T_f - \hat{T}_{fab} = T_f - T_f \left(\frac{k_a + k_b}{2}\right)$$

Or

$$Error = T_f \times \ \left(1 - \frac{k_a + k_b}{2}\right)$$

To size this error, if devices A and B have clocks, where each are 20 ppm away from the nominal clock in directions make their combined error additive and equal to 40 ppm, then k_a and k_b might both be 0.99998 or 1.00002.

Even with a fairly large UWB operating range of say 100 m, the TOF is just 333 ns, so the error is:
$20\times10^{-6} \times 333\times10^{-9}$ seconds, which 6.7×10^{-12} seconds or 6.7 picoseconds. Thus the precision of determining the arrival time of the RMARKER is the more significant source of error.

Again note that these error levels do NOT require the use of the same response time at each end.