The questions are:

1. How long to measure the signal?

2. How many measurements per event are optimal?

3. How congregated about the zero-crossings the measurements should be?

The analytical expression of the variance of the estimate is[1]

$$
\begin{cases}
\mathrm{var}\,[\hat{\omega}] & = {}^{\nu}\!/\!\sum_j x_j \mathrm{var}_w\,[t], \\
\mathrm{var}_w\,[t] & = \sum_i w_i\,(t_i - \langle t\rangle_w)^2\,, \quad \langle t\rangle_w = \sum_i t_i w_i, \\
w_i & = \frac{x_i}{\sum_j x_j},\ \ x_i = (N_0 P \exp(\lambda t_i))^2 \cos^2(\omega t_i + \phi) = \left(\mu'_\phi(t_i)\right)^2.
\end{cases}
\tag{1}
$$

As one can see, the variance inversely depends on the spread of the predictor variable. This spread is determined mainly by the total duration of measurement, and hence the spin tune decoherence time. Variation in the number of polarimetry measurements per event has only a second-order effect on $\mathrm{var}_w\,[t]$.

The two questions that follow both condition the measurement distribution across time. The number of measurements per event determine, with regard to this statistical aspect, the number of events per second, and hence the number of events that can fit within a compaction region. If that number is large, there will be fewer (but more precise) measurements within the high-information regions, and we will have to increase the compaction factor in order to collect enough events, at which point our use of the beam is sub-optimal.
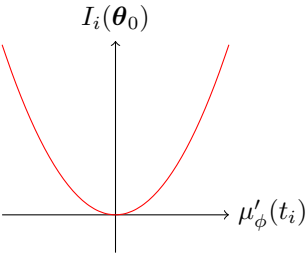
1 Spin tune decoherence time

We have the Cramér-Rao inequality to tell us what's the minimum variance of an estimator is possible:

$$
\mathrm{var}\,[\hat{\omega}] \geq \frac{1}{I(\omega)}.
$$

Fisher information is additive, and what I call *point Fisher information* can be expressed as:

$$
I_i(\boldsymbol{\theta}_0) = \frac{1}{\nu}
\begin{pmatrix}
\left(\sqrt{2}\cdot\mu'_\phi(t_i)\right)^{-2} & 0 & 0 \\
0 & t_i^2 & t_i \\
0 & t_i & 1
\end{pmatrix}
\cdot \left(\mu'_\phi(t_i)\right)^2.
$$

---

[1] This expression consistently overestimates the standard error by a factor of 2; however, when used on a function with $x_i = (N_0 P)^2$ (linear regression), it yields the same expression as the fitter. The fitter also computes the standard error according to some formula. As far as I can tell, the difference between this one and the one the fitter uses (which I haven't seen yet), is the $sum_j x_j$ factor in the denominator.