Technical Appendix

1. l^2 **Norm**

Consider the case where a physical observable f_{obs}, such as the stress of a material, is measured in an experiment. This observable can then be modeled with a governing equation f which takes a set of parameters θ as its arguments. For example, the stress σ_{obs} might be measured and described with an elastic governing equation which takes the Young's Modulus of the material E as its argument, such as $\sigma(E)$. As the goal of the observation of f_{obs} is to obtain the θ which best explains the data (such as the best fitting Young's Modulus), we seek to minimize the error between the model prediction f and the data f_{obs}. Such a problem can be solved as an optimization problem, where a distance metric is used to define the distance, or error, between the model prediction and the data. The most common distance metrics are the l^1 norm (Manhattan distance) and the l^2 norm (Euclidian distance, also referred to as the squared error). We use the l^2 norm as it is the most commonly used in these types of optimization problems in the AFM community. Also, the l^2 norm provides more general solutions, thus avoiding overfitting – although it can suffer performance issues when working with data with significant outliers[1]. As the l^2 norm sums over all the observations (typically instances of time) in the data, it can be thought of as a function of the parameter set θ.

$$l^2(\theta) = \sum_n [f(n, \theta) - f_{obs}(n)]^2 \tag{S1}$$

2. **Log Likelihood**

In realistic conditions, the observed data f_{obs} will have some experimental noise. Here, we treat the noise as a Gaussian process with a known standard deviation s. Then, for a single sample n_0 of the data, we can imagine the probability of correctly modeling f_{obs} with the parameter set θ which is given as $P(f_{obs}(n_0) | \theta)$. As the noise is generated by a Gaussian distribution, this probability will follow the same form. Then, as we previously sought to minimize the l^2 norm with respect to θ, we can now seek to maximize this probability, at least for the single sample n_{01} .

$$P(f_{obs}(n_0) | \theta) = \frac{1}{\sqrt{2\pi s^2}} e^{-\frac{(f(n_0, \theta) - f_{obs}(n_0))^2}{2s^2}}$$ (S2)

Then, the probability of correctly modeling the entire dataset $P(f_{obs} | \theta)$ will be given as the joint probability distributions for every sample n. The resulting probability distribution can be significantly simplified if we assume that the distributions for each instant in time are independent and identically distributed (i.i.d.) which results in a product of the individual distributions[2–5].

$$P(f_{obs} | \theta) = P(f_{obs}(1) \cap f_{obs}(2) \cap ... f_{obs}(N) | \theta) = \prod_{n}^{N} P(f_{obs}(n) | \theta)$$ (S3)

Again, this probability distribution governs the probability, or likelihood, that the model has correctly described the data. If the model perfectly describes the data, then this distribution will be maximized. In this case, it is highly *likely* that the model describes the data. Hence, this is commonly referred to as the likelihood. A far more commonly used quantity is the log likelihood $L(f_{obs} | \theta)$, which is given by the logarithm of the likelihood distribution. The product in the likelihood turns into a sum in the log likelihood, thus greatly simplifying the math when working with the quantity.

$$L(f_{obs} | \theta) = \sum_{n} -\frac{\log(2\pi s^2)}{2} - \frac{(f(n, \theta) - f_{obs}(n))^2}{2s^2}$$ (S4)

Here, the connection between the l^2 norm and the log likelihood become obvious as seen in eqn. S5. We can now clearly see that minimizing the l^2 norm is equivalent to maximizing the log likelihood. As the logarithm is a monotonically increasing function, maximizing the log likelihood is then also equivalent to maximizing the likelihood. Thus, the θ that either minimizes the l^2 norm or maximizes the likelihood / log likelihood is commonly referred to as the maximum likelihood estimate (MLE) θ_{MLE}.

$$L(f_{obs}|\theta) = -\frac{l^2(\theta)}{2s^2} - \sum_n \frac{\log(2\pi s^2)}{2} \tag{S5}$$

3. Fisher Information, Observed Fisher Information, D-Optimality Criterion

Then, as we seek to perform experiments so that we obtain an accurate estimate of θ_{MLE} for the physical observable f_{obs}, we might imagine that some data might be more useful than others. For example, if the data is heavily polluted with noise, determining θ_{MLE} may be impossible. Thus, to measure the quality of a dataset for determining θ_{MLE}, we turn to information theory. Here, we will derive the Fisher Information and Observed Fisher Information using the results of the previous section.

To begin, we recall the hypothetical model f for the data f_{obs}. First, consider the case where the model f is entirely incapable of describing the features of f_{obs}. Here, one might have selected an entirely inappropriate model for the analysis! For instance, one could be attempting to model the data of the stress of an elastic material (with Young's Modulus E) under a constant strain input $\varepsilon(t \geq 0) = \varepsilon_0$ by using a viscous fluid model (with viscosity μ as the model parameter). Obviously in this ridiculous hypothetical case, the model and the data would never match, thus the log likelihood would be invariant under changes in the viscosity as seen in eqn.'s S6-S8. Furthermore, one could determine that the data does not contain any information about the viscosity of the material.

$$L(\sigma_{obs}|\mu) = \sum_n -\frac{\log{(2\pi s^2)}}{2} - \frac{(\sigma(n,\mu) - \sigma_{obs}(n))^2}{2s^2} \tag{S6}$$

$$L(\sigma_{obs}|\mu) = \sum_n -\frac{\log{(2\pi s^2)}}{2} - \frac{(\mu\dot{\varepsilon}(n) - E\varepsilon(n))^2}{2s^2} \tag{S7}$$

$$L(\sigma_{obs}|\mu) = \sum_n -\frac{\log{(2\pi s^2)}}{2} - \frac{\varepsilon_0^2 E^2}{2s^2} \tag{S8}$$

As $L(f_{obs}|\theta)$ compares the model to the data, we see that if $\nabla_\theta L(f_{obs}|\theta) = 0$ (as in the previous example) then not only is the model incapable of describing the data, but also the data does not provide any information for determining θ_{MLE}. Conversely, if a change in θ results in a significant change in $L(f_{obs}|\theta)$, then the data provides a similarly significant amount of information about θ_{MLE}. Thus, we see that the sensitivity of the log likelihood to changes in θ serves as a way to measure the amount of information that the dataset contains about θ. We can mathematically define this sensitivity as the magnitude of the derivative of the log likelihood with respect to θ. This quantity is referred to as the Fisher Information $I(\theta)$ as seen in eqn. S9[2–6]. Here, the square is used to take the magnitude of the gradient. We could have just as well taken the absolute value; however, this would have resulted in a non-differentiable function. Finally, the expected value $E[\,]$ is taken as the log likelihood is a probability distribution.

$$I(\theta) = E\left[\left(\nabla_\theta L(f_{obs}|\theta)\right)^2\right] \tag{S9}$$

Though straightforward, this definition is not practical as it explicitly depends on knowledge of the true parameter set, which is unknown. Hence, calculating the expected value is impossible. Instead, the Observed Fisher Information $J(\theta_{MLE})$ can be used as an estimate for the true Fisher Information[2–4].

$$J(\theta_{MLE})_{i,j} = \left[-\partial_{\theta_i}\partial_{\theta_j} L(f_{obs}|\theta)\right]_{\theta = \theta_{MLE}} \tag{S10}$$

Here, the Observed Fisher Information calculates the curvature of the log-likelihood with respect to the parameter set, evaluated at the maximum likelihood estimate θ_{MLE}. If the log likelihood is sharply defined

at the maximum likelihood estimate, then changes in the θ will result in significant changes in $L(f_{obs}|\theta)$, and hence, the Observed Fisher Information is large. In general, $J(\theta_{MLE})_{i,j}$ defines the matrix elements for a matrix of size $m \times m$ where m is the dimension of θ. Various elements of $J(\theta_{MLE})_{i,j}$ can be used to quantify information, here we consider the D-optimality criterion, defined as the determinant of $J(\theta_{MLE})$, for consistency with the study of the characteristics of step-like experiments[2,4]. While several technical details have been left out in obtaining $J(\theta_{MLE})_{i,j}$ from $I(\theta)$, further details can be found elsewhere[4–8].

4. **Observed Fisher Information for Generalized Maxwell Model**

The elements of $J(\theta_{MLE})_{i,j}$ for a generalized Maxwell model with N relaxation times can be calculated following eqn. S10 as seen below.

$$J\left(G_{e_{MLE}}, G_{1_{MLE}}, \ldots \tau_{N_{MLE}}\right) = \left[-\begin{bmatrix} \partial^2_{G_e} & \partial_{G_e}\partial_{G_1} & \cdots & \partial_{G_e}\partial_{\tau_N} \\ \partial_{G_1}\partial_{G_e} & \partial^2_{G_1} & & \vdots \\ \vdots & & \ddots & \vdots \\ \partial_{\tau_N}\partial_{G_e} & \cdots & \cdots & \partial^2_{\tau_N} \end{bmatrix} L\left(\sigma_{obs}|G_e, G_1, \ldots \tau_N\right) \right]_{G_{e_{MLE}}, G_{1_{MLE}}, \ldots \tau_{N_{MLE}}} \quad (S11)$$

5. **Mechanical Models**

The constitutive equation for linear viscoelasticity is given as an arbitrary order differential equation which relates the stress and the strain of the material. Such an equation is generally solved by the Boltzmann integral which relates the stress σ to a convolution of the viscoelastic modulus Q and the strain ε[9–13].

In these types of equations, the modulus is referred to as the kernel of the integral and represents the impulse response of the material such that ($Q(t) = Q(t) * \delta(t)$).

$$\sigma(t) = \int du \, Q(t-u)\varepsilon(u) \quad (S12)$$

Then, using the models derived by Lee and Radok, this equation can be equivalently obtained in terms of force f and indentation h between the viscoelastic half-space and a rigid indenter of various geometries as seen in eqn. S13[14]. The parameters α and β depend on the geometry of the indenter and can be found in other sources[15–17].

$$\frac{f(t)}{\alpha} = \int du\, Q(t-u)h^{\beta}(u) \tag{S13}$$

6. Modified Fourier Transform

Recently, the Z-transform (ZT) was used to directly obtain Q from force-indentation data[18,19]. The ZT can be thought of as a discrete analog to the Laplace transform. Recall that the Fourier transform of a function can be directly obtained from its Laplace transform by setting the complex Laplace domain variable $s = i\omega$. Here, the discrete Fourier transform (DFT) can similarly be obtained from the ZT, by setting $z = i\omega$[18,19]. In both of these cases, the real part of the complex variable is removed. If the real part were kept, the result would give the modified Fourier transform (MFT) (in the case of $s = a + i\omega$) and the modified discrete Fourier transform (MDFT) (in the case of $z = \ln\left(r_0\right) + i\omega$). As both the Z and Laplace transforms deal with the entirety of the complex plane, the modified transforms restrict this to a single line and circular contour, in the continuous and discrete cases, respectively. Thus, it is far more practical to use the MDFT to analyze the discrete force-indentation data.

$$T_{DFT}\{f[n]\}(\omega) = \sum_n f[n]e^{-i\omega n} \tag{S14}$$

$$T_{ZT}\{f[n]\}(z) = T_{ZT}\{f[n]\}(\omega, r) = \sum_n f[n]z^{-n} = \sum_n f[n]r^{-n}e^{-i\omega n} \tag{S15}$$

$$T_{MDFT}\{f[n]\}(\omega) = \sum_n f[n]r_0^{-n}e^{-i\omega n} = T_{DFT}\{f[n]r_0^{-n}\}(\omega) \tag{S16}$$

By applying the MDFT to eqn. S13, the convolution of Q and h^{β} turns into a multiplication of their transforms.

$$T_{MDFT}\left\{\frac{f[n]}{\alpha}\right\}(\omega) = T_{MDFT}\{Q[n]\}(\omega)\,T_{MDFT}\{h^{\beta}[n]\}(\omega) = Q(\omega)\,T_{MDFT}\{h^{\beta}[n]\}(\omega) \tag{S17}$$

This multiplication can then be directly manipulated to obtain Q from the force-indentation data as seen in eqn. S18[19]. Of course, such a technique is not restricted to force-indentation data. The general definition of the Boltzmann integral can also be used, thus obtaining Q from stress-strain data.

$$Q(\omega) = \frac{T_{MDFT}\left\{\dfrac{f[n]}{\alpha}\right\}(\omega)}{T_{MDFT}\{h^{\beta}[n]\}(\omega)} \tag{S18}$$

$$Q(\omega) = \frac{T_{MDFT}\{\sigma[n]\}(\omega)}{T_{MDFT}\{\varepsilon[n]\}(\omega)} \tag{S19}$$

While using the DFT may seem to be more convenient, doing so would result in an inaccurate characterization as the force-indentation data is numerically unbounded and does not represent a periodic signal[18,19]. The MDFT can be calculated from the DFT as seen in eqn. S16, though it should be noted that careful selection of r_0 is required to avoid errors in this inversion process[19]. It is typical for researchers to seek the storage and loss moduli of a material as the end goal of its viscoelastic characterization. These quantities are defined as the real and imaginary parts of the Fourier spectrum of Q. As the MDFT operates in the modified Fourier domain, the correspondence between the real and imaginary parts of $Q(\omega)$ obtained through this method do not exactly correspond to the real storage and loss moduli. Though they often are quite close, inaccuracies compound for excessively small and large values of r_0[18]   .

7. Dimensionless Model Derivation: Time Domain

The generalized Maxwell model is one of the simplest linear viscoelastic models that is capable of representing the canonical rheological behaviors of creep and stress relaxation. For these reasons, this model is a popular choice among those seeking to characterize viscoelastic material using force-indentation or stress-strain experiments.
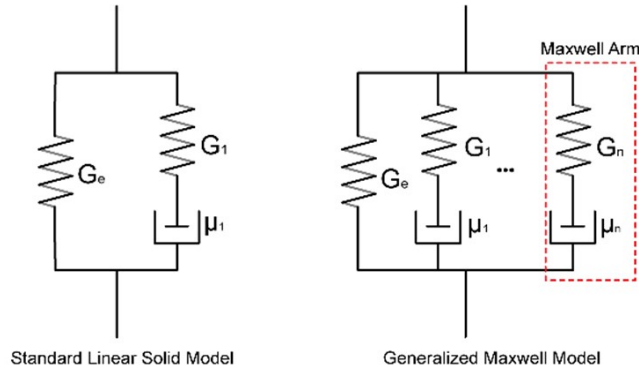
Figure 1: The Standard Linear Solid (SLS) model (seen on the left) is obtained as a special case of the generalized Maxwell model (seen on the right), where the number of relaxation times is set to 1.

Seen in Fig. 1, the generalized Maxwell model is comprised of a series of parallel 'Maxwell arms' made of an elastic spring with a modulus G in series with a viscous dashpot with a viscosity μ. The rate at which the stress in the nth Maxwell arm is governed by the ratio of μ_n/G_n and is referred to as the relaxation time τ_n. The impulse response of such a material can be found from solving the stress-strain equation for the linear time invariant (LTI) system and is given as eqn. S20[9–11,13].

$$Q(t) = G_e\delta(t) + \sum_{n}^{N}\left[G_n\delta(t) - \frac{G_n}{\tau_n}e^{-\frac{t}{\tau_n}}\right] \qquad (S20)$$

The modified Fourier domain correspondence of Q for this model can be obtained by solving the same equation using discrete signals and difference equations which results in eqn. 21[18,19]. It is an interesting note that in the limit of small ω, eqn. S21 approaches the equivalent continuous Fourier domain correspondence.

$$Q(\omega) = G_e + \sum_n^N \left[G_n - \frac{G_n}{1 + \frac{\tau_n}{\Delta t}\left(1 - \frac{e^{-i\omega}}{r_0}\right)} \right]$$

(S21)

8. Derivation of Dimensionless Models

Due to the similarity in the structures of eqn.'s S12 and S13, we consider the quantities f/α and h^β to be equivalent to stress and strain respectively. While the indentation h in typical AFM force-indentation experiments typically does not perfectly follow a line, the 'strain', given as h^β, typically is closely linear as seen in Fig. 2.
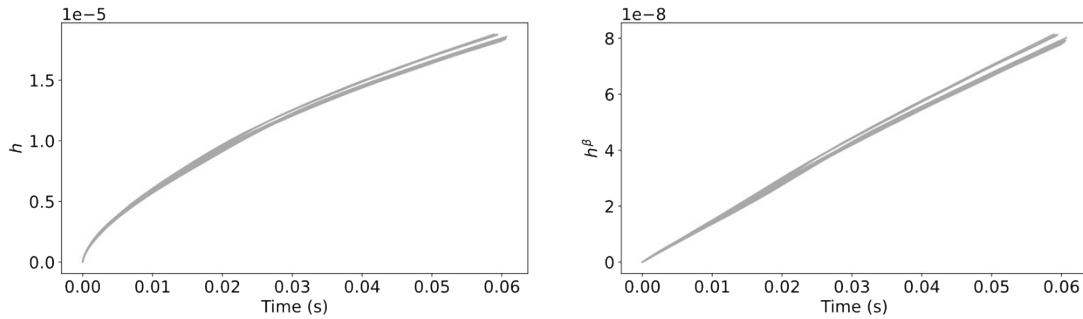
Figure 2: Indentation h and 'strain' h^β as functions of time obtained from an AFM experiment with a spherical probe ($\beta = 3/2$).

Then, assuming the 'strain' of a force-indentation experiment follows a linear function in time, we can write.

$$h^\beta(t) = \varepsilon_0\, t$$

(S22)

Substituting this linearized strain as well as the $Q(t)$ of a generalized Maxwell model, we can obtain the behavior of the stress / force for our idealized force-indentation curves.

$$\sigma(t) = \frac{f(t)}{\alpha} = \int du \left(G_e \delta(t-u) + \sum_n^N \left[G_n \delta(t) - \frac{G_n}{\tau_n} e^{-\frac{t-u}{\tau_n}} \right] \right) \varepsilon_0 u \tag{S23}$$

Which can be solved exactly to yield eqn. S24.

$$\sigma(t) = \frac{f(t)}{\alpha} = \varepsilon_0 \left(G_e t + \sum_n^N G_n \tau_n \left(1 - e^{-\frac{t}{\tau_n}} \right) \right) \tag{S24}$$

To isolate the behavior of the model parameters with respect to the experimental conditions, we can

substitute the following normalizations where   is the length of the experiment in time.

$$\widetilde{G}_n = \frac{G_n}{G_e + \sum_n G_n} \tag{S25}$$

$$\widetilde{\tau}_n = \frac{\tau_n}{} \tag{S26}$$

$$\widetilde{t} = \frac{t}{} \tag{S27}$$

Performing these substitutions and further normalizing by ε_0 yields the following equation which is

essential in obtaining the results in the main paper.

$$\widetilde{\sigma}(\widetilde{t}) = \frac{f(\widetilde{t})}{\alpha \, \varepsilon_0 \, G_g} = \widetilde{t} + \sum_n^N \widetilde{G}_n \widetilde{\tau}_n \left(1 - e^{-\frac{\widetilde{t}}{\widetilde{\tau}_n}} \right) - \widetilde{G}_n \widetilde{t} \tag{S28}$$

A similar approach can be taken to obtain the dimensionless form of $Q(\omega)$.

$$\bar{Q}(\omega) = \frac{Q(\omega)}{G_g} = 1 - \sum_{n}^{N} \frac{\tilde{G}_n}{1 + \frac{\tilde{\tau}_n}{\Delta t}\left(1 - \frac{e^{-i\omega}}{1.001}\right)}$$ (S29)

Note that by returning to eqn. S24, it is evident that the timescales t and τ_n are both scaled by the strain rate ε_0. As we mention in the main paper, this results in an equivalent nondimensional form of eqn. S28 where the dimensionless timescales are given by eqn.'s S26a and S27a. Interpreting these nondimensional parameters allows us to equivalently obtain an upper bound for obtainable relaxation times in terms of the strain rate. Specifically, to keep the dimensionless time constant $\tilde{\tau}_n$ in the information rich regime of the time domain experiments ($\tilde{\tau}_n < 0.1$), one requires that $\tau_n < 1/10\varepsilon_0$ thus allowing the specification of an optimal strain rate for the detection of specific relaxation times.

$$\tilde{\tau}_n = \tau_n \varepsilon_0$$ (S26a)

$$\tilde{t} = t\varepsilon_0$$ (S27a)

9. Dimensionless Model Parameters

As the arm moduli $\tilde{G}_n$ have been normalized by the instantaneous modulus ($G_g = G_e + \sum_{n} G_n$), they can obtain a minimum value of 0 which of course corresponds to a modulus of 0. The sum of all of the $\tilde{G}_n$'s can be, at most, 1 which corresponds to a material with $G_e = 0$, which is often interpreted as a fluid (i.e. the material will continue to relax its stress until it reaches a state of zero stress)[11]. The normalized relaxation times $\tilde{\tau}_n$ range between 0 and ∞, though we are mostly interested in their behavior within the observable timeframe of the experiment. Hence, we typically only consider $\tilde{\tau}_n$'s between 0 and 1. Similar

to the relaxation times, the time axis $\tilde{t}$ has been normalized by the experiment length, thus restricting $\tilde{t}$ to be between 0 and 1 with 1 corresponding to the end of the experiment.

10. Surfaces of the l^2 Norm and Maximum Log Likelihood for the Dimensionless Models

The l^2 norm for the normalized time (eqn. S28) and frequency (eqn. S29) domain models can now be determined. Again, the l^2 norm is a function of the model parameters θ and the successful identification of θ from an optimization algorithm depends on its structure.

$$l_t^2(\theta) = \sum_{\tilde{t}} \left[\tilde{\sigma}(\tilde{t},\theta) - \tilde{\sigma}_{obs}(\tilde{t}) \right]^2 \tag{S30}$$

$$l_\omega^2(\theta) = \sum_{\omega} \left[\tilde{Q}(\omega,\theta) - \tilde{Q}_{obs}(\omega) \right]^2 \tag{S31}$$

We now perform parameter sweeps for these norms over a range of parameters for a model with a single relaxation time. In both cases, $\tilde{G}_1$ and $\tilde{\tau}_1$ are both varied between 0 and 1 and the resulting model predictions ($\tilde{\sigma}(\tilde{t},\tilde{G}_1,\tilde{\tau}_1)$ and $\tilde{Q}(\omega,\tilde{G}_1,\tilde{\tau}_1)$) are compared to a simulated dataset with the parameters $\tilde{G}_{sim},\tilde{\tau}_{sim}$. The resulting surfaces are plotted below.

In Fig. 3, we have $\tilde{G}_{sim} = \tilde{\tau}_{sim} = 0.5$ with these 'true' values marked with an X in the center of these surfaces. The left column of plots shows the surfaces of l_t^2 with an increasingly large amount of noise added to the simulated data. In the right column, the l_ω^2 surfaces are shown, with the same increasing noise. As discussed in the main paper, the time domain contains a valley like formation with values below the machine precision level, thus indicating that optimization will likely fail in these areas. As the standard deviation of the noise increases, so too does the size of the valley. In comparison, the minimum in the frequency domain remains well defined even in the presence of considerable noise.
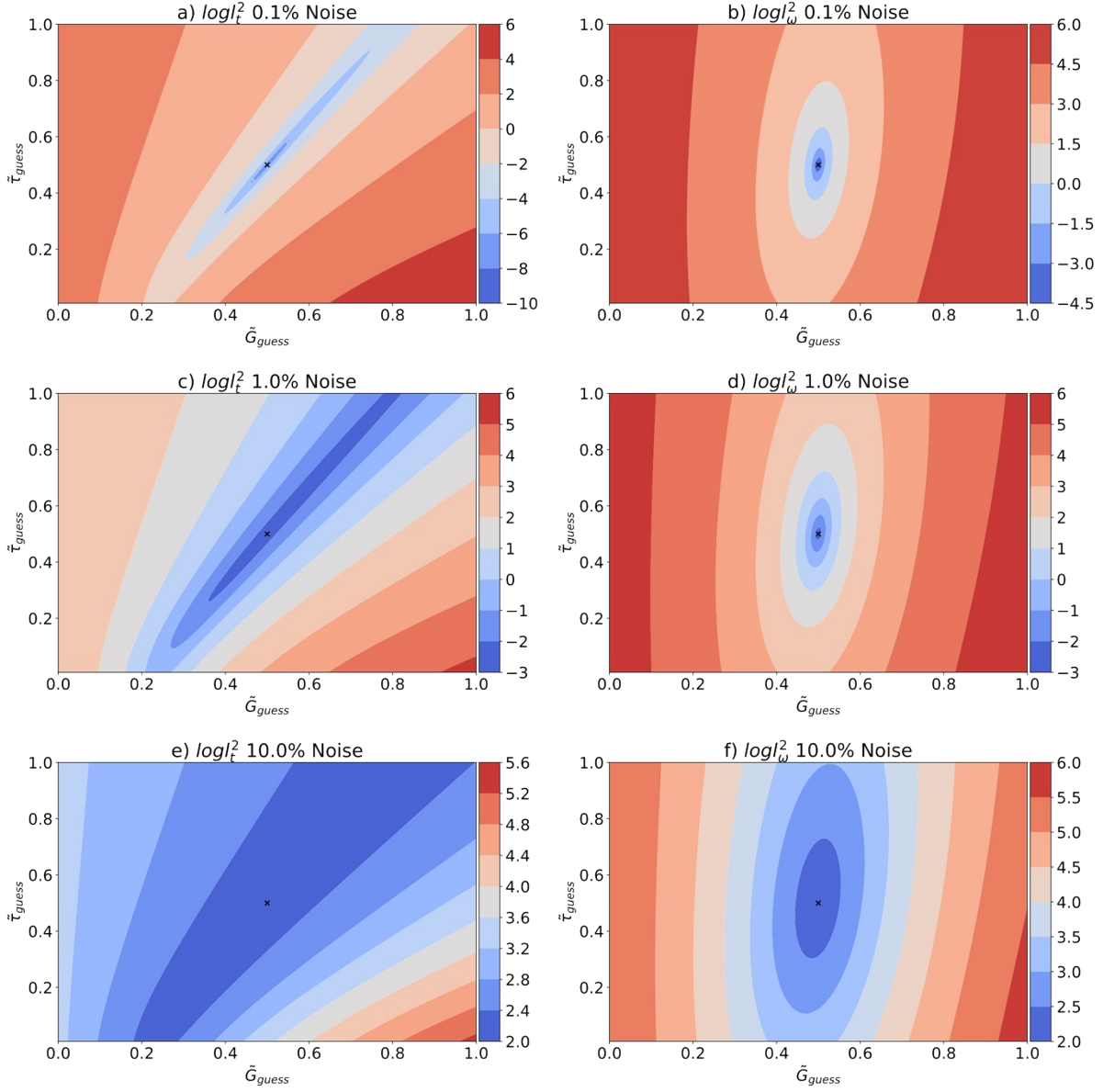
Figure 3: Surfaces of l_t^2 and l_ω^2 for a range of parameter space. Compared to simulated dataset ($\tilde{G}_{sim} = \tilde{\tau}_{sim} = 0.5$) with an increasing amount of noise.

Next, we perform a similar calculation; however, the dataset is now simulated with a relaxation time $\tilde{\tau}_{sim} = 0.1$. As we have shown in the main paper, relaxation times that are less than $/10$ are most readily

obtained from the time domain. As expected, the size of the valley in l_t^2 is significantly reduced when compared to the previous case shown in Fig. 4. Furthermore, the expansion of the valley in l_t^2 is further restricted in this case – though it still indicates there will be considerable problems when optimizing.
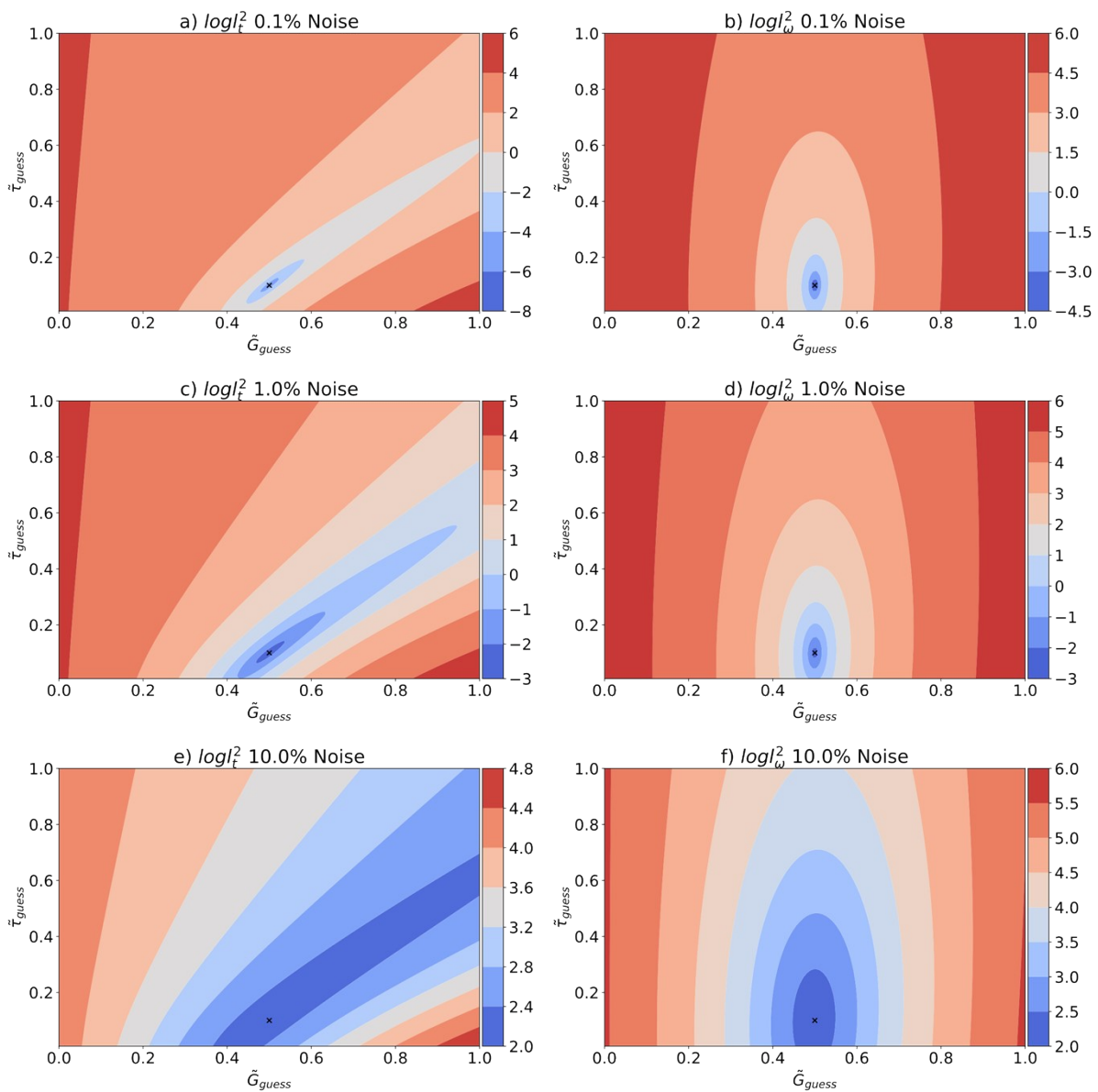
Figure 4: Surfaces of l_t^2 and l_ω^2 for a range of parameter space. Compared to simulated dataset ($\tilde{G}_{sim} = 0.5,\ \tilde{\tau}_{sim} = 0.1$) with an increasing amount of noise.

As done in the main paper, we can further contrast these two approaches by using the log likelihood. Here, we seek to describe a material with two relaxation times by using a model with a single relaxation time (SLS model). In the two figures below, we provide the normalized log likelihood ($L(\tilde{\sigma}_{obs}|\theta_{MLE})$ and $L(\tilde{Q}_{obs}|\theta_{MLE})$) for the maximum likelihood SLS estimate of various parameters ($\tilde{\tau}_1, \tilde{\tau}_2$ and $\tilde{G}_1, \tilde{G}_2$) of the two arm model in both the time (Fig. 5) and frequency (Fig. 6) domains.
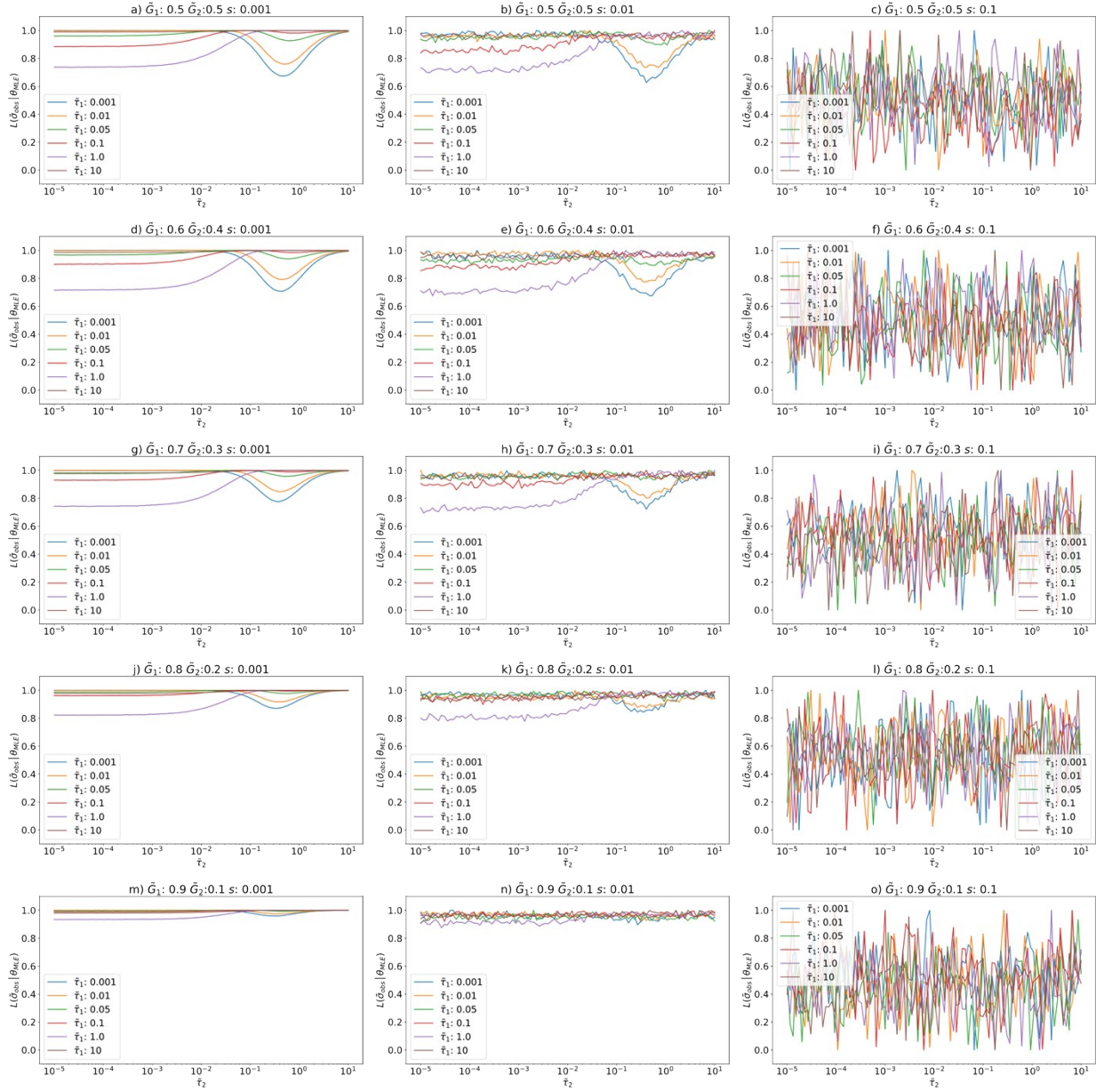
Figure 5: Values of the normalized maximum log likelihood estimate for an SLS model describing a material with two relaxation times $\tilde{\tau}_1, \tilde{\tau}_2$ and two moduli $\tilde{G}_1, \tilde{G}_2$ in the time domain. Moving from left to right, the columns correspond to data with noise of an increasing magnitude s. Each row corresponds to sets of values of $\tilde{\tau}_1, \tilde{\tau}_2$ and $\tilde{G}_1, \tilde{G}_2$ with the difference between $\tilde{G}_1$ and $\tilde{G}_2$ increasing from top to bottom.
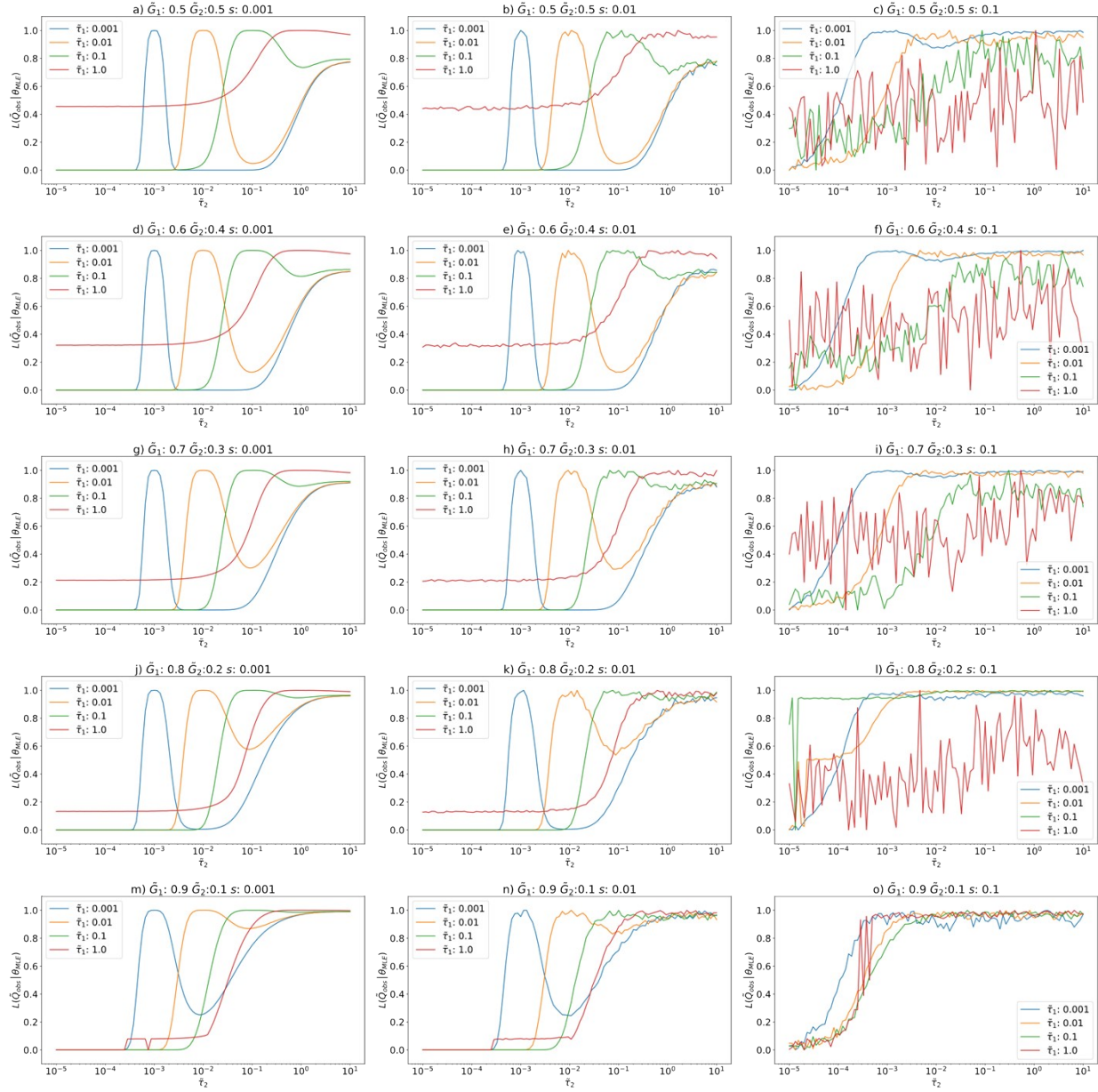
Figure 6: Values of the normalized maximum log likelihood estimate for an SLS model describing a material with two relaxation times $\tilde{\tau}_1, \tilde{\tau}_2$ and two moduli $\tilde{G}_1, \tilde{G}_2$ in the frequency domain. Moving from left to right, the columns correspond to data with noise of an increasing magnitude s. Each row corresponds to sets of values of $\tilde{\tau}_1, \tilde{\tau}_2$ and $\tilde{G}_1, \tilde{G}_2$ with the difference between $\tilde{G}_1$ and $\tilde{G}_2$ increasing from top to bottom.

In both cases, $L(\tilde{\sigma}_{obs}|\theta_{MLE})$ and $L(\tilde{Q}_{obs}|\theta_{MLE})$ form peaks where the values of $\tilde{\tau}_1$ and $\tilde{\tau}_2$ overlap. As discussed in the main paper, this is because as the two relaxation times come closer together, their behavior can be better approximated as a single relaxation time. Hence, for values of $\tilde{\tau}_1$ and $\tilde{\tau}_2$ that are close together, $L(\tilde{\sigma}_{obs}|\theta_{MLE})$ and $L(\tilde{Q}_{obs}|\theta_{MLE})$ will be maximized. With this in mind, we can determine that the superior method of characterization will be the one which offers the greatest distinction between relaxation times. We can see that as the width of the peaks in $L(\tilde{Q}_{obs}|\theta_{MLE})$ are narrower than $L(\tilde{Q}_{obs}|\theta_{MLE})$, the frequency domain offers a greater sensitivity for resolving multiple relaxation times from force-indentation experiments. A similar conclusion can be drawn from the decreased amplitude in $L(\tilde{\sigma}_{obs}|\theta_{MLE})$ as compared to $L(\tilde{Q}_{obs}|\theta_{MLE})$.

When the modulus associated with a relaxation time is decreased, the effect of that relaxation time is significantly reduced. In comparing Fig. 5a to 5m, for example, this effect is made obvious: $L(\tilde{\sigma}_{obs}|\theta_{MLE})$ is effectively flattened for the case when $\tilde{G}_1$ is about an order of magnitude greater than $\tilde{G}_2$. Thus, determining $\tilde{\tau}_2 = 0.1$s when $\tilde{\tau}_1 = 0.001$s for $\tilde{G}_1 = 0.9$ and $\tilde{G}_2 = 0.1$ is as difficult as determining $\tilde{\tau}_2 = 0.01$s when $\tilde{\tau}_1 = 0.001$s for $\tilde{G}_1 = \tilde{G}_2 = 0.5$. However, these issues are significantly reduced in the frequency domain as seen in Fig. 6a and 6m, for example.

11. Discrete Convolution Error

It is uncommon for researchers to assume that the indentation in an AFM experiment follows a prescribed functional form as we have done in the derivation of the linearized model. In practice, the convolution definition given in eqn. S13 is used more frequently as it accommodates force and indentation data of almost any arbitrary form, without the need for assuming one as we have done here. However, as the experimental data is inherently discretized, one must use a *discrete* convolution operation rather than a *continuous* one. In this case, such an approximation is only valid when the convolved functions are band

limited (i.e. their frequency components lie within a finite region of the Fourier domain)[20]. Of course, one may be able to perform experiments which use band limited strain/indentation signals such as sines or cosines; however, most signal shapes used in force-indentation experiments are not exactly so. Though, by admission, the family of ramp-like signals that are more commonly used in these experiments approximately satisfy this criterion due to their rapid attenuation in the frequency domain. However, this issue still remains for the viscoelastic modulus. In particular, band limited moduli may not correspond to physical behaviors of materials.

To address this question as generally as possible, we should first consider the general shape of the viscoelastic modulus in the frequency domain. The behavior of such functions near the origin either starts from zero, corresponding to a fluid, or a finite nonzero value, corresponding to a solid. Following this, the function tends towards some nonzero value at high frequency. This general behavior can be represented as a superposition of a positive constant (DC shift), representing the immediate response of the material, with some function of frequency, representing the transient response of the material. Indeed, such decompositions are commonplace in the general theory of linear viscoelasticity[12,13]. In this case, the frequency components associated with the positive constant result in a Dirac delta function in the time domain, which is analytically equivalent to a scaling of the strain\indentation input with the positive constant. The remaining transient portion of the modulus must then be convolved with the input which is where the issue of band limiting arises.

To further assess these issues, we must assume that the transient behavior follows that of a generalized Maxwell model. For this case, the constitutive equation is again decomposed into immediate and transient components as discussed above which results in eqn. S33. A brief inspection of the spectrum of e^{-t/τ_n} in Fig. 5 reveals that it has a decaying behavior as the magnitude of the frequency increases. Thus, we may posit that approximating e^{-t/τ_n} as being band limited may not introduce significant errors in the

calculation of the convolution. To fully determine if this is true, we first can calculate the spectrum of

e^{-t/τ_n} given as q as seen below.

$$\frac{f(t)}{\alpha} = \int du \left(G_e \delta(t-u) + \sum_n^N \left[G_n \delta(t) - \frac{G_n}{\tau_n} e^{-\frac{t-u}{\tau_n}} \right] \right) h^\beta(u) \tag{S32}$$

$$\frac{f(t)}{\alpha} = \left(G_e + \sum_n^N G_n \right) h^\beta(t) - \sum_n^N \frac{G_n}{\tau_n} \int du \, e^{-\frac{t-u}{\tau_n}} h^\beta(u) \tag{S33}$$

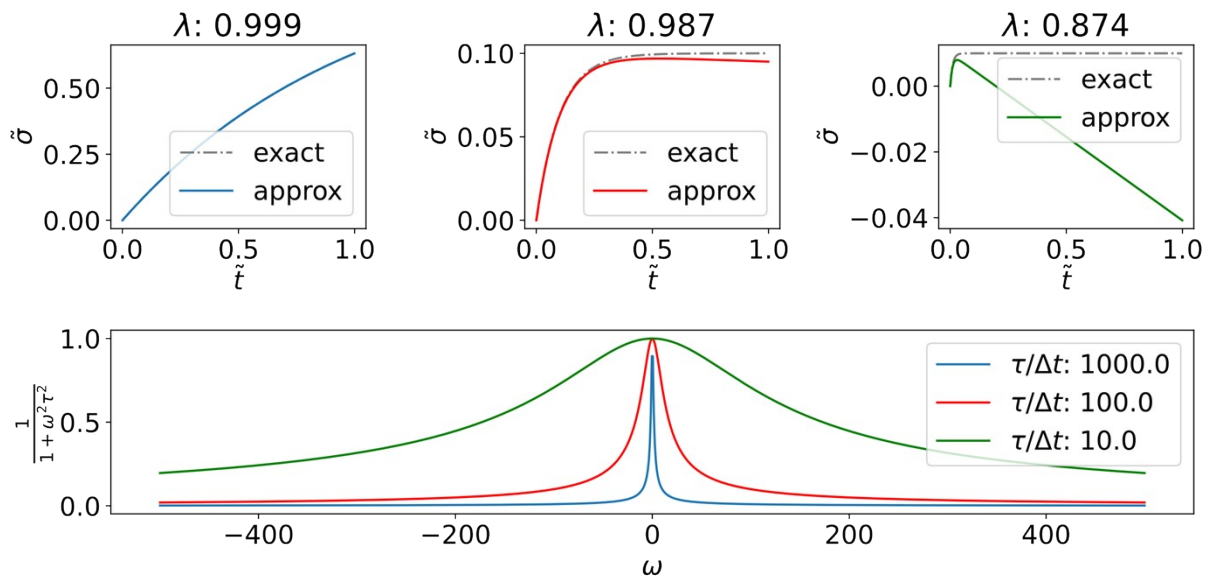$$q(\omega) = \frac{1}{1 + i\omega\tau} \tag{S34}$$

Figure 7: Three force-indentation curves calculated exactly (shown in grey) and approximately using the

discrete convolution (shown in color). The percent of the total energy of the spectrum contained in the

Nyquist band limits is included to demonstrate how much energy is needed to be captured in order for

this approximation to be valid. The magnitudes of the spectra for these three force-indentation curves

are also included, demonstrating that larger relaxation times relative to the sampling timestep cause the spectra to be better approximated as band limited.

As band limited spectra have their entire energy contained within a finite frequency range, we can assess the ability of q to be approximated as a band limited spectrum by looking at the amount of its energy that is contained in the Nyquist frequency bands ($-1/2\Delta t$ to $1/2\Delta t$). For q to be determined to be approximately band limited, it should then have most of its energy contained in this range. To begin, we first determine the energy of q.

$$|q(\omega)| = \frac{1}{1 + \omega^2\tau^2} \tag{S35}$$

Then, assuming that the spacing between the frequencies are sufficiently small (i.e., the experiment is sufficiently long), the total energy contained between the frequencies $-\omega_n$ and ω_n can be obtained as an integral of $|q(\omega)|$.

$$E(\omega_n) = \int_{-\omega_n}^{\omega_n} d\omega \frac{1}{1 + \omega^2\tau^2} \tag{S36}$$

As previously mentioned, a band limited signal should have 100% of its energy contained in a given frequency range, say $-\omega_n$ and ω_n. Thus, the ratio λ of the energy in this frequency range to the energy in all the frequencies must be 1, or as close to it as possible for the approximate case.

$$\lambda = \frac{E(\omega_n)}{E(\infty)} = \frac{E\left(\frac{1}{2\Delta t}\right)}{E(\infty)} \tag{S37}$$

These energies can then be calculated from the integral in eqn. S36.

$$\lambda = \frac{2\text{atan}\left(\dfrac{\tau}{2\Delta t}\right)}{\pi} \tag{S38}$$

Finally, to maximize λ, we obtain an inequality between τ and Δt which gives the final eqn. S39.

$$\tau > 2\Delta t \tan\left(\frac{\lambda\pi}{2}\right) \tag{S39}$$

As demonstrated in Fig. 5, the convolution works best for values of $\lambda > 0.999$ which corresponds to roughly $\tau > 1000\,\Delta t$. Therefore, when using the convolution definition of the generalized Maxwell model, one can only use relaxation times that are thousands of times larger than the sampling timestep. Thus, an even more restrictive lower limit can be placed on the relaxation times that can be determined in the time domain if the convolution definition is used.

12. Noise Carried Through Spectral Inversion with MDFT

As mentioned in Section 6, the viscoelastic modulus of a material can be directly obtained from force-indentation indentation by applying the modified discrete Fourier transform (MDFT) to the data. Once transformed, the force and indentation spectra can be divided to obtain the modulus; however, this division can magnify the effects of even the slightest noise in the data. For example, Fig. 6 shows the spectra of the force and indentation from an AFM experiment. Although the individual spectra appear to be free from noise, the resulting viscoelastic modulus is severely distorted at high frequencies.
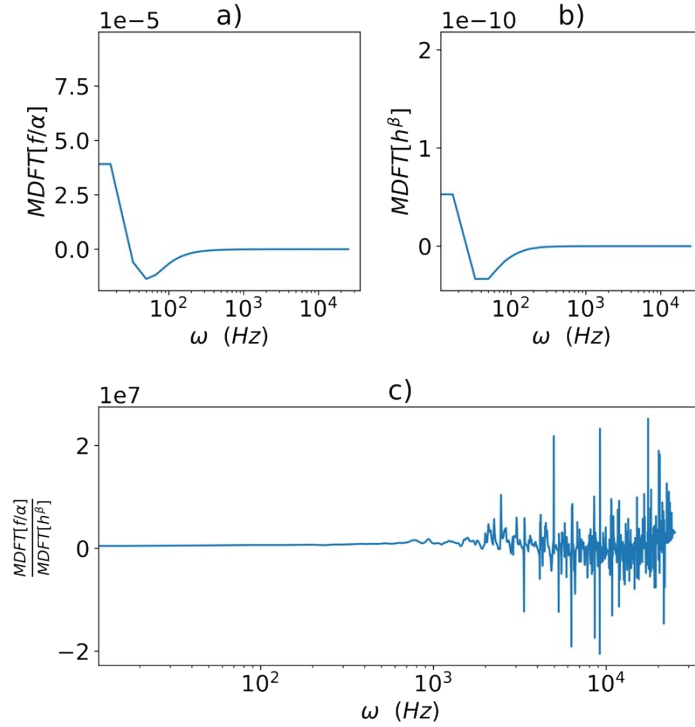
Figure 8: Spectra of a) force and b) indentation data from an AFM experiment obtained from the MDFT as described in Section 6. Though the individual spectra seem noticeably free from noise, dividing them to obtain the c) viscoelastic modulus magnifies the high frequency noise.

To understand how this noise manifests in the frequency domain, we use the more generalized stress-strain relationship from eqn. S12. Here, consider measuring the stress and strain of a material in an experiment where the data is polluted with noise x which is generated by a mean zero Gaussian process with a standard deviation s. The resulting stress-strain equation will be given as eqn. S40.

$$\sigma(t) + x(t) = \int du\, Q(t-u)\, (\varepsilon(u) + x(u))$$

(S40)

As the data has been polluted with noise, the value of Q obtained from inverting eqn. S40 will be incorrect, hence we denote it as Q'. The result can be seen in eqn. S41 where the notation f_ω is used to denote the MDFT of some signal f.

$$\sigma_\omega + x_\omega = Q_\omega(\varepsilon_\omega + x_\omega)$$

(S41)

$$Q'_\omega = \frac{\sigma_\omega + x_\omega}{\varepsilon_\omega + x_\omega}$$

(S42)

However, the resulting equation is invalid as the spectrum of a random process such as x is undefined.

Instead, we will use the power spectrum of $Q'_\omega 21$.

$$|Q'_\omega|^2 = \frac{\sigma_\omega + x_\omega}{\varepsilon_\omega + x_\omega}\left(\frac{\sigma_\omega^* + x_\omega^*}{\varepsilon_\omega^* + x_\omega^*}\right) = \frac{|\sigma_\omega|^2 + |x_\omega|^2 + \sigma_\omega^* x_\omega + \sigma_\omega x_\omega^*}{|\varepsilon_\omega|^2 + |x_\omega|^2 + \varepsilon_\omega^* x_\omega + \varepsilon_\omega x_\omega^*}$$

(S43)

As the power spectrum of a random variable is its autocorrelation in the time domain, we can then substitute the variance s^2 of x for $|x_\omega|^2$.

$$|Q'_\omega|^2 = \frac{|\sigma_\omega|^2 + s^2 + \sigma_\omega^* x_\omega + \sigma_\omega x_\omega^*}{|\varepsilon_\omega|^2 + s^2 + \varepsilon_\omega^* x_\omega + \varepsilon_\omega x_\omega^*}$$

(S44)

To completely remove the dependence of Q'_ω on the random variables, we take the expected value. To do so, we first need the following identity of the expected value of the spectra of x_ω and x_ω^*.

$$E(x_\omega) = E\left[\sum_n x[n]r^{-n}e^{-i\omega n}\right] = \sum_n r^{-n}e^{-i\omega n}E[x[n]] = 0$$

(S45)

$$E(x_\omega^*) = E\left[\sum_n x[n]r^{-n}e^{i\omega n}\right] = \sum_n r^{-n}e^{i\omega n}E[x[n]] = 0$$

(S46)

Thus, the expected value of the magnitude of the modulus is given below.

$$E\left(\left|Q_\omega'\right|^2\right) = \frac{\left|\sigma_\omega\right|^2 + s^2}{\left|\varepsilon_\omega\right|^2 + s^2}$$

(S47)

To understand how this erroneous modulus behaves with respect to the amplitude of the noise in the data, we can take a Maclaurin expansion of eqn. S47 to obtain a simple scaling behavior which converges for all values of $s^2 < \left|\varepsilon_\omega\right|^2$ (i.e. reasonable amounts of noise).

$$E\left(\left|Q_\omega'\right|^2\right) = \frac{\left|\sigma_\omega\right|^2}{\left|\varepsilon_\omega\right|^2} + \sum_{n=1}^{\infty} (-1)^n \frac{\left|\sigma_\omega\right|^2 - \left|\varepsilon_\omega\right|^2}{\left(\left|\varepsilon_\omega\right|^2\right)^{n+1}} s^{2n}$$

(S48)

Since the division of $\left|\sigma_\omega\right|^2$ and $\left|\varepsilon_\omega\right|^2$ gives the true value of $\left|Q_\omega\right|^2$, we can rewrite eqn. S48 to give the following, which is interpreted as the expected value of the error in obtaining Q_ω due to noise.

$$E\left(\left|Q_\omega'\right|^2 - \left|Q_\omega\right|^2\right) = \delta_Q = \sum_{n=1}^{\infty} (-1)^n \frac{\left|Q_\omega\right|^2 - 1}{\left|\varepsilon_\omega\right|^{2n}} s^{2n}$$

(S49)

Thus, we deduce that the error in the modulus scales according to the rule in eqn. S50.

$$\delta_Q \sim \frac{s^2}{\left|\varepsilon_\omega\right|^2}$$
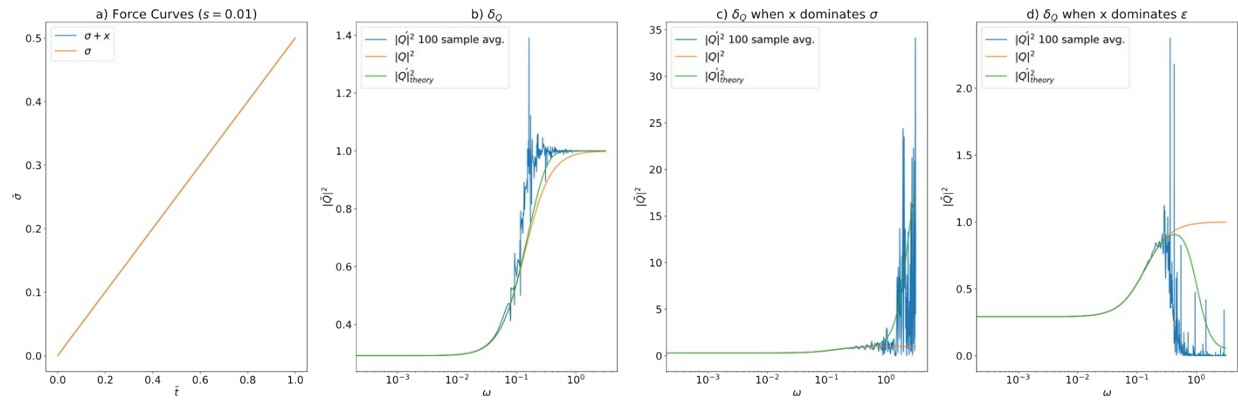
(S49)

Figure 9: a) force curves with and without noise shown in the blue and orange curves, respectively. Even in the case where the amplitude of the noise is small relative to the full scale of the data, the b) error seen when inverting the force curves to obtain the viscoelastic modulus is large. Here, a 100-sample average (shown in blue) was taken to obtain an approximate expected value for this behavior. Two more plots are included which demonstrate the effect of noise when it dominates the c) stress signal and d) the strain signal.

As seen previously in Fig. 6, the spectra of typical AFM force and indentation data strongly decay at high frequencies. Thus, one can expect that since the error has an inverse dependence on the magnitude of the strain, the error will increase with frequency as seen in Fig. 7b. Most noticeable is the fact that even though the amplitude of the noise used in this example is small, the expected inverted spectrum (green) noticeably deviates from the true behavior (orange). We then perform a 100-sample average to gain a practical determination of this expected value as seen in the blue curve. While the effects of the noise in this average are even more pronounced than expected, the spectral averaging considerably reduces the error when compared to individual, non-averaged spectra. Further examples have been provided in Fig. 7c and 7d where the noise is assumed to dominate the stress only (Fig. 7c) and strain only (Fig. 7d). In these cases, the error due to the noise is considerably worse.
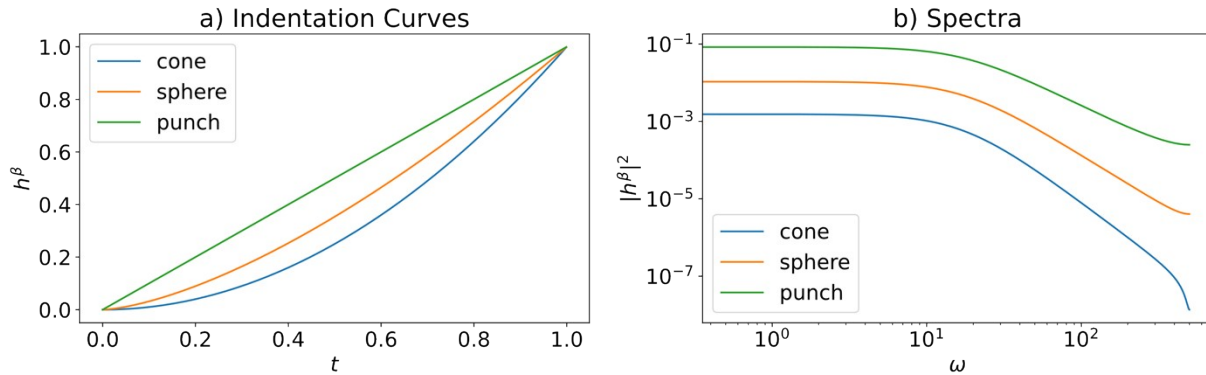
Figure 10: a) 'strain' curves h^β generated for linear indentations for different indenter profiles (flat punch $\beta = 1$, sphere $\beta = 3/2$, and cone $\beta = 2$) and b) the magnitude of their spectra in the modified Fourier domain.

Fig. 10 demonstrates various 'strain' profiles given by h^β for a linear indentation. Here, the different profiles for different indenter geometries are included as well as the magnitudes of their spectra. We see that the conical profile has a spectrum that decays most rapidly whereas the punch decays the slowest. Thus, for the same indentation form, a conical indenter will have the most error in the presence of noise whereas the flat punch will have the least. More generally, less divergent excitations will have a less rapid attenuation of their magnitude in the frequency domain. Therefore, we can expect that experiments with more gradual excitations will have less high frequency noise than those with more rapid, aggressive ones.

References

1 S. L. Brunton and J. N. Kutz, *Data-driven science and engineering: machine learning, dynamical systems, and control*, Cambridge University Press, Cambridge, 2019.
2 I. J. Myung, *Journal of Mathematical Psychology*, 2003, **47**, 90–100.
3 K. Fujita, K. Okada and K. Katahira, *The Fisher information matrix: A tutorial for calculation for decision making models*, PsyArXiv, 2022.
4 K. Vemaganti, S. Madireddy and S. Kedari, *Mech Time-Depend Mater*, 2020, **24**, 1–24.

5 A. C. Atkinson, *Journal of Statistical Planning and Inference*, 2008, **138**, 56–64.

6 J. Mulder and W. J. van der Linden, *Psychometrika*, 2009, **74**, 273–296.

7 J. I. Myung and M. A. Pitt, *Psychological Review*, 2009, **116**, 499–518.

8 J. Andere-Rendon, D. C. Montgomery and D. A. Rollier, *Journal of Quality Technology*, 1997, **29**, 451–463.

9 W. N. Findley, J. S. Lai and K. Onaran, *Creep and relaxation of nonlinear viscoelastic materials: with an introduction to linear viscoelasticity*, Dover, New York, 1989.

10 N. W. Tschoegl, .

11 D. R. Bland, *The theory of linear viscoelasticity*, Dover Publications, Inc, Mineola, New York, 2016.

12 J. M. Golden and G. A. C. Graham, *Boundary Value Problems in Linear Viscoelasticity*, Springer Berlin Heidelberg, Berlin, Heidelberg, 1988.

13 R. M. Christensen, *Theory of viscoelasticity*, Dover Publications, Mineola, N.Y, 2nd ed., 2003.

14 E. H. Lee and J. R. M. Radok, *Journal of Applied Mechanics*, 1960, **27**, 438–444.

15 Y. M. Efremov, T. Okajima and A. Raman, *Soft Matter*, 2020, **16**, 64–81.

16 R. Garcia, *Chem. Soc. Rev.*, 2020, **49**, 5850–5884.

17 M. McCraw, B. Uluutku and S. Solares, *Rep. Mech. Eng.*, 2021, **2**, 156–179.

18 B. Uluutku, E. A. López-Guerra and S. D. Solares, *Beilstein J. Nanotechnol.*, 2021, **12**, 1063–1077.

19 B. Uluutku, M. R. McCraw and S. D. Solares, *Journal of Applied Physics*, 2022, **131**, 165101.

20 N. Chacko, M. Liebling and T. Blu, *J. Opt. Soc. Am. A*, 2013, **30**, 2012.

21 A. V. Oppenheim and R. W. Schafer, *Digital signal processing*, Prentice-Hall, Englewood Cliffs, N.J, 1975.