Supplementary information

Statistical nature of the incipient plasticity in amorphous alloys

Shankha Nag[1], R. L. Narayan[2], Jae-il Jang[3], C. Mukhopadhyay[4], and Upadrasta Ramamurty[5*]

[1]School of Mechanical Engineering, École Polytechnique Fédérale de Lausanne, Lausanne, CH-1015, Switzerland.
[2]Department of Materials Science and Engineering, Indian Institute of Technology, Delhi, 110016, India.
[3]Division of Materials Science and Engineering, Hanyang University, Seoul 04763, South Korea.
[4]Department of Management Studies, Indian Institute of Science, Bangalore-560012, India.
[5]School of Mechanical and Aerospace Engineering, Nanyang Technological University, Singapore 639798.

[*]Corresponding author; e-mail: uram@ntu.edu.sg

Note: Due to the extensive use of mathematical equations in the paper, some of symbols used in the main manuscript may have been repeated or used to define some other variable in the supplementary information (SI). To avoid confusion, the nomenclature provided below serves as a reference only for the symbols used in SI.

Nomenclature

$\mathcal{D}$	Dataset
d	Data point
$f(x)$	An arbitrary probability density function used for fitting $\mathcal{D}$
θ	Parameter space for $f(x)$
μ	Mean of a gaussian distribution
S	Standard deviation of a gaussian distribution
$\widehat{(.)}$	Maximum Likelihood estimate of any parameter, such as θ, μ, σ
$L(.)$	Likelihood function
ψ, λ	Subspaces of θ
$L_p(.)$	Profile likelihood.

n	Sample size
β	Scale parameter
m	Weibull modulus
α	Location parameter of Weibull distribution
$f_M(x)$	Density function of a mixture model
p	Proportion of a component distribution in a mixture model
z_i	Indicator variable corresponding to every element of $\mathcal{D}$
$\mathcal{C}$	Organized and complete data
$E(\cdot)$	Expectation operator
ϕ	Choice of θ
$Q(.)$	Expected likelihood
${c_i}^k$	Alternate notation for $E\langle z_i \middle| \hat{\theta}^k \rangle$
s	Vector representing collection of sample sizes
s_i	i^{th} sample size in s
θ_0	Initial parameter space
θ_{mean}	Mean of $\hat{\theta}$ of s
θ_{var}	Variance of $\hat{\theta}$ of s
K	Absolute maximum distance between the empirical distribution function and the cumulative distribution function of the reference distribution
H_0	Null hypothesis
α_s	Significance level
K_{crit}	Critical value of K
p-value	Threshold value of α_s
V_d	Total deformed volume underneath the indenter
$\bar{V}_d^y$	Average deformed volume of the material at first pop-in
V_e	Fraction of $\bar{V}_d^y$ which has optimum conditions for shear band formation
ρ_d	Density of defects
N_{STZ}	Number of activable STZs
P	Load

h	Displacement
R_i	Indenter radius
$\overline{E_r}$	Reduced modulus
ν	Poisson's ratio
$\overline{E_\iota}$	Elastic modulus of the indenter
$\overline{E_s}$	Elastic modulus of the sample
r', z'	Normalized spacial coordinates for stress contours
P_m	Mean pressure
A	Contact radius of indenter with surface
u'	Normalized displacement
σ'	Normalized Stress
τ'	Normalized Shear stress
τ_{max}	Maximum of maximum shear stress field
σ'_m	Hydrostatic stress
P	Indentation load
P_{FP}	First pop-in load
$\dot{P}$	Loading rate
h	Indentation depth
R_i	Indenter tip radius
τ	Shear stress
τ_y	Shear yield strength
τ_{max}	Maximum shear stress underneath the spherical indenter
$\overline{\tau_y}$	Mean (or average) value of the shear yield strength

S 1. Statistical procedures employed

S1.1 Sample size optimization procedure

A vector, $s = [s_1, s_2, s_3, \dots, s_m]$ is chosen, which signifies a collection of sample sizes. Then, N random samples of datapoints corresponding to each sample size s_i are generated from $f(d|\theta_0)$ where d and θ_0 represent data and initial parameter space, respectively. For every sample, $\hat{\theta}$ is calculated. Then, the mean, θ_{mean}, and variance, θ_{var}, for each sample size s_i, is determined from the following:

$$\theta_{mean} = \frac{1}{N}\sum_{j=1}^{N}\hat{\theta}_j, \qquad \forall s_i \tag{S1}$$

$$\theta_{var} = \frac{1}{N}\sum_{j=1}^{N}\left(\hat{\theta}_j - \theta_0\right)^2, \qquad \forall s_i. \tag{S2}$$

With increasing sample size, θ_{mean} would tend to θ_0 whereas θ_{var} would tend to zero asymptotically and the optimum sample size for the chosen statistical model corresponds to one where $(\theta_{mean} - \theta_0)$ or θ_{var} is relatively small. The sample size optimization test for the single component 3-parameter Weibull model is shown in Fig. S1. The largest spread observed is in the scale parameter, which is 10% of the reference value. We have conducted similar sample optimization tests for all the statistical models considered in this study (for both single component and two component distributions).

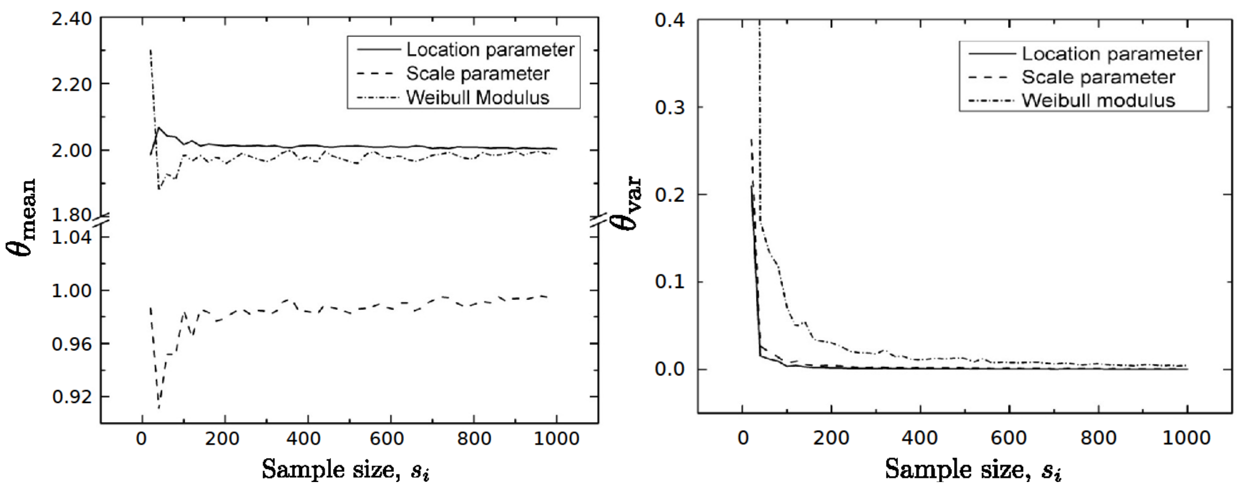

Fig. S1. Variations of (a) θ_{mean} and (b) θ_{var} with the sample size, s_i for the different parameters of 3-parameter Weibull model (θ_0: Location parameter, α = 2, Scale Parameter, β = 1, Weibull Modulus, m = 2). θ_{var} is negligible for all values of s_i > 100 for all the sample optimization tests conducted in this study.

S1.2.1 Statistical inference through maximum likelihood (ML) approach:

Let $\mathcal{D} = \{d\}$ be a given set of data, where d represents individual data points. $\mathcal{D}$ is fitted with an arbitrary probability density function $f(x|\theta)$, where θ is the parameter space. For example, if we choose Gaussian distribution as the fitting model, then

$$f(x|\mu, S) = \frac{1}{S\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{S}\right)^2}, \tag{S3}$$

where μ and S are the mean and standard deviation. Note that in this example, μ and S, represent θ for the Gaussian distribution. For a given $\mathcal{D}$ and $f(x|\theta)$, we are interested in finding the best fitting parameters, $\hat{\theta}$, which are also referred to as Maximum Likelihood Estimates (MLEs). The ML approach defines a likelihood function, $L(\theta)$, for $f(x|\theta)$, as follows,

$$L(\theta|\mathcal{D}) = \prod_i f\,(d_i|\theta). \tag{S4}$$

Note that L is a function of θ for a given $\mathcal{D}$. Our desired fitting parameters or $\hat{\theta}$ are the parameters for which the $L(\theta)$ attains global maximum.

S1.2.2 Concept of Profile Likelihood:

If ψ and λ are subspaces of θ, such that, $\psi \cup \lambda = \theta$, the profile likelihood of ψ, $L_p(\psi)$, is defined from the following. (i) Two functions $L_\psi(\lambda) = L(\lambda|\psi)$ and $L_\lambda(\psi) = L(\psi|\lambda)$, are defined as the likelihood functions, where one parameter is held fixed while the other is varied. (ii) $\hat{\lambda}_\psi$ is defined as the value of λ at which the function $L_\psi(\lambda)$ attains a global maximum, i.e., $\hat{\lambda}_\psi = argmax_\lambda L_\psi(\lambda)$. Then, the profile likelihood $L_p(\psi)$ is $L_{\hat{\lambda}_\psi}(\psi)$. From this, $\hat{\theta}$ is deduced as follows. Let $\hat{\psi}$ be the parameter ψ for which the profile likelihood $L_p(\psi)$ attains a global maximum. Then $\hat{\theta} = \hat{\psi} \cup \hat{\lambda}_{\hat{\psi}}$.

S1.2.3. MLE for single component distribution functions

For Gaussian and Lognormal distributions, the probability density functions (PDFs), $f(x)$, are given as:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \tag{S5}$$

$$f\ (x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(-\frac{(ln\, x-\mu)^2}{2\sigma^2}\right), \quad x > 0 \tag{S6}$$

where μ and σ are mean and standard deviation. In the case of 3 parameter Weibull distribution, $f(x)$ is given as,

$$f(x) = \begin{cases} \frac{m}{\beta}\left(\frac{x-\alpha}{\beta}\right)^{m-1} \exp\left(-\left(\frac{x-\alpha}{\beta}\right)^{m}\right), & x \geq 0 \\ 0 \qquad , & x < 0 \end{cases} \tag{S7}$$

where α and β are is location and scale parameters respectively, and m is the Weibull modulus. For 2 parameter Weibull distribution, $\alpha = 0$. The PDF for the two-component mixture models, f_M, is obtained by weighted linear combination of each component, as following:

$$f_M(x|p, \theta_1, \theta_2) = pf(x|\theta_1) + (1-p)f(x|\theta_2) \tag{S8}$$

where p is the proportion of each component models of the mixture $f(x|\theta_1)$ and $f(x|\theta_2)$; $0 < p < 1$.

Closed form solutions of MLEs exist for Gaussian and Lognormal distributions and are as follows.

MLE for Gaussian distribution:

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} d_i\ , \qquad \hat{S}^2 = \frac{1}{n}\sum_{i=1}^{n}\left(d_i - \hat{S}\right)^2 \tag{S9}$$

MLE for Lognormal distribution:

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} \ln d_i, \qquad \hat{S}^2 = \frac{1}{n}\sum_{i=1}^{n}\left(\ln d_i - \hat{S}\right)^2, \tag{S10}$$

where n is the sample size.

MLE for 2-parameter Weibull distribution:

This model does not have a closed form solution and must be obtained via a numerical approach. m and β are Weibull modulus and scale parameter, respectively. $\widehat{m}$ is estimated iteratively through Regula Falsi from the equation:

$$\frac{1}{\widehat{m}} - \frac{\sum_{i=1}^{n} u_i{}^{\widehat{m}} \ln d_i}{\sum_{i=1}^{n} d_i{}^{\widehat{m}}} + \frac{\sum_{i=1}^{n} \ln d_i}{n} = 0, \text{ and} \tag{S11}$$

$$\hat{\beta}^{\widehat{m}} = \frac{\sum_{i=1}^{n} d_i{}^{\widehat{m}}}{n} \tag{S12}$$

MLE for the 3-parameter Weibull distribution:

For a known location parameter α, the 3-parameter Weibull distribution is equivalent to the 2-parameter Weibull distribution with d_i substituted by $d_i - \alpha$. Since α is a non-regular parameter[1], $\hat{\alpha}$ cannot be estimated via differential calculus. Hence, we estimate $\hat{\alpha}$ (and corresponding $\hat{\beta}$ and $\widehat{m}$ from eq. S11 and S12) by studying the profile likelihood of α in the interval [0, d_1] discretized in steps of 0.01, where d_1 is the first order statistic.

S1.2.4. MLE for mixture models via Expectation–Maximization (EM) Algorithm

$\mathcal{D}$ is fitted with a mixture model with density function, $f_M(x)$, of the form $f_M(x|p, \theta_1, \theta_2) = pf(x|\theta_1) + (1-p)f(x|\theta_2)$,. In this study a simple mixture model, with two components, has been chosen. Also, the component models are of the same type albeit with different parameters, θ_1 and θ_2. Note that, in general, the application of EM Algorithm is not restricted to this form of a mixture model. $f_M(x)$ has twice the number of parameters as the component models in addition to p. Therefore, maximization of the Likelihood function of $f_M(x)$ over such parameter space is cumbersome. This problem of ML estimation is assuaged by the EM Algorithm which enables us to exploit the numerical simplicity of finding the MLEs of the component models.

In EM Algorithm, $\mathcal{D}$ is considered incomplete as we do not know from which of the two mixture components an element in $\mathcal{D}$ is sampled. Therefore, we augment $\mathcal{D}$ with an indicator variable, z_i, corresponding to every element of $\mathcal{D}$. The resultant organized and complete data is denoted as $\mathcal{C}$.

$$z_i|\mathcal{D} = \begin{cases} 1, & \text{if } d_i \sim f(x|\theta_1) \\ 0, & \text{if } d_i \sim f(x|\theta_2) \end{cases}. \tag{S13}$$

---

[1] Regular parameter is one which satisfies the regularity conditions provided in Ref.[11].

In the above equations, the operator "∼" implies that the variable on LHS was sampled from the component in the RHS. The expectation of the indicator variable z_i, given $\mathcal{D}$, for a certain choice of parameters $\hat{p}, \widehat{\theta_1}, \widehat{\theta_2} \in \phi$ is defined as,

$$E_\phi(z_i|\mathcal{D}) = \frac{\hat{p} f(d_i|\widehat{\theta_1})}{f_M(d_i|\hat{p},\widehat{\theta_1},\widehat{\theta_2})}, \tag{S14}$$

where $E(\cdot)$ is the expectation operator and its subscript represents the choice of θ (which is ϕ in this case) on which the operation is performed. From the definition of z_i, the likelihood function for $f_M(x|\theta)$, given $\mathcal{C}$, can be expressed as,

$$L(\theta|\mathcal{C}) = \prod_i f\,(d_i|\theta_1)^{z_i} f(d_i|\theta_2)^{(1-z_i)}, \tag{S15}$$

where $p, \theta_1, \theta_2 \in \theta$. Note that $L(\theta|\mathcal{C})$ can now be written as $L(\theta_1, \theta_2|\mathcal{C})$. We define the Expected Likelihood, $Q(\theta|\phi)$, of $L(\theta|\mathcal{C})$, given $\mathcal{D}$, for a certain choice of parameters, ϕ, as,

$$Q(\theta|\phi) = E_\phi(L(\theta|\mathcal{C})|\mathcal{D}). \tag{S16}$$

Since z_i is the only random variable in eq. S15, the expectation operator will only operate on it when the expectation is estimated. Therefore, we can write,

$$\log(Q(\theta|\phi)) = \sum_i E_\phi\,(z_i|\mathcal{D}) log(f(d_i|\theta_1)) + (1 - E_\phi(z_i|\mathcal{D})) log(f(d_i|\theta_2)). \tag{S17}$$

Given a certain choice of parameters, ϕ, $Q(\theta|\phi)$ can be formulated. This is referred to as the Expectation (or E) step of the EM algorithm.

The choice of θ that has to be used as ϕ and the utility of E-step in finding $\hat{\theta}$ will now be addressed. An iterative algorithm is formulated, where $\phi = \phi^1$ is the initial guess of the fitting parameters of $f_M(x|\theta)$, given $\mathcal{D}$, and is designated as θ_{init}. On substituting the values of θ_{init} in eq. S14 and S17, the expected likelihood $Q(\theta|\theta_{init})$ can be formulated. Then, the $\widehat{\theta_1}$ and $\widehat{\theta_2}$ that maximizes $log(Q(\theta|\phi))$ can be determined. This is referred to as the Maximization (or M) step. Since the first and second term in eq. S17 depends only on θ_1 and θ_2, the maximization can be decoupled to two parts, one for the MLEs $\widehat{\theta_1}$ and other for the MLEs $\widehat{\theta_2}$, which simplifies the maximization step.

Subsequently, we update ϕ from ϕ^1(=θ_{init}) to ϕ^2 with the MLEs $\widehat{\theta_1}$ and $\widehat{\theta_2}$, calculated from the previous step and $\hat{p} = \sum_i E_{\phi^1}(z_i)/n(\mathcal{D})$ as proportion, where $n(\mathcal{D})$ is the number of elements in the dataset $\mathcal{D}$. Using newly updated ϕ^2, we formulate the next E-step and then perform the M-step to update ϕ^2 to ϕ^3 and so on, until convergence is attained. In the present study, we define convergence to occur when the Eucledian distance between ϕ^k and ϕ^{k+1} is below a tolerance of 10^{-4} , where k is the iteration counter. Our solution for $\hat{\theta}$ would then be ϕ^{k+1}. In Fig. S2, the EM algorithm is illustrated with a flowchart.

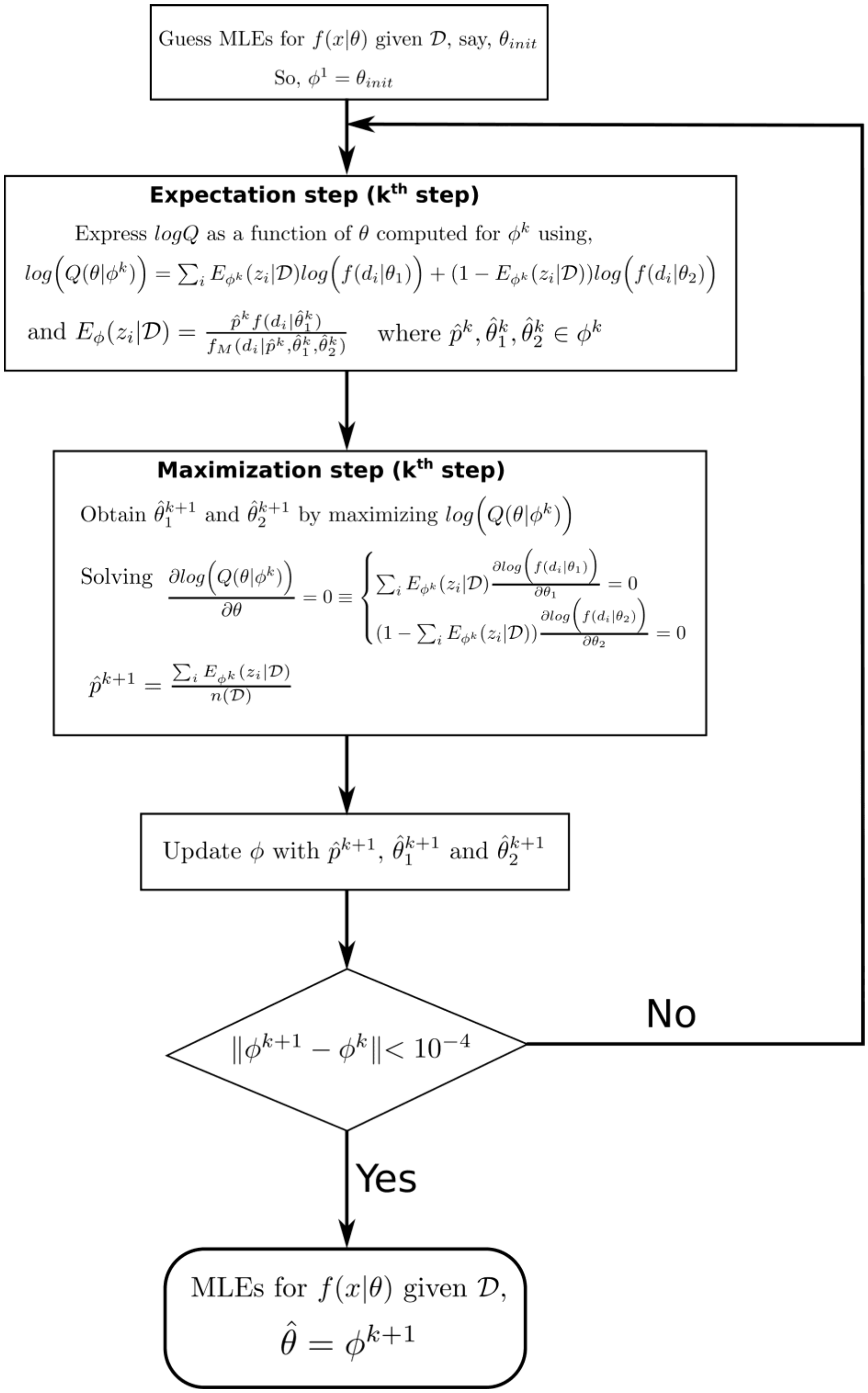

Fig. S2. A Flowchart describing the implementation of the EM Algorithm.

It must be acknowledged that the output of the EM algorithm is dependent on the appropriateness of choice of ϕ^1. In the present study, instead of arbitrarily guessing ϕ^1, Fuzzy

C-Means Clustering algorithm (integrated in Matlab) [2] is employed, whereby the membership grades were directly assigned to $E_{\phi^1}(z_i)$ for the 1st iteration.

The M-step for 2-component mixture of the different statistical models used in this study are discussed next. To maintain brevity, $E\langle z_i \middle| \hat{\theta}^k \rangle$ is denoted as ${c_i}^k$ in the rest of the text.

M-step for 2-component mixture Gaussian distribution:

$$\hat{\mu}^k = \frac{\sum_{i=1}^{n} {c_i}^k d_i}{\sum_{i=1}^{n} {c_i}^k}, \qquad \hat{S}^k = \sqrt{\frac{\sum_{i=1}^{n} {c_i}^k (d_i - \hat{S}^k)^2}{\sum_{i=1}^{n} {c_i}^k}} \tag{S18}$$

M-step for 2-component mixture Lognormal distribution:

$$\hat{\mu}^k = \frac{\sum_{i=1}^{n} {c_i}^k \ln d_i}{\sum_{i=1}^{n} {c_i}^k}, \qquad \hat{S}^k = \sqrt{\frac{\sum_{i=1}^{n} {c_i}^k (\ln d_i - \hat{S}^k)^2}{\sum_{i=1}^{n} {c_i}^k}} \tag{S19}$$

M-step for 2-component mixture 2-parameter Weibull distribution:

$\widehat{m}^k$is estimated iteratively through Regula Falsi from the following equations.

$$\frac{1}{\widehat{m}^k} - \frac{\sum_{i=1}^{n} {c_i}^k {d_i}^{\widehat{m}^k} \ln d_i}{\sum_{i=1}^{n} {c_i}^k {d_i}^{\widehat{m}^k}} + \frac{\sum_{i=1}^{n} {c_i}^k \ln d_i}{\sum_{i=1}^{n} {c_i}^k} = 0 \tag{S20}$$

$$\hat{\beta}^k = \sqrt[\widehat{m}^k]{\frac{\sum_{i=1}^{n} {c_i}^k {d_i}^{\widehat{m}^k}}{\sum_{i=1}^{n} {c_i}^k}} \tag{S21}$$

M-step for 2-component mixture 3-parameter Weibull distribution:

The 2-component mixture 3-parameter Weibull distribution has two location parameters, α_1 and α_2, both of which are non-regular. The domain of these parameters are defined as $0 \le \alpha_1 < \min(d_1, \alpha_2)$ and $0 \le \alpha_2 < d_{n-2}$, where $d_1 < d_2 < \cdots < d_n$ (depicted in Fig. S3). Therefore, we have studied the profile likelihood surface of (α_1, α_2) to estimate the MLEs for this mixture model. However, the implementation of this scheme entails the following intricacies.

The profile likelihood of α_2 is discontinuous at every $\alpha_2 = d_i$. However, the profile likelihood of α_1 given any α_2 is never discontinuous. So we discretize α_2 piecewise over its domain in the following intervals: $[0\ d_1)$, $[d_1 d_2)$, …, $[d_i d_{i+1})$, …, $[d_{n-3} d_{n-2})$. Each interval of α_2 is discretized in steps of 0.01 (if length of an interval is less than 0.03 then the interval is discretized in 3 sub-intervals). Now for every α_2, we study the profile likelihood of α_1 over the domain $0 \leq \alpha_1 < \min(d_1, \alpha_2)$ discretized in steps of 0.01, whereby we get the estimates of $\widehat{\alpha_1}$, $\hat{\beta}$ and $\hat{m}$ ($\hat{\beta}$ and $\hat{m}$ are estimated at fixed (α_1,α_2) via the EM algorithm). This, in turn, leads to the profile likelihood of α_2, which is $\sup_{\alpha_1,\beta,m} L(\alpha_1, \alpha_2, \beta, m|\mathcal{D})$. Global maxima of the profile likelihood of α_2 so estimated corresponds to the MLEs of the 2-component mixture 3-parameter Weibull distribution.

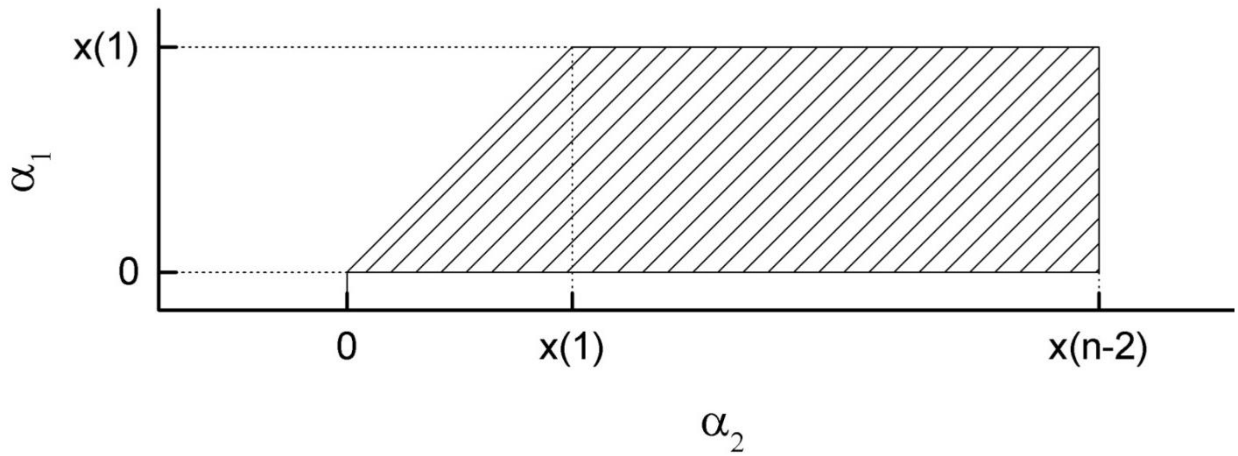

Fig. S3. Domain of (α_1, α_2) over which the profile likelihood of these non-regular parameters is examined for ML estimation of the 2-component 3-parameter Weibull model.

S1.3. Akaike Information Criterion (AIC)

The values of AIC were computed using the following equation:

$$\mathrm{AIC} = -2\ln L(\hat{\theta}) + 2\gamma, \tag{S22}$$

where γ is the number of independent parameters in the model and $L(\hat{\theta})$ is the maximum likelihood. *The model which yields the lowest AIC for a given dataset is the best fitting model for that particular dataset.*

S1.4. Kolmogorov-Smirnov (KS) test

The KS test is a non-parametric hypothesis test that determines whether a sample data can be compared with specific distributions. This test is primarily preferred over other goodness-of-fit tests because it is exact and the test statistic obtained is independent of the underlying cumulative distribution function tested [1]. The statistic obtained from this test is the absolute maximum distance between the empirical distribution function (i.e. obtained from experimental data) and the cumulative distribution function of the reference distribution, say K. Thereafter, based on the value of K, the null hypothesis, H_0, which is the assertion that the empirical data follows the reference distribution, can be tested. Note that the closer the value of K is to 0, the higher the likelihood that the empirical distribution follows the reference distribution. However, the acceptance or rejection of H_0 will depend on the tolerance (confidence bounds) chosen for the value of K. This tolerance is also known as significance level, α_s. In most applications α_s is chosen as 5% although a stricter value of 1% can also be chosen for critical cases. Then, for each α_s, a table containing the critical values of K, K_{crit}, for different data sizes is created. H_0 is accepted for all measured values of K that are lesser than K_{crit} at the chosen α_s [2]. Alternately, this hypothesis testing can also be performed on the basis of the p-value obtained from the KS test. The p-value denotes the threshold value of α_s such that H_0 will be accepted for all values of α_s that are less than the p-value. For instance, if $p = 0.02$, H_0 will be accepted at all $\alpha_s < 0.02$ and rejected for $\alpha_s > 0.02$. In the current study, α_s is chosen as 0.05 and hence if the KS test for any specific model on the datasets returns $p < 0.05$, that particular model can be rejected[2].

S2. Number of activable STZs

An estimate of the total deformed volume, V_{d}, at any given h is [3]:

$$V_{\mathrm{d}} = \frac{\pi h}{6}\left[3{R_{\mathrm{i}}}^2 + h^2\right]. \tag{S23}$$

For a spherical indenter [3],

$$P = \frac{4}{3} E_r \sqrt{R_{\mathrm{i}} h^3}. \tag{S24}$$

[2]Here one must take note of the caveat that the KS test, by virtue of being a hypothesis test, tells whether a model is acceptable or not; it never comments on how good a model fits the data. Therefore, comparing p-value of KS test <u>is not</u> a model selection procedure.

Combining Eqs. (S23) and (S24) to eliminate h, gives V_d as,

$$V_d = \frac{\pi}{6}\left[3R_i{}^2 . \frac{1}{R_i^{\frac{1}{3}}} . \left[\frac{3P}{4E_r}\right]^{\frac{2}{3}} + \frac{1}{R_i}\left[\frac{3P}{4E_r}\right]^2\right]. \tag{S25}$$

For $P = P_{FP}$ and $\tau_y = \tau_{max}$, the average total volume of the material under the indenter that deforms at P_{FP}, $\bar{V}_d^y$, can now be written in terms of the mean value of shear stress, $\bar{\tau_y}$, as (see section 2),

$$\bar{V}_d^y = \frac{\pi}{6}\left[85.69 R_i^{\frac{7}{3}} . \left[\frac{\bar{\tau_y}}{R_i E_r}\right]^2 + 2.3 \times 10^4 R_i \left[\left(\frac{\bar{\tau_y}}{R_i E_r}\right)^3\right]^2\right]. \tag{S26}$$

Using Eq. (S25), an estimate of $\bar{V}_d^y$ can be obtained for indenters with different R_i. However, only a small fraction of this volume, designated as representative volume, V_e, which is approximately equal to ~ 0.01 $\bar{V}_d^y$, provides the optimum conditions for nucleating a shear band [4]. Therefore, any STZ activity, which affects the formation of a shear band, must occur within V_e. Assuming the density of defects, ρ_d, to be approximately equal to 10^{20} m^{-3} [5,6], the number of activable STZs, N_{STZ}, is given as $N_{STZ} = V_e . \rho_d$. Table S1 lists the deformed volumes and N_{STZ} for 3 representative datasets and their corresponding R_i. In specimens indented by a large indenter (R_i ~ 31.5 µm), 388 STZs can potentially get activated whereas only 6 STZs can be accessed during a single indentation when a smaller tip (R_i ~ 5.75µm) is used.

Dataset	**R_i (µm)**	**$\dot{P}$ (mN/s)**	**E_r (GPa)**	**$\bar{\tau_y}$ (GPa)**	**$\bar{V}_d^y$ (m^3)**	**V_e (m^3)**	**Number of STZs, N_{STZ}**
AC1	31.5	1	89	1.5	$3.882\text{x}10^{-16}$	$3.882\text{x}10^{-18}$	388
AC5	5.75	1	89	2.4	$6.035\text{x}10^{-18}$	$6.035\text{x}10^{-20}$	6
A1	1	0.4	97.9	3.5	$5.572\text{x}10^{-20}$	$5.573\text{x}10^{-22}$	0.06

Table S1: The average deformed volume and number of STZs in samples tested with different indenters.

S3. Shear band trajectories and Hertzian Contact relations

According to the Hertz theory of contact [3,7], the relationship between load, P, and displacement, h, is,

$$P = \frac{4}{3}\bar{E}_r\sqrt{R_i}.h^{\frac{3}{2}} \qquad \text{(S27)}$$

where the reduced modulus, $\overline{E_r}$, which accounts for the elastic deformation in both the indenter and the specimen, is given by,

$$\frac{1}{\overline{E_r}} = \frac{1-\nu_s^2}{\overline{E_s}} + \frac{1-\nu_i^2}{\overline{E_i}} \qquad \text{(S28)}$$

where R_i is the indenter radius, $\bar{E}$ and ν are the elastic modulus and Poisson's ratio. The subscripts 's' and 'i' refer to the sample and the indenter, respectively. For the diamond tipped indenter, $\overline{E_i}$ = 1141 GPa and ν_i = 0.07.

The Hertzian contact relations describing the stress field underneath the spherical indenter are listed below (these are taken from Ref. [8]). In the present study, all stress and spatial variables (like r, z etc.) associated with an apostrophe (′) have been normalized with mean contact pressure (P_m) and mean contact radius (a), respectively.

$$a = (0.75PR_i/\overline{E_r})^{1/3} \qquad \text{(S29)}$$

$$P_m = P/(\pi a^2) \qquad \text{(S30)}$$

$$\sigma'_r = \frac{3}{2}\left\{\frac{1-2\nu}{3r'^2}\left[1-\left(\frac{z'}{\sqrt{u'}}\right)^3\right] + \left(\frac{z'}{\sqrt{u'}}\right)^3\frac{u'}{u'^2+z'^2} + \frac{z'}{\sqrt{u'}}\left[u'\frac{(1-\nu)}{1+u'} + (1+\nu)\sqrt{u'}\tan^{-1}\left(\frac{1}{\sqrt{u'}}\right) - 2\right]\right\} \qquad \text{(S31)}$$

$$\sigma'_\theta = -\frac{3}{2}\left\{\frac{1-2\nu}{3r'^2}\left[1-\left(\frac{z'}{\sqrt{u'}}\right)^3\right] + \frac{z'}{\sqrt{u'}}\left[2\nu + u'\frac{(1-\nu)}{1+u'} - (1+\nu)\sqrt{u'}\tan^{-1}\left(\frac{1}{\sqrt{u'}}\right)\right]\right\} \qquad \text{(S32)}$$

$$\sigma'_z = -\frac{3}{2}\frac{z'^3}{\sqrt{u'}(u'^2+z'^2)} \qquad \text{(S33)}$$

$$\tau'_{rz} = -\frac{3}{2}\frac{r'z'^2\sqrt{u'}}{\left(u'^2+z'^2\right)(1+u')}, \tag{S34}$$

where *u'* is the normalized displacement, defined as

$$u' = 0.5\left[r'^2 + z'^2 - 1 + \sqrt{\left(r'^2 + z'^2 - 1\right)^2 + 4z'^2}\right]. \tag{S35}$$

$$\text{Maximum shear stress, } \tau' = \sqrt{\left(\frac{\sigma'_r-\sigma'_z}{2}\right)^2 + \tau'^{\,2}_{rz}} \tag{S36}$$

$$\text{Maximum of maximum shear stress field} = \tau_{max} = 0.31\left(\frac{3}{2}P_m\right) = 0.31\left(\frac{6\bar{E}_r^2}{\pi^3 {R_i}^2}P\right) \tag{S37}$$

$$\text{Hydrostatic stress, } \sigma'_m = \frac{\sigma'_r+\sigma'_\theta+\sigma'_z}{3} \tag{S38}$$

Using eqs. (S31-S35) and (S36), four discrete contours of τ_{max}, A, B, C, D are plotted and displayed in Fig. S4[3]. These contours, labeled A, B, C and D, represent 2-dimensional projections of shear planes where plasticity could potentially initiate, as long as the Mohr-Coulomb yield criterion, which was found to capture the pressure sensitivity of plastic flow in MGs, [9,10], is satisfied.

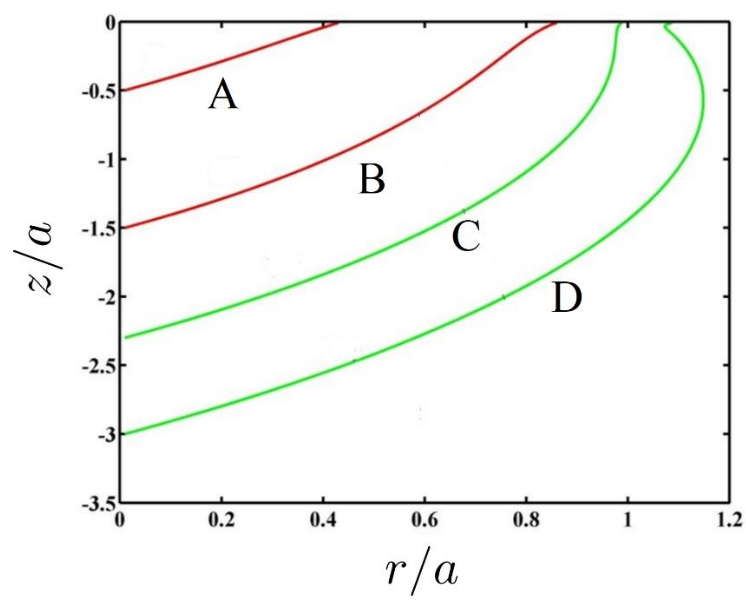

Fig. S4. Contours of τ_{max} in a material calculated from the Hertzian contact relations for a sphere

[3] Although there are different alternatives of shear band trajectories proposed in the literature, they vary only marginally from the one shown here and do not affect the outcome of our discussion

in elastic contact with a flat surface.

References:

[1] J.J. Filliben, W.F. Guthrie, N.A. Heckert, J.D. Splett, N.-F. Zhang, NIST/SEMATECH e-Handbook of Statistical Methods, 2010. http://www.itl.nist.gov/div898/handbook/.

[2] F.J. Massey, The Kolmogorov-Smirnov Test for Goodness of Fit, J. Am. Stat. Assoc. 46 (1951) 68–78. doi:10.1080/01621459.1951.10500769.

[3] A.C. Fischer-Cripps, Introduction to Contact Mechanics, Springer, 2000. http://books.google.co.in/books?id=oLBQKmIfEzsC.

[4] C.E. Packard, C.A. Schuh, Initiation of shear bands near a stress concentration in metallic glass, Acta Mater. 55 (2007) 5348–5358. doi:10.1016/j.actamat.2007.05.054.

[5] J.Q. Wang, J.H. Perepezko, Focus: Nucleation kinetics of shear bands in metallic glass, J. Chem. Phys. 145 (2016). doi:10.1063/1.4966662.

[6] J.H. Perepezko, S.D. Imhoff, M.-W. Chen, J. Wang, S. Gonzalez, Nucleation of shear bands in amorphous alloys., Proc. Natl. Acad. Sci. U. S. A. 111 (2014) 3938–42. doi:10.1073/pnas.1321518111.

[7] K.L. Johnson, Contact mechanics, Cambridge Univ. Cambridge. (1995).

[8] P. Wriggers, Introduction to Contact Mechanics, in: F.F. Ling (Ed.), Comput. Contact Mech., Second Edi, Springer, 2006: pp. 11–29. doi:10.1007/978-3-540-32609-0_2.

[9] A.C. Lund, C.A. Schuh, The Mohr-Coulomb criterion from unit shear processes in metallic glass, Intermetallics. 12 (2004) 1159–1165. doi:10.1016/j.intermet.2004.07.001.

[10] A.C. Lund, C.A. Schuh, Yield surface of a simulated metallic glass, Acta Mater. 51 (2003) 5399–5411. doi:10.1016/S1359-6454(03)00396-3.

[11] Hogg R, McKean J, Craig A, Introduction to Mathematical Statistics; Chapter 6, Sixth Edit, Pearson Education, Upper Saddle River, N.J., 2005. doi:10.2307/2987351.