

MASKE test *basic-Ca(OH)2-CaCO3*

Enrico Masoero

Newcastle University, U.K., December 25, 2020

Abstract The document describes the inputs and outputs of the MASKE test simulation called *basic-Ca(OH)2-CaCO3*. The rationale of the simulation is explained first, followed by a description of the inputs for the chemistry database, the initial configuration, the interaction potentials, and the MASKE input script. The outputs of the simulations show the kinetics of Ca(OH)_2 dissolution and CaCO_3 precipitation, following also the composition of the solution with respect to the relevant equilibrium constants.

Contents

1	General description	1
2	Chemistry database: relevant entries	2
3	Rate of the backward reaction	5
4	Net rate of a reaction	6
5	A particular case: dissolution and precipitation	6
6	Dissolution of a specifically located molecule (explicit in MASKE)	7
6.1	Inverse dissolution rate and net rate, for MASKE	11
7	Nucleation of a specifically located molecule (explicit in MASKE)	12

1 General description

The simulation describes an initial layer of Ca(OH)_2 in pure water at equilibrium with atmospheric CO_2 . Some strong assumptions are made regarding the solution chemistry, since here we do not simulate the equilibration of the solution and of its pH (for that, see examples where MASKE is coupled with PHREEQC). In particular, the formation of Ca(CO)_3 is linked to the reaction of carbonate ions CO_3^{2-} , and an initial concentration of such ions is given. This initial concentration is then let to deplete as Ca(CO)_3 precipitates; in reality, CO_3^{2-} can be replenished both by additional CO_2 dissolving and by other species reacting to form it (carbonic acid and bicarbonate). In the results section below, various initial concentrations of CO_3^{2-} are sampled, as well as various ratios of solution volume per unit surface area of Ca(OH)_2 .

2 Chemistry database: relevant entries

The following molecules are included in the MASKE simulation. The names of the molecules match the names of the corresponding species in the PHREEQC database *CEMDATA18.1*.

- *Ca(OH)2* – solid Portlandite (calcium hydroxide)
Being a solid, only its foreground linear sizes are required in MASKE. It is assumed that the molecule is a cube, hence its linear sizes are obtained as cubic root of the molecular volume. This latter volume, is obtained from the molar mass (74.093 g/mol) and the density (2.21 g/cm³).
- *Ca(OH)2* – solid Portlandite (calcium hydroxide)
Being a solid, only its foreground linear sizes are required in MASKE. It is assumed that the molecule is a cube, hence its linear sizes are obtained as cubic root of the molecular volume. This latter volume, is obtained from the molar mass (74.093 g/mol) and the density (2.21 g/cm³).



with R indicating the rates of the forward and backward reactions.

Transition State Theory (TST) invokes the linear rate theory and assumes the existence of an intermediate state, the activated complex AB_l , which is at equilibrium with the reactants A and B (in the next sections, for the backward reaction, we will postulate an activated complex AB_r with same free energy as AB_l but in equilibrium with the products P instead). Therefore the rate of product formation can be written as:

$$R^+ = \frac{dP}{dt} = k^+ [AB_l] \quad (2)$$

where k^+ is the forward rate constant, in t^{-1} units, and the square brackets indicate concentrations. Thus R is a number of events per unit time, per unit volume.

Furthermore, since AB_l is at equilibrium with A and B, one can introduce the equilibrium constant between reactants and activated complex using activities, $a_A = \gamma_A \frac{[A]}{c}$, where γ is the dimensionless coefficient of activity and c is the standard state concentration (usually, 1 mol per litre; notice that activity is linked to how far the solution is from its standard state, due to the ideal-mixture contribution of concentration, and the interactions accounted for by

γ):

$$K^\ddagger = \frac{a_{AB_l}}{a_A a_B} = \frac{\gamma^\ddagger}{\gamma_A \gamma_B} \cdot \frac{[AB_l]/c_{AB}}{[A]/c_A [B]/c_B} \quad (3)$$

γ^\ddagger is the activity coefficient of the activated complex. All c 's are the same, but I have used subscripts to track whether leftover c 's in the final rate equations will be due to reactants, products, or the activated complex, which will help generalise to higher order reactions too. The literature can be quite messy here, because equilibrium constants are sometimes treated correctly as ratios of activity products, and sometimes treated loosely as ratios of concentrations (equilibrium constant of concentration K_c). Here, for clarity, I will stick to the proper definition of K based on activities and not just on concentrations.

Thus we can rewrite:

$$R^+ = k^+ \frac{\gamma_A \gamma_B}{\gamma^\ddagger} \frac{c_{AB}}{c_A c_B} K^\ddagger [A][B] \quad (4)$$

The rate constant k^+ is simply the frequency at which the activated complex turns into products:

$$k^+ = \kappa \nu \quad (5)$$

ν is the frequency of the vibrational mode causing the conversion of AB_l into P (units of number per unit time). κ is a transmission coefficient between 0 and 1 quantifying the fraction of vibrations towards P effectively resulting in P. Variational TST allows a quantitative treatment of κ , but it is proven that $\kappa = 1$ yields an upper limit of the rate (*e.g.* see Lasaga's book chapter on TST), and this is the typical assumption in classical TST.

So we got to a forward rate equation:

$$R^+ = \kappa \nu \frac{\gamma_A \gamma_B}{\gamma^\ddagger} \frac{c_{AB}}{c_A c_B} K^\ddagger [A][B] \quad (6)$$

If we take $\kappa = 1$, we still have to quantify, or get rid of, ν and K^\ddagger . Statistical mechanics tells us that equilibrium constants are related to partial partition functions per unit volume q_i/V , where i is the generic chemical species (A, B, or AB_l for us), and V is the volume of the system:

$$K = \frac{(q_{AB_l}/V)^{n_{AB_l}}}{(q_A/V)^{n_A} (q_B/V)^{n_B}} \quad (7)$$

To be general, in the equation above I have assumed a more generic reaction $n_A A + n_B B \rightarrow n_{AB_l} AB_l$. In the simpler case I started this document with, all $n = 1$. Notice that I have seen the relationship between K and q proven only for ratios of concentrations in K (typically for gases, only with collisions

and no strong interactions between molecules), never for ratios of activities. However, Lasaga in his Kinetic Theory of the Earth Sciences book, uses the same equation also for proper K defined as ratios of activities. I would love to see a derivation of the $K - q$ relationship also for activities one day...

For our case of reactants going to an activated complex, we have:

$$K^\ddagger = \frac{(q^\ddagger/V)}{(q_A/V)(q_B/V)} \quad (8)$$

Even if we do not have n here any more, we lose no generality because the activated complex is only one, thus $n^\ddagger = 1$ necessarily, whereas keeping n for A and B would not alter the following derivation at all.

A cool step here is that the partial partition function of the activated complex q^\ddagger can be split into two part: the first one q'^\ddagger dealing with all the vibrations that do NOT go towards P, hence proper harmonic modes (in the limit of infinitesimal vibrations); the second one accounting for that only mode that goes towards P. This mode corresponds to a saddle direction in the energy landscape, so it has no stiffness associated: it is a pure translation. The partial partition function per unit volume for the pure translation can be computed analytically (see Lasaga's book, chapter on TST), leading to:

$$K^\ddagger = K'^\ddagger \frac{k_B T}{h\nu} = \frac{(q'^\ddagger/V)}{(q_A/V)(q_B/V)} \frac{k_B T}{h\nu} \quad (9)$$

Substituting K^\ddagger into the R^+ we get:

$$R^+ = \kappa \frac{k_B T}{h} \frac{\gamma_A \gamma_B}{\gamma^\ddagger} \frac{c_{AB}}{c_A c_B} K'^\ddagger [A][B] \quad (10)$$

The last step is to find a simpler expression of K'^\ddagger . This is an equilibrium constant from the reactant to the activated complex, except that the pure translational mode of the complex is removed. This is therefore the constant for two energy minima, which is related to the difference in Gibbs free energy by the usual equation:

$$K = \exp\left(-\frac{\Delta G}{k_B T}\right) \quad (11)$$

This leads to:

$$\begin{aligned} R^+ &= \kappa \frac{k_B T}{h} \frac{\gamma_A \gamma_B}{\gamma^\ddagger} \frac{c_{AB}}{c_A c_B} \exp\left(-\frac{\Delta G^\ddagger}{k_B T}\right) [A][B] = \\ &= \kappa \frac{k_B T}{h} \frac{c_{AB}}{\gamma^\ddagger} \exp\left(-\frac{\Delta G^\ddagger}{k_B T}\right) Q_{reac} \end{aligned} \quad (12)$$

where I generalised by using the activity product of the reactants Q_{reac} . ΔG^\ddagger is the standard activation energy to go from reactants to activated complex.

Notice that the R is still, consistently, a number of events (t^{-1} emerging from $k_B T/h$) per unit volume (from the standard state concentration of the activated complex c_{AB} , usually 1 mol/L when evaluating ΔG^\ddagger , so the rate is in moles per litre too).

The resulting equation features:

- the transmission coefficient κ , usually taken equal to 1 to obtain an upper bound rate in classical chemical kinetics;
- the activity coefficient of the activated complex γ^\ddagger . This is simply 1 when the activated complex is charge neutral (by Debye-Huckel theory for activity coefficients), and this is for example the case of dissolution and precipitation of charge-neutral minerals. However reaction such as speciation might have non-neutral complexes, thus requiring a calculation of γ^\ddagger . This is not too difficult though: one knows the concentrations of A, B, and Q in solution, missing only the concentration of AB_l . However, if you know ΔG^\ddagger , then you also know K'^\ddagger , and this latter is linked to the activities of A, B, and AB. These activities involve activity coefficients, which are functions of all concentrations of A, B, AB, and P (by Debye-Huckel). Hence we end up with one nonlinear equation with one variable only: $[AB_l]$.
- the standard activation energy ΔG^\ddagger
- the activities of the reagents A and B. If there were n_A and n_B reacting, those n would simply be in the exponent of the activities. The rest of the equation would be instead unaltered.

3 Rate of the backward reaction

An assumption behind TST is that the products are also at equilibrium with an activated complex AB_r which has same Gibbs free energy as AB_l but can have a different concentration and activity. This is important because, if one considers only one complex at equilibrium with both reactants and products, than everything would be in equilibrium, hence no kinetics. This caveat is called the quasi-equilibrium hypothesis. Using it, one can follow the same derivation as per the forward reaction, and obtain the rate of the inverse reaction:

$$R^- = \kappa \frac{k_B T}{h} \frac{c_{AB}}{\gamma^-} \exp\left(-\frac{\Delta G^-}{k_B T}\right) a_P \quad (13)$$

ΔG^- and γ^- are analogous to ΔG^\ddagger and γ^\ddagger , but for the inverse reaction from P to AB_r .

The activity coefficient $\gamma^=$ is to be found similarly to γ^\ddagger , starting from the concentration of P and the value of $\Delta G^=$ (and thus the equilibrium constant between P and AB_r), unless the complex is charge neutral, in which case $\gamma^= = 1$.

For the backward activation energy, instead, we can exploit the assumption that AB_l and AB_r have same Gibbs free energy, to write:

$$\Delta G^= = \Delta G^\ddagger - \Delta G_r \quad (14)$$

ΔG_r is the free energy change of the transformation of reactants A and B to products. This is related to the equilibrium constant of the reactions:

$$K_{eq} = \left(\frac{a_P}{a_A a_B} \right)_{eq} = \exp\left(-\frac{\Delta G_r}{k_B T}\right) \quad (15)$$

Substituting into the rate equation:

$$R^- = \kappa \frac{k_B T}{h} \frac{c_{AB}}{\gamma^=} \exp\left(-\frac{\Delta G^\ddagger}{k_B T}\right) \frac{Q_{prod}}{K_{eq}} \quad (16)$$

where I generalised by using the activity product of the products Q_{prod} . Also this R , of course, is in moles per litre from the standard state concentration of the activated complex c_{AB} , which is usually 1 mol/L when evaluating ΔG^\ddagger .

4 Net rate of a reaction

The net rate of the forward reaction is simply $R_{net}^+ = R^+ - R^-$. A way to simplify it a bit is to take $\gamma^\ddagger = \gamma^=$, possibly using the sum of concentrations of AB_l and AB_r to obtain it (*i.e.* assimilating the complexes that are at equilibrium with reactants and products). This leads to:

$$R_{net}^+ = \kappa \frac{k_B T}{h} \frac{c_{AB}}{\gamma^\ddagger} \exp\left(-\frac{\Delta G^\ddagger}{k_B T}\right) \left(Q_{reac} - \frac{Q_{prod}}{K_{eq}} \right) \quad (17)$$

5 A particular case: dissolution and precipitation

In solid-liquid reactions, the thermodynamic and kinetic quantities are usually referred to the dissolution reaction, where the reactant is the solid, and the products are the dissolved molecules. The forward rate is therefore a dissolution rate:

$$R^+ = \kappa \frac{k_B T}{h} \frac{c_{AB}}{\gamma^\ddagger} \exp\left(-\frac{\Delta G^\ddagger}{k_B T}\right) Q_{reac} = \kappa \frac{k_B T}{h} \frac{c_{AB}}{\gamma^\ddagger} \exp\left(-\frac{\Delta G^\ddagger}{k_B T}\right) \quad (18)$$

because the activity of the reactant, *i.e.* the solid, is 1. The inverse reaction, instead, is usually the precipitation, with rate:

$$\begin{aligned}
R^- &= \kappa \frac{k_B T}{h} \frac{c_{AB}}{\gamma^-} \exp\left(-\frac{\Delta G^\ddagger}{k_B T}\right) \frac{Q_{prod}}{K_{eq}} = \\
&= \kappa \frac{k_B T}{h} \frac{c_{AB}}{\gamma^-} \exp\left(-\frac{\Delta G^\ddagger}{k_B T}\right) \frac{Q_{prod}}{Q_{prod,eq}} = \\
&= \kappa \frac{k_B T}{h} \frac{c_{AB}}{\gamma^-} \exp\left(-\frac{\Delta G^\ddagger}{k_B T}\right) \beta
\end{aligned} \tag{19}$$

where the supersaturation β is the ratio of the activity product of the products, here the molecules in solution, divided by the activity product of the same at equilibrium. This equation exploits the fact that the activity of the reactants at equilibrium, within K_{eq} , is 1, because the reactant is the solid.

Thus we can define the net rate per unit volume (moles per litre if c_{AB} is in mol per litre as usual in the evaluation of ΔG^\ddagger) of the forward, dissolution, reaction:

$$\begin{aligned}
R_{net}^+ = R_{net}^{diss} &= \kappa \frac{k_B T}{h} \frac{c_{AB}}{\gamma^\ddagger} \exp\left(-\frac{\Delta G^\ddagger}{k_B T}\right) \left(Q_{reac} - \frac{Q_{prod}}{K_{eq}} \right) = \\
&= \kappa \frac{k_B T}{h} \frac{c_{AB}}{\gamma^\ddagger} \exp\left(-\frac{\Delta G^\ddagger}{k_B T}\right) (1 - \beta)
\end{aligned} \tag{20}$$

Similarly, the rate per unit volume (moles per litre if c_{AB} is in mol per litre as usual in the evaluation of ΔG^\ddagger) of the inverse, precipitation, reaction, is:

$$R_{net}^- = R_{net}^{prec} = \kappa \frac{k_B T}{h} \frac{c_{AB}}{\gamma^\ddagger} \exp\left(-\frac{\Delta G^\ddagger}{k_B T}\right) (\beta - 1) \tag{21}$$

6 Dissolution of a specifically located molecule (explicit in MASKE)

Let's consider the case of particle dissolution in MASKE, where the reactants of type A are explicitly considered as particles, whereas reactants B and products P may be implicit. For a box of volume V containing N particles of type A, the total rate of dissolution, in number of events per unit time (NOT per unit volume) is:

$$\mathcal{R}_{tot}^{diss} = \sum_N \mathcal{R}_i^{diss}(\mathbf{r}) \tag{22}$$

where r is the position vector (rates can depend on local concentrations or interactions, which will be expressed explicitly and not averaged out in the activity coefficients).

The same rate can be expressed using the equation for R^+ (per unit volume) derived before. To this end, we consider that the i^{th} molecule of type A has a location-dependent Voronoi volume $V_i(r)$, so that its concentration is $[A]_i(r) = 1/V_i(r)$. Therefore:

$$\mathcal{R}_{tot}^{diss} = \sum_N R_i^+(r) V_i(r) \quad (23)$$

It is implicitly assumed that R^+ is the same for every r in V_i , which is reasonable if A is a molecule in solution that feels a uniform mean field of interactions and concentrations around it. We will now derive the rate-per-particle equation under this assumption, but later we will consider the case of A being part of a solid lattice, which will require the equation to be re-formulated in terms of average R^+ within V_i . The good news is that, with some quite reasonable assumptions, we will end up with the same equation after all.

Starting with the case of A being an explicit molecule in solution, using Eq. 12 (the first part, with concentrations and not activities), we can write:

$$\begin{aligned} R_i^+(r) &= \kappa \frac{k_B T}{h} \frac{\gamma_A \gamma_B(r)}{\gamma^\ddagger} \frac{c_{AB}}{c_A c_B} \exp\left(-\frac{\Delta G^\ddagger}{k_B T}\right) [A]_i [B](r) = \\ &= \kappa \frac{k_B T}{h} \frac{\gamma_A(r)}{\gamma^\ddagger} \frac{c_{AB}}{c_A} \exp\left(-\frac{\Delta G^\ddagger}{k_B T}\right) [A]_i a_B(r) \\ &= \kappa \frac{k_B T}{h} \frac{\gamma_A(r)}{\gamma^\ddagger} \frac{c_{AB}}{c_A} \exp\left(-\frac{\Delta G^\ddagger}{k_B T}\right) a_B(r) \frac{1}{V_i(r)} \end{aligned} \quad (24)$$

Substituting into the previous equations gives:

$$\mathcal{R}_{tot}^{diss} = \sum_N \kappa \frac{k_B T}{h} \frac{\gamma_A(r)}{\gamma^\ddagger} \frac{c_{AB}}{c_A} \exp\left(-\frac{\Delta G^\ddagger}{k_B T}\right) a_B(r) \quad (25)$$

The activity coefficient $\gamma_A(r)$ can be approximated explicitly during the simulation, if we assume that the excess enthalpy H associated with the finite concentration of A can be expressed as a sum of interaction energy with other explicit phases, U , and surface energy with the implicit species, $\sigma\Omega$, where σ is the interfacial energy between A and the implicit solution. If we also assume that the excess enthalpy changes negligibly with temperature, then we can write:

$$\frac{d \ln(\gamma)}{dT} = -k_B T^2 H \rightarrow \ln(\gamma) = \int -\frac{1}{k_B T^2} H dT = \frac{H}{k_B T} \quad (26)$$

Using this expression in combination with Eqs. 27, we can rewrite:

$$\mathcal{R}_{tot}^{diss} = \sum_N \kappa \frac{k_B T}{h} \frac{1}{\gamma^\ddagger} \frac{c_{AB}}{c_A} \exp\left(-\frac{\Delta G^\ddagger}{k_B T}\right) \exp\left(\frac{U + \sigma\Omega}{k_B T}\right) a_B(r) \quad (27)$$

Notice that favourable interactions U and favourable interfaces $\sigma\Omega$ would have a negative sign, thus making the removal of A to obtain P less frequent.

Finally, we can compare Eq. 22 with Eq. 27 to obtain the deletion rate of a particle A:

$$\mathcal{R}_i^{diss}(\mathbf{r}) = \kappa \frac{k_B T}{h} \frac{1}{\gamma^\ddagger} \frac{c_{AB}}{c_A} \exp\left(-\frac{\Delta G^\ddagger}{k_B T}\right) \exp\left(\frac{U + \sigma\Omega}{k_B T}\right) a_B(r) \quad (28)$$

Notice that, dimensionally, \mathcal{R}_i^{diss} is indeed a number per unit time, because the ratio of standard state concentrations c simplifies to 1 (if both are 1 mol/L). I am not sure what would happen (nor mean) if we also treated B as explicit, but this is not what I am planning to do in MASKE anyhow, so I will need philosophical questions aside for now. It is interesting to note that the final equation also accounts for the possibility that also other ions in solution, B, are consumed during the reaction causing A to be deleted (*e.g.* an alkali ion B contributing to the dissolution of metakaolin..).

The special case of species A being a solid, leads to two slight changes of the equation above. First, c_A is not anymore 1 mol/L but the inverse of the molar volume $V_{m,A}$ in mol/L. Second, the concentration of A is still 1 molecule per molecular volume $1/v_{A,1}$ (because the particle, wherever it is attached to the solid lattice, bulk or surface, spends most of the time in an energy minimum corresponding to a lattice site, where the density is the same as in the bulk) but instead the Voronoi volume associated to a particle depends on its position in the solid: for a bulk particle, that will coincide with $v_{A,1}$, linked to $V_{m,A}$, but for a particle on the surface the tributary volume would be its whole catchment basin in the potential energy surface (*i.e.* the whole part of the PES that the particle can explore with its vibrations). Then the rate equation can be rewritten as:

$$\mathcal{R}_i^{diss}(\mathbf{r}) = \kappa \frac{k_B T}{h} \frac{1}{\gamma^\ddagger} V_{m,A} c_{AB} \exp\left(-\frac{\Delta G^\ddagger}{k_B T}\right) \exp\left(\frac{U + \sigma\Omega}{k_B T}\right) a_B(r) \frac{v_A}{v_{A,1}} \quad (29)$$

Once again, it is worth remarking that \mathcal{R}_i^{diss} is still in units of number of reactions per unit time.

However, this last equation is still based on the initial assumption that the rate per unit volume is constant within the Voronoi volume of the particle. If the particle is part of a solid lattice, such assumption does not hold, because even if we accept all the terms in the prefactor to be constant, U in the exponent is certainly not a constant within the catchment volume v_A of the particle.

If we consider a non-constant rate with r in v_A , we can start by rewriting the one-particle rate equation in a more general form as:

$$\mathcal{R}_i^{diss}(\mathbf{r}) = \langle R_i^+(r) \rangle v_A \quad (30)$$

where the angle brackets indicate the ensemble average of the rate-per-unit-volume over v_A . Notice that this is not simply the average of the rate computed at all possible positions in r , but instead the average value of rate displayed by a particle that can explore all positions in v_A - the key difference is that the portion of v_A that the particle will actually explore is biased by the energy landscape, so that the most probable positions will dominate the rate.

Because we are considering the particle as part of a solid lattice, the term V_i in Eq. 24 is equal to v_A , which thus cancels out as before. This means that the whole derivation above remains the same, except for the ensemble average, thus Eq. 30 simply becomes:

$$\begin{aligned} \mathcal{R}_i^{diss}(\mathbf{r}) &= \left\langle \kappa \frac{k_B T}{h} \frac{1}{\gamma^\ddagger} V_{m,A} c_{AB} \exp\left(-\frac{\Delta G^\ddagger}{k_B T}\right) \exp\left(\frac{U+\sigma\Omega}{k_B T}\right) a_B(r) \right\rangle \frac{v_A}{v_{A,1}} \\ &= \kappa \frac{k_B T}{h} \frac{1}{\gamma^\ddagger} V_{m,A} c_{AB} \exp\left(-\frac{\Delta G^\ddagger}{k_B T}\right) a_B(r) \left\langle \exp\left(\frac{U+\sigma\Omega}{k_B T}\right) \right\rangle \frac{v_A}{v_{A,1}} \end{aligned} \quad (31)$$

where we assumed that every term in the equation is fairly constant within v_A except for the excess energy in the exponential. One way to compute the ensemble average is *via* time average from a dynamic simulation, so that the particle will indeed spend more time in the most likely positions. This however is expensive to repeat for each particle. An alternative is to estimate the average using statistical mechanics and some assumptions. If particle A can occupy N_s energy levels in v_A , we can write:

$$\left\langle \exp\left(\frac{U+\sigma\Omega}{k_B T}\right) \right\rangle = \frac{\sum_{i=1}^{N_s} \exp\left(\frac{U+\sigma\Omega}{k_B T}\right) \exp\left(\frac{-E_i}{k_B T}\right)}{\sum_{i=1}^{N_s} \exp\left(\frac{-E_i}{k_B T}\right)} \quad (32)$$

Notice however that $E = U + \sigma\Omega$, thus the numerator reduces to a sum of ones:

$$\left\langle \exp\left(\frac{U+\sigma\Omega}{k_B T}\right) \right\rangle = \frac{N_s}{\sum_{i=1}^{N_s} \exp\left(\frac{-E_i}{k_B T}\right)} = \frac{N_s}{\frac{1}{\exp\left(\frac{U_1+\sigma\Omega_1}{k_B T}\right)} + \frac{1}{\exp\left(\frac{U_2+\sigma\Omega_2}{k_B T}\right)} + \dots + \frac{1}{\exp\left(\frac{U_{N_s}+\sigma\Omega_{N_s}}{k_B T}\right)}} \quad (33)$$

This last equation is simply the harmonic mean of $\exp\left(\frac{U+\sigma\Omega}{k_B T}\right)$ in v_A . The key property of a harmonic mean, is that it attributes more weight to the smallest terms. In other words, states with low U will be governing the mean, and thus

the rate. Taking this concept to the extreme, we can approximate the whole mean with the minimum U in v_A , to be computed *via* energy minimisation. In this way we are back to exactly Eq. 30, if U there is taken as the result of energy minimisation. (I have thought of possibly adding the average energy of a harmonic oscillator, which is equal to simply $k_B T$, to refine the average energy, but that would add a $+1$ to the exponential, causing troubles to retrieve the usual $(1 - \beta)$ from TST, so I will avoid for now).

A last point to notice is that the expression for the rate of dissolution for an isolated particle A, Eq. 28, has 1 instead of the $v_A/v_{A,1}$ ratio (due to concentration and Voronoi volume being related), and the prefactor differs from that of A being attached to a crystal lattice by a factor c_A *vs.* $V_{m,A}$. The reason for the latter is that the standard states are different. However, if we assume that the activated complex is the same (not necessarily true), the natural log of the ratio between c_A and $1/V_{m,A}$ is the free energy difference between an ideal solution in standard state c_A and the same solution at concentration $1/V_{m,A}$. This would effectively mean that the ΔG^\ddagger for the isolated particle is related to the activation energy for the solid, corrected by a difference in energy that is indeed the concentration ratio above. Bottom line is that one can still use exactly the same equation as derived for the solid, with $V_{m,A}$, also for the isolated particle (under the assumption that the activated complex is the same, and if the equation for the solid assumes $v_A/v_{A,1}=1$). In the correction to ΔG mentioned above I did not consider the activity coefficient of the molecule in solution, because it can still be explicitly calculated *via* the $U + \sigma\Omega$ term in a simulation where A is an explicit particle.

6.1 Inverse dissolution rate and net rate, for MASKE

The inverse process to dissolution of A forming products P, is the precipitation of A starting from the products themselves, anywhere within the volume that was originally occupied by the particle A and that its dissolution would free. Indeed, precipitating in other parts of the box volume, would constitute a separate nucleation event, and not the opposite of the dissolution. The general expression for the inverse rate has been derived already, in Eq. 16, here written again assuming $\gamma^- = \gamma^\ddagger$:

$$R^- = \kappa \frac{k_B T}{h} \frac{c_{AB}}{\gamma^\ddagger} \exp\left(-\frac{\Delta G^\ddagger}{k_B T}\right) \frac{Q_{prod}}{K_{eq}} \quad (34)$$

This is a rate of nucleation per unit volume, hence must be multiplied by a certain volume. The volume corresponding to the dissolution is the catchment

volume v_A of a particle of type A, giving:

$$\mathcal{R}_i^{-diss}(\mathbf{r}) = \kappa \frac{k_B T}{h} \frac{1}{\gamma^\ddagger} c_{AB} \exp\left(-\frac{\Delta G^\ddagger}{k_B T}\right) \frac{Q_{prod}}{K_{eq}} v_A \quad (35)$$

If v_A is in litres and c_{AB} , as usual, in mol/L, then \mathcal{R}_i^{-diss} is in moles per unit time. Taking the moles and combining them with v_A , if A is unimolecular (need to derive what happens if A is not unimolecular), a molar volume appears again yielding:

$$\mathcal{R}_i^{-diss}(\mathbf{r}) = \kappa \frac{k_B T}{h} \frac{1}{\gamma^\ddagger} c_{AB} \exp\left(-\frac{\Delta G^\ddagger}{k_B T}\right) \frac{Q_{prod}}{K_{eq}} V_{m,A} \frac{v_A}{v_{A,1}} \quad (36)$$

$v_{A,1}$ is the actual bulk volume of 1 particle A in the solid, whose molar volume is indeed $V_{m,A}$. If the catchment basin of A extends beyond the $v_{A,i}$, the probability to nucleate a particle that will eventually relax to occupy the position of minimum in the basin is higher.

This inverse rate, together with the forward rate, gives the net rate:

$$\begin{aligned} \mathcal{R}_{i,net}^{diss}(\mathbf{r}) &= \mathcal{R}_i^{diss} - \mathcal{R}_i^{-diss} = \\ &= \kappa \frac{k_B T}{h} \frac{1}{\gamma^\ddagger} V_{m,A} c_{AB} \exp\left(-\frac{\Delta G^\ddagger}{k_B T}\right) \left[a_B(r) \exp\left(\frac{U+\sigma\Omega}{k_B T}\right) - \frac{Q_{prod}}{K_{eq}} \right] \frac{v_A}{v_{A,1}} \end{aligned} \quad (37)$$

It is customary to account for any effect of participating ions activity (ions B) within the prefactor, making the supersaturation of the solution β appear in the equation:

$$\begin{aligned} \mathcal{R}_{i,net}^{diss}(\mathbf{r}) &= \kappa \frac{k_B T}{h} \frac{1}{\gamma^\ddagger} V_{m,A} c_{AB} \exp\left(-\frac{\Delta G^\ddagger}{k_B T}\right) a_B(r) \frac{v_A}{v_{A,1}} \left[\exp\left(\frac{U+\sigma\Omega}{k_B T}\right) - \frac{Q_{prod}/a_B(r)}{K_{eq}} \right] = \\ &= \kappa \frac{k_B T}{h} \frac{1}{\gamma^\ddagger} V_{m,A} c_{AB} \exp\left(-\frac{\Delta G^\ddagger}{k_B T}\right) a_B(r) \frac{v_A}{v_{A,1}} \left[\exp\left(\frac{U+\sigma\Omega}{k_B T}\right) - \beta \right] \end{aligned} \quad (38)$$

For the special case of particles with no excess enthalpy (*e.g.* those on a kink of a Kossel crystal), the last parenthesis yields the usual term $(1 - \beta)$.

Notice that an a priori knowledge of the catchment volume v_A of each particle would be required, in order to apply the last rate equation. Usually we do not know it and sampling it would be costly. If the interaction potentials are short ranged, a good approximation would be $v_A = v_{A,1}$ for any particle. If A is an isolated ion in solution, the same derivation would apply.

7 Nucleation of a specifically located molecule (explicit in MASKE)

Let's consider the case of particle nucleating in MASKE, where the reactants of type P are implicit in the solution, and the products may be explicit (A)

or implicit too (B). The rate to form an explicit particle A within a volume ΔV , at a specific location r , is obtained from the usual forward rate per unit volume:

$$\mathcal{R}^{nucl}(\mathbf{r}) = \kappa \frac{k_B T}{h} \frac{1}{\gamma^\ddagger} c_{AB} \exp\left(-\frac{\Delta G^\ddagger}{k_B T}\right) Q_{reac} \Delta V \quad (39)$$

ΔV is a sampling volume chosen by the user of the Kinetic Monte Carlo code. It should be sufficiently small to neglect any gradient of temperature and Q_{reac} within it, and such that adjacent ΔV around different points do not overlap and eventually cover the whole available space. (I have been tempted to pre-multiply by $(1 - \eta)$, packing fraction, but then I could not nucleate on top of the radius of a surface particle if ΔV is very small, with issues related to comparing results with straight *vs.* net rates later down the line; so for now I did not use η , letting local packing be accounted for by the fact that a particle settling down into a highly packed area will be rapidly deleted, or feel a strong U in the net rate).

Units wise, ΔV above must be expressed in litres per moles, *i.e.* as a molar volume, to simplify the units of c_{AB} . Notice that here, ΔG^\ddagger is that of the forward reaction, which is nucleation. The equation to relate it to the rate of dissolution, *via* the equilibrium constant, was shown in a previous section and will be used later below to discuss the final equations.

The inverse process is deletion of a particle that has formed anywhere within the sampled ΔV . The deletion rate can be obtained following the same logics as in the previous section, but with two points to note: (1) the concentration of A is still given by $1/V_i(r)$ (cf. Eq. 24), so the concentration will remain in the equation along with ΔV ; if A is part of a solid lattice, the concentration becomes $1/V_{m,A}$ (the same discussion and issue of A being in solution *vs.* part of a lattice apply as above); (2) the volume over which the deletion rate should be averaged is now ΔV . This yields a factor $\Delta V/V_{m,A}$ (for A pertaining to a solid) appearing in the derivation, which otherwise remains the same as before until Eq. 30, leading to:

$$\mathcal{R}_i^{diss}(\mathbf{r}) = \kappa \frac{k_B T}{h} \frac{1}{\gamma^\ddagger} c_{AB} \exp\left(-\frac{\Delta G_{diss}^\ddagger}{k_B T}\right) \exp\left(\frac{U + \sigma \Omega}{k_B T}\right) a_B(r) \Delta V \quad (40)$$

Notice that $V_{m,A}$ has been simplified away here, compared to the rate equation for dissolution, because of the ration of volumes mentioned just before the equation. Also, the activation energy is that of dissolution, ΔG_{diss}^\ddagger (also the activity coefficient, but here we will assume it is the same for nucleation and dissolution). To relate it to that of nucleation, we can write (as explained

in a section above) $\Delta G_{diss}^\ddagger = \Delta G^\ddagger - \Delta G_r = \Delta G^\ddagger + k_B T \ln(K_{eq})$, where K_{eq} is that of the nucleation reaction, so $a_A a_B / a_P$ in this case, where A and B are products of nucleation and P is the reactant. Substituting into the rate equation we obtain:

$$\mathcal{R}_i^{diss}(\mathbf{r}) = \kappa \frac{k_B T}{h} \frac{1}{\gamma^\ddagger} c_{AB} \exp\left(-\frac{\Delta G^\ddagger}{k_B T}\right) \exp\left(\frac{U + \sigma \Omega}{k_B T}\right) \frac{a_B(r)}{K_{eq}} \Delta V \quad (41)$$

The net rate of nucleation can thus be written:

$$\begin{aligned} \mathcal{R}_{i,net}^{nuc}(\mathbf{r}) &= \mathcal{R}_i^{nuc} - \mathcal{R}_i^{dis} = \\ &= \kappa \frac{k_B T}{h} \frac{1}{\gamma^\ddagger} c_{AB} \exp\left(-\frac{\Delta G^\ddagger}{k_B T}\right) \left[Q_{reac} - \frac{a_B(r)}{K_{eq}} \exp\left(\frac{U + \sigma \Omega}{k_B T}\right) \right] \Delta V \end{aligned} \quad (42)$$

What U should we take? After long pondering, I think one should run an unbound energy minimisation and take the minimum U within the catchment basin to which ΔV pertained. In this way, for a particle close to the surface of a dissolving solid, we will end up with the same U as in the net dissolution rate, and with the integral of corresponding ΔV 's summing up to catchment volume v_A , just like in the net dissolution rate. In this way the net nucleation rate will be exactly one minus the net dissolution rate.

Remember that ΔG^\ddagger (and γ^\ddagger indeed) are for the nucleation reaction. If we were to express the equation instead using ΔG_{diss}^\ddagger , using the relationship $\Delta G^\ddagger = \Delta G_{diss}^\ddagger - \Delta G_{r,diss} = \Delta G_{diss}^\ddagger + k_B T \ln(K_{eq,diss})$ (see previous section on net rate), we would obtain (assuming same γ for dissolution and nucleation):

$$\begin{aligned} \mathcal{R}_{i,net}^{nuc}(\mathbf{r}) &= \mathcal{R}_i^{nuc} - \mathcal{R}_i^{dis} = \\ &= \kappa \frac{k_B T}{h} \frac{1}{\gamma^\ddagger} c_{AB} \exp\left(-\frac{\Delta G_{diss}^\ddagger}{k_B T}\right) \left[\frac{Q_{reac}}{K_{eq,diss}} - a_B(r) \exp\left(\frac{U + \sigma \Omega}{k_B T}\right) \right] \Delta V \end{aligned} \quad (43)$$

where K_{eq} and $K_{eq,diss}$, multiplied together, gave 1. If we put $a_B(r)$ in the prefactor, and thus to divide Q_{reac} too, we end up with the usual $(\beta - 1)$, with β being the supersaturation of the solution, and with the same prefactor as for the dissolution case discussed before. The only remaining difference is that the sampled volume now is ΔV and not v_A as for the dissolution, but eventually all ΔV within a simulation box volume will be sampled, thus the total rate to nucleate anywhere in the simulation box will not depend on the arbitrary parameter ΔV (except for the precision with which the energy landscape is sampled).

The last equation is conceptually interesting, to link back to usual TST, but for the purpose of simulation it is always best to work with forward reactions only, thus Eq. 42 with all parameters for nucleation will be the useful one.

Notice that, with the equations introduced in this document, running a simulation with net rate or with straight forward and backward rates should give the same results, as long as the catchment volume v_A is correctly estimated in the dissolution rate equation. If not, one can use a lower bound estimate for the dissolution rate using $v_A = v_{A,1}$, and an upper bound based on the interaction threshold (*e.g.* , for a spherical potential with interaction threshold t , the upper bound of v_A could be a sphere with radius t). If estimations of v_A are used in the dissolution rate, the resulting simulation will not in general yield equilibrium at $\beta = 1$. This is true both for a simulation using straight dissolution and nucleation rates, and for a simulation using net dissolution and nucleation rates. One possible way to estimate a meaningful average v_A would be to run simulations and fitting v_A to the target of equilibrium at $\beta = 1$ for appropriate crystal topographies. If instead one uses only net dissolution, equilibrium would be at $\beta = 1$ for any v_A but the rate prefactor would be approximated. If one uses only a net nucleation rate, equilibrium would be at $\beta = 1$ and the rate would be correct (because nucleation samples the PES).

The algorithms to implement the equations obtained thus far would require only a measurement of U for the dissolution for an energy minimised structure, and an unbounded (*i.e.* able to explore the whole space if necessary, differently from Shvab) minimisation of trial particles placed in all possible ΔV 's for the nucleation part. One beneficial result should be that the dependence of the results on ΔV should be less drastic than in Shvab, because now ΔV will only determine the precision with which we estimate the catchment volume v_A , which goes linearly in the prefactor, and not the value of U that in Shvab went in the exponential and got integrated there.