1 LSH bound

Here we repeat the definition given in the paper. For a domain S with distance measupre D, a LSH family is:

Definition 1. *$\mathcal{H} = \{h : S \to U\}$ is called a (r_1, r_2, p, q)-sensitive LSH function family for D if for any two points $x, y \in S$, one function h chosen uniformly at random from $\mathcal{H}$ satisfies:*

- *if $D(x, y) \le r_1$, then $P_{h\in\mathcal{H}}[h(x)=h(y)] \ge p$,*
- *if $D(x, y) \ge r_2$, then $P_{h\in\mathcal{H}}[h(x)=h(y)] \le q$.*

Now we let $r = r_1$, $cr = r_2 (c > 1)$, $p > q$, $h_1, \ldots, h_K \in \mathcal{H}$. Construct $g_i(x) = [h_1(x), \ldots, h_K(x)]$, select L different functions $g_1, \ldots, g_L$. For one point x, hash x into ALL L buckets, denoted by $g_1(x), \ldots, g_L(x)$. x and y is called to "collide" if they collide under any of the $g_1, \ldots, g_L$ functions.

Theorem 1.

- *Under all previous definitions, if $D(x, y) \le r$, then $P(x \text{ collide with } y) \ge 1 - (1 - p^K)^L$.*
- *Under all previous definitions, if $D(x, y) \ge cr$, then $P(x \text{ collide with } y) \le 1 - (1 - q^K)^L$.*

Proof. For (1),

$$\begin{aligned} P(x \text{ collide with } y) &= 1 - P(x \text{ does not collide with } y) && (1)\\ &= 1 - \prod_{i=1}^{L}(1 - P(g_i(x) = g_i(y))) && (2)\\ &\ge 1 - (1 - p^K)^L. && (3) \end{aligned}$$

For (2), similarly,

$$\begin{aligned} P(x \text{ collide with } y) &= 1 - P(x \text{ does not collide with } y) && (4)\\ &= 1 - \prod_{i=1}^{L}(1 - P(g_i(x) = g_i(y))) && (5)\\ &\le 1 - (1 - q^K)^L. && (6) \end{aligned}$$

□

Theorem 2. *For $0 < \epsilon < 0.5$, let $C = \log_\epsilon(1 - \epsilon)$, then $0 < C < 1$, pick $K = \max\{1, \lceil \log_{q/p} C \rceil\}$, pick $L = \frac{1}{\log_\epsilon(1-p^K)}$, under the previous definitions, we must have*

- *For all x, y satisfying $D(x, y) \le r$, $P(x \text{ collide with } y) \ge 1 - \epsilon$.*
- *For all x, y satisfying $D(x, y) \ge cr$, $P(x \text{ collide with } y) \le \epsilon$.*

The first part of theorem 2 can be easily proven by the choice of L. To prove the second part, notice the choice of K satisfies

$$\left(\frac{q}{p}\right)^K \le \log_\epsilon(1 - \epsilon). \tag{7}$$

This implies

$$(1 - p^K)\left(\frac{q}{p}\right)^K \le \log_\epsilon(1 - \epsilon), \tag{8}$$

which is

$$(1-p^K)q^K \leq p^K \log_\epsilon(1-\epsilon). \tag{9}$$

Divide both sides by $1 > (1-p^K) > 0$, we have

$$q^K \leq \left(\frac{1}{1-p^K} - 1\right) \log_\epsilon(1-\epsilon). \tag{10}$$

Then divide both sides by $\left(\frac{1}{1-p^K} - 1\right) > 0$, and re-arrange terms, we have

$$\frac{(1-p^K)-1}{1-\frac{1}{1-p^K}} \leq \log_\epsilon(1-\epsilon). \tag{11}$$

Using the fact that $1 - \frac{1}{x} \leq \ln x \leq x - 1$ for any $x > 0$, we slightly make the nominator smaller, and the denominator bigger, which results in:

$$\log_{(1-p^K)}(1-q^K) = \frac{\ln(1-q^K)}{\ln(1-p^K)} \leq \frac{(1-q^K)-1}{1-\frac{1}{1-p^K}} \leq \log_\epsilon(1-\epsilon). \tag{12}$$

This is:

$$(1-q^K) \geq (1-p^K)^{\log_\epsilon(1-\epsilon)} = (1-\epsilon)^{\log_\epsilon(1-p^K)}. \tag{13}$$

Because the choice of L satisfying $L = \frac{1}{\log_\epsilon(1-p^K)}$, we have $\frac{1}{L} = \log_\epsilon(1-p^K)$. In other words,

$$(1-q^K) \geq (1-\epsilon)^{\log_\epsilon(1-p^K)} = (1-\epsilon)^{1/L}. \tag{14}$$

which implies

$$P(x \text{ collide with } y) \leq 1 - (1-q^K)^L \leq \epsilon. \tag{15}$$

2 Physics background

2.1 Equations

This is the general form of update.

$$\vec{u}(\vec{i}, j+1) = \vec{u}(\vec{i}, j) + \delta_t Q(\{\vec{u}(\vec{i'}, j), \vec{i'} \in N(\vec{i})\}). \tag{16}$$

The discretization of the following two physics equations, namely the Cahn–Hilliard equation and the Allen-Cahn equation have this general form. The Cahn-Hilliard equation looks like:

$$\frac{\partial u}{\partial t} = \nabla \cdot \left(M \nabla \frac{1}{N} \frac{\delta F}{\delta u}\right). \tag{17}$$

And the Allen-Cahn equation has the general form:

$$\frac{\partial v}{\partial t} = -L \, \frac{\delta F}{\delta v}. \tag{18}$$

2.2 Examples

Example 1: Material Grain Growth. In materials science, grain growth is the change of the grains shape (crystallites) in materials. This occurs when the recovery and recrystallisation are complete and further reduction in the internal energy can only be achieved by reducing the surface energy of the grain boundary. The understanding of the factors influencing the evolution of a grain structure is of great scientific and technological importance. In this paper, we use the grain growth model given by Fan and Chen [1]. In the model, each grain is described by one order parameter η_i which takes the value one inside a designated grain and the value zero outside. The evolution function is described by the non-conserved Allen-Cahn equation in the form of:

$$\frac{\partial \eta_i}{\partial t} = -L_i \frac{\delta F}{\delta \eta_i}. \tag{19}$$

Here L_i is the mobility coefficient, N is the number of grains and F is the free energy function which takes the form:

$$F = \int_V \left[f(\eta_1, \eta_2, ...\eta_N) + \sum_{i=1}^{N} \frac{\kappa_i}{2} |\nabla \eta_i|^2 \right] dV \tag{20}$$

Here κ_i are the gradient energy coefficients and f is the local free energy density. The specific form of local free energy which is independent of orientation is given as:

$$f(\eta_1, \eta_2, ...\eta_N) = \sum_{i=1}^{N} (-\frac{A}{2}\eta_i^2 + \frac{B}{4}\eta_i^4) + \sum_{i=1}^{N} \sum_{j=i+1}^{N} \eta_i^2 \eta_j^2 \tag{21}$$

in which A and B are positive constants. Substituting Eq 21 and Eq 20 into Eq 19, we can have the governing equation for evolution as:

$$\frac{\partial \eta_i}{\partial t} = -L_i(-A\eta_i + B\eta_i^3 + 2\eta_i \sum_{i \neq j}^{N} \eta_j^2 - \kappa_i \nabla^2 \eta_i)$$

In this equation, we can see that it matches the general form of the PDE given in Equation 1 in the main text in this way: $\frac{\partial \vec{u}(\vec{p},t)}{\partial t}$ is $(\frac{\partial \eta_1}{\partial t}, \ldots, \frac{\partial \eta_N}{\partial t})$. $D(\vec{u})$ can be viewed as the total of all the terms that has η_i or η_i^2 or η_i^3 as base. $G(\vec{u})\nabla F(\vec{u})$ is not presented here. And $I(\vec{u})\nabla^2 H(\vec{u})$ can be viewed as the sum of terms $\kappa_i \nabla^2 \eta_i$.

Example 2: Nanovoid Evolution. Nanovoid evolution incorporates a coupled set of Cahn–Hilliard (Equation 17) and Allen–Cahn equations (Equation 18) to capture the processes of point defect generation and recombination, annihilation of defects at sinks. The phase-field model includes 3 field variables, c_v, c_i, and η, which are defined to describe the void fraction, the interstitial fraction and the void cluster concentration. These variables and vary both spatially and temporally on a 2-dimensional space. The free energy function F in here is described as:

$$F = N \int_V \left[h(\eta) f^s(c_v, c_i) + j(\eta) f^v(c_v, c_i) + \frac{\kappa_v}{2} |\nabla c_v|^2 + \frac{\kappa_i}{2} |\nabla c_i|^2 + \frac{\kappa_\eta}{2} |\nabla \eta|^2 \right] dV. \tag{22}$$

Here, $f^s(c_v, c_i)$ is the contribution term from the solid phase. $h(\eta) = (\eta - 1)^2$, $f^i(c_v, c_i)$ is the contribution term from the void phase, and $j(\eta) = \eta^2$. We use the formulation from [2] for f^s and f^v: $f^s(c_v, c_i) = E_v^f c_v + E_i^f c_i + k_B T[c_v \ln c_v + c_i \ln c_i + (1 - c_v - c_i) \ln(1 - c_v - c_i)]$. $f^v(c_v, c_i) = (c_v - 1)^2 + c_i^2$. Substituting above formulas into Eq 17, we can have the governing equation for evolution of c_v as:

$$\frac{\partial c_v}{\partial t} = M_v \nabla^2 \left[(\eta - 1)^2 (E_v + k_B T \ln(c_v) - k_B T \ln(1 - c_v - c_i)) + \eta^2 2(c_v - 1) - \kappa_v \nabla^2 c_v \right] \tag{23}$$

In this equation, we can see that it matches the general form of the PDE given in Equation 1 in the main text in this way: $\frac{\partial \vec{u}(\vec{p},t)}{\partial t}$ is $\frac{\partial c_v}{\partial t}$, $\frac{\partial c_i}{\partial t}$ and $\frac{\partial \eta}{\partial t}$. And $D(\vec{u}) + G(\vec{u})\nabla F(\vec{u})$ is omitted here because the equation doesn't contain any terms like them. And $I(\vec{u})\nabla^2 H(\vec{u})$ is the entire right hand side of the equation.

The discrete form is

$$c_{vt+1} = c_{vt} + dtM_v\nabla^2 N\left[h(\eta_t)\frac{df^s(c_{vt}, c_{it})}{dc_{vt}} + j(\eta_t)\frac{df^v(c_{vt}, c_{it})}{dc_{vt}} - \kappa_v\nabla^2 c_{vt}\right]$$

where c_{vt} means c_v values at time t, c_{it} means c_i values at time t, η_t means η values at time t, $h(\eta_t)\frac{df^s(c_{vt},c_{it})}{dc_{vt}} = (\eta_t - 1)^2(E_v + k_BT\ln(c_{vt}) - k_BT\ln(1 - c_{vt} - c_{it}))$, and $j(\eta_t)\frac{df^v(c_{vt},c_{it})}{dc_{vt}} = \eta_t^2 2(c_{vt} - 1)$.

Similarly, we can get the governing equation for c_i as:

$$\frac{\partial c_i}{\partial t} = M_i\nabla^2\left[(\eta - 1)^2(E_i + k_BTln(c_i) - k_BTln(1 - c_v - c_i)) + \eta^2 2c_i - \kappa_i\nabla^2 c_i\right] \tag{24}$$

And its discrete form can be written as:

$$c_{it+1} = c_{it} + dtM_i\nabla^2\left[h(\eta_t)\frac{df^s(c_{vt}, c_{it})}{dc_{it}} + j(\eta_t)\frac{df^v(c_{vt}, c_{it})}{dc_{it}} - \kappa_i\nabla^2 c_{it}\right]$$

where c_{vt} means c_v values at time t, c_{it} means c_i values at time t, η_t means η values at time t, $h(\eta_t)\frac{df^s(c_{vt},c_{it})}{dc_{it}} = (\eta_t - 1)^2(E_i + k_BT\ln(c_{it}) - k_BT\ln(1 - c_{vt} - c_{it}))$, and $j(\eta_t)\frac{df^v(c_{vt},c_{it})}{dc_{it}} = \eta_t^2 2c_{it}$.

Finally, we can get the governing equation for evolution of η as:

$$\frac{\partial \eta}{\partial t} = -LN(2(\eta - 1)f^s(c_v, c_i) + 2\eta f^v(c_v, c_i) - \kappa_\eta\nabla^2\eta) \tag{25}$$

And its discrete form is:

$$\eta_{t+1} = \eta_t - dtLN(2(\eta - 1)f^s(c_{vt}, c_{it}) + 2\eta_t f^v(c_{vt}, c_{it}) - \kappa_\eta\nabla^2\eta_t)$$

where c_{vt} means c_v values at time t, c_{it} means c_i values at time t, η_t means η values at time t, $f^s(c_v, c_i) = E_v^f c_v + E_i^f c_i + k_BT[c_v\ln c_v + c_i\ln c_i + (1 - c_v - c_i)\ln(1 - c_v - c_i)]$. $f^v(c_v, c_i) = (c_v - 1)^2 + c_i^2$.

References

[1] D. Fan and L.-Q. Chen. Computer simulation of grain growth using a continuum field model. *Acta Materialia*, 45(2):611–622, 1997.

[2] P. C. Millett, A. El-Azab, S. Rokkam, M. Tonks, and D. Wolf. Phase-field simulation of irradiated metals: Part i: Void kinetics. *Computational materials science*, 50(3):949–959, 2011.