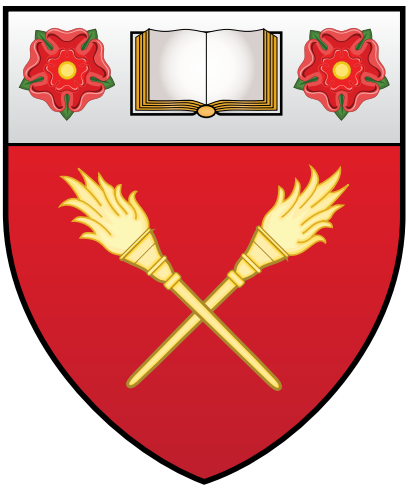
Dislocation Based Modelling of Fusion Relevant Materials.



Daniel Celis Garza

University of Oxford

Harris-Manchester College

Department of Materials

Supervisors: Edmund Tarleton & Angus Wilkinson

A thesis submitted for the degree of

Doctor in Philosophy

Trinity Term 2020

Dedication

"Ohana significa familia, y tu familia nunca te abandona, ni te olvida."
"Ohana means family, family means nobody gets left behind, or forgotten."
– Stitch, Lilo & Stitch

Malle santa, cuando te hablé desde el trabajo para darte la noticia me dijiste que siempre lo supiste. Recuerdo que lloraste en el teléfono, se te cerró la garganta y a mí también. A diferencia de otras veces, no pudimos platicar mucho porque nos quedamos sin aliento. Así que le hablaste a abuelita Raquel mientras le hablé a papá. Madre, David y yo te debemos tanto que no podemos repagar en mil vidas, pero ten por seguro que eres nuestro ejemplo a seguir. Nuestra madre chingona y chambeadora, solo hay una y como ella no hay ninguna. Sin tí el mundo sería un lugar más cruel y pobre, no merece un ángel tan grande y puro como tú.

Dad, "Igualito que tu jefe, wey." con tu risa característica fue lo primero que dijiste cuando te avisé "Sí dad, igualito que mi jefe." entre risas respondí. Como siempre, no platicamos mucho pero esta vez porque le querías avisar a abuelita Teté y a mis tíos, y yo le quería avisar a David. Pensé que seguirías aquí para carcajearte al verme en la túnica ridícula, así como lo hicimos nosotros cuando te vimos a tí en la tuya. Escribo esta dedicatoria antes de acabar porque una promesa es una promesa y esta madre la voy a acabar. No te tendré para pedirte consejos y contarte mis avances y tropiezos. Pero a veces pretendo que me escuchas mientras intento comprender o arreglar algo. Te queremos y te extrañamos muchísimo. Buenas noches, dad.

David, "Te mamaste we." me dijiste cuando me abrazaste al llegar a casa después del trabajo el día que me aceptaron. Me has hecho un chingo de falta ojete, te extraño mucho we. Perdón por no hablar tan seguido, pero me duele colgar we. No estoy tan chisqueado como mamá pero siento feo cuando terminamos de platicar. Me gusta mucho ver tus streams porque me recuerda un poco a sentarme a tu lado a verte jugar… o a cuando veíamos Twitch juntos y veíamos a TB missing legal.

Esto es para mi Ohana por sangre y por elección. Ustedes creyeron en mí cuando yo no lo hacía. Me empujaron a ser mejor. Me extendieron la mano cuando nadie más lo hizo. Me hicieron reir cuando solo sabía llorar. Gracias por hacerme quien soy.

This is for my Ohana by blood and by choice. You believed in me when I did not. You pushed me to be better. Offered me a helping hand when nobody would. Made me laugh when all I knew was sorrow. Thank you for making me who I am.

Acknowledgements

"You can always judge a man by the quality of his enemies."
– Oscar Wilde

I did not get here alone. Truly I don't have much idea what I did to deserve the help and encouragement of such an eclectic mix of awesome people. The list is long, but like the proverbial beating of a butterfly's wings, there is no telling where I'd be without the aid and support of these people.

- Ed and Angus
- Fusion CDT
- Harris-Manchester
- CONACYT
- Anatoly Kolomeisky, John F. Stanton
- Isla
- Clan del Rano Sentado
- Goosehood
- Kopse Lane Krump v1.0 v2.0
- Kristmas Karnage: n
- OUPLC: Perhaps the heaviest things that we lift are not our weights, but our feels.
- Damian Rouson
- Bruce Bromage, Haiyang Yu, Fenxian Liu, Daniel Hortelano-Roig, Junnan Jiang

I guess I'll content myself with being judged unfavourably by the world.

Contents

List of Figures

List of Tables

List of Algorithms

Outline

1. Lit review
2. DDD-FEM
3. Analytic tractions
    (a) Traction errors
    (b) Reaction force errors
    (c) Simulations
    (d) Recommendations and Conclusions
4. Parallelising tractions
    (a) Algorithms
    (b) Performance
    (c) Simulations
    (d) Recommendations and Conclusions
5. EasyDD v2.0
    (a) Adaptive integration
    (b) New mobility law
    (c) Surface velocities
    (d) Collisions
    (e) Code redesign
6. Future work
    (a) Julia redesign
    (b) Preliminary comparisons
7. Conclusions

8. Appendices

    (a) Collaboratory summary

Chapter 0

Preface

0.1 Notation

For the sake of clarity the following notation conventions have been used:

- Tensors are denoted by sans serif bold italics, $\boldsymbol{T}$.
- Matrices are denoted by serif bold roman, $\mathbf{M}$.
- Vectors are denoted by serif bold italics, $\boldsymbol{V}$.
- Scalars are denoted by serif italics, S.
- Host variables are denoted by ${}^{\mathrm{h}}a$.
- Global device variables are denoted by ${}^{\mathrm{d}}a$.
- Thread variables are denoted by ${}^{\mathrm{t}}a$.

0.2 Conventions

All angles are given in radians unless stated otherwise. All pseudo-code is C-style and assuming C conventions (row-order, 0-start indexing, memory allocation).

Elements, be they Surface Elements (SEs), Finite Elements (FEs), dislocation line segments or else, are denoted by e and their total as E. Similarly nodes are denoted by n and their total N. Parallel indices are denoted by idx to distinguish them from regular indices/counters i, j, k. For disambiguation purposes the exponential function will de denoted as $\exp(x)$.

0.3 Typesetting

This document was typeset using a custom-made LaTeX document class created by the author, compiled with XeLaTeX, and the bibliography was produced with BibTeX. The custom document class can be found in https://github.com/dcelisgarza/latex_temp

0.4 Diagrams

All diagrams are vector graphics drawn using the open-source image editing software InkScape for its ease of use, elegant simplicity, high quality outputs, and "Draw Freely" philosophy.

Chapter 1

Introduction

1.1 Section 1

1.1.1 Subsection 1

Chapter 2

Coupling Discrete Dislocation Dynamics to Finite Element Methods

2.1 Superposition Scheme

Coupling Discrete Dislocation Dynamics (DDD) to Finite Element Methods (FEMs) [1] is important to properly simulate micromechanical tests because DDD provides us with a more precise set of inputs and greater granularity for solving the FE problem. This can be achieved by using a so-called superposition scheme fig. 2.1 that enables the independent solution of both problems, whilst feeding information from one to the other in a continuous feedback loop.

2.2 Extracting Surface Nodes

The FEM coupler arranges the nodes starting on the xz-plane where $y = 0, \ldots, n\mathrm{d}y,\ n \in \mathbb{N}$. However in order to couple DDD to FEM we only require the surface nodes where displacements are not calculated. Because we're working with rectangular prisms, we can easily pick out the surface nodes using a search algorithm with a logical mask. MATLAB and Fortran provide vector intrinsics that allow one to do so. Figure 2.2a illustrates only the surface nodes according to our implementation's node arrangement—which is the xz-plane going from $y = y_{\min} \rightarrow y = y_{\max}$. However, due to the nature of the analytical solutions in chapter 3, we need all the surface nodes of the rectangular faces for which there are no displacements. This means that edge nodes are shared between 2 adjacent faces and corner nodes between 3 adjacent faces. In order to properly apply the logical mask to find *only* the surface nodes need to know that each FEs' nodes are numbered according to

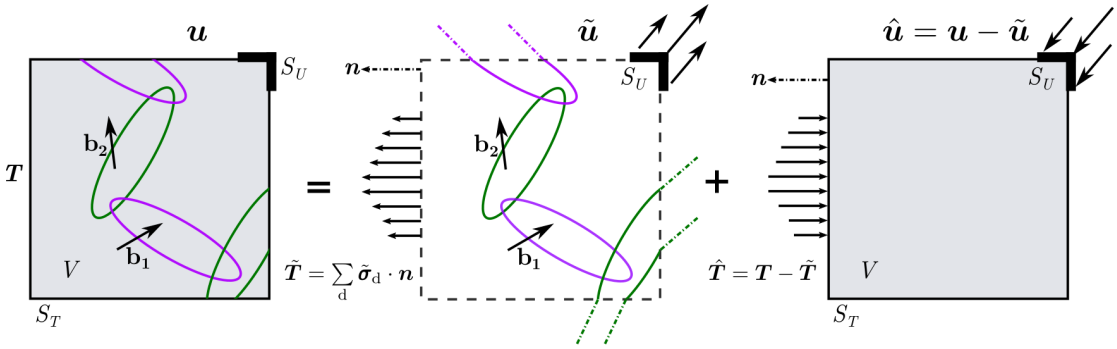
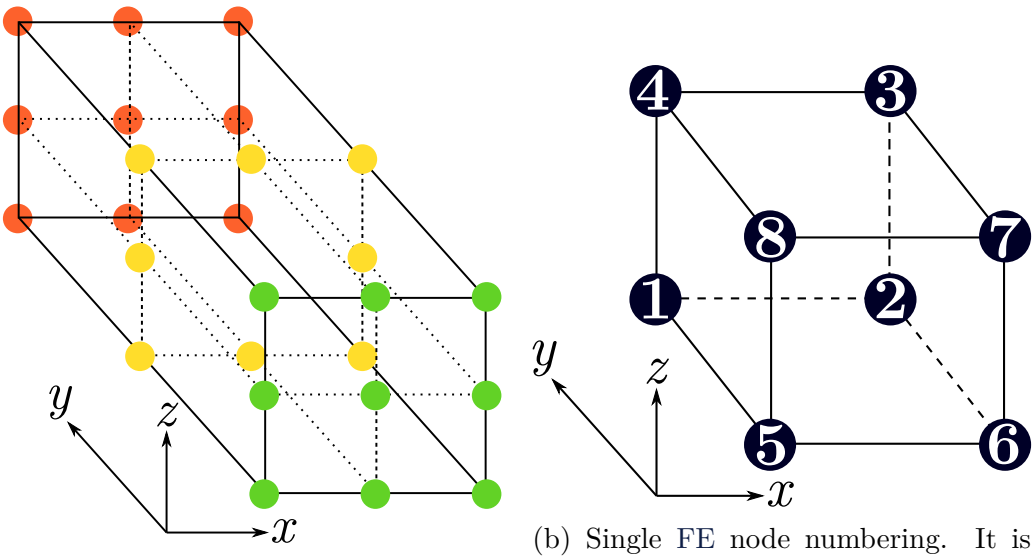


Figure 2.1: The dislocation ensemble in a volume V is bounded by surface S. First, the traction field $\sum_{\mathrm{d}} \tilde{\boldsymbol{\sigma}}_{\mathrm{d}}$ due to the dislocation ensemble is evaluated at the surface. Then, a traditional FEM or Boundary Element Method (BEM) calculates the image traction field $\hat{\boldsymbol{\sigma}} \times \boldsymbol{n}$. Which is then fed back to the DDD problem to evolve the dislocation positions and repeat the cycle. Image edited from [1].



(a) Arrangement of the surface nodes of our FE model. FE nodes are arranged in chunks of $\Delta x \Delta z$ nodes in our implementation, but we only want the surface nodes.

(b) Single FE node numbering. It is necessary to know which FE plane corresponds to which node labels. This lets us design the auxiliary matrix that selects nodes according to the planes we want to extract.

Figure 2.2: FE node arrangement for coupling to DDD.

fig. 2.2b.

Using fig. 2.2 one can work out which nodes are of interest to whichever surface is being extracted. The ordering of the nodes in the final array will depend on the definition of the problems in chapter 3.

Node selection remains an expensive operation and minimising array indexing is of the utmost importance for the best perfomance. Selecting nodes in the traditional sense, i.e. with code branching such as `if statements` or `case selection` is unmantainable, verbose and very prone to mistakes. The issue was solved by introducing an auxiliary matrix which defines various parameters that aid node selection and greatly reduces code size, improves readability, and eliminates the need for code branching. The matrix can be constructed utilising fig. 2.2 in order to know which nodes correspond to which FE planes. The p^{th} column of the matrix[1] corresponds to the p^{th} plane (according to an arbitrary plane numbering) and is defined as,

$$\boldsymbol{V_p}^{\mathsf{T}} = \begin{bmatrix} L_{1p} & L_{2p} & \cdots & L_{Np} & A_p & C_p \end{bmatrix}, \tag{2.1}$$

where L_{np} is the numeric label for node n as given by fig. 2.2, A_p is the area of the plane, and C_p is the numeric label of the orthogonal coordinate to the plane $C_p = 1,\ 2,\ 3$ for the $x,\ y,\ z$ coordinates respectively. A_p lets us segment our output and transitional arrays so that the only data being modified is that which corresponds to the correct plane and C_p lets us know which coordinate we must use in our selection criteria. Using our particular node labelling scheme (with dimensions $\Delta x,\ \Delta y,\ \Delta z$ respectively in the $x,\ y,\ z$ directions), the matrix is defined as,

$$\mathbf{V} = \begin{bmatrix} 5 & 2 & 6 & 1 & 5 & 4 \\ 1 & 6 & 5 & 2 & 6 & 3 \\ 8 & 3 & 7 & 4 & 1 & 8 \\ 4 & 7 & 8 & 3 & 2 & 7 \\ \Delta y\Delta z & \Delta y\Delta z & \Delta x\Delta z & \Delta x\Delta z & \Delta x\Delta y & \Delta x\Delta y \\ 1 & 1 & 2 & 2 & 3 & 3 \end{bmatrix}. \tag{2.2}$$

The information codified in eq. (2.2) lets us index and process only the necessary columns to extract the surface nodes we're interested in. The advantage of this setup over a naïve implementation is that it can be relatively easily expanded, maintained, and is general enough that it lends itself to a variety of selection criteria. The columns from left to right (1 to 6) represent: face 1 $\equiv \min(x)$, yz-

[1] MATLAB uses column-major ordering, so this gives us the best performance for vectorised code.

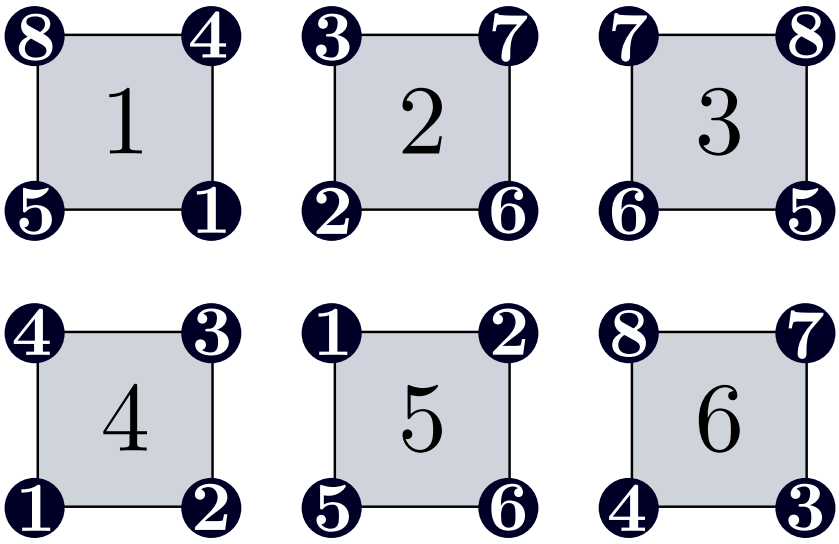


Figure 2.3: Self-consistent, chirality preserving surface planes. Plane and node numbering according to the columns of eq. (2.2) and fig. 2.2.

plane; face $2 \equiv \max(x)$, yz-plane; face $3 \equiv \min(y)$, xz-plane; face $4 \equiv \max(y)$, xz-plane; face $5 \equiv \min(z)$, xy-plane; face $6 \equiv \max(z)$, xy-plane.

It is important to make sure the nodes are extracted in a self consistent manner because the problem definitions in chapter 3 assume a specific node ordering. In particular, the labelling chirality must remain the same. This is due to the fact that the problems in chapter 3 utilise internal coordinates derived from a set of basis vectors, and scalar projections onto them. If the relationship between the vectors is kept, but the chirality is different (opposite) the calculated quantities will have the wrong sign. If however the relationship between the vectors is different to the problem formulation, the program will crash due to division by 0. Self-consistency was achieved by placing an imaginary observer inside the FE model facing the $\min(x)$, yz-plane (face 1) and rotating it to view all 6 planes according to fig. 2.3.

2.3 Mapping Forces

$\hat{\boldsymbol{F}}$ only cares about the global node numbers, of which there are M. The columns of $\mathbf{N_L}$, $\mathbf{F_N}$ represent the 4 nodes of a SE, the rows represent the SE they are part of. The column vector $\boldsymbol{x}$ represents the x, y, z-coordinates of the node-element or global node corresponding to their subscript. The column vector $\boldsymbol{\gamma_t}$ is the list

Algorithm 2.1 If $\hat{\boldsymbol{F}}$'s columns are arranged the same way as $\boldsymbol{\gamma_t}$.

▷ Loop through the array containing the node labels of the relevant surface nodes.
for $i = 0;\ i < \text{length}(\boldsymbol{\gamma_t});\ i++$ **do**
▷ Save the global node label for the current iteration.
$n \leftarrow \boldsymbol{\gamma_t}[i]$
▷ Use the node label to find a vector with the linearised index of all surface nodes whose labels correspond to the node whose forces we want to pass on to the FEM coupler.
$\boldsymbol{L} \leftarrow \text{find}(\mathbf{N_L} == n)$
▷ Loop over coordinates.
for $k = 0;\ k < 3;\ k++$ **do**
▷ Use global node label vector to index the force array from the analytical force calculation due to dislocations on surface nodes. Multiplied by 3 because there are three coordinates per node. We sum the forces from the analytical calculation because the same global node can be part of multiple surface elements. We add k because the $x,\ y,\ z$ coordinates are consecutively stored in $\mathbf{F_n}$.
$\hat{\mathbf{F}}[3 \times \boldsymbol{L} + k] \leftarrow \hat{\mathbf{F}}[3 \times \boldsymbol{L} + k] + \sum \mathbf{F_n}[3 \times \boldsymbol{L} + k]$
end for
end for

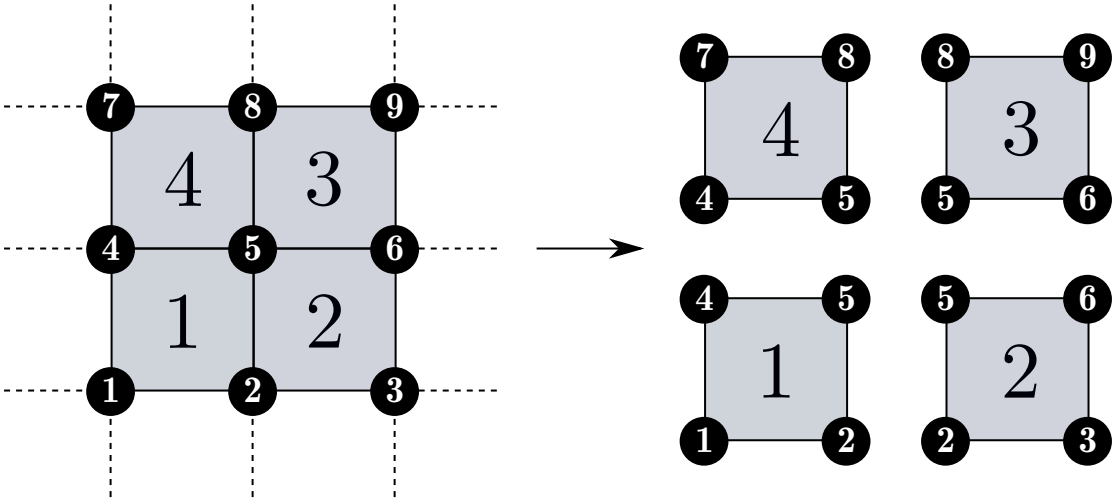


Figure 2.4: FE nodes shared by multiple SEs.

of global nodes for which the forces induced by dislocations must be calculated.

$$\boldsymbol{x_{e,n}}^{\mathsf{T}} \equiv \begin{bmatrix} x_{en} & y_{en} & z_{en} \end{bmatrix}, \quad \hat{\boldsymbol{x}}_{\boldsymbol{n}}^{\mathsf{T}} \equiv \begin{bmatrix} x_n & y_n & z_n \end{bmatrix} \tag{2.3}$$

$$\mathbf{N_L} = \begin{bmatrix} l_{1,1} & l_{1,2} & l_{1,4} & l_{1,5} \\ l_{2,2} & l_{2,3} & l_{2,5} & l_{2,6} \\ l_{3,5} & l_{3,6} & l_{3,8} & l_{3,9} \\ l_{4,4} & l_{4,5} & l_{4,7} & l_{4,8} \end{bmatrix}, \quad \boldsymbol{\gamma} = \begin{bmatrix} l_1 \\ l_2 \\ \vdots \\ l_9 \end{bmatrix} \tag{2.4}$$

$$\mathbf{F_e} = \begin{bmatrix} \boldsymbol{x_{1,1}} & \boldsymbol{x_{1,2}} & \boldsymbol{x_{1,3}} & \boldsymbol{x_{1,4}} \\ \boldsymbol{x_{2,1}} & \boldsymbol{x_{2,2}} & \boldsymbol{x_{2,3}} & \boldsymbol{x_{2,4}} \\ \boldsymbol{x_{3,1}} & \boldsymbol{x_{3,2}} & \boldsymbol{x_{3,3}} & \boldsymbol{x_{3,4}} \\ \boldsymbol{x_{4,1}} & \boldsymbol{x_{4,2}} & \boldsymbol{x_{4,3}} & \boldsymbol{x_{4,4}} \end{bmatrix}, \quad \hat{\boldsymbol{F}} = \begin{bmatrix} \hat{\boldsymbol{x}}_{\boldsymbol{1}} \\ \hat{\boldsymbol{x}}_{\boldsymbol{2}} \\ \hat{\boldsymbol{x}}_{\boldsymbol{3}} \\ \hat{\boldsymbol{x}}_{\boldsymbol{4}} \end{bmatrix}. \tag{2.5}$$

Chapter 3

Analytical Forces Induced by Dislocations on Linear Rectangular Surface Elements

3.1 Forces Exerted by a Dislocation Line Segment on Linear Rectangular Surface Elements

Coupling DDD to FEM requires the traction field $\boldsymbol{\sigma}^{\infty}(\boldsymbol{x}) \cdot \boldsymbol{n}(\boldsymbol{x})$ to be distributed among the set of relevant discrete nodes of a FE or Boundary Element (BE) model. This is usually achieved as follows,

$$\boldsymbol{F}^{(m)} = \int_{S^e} \left[\boldsymbol{\sigma}^{\infty}(\boldsymbol{x}) \cdot \boldsymbol{n}(\boldsymbol{x})\right] N_m(\boldsymbol{x}) \, \mathrm{d}S_e \,, \tag{3.1}$$

where $\mathrm{d}S_e$ is the infinitesimal surface element with surface area S_e. $N_m(\boldsymbol{x})$ are so-called shape functions (interpolation functions) that distribute the traction field among the surface element's nodes.

The problematic singularity associated with the classical Volterra dislocation is avoided by using the non-singular formulation of Cai et al. [2]. The stress field of a dislocation in a homogenous infinite linear elastic domain can be calculated as a contour integral along the loop [3],

$$\sigma_{ij}^{\infty}(\boldsymbol{x}) = C_{ijkl} \oint \epsilon_{lnh} C_{pqmn} \frac{\partial G_{kp}(\boldsymbol{x} - \boldsymbol{x}')}{\partial x_q} b_m \mathrm{d}x'_h \,, \tag{3.2}$$

where C_{ijkl} is the elastic stiffness matrix, ϵ_{lnh} the permutation operator, $\boldsymbol{b}$ the Burgers vector, $\boldsymbol{x}'$ the coordinate that spans the dislocation, and $G_{kp}(\boldsymbol{x} - \boldsymbol{x}')$ is Green's function of elasticity [3]. $G_{kp}(\boldsymbol{x} - \boldsymbol{x}')$ is defined as the displacement component in the x_k direction at point $\boldsymbol{x}$ due to a force applied in the x_p direction

at point $\boldsymbol{x}'$. The traditional singularity comes from taking the Burgers vector distribution as a delta function. Cai et al. [2] proposed an alternative definition of $G_{kp}(\boldsymbol{x}-\boldsymbol{x}')$ which has a wider isotropic spread mainly localised in a radius a around the dislocation core,

$$G_{ij}(\boldsymbol{x}-\boldsymbol{x}') = \frac{1}{8\pi\mu}\left[\delta_{ij}\partial_{pp} - \frac{1}{2(1-\nu)}\partial_{ij}\right] R_a\,, \tag{3.3}$$

where μ, ν are the isotropic shear modulus and Poisson's ratio respectively, δ_{ij} is the Kronecker Delta, $\partial_{x_1\dots x_n} \equiv \dfrac{\partial^n}{\partial x_1 \dots \partial x_n}$. R_a is the defined as,

$$\begin{aligned} R_a(\boldsymbol{x}) &= R(\boldsymbol{x}) * w(\boldsymbol{x}) = \int R(\boldsymbol{x}-\boldsymbol{x}')w(\boldsymbol{x}')\mathrm{d}^3\boldsymbol{x}' \\ &= \sqrt{R^2+a^2} \qquad (3.4a) \\ w(\boldsymbol{x}) &= \frac{15a^4}{8\pi(R^2+a^2)^{7/2}}\,, \qquad (3.4b) \end{aligned}$$

where $w(\boldsymbol{x})$ is the isotropic Burgers vector distribution derived in the appendix of [2], $\boldsymbol{x} = (x, y, z)$ and $R(\boldsymbol{x}) = \sqrt{x^2+y^2+z^2}$.

For two dislocation nodes (1, 2) connected by straight line segments eq. (3.2) becomes,

$$\begin{aligned} \boldsymbol{\sigma}^{(12)}(\boldsymbol{x}) = &- \frac{\mu}{8\pi}\int_{\boldsymbol{x_1}}^{\boldsymbol{x_2}} \left(\frac{2}{R_a^3} + \frac{3a^2}{R_a^5}\right)\left[(\boldsymbol{R}\times\boldsymbol{b})\otimes\mathrm{d}\boldsymbol{x}' + \mathrm{d}\boldsymbol{x}'\otimes(\boldsymbol{R}\times\boldsymbol{b})\right] \\ &+ \frac{\mu}{4\pi(1-\nu)}\int_{\boldsymbol{x_1}}^{\boldsymbol{x_2}} \left(\frac{1}{R_a^3} + \frac{3a^2}{R_a^5}\right)\left[(\boldsymbol{R}\times\boldsymbol{b})\cdot\mathrm{d}\boldsymbol{x}'\right]\mathbf{I}_2 \\ &- \frac{\mu}{4\pi(1-\nu)}\int_{\boldsymbol{x_1}}^{\boldsymbol{x_2}} \frac{1}{R_a^3}\left[(\boldsymbol{b}\times\mathrm{d}\boldsymbol{x}')\otimes\boldsymbol{R} + \boldsymbol{R}\otimes(\boldsymbol{b}\times\mathrm{d}\boldsymbol{x}')\right] \\ &+ \frac{\mu}{4\pi(1-\nu)}\int_{\boldsymbol{x_1}}^{\boldsymbol{x_2}} \frac{3}{R_a^5}\left[(\boldsymbol{R}\times\boldsymbol{b})\cdot\mathrm{d}\boldsymbol{x}'\right]\boldsymbol{R}\otimes\boldsymbol{R} \end{aligned}\,. \tag{3.5}$$

3.1.1 Resolving Singularities when Dislocation Line Segments are Parallel to Surface Elements

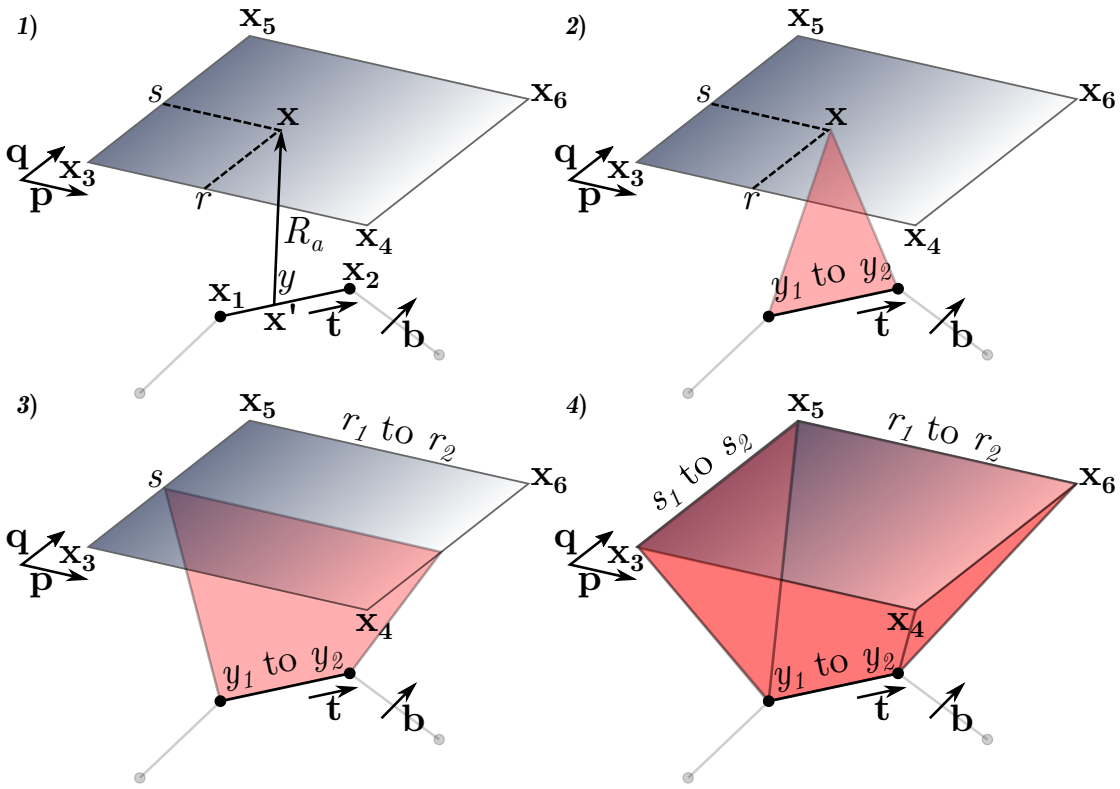


Figure 3.1: Diagram of the line integral method used to find analytical expressions for the forces exerted by dislocations on linear rectangular SEs [1]. *1)* For any given point $\boldsymbol{x}$ on the SE and any given point $\boldsymbol{x}'$ on the dislocation line segment, define distance R_a. *2)* Integrate from $x_1 \to x_2$ along line direction $\boldsymbol{t}$. *3)* Integrate from $r_1 \to r_2$ along vector $\boldsymbol{p}$. *4)* Integrate from $s_1 \to s_2$ along vector $\boldsymbol{q}$.

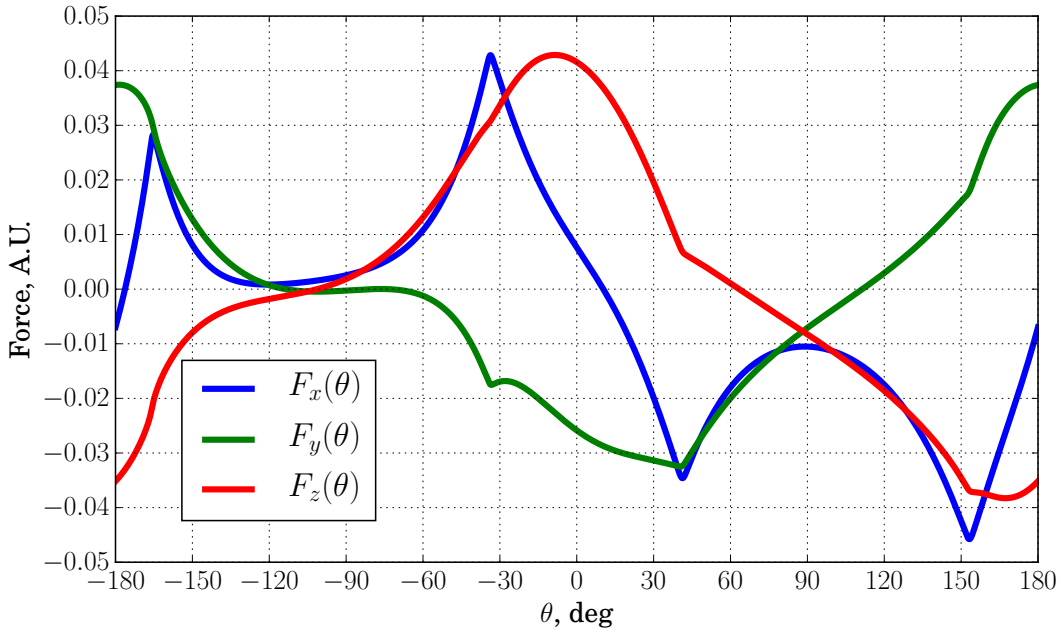


Figure 3.2: Effects of rotating a single dislocation line segment on the forces exerted by it on a linear rectangular SE. The specific values of this function are not known *a priori*, all that is known is that it must be periodic ($T = 2\pi$) and have finite maximum and minimum values. The singularity is avoided by perturbing the angle $\theta = 0 \rightarrow \theta = \pm\epsilon\,, \epsilon \gtrsim 0$.

Chapter 4

Parallelisation of the Analytical Forces Induced by Dislocations on Surface Elements

4.1 Parallelisation on Graphics Processing Units

Efficient parallelisation requires coalesced memory access, which means we have to be extremely careful when mapping CPU memory to device memory. The fact that threads work "simultaneously"[1] means that in order to obtain good performance, data which is to be "simultaneously" loaded into each thread must be contiguous. This maximises cache memory use and therefore reduces slow memory fetch operations to global or shared memory.

The parallelisation was done only over the SEs in order to avoid the undesirable and inefficient GPU branching that would occur under other schemes.

The most natural form of parallelisation is to have blocks of $4n$ threads where $n \leq 8 \in \mathbb{N}$. This also fits nicely into the 32 thread per warp paradigm.

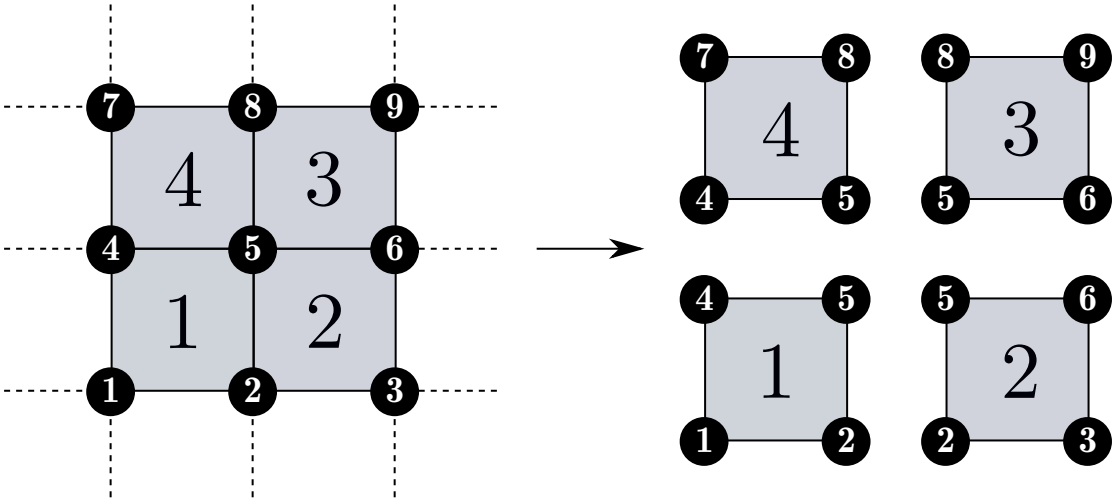


Figure 4.1: Each linear rectangular SE is mapped to one thread.

Algorithm 4.1 Elements in host $\mapsto$ device.

function ELEMENT_HOST_DEVICE_MAP(${}^{\mathrm{h}}\boldsymbol{X}[N][3 \times E]$)
    ▷ Input, temporary, and output indices set to zero.
    $i, j, k \leftarrow 0$
        ▷ ${}^{\mathrm{d}}\boldsymbol{X}$ accomodates all 3 coordinates of all N nodes in all E elements.
    ${}^{\mathrm{d}}\boldsymbol{X} \leftarrow \mathrm{malloc}(3 \times E \times N)$
    for $(n = 0; n < N; n++)$ **do** ▷ Loop over nodes in element.
        ▷ Set output index to point at the n^{th} node of the first coordinate of the first element.
        $k \leftarrow j$
        for $(c = 0; c < 3; c++)$ **do** ▷ Loop over coordinates.
            ▷ Set input index to point at the c^{th} coordinate of the first element of the n^{th} node.
            $i \leftarrow c$
            for $(e = 0; e < E; e++)$ **do** ▷ Loop over elements.
                ${}^{\mathrm{d}}\boldsymbol{X}[k + e] \leftarrow {}^{\mathrm{h}}\boldsymbol{X}[n][i]$
                    ▷ Advance output index to point at the n^{th} node of the c^{th} coordinate of the first element.
                $i \leftarrow i + 3$
            end for
                ▷ Advance output index to point at the n^{th} node of the c^{th} coordinate of the first element.
            $k \leftarrow k + E \times N$
        end for
        ▷ Advance temporary index to point at the $(n + 1)^{\mathrm{th}}$ node of the first coordinate of the first element.
        $j \leftarrow j + E$
    end for
    return ${}^{\mathrm{d}}\boldsymbol{X}$
end function

Algorithm 4.2 Elements in device $\mapsto$ thread.

GPU function ELEMENT_DEVICE_THREAD_MAP(${}^{\mathrm{d}}\boldsymbol{X}[3 \times E \times N]$, ${}^{\mathrm{t}}\boldsymbol{X}[N][3]$)
    ▷ Thread, input indices.
    idx, $i \leftarrow \text{threadIdx}.x + \text{blockIdx}.x \times \text{blockDim}.x$
    for $(c = 0;\ c < 3;\ c++)$ **do** ▷ Loop over elements.
        for $(n = 0;\ n < N;\ n++)$ **do** ▷ Loop over nodes in element.
            ${}^{\mathrm{t}}\boldsymbol{X}[n][c] \leftarrow {}^{\mathrm{d}}\boldsymbol{X}[i + n \times E]$
        end for
        $i \leftarrow i + E \times N$ ▷ Advance input index to the $(n+1)^{\text{th}}$ node of element i.
    end for
    return ${}^{\mathrm{t}}\boldsymbol{X}[n][c]$
end GPU function

Algorithm 4.3 Force in thread $\overset{+}{\mapsto}$ device.

GPU function ADD_FORCE_THREAD_DEVICE(${}^{\mathrm{t}}\boldsymbol{F_n}[N][3]$, ${}^{\mathrm{t}}\boldsymbol{F_e}[3]$, ${}^{\mathrm{d}}\boldsymbol{F_n}[3 \times E \times N]$,
${}^{\mathrm{d}}\boldsymbol{F_e}[3 \times E]$, idx)
    ▷ Nodal force.
    ▷ Set output index to whatever parallelisation index is given to the function. In this case the same index we use for the main parallelisation.
    $i \leftarrow$ idx
    for $(n = 0;\ n < N;\ n++)$ **do** ▷ Loop over nodes.
        for $(c = 0;\ c < 3;\ c++)$ **do** ▷ Loop over coordinates.
            ▷ Ensure nodal forces are correctly added and mapped from local thread memory to global device memory.
            atomicAdd(${}^{\mathrm{d}}\boldsymbol{F_n}[i + c \times E \times N]$, ${}^{\mathrm{t}}\boldsymbol{F_n}[n][c]$)
        end for
            ▷ Displace output index to point at the first coordinate of the $(n+1)^{\text{th}}$ node of the idx$^{\text{th}}$ SE.
        $i \leftarrow i + E$
    end for
    ▷ Total force.
    $i \leftarrow$ idx ▷ Reset output index.
    for $(c = 0;\ c < 3;\ c++)$ **do** ▷ Loop over coordinates.
        atomicAdd(${}^{\mathrm{d}}\boldsymbol{F_e}[i]$, ${}^{\mathrm{t}}\boldsymbol{F_e}[c]$)
        ▷ Advance the output index to point at the $(i+1)^{\text{th}}$ coordinate of the idx$^{\text{th}}$ SE.
        $i \leftarrow i + E$
    end for
    return ${}^{\mathrm{d}}\boldsymbol{F_n}$, ${}^{\mathrm{d}}\boldsymbol{F_e}$
end GPU function

Algorithm 4.4 Force (parallelise over dislocations) in thread $\overset{+}{\mapsto}$ device.

GPU function DLN_ADD_FORCE_THREAD_DEVICE(${}^{\mathrm{t}}\boldsymbol{F_n}[N][3]$, ${}^{\mathrm{t}}\boldsymbol{F_e}[3]$, ${}^{\mathrm{d}}\boldsymbol{F_n}[3 \times E \times N]$, ${}^{\mathrm{d}}\boldsymbol{F_e}[3 \times E]$, k)
    ▷ Nodal force.
    ▷ Set output index to correspond to whatever SE we're on. By parallelising over dislocation lines, we must loop through SEs in the main code, this is the value of k we provide. Multiply by 3 and N because each SE has 3 coordinates and N nodes.
    $i \leftarrow 3 \times N \times k$
    $j \leftarrow 0$ ▷ Auxiliary index.
    for $(n = 0;\ n < N;\ n++)$ **do** ▷ Loop over nodes.
        for $(c = 0;\ c < 3;\ c++)$ **do** ▷ Loop over coordinates.
            ▷ Ensure nodal forces are correctly added and mapped from local thread memory to global device memory.
            atomicAdd(${}^{\mathrm{d}}\boldsymbol{F_n}[i + j + c]$, ${}^{\mathrm{t}}\boldsymbol{F_n}[n][c]$)
        end for
        ▷ Displace auxiliary index to point at the first coordinate of the $(n+1)^{\text{th}}$ node of the k^{th} SE.
        $j \leftarrow j + 3$
    end for
    ▷ Total force.
    ▷ Set auxiliary index to point at the first coordinate of the k^{th} SE.
    $j \leftarrow 3 \times k$
    for $(c = 0;\ c < 3;\ c++)$ **do** ▷ Loop over coordinates.
        atomicAdd(${}^{\mathrm{d}}\boldsymbol{F_e}[j + c]$, ${}^{\mathrm{t}}\boldsymbol{F_e}[c]$)
    end for
    return ${}^{\mathrm{d}}\boldsymbol{F_n}$, ${}^{\mathrm{d}}\boldsymbol{F_e}$
end GPU function

Algorithm 4.5 Nodal force in device $\mapsto$ host.

function FX_DEVICE_HOST_MAP($^{\mathrm{d}}\boldsymbol{F_n}[3 \times E \times N]$, $^{\mathrm{h}}\boldsymbol{F_n}[N][3 \times E]$)
  $i,\ j \leftarrow 0$ ▷ Set input and output indices to zero.
  for $n = 0;\ n < N;\ n++$ **do** ▷ Loop over nodes.
    $j \leftarrow 0$ ▷ Reset output index to point at the first element.
    for $e = 0;\ e < E;\ e++$ **do** ▷ Loop over elements.
      for $c = 0;\ c < 3;\ c++$ **do** ▷ Loop over coordinates.
        $^{\mathrm{h}}\boldsymbol{F_n}[n][j + c] =^{\mathrm{d}} \boldsymbol{F_n}[i + e + c \times E \times N]$
      end for
      ▷ Advance output index to point at the first coordinate of the $(e+1)^{\text{th}}$ element.
      $j \leftarrow j + 3$
    end for
    ▷ Advance input index to point at the first coordinate of the $(n+1)^{\text{th}}$ node of the first element.
    $i \leftarrow i + E$
  end for
  return $^{\mathrm{h}}\boldsymbol{F_n}$
end function

4.1.1 Data Mapping

4.1.1.1 Elements: Host $\mapsto$ Device

4.1.1.2 Elements: Device $\mapsto$ Thread

4.1.1.3 Force: Thread $\overset{+}{\mapsto}$ Device

4.1.1.4 Force (Parallelise over Dislocations): Thread $\overset{+}{\mapsto}$ Device

4.1.1.5 Nodal Force: Device $\mapsto$ Host

4.1.1.6 Total Force: Device $\mapsto$ Host

$$\boldsymbol{X}_{en} \coloneqq [x_{en},\ y_{en},\ z_{en}] \tag{4.1a}$$

$$\boldsymbol{X}_{(1\to E)n} \mapsto \boldsymbol{X}_n$$

$$\boldsymbol{X}_n \coloneqq [x_{1n},\ y_{1n},\ z_{1n}, \ldots, x_{En},\ y_{En},\ z_{En}] \tag{4.1b}$$

$$\boldsymbol{X}_{1\to N} \mapsto \boldsymbol{X}^{\mathrm{SE}}$$

$$\boldsymbol{X}^{\mathrm{SE}} \coloneqq \begin{bmatrix} x_{11},\ \ldots,\ x_{E1},\ x_{12},\ \ldots,\ x_{E2},\ \ldots,\ x_{1N},\ \ldots,\ x_{EN}, \\ y_{11},\ \ldots,\ y_{E1},\ y_{12},\ \ldots,\ y_{E2},\ \ldots,\ y_{1N},\ \ldots,\ y_{EN}, \\ z_{11},\ \ldots,\ z_{E1},\ z_{12},\ \ldots,\ z_{E2},\ \ldots,\ z_{1N},\ \ldots,\ z_{EN} \end{bmatrix} \tag{4.1c}$$

where e is the SE label, n the node label, E the number of SEs in the scope. N

[1] Not quite but essentially simultaneously.

Algorithm 4.6 Total force in device $\mapsto$ host.

function FTOT_DEVICE_HOST_MAP(${}^{\mathrm{d}}\boldsymbol{F_e}[3 \times E]$, ${}^{\mathrm{h}}\boldsymbol{F_e}[3 \times E]$)
  $i \leftarrow 0$ ▷ Set output index to zero.
  for $e = 0$; $e < E$; $e++$ **do** ▷ Loop over elements.
    for $c = 0$; $c < 3$; $c++$ **do** ▷ Loop over coordinates.
      ${}^{\mathrm{h}}\boldsymbol{F_e}[i + c] = {}^{\mathrm{d}}\boldsymbol{F_e}[e + c \times E]$
    end for
    ▷ Advance the output index to point at the first coordinate of the $(e+1)^{\text{th}}$ surface element.
    $i \leftarrow i + 3$
  end for
  return ${}^{\mathrm{h}}\boldsymbol{F_e}$
end function

the number of nodes in a SE.

Data-mapping according to **??** and relabelling the nodes so they go from $0 \rightarrow N-1$, the data from fig. 4.1 would be arranged in device memory like eq. (4.2),

$$\boldsymbol{X}^{\mathrm{SE}} = \Big[\underbrace{h_x,\, i_x,\, f_x,\, g_x}_{\mathbf{x_0}},\, \underbrace{a_x,\, b_x,\, i_x,\, h_x}_{\mathbf{x_1}},\, \underbrace{i_x,\, d_x,\, e_x,\, f_x}_{\mathbf{x_2}},\, \underbrace{b_x,\, c_x,\, d_x,\, i_x}_{\mathbf{x_3}},\, \ldots\ y\text{-coord}\ \ldots,\, \ldots\ z\text{-coord}\ \ldots\Big] \tag{4.2}$$

In the GPU, each thread will cater to one SE at a time. This means that each thread will have to extract the relevant data from the 1D array with length $3 \times E \times N$ into four 1D arrays of length 3. The purpose of CNE mapping is to provide the warp with coalesced memory access. This is achieved via **??**.

Since **??** is performed in a CUDA GPU, threads in a single block execute sequentially from,

$$\text{threadIdx}.x = 0 \rightarrow \text{threadIdx}.x = \text{blockDim}.x - 1\,, \tag{4.3}$$

while threads in different blocks execute in parallel. Coalesced memory access is ensured by having each block load a cache line whose entries are contiguously accessed by the threads in the block. Using the same notation as **??**, full cache line utilisation (optimal cache use) is achieved if cache lines can accomodate l entries

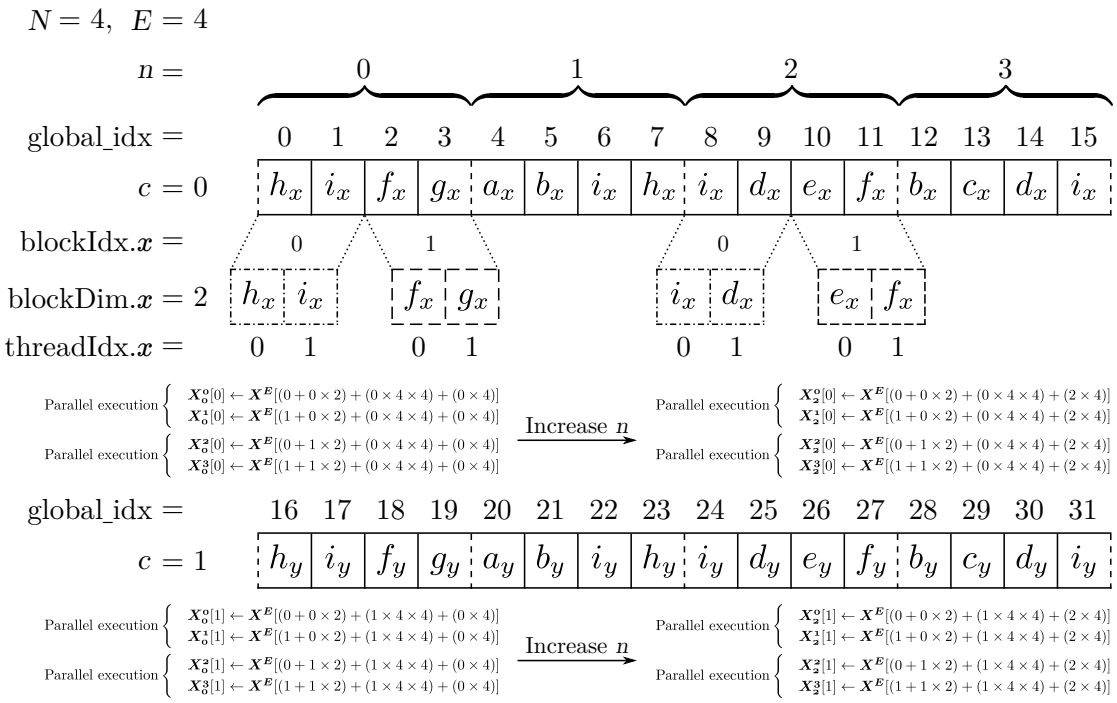


Figure 4.2: Minimum working complete example of the backward coodinate-node-element-map. The explicit calculations of global indices correspond to subsituting the values for "idx" and "idxi" as in **??**. $\boldsymbol{X}_{\boldsymbol{a}}^{\boldsymbol{b}}$ denotes a 1D array of length three containing the xyz-coordinates of node a of element b. Each thread concerns itself with only one element at a time. The dash-dot and dashed boxes represent cache lines for a thread block, the dotted line represents a memory fetch, and the dashed lines in $\boldsymbol{X}^{\boldsymbol{E}}$ represent steps of E entries (the start of the data for the next node of the element type we're dealing with). The memory operations are not shown twice to minimuse redundancy.

given by eq. (4.4),

$$l = \begin{cases} a \times N \times E & , a > 0 \in \mathbb{N} \\ & \text{or} \\ \dfrac{1}{2^a} \times N \times E & , a \geq 0 \in \mathbb{N},\ N \times E \equiv 0 \pmod{2^a}. \end{cases} \tag{4.4}$$

Figure 4.2 shows an example of **??** up to the y-coordinate.

4.1.1.7 Node Coordinate Element Map

The node-coordinate-element (NCE) data mapping in eq. (4.5) is carried out by algorithm 4.7, each thread looks after a given SE.

Algorithm 4.7 NCE data mapping.

```
for all n nodes ∈ surface element do
    for all c coordinates ∈ [x, y, z] do
        for all e surface elements ∈ surface mesh section do
            list.append(data of the n^th node with c^th coordinate of the e^th SE)
        end for
    end for
end for
```

$$\boldsymbol{X}_{en} \coloneqq [x_{en},\, y_{en},\, z_{en}] \tag{4.5a}$$

$$\boldsymbol{X}_{(1\to E)n} \mapsto \boldsymbol{X}_n$$

$$\boldsymbol{X}_n \coloneqq [x_{1n}, \ldots,\, x_{En},\, y_{1n}, \ldots,\, y_{En},\, z_{1n}, \ldots,\, z_{En}] \tag{4.5b}$$

$$\boldsymbol{X}_{1\to N} \mapsto \boldsymbol{X}^{\mathrm{SE}}$$

$$\boldsymbol{X}^{\mathrm{SE}} \coloneqq \begin{matrix} [x_{11}, \ldots,\, x_{E1},\, y_{11},\, \ldots,\, y_{E1},\, z_{11}, \ldots,\, z_{E1} \\ \ldots, \\ x_{1M}, \ldots,\, x_{EN},\, y_{1N}, \ldots,\, y_{EN},\, z_{1N}, \ldots,\, z_{EN}] \end{matrix} \tag{4.5c}$$

where e is the surface **e**lement, n the **n**ode, E the total number of SEs in scope, N the total number of nodes in each SE.

Data-mapping according to algorithm 4.7 and relabelling the nodes so they go from $0 \to N-1$, the data from fig. 4.1 would be arranged in device memory like eq. (4.6),

$$\boldsymbol{X}^{\mathrm{SE}} = \begin{bmatrix} \underbrace{h_x,\, i_x,\, f_x,\, g_x,\, h_y,\, i_y,\, f_y,\, g_y,\, h_z,\, i_z,\, f_z,\, g_z}_{\mathbf{x_0}}\,, \\ \ldots\ \mathbf{x_1},\, xyz\text{-coords}\ \ldots, \\ \ldots\ \mathbf{x_2},\, xyz\text{-coords}\ \ldots, \\ \ldots\ \mathbf{x_3},\, xyz\text{-coords}\ \ldots] \end{bmatrix} \tag{4.6}$$

4.1.1.8 Resolving Data Write Conflicts

Data write conflicts can be a problem in parallel applications where global data is changed by multiple threads within the same clock tick. This can be avoided with atomic operations.

4.1.2 Resolving Parallel Dislocation Line Segments to Surface Elements

Dealing with the special case when the dislocation line segment $\boldsymbol{t}$ is parallel to the SE is relatively trivial when in a serial program. By the mean value theorem, we

can slightly perturb $\boldsymbol{t}$ by rotating it by a small angle around the midpoint of $\boldsymbol{t}$ with respect to the axis of rotation defined by $\boldsymbol{t} \times \boldsymbol{n}$. In contrast to serial code, program branches can have a serious effect on parallel performance due to warp divergence. Due to the complexity of the force calculation and relative rarity of the edge case, there is no branching behaviour in the parallel code until *after* the calculation is performed. Where algorithm 4.8 is performed.

Algorithm 4.8 Resolving cases when $\boldsymbol{t} \parallel \boldsymbol{n}$ on GPUs.

for all surface elements *and* line segments **do**
    ...
    if $\boldsymbol{t}_{\text{thread}} \parallel \boldsymbol{n}_{\text{thread}}$ **then**
        $\boldsymbol{x}_{\text{buffer}} \leftarrow \boldsymbol{x}_{\text{thread}}$
        $\boldsymbol{t}_{\text{buffer}} \leftarrow \boldsymbol{t}_{\text{thread}}$
    else
        $\boldsymbol{F}_{\text{total}} += \boldsymbol{F}_{\text{thread}}$
    end if
end for
return to serial code
if $\boldsymbol{x}_{\text{buffer}} \neq$ empty **then**
    for all surface elements *and* line segments **do**
        perform calculation for $\boldsymbol{t} \parallel \boldsymbol{n}$
    end for
end if

4.1.2.1 Implementation

The node mapping would be the same as **??**. The parallelisation would be on individual packets of a SE with a slightly rotated dislocation line segment, as in a single iteration of the serial code where the dislocation line segment is rotated. The forces would be averaged with atomic operations and `__syncthreads()` before they are added to the total nodal forces.

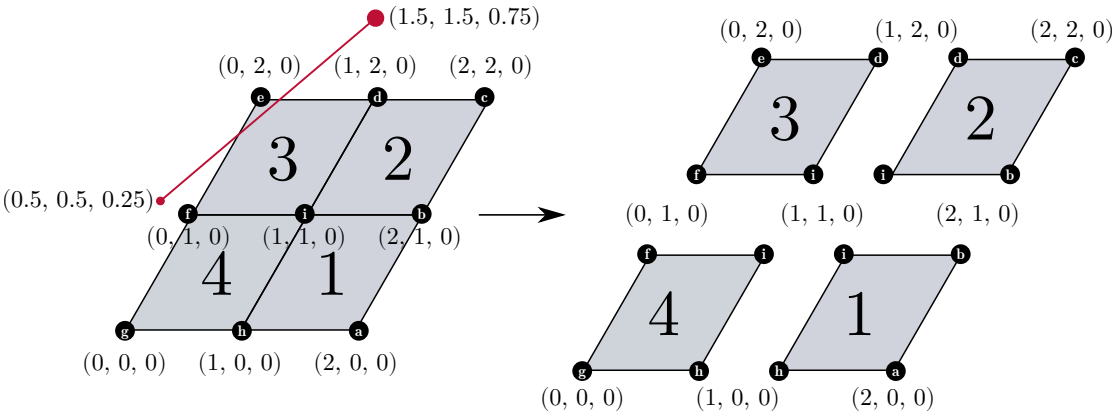


Figure 4.3: Test case for CUDA development. Dislocation line not shown in node separation as the parallelisation happens across surface elements. Not to perspective or scale.

4.1.3 Test Case

According to **??** and the node labelling scheme of fig. 3.1 the data mapping would look like:

$$\boldsymbol{X}^{\mathrm{SE}} = \left[\begin{array}{c} \overbrace{\underbrace{1,1,0,0}_{\mathbf{x_3}},\underbrace{2,2,1,1}_{\mathbf{x_4}},\underbrace{1,1,0,0}_{\mathbf{x_5}},\underbrace{2,2,1,1}_{\mathbf{x_6}}}^{x\text{-coord}}, \\ \overbrace{\underbrace{0,1,1,0}_{\mathbf{x_3}},\underbrace{0,1,1,0}_{\mathbf{x_4}},\underbrace{1,2,2,1}_{\mathbf{x_5}},\underbrace{1,2,2,1}_{\mathbf{x_6}}}^{y\text{-coord}}, \\ \overbrace{\underbrace{0,0,0,0}_{\mathbf{x_3}},\underbrace{0,0,0,0}_{\mathbf{x_4}},\underbrace{0,0,0,0}_{\mathbf{x_5}},\underbrace{0,0,0,0}_{\mathbf{x_6}}}^{z\text{-coord}} \end{array}\right] \tag{4.7}$$

In general the number of indices in the array would be $3 \times e \times n$ where e is the number of surface elements, n the number of nodes per element and 3 the number of spatial dimensions.

4.2 Thread Block Size Optimisation

http://docs.nvidia.com/cuda/cuda-c-best-practices-guide/index.html#memory-optimizations
http://docs.nvidia.com/cuda/cuda-c-best-practices-guide/index.html#execution-configuration-optimizations

Bibliography

[1] S Queyreau, J Marian, BD Wirth, and A Arsenlis. Analytical integration of the forces induced by dislocations on a surface element. *Modelling and Simulation in Materials Science and Engineering*, 22(3):035004, 2014.

[2] Wei Cai, Athanasios Arsenlis, Christopher R Weinberger, and Vasily V Bulatov. A non-singular continuum theory of dislocations. *Journal of the Mechanics and Physics of Solids*, 54(3):561–587, 2006.

[3] Toshio Mura. *Micromechanics of defects in solids*. Springer Science & Business Media, 2013.

Appendix A

Best Practices

This chapter defines the best practices that will make improve the development and testing pipeline by defining rules and standards that facilitate collaboration.

A.1 Filesystem

All files must be within `/project_name`.

1. All files must be within `/project_name`,
   `~/` = `project_name/`,
   `~/~/` = `project_name/src/` or `project_name/dev/`.
2. Release code must be within `~/src/`.
3. Development code must be within `~/dev/`.
4. Test data must be within `~/tests/`.
5. Documentation must be within `~/~/doc/`.
6. Examples must be within `~/~/exmp/`.
7. External libraries must be within `~/lib/`.
8. Generated images must be within `~/~/images/`.
9. Old versions recordkeeping must be within `~/prv/va.b.c/`.

A.2 Versioning

1. Use a version control system like GitHub or PasteBin.

2. There must be a master branch that is only changed when the code is stable and bug free.

3. Development branches should be exploited as seen fit without making things overly convoluted.

4. Commits must be as bug free and regular as possible. When to commit is left to the developer's discresion.

5. Commit messages should be as descriptive as possible.

6. Versions should be specified as `va.b.c` where `a`, `b`, `c` = integers. The three levels are `a` = release version (usable, bug free code), `b` = beta version (code that is undergoing testing), `c` = alpha version (code that is under active development).

A.3 Documentation

A.3.1 Commenting

Every codefile must be appropriately commented by meeting the following guidelines.

1. The start of each codefile must have a heading detailing the creator, date of creation and edit history (date and name of editor).

2. Below the heading there must be a general explanation of the code. It must state any procedures, structures, objects and how they are to be utilised. Any backward or forward dependencies must be stated.

3. Below the description and edit history any relevant literature must be mentioned (dois are preferred). Must be as detailed as possible, include equation numbers/ranges if necessary.

4. The start of every procedure has an explanation of its purpose, inputs, outputs and inputs-outputs.

5. Particularly complicated code blocks must have an in-depth explanation of what it does. Comment each line if necessary.

6. Corrections or additions must be explicitly bounded by comments at the start and end of the change. Both bounding comments must have the author's name and the date. Below the starting comment, there should be an explanation of the change. Any punctual comments can be made as normal.

A.3.2 README

Every codefile must have an associated README `.tex` document that documents the codefile's contents. It must meet the following guidelines as appropriate.

1. The name must be that of the file it documents (minus the extension of course).
2. Description and overall explanation of the codefile's purpose.
3. Overall flow chart or pseudo code describing the file's purpose.
4. Document the codefile's procedures. This means describing and explaining their corresponding inputs, outputs, inputs-outputs, forwards and backwards dependencies, and flow charts or pseudo codes.
5. Unit test designs and results for each procedure. If appropriate also include those of integral tests.

The codefile may also be appended at the end of documentation if desired (the `minted` package is highly recommended).

A.4 Modularisation

1. Code repetition must be kept to a *strict* minimum. Any piece of code that will be reused must be modularised.
2. Procedures must be as self-sufficient as possible *without* repeating code. If repeating code is necessary, replace it with a procedure that is to be repeatedly called instead. Minimising repeated code $\ggg$ procedure self-sufficiency.

A.5 Coding Style

All names must meet the following guidelines.

1. Indent appropriately. Four space tabs are a good compromise between code necking and readability.
2. Minimise the use of nested code blocks, use intrisics, libraries or create procedures instead.
3. Break up lines that are uncomfortably long, typically anything over 80–100 characters.

4. Names should be appropriately descriptive and human readable.

5. All code and names must be systematic and logical.

6. the use of upper cases should be reserved for parameters (`const` variables in C).

7. Delimit words with "`_`" *not* case changes.

8. Long and descriptive $\ggg$ short and cryptic.

A.5.1 Filenames

1. If old versions are to be kept, the old *stable* versions of file must have the date of last modification appended *suffixed* after the file extension in the `_yyyymmdd` format. For example, if the stable version of the release code `hello_world_parallel.c` was last modified on April 25, 2017 it should be archived as
`hello_world_parallel.c_20170425`.

A.5.2 Variables, Structures and Objects

1. Only counters and indices can be single letter variables.

2. Structure and object *definitions* are *suffixed* with `_s` and `_o` respectively.

3. Inputs, outputs and input-outputs to procedures must be *prefixed* with `i_`, `o_` and `io_` respectively.

A.5.3 Procedures

1. Functions and subroutines must be *prefixed* with `f_` and `s_` respectively.

Appendix B

Coupling Discrete Dislocation Dynamics to Finite Element Methods

Appendix C

Implementation of Analytical Forces Induced by Dislocations on Linear Rectangular Surface Elements

C.1 Serial C Code MEX File

C.2 Parallel CUDA C Code MEX File

Appendix D

Talks

D.1 Durham July 12–14 2017

Bridging the gap between the microscopic and macroscopic world is an on-going challenge for science and technology. If we hope to understand complex emergent phenomena we need to study systems that blur the line between micro and macro. In materials science, one such area is the study of extended defects called dislocations; whose nucleation and movement mediate the permanent deformation of materials. These large, high energy defects present very complex and long ranged interactions with each other, crystal boundaries, impurities, free surfaces, and themselves. As such, their dynamics are difficult to study experimentally. So we make justified assumptions and simplifications and create models that let us study them in detail. However if our models are to prove useful in real applications, they must be continually refined and improved by weakening assumptions and removing simplifications. Unfortunately, with increased refinement comes increased computational cost and new challenges. Therefore, finding faster alternatives that do not sacrifice accuracy is of the utmost importance—better yet if the alternatives are more accurate or exact. With the advent of increasingly accessible graphics processor units (GPUs) typically used in video gaming, the power of parallel processing is no longer exclusive to researchers with access to supercomputers. In this talk we will discuss the GPU implementation of exact solutions for the forces dislocations exert on the surfaces of materials.

Index