

第8回
葛ユニット
ワークショップ

参加者ワークシートへの回答例
2023.8.7-8, 青木学聡(名古屋大学)

自己紹介

- 氏名
青木 学聡 (あおき たかあき)
- 所属
名古屋大学
- 研究分野・職種
情報戦略室室長 - 大学全体の情報基盤整備統括
(10年程前まで)ナノ加工・計測技術の開発(特に大規模MDシミュレーション)
- 参加動機
これからの大学の研究者が持つべき「マネジメントスキル」を定義し、普及させること
その中での、情報基盤の位置づけ

ループリック作業を 始める前に

当たり前だと思っていた「研究ルーチン」を見直してみよう

あなたの研究ルーチンを記述してみよう

planning

organizing

analyzing

sharing or
publishing

データを生成
・入手する

生成・入手したデータ
に名前を付け整理する

データを分析し
結果を得る

レポートを書く

- どのようなデータ?
- どのように生成・入手する?
- etc. . . .

- 入手データはどこに保存する
- データ命名方法
- 分類方法
- etc. . . .

- 分析方法
- 使用する機材
- 得られるデータ
- etc. . . .

- 誰と結果を共有する?
- 次の研究ルーチンに向けて何をする?
- etc. . . .

私の研究を一言で . . .

- 2020.4～ 名古屋大 情報連携推進本部
- ～2020.3 京大 情報環境機構、工学研究科附属情報センター
- 2000年前後～2015年頃: クラスタイオンビームを応用するナノ科学

原子集団を壊さず
そのままぶつける



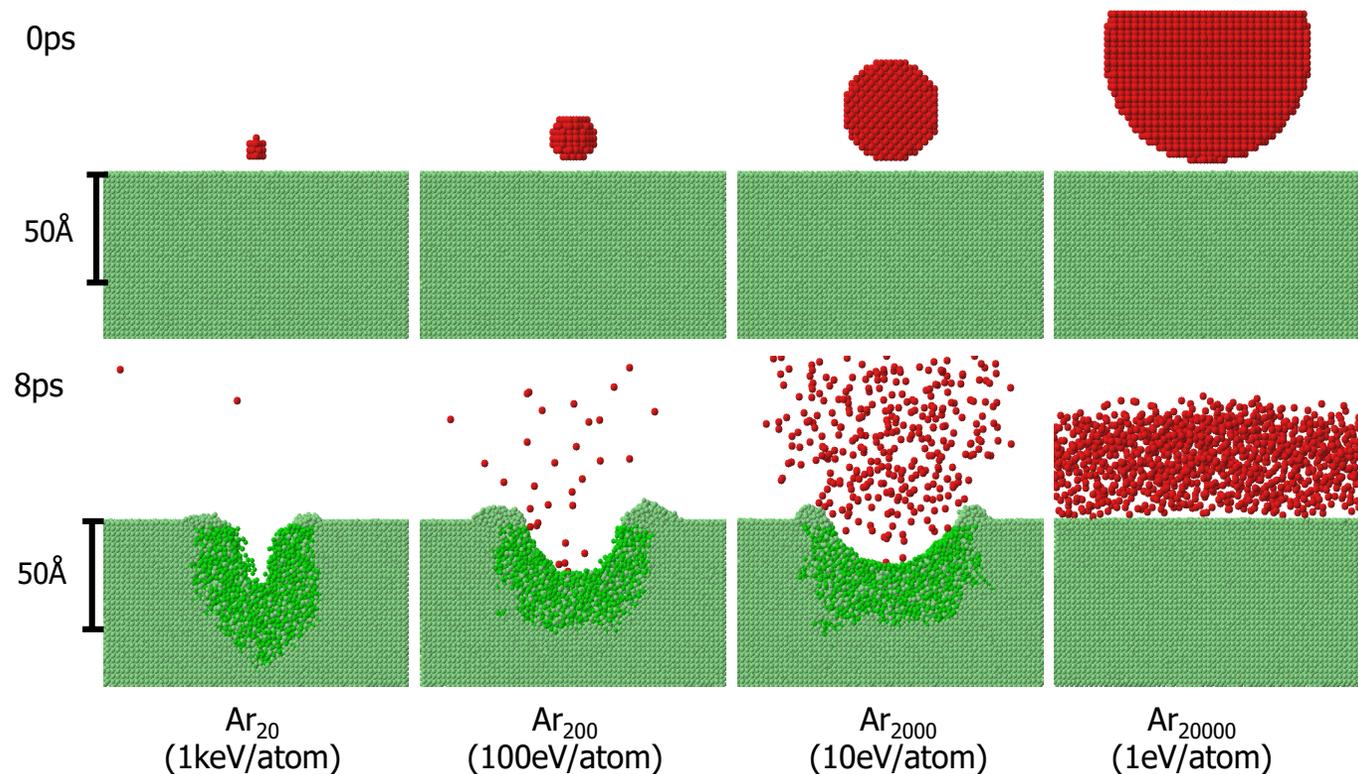
衝突現象そのものが面白い
(塊の大きさという新しいパラメータ)



材料表面の性質を
変える新しいツール

Ar_n 20keV \rightarrow Si(100)

T.Aoki et al. / Nucl. Instr. and Meth. B202 (2003) 278–282



Planning Data

- どのようなデータ?

分子動力学シミュレーションによる、原子衝突過程の時空間データ

- どのように生成・入手する?

- 何を使って?

LAMMPSシミュレータでシミュレーション(以前は、自作ソフトウェア)

研究室でのPCクラスタ、または大学のスパコンで実行

- どんな準備をして?

MD固有の設定パラメータ(時間刻みと繰り返し数)

シミュレーション直前の状態作成(衝突させる原子(1-100万原子)と衝突を受けるターゲット(10万原子-10億原子))

Organizing Data

- 入手データはどこに保存する
研究室のストレージに保存(当時は合計20TiBまで拡大)
- データ命名方法、分類方法
 - シミュレーション条件ごとに命名、フォルダ分類
Ex Ar1000 10keV を Si2000000 のターゲットに衝突させる
Ar1000_10keV_Si2M/
 - 複数のパラメータを変えた統計を取る場合、枝番や詳細情報も可能な範囲でフォルダ名に記載する
ex. 衝突角度を 90度(垂直), 80度、70度、60度・・・ と変えるなら
Ar1000_10keV_Si2M_90deg/

Analyzing Data

- 分析方法・得られるデータ
 - MDから得られるのは、「その瞬間における原子の位置と速度」なので、意味のある情報を取り出す必要がある
 - 「入射原子の軌跡(位置、速度、エネルギー)」「衝突を受けた標的に表面形状」 etc.
- 分析ツール
 - 自作の分析ツール(python, C++)を開発
 - 個人のgithub リポジトリで基本ツール(cf. 生データからのフィルタ、基本統計情報、スナップショット描画)を管理
 - 解析データは、「目的毎に」元データと一緒に(または元データへのリンク)に保存
 - 晩年は Matplotlib のデータ描画までをセットに
 - 以前は Origin (データ描画ツール) や excel との併用

Sharing or Publishing Data

- 誰と結果を共有する？
データを収めるストレージは研究室LAN内で共有可能
(実は、青木以外にこのデータを分析できる、協業できる人がいなかった。
また、研究室外でのデータ共同利用の実績もなかった)
- 次の研究ルーチンに向けて何をする？
 - データをもとに論文を執筆する場合は、容量の許す範囲で、元データも論文執筆フォルダにコピーしておく。
 - 研究室内のレポートの場合、コピーではなく参照先の明記で済ませることもある。
 - (データ参照先の記載)コピーできない場合(1M個の原子座標データは1G、時系列を1000点収集すると1Tになる)、参照先データの所在をメモ

おおよそ書いたら 「ループリック」 で深掘りしよう

上に書いた研究ルーチンで
書けたこと、書けなかったことを整理する

以下、ループリックでの
自己評価・記載事項

Planning For Data

Data to be obtained

評価: Developing

- シミュレーションの元データと元データからの解析結果を取得する方針を説明している

Planning For Data Methodology for Obtaining Data

評価: Developing

- データの生成、解析方法、保存場所を説明できている

Planning For Data

Materials, Equipment and Software for Data

Aquisition

評価: Development

データ生成、解析のためのソフトウェアを用意し、開発できる体制が整備されていました

Planning For Data Team Members

評価: Completed

データ生成、分析とともに、その結果を評価する体制が研究室で
成立していました。
(既に終了、論文化された研究なので当然?)

Planning For Data Funding for Data Acquisition

評価: NA

Planning For Data Data re-use

評価: beginning

研究室内のファイルサーバにデータを保管していたが、それ以上の利用・再利用は想定していなかった

Planning For Data

Open versus closed data

評価: NA

Planning For Data

Planning for publishing and Sharing Data

評価: NA

Organizing Data Metadata

評価: Developing

シミュレーションソフトウェアからの出力データをもとに必要な分析ができるよう、データフォーマット等を説明・理解できる資料があります。

(汎用的な機械可読形式への対応は部分的)

Organizing Data Data Collection

評価: Enhancing

一人での研究とはいえ、シミュレーション条件や実施する分析方法が整理されています。

Organizing Data Materials Acquisition

評価: NA

Organizing Data Digitalization

評価: Beginning

シミュレーションの実施なので、データはすべてデジタル化されていることが前提

Organizing Data

Data Storage

評価: Developing

実施するシミュレーションの規模に応じ、研究室ファイルサーバの容量を拡張しました。

Organizing Data Storage of Physical Materials

評価: NA

Organizing Data File Format

評価: Beginning

MDシミュレーションソフトウェアは広く普及しているものを利用したため、データ形式は共通

(ただし、古い自作MDからの解析ツールを引き継ぐ関係から、解析用には独自のデータ形式に変換している)。

Organizing Data Version Control

評価: NA

Organizing Data Backup

評価: Beginning

データの容量を考慮するとバックアップ用のストレージを持つことは非現実的と判断した

Analyzing Data Software for Analysis

評価: Developing

分析用のソフトウェアは自作し、効率的な研究が進められる状態である

Analyzing Data

Data Processing and Analysis

評価: Developing

MD結果からの「生データ」から新たな知見を見出すためのツール・ライブラリ群が整備されています。

Analyzing Data Intermediate Data

評価: enhancing

生データからの「抽出」「統計処理」「グラフ描画」という段階を設計し、それぞれの中間データを残す、さらにmakeコマンドで再利用できるように配慮している。

Analyzing Data

Data Cleaning

評価: NA

Sharing or Publishing Data Licensing

評価: NA

Sharing or Publishing Data Metadata Schema

評価: NA

Sharing or Publishing Data Collection of Materials into Datasets

評価: NA

Sharing or Publishing Data Documentation

評価: NA

Sharing or Publishing Data

Data Completeness, validity, reliability, leganity,
and ethics

評価: NA

Sharing or Publishing Data

Preserving, Returning, or Disposing of Research
Materials

評価: NA

Sharing or Publishing Data

Publishing and Sharing Data

評価: NA