Reproducible Workflows for Mass Spectrum Analysis in Atom Probe Tomography

Clara Tan

Supervised by Prof. Simon Ringer
& Dr. Anna Ceguerra

THE UNIVERSITY OF SYDNEY

ammrf
Australian Microscopy
& Microanalysis Research Facility

The scientific method

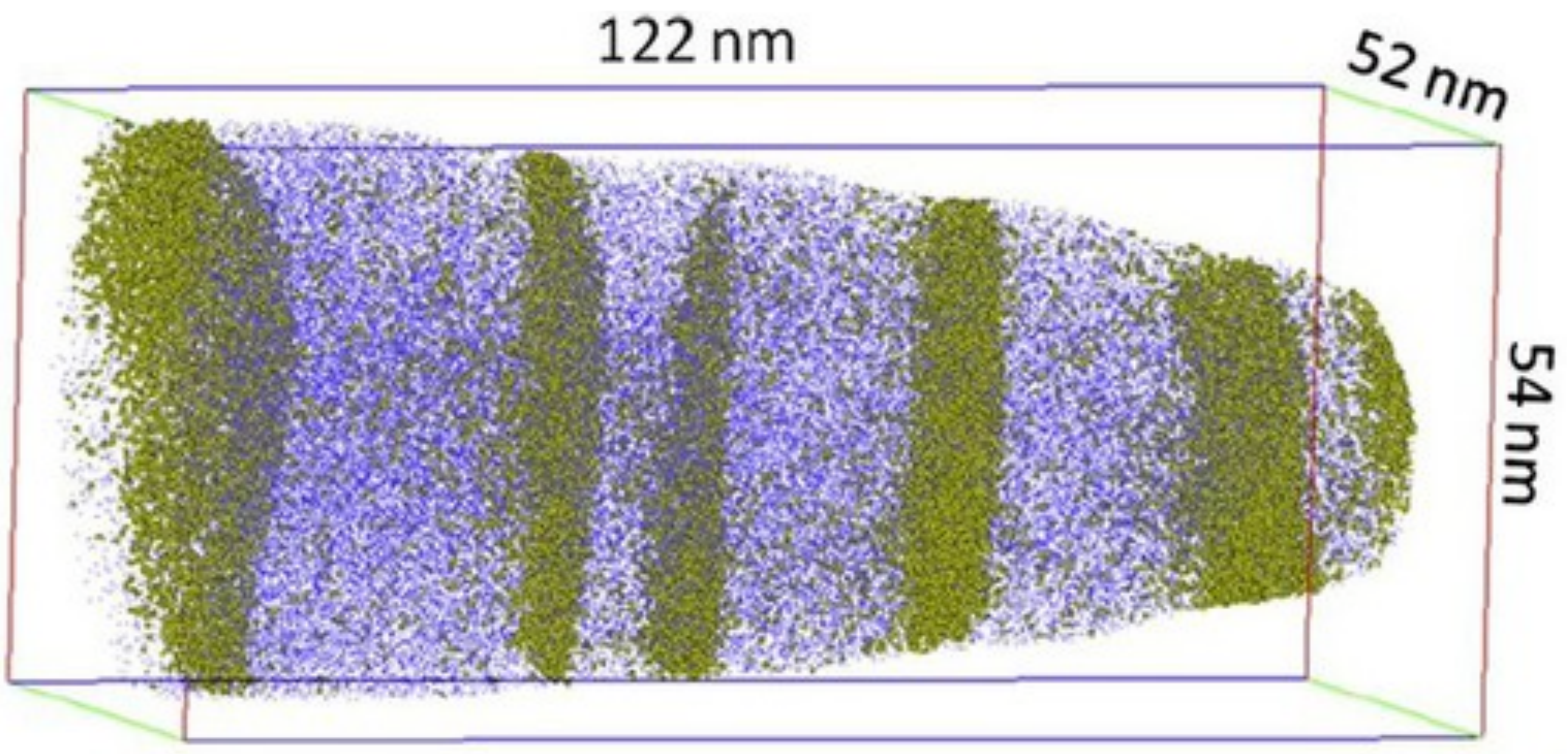
question → hypothesis → test

- *systematic pursuit of testable explanations*
- repeatable, reproducible and peer-reviewed
- objective and empirical

Reproducibility in computation

- methods *not* obvious from article’s text
- difficult to verify, reproduce
- most journals do not require source code

Atom probe tomography

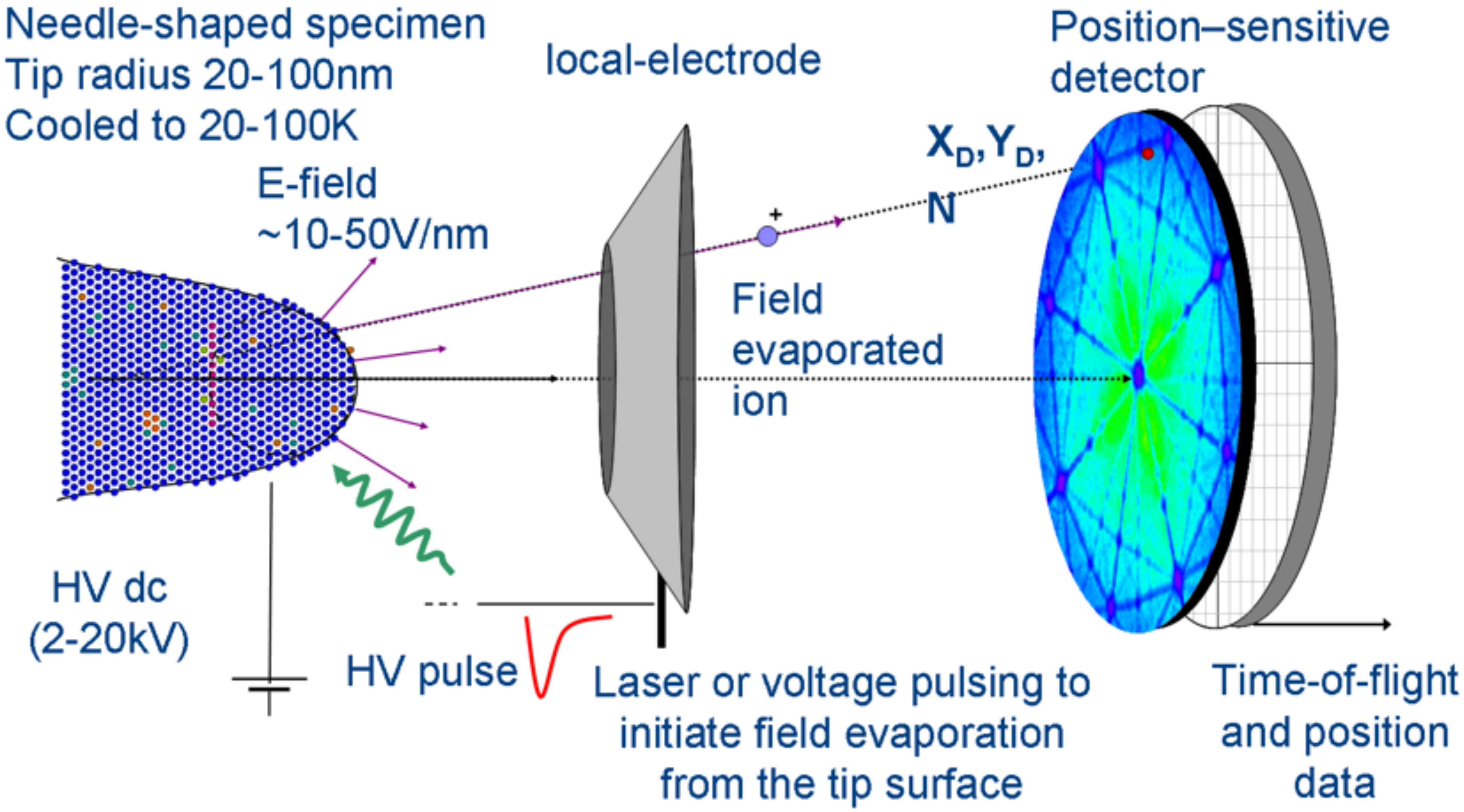


● C

● Fe

Li, et al. 2011. ‘Atomic-scale mechanisms of deformation-induced cementite decomposition in pearlite.’

Experiment method



Needle-shaped specimen
Tip radius 20-100nm
Cooled to 20-100K

E-field
~10-50V/nm

local-electrode

Position–sensitive detector

X_D, Y_D, N

Field evaporated ion

HV dc
(2-20kV)

HV pulse

Laser or voltage pulsing to initiate field evaporation from the tip surface

Time-of-flight and position data

http://atomprobe.materials.ox.ac.uk/uploads/images/APT.png

Chemical identity assignment

time-of-flight → mass-to-charge

(s) *(m/z or Da)*

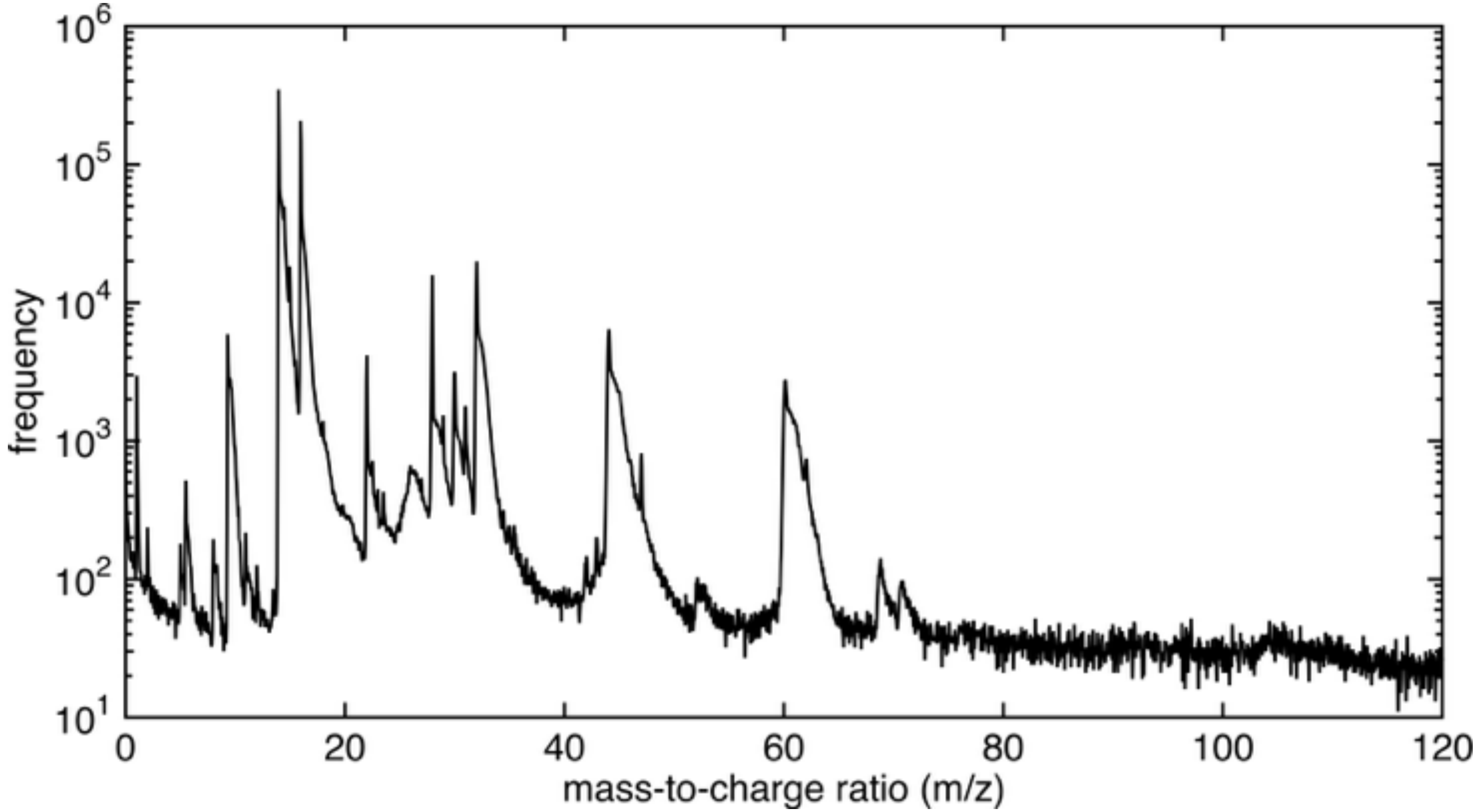
every element has a unique mass (amu)

two elements can have the same m/z

Si $28\ amu / 2^{+} = 14\ m/z$

N $14\ amu / 1^{+} = 14\ m/z$

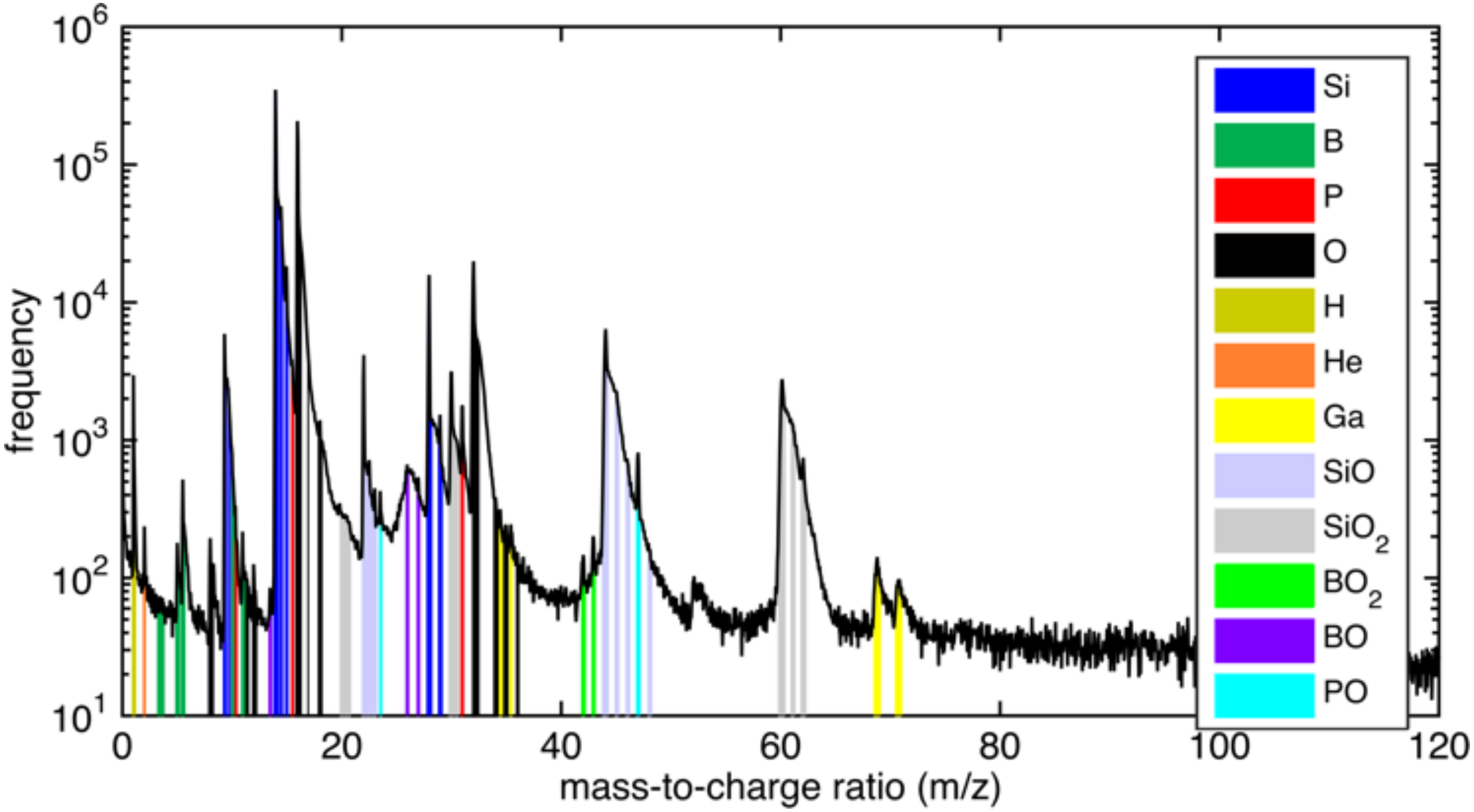
Mass spectrum analysis



Courtesy of Keita Nomoto

peaks *background* *noise*

Mass spectrum analysis



Courtesy of Keita Nomoto

identify peaks *background label* *range noise*

Background Problem Proposal Results Conclusion

At the moment, every user assigns identities and ranges their own way.

- ‘Blue Mountains Study’ - Hudson et al. 2011
    - 20 users
    - Same dataset
    - Mean deviation of 10.74%

It is difficult to know what the user has done to achieve the results.

- Procedure is manual and ‘hidden’ in software

- Not feasible to describe all steps
  - many steps
  - not easy to quantify or explain
  - ad hoc and heuristic approach

Problem

1. Analysis is not reproducible
2. Analysis is not transparent

Aim

To enable transparent and reproducible mass spectrum analysis in atom probe tomography.

Proposal: New graphical user interface (GUI) for a new method-based approach

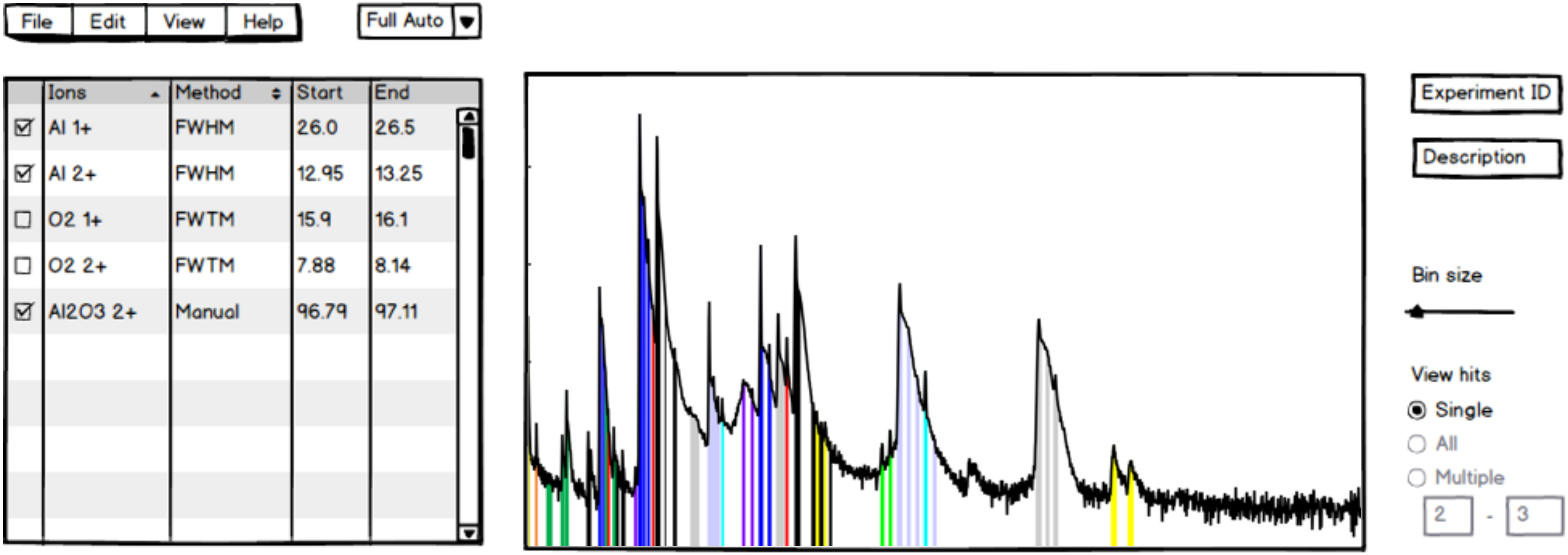
Thesis:

- Enable transparent and reproducible analysis
- Facilitate automated and manual methods (long-term - proliferate auto)
- Focus on usability and sustainable code

Method

- Python & PyQt
- User interviews
- Agile development methodology
- Model-view-controller
- ‘Clean Code’, PEP Python Style Guide

GUI design



	Ions	Method	Start	End
☑	Al 1+	FWHM	26.0	26.5
☑	Al 2+	FWHM	12.95	13.25
□	O2 1+	FWTM	15.9	16.1
□	O2 2+	FWTM	7.88	8.14
☑	Al2O3 2+	Manual	96.79	97.11

Bin size

View hits

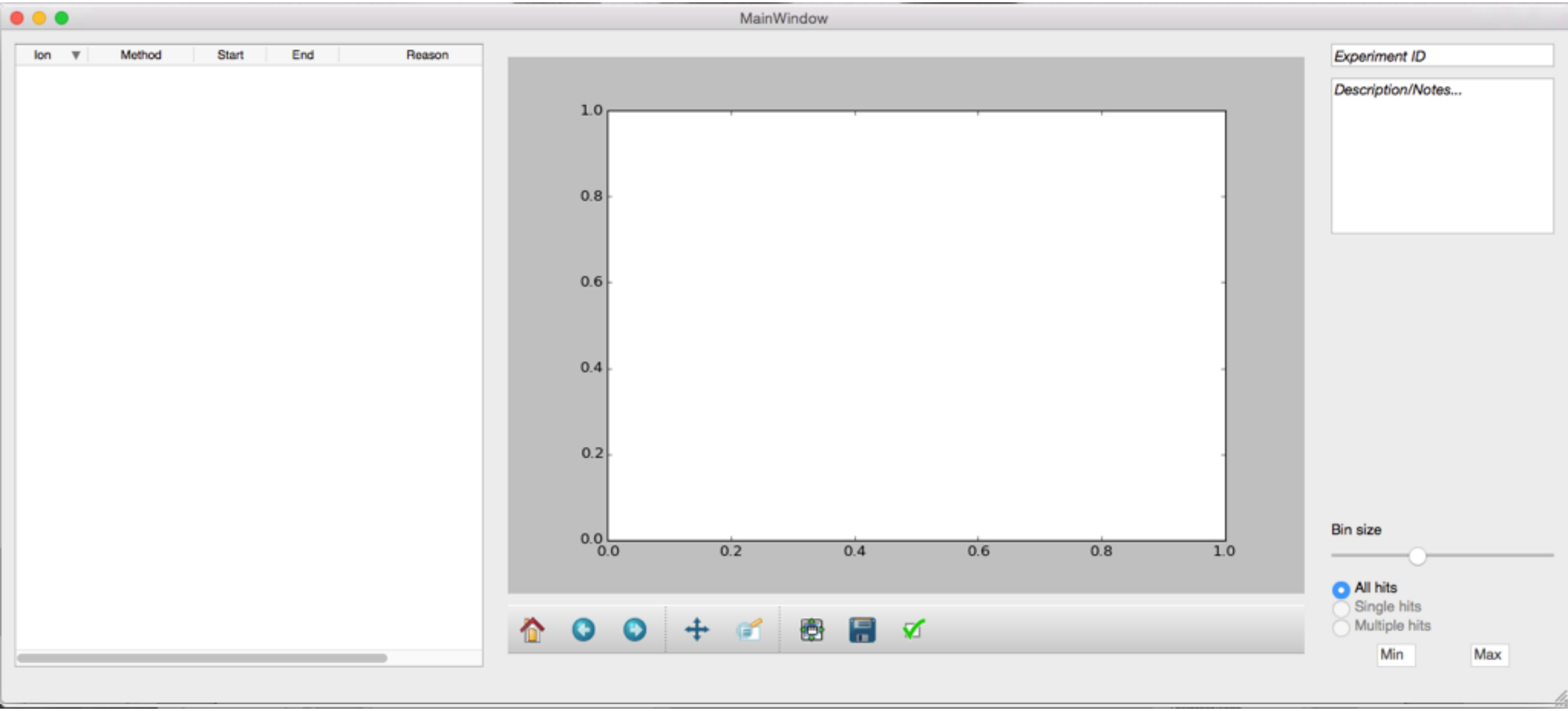
● Single

○ All

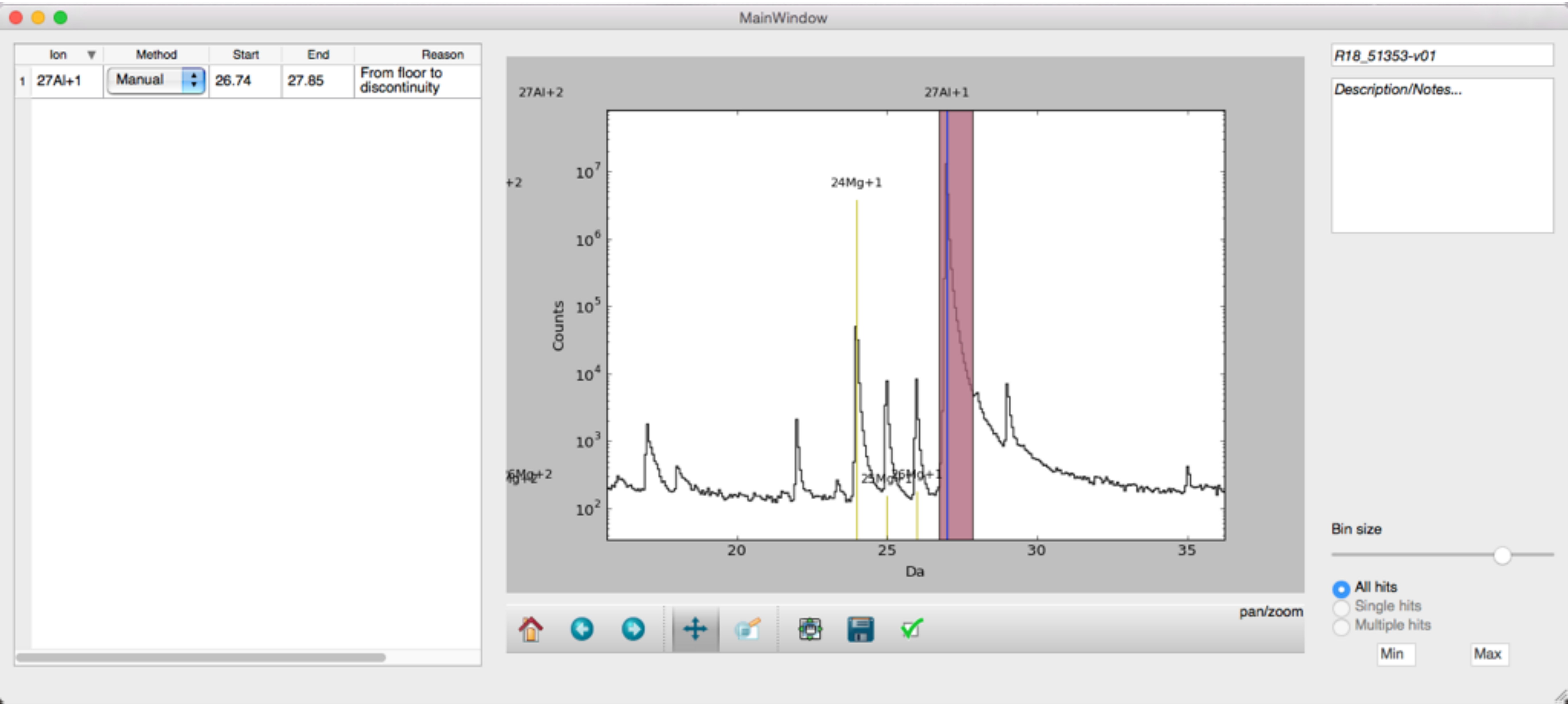
○ Multiple

2 - 3

Normal functions



Normal functions



	Ion	Method	Start	End	Reason
1	27Al+1	Manual	26.74	27.85	From floor to discontinuity

R18_51353-v01

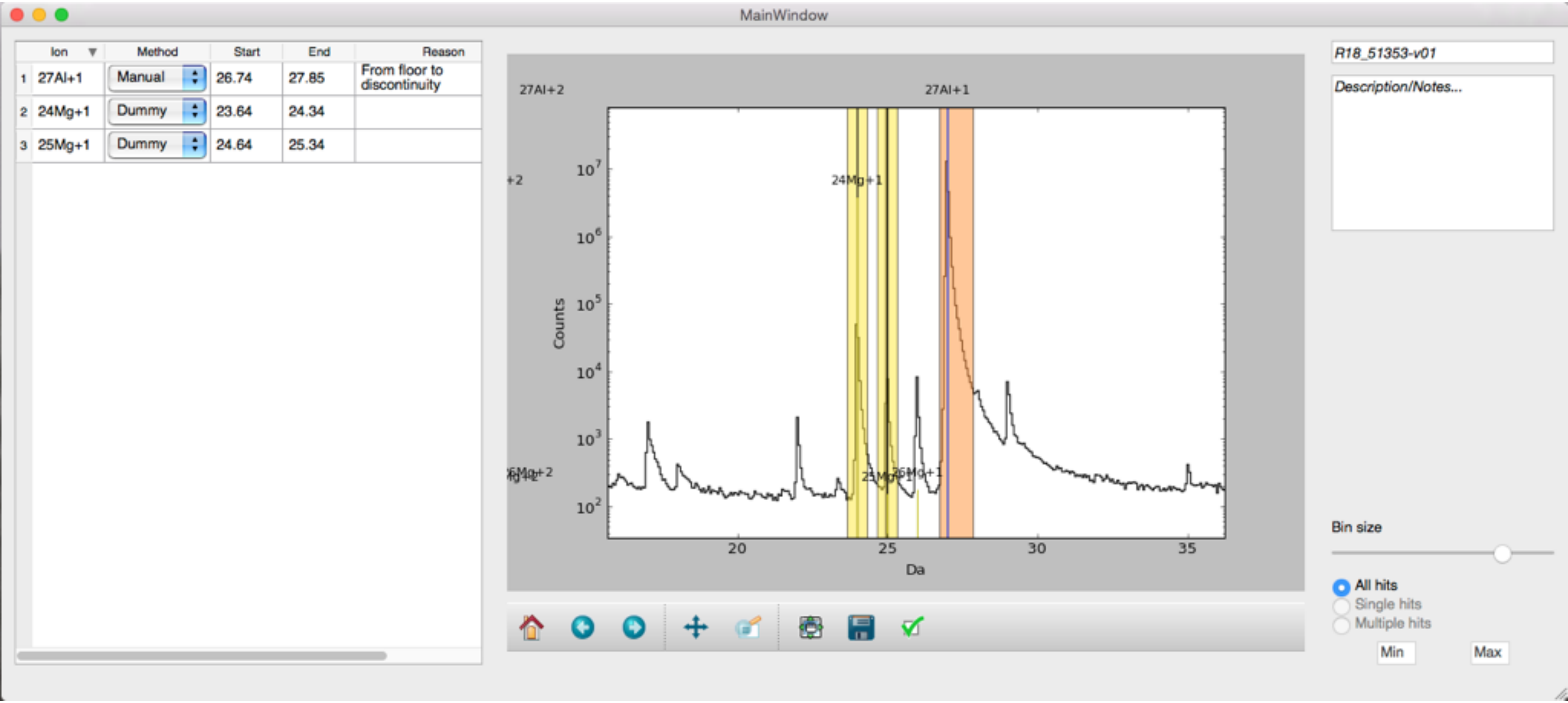
Description/Notes...

Bin size

● All hits
○ Single hits
○ Multiple hits

Min Max

Automatic range methods



MainWindow

	Ion	Method	Start	End	Reason
1	27Al+1	Manual	26.74	27.85	From floor to discontinuity
2	24Mg+1	Dummy	23.64	24.34	
3	25Mg+1	Dummy	24.64	25.34	

R18_51353-v01

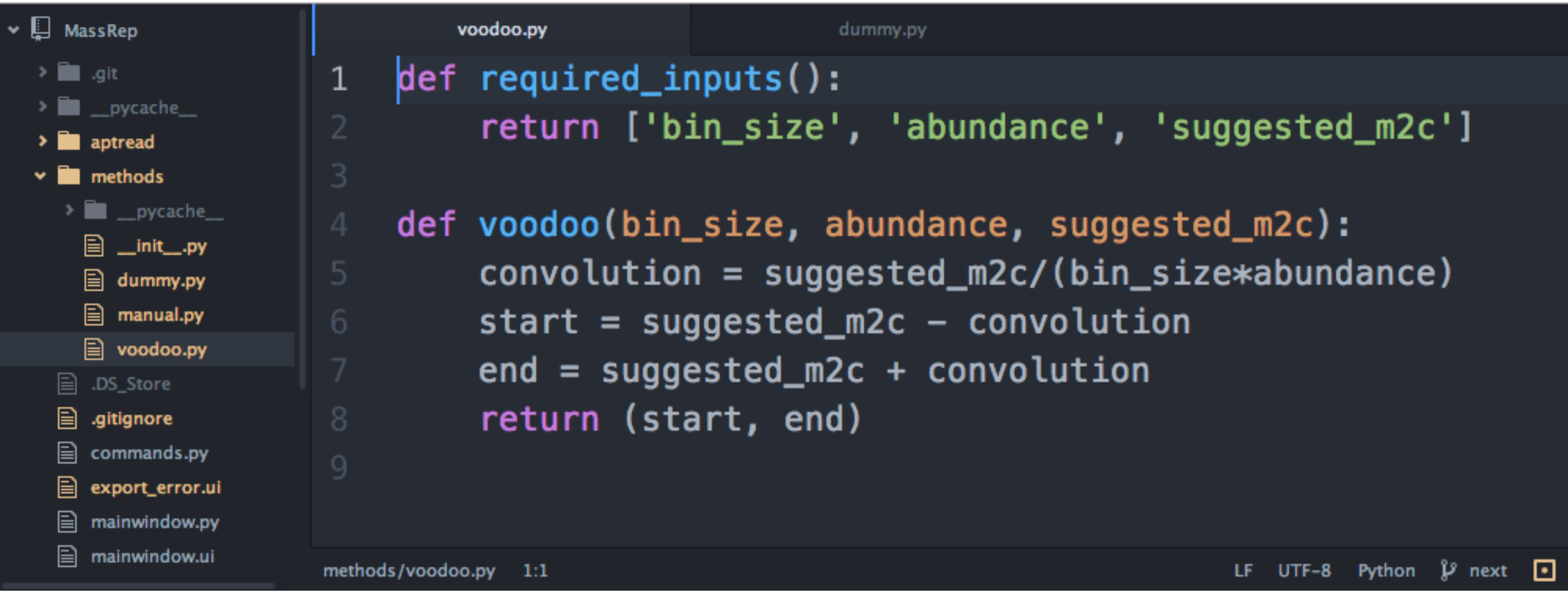
Description/Notes...

Bin size

● All hits
○ Single hits
○ Multiple hits

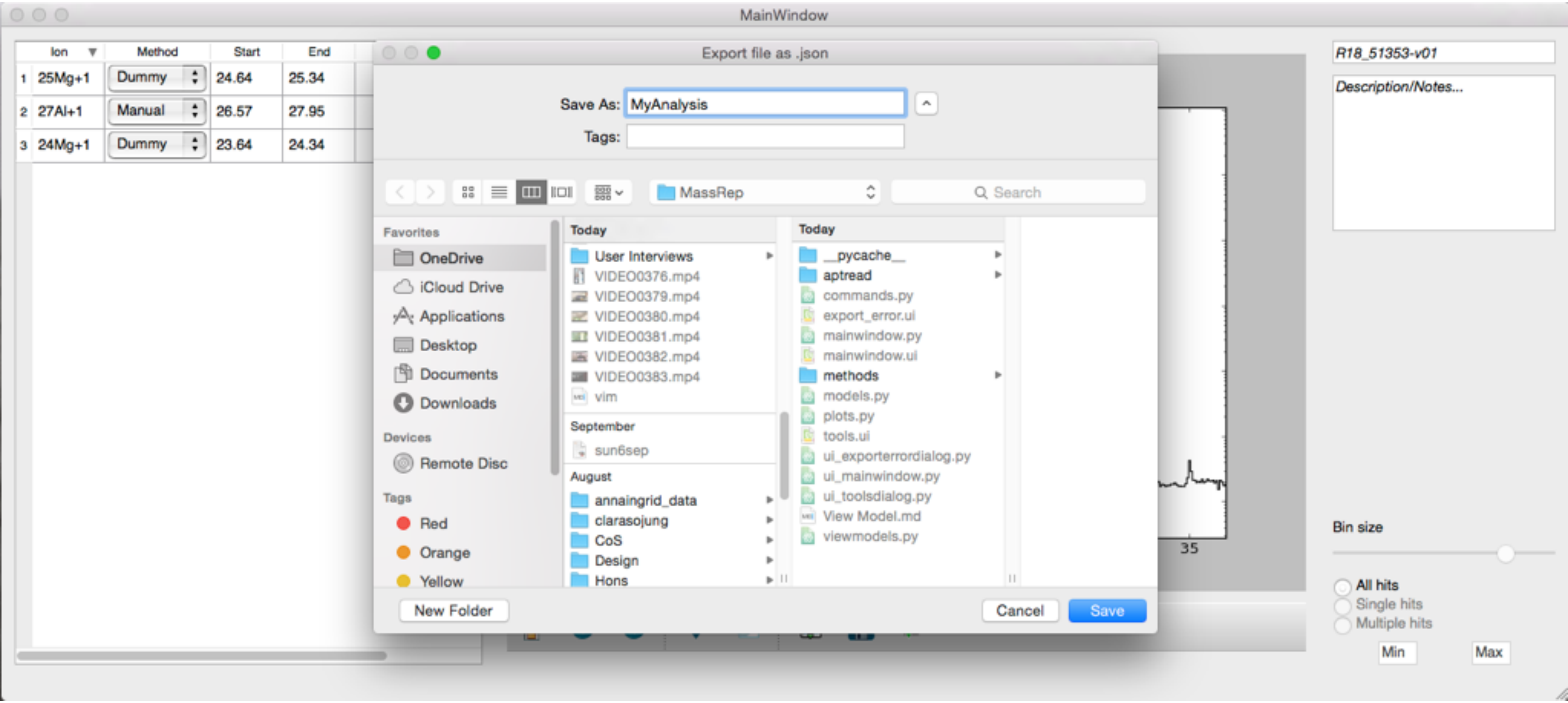
Min Max

Auto-method API



```python
def required_inputs():
    return ['bin_size', 'abundance', 'suggested_m2c']

def voodoo(bin_size, abundance, suggested_m2c):
    convolution = suggested_m2c/(bin_size*abundance)
    start = suggested_m2c - convolution
    end = suggested_m2c + convolution
    return (start, end)
```

Reproducible workflow file



MainWindow

	Ion	Method	Start	End
1	25Mg+1	Dummy	24.64	25.34
2	27Al+1	Manual	26.57	27.95
3	24Mg+1	Dummy	23.64	24.34

Export file as .json

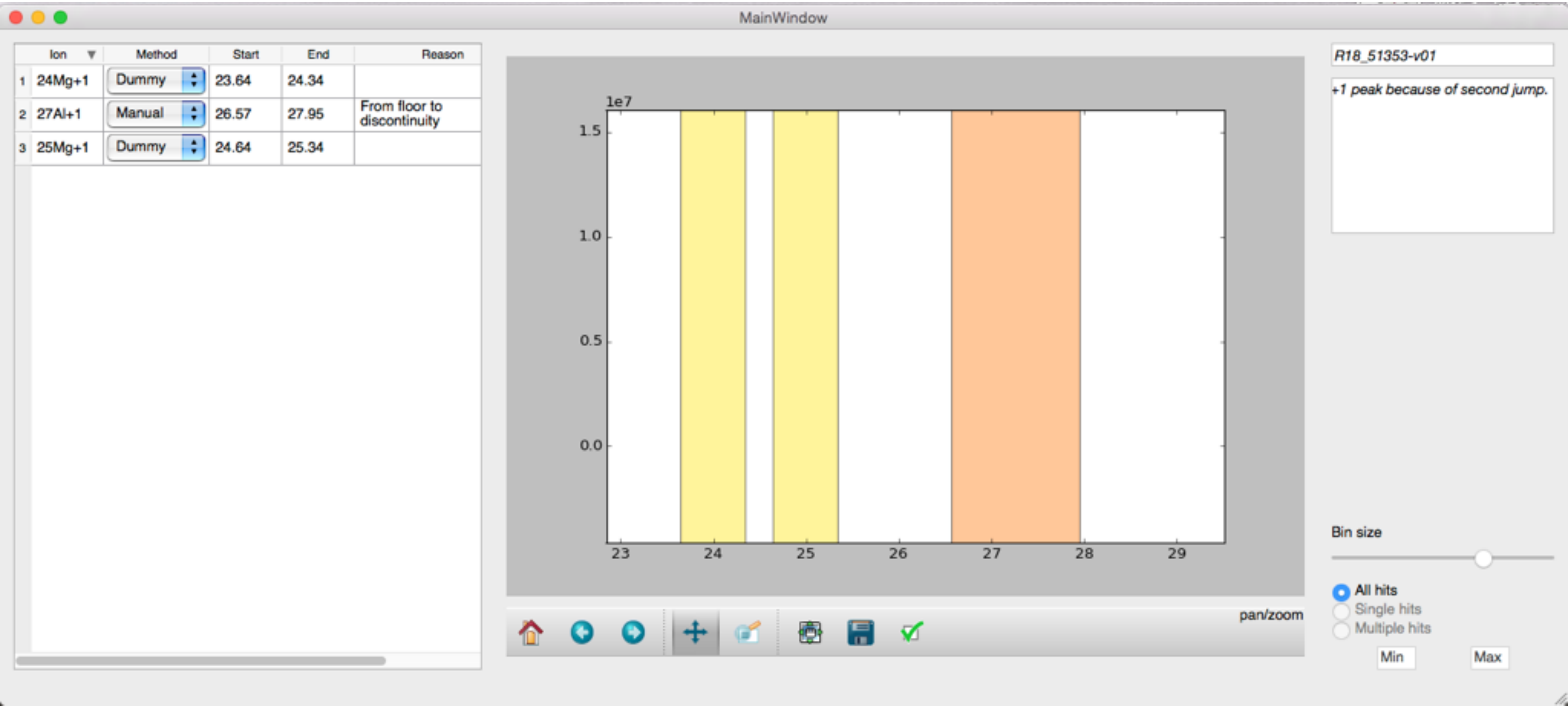
Save As: MyAnalysis

Tags:

R18_51353-v01

Description/Notes...

Reproducible workflow file



	Ion ▼	Method	Start	End	Reason
1	24Mg+1	Dummy	23.64	24.34	
2	27Al+1	Manual	26.57	27.95	From floor to discontinuity
3	25Mg+1	Dummy	24.64	25.34	

R18_51353-v01

+1 peak because of second jump.

Bin size

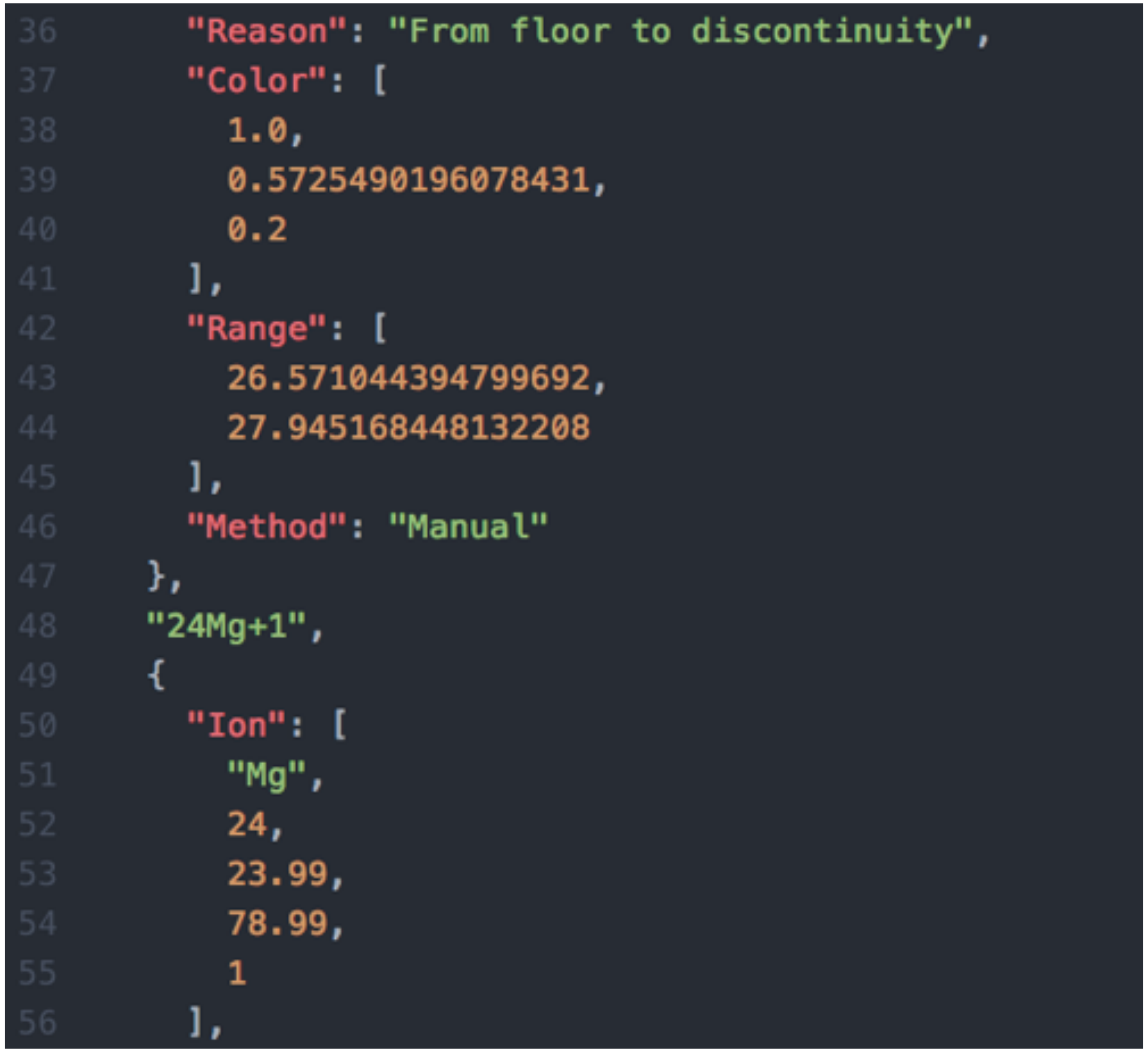
● All hits
○ Single hits
○ Multiple hits

Min Max

Transparent analysis

Workflow file

human readable
analysis + notes

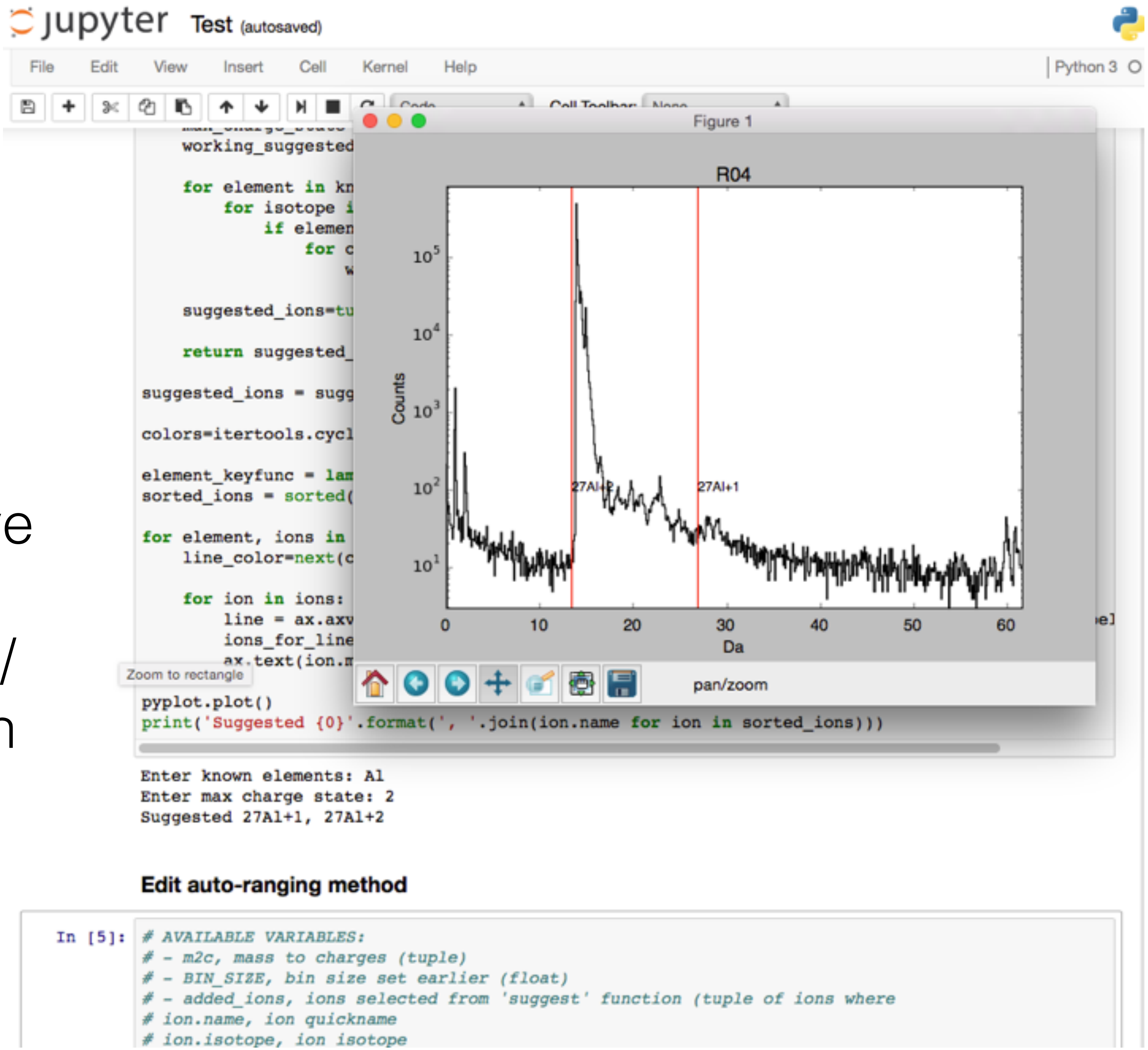


```
      "Reason": "From floor to discontinuity",
      "Color": [
        1.0,
        0.5725490196078431,
        0.2
      ],
      "Range": [
        26.571044394799692,
        27.945168448132208
      ],
      "Method": "Manual"
    },
    "24Mg+1",
    {
      "Ion": [
        "Mg",
        24,
        23.99,
        78.99,
        1
      ],
```

iPython notebook

send & receive

easy method/ user research



```
    working_suggested
    
    for element in kn
        for isotope i
            if elemen
                for c
                    w
    
    suggested_ions=tu
    
    return suggested_

suggested_ions = sugg

colors=itertools.cycl

element_keyfunc = lam
sorted_ions = sorted(

for element, ions in 
    line_color=next(c
    
    for ion in ions:
        line = ax.axv
        ions_for_line
        ax.text(ion.m

pyplot.plot()
print('Suggested {0}'.format(', '.join(ion.name for ion in sorted_ions)))
```

```
Enter known elements: Al
Enter max charge state: 2
Suggested 27Al+1, 27Al+2
```

Edit auto-ranging method

```
In [5]: # AVAILABLE VARIABLES:
        # - m2c, mass to charges (tuple)
        # - BIN_SIZE, bin size set earlier (float)
        # - added_ions, ions selected from 'suggest' function (tuple of ions where
        # ion.name, ion quickname
        # ion.isotope, ion isotope
```

Background Problem Proposal ***Results*** Conclusion

Conclusion

- Reproducibility is fundamental to research
- Mass spectrum analysis in atom probe tomography
  - is performed ad-hoc and heuristically
  - is not transparent or reproducible

Conclusion

✓ Problem and stakeholder research

✓ A new GUI — a new approach
- enables transparent and reproducible analysis
- encourages use of automatic methods

✓ An iPython notebook
- a lightweight tool for method/user research

Thank You

THE UNIVERSITY OF SYDNEY

ammrf
Australian Microscopy
& Microanalysis Research Facility

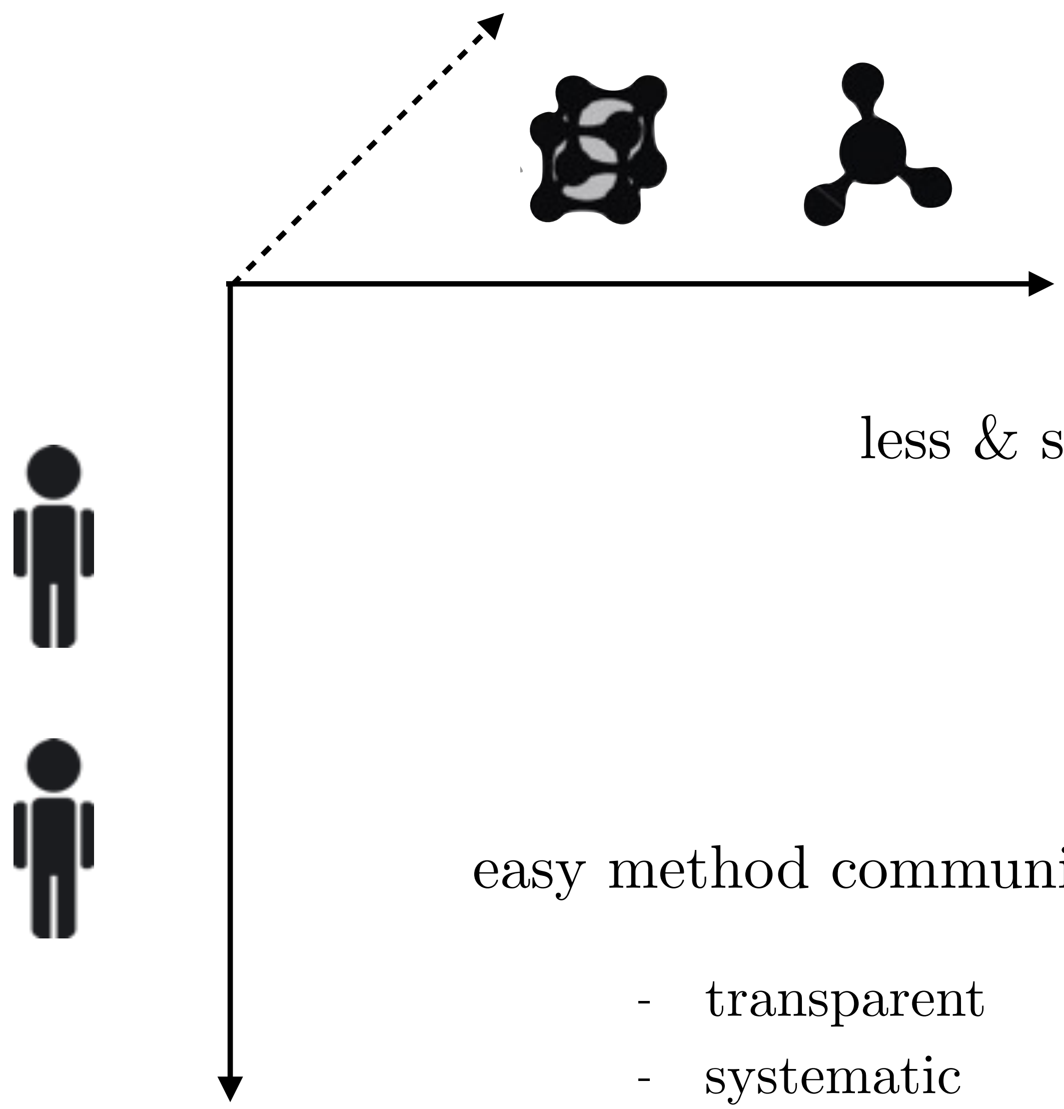
“A scientist builds in order to learn; an engineer learns in order to build.”

– Fred Brooks Jr., architect of the IBM System/360

[extra slides]

The scientific problem model

accuracy



consistency

less & standard/automated methods

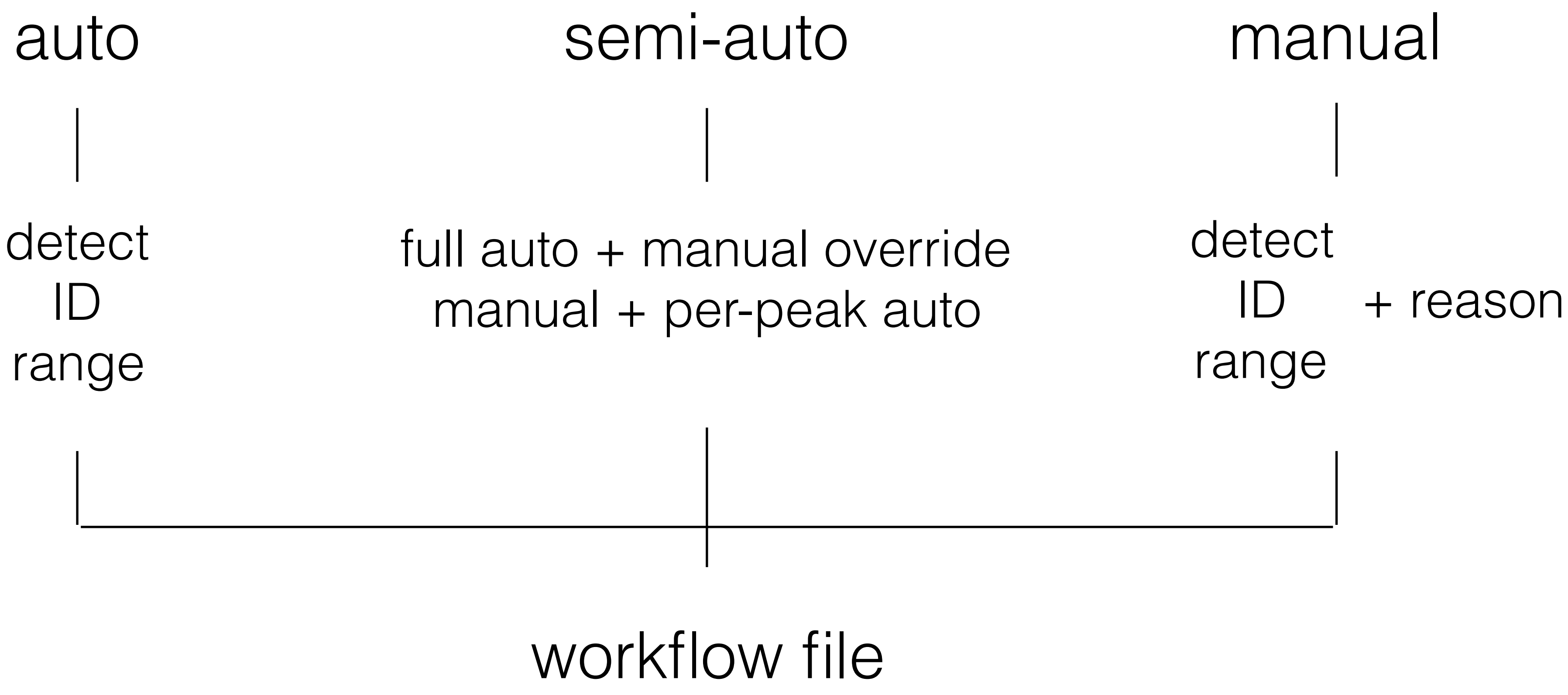
easy method communication

- transparent
- systematic
- standardised
- common language

reproducibility

A new workflow

method-based



auto

detect
ID
range

semi-auto

full auto + manual override
manual + per-peak auto

manual

detect
ID
range
+ reason

workflow file

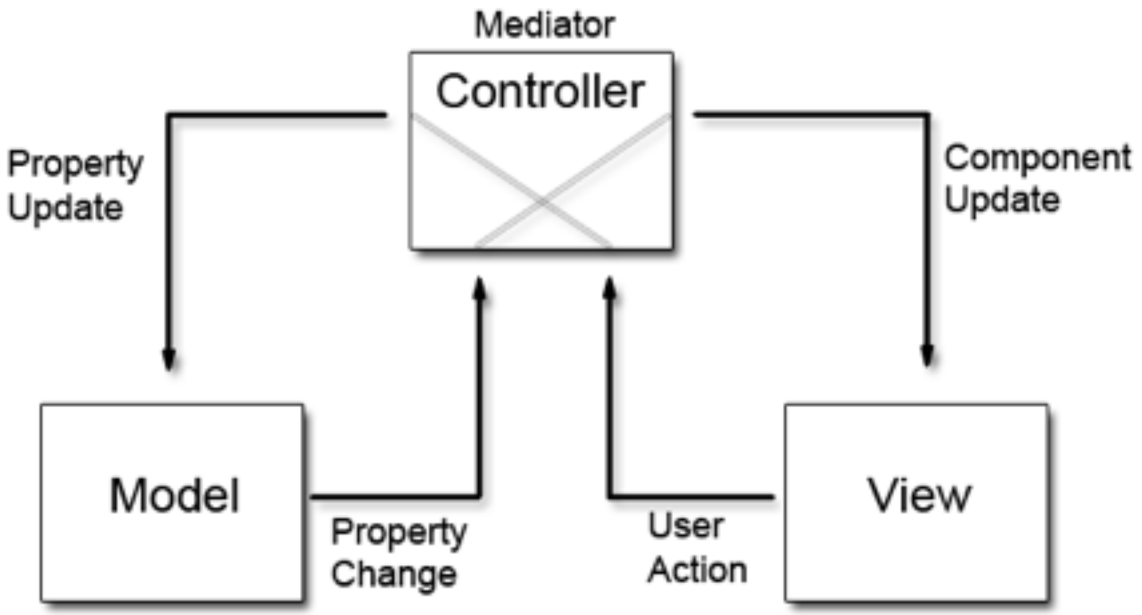
transparent, recorded

Sustainable code

‘Clean Code’ ✓

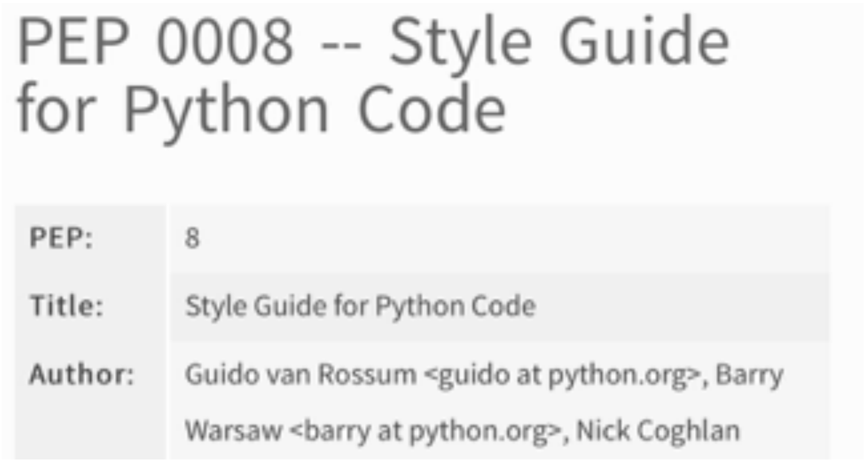
Robert C. Martin Series
Prentice Hall
Clean Code
A Handbook of Agile Software Craftsmanship
Foreword by James O. Coplien
Robert C. Martin

Model-view-controller ✓



http://www.codeproject.com/KB/tips/ModelViewController/Figure4.gif

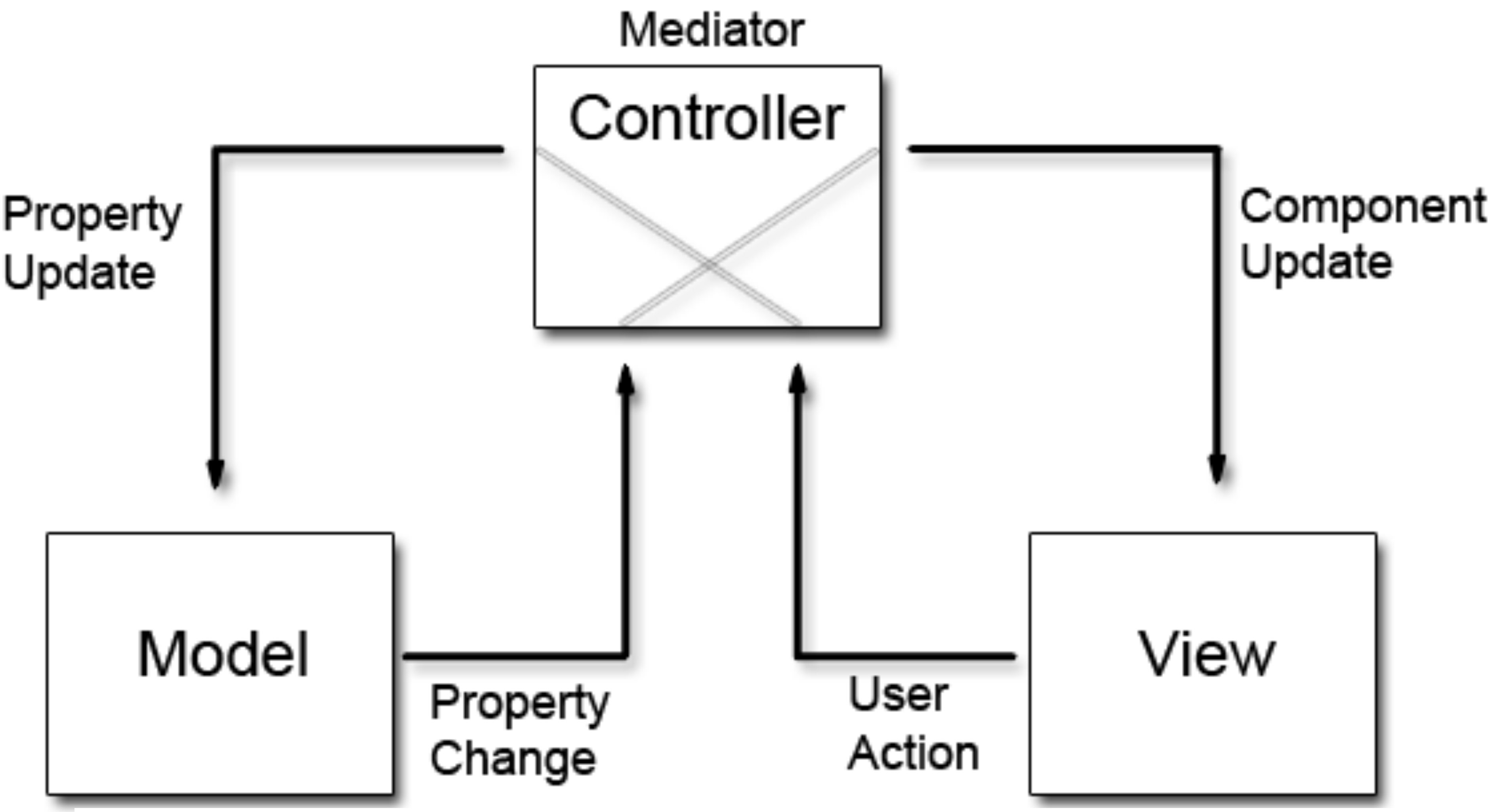
Python Style Guide ✓



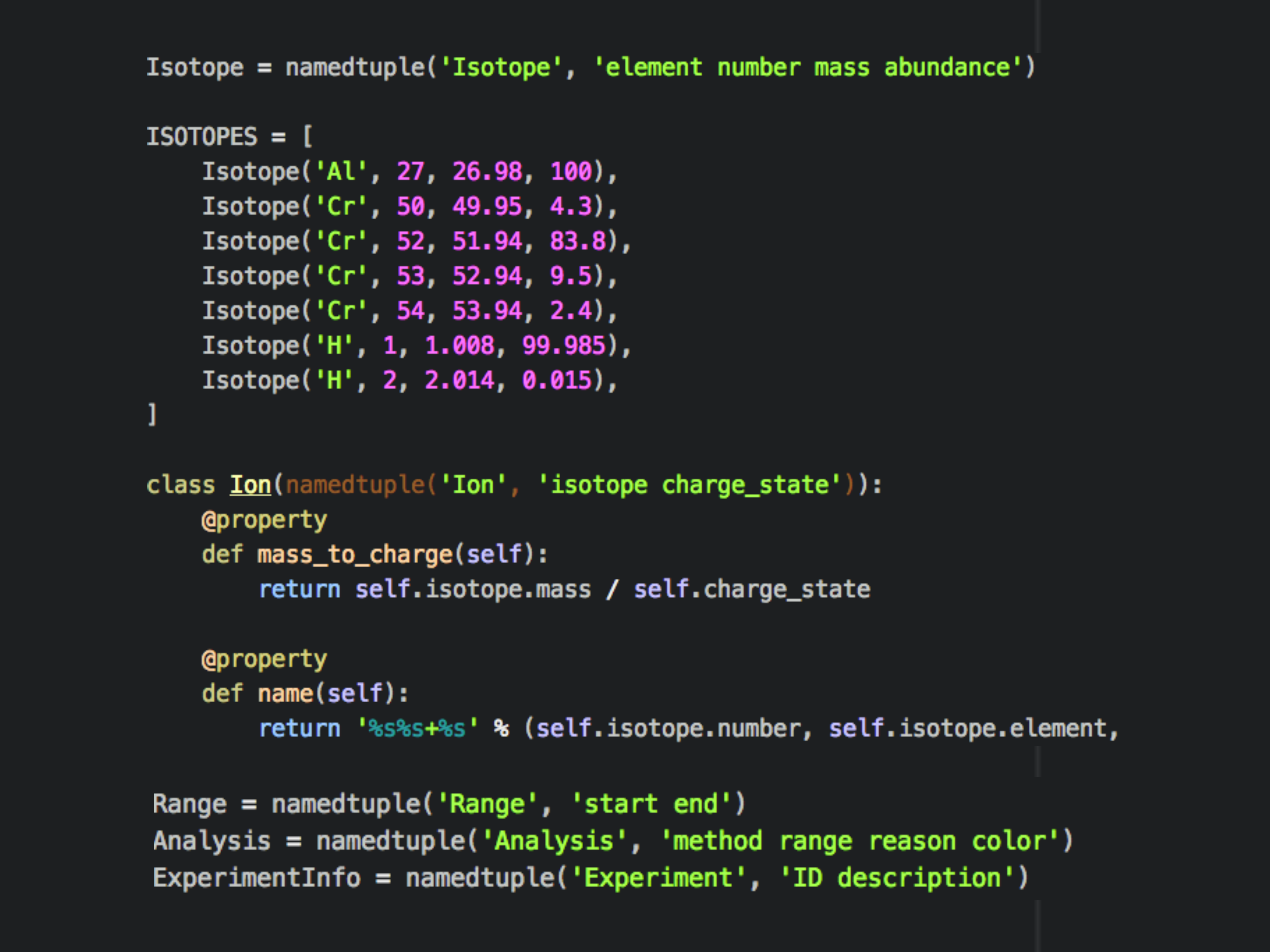
PEP 0008 -- Style Guide for Python Code

PEP:	8
Title:	Style Guide for Python Code
Author:	Guido van Rossum <guido at python.org>, Barry Warsaw <barry at python.org>, Nick Coghlan

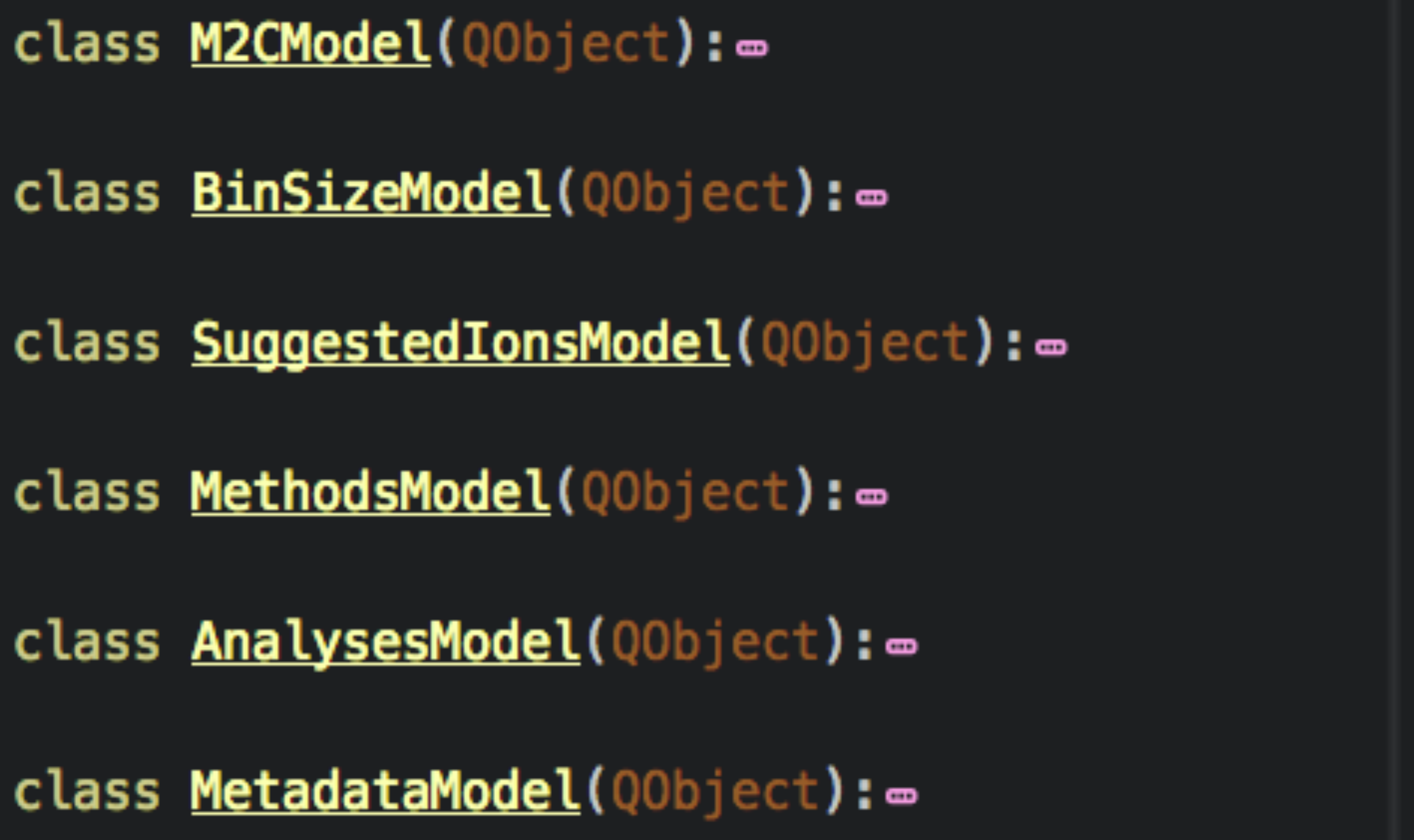
Model-view-controller



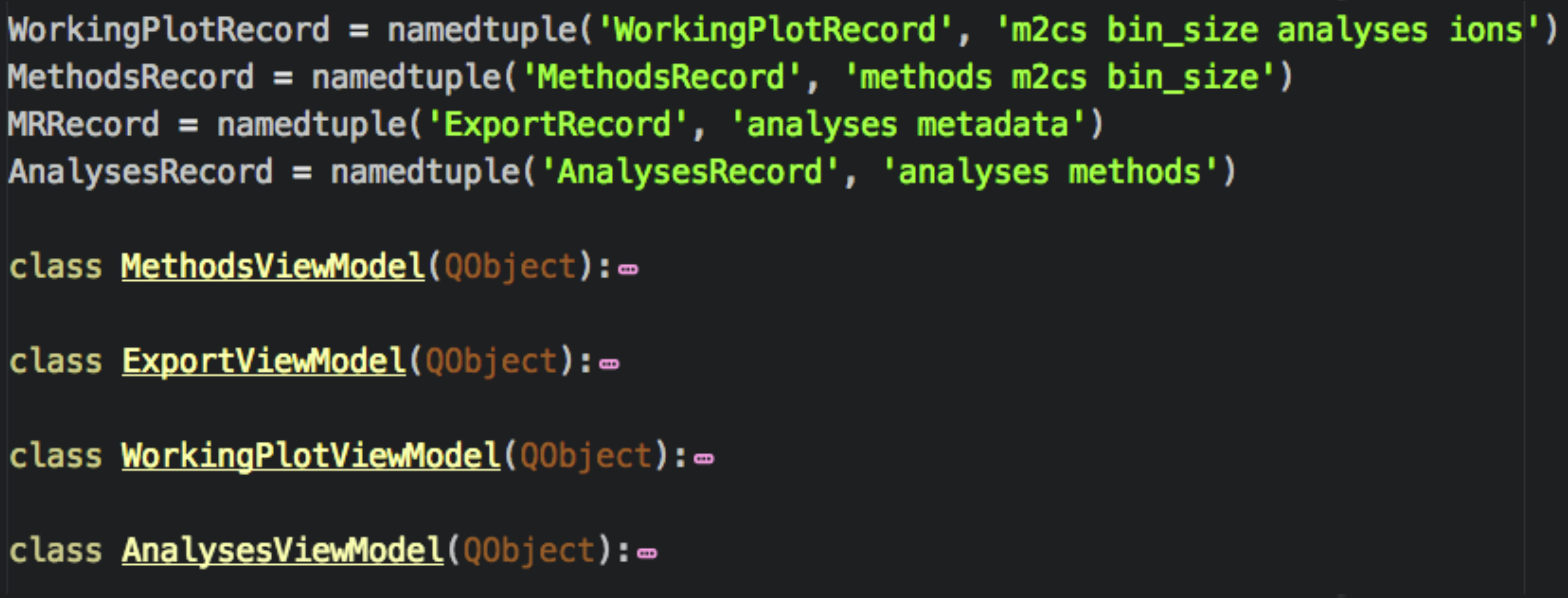
http://www.codeproject.com/KB/tips/ModelViewController/Figure4.gif



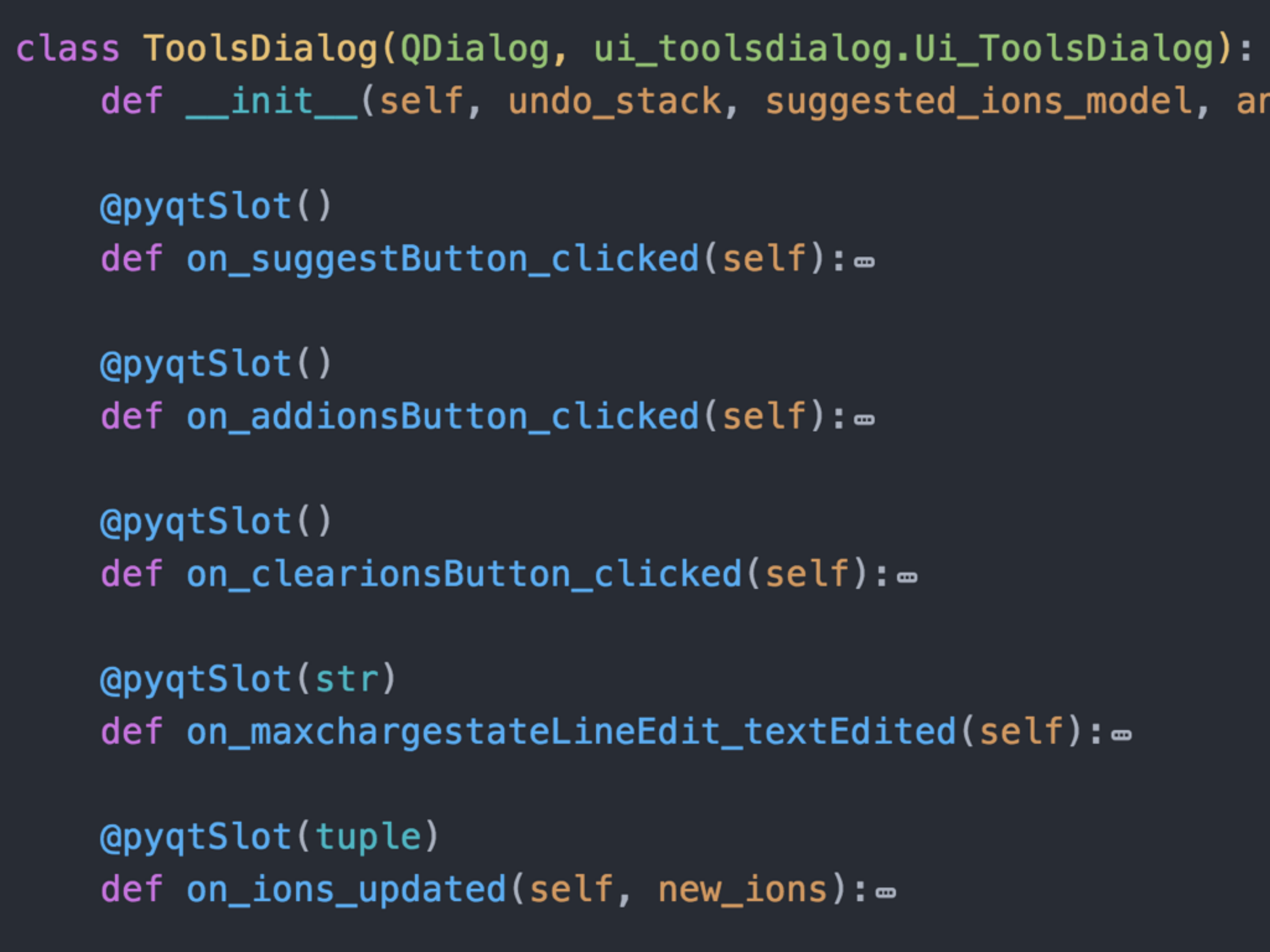
```
Isotope = namedtuple('Isotope', 'element number mass abundance')

ISOTOPES = [
    Isotope('Al', 27, 26.98, 100),
    Isotope('Cr', 50, 49.95, 4.3),
    Isotope('Cr', 52, 51.94, 83.8),
    Isotope('Cr', 53, 52.94, 9.5),
    Isotope('Cr', 54, 53.94, 2.4),
    Isotope('H', 1, 1.008, 99.985),
    Isotope('H', 2, 2.014, 0.015),
]

class Ion(namedtuple('Ion', 'isotope charge_state')):
    @property
    def mass_to_charge(self):
        return self.isotope.mass / self.charge_state

    @property
    def name(self):
        return '%s%s+%s' % (self.isotope.number, self.isotope.element,

Range = namedtuple('Range', 'start end')
Analysis = namedtuple('Analysis', 'method range reason color')
ExperimentInfo = namedtuple('Experiment', 'ID description')
```



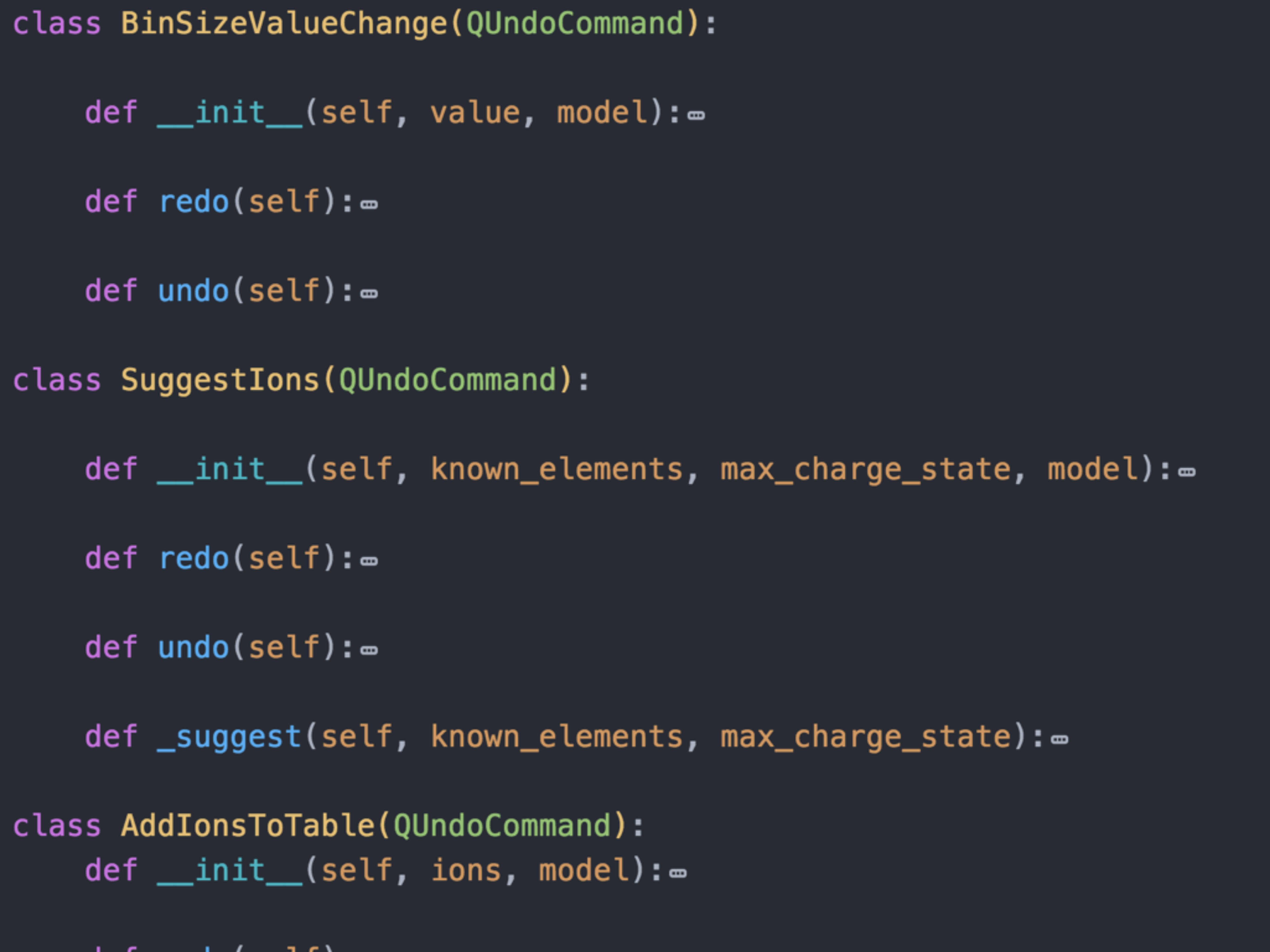
```python
class M2CModel(QObject):

class BinSizeModel(QObject):

class SuggestedIonsModel(QObject):

class MethodsModel(QObject):

class AnalysesModel(QObject):

class MetadataModel(QObject):
```



```python
WorkingPlotRecord = namedtuple('WorkingPlotRecord', 'm2cs bin_size analyses ions')
MethodsRecord = namedtuple('MethodsRecord', 'methods m2cs bin_size')
MRRecord = namedtuple('ExportRecord', 'analyses metadata')
AnalysesRecord = namedtuple('AnalysesRecord', 'analyses methods')

class MethodsViewModel(QObject):

class ExportViewModel(QObject):

class WorkingPlotViewModel(QObject):

class AnalysesViewModel(QObject):
```

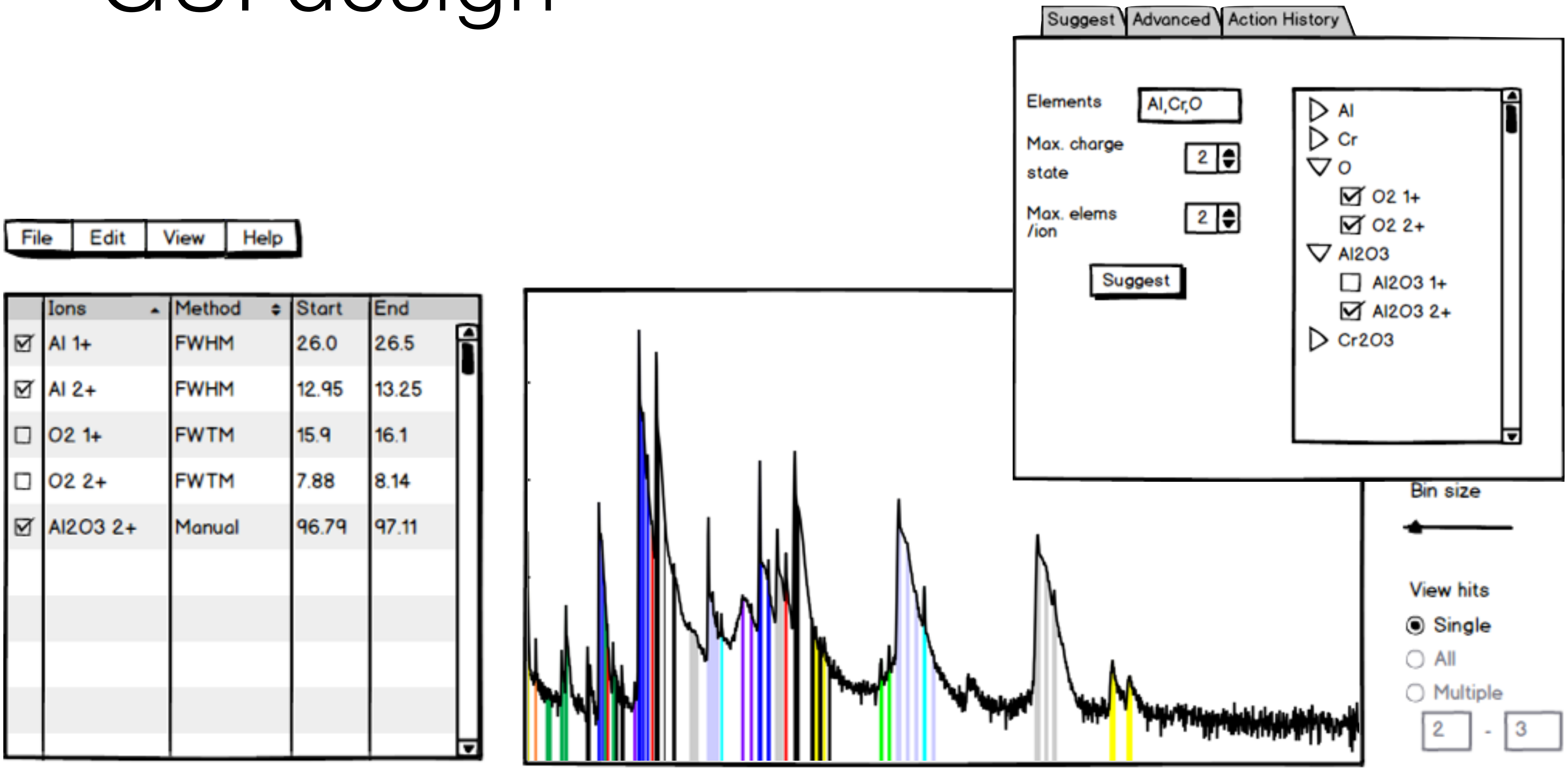


```
class ToolsDialog(QDialog, ui_toolsdialog.Ui_ToolsDialog):
    def __init__(self, undo_stack, suggested_ions_model, ar

    @pyqtSlot()
    def on_suggestButton_clicked(self):⋯

    @pyqtSlot()
    def on_addionsButton_clicked(self):⋯

    @pyqtSlot()
    def on_clearionsButton_clicked(self):⋯

    @pyqtSlot(str)
    def on_maxchargestateLineEdit_textEdited(self):⋯

    @pyqtSlot(tuple)
    def on_ions_updated(self, new_ions):⋯
```



```python
class BinSizeValueChange(QUndoCommand):

    def __init__(self, value, model):⋯

    def redo(self):⋯

    def undo(self):⋯

class SuggestIons(QUndoCommand):

    def __init__(self, known_elements, max_charge_state, model):⋯

    def redo(self):⋯

    def undo(self):⋯

    def _suggest(self, known_elements, max_charge_state):⋯

class AddIonsToTable(QUndoCommand):
    def __init__(self, ions, model):⋯

    def redo(self):
```

GUI design



	Ions	Method	Start	End
☑	Al 1+	FWHM	26.0	26.5
☑	Al 2+	FWHM	12.95	13.25
□	O2 1+	FWTM	15.9	16.1
□	O2 2+	FWTM	7.88	8.14
☑	Al2O3 2+	Manual	96.79	97.11

Elements: Al,Cr,O

Max. charge state: 2

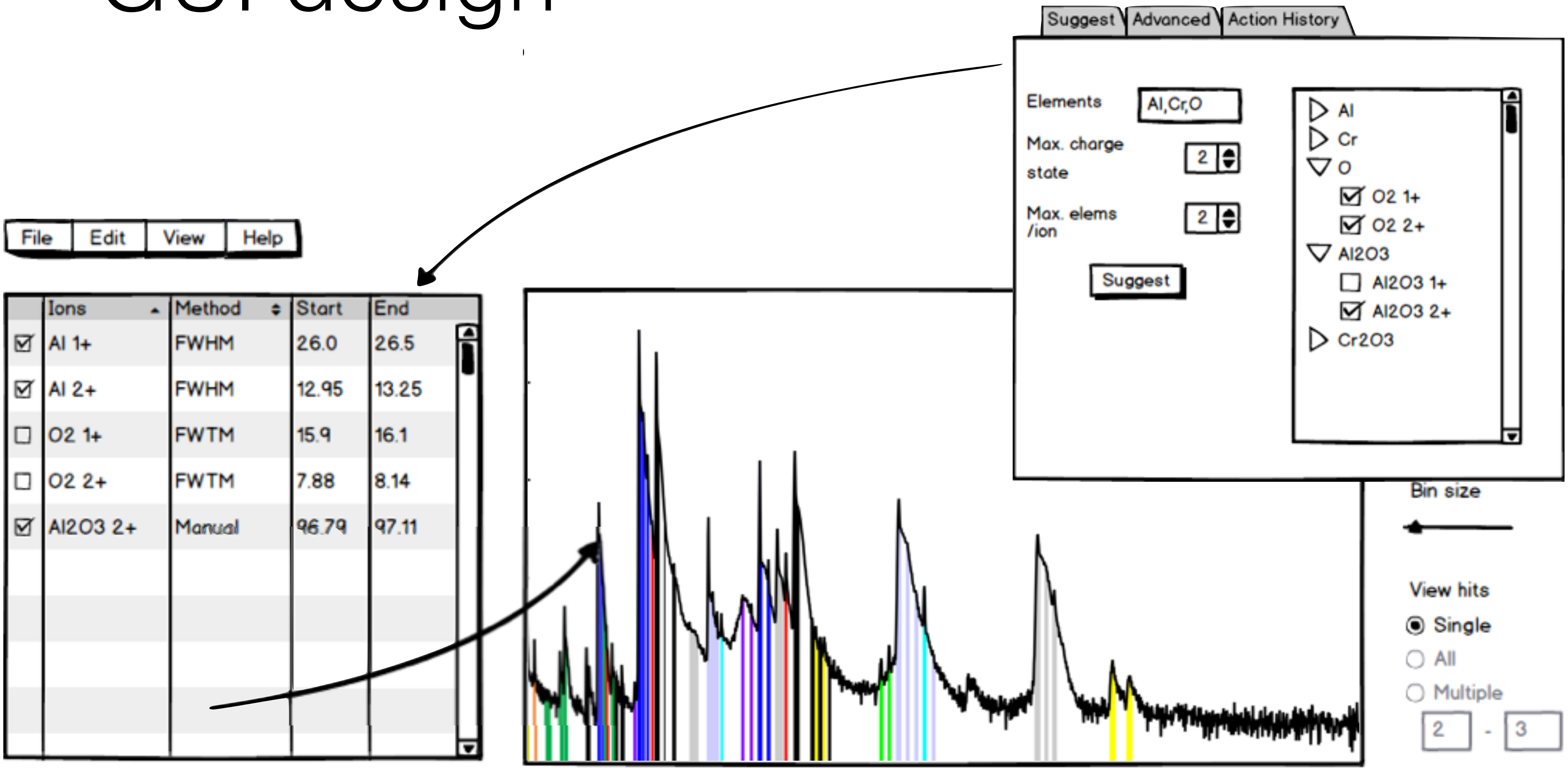
Max. elems /ion: 2

- ▷ Al
- ▷ Cr
- ▽ O
  - ☑ O2 1+
  - ☑ O2 2+
- ▽ Al2O3
  - □ Al2O3 1+
  - ☑ Al2O3 2+
- ▷ Cr2O3

View hits

● Single
○ All
○ Multiple 2 - 3

GUI design



	Ions	Method	Start	End
☑	Al 1+	FWHM	26.0	26.5
☑	Al 2+	FWHM	12.95	13.25
□	O2 1+	FWTM	15.9	16.1
□	O2 2+	FWTM	7.88	8.14
☑	Al2O3 2+	Manual	96.79	97.11

Elements: Al,Cr,O

Max. charge state: 2

Max. elems /ion: 2

- ▷ Al
- ▷ Cr
- ▽ O
  - ☑ O2 1+
  - ☑ O2 2+
- ▽ Al2O3
  - □ Al2O3 1+
  - ☑ Al2O3 2+
- ▷ Cr2O3

Bin size

View hits

● Single

○ All

○ Multiple 2 - 3

Deliverables

- problem identification & problem model
- user interviews & ‘user stories’
- the ‘auto-manual’ workflow (designed by Ingrid McCarroll)
- GUI, its framework and its code architecture
- Code Style Guide
- iPython notebook for easy send/receive research
- recommendations & future work

Learning outcomes

- GUI programming is very difficult
- Good software/product design process is even more critical
- Consider all stakeholders and their interests
- Consider the logic of the problem and compare solutions that improve the problem or negate the problem

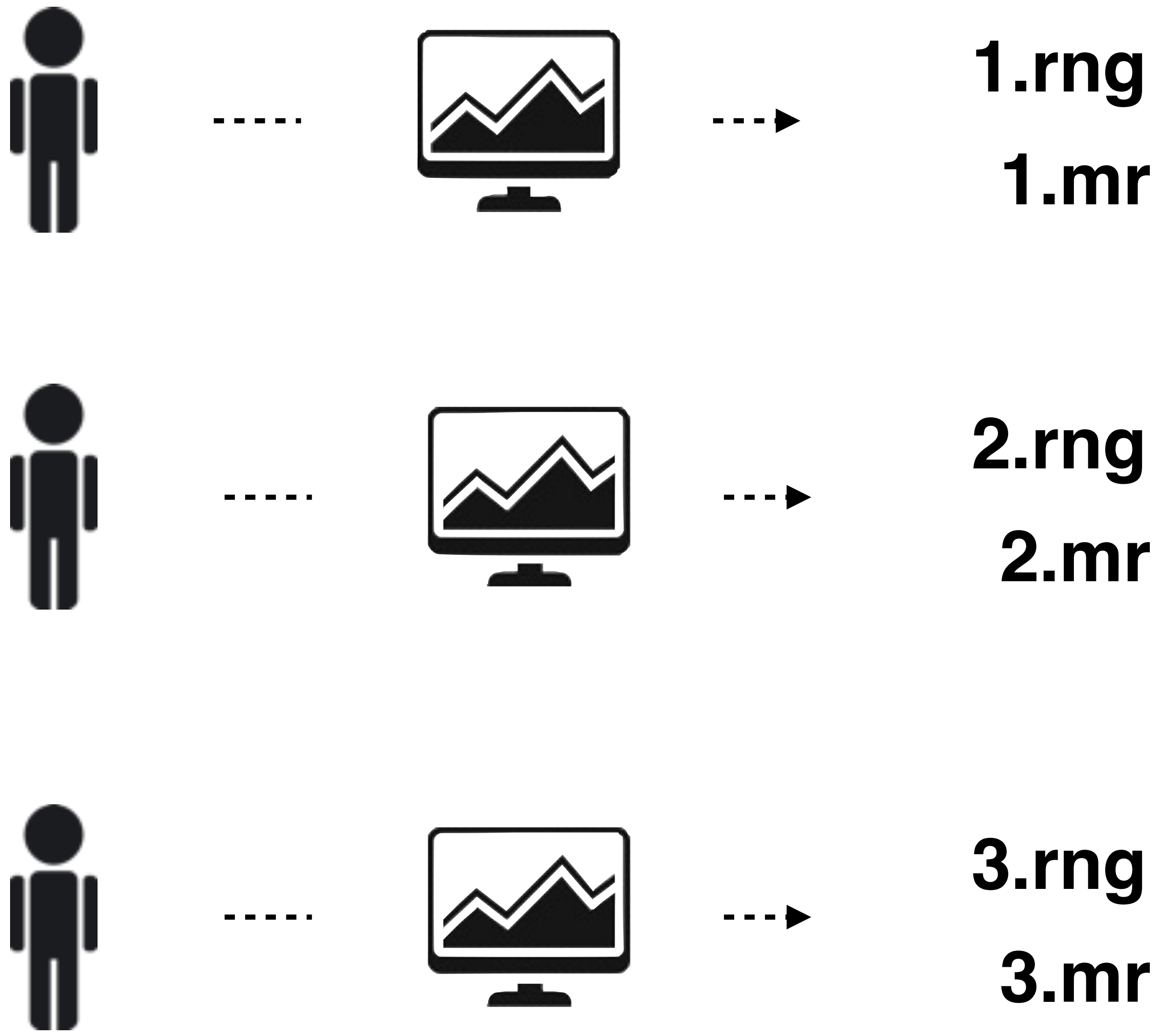
different

same

own

1.rng 2.rng 3.rng

1.json 2.json 3.json



1.rng

1.mr

2.rng

2.mr

3.rng

3.mr