Penn State University

PRINCETON UNIVERSITY

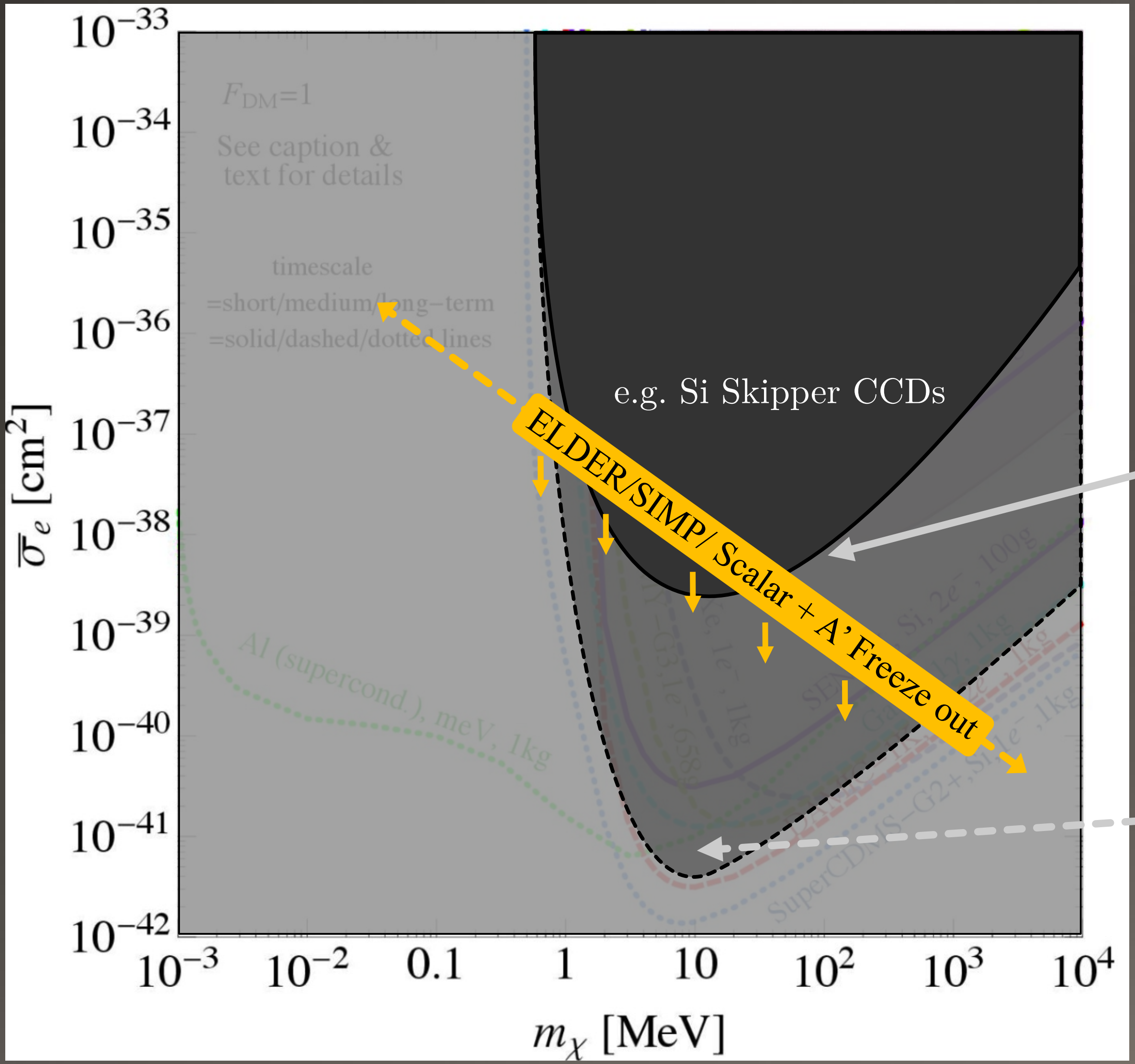
AI-Driven Material Discovery for Next-Generation Directional Dark Matter Detectors

CARLOS BLANCO

Stockholm University

Oskar Klein centre

Outlook and Potential Reach



[US Cosmic Visions: Community Report: 1707.04591]

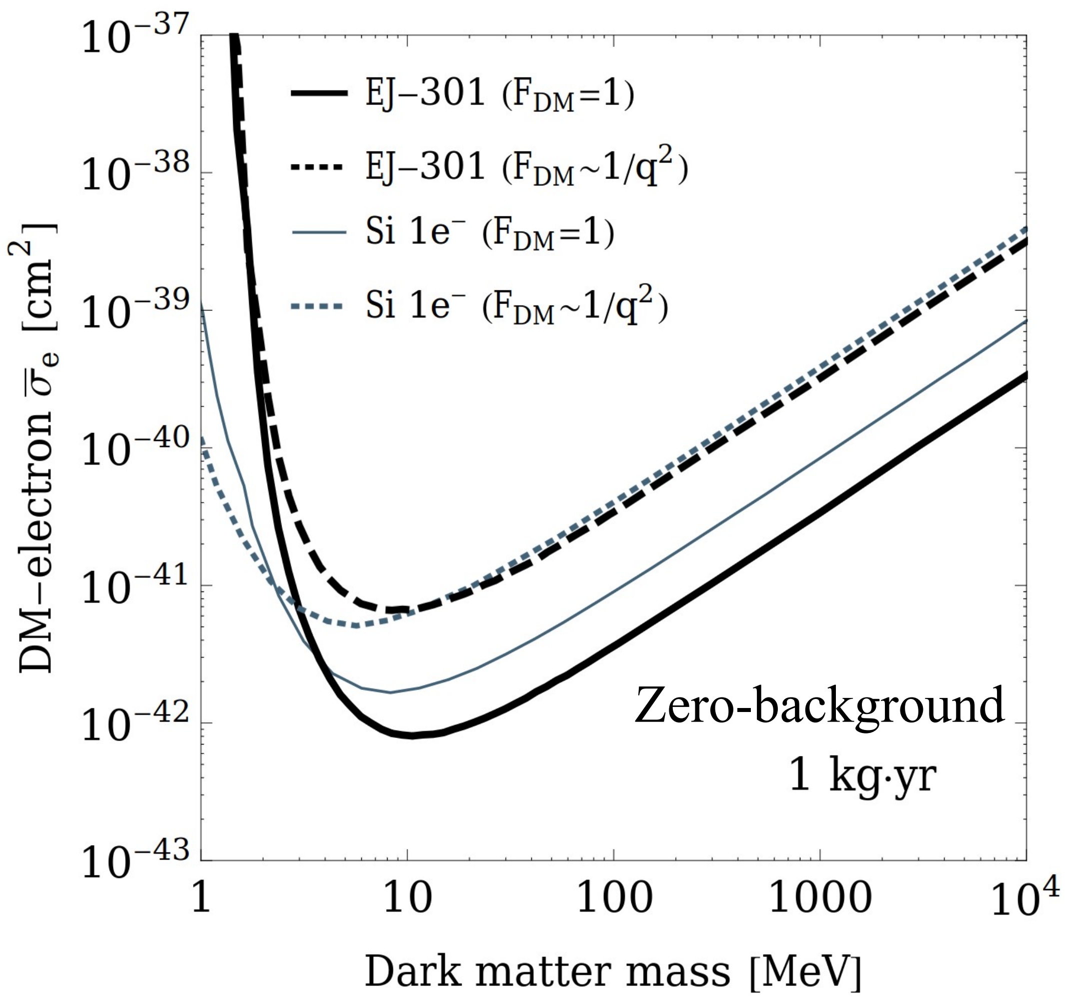
Present day:

Experimental results

Zero-background projection

Outlook and Potential Reach

[**CB**, Collar, Kahn, Lillard: 1912.02822]



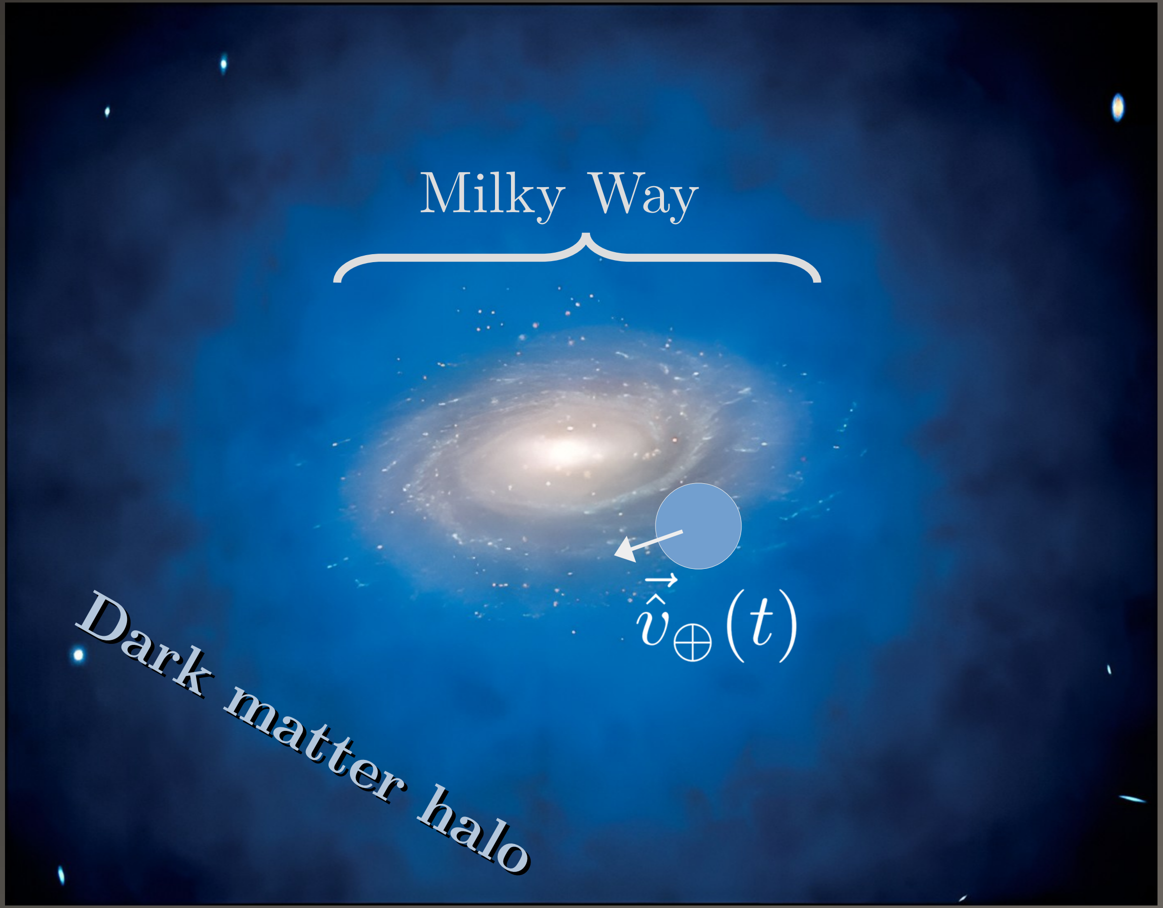
The obstacle

Backgrounds

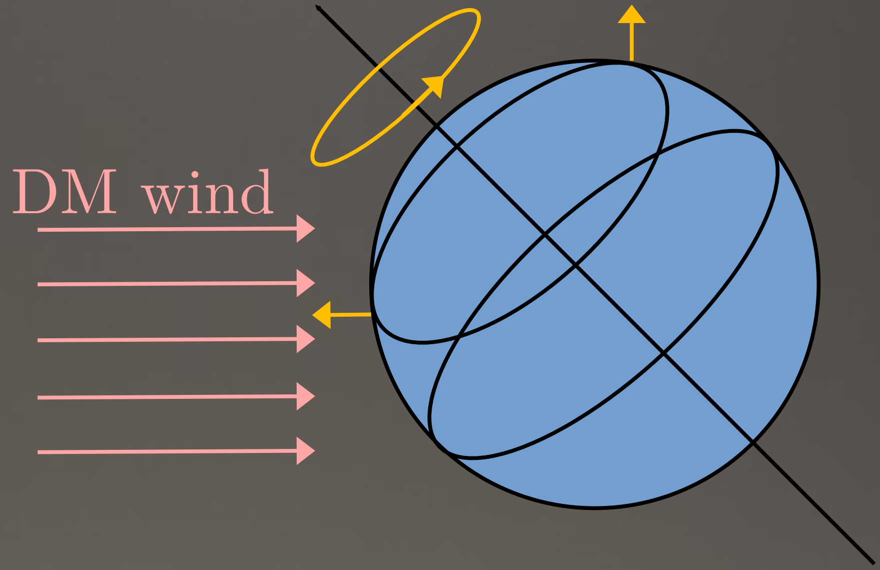
4 orders of magnitude of potential reach awaiting

Pound (kg) for Pound (kg) molecules produce about as much signal as e.g. Si.

Directional Detection



Effective dark matter "wind" from relative motion

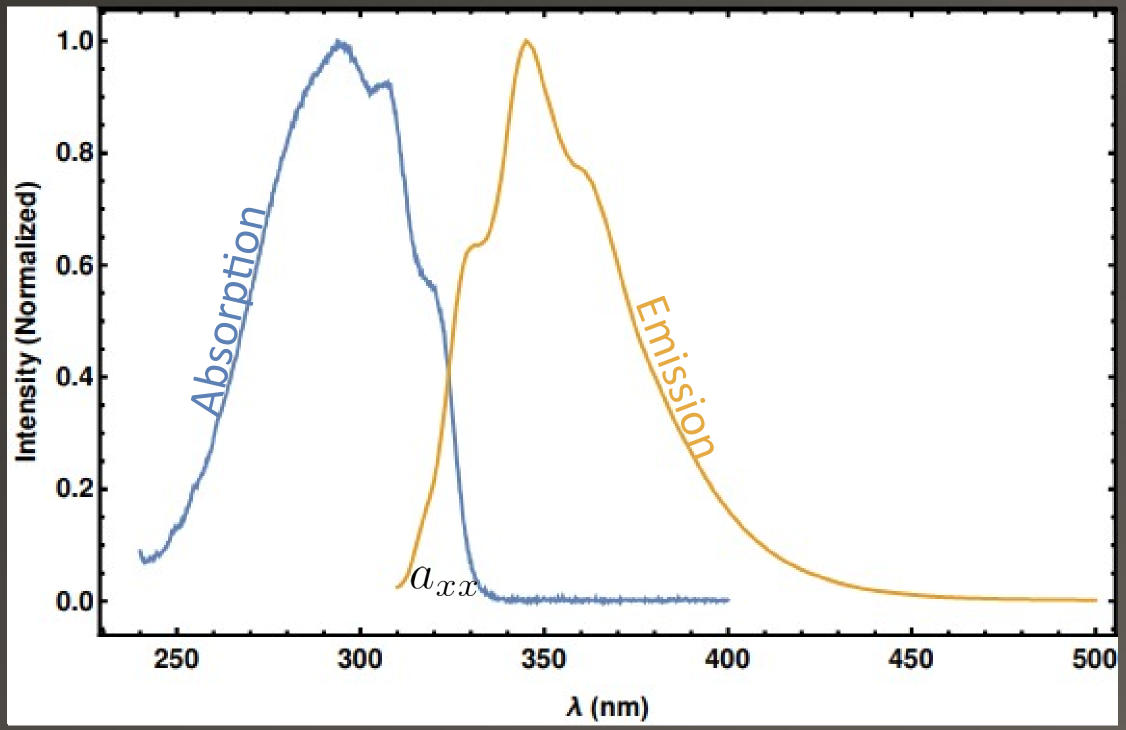


Change in relative orientation between detector and dark matter wind leads to ***daily modulation***

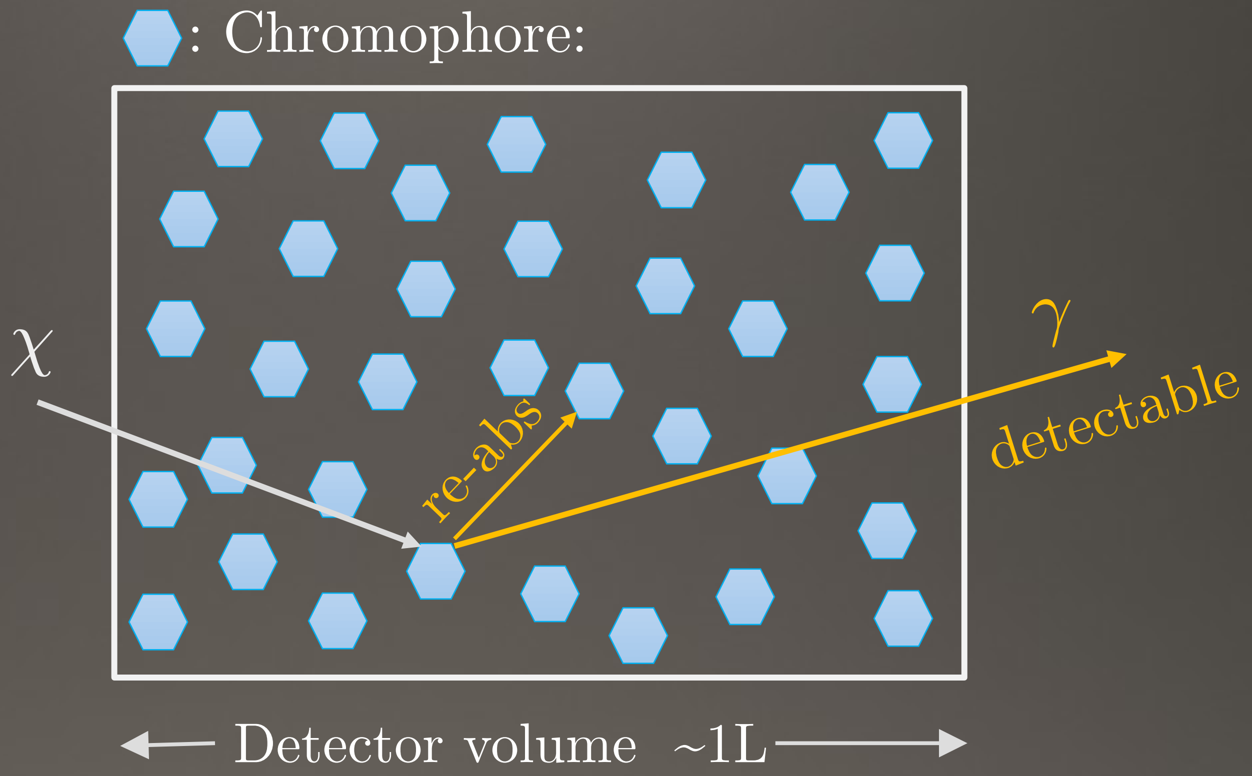
Fluorescence with DM

fluorescence spectra

Absorption/emission Probability



Decreasing energy (E) →

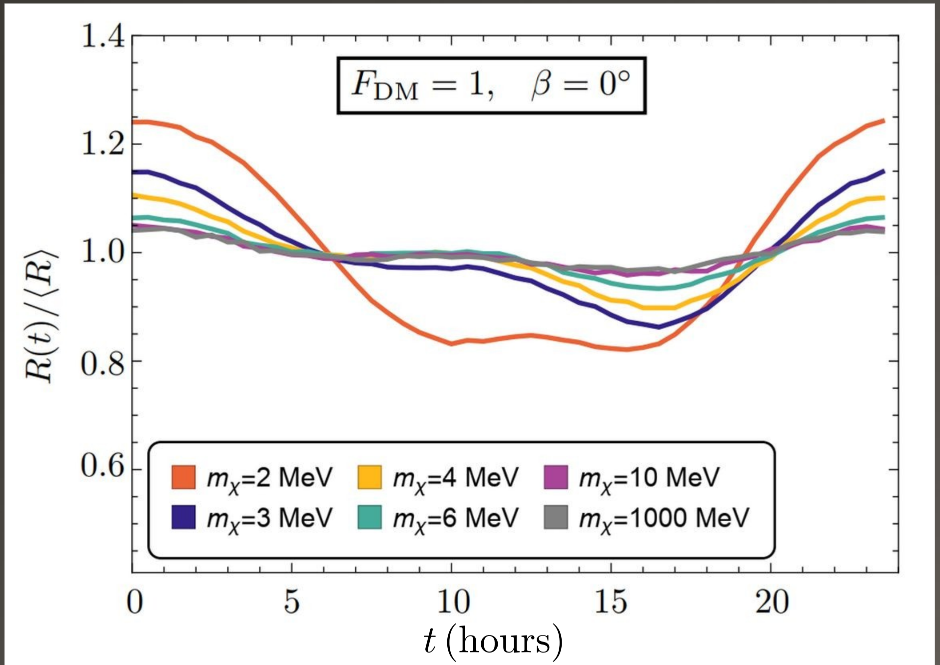


Probability for the photon to free stream

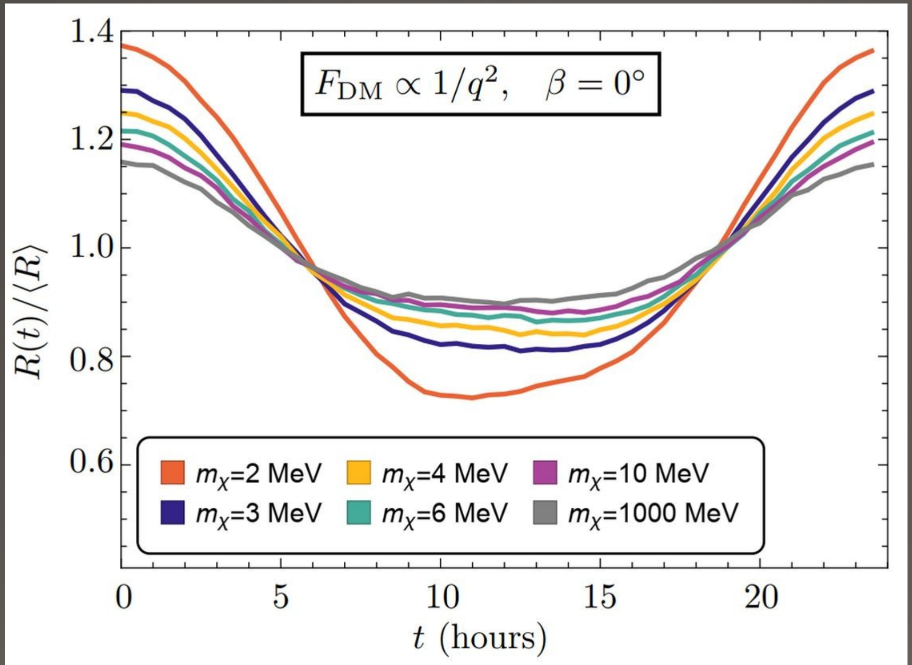
$\Phi_{\mathrm{FB}} \sim (1 - a_{xx})$ e.g. molecular crystals: $\Phi_{\mathrm{FB}} \approx 65\%$

Daily Modulation

(Contact interaction)



(Long-range interaction)



[CB, Kahn, Lillard, McDermott: 2103.08601]

Predicted rate changes by up to 70% throughout the day.

That's a verifiable signal!

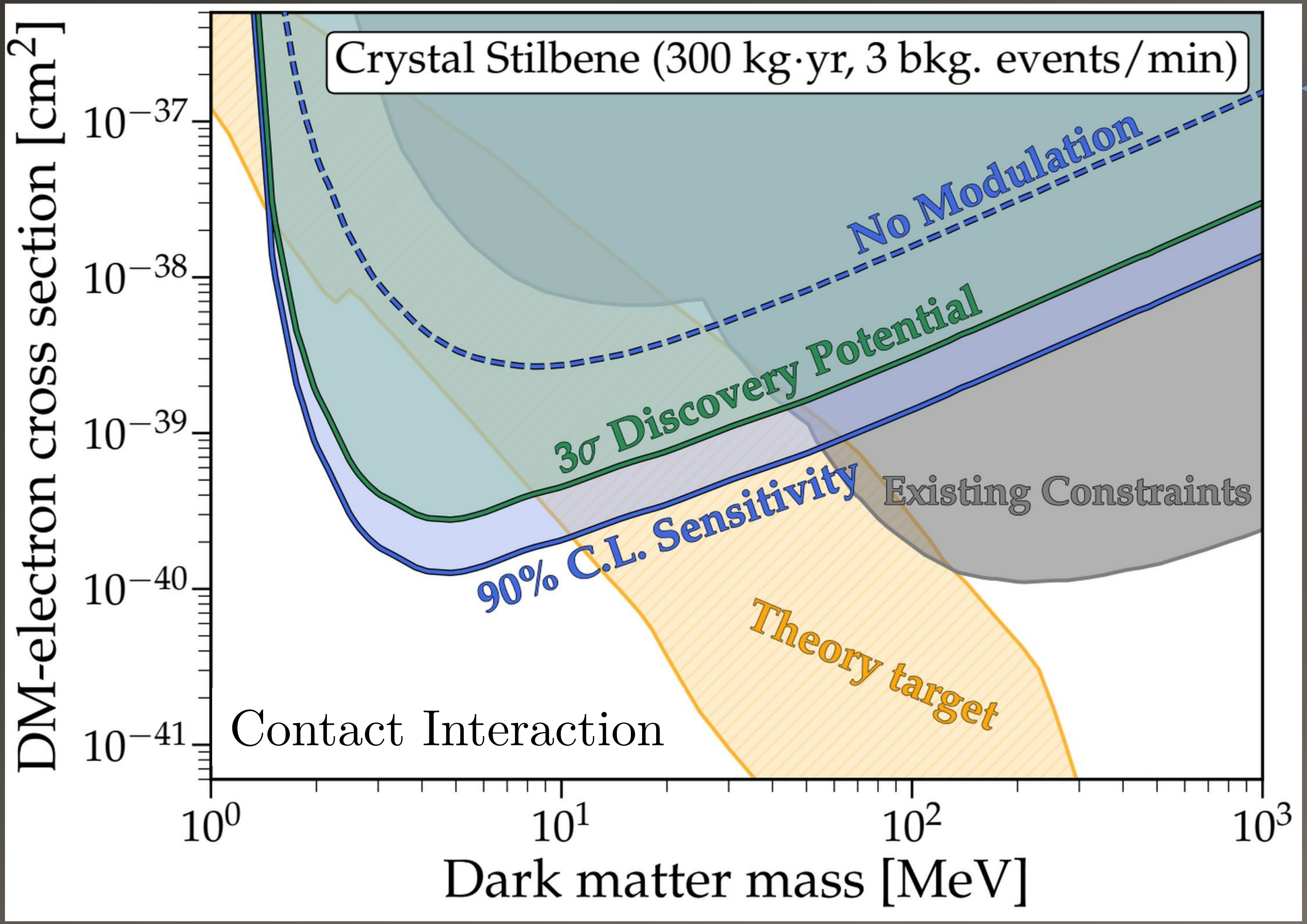
Modulation amplitude remains as high as 10% even at the highest masses. This is due to the ***fundamental anisotropy of the molecular form factor***.

$$f_{\rm RMS} = [5\% - 25\%]$$

$$N_\sigma \sim f_{\rm RMS}\sqrt{N_{\rm events}}$$

[CB, Foster, Kahn, Lillard: 2602.15947]

Sensitivity & Reach



Assuming realistic backgrounds

Exclusion w/o modulation

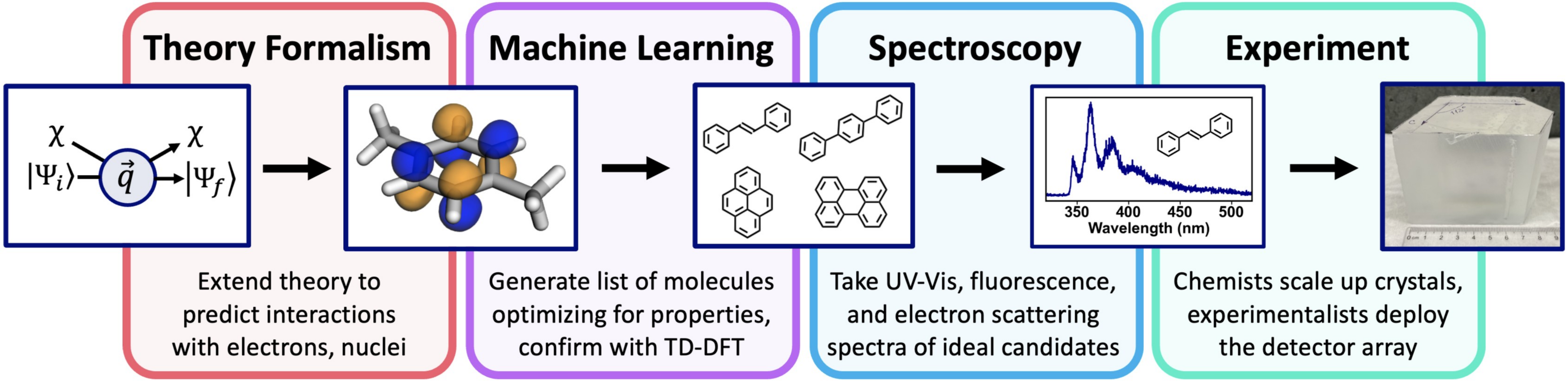
Discovery potential w/ mod.

Exclusion w/ modulation

100 kg t-stilbene crystals exposed for 3 years w/ background rate of 3 events/min dominated by radioactivity of ^{14}C, Th, & U.

Consistent w/ Borexino $^{14}C/^{12}C \sim 10^{-18}$ (0.01/min/kg)

Detector Development Pipeline



Inventing Optimal Targets

Problem: Chemical space is unreasonably large

How many molecules possible with
C, O, N, F, H?

< 9 atoms: 100s of Thousands (DFT Computable)

< 30 atoms: 100s of Billions (Intractable)
...toluene has 15, xylene has 18, t-stilbene has 26

Method

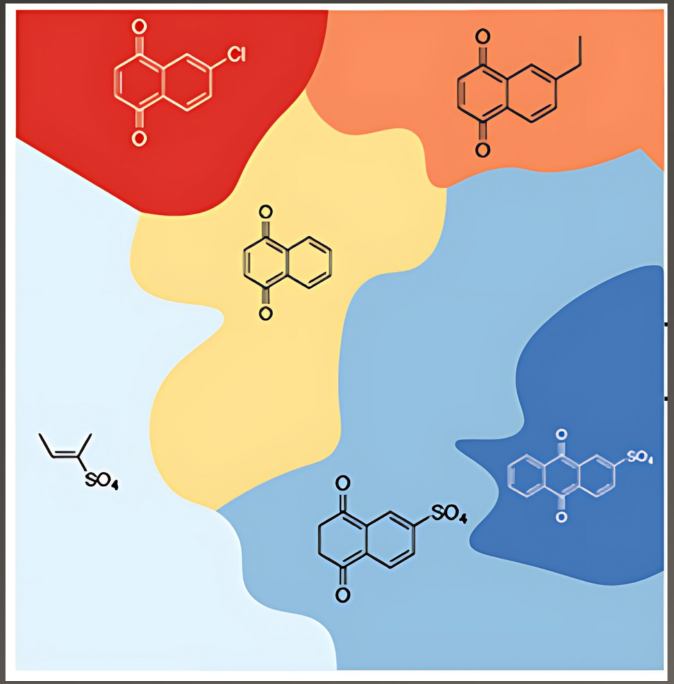
1. Look for known favorable properties - *cheminformatics*
2. Extra(intra)polate onto new molecules – *machine learning*

ML for DM Direct Detection

Property prediction — Energies & Matrix elements

Molecular Generation — Sample latent space → new molecules

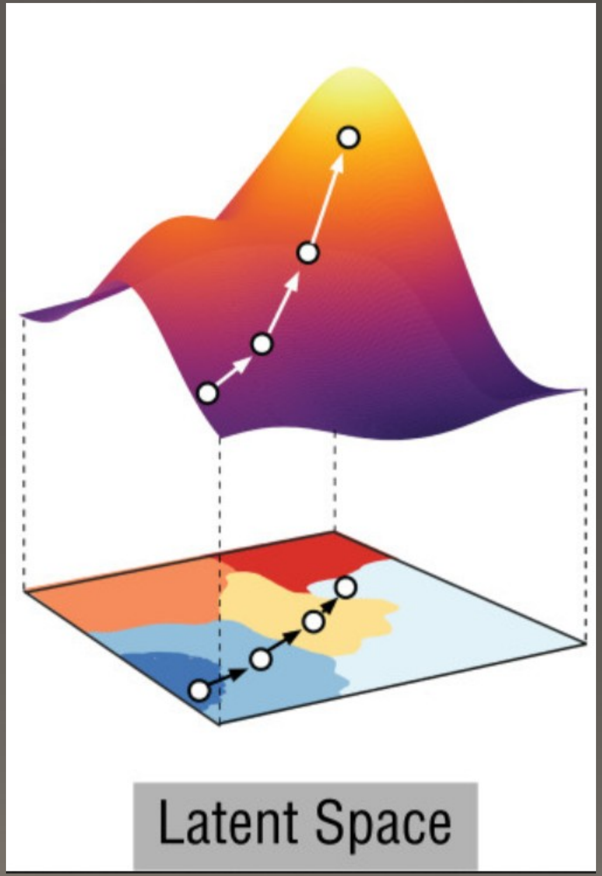
Map space of molecules into a smooth "latent space"



[Gomez-Bombarelli,ACS central science 4.2 (2018)]

Property prediction

Energies & Matrix elements

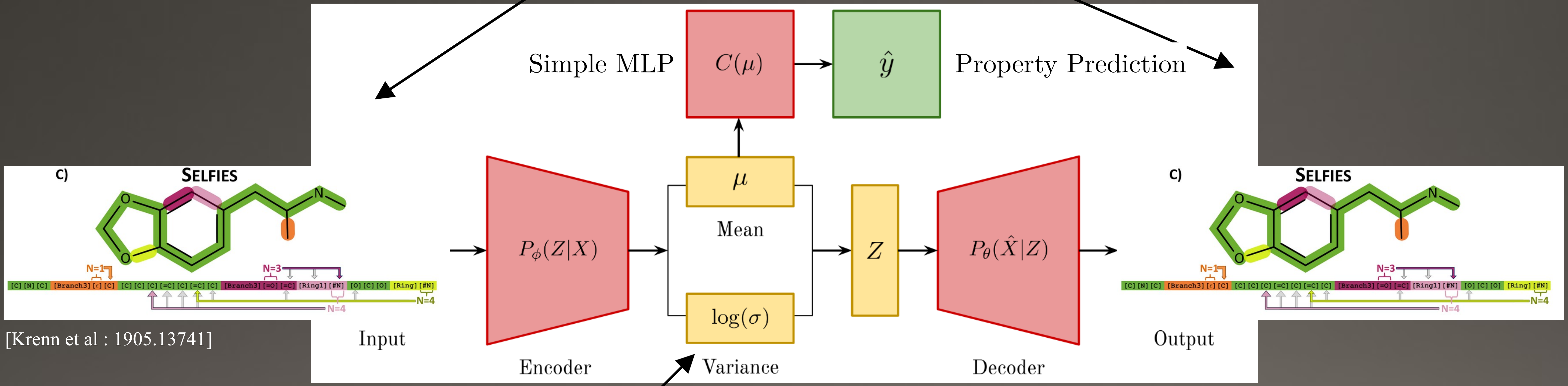


ML for DM Direct Detection

Molecular Space: Variational Autoencoders

Energies & Matrix elements Sample latent space → new molecules

Discrete and sparse data space



Continuous latent space

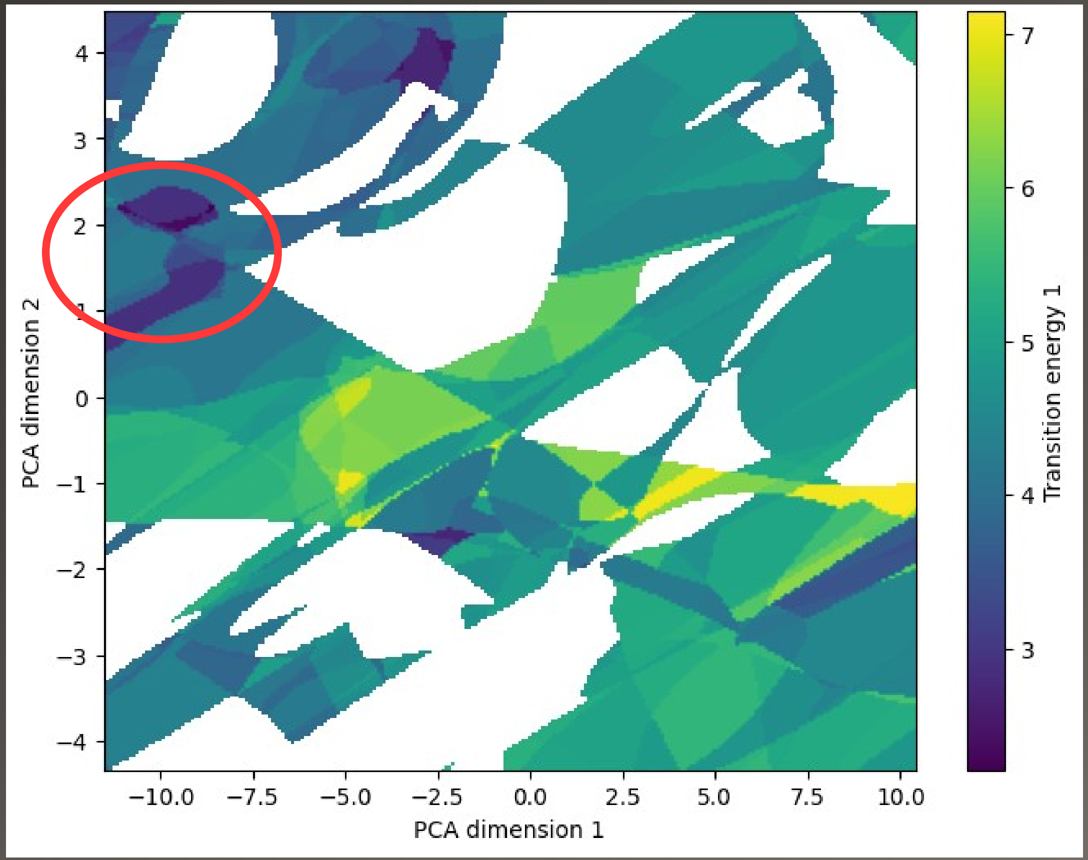
ML for DM Direct Detection

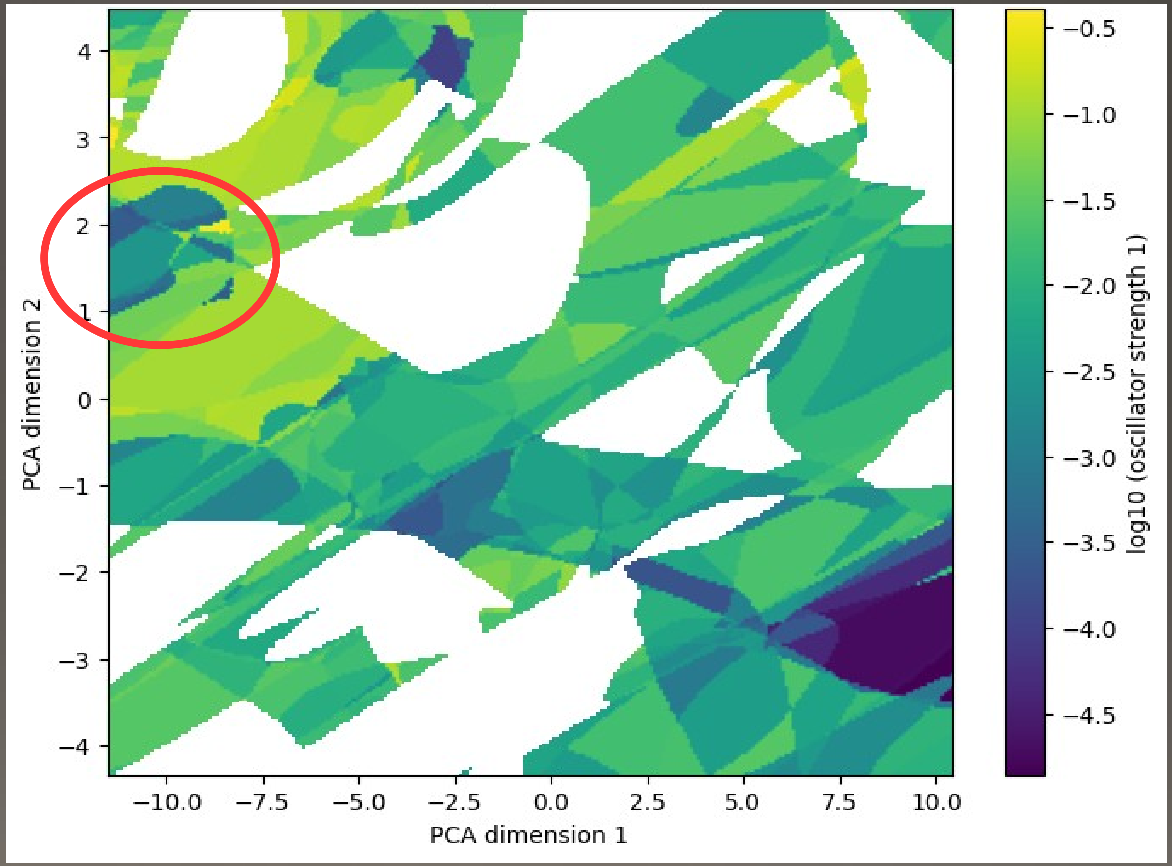
Property prediction: Molecular latent space

Energies & Matrix elements Sample latent space → new molecules

ΔE

$\langle r_{01} \rangle \approx |f_{0,1}|/q$





[C. Cook*, CB, J. Smirnov 2501.00091]

Clustering: Results

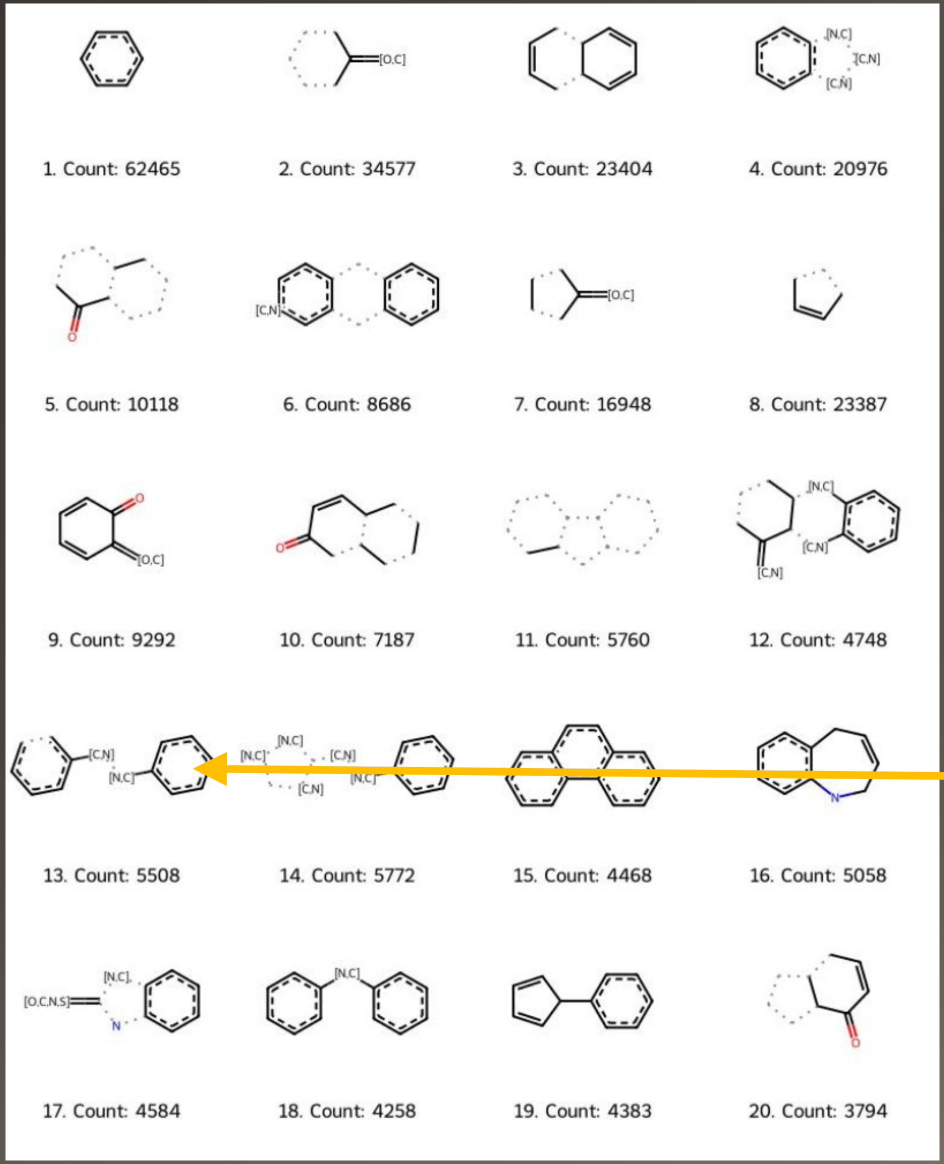
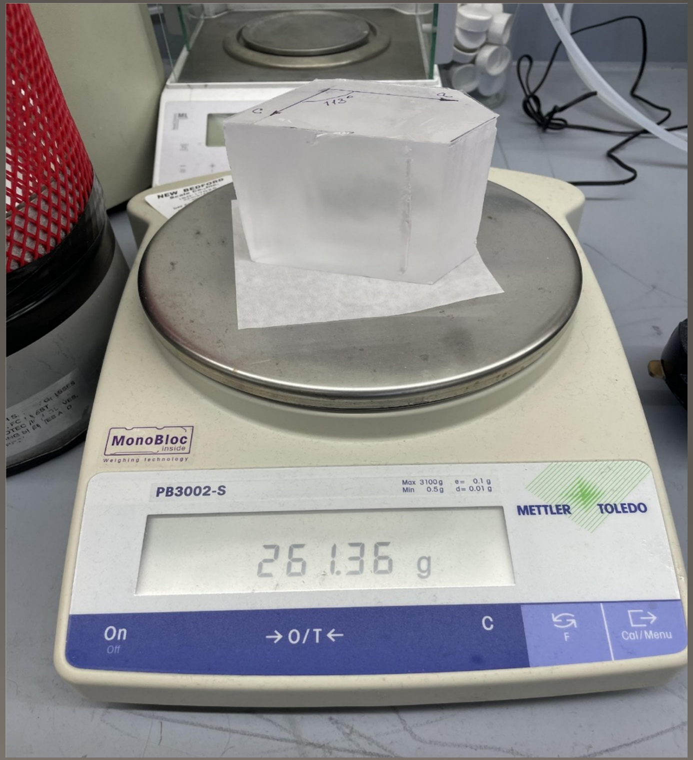


FIG. 7. Top 20 molecular motifs ranked by score. Note that certain atoms can be substituted by [C,N,O,S] as long as the structure remains isoelectronic. Dotted lines indicate that the bond could be a double or single bond.

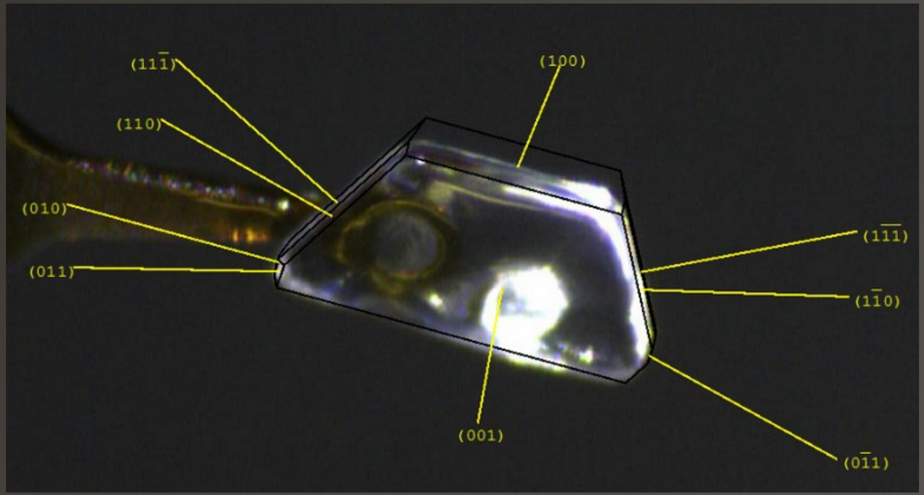
[C. Cook*, CB, J. Smirnov 2501.00091]

In hindsight, we're probably on the right track.



Characterizing t-stilbene

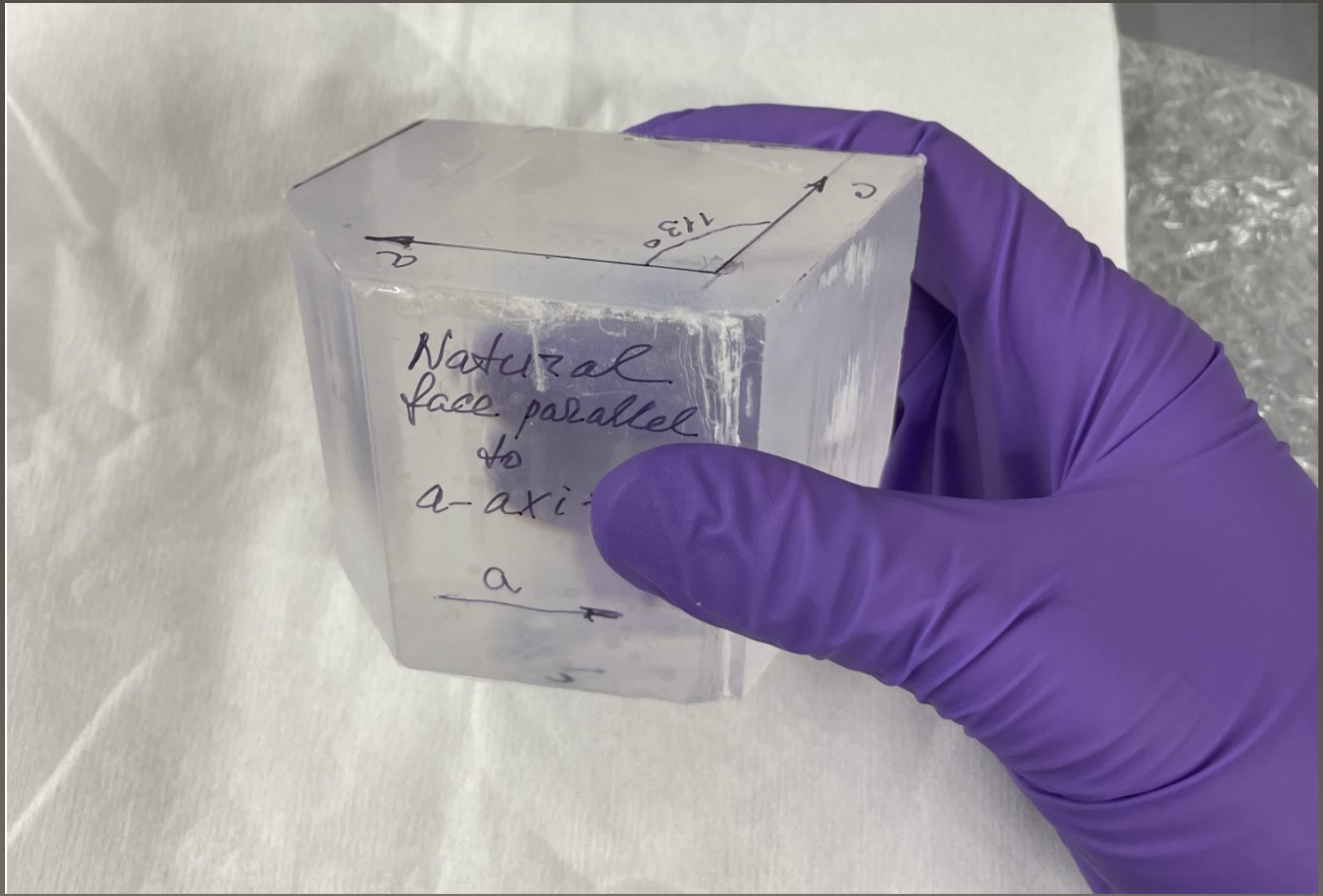
Calibrate at mg-scale



Credit: Dane Johnson
Freedman Group (MIT)

Natalia Zaitseva (LLNL)

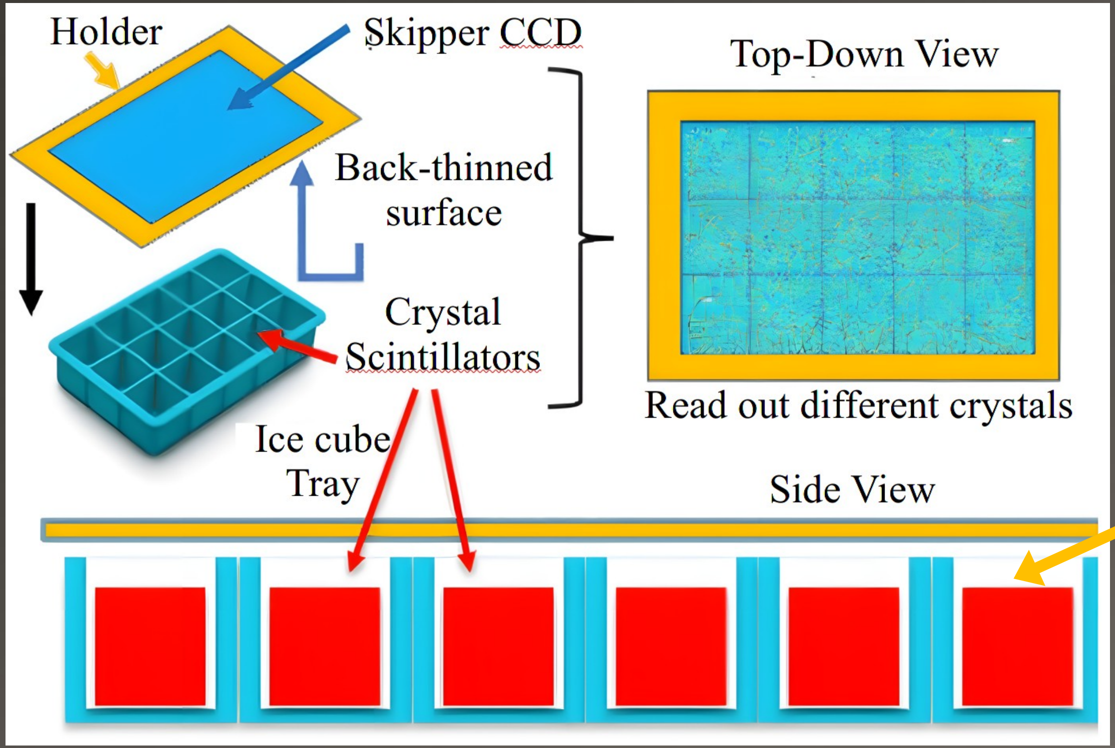
Deploy up to kg-scale

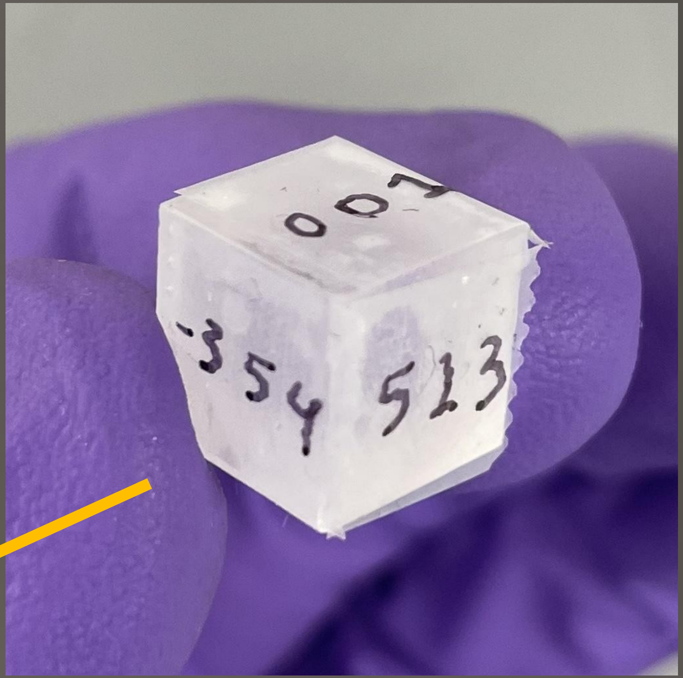


Experimental Deployment

DIANA*

Daily modulation in an Intrinsically ANisotropic Array



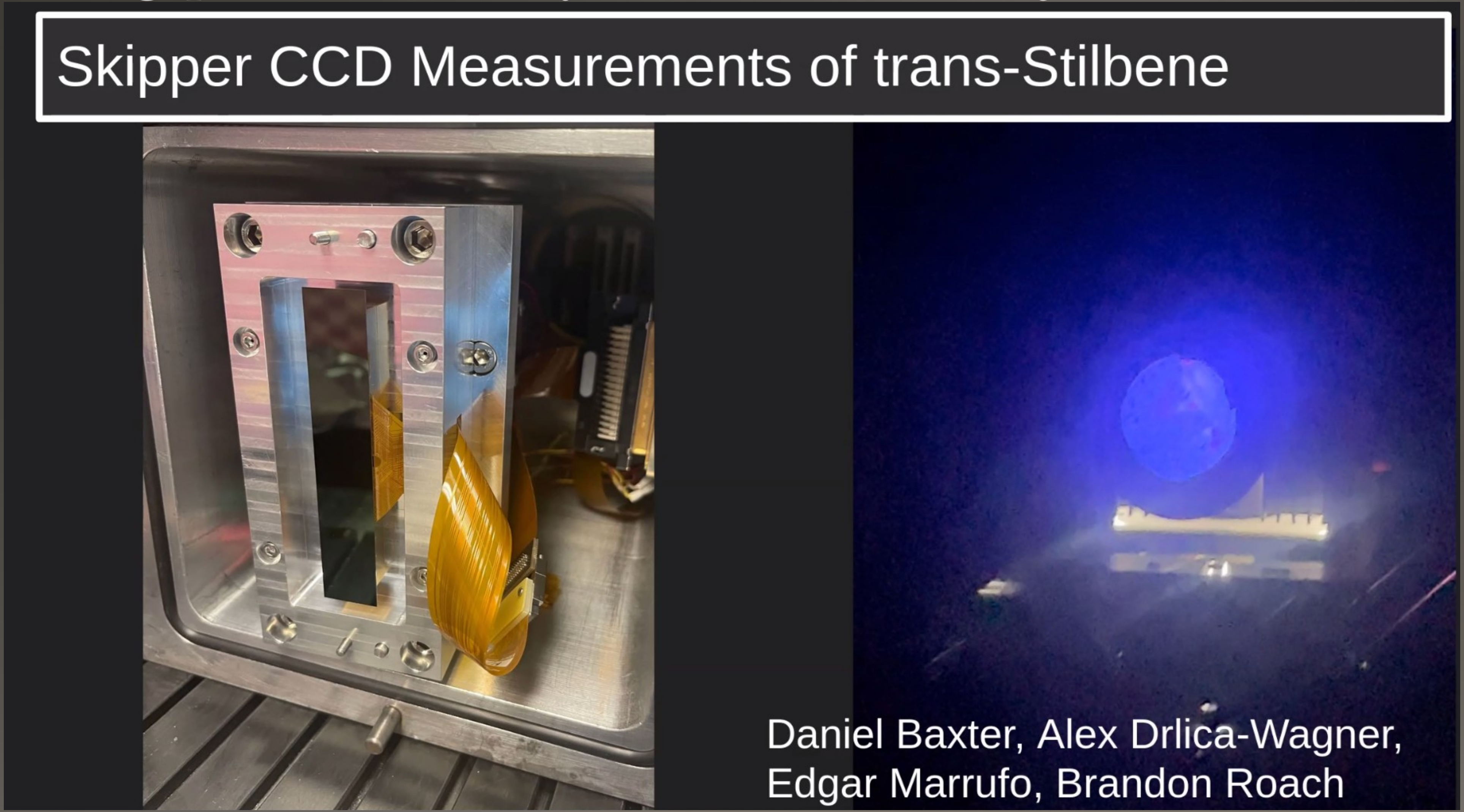


Many crystal samples can be read out by a single skipper CCD.

Collaboration: FermiLab, U. Toronto, MIT, UIUC, U. Oregon, Penn State

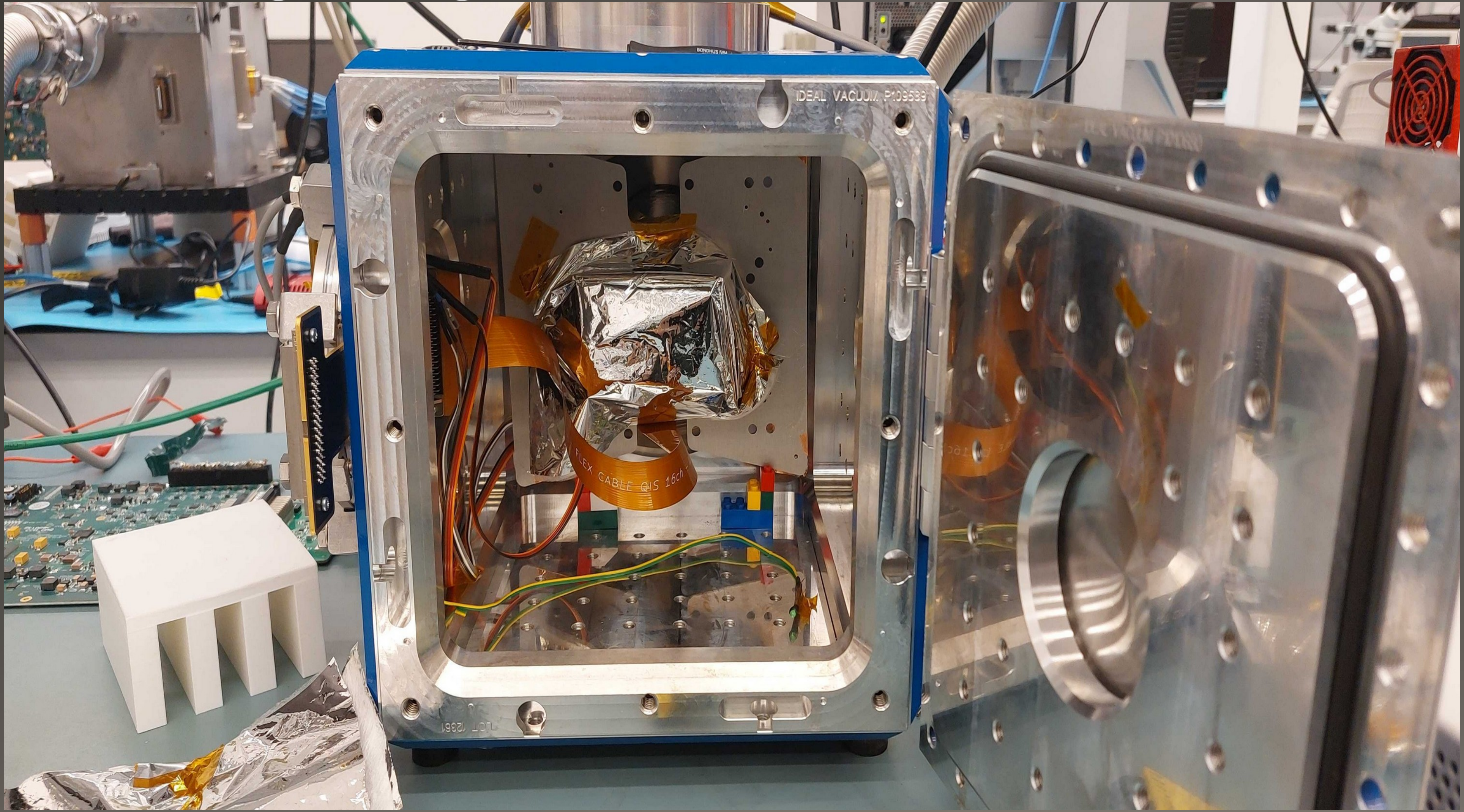
Astro-Skipper CCDs as Readout

Taking pictures of crystals with a *very* sensitive CCD



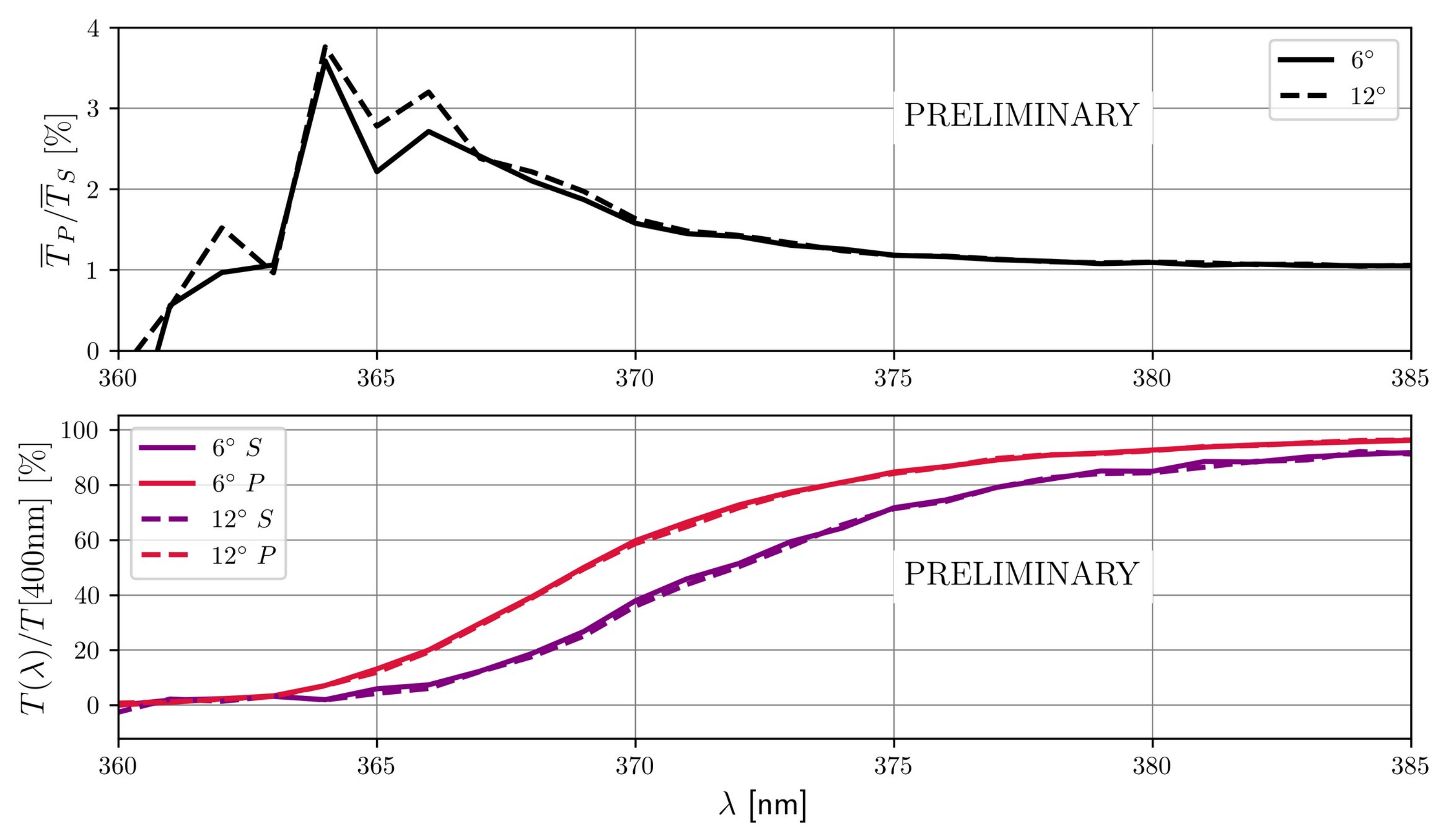
Astro-Skipper CCDs as Readout

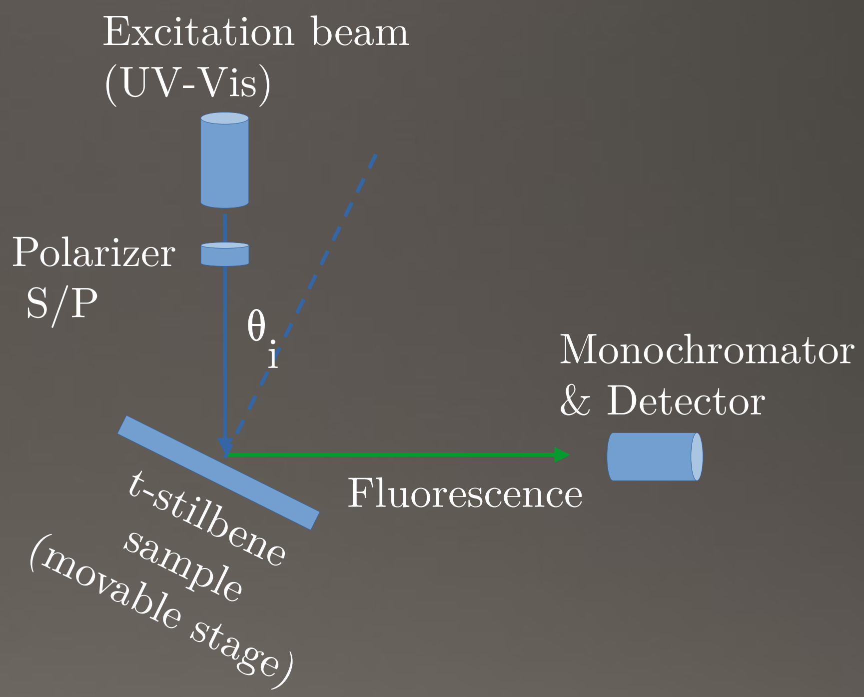
Now taking background measurements at Fermilab



Calibrating Anisotropic Materials

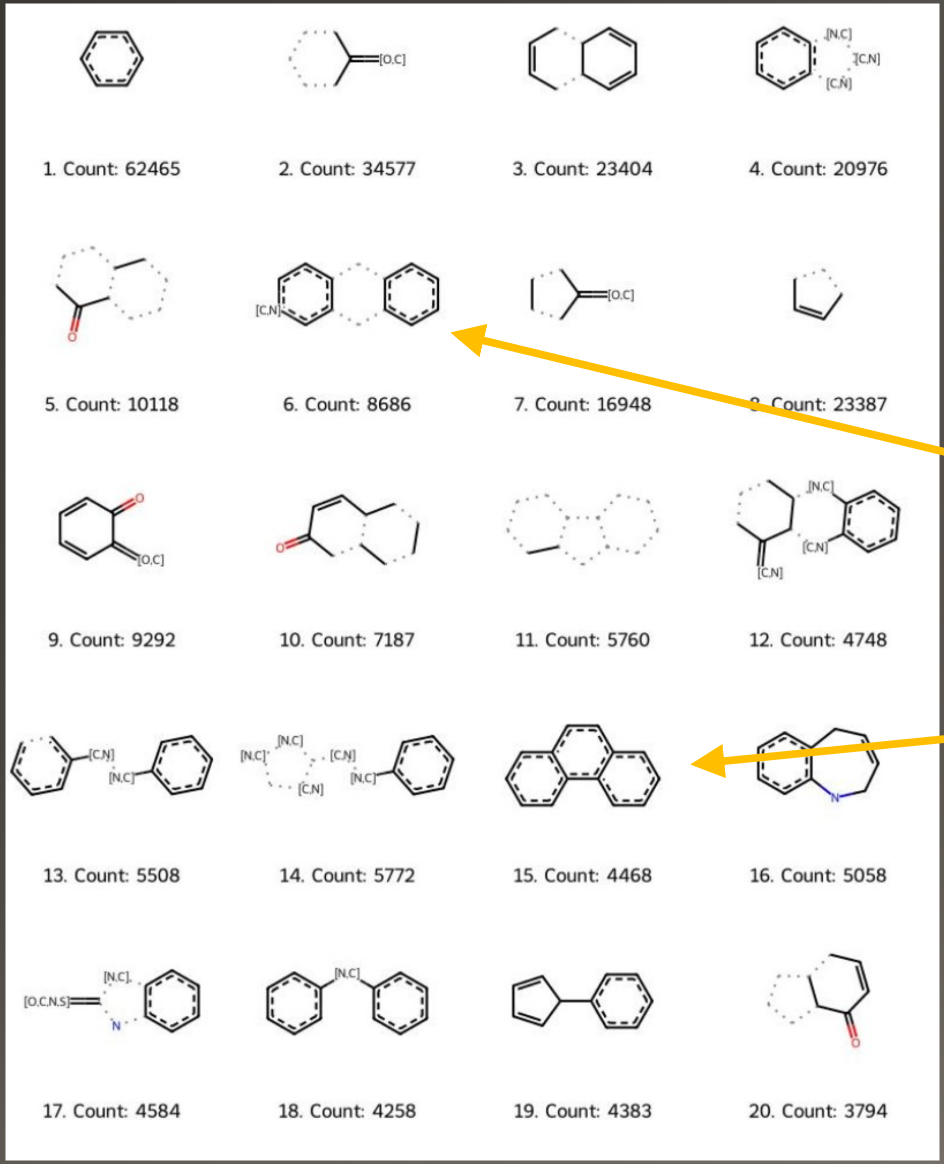
Calibration of Anisotropy via Photoabsorption

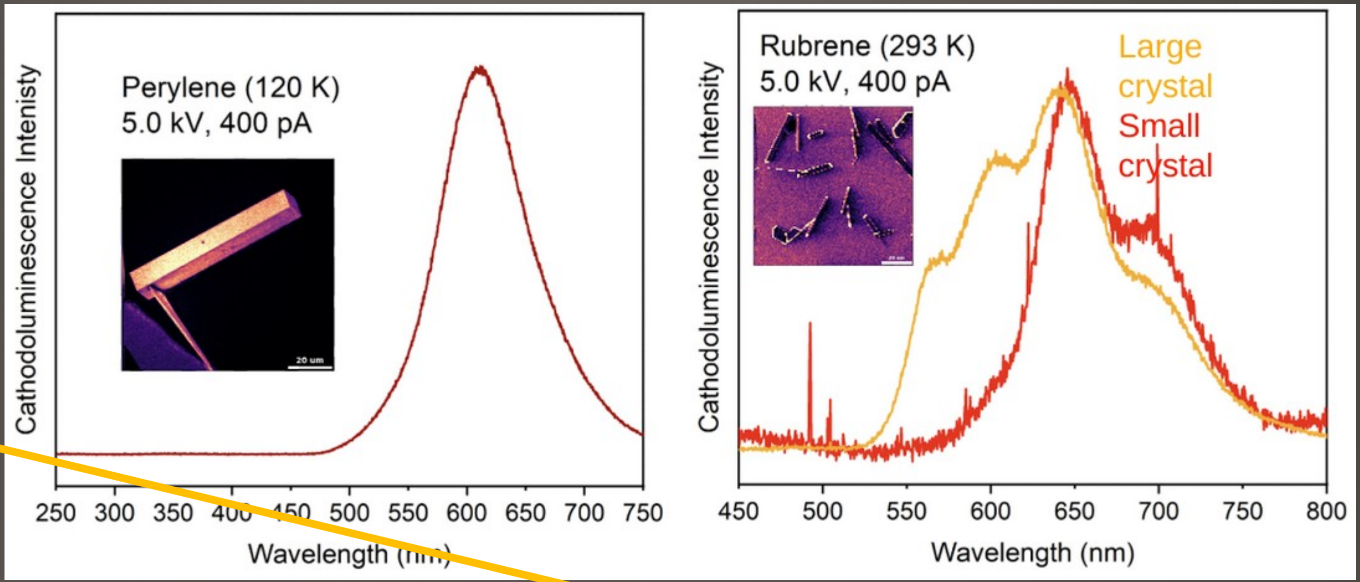


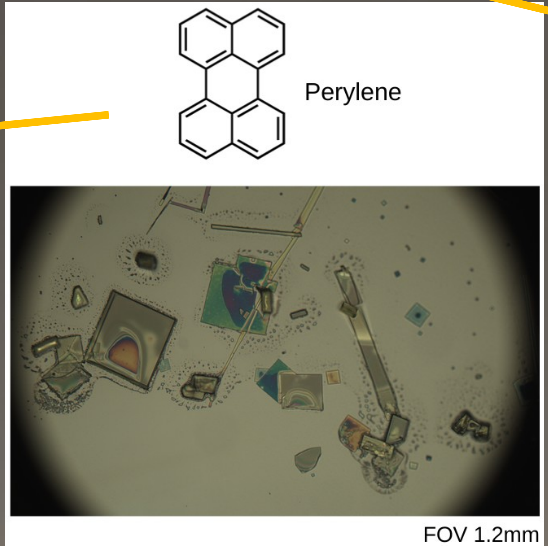


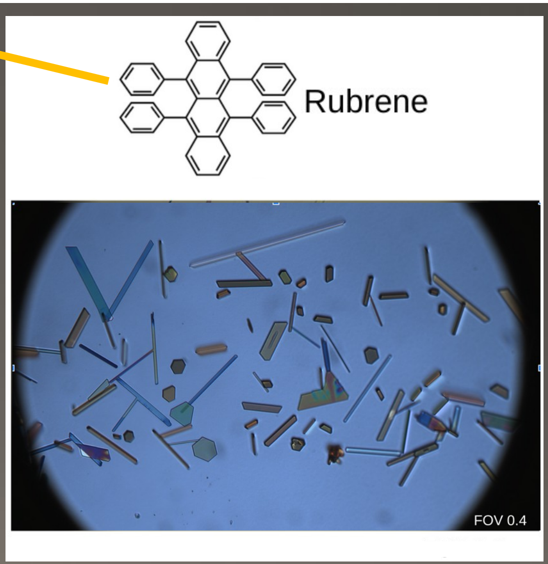
$A_{\mathrm{UV}} \sim n_{\mathrm{mol}} |\vec{\epsilon} \cdot \langle \Psi_f | \mathbf{r} | \Psi_i \rangle|^2$ $\qquad$ $A_{\mathrm{UV}} \approx \mathrm{const} - T_{\mathrm{UV}}$

New Target Materials





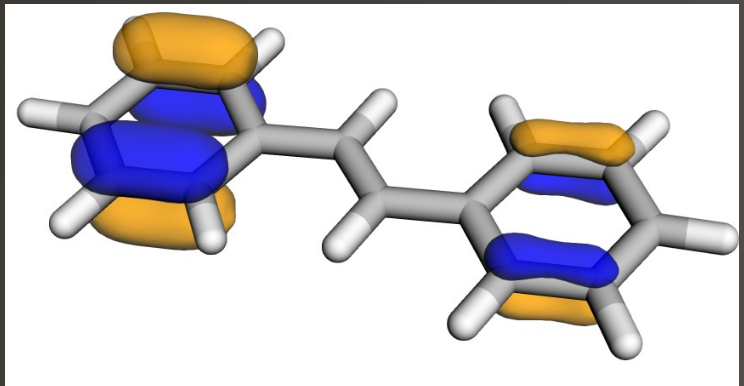




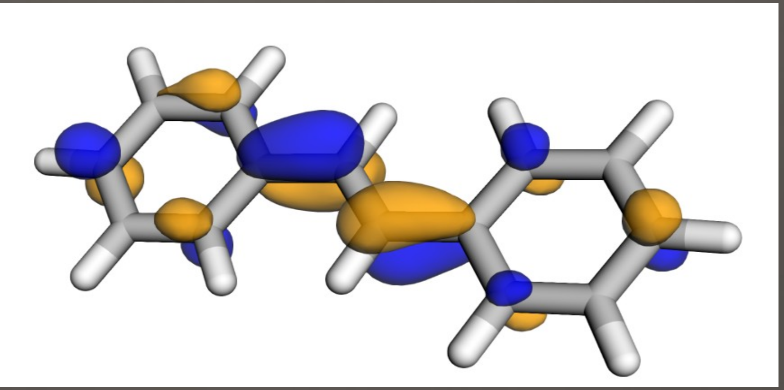
[C. Cook*, CB, J. Smirnov 2501.00091]

Credit: Dane Johnson (MIT)

Validating AI Predictions



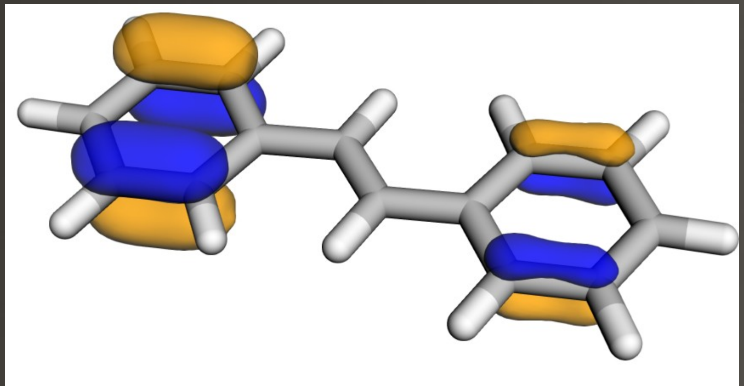
Ground State



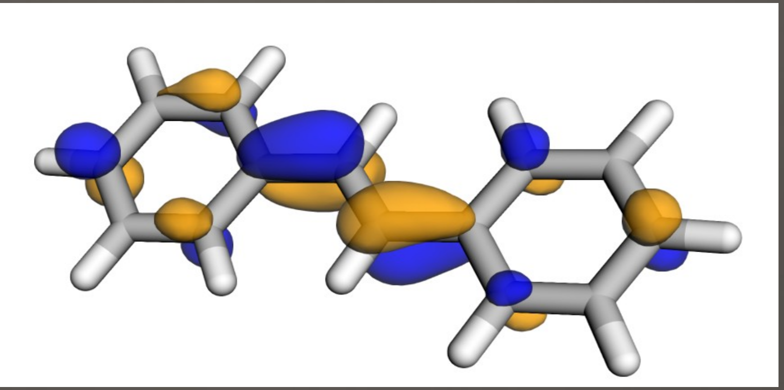
Excited State

Use time-dependent DFT to compute accurate many-body states

Validating AI Predictions

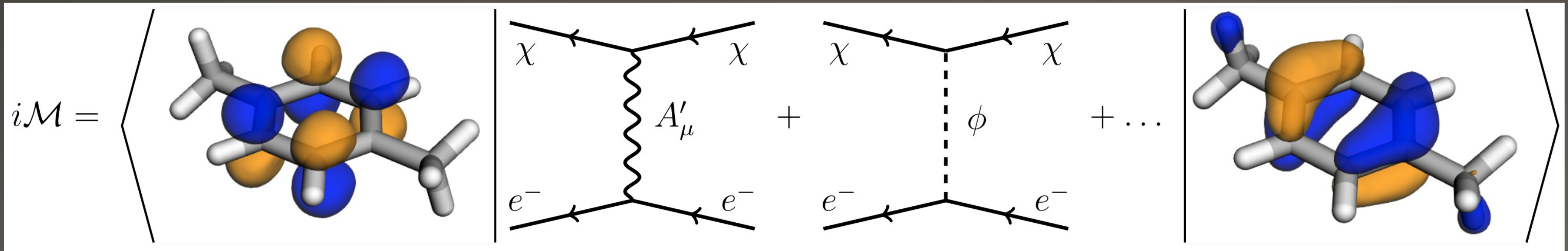


Ground State



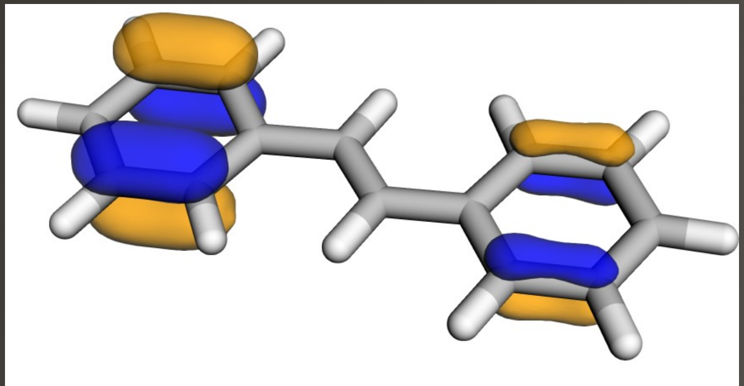
Excited State

Use time-dependent DFT to compute accurate many-body states

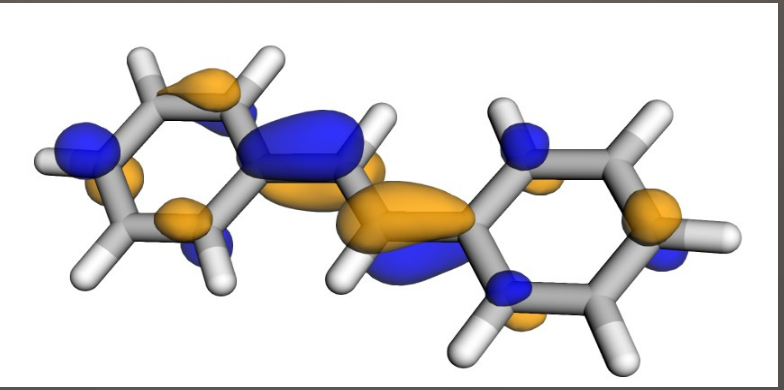


Then compute full matrix elements

Validating AI Predictions

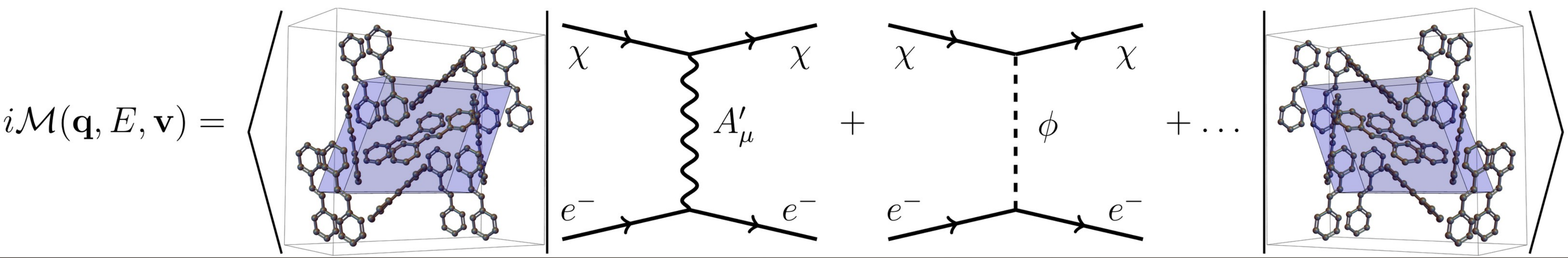


Ground State



Excited State

Use time-dependent DFT to compute accurate many-body states

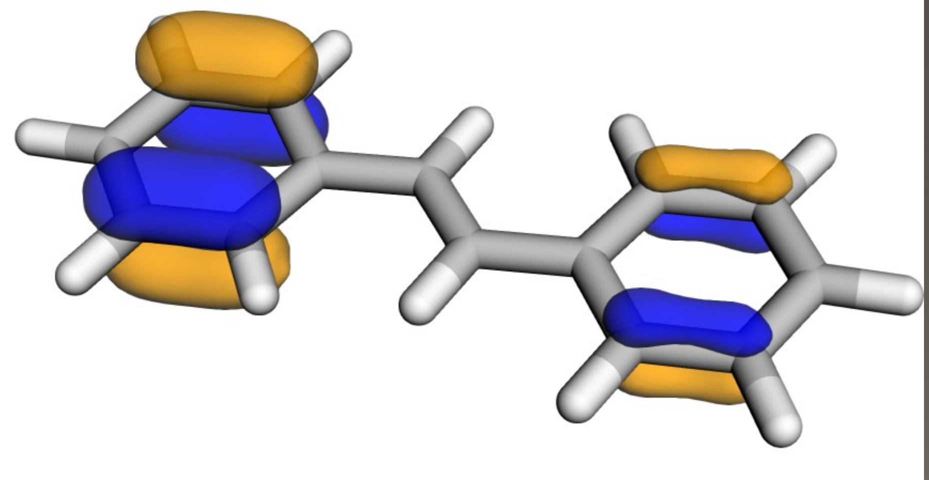


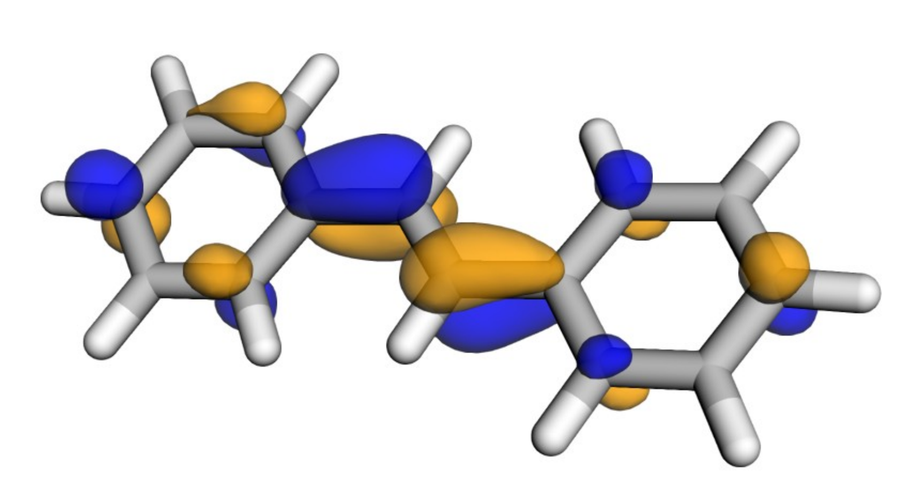
Then compute full matrix elements... accounting for crystal structure

Conclusions

1) We have done an extremely effective job looking for WIMPs above a GeV, now we must look beyond.

2) By developing the formalism that describes the interaction between dark matter and molecules, we can develop detection strategies capable of *delving deep* and *searching wide* across the dark matter parameter space.





ML for DM Direct Detection

Property prediction

Energies & Matrix elements

Step 1. Small molecules : Train on exhaustive database ($<$ 9 atoms)

Characterize neural nets & architecture

$\rightarrow$ Prove it's possible to learn from small subsample

Dataset of energies + oscillator strengths:

QM9: "All" 134K organic mols up to 9 heavy atoms - *computed* using DFT

[C. Cook*, CB, J. Smirnov 2501.00091]

ML for DM Direct Detection

Property prediction

Energies & Matrix elements

Step 1. Small molecules : Train on exhaustive database ($<$ 9 atoms)

Characterize neural nets & architecture

→ Prove it's possible to learn from small subsample

Dataset of energies + oscillator strengths:

QM9: "All" 134K organic mols up to 9 heavy atoms - *computed* using DFT

Step 2. Large molecules: Train on dataset up to 10s of atoms

Scale architecture

→ Generalize property prediction

Dataset of energies + oscillator strengths:

PubChemQ: "Sparse" ~3M organic molecules - *computed* using DFT

[C. Cook*, CB, J. Smirnov 2501.00091]

Regression Model in Action

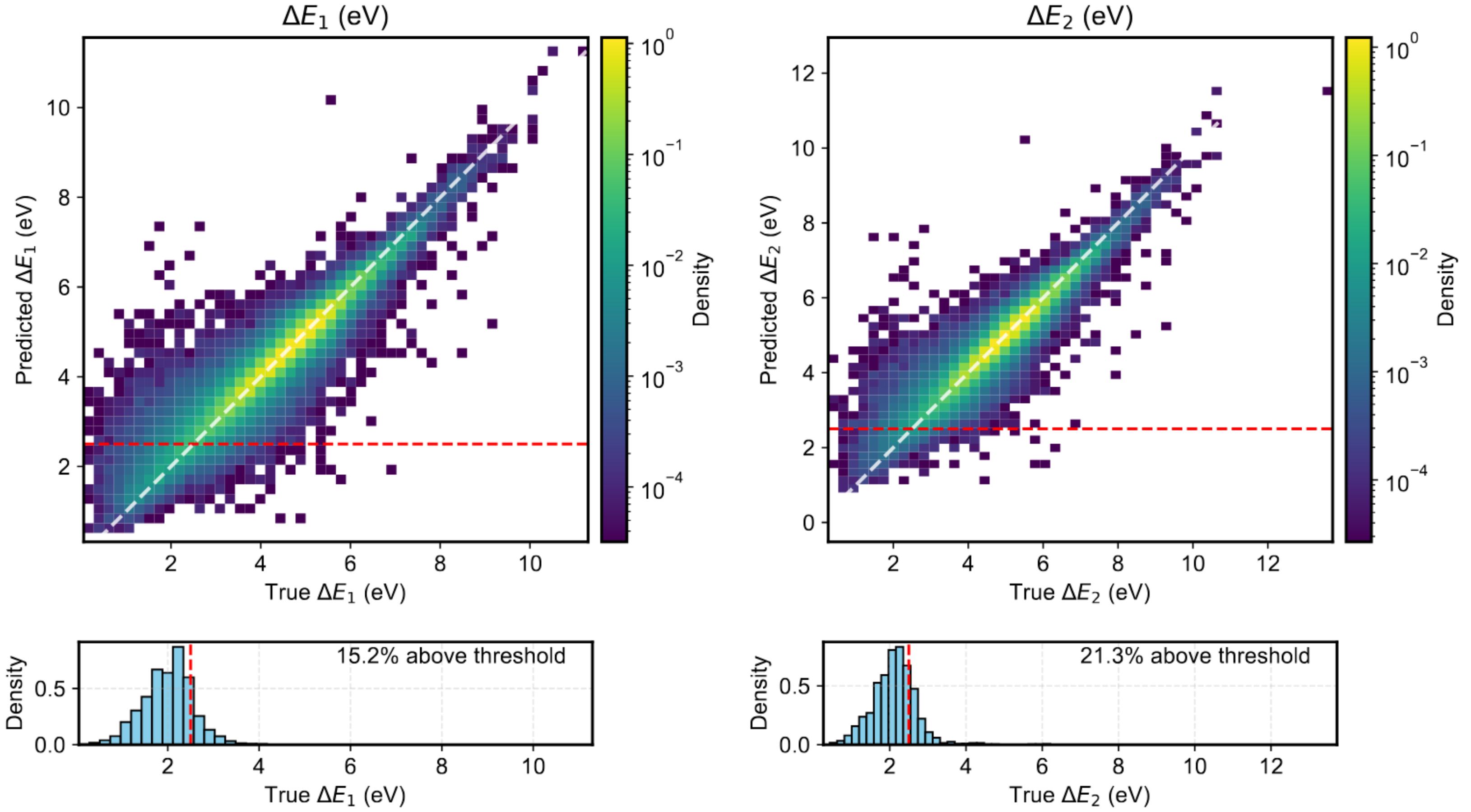


FIG. 19. We show the density of predicted values vs true values for the transition energies of our validation set, with a red line that corresponds to the threshold placed on the predicted value. The Histogram shows the distribution of true values for molecules whose predicted value is above the threshold.

[C. Cook*, CB, J. Smirnov 2501.00091]

Regression Model in Action

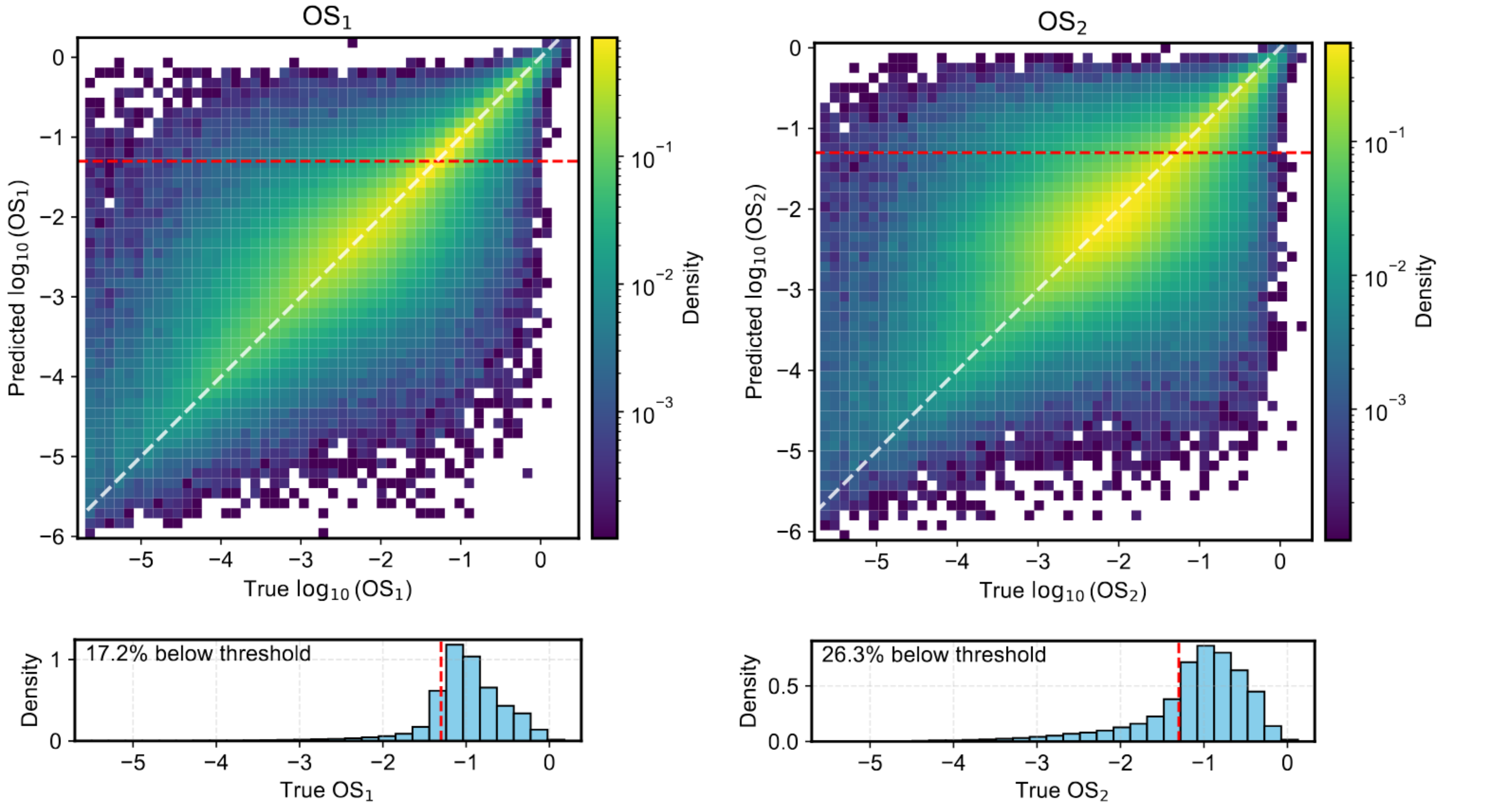


FIG. 20. We show the density of predicted values vs true values for the oscillator strength of our validation set, with a red line that corresponds to the threshold placed on the predicted value. The Histogram shows the distribution of true values for molecules whose predicted value is above the threshold.

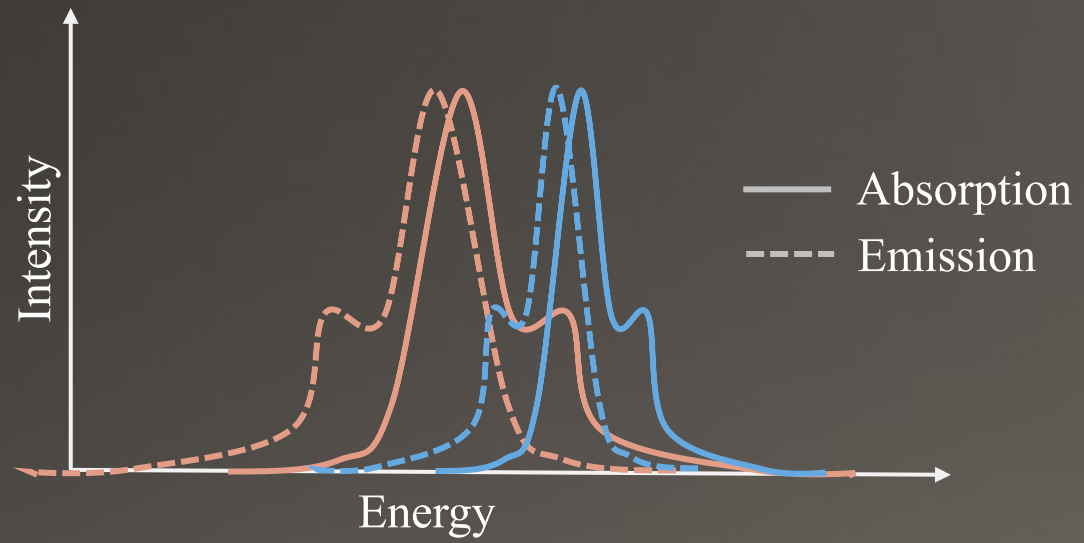
[C. Cook*, CB, J. Smirnov 2501.00091]

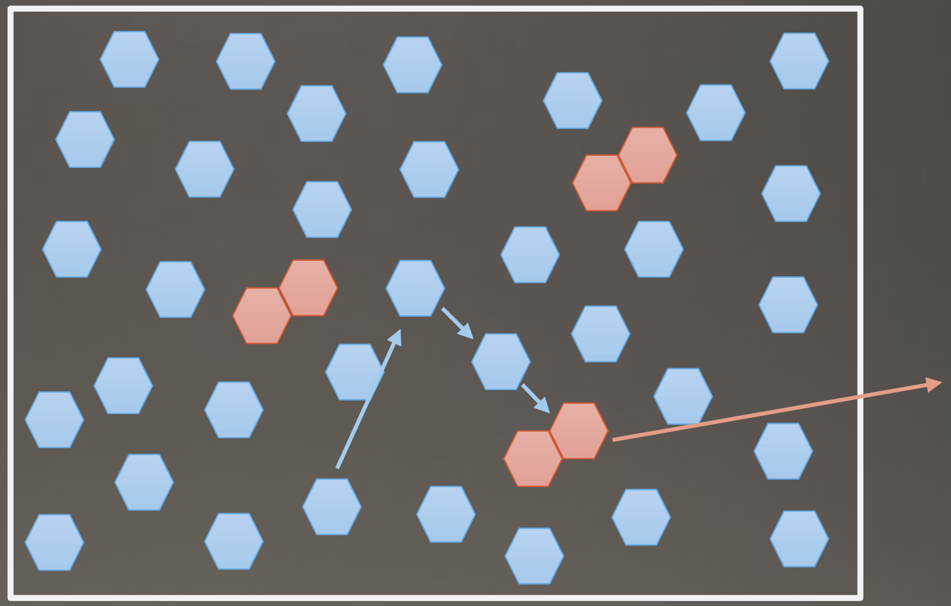
arXiv: 1912.02822

Fluorescence: Binary Scintillators

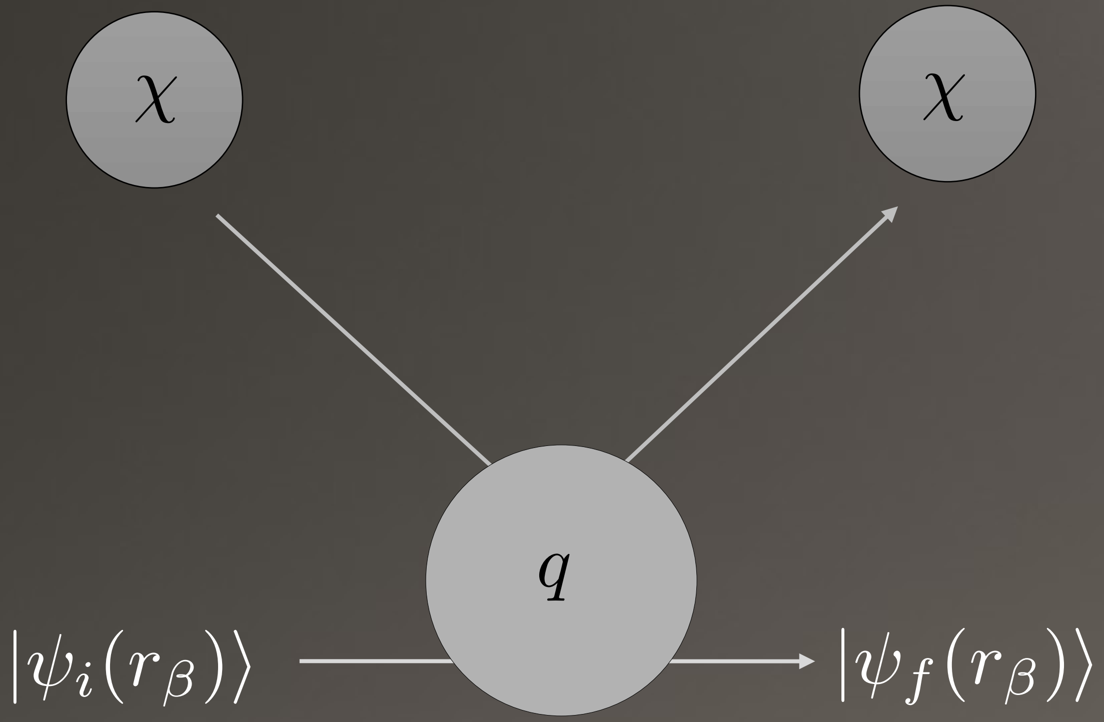
Solvent: Primary target starts the signal

Solute: Dilute fluor gets the signal out of the bulk





Electron Recoil: Charge Signal



Electron scattering

$$\Delta E_r = (m_\chi^2/m_{\mathrm{T}}) \times 10^{-6}$$

$$\Delta E \sim \mathcal{O}(\text{few eV}) \left(\frac{m_\chi}{1\,\text{MeV}}\right)^2$$

What has such transition energies?

- Semiconductor band gaps
- Maybe atomic ionization

Electrons in crystals (exciton generation)

$$|\psi_i\rangle \sim u_v(r)e^{ik'\cdot r} \qquad |\psi_f\rangle \sim u_c(r)e^{ik\cdot r}$$

Electrons in atoms (ionization)

$$|\psi_i\rangle \sim \psi_{\mathrm{STO}}(r_\beta) \qquad |\psi_f\rangle \sim e^{ik\cdot r},\ r \gg a_0$$

Trans-Stilbene

s	Platt Symbol	Symmetry	ΔE [eV]	Configuration amplitudes			
s_1	^1B	B_u	4.240	$d_{7,8} = 0.94,$	$d_{4,11} = -0.24$		
s_2	$^1G^-$	B_u	4.788	$d_{7,10} = 0.53,$	$d_{5,8} = 0.53,$	$d_{6,11} = 0.37,$	$d_{4,9} = -0.37$
s_3	$^1G^-$	A_g	4.800	$d_{7,9} = 0.53,$	$d_{6,8} = 0.53,$	$d_{5,11} = 0.37,$	$d_{4,10} = -0.37$
s_4	$^1(C,H)^+$	A_g	5.137	$d_{7,11} = 0.41,$	$d_{5,9} = -0.41,$	$d_{6,10} = -0.41,$	$d_{4,8} = -0.59$
s_5	$^1H^+$	B_u	5.791	$d_{5,10} = 0.54,$	$d_{6,9} = 0.54,$	$d_{7,12} = 0.33,$	$d_{3,8} = 0.33$
s_6	$^1G^+$	A_g	6.264	$d_{7,9} = 0.68,$	$d_{6,8} = -0.68$		
s_7	$^1C^-$	A_g	6.013	$d_{7,11} = 0.66,$	$d_{4,8} = 0.54,$		
s_8	$^1G^+$	B_u	6.439	$d_{7,10} = 0.65,$	$d_{5,8} = -0.65$		

Table 1: The first eight excited states $s_{n=1\ldots8}$, with their energy eigenvalues $\Delta E(s_n)$ with respect to the ground state and coefficients $d_{ij}^{(n)}$ as calculated by Ting and McClure.

$$|s_n\rangle = \sum_{i,j>i} d_{ij}^{(n)} |\psi_i^j\rangle,$$

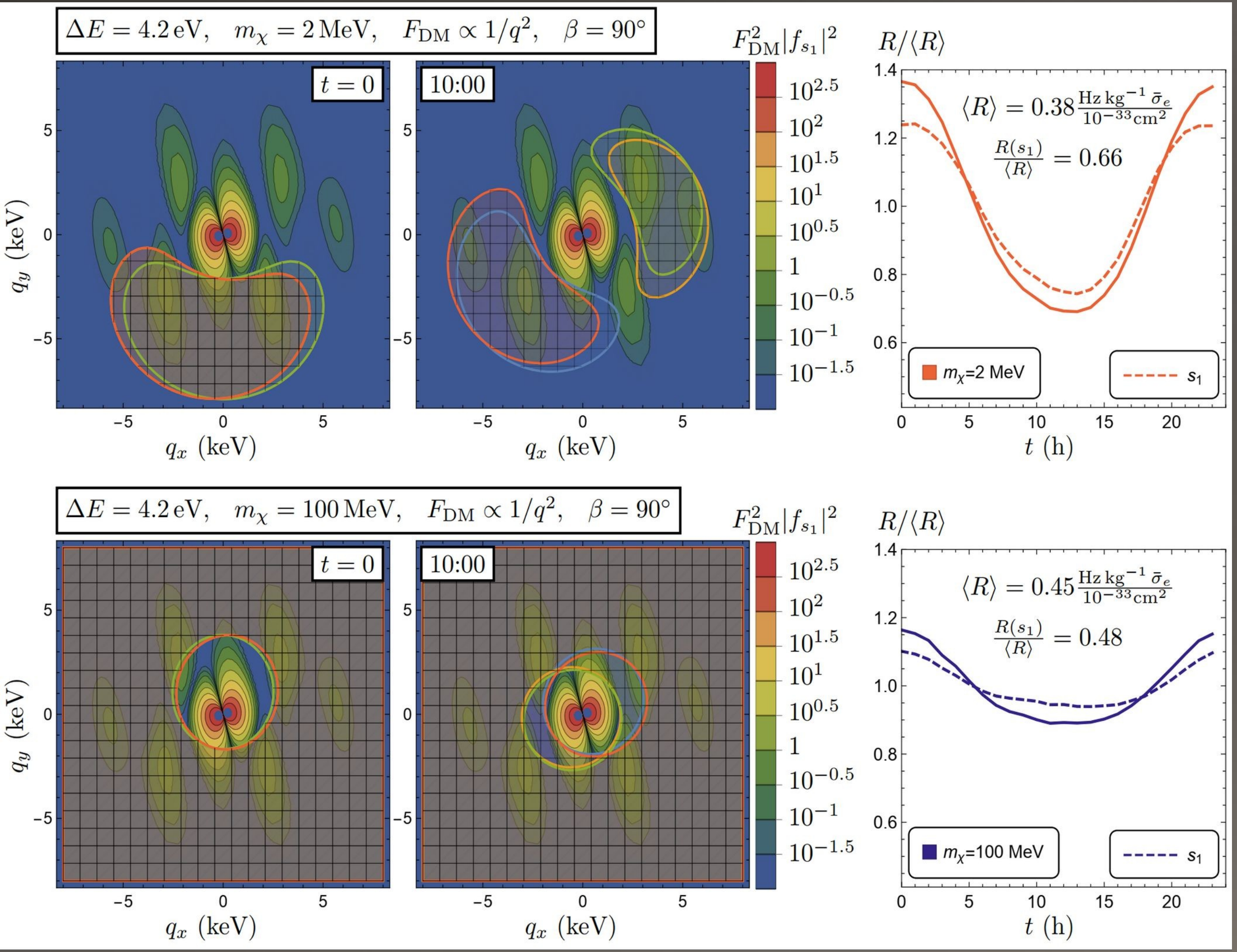
$$\sum_{ij} |d_{ij}^{(n)}|^2 = 1.$$

$$f_{g\to s_n}(\vec{q}) = \left\langle \psi_{s_n}(\vec{r}_1 \ldots \vec{r}_{14}) \left| \sum_{m=1}^{14} e^{i\vec{q}\cdot\vec{r}_m} \right| \psi_G(\vec{r}_1 \ldots \vec{r}_{14}) \right\rangle$$

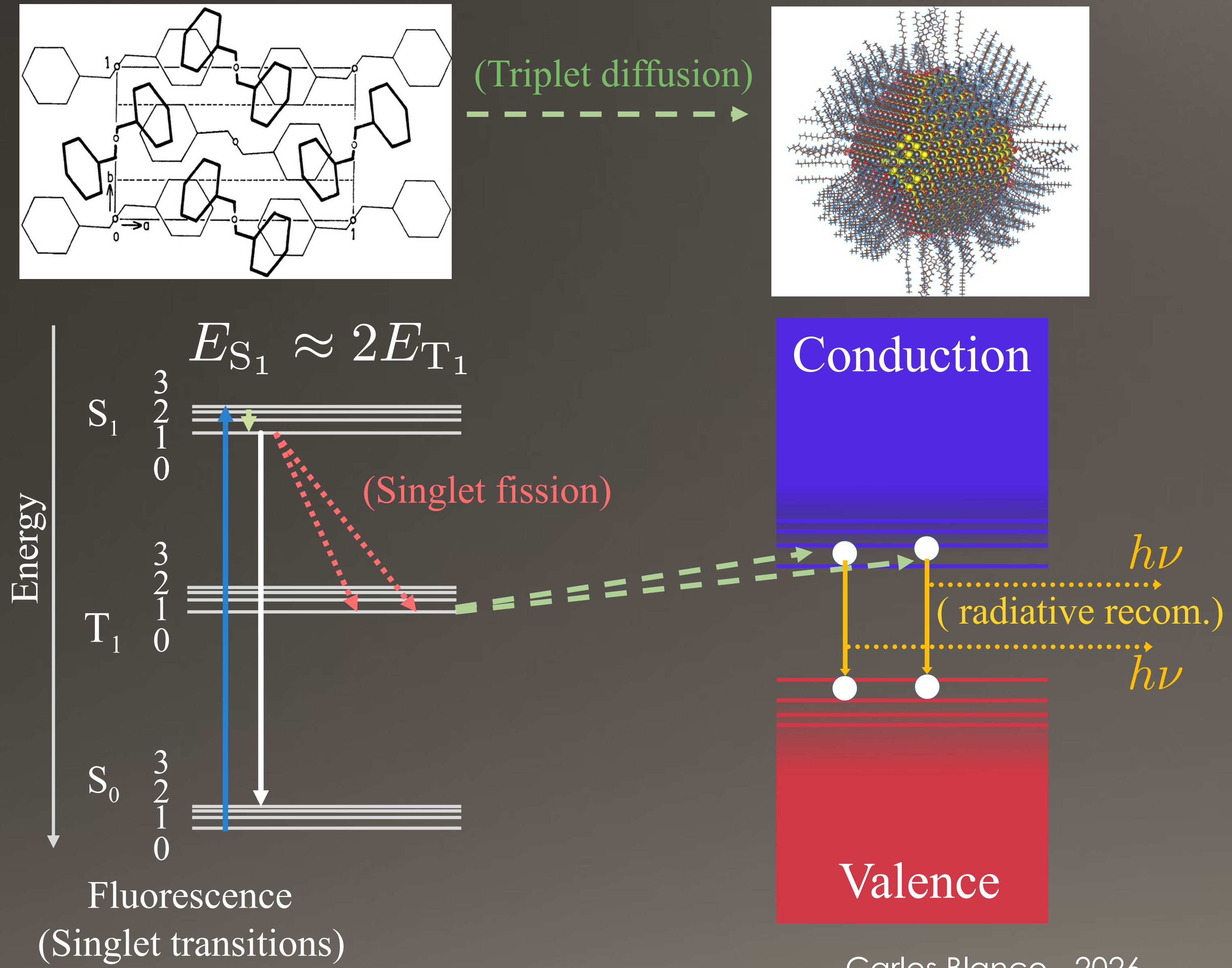
$$= \sum_{ij} d_{ij}^{(n)} \left\langle \psi_i^j \left| e^{i\vec{q}\cdot\vec{r}} \right| \psi_G \right\rangle$$

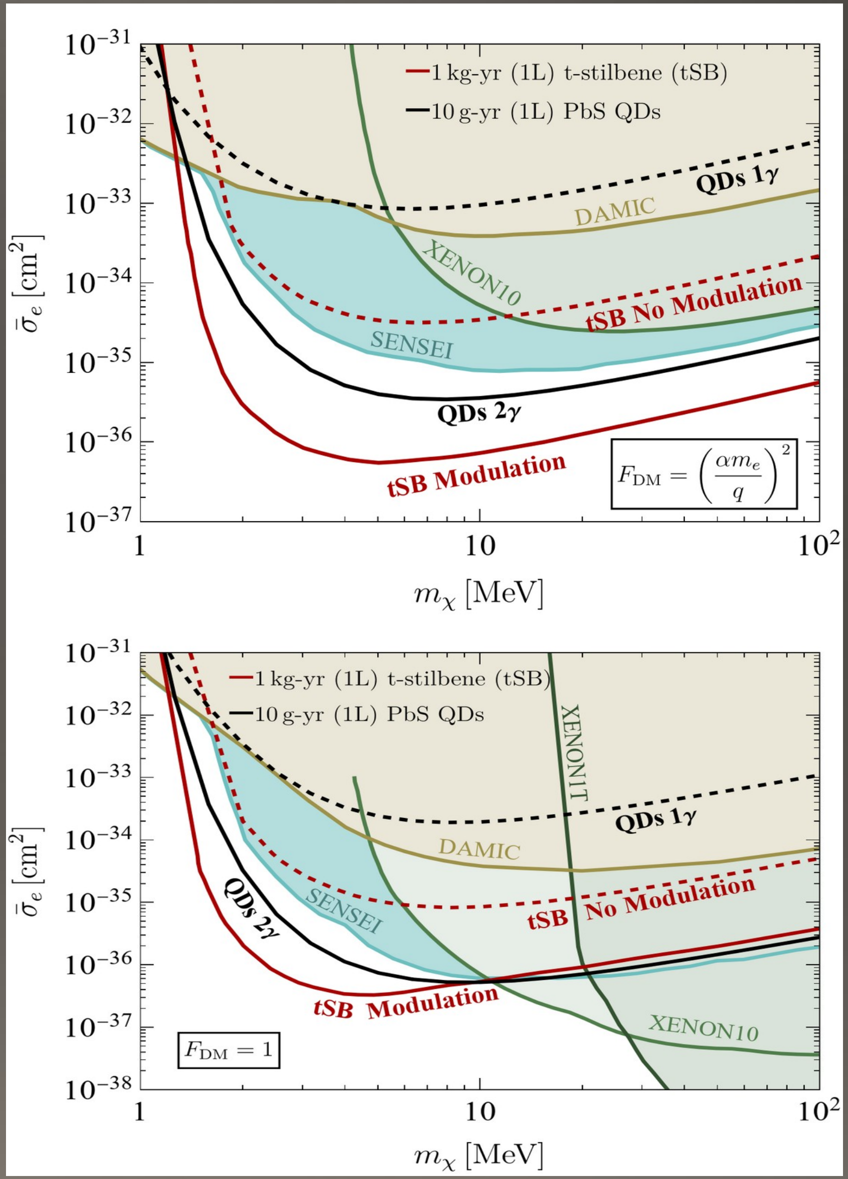
$$= \sqrt{2} \sum_{ij} d_{ij}^{(n)} \langle \Psi_j(\vec{r}) | e^{i\vec{q}\cdot\vec{r}} | \Psi_i(\vec{r}) \rangle.$$

Daily Modulation: Light Mediator

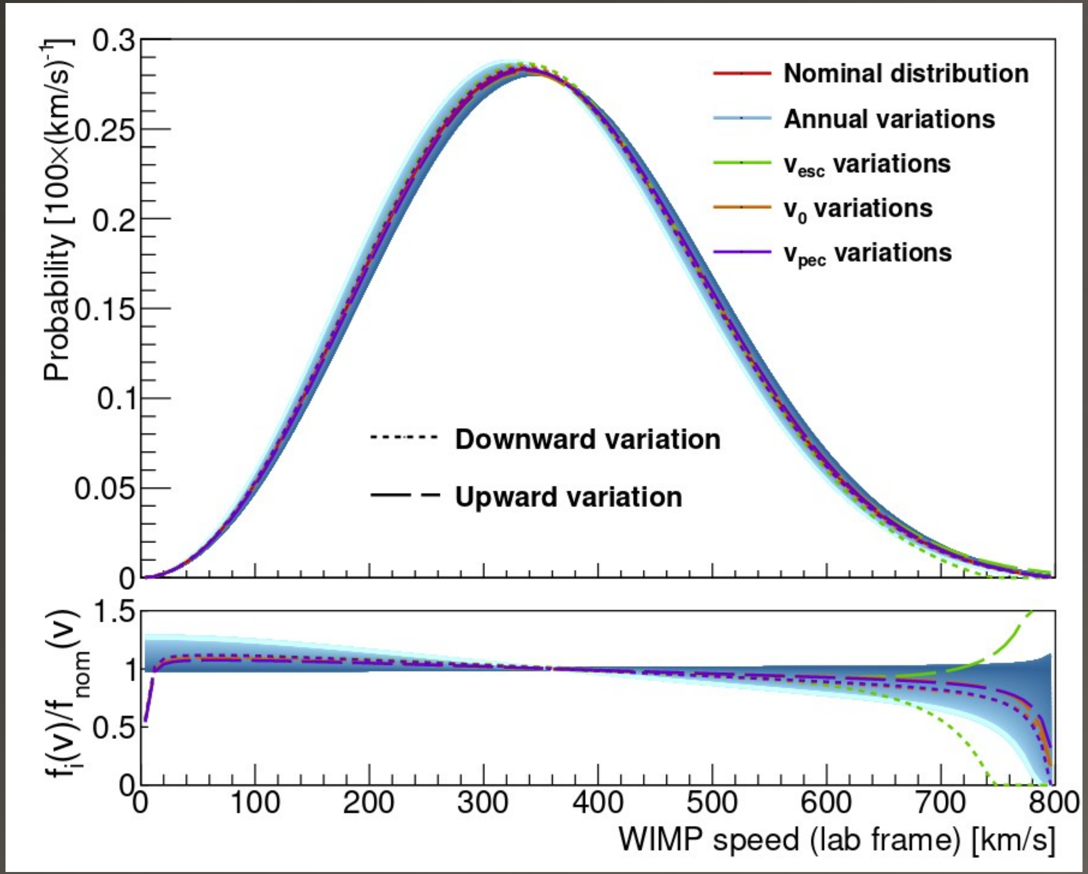


Same as previous figures (top) for a light mediator DM form factor $F_{\text{DM}} = (\alpha m_e/q)^2$. Here, the contour plots show $F^2_{\text{DM}}|f(s_1)|^2$ which appears in the rate integrand; the scattering is dominated by the smallest kinematically-allowed q. **Top:** Molecular form factors with $q_z = 0$ and rate modulations for $m_\chi = 2$ MeV. **Bottom:** Molecular form factors with $q_z = 0$ and rate modulations for $m_\chi = 100$ MeV.

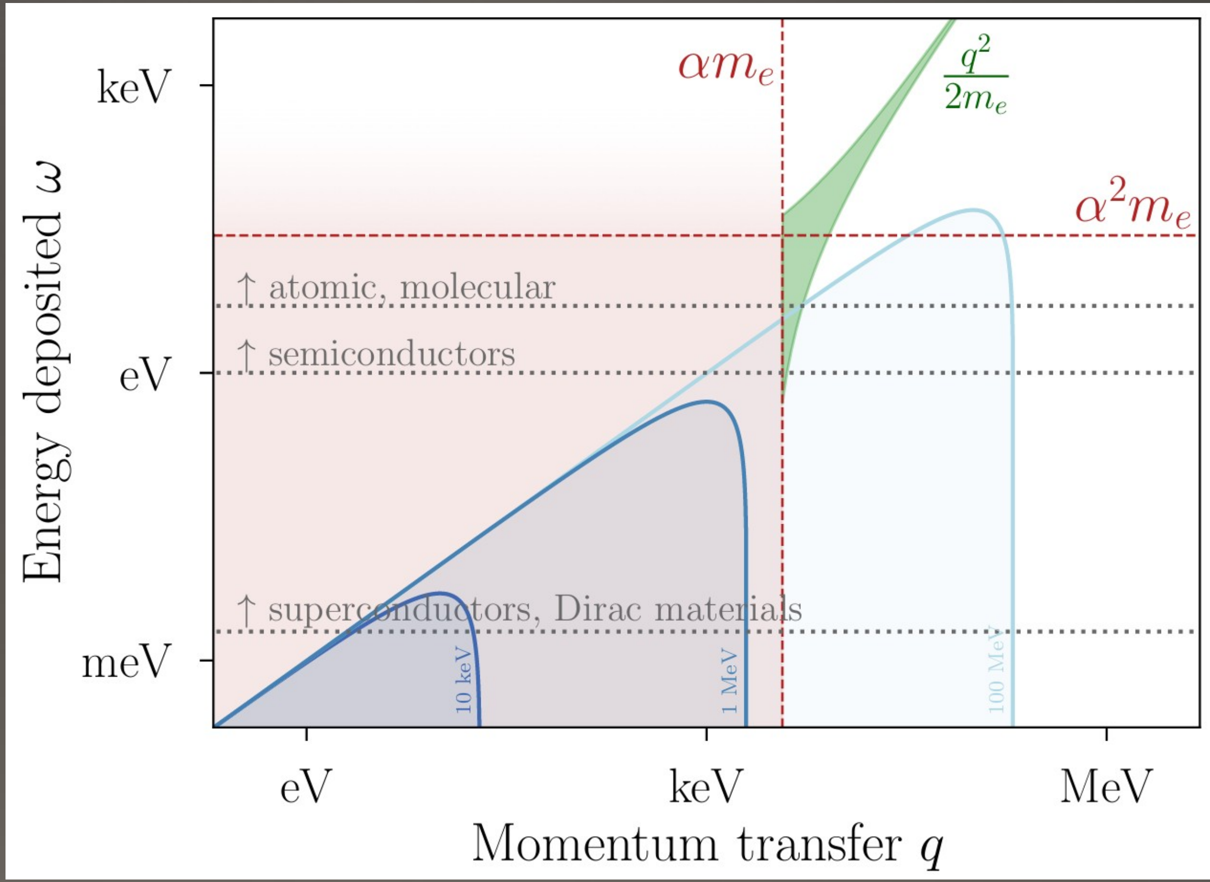

(Triplet diffusion)
$E_{S_1} \approx 2E_{T_1}$
Energy
S_1
T_1
S_0
3
2
1
0
(Singlet fission)
Fluorescence
(Singlet transitions)
Conduction
Valence
$h\nu$
(radiative recom.)
$h\nu$


1 kg-yr (1L) t-stilbene (tSB)
10 g-yr (1L) PbS QDs
QDs 1γ
DAMIC
XENON10
tSB No Modulation
SENSEI
QDs 2γ
tSB Modulation
$F_{DM} = \left(\frac{\alpha m_e}{q}\right)^2$
$\bar{\sigma}_e$ [cm^2]
m_χ [MeV]
1 kg-yr (1L) t-stilbene (tSB)
10 g-yr (1L) PbS QDs
XENON1T
QDs 1γ
DAMIC
tSB No Modulation
QDs 2γ
SENSEI
tSB Modulation
XENON10
$F_{DM} = 1$
$\bar{\sigma}_e$ [cm^2]
m_χ [MeV]

Local DM Phase Space



Baxter, D., et al. "*Recommended conventions for reporting results from direct dark matter searches*." The European Physical Journal C 81.10 (2021): 1-19.



Lin, Tongyan. "*Sub-GeV dark matter models and direct detection*." SciPost Physics Lecture Notes (2022): 043.