

PDFgui: A Practical Platform for Small-Box Modeling and Real-Space Structure Refinement of PDF Data

Emil S. Bozin

Institute of Physics Belgrade

29 September 2026 · 20:30 · 1 h · Business Lounge, CC Academia

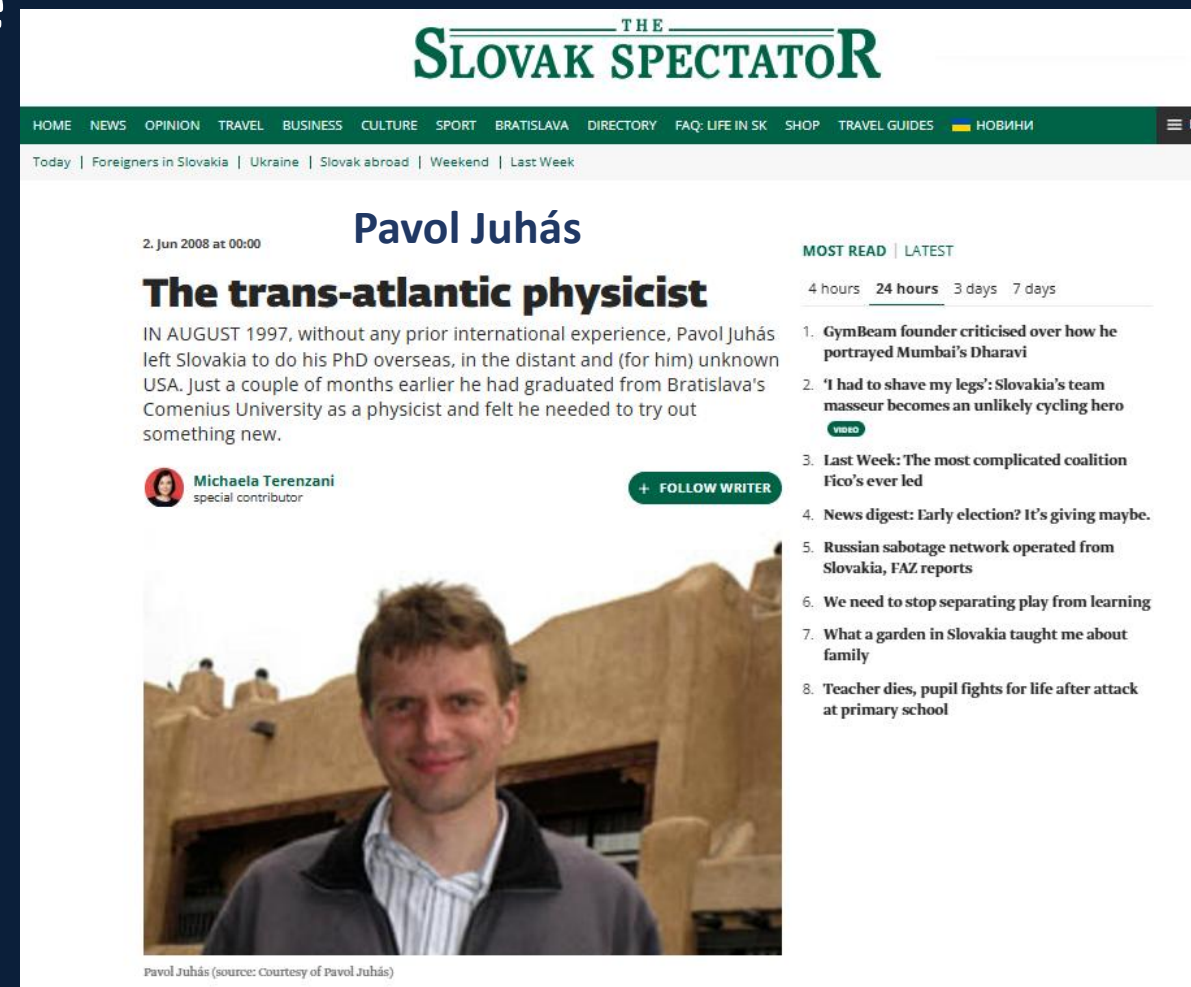


PDFgui: A Practical Platform for Small-Box Modeling and Real-Space Structure Refinement of PDF Data

Emil S. Bozin

Institute of Physics Belgrade

29 September 2026 · 20:30 · 1 h · Business Lounge, CC Academia



Four aspects of small box modeling

01

Small-box modelling

What PDFgui is for — and what it is not

02

What the calculator knows

Thermal, $Q_{\max}$, Q_{damp} , size, correlated motion

03

Parameters and project structure

Datasets, phases, handles, wizards

04

A hidden local state

AgGaTe₂ — NTE as a nanoscale distortion

Where small-box PDF lives

Non-crystalline

liquids, glasses,
polymers

Nanoscale

particles, cores,
interfaces

Disordered crystals

nanoscale
heterogeneities

S.J.L. Billinge & I. Levin, Science 316, 561 (2007) — the average cell is not always the structure that functions.

Molecule → cluster → nanoparticle → nano-domain crystal → periodic crystal



PDF keeps the diffuse intensity on equal footing with Bragg. Small-box modelling asks: how far can one (or a few) unit cells, plus ADPs and an envelope, explain $G(r)$ as a function of r ?

Three levels of model — climb them in order

Small box

PDFgui / DiffPy-CMI / TOPAS

Up to a few hundred atoms.
Rietveld-type parameters.
Least squares vs $G(r)$.
Fit as a function of r -range.

Large box

RMCProfile / EPSR / DISCUS

20 000+ atoms.
Fit $F(Q)$, $G(r)$, Bragg.
A snapshot, not a free-energy
minimum. Constrain hard.

Higher-level / ab initio

DiffPy-CMI, DFT, MD

Diameter, stacking-fault
probability, layer spacing.
Or forces instead of free atoms.

Three levels of model — climb them in order

Small box

PDFgui / DiffPy-CMI / TOPAS

Up to a few hundred atoms.
Rietveld-type parameters.
Least squares vs $G(r)$.
Fit as a function of r -range.

Large box

RMCProfile / EPSR / DISCUS

20 000+ atoms.
Fit $F(Q)$, $G(r)$, Bragg.
A snapshot, not a free-energy
minimum. Constrain hard.

Higher-level / ab initio

DiffPy-CMI, DFT, MD

Diameter, stacking-fault
probability, layer spacing.
Or forces instead of free atoms.

PDFgui is a GUI on the PDFfit2 engine

PDFfit2 does the work

C++ engine, Python-accessible. Calculates $G(r)$ from a 3-D cell and least-squares-fits it to data.

PDFgui organises the science

Projects (.ddp), journals, live plots, parametric plots, T / doping / r-series wizards.

Symmetry is a constraint, not a prison

Coordinates and U_{ij} can follow any space group — usually lower than the average cell.

It will not fix bad data

PDFgui is a calculator and a fitter. Reduction (PDFgetX3 / Gudrun) is a separate claim.

The small-box contract

Periodic box

P1 periodic boundary conditions. Translational symmetry is a metric, not a claim, if r_{MAX} is a few cells.

Few hundred atoms

One cell, a supercell, or an expanded asymmetric unit — not a 20 000-atom snapshot.

Symmetry $\leq$ average

Local symmetry breaking is the point. Built-in generators for coordinates and ADPs.

Least squares vs r-range

High-r $\approx$ average structure. Low-r $\approx$ local structure. The residual as a function of r is data.

Effects PDFgui accounts for

Thermal broadening

Anisotropic Uij. Peaks widen with uncorrelated static + dynamic disorder.

Correlated motion

Nearest neighbours vibrate together. delta1, delta2, sratio, rcut sharpen the first shells.

Finite Q_{MAX}

Sinc convolution. r-resolution $\approx \pi/Q_{\text{max}}$. Termination ripples if the tail is thin.

Q-resolution (Q_{DAMP})

G(r) is multiplied by a real-space Gaussian. That is field-of-view, not particle size.

Particle size

spdiameter: spherical envelope only. Non-spherical shapes need another tool.

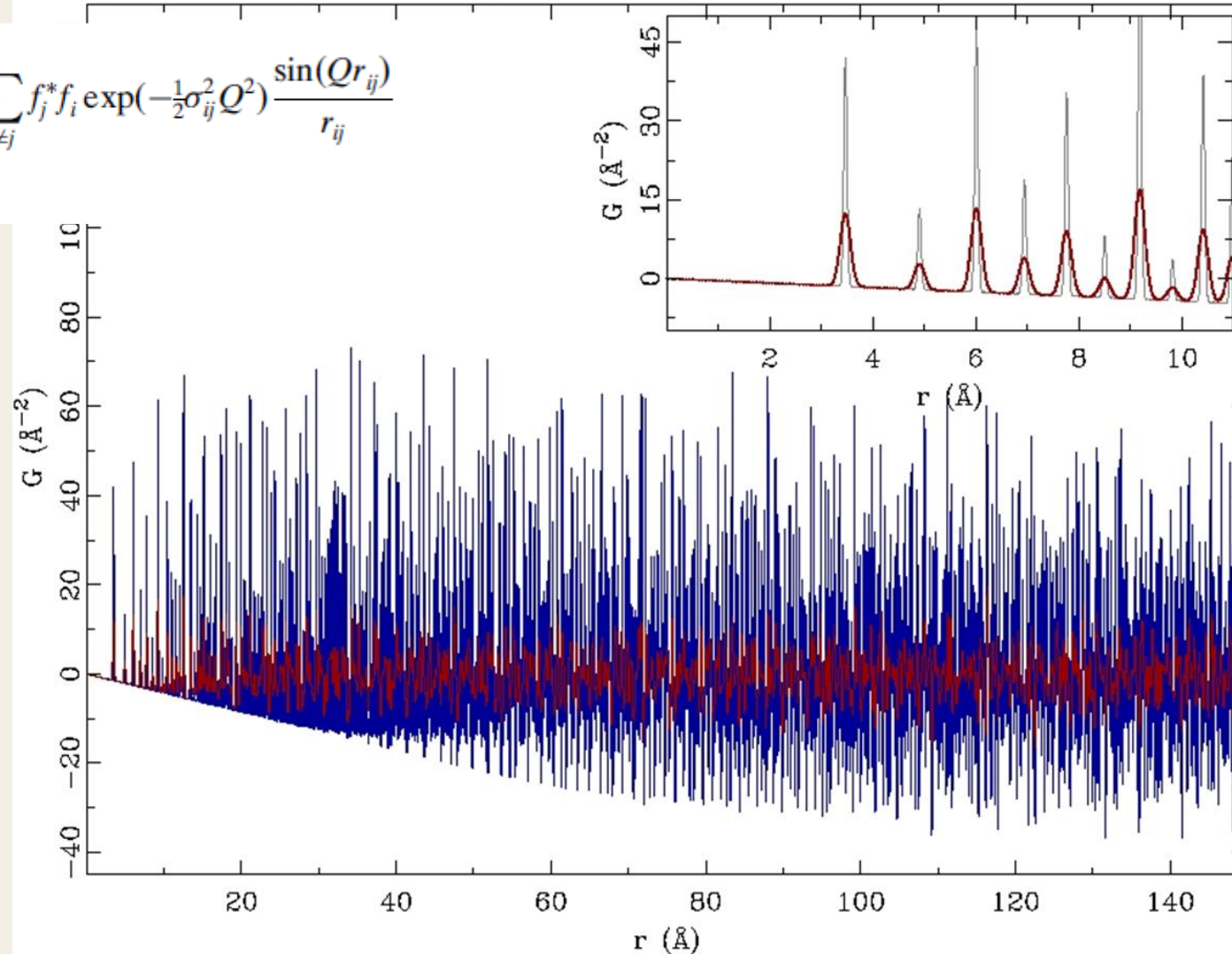
Radiation contrast

X-ray $\approx Z$; neutron = bcoh. Isotopes and ions: dummy species in the lookup table.

Thermal broadening

Thermal vibrations of atoms broaden PDF peaks anisotropic atomic displacement parameters (ADPs)

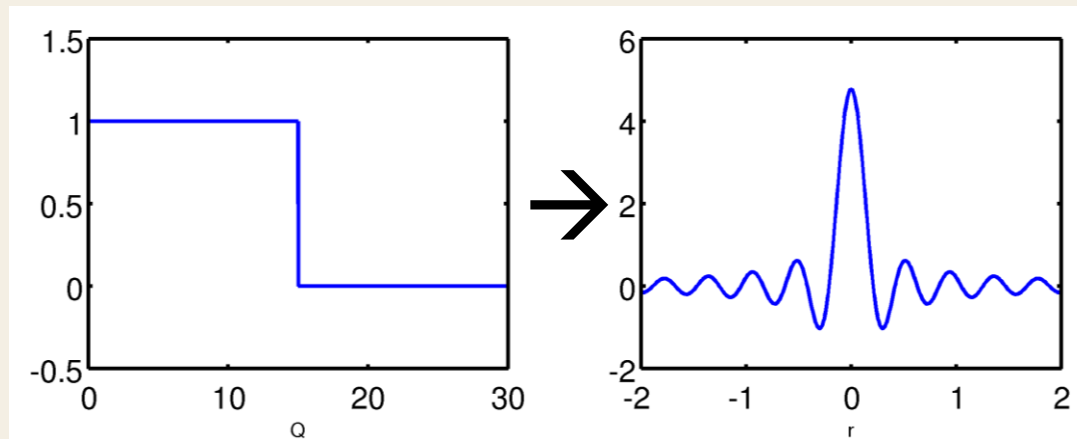
$$F(Q) = \frac{1}{N\langle f \rangle^2} \sum_{i \neq j} f_j^* f_i \exp(-\frac{1}{2} \sigma_{ij}^2 Q^2) \frac{\sin(Qr_{ij})}{r_{ij}}$$



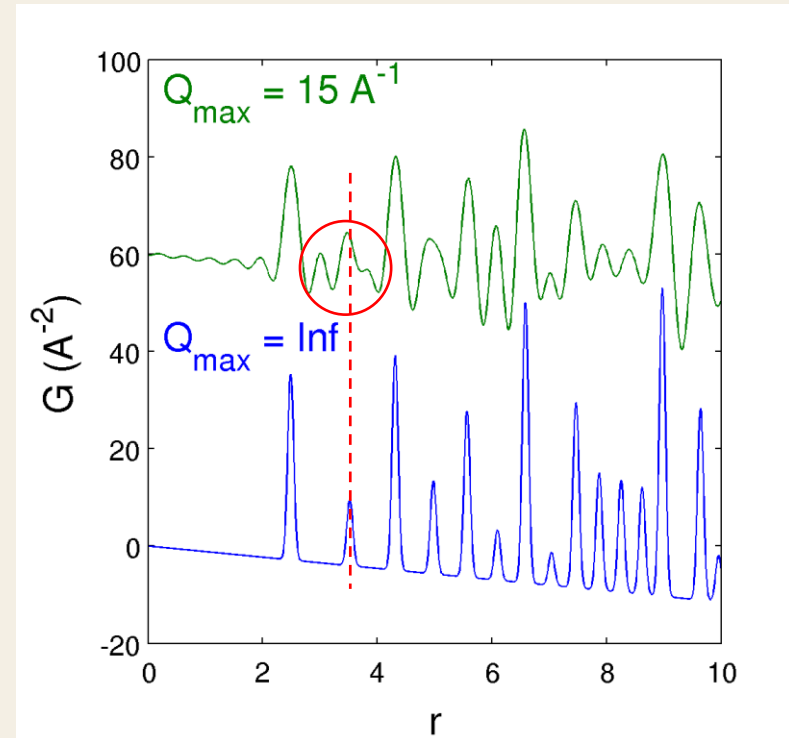
PDFgui accounts
for ADPs

Finite Q_{MAX}

$$G(r) = \frac{2}{\pi} \int_0^{Q_{\text{max}}} F(Q) \sin Qr \, dQ$$



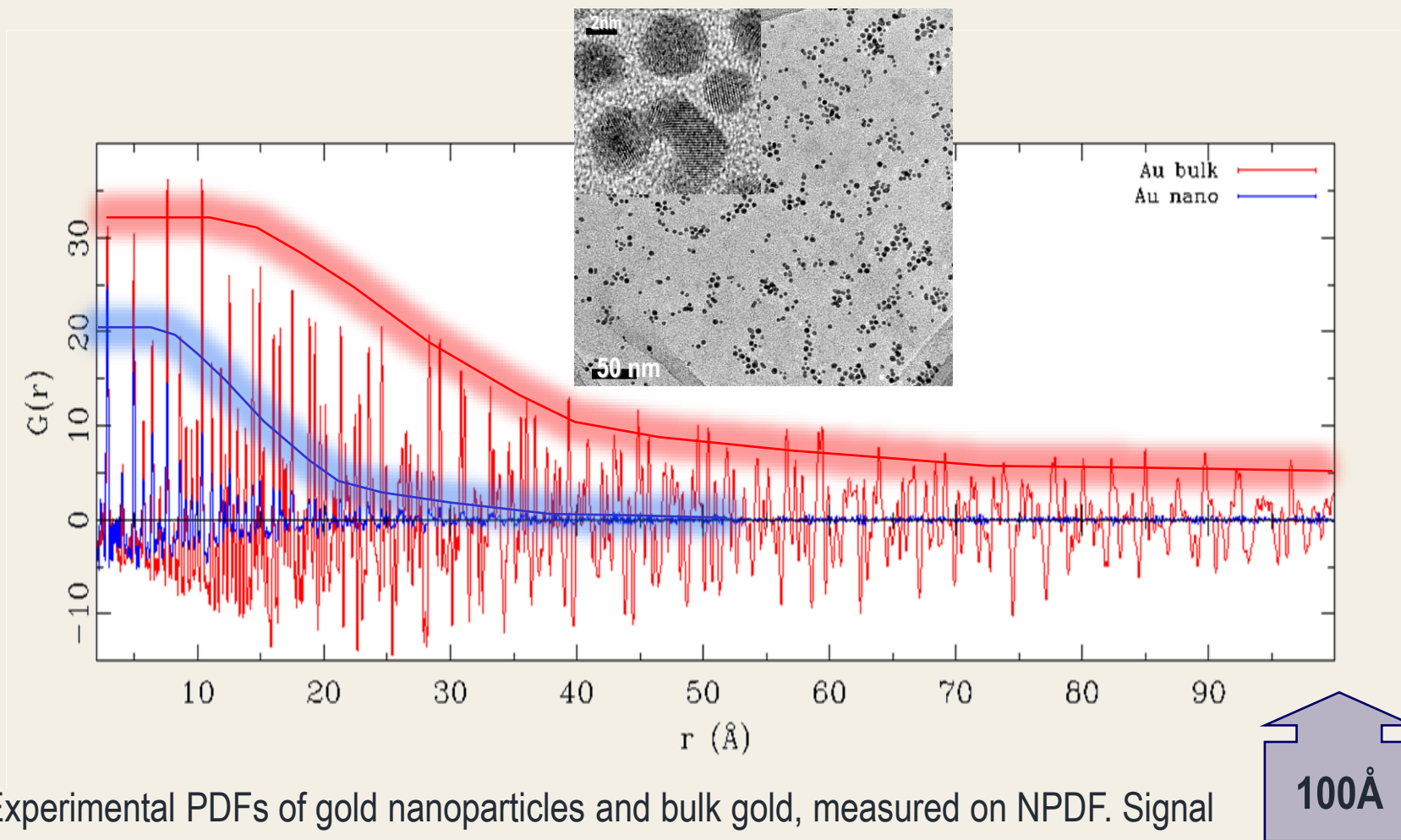
simulated nickel PDF



- ideal $F(Q)$ is multiplied by a step function
- $G(r)$ gets convoluted with a sinc function $\text{sinc}(r) = \sin(Q_{\text{max}} r) / r \rightarrow r\text{-resolution} \approx \pi/Q_{\text{max}}$
- good r -resolution of G requires large Q_{max}
 $Q = 4\pi \sin \theta / \lambda \rightarrow$ best results with TOF neutrons or high-energy X-rays

Particle size

K.L. Page et al., Chem. Phys. Lett. **393**, 385 (2004).

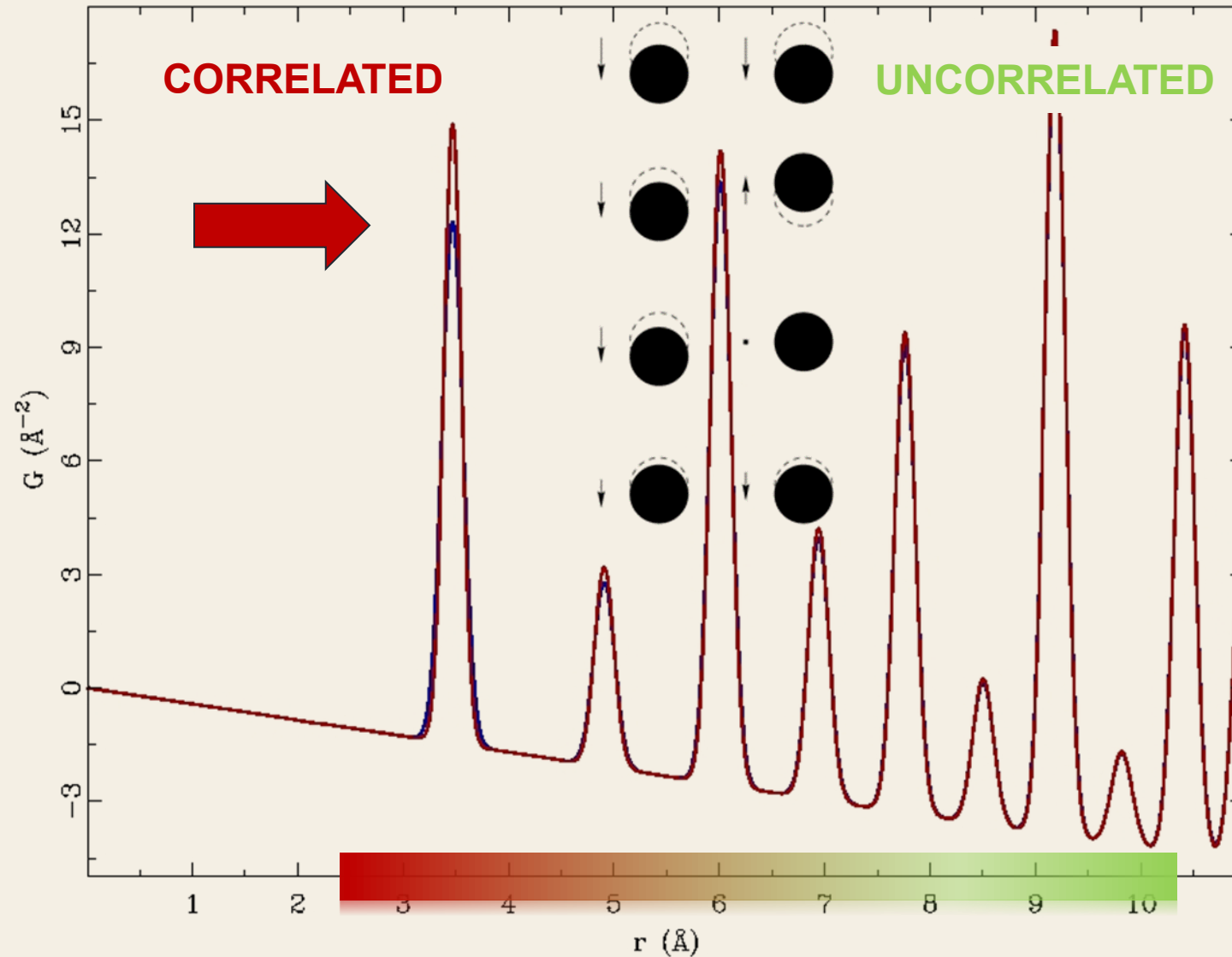


Experimental PDFs of gold nanoparticles and bulk gold, measured on NPDF. Signal damping depends on both **shape** & **size**!

PDFgui accounts for spherical shapes only

Correlated motion

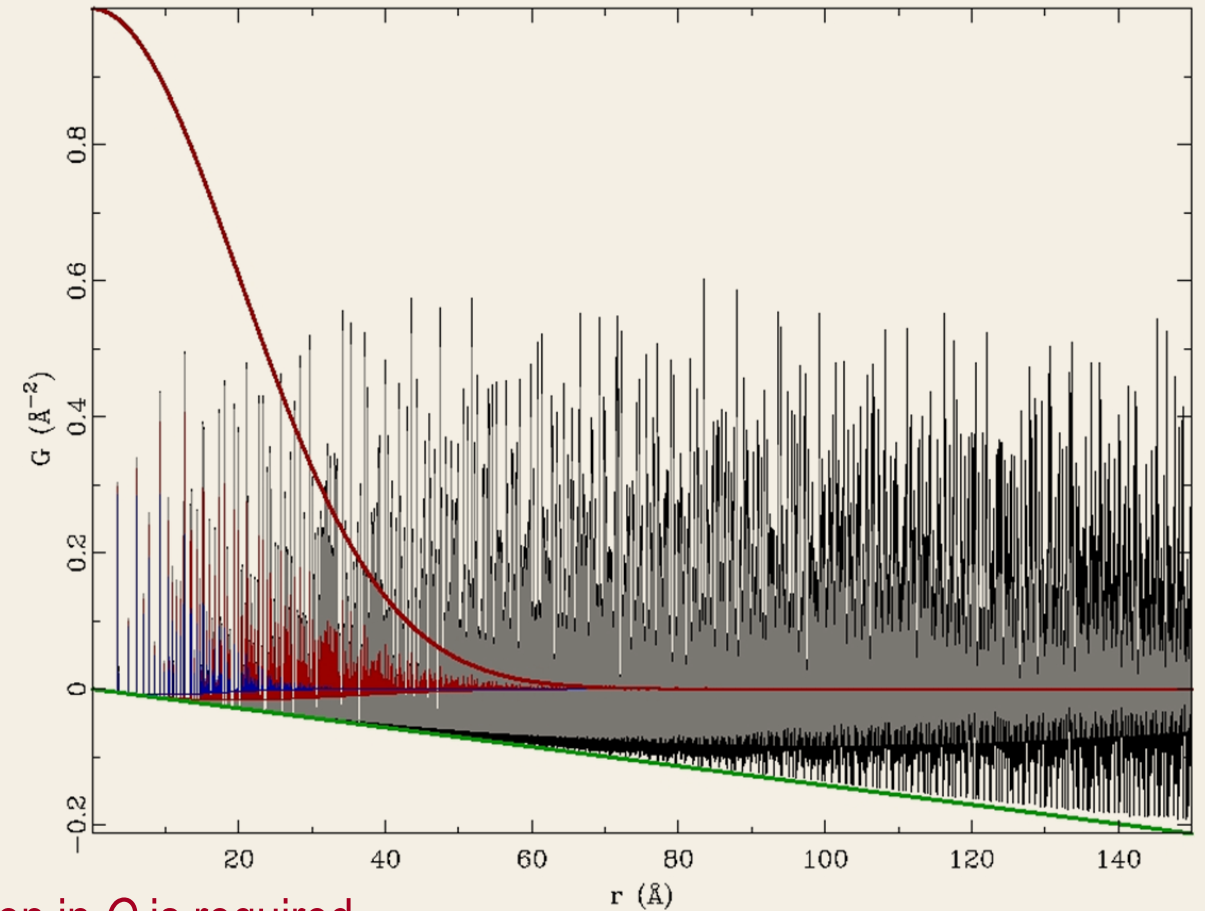
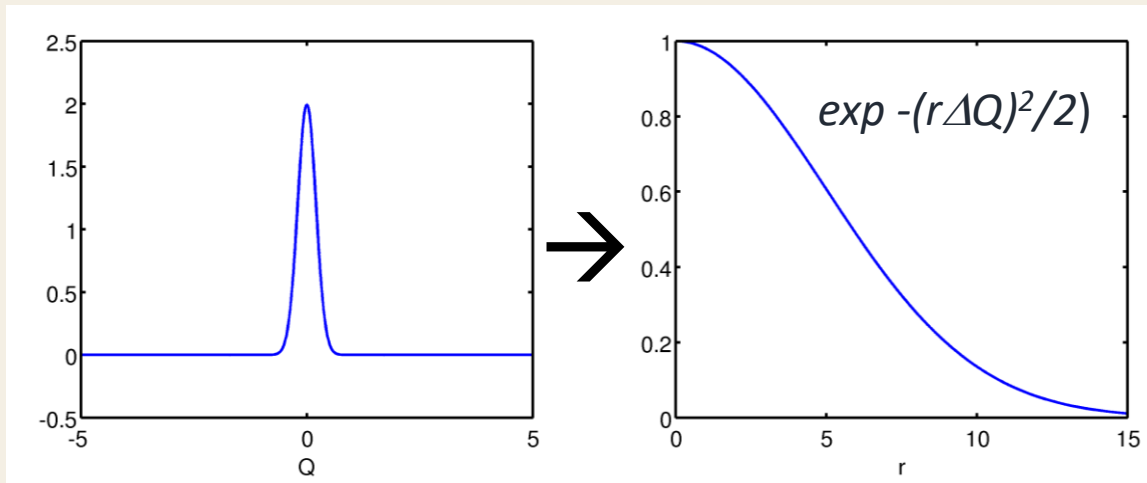
Nearest neighbors tend to vibrate in correlated manner this sharpens the NN peak(s) in PDF relative to distant peaks



PDFgui
Incorporates empirical corrections

Q-resolution (Q_{DAMP})

- Ideal $F(Q)$ is convoluted by Gaussian to simulate finite Q resolution
- $G(r)$ gets multiplied by real-space Gaussian with reciprocal width



For $G(r)$ to have good r -range high resolution in Q is required
Q-resolution defines PDF “field of view”

Radiation contrast

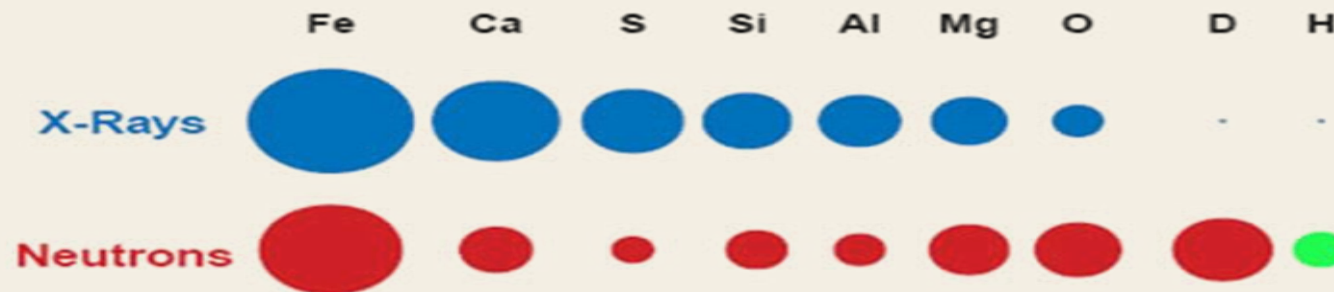
- Considering scattering contrast
- Considering absorption
- Considering isotopes and resonances

hydrogen 1 H 1.0079																	helium 2 He 4.0026						
lithium 3 Li 6.941	beryllium 4 Be 9.0122																	boron 5 B 10.811	carbon 6 C 12.011	nitrogen 7 N 14.007	oxygen 8 O 15.999	fluorine 9 F 18.998	neon 10 Ne 20.180
sodium 11 Na 22.990	magnesium 12 Mg 24.305																	aluminum 13 Al 26.982	silicon 14 Si 28.086	phosphorus 15 P 30.974	sulfur 16 S 32.065	chlorine 17 Cl 35.453	argon 18 Ar 39.948
potassium 19 K 39.098	calcium 20 Ca 40.078	scandium 21 Sc 44.956	titanium 22 Ti 47.867	vanadium 23 V 50.942	chromium 24 Cr 51.996	manganese 25 Mn 54.938	iron 26 Fe 55.845	cobalt 27 Co 58.933	nickel 28 Ni 58.693	copper 29 Cu 63.546	zinc 30 Zn 65.39	gallium 31 Ga 69.723	germanium 32 Ge 72.61	arsenic 33 As 74.922	selenium 34 Se 78.96	bromine 35 Br 79.904	krypton 36 Kr 83.80						
rubidium 37 Rb 85.468	strontium 38 Sr 87.62	yttrium 39 Y 88.906	zirconium 40 Zr 91.224	niobium 41 Nb 92.906	molybdenum 42 Mo 95.94	technetium 43 Tc [98]	ruthenium 44 Ru 101.07	rhodium 45 Rh 102.91	palladium 46 Pd 106.42	silver 47 Ag 107.87	cadmium 48 Cd 112.41	indium 49 In 114.82	tin 50 Sn 118.71	antimony 51 Sb 121.76	tellurium 52 Te 127.60	iodine 53 I 126.90	xenon 54 Xe 131.29						
cesium 55 Cs 132.91	barium 56 Ba 137.33	71 Lu 174.97	72 Hf 178.49	73 Ta 180.95	74 W 183.84	75 Re 186.21	76 Os 190.23	77 Ir 192.22	78 Pt 195.08	79 Au 196.97	80 Hg 200.59	81 Tl 204.38	82 Pb 207.2	83 Bi 208.98	84 Po [209]	85 At [210]	86 Rn [222]						
francium 87 Fr [223]	radium 88 Ra [226]	89-102 * *	103 Lr [262]	104 Rf [261]	105 Db [262]	106 Sg [266]	107 Bh [264]	108 Hs [265]	109 Mt [268]	110 Uun [271]	111 Uuu [272]	112 Uub [277]	114 Uuq [289]										

* Lanthanide series

** Actinide series

lanthanum 57 La 138.91	cerium 58 Ce 140.12	praseodymium 59 Pr 140.91	neodymium 60 Nd 144.24	promethium 61 Pm [145]	samarium 62 Sm 150.36	europium 63 Eu 151.96	gadolinium 64 Gd 157.25	terbium 65 Tb 158.93	dysprosium 66 Dy 162.50	holmium 67 Ho 164.93	erbium 68 Er 167.26	thulium 69 Tm 168.93	ytterbium 70 Yb 173.04
actinium 89 Ac [227]	thorium 90 Th 232.04	protactinium 91 Pa 231.04	uranium 92 U 238.03	neptunium 93 Np [237]	plutonium 94 Pu [244]	americium 95 Am [243]	curium 96 Cm [247]	berkelium 97 Bk [247]	californium 98 Cf [251]	einsteinium 99 Es [252]	fermium 100 Fm [257]	mendelevium 101 Md [258]	nobelium 102 No [259]



PDFgui handles
xPDF & nPDF

Parameters attached to a DATASET

 r_{MIN} , r_{MAX}

User-selected fit range. Fixed. Narrow ranges correlate parameters.

 Q_{MAX}

Upper limit of the Fourier integral. Fixed. Sets r-resolution.

Qdamp

Gaussian damping from limited Q-resolution. Refine on a calibrant; fix on the sample.

QbroadExtra high-r peak broadening. Only matters at large r_{MAX} . Same protocol as Qdamp.**dscale**

Overall scale of that dataset. Refined. Can trade with pscale.

RadiationNeutron or **X**-ray declaration. Chooses the lookup table.

Parameters attached to a PHASE

pscale

Phase scale. Often correlated with dscale.

delta1 / delta2

1/r and 1/r² peak sharpening. Strongly correlated.

a, b, c, α , β , γ

Lattice. Refine with symmetry in mind.

sratio, rcut

Rigid-unit sharpening inside a cutoff.

x,y,z[n], occ[n]

Fractional coordinates and occupancy.

spdiameter

Spherical nanoparticle diameter (Å).

u11...u66[n]

Anisotropic ADPs, symmetry-constrained.

stepcut

Hard truncate of G(r) above r (Å).

Handles, correlations, declarations

Syntax @pn

Parameters are assigned by handle: @1, @55, @321. Numbers need not be consecutive. Variables that share a handle share a parameter. Do not assign the same number to incompatible types.

Correlations

PDFgui flags parameter correlations if $|\rho| > 0.8$. Narrow r-ranges make lattice, scale, ADPs and delta parameters trade. Qdamp and spdiameter are notorious partners. Refine the instrument on a calibrant first.

Dummy species

X-rays: Z from the element label. Ions → a neighbour with the right electron count. Neutrons: natural-abundance b_{COH} . Isotopes → invent a two-letter code and edit the lookup table.

What you actually need to start

01

Data

.gr files plus Qmax, radiation, chemistry, number density.

02

A starting cell

Space group, lattice, asymmetric unit, occupancies. CIF / xyz / PDFfit.

03

A calibrant protocol

Qdamp (and Qbroad if rMAX is large) from Ni or Si in the same geometry.

04

A plan for r

High-r first if you want the average; low-r if you are hunting a local distortion.

Hidden local state in AgGaTe₂

Diamondoid thermoelectrics — why is κL ultralow, and smaller in the Ag variant than in Cu?

Average (Rietveld)

Ag variant: negative thermal expansion along c. Cu variant: ordinary. Ag ADPs inflated in the ab plane — in-plane nanoscale disorder written as ‘thermal’.

Local (PDFgui)

Short-range distortion of AgTe₄ tetrahedra, correlated over a nanometre. Local rotations produce the macroscopic NTE. The distortion emerges continuously on warming.

The moral

NTE is a bulk meter of a nanoscale distortion. The average cell told you something was wrong. Only an r-dependent small-box fit named it.

The paper, and why it needed PDFgui

Hongyao Xie et al., Adv. Mater. 34, 2202255 (2022)

Discovery via elaborate PDFgui modelling: local AgTe₄ rotations, nanometre correlation length, continuous emergence on warming. NTE is the macroscopic integral of that local texture.

Co-refinement

X-ray and neutron $G(r)$ on one model when contrast disagrees.

Phase analysis

Ni + Si mixture: two cells, one dscale logic.

T / doping series

Linked fits. Output of T_n seeds T_{n+1} .

Box-car r-series

Walk a 4 Å window. $R_w(r)$ estimates a correlation length.

Small-box software, without blushing

PDFgui

- + Open source and free
- + Simple GUI, shareable .ddp projects
- + Wizards for T, doping, r-series
- + Symmetry constraints out of the box
- Slow at high r
- Cumbersome exports
- Spherical NP envelope only

TOPAS PDF

- + Very fast
- + Custom output functions
- + Easy parameter algebra
- Commercial
- Steeper curve; write your macros

Also in the family: DiffPy-CMI (scripted, composable), GSAS-II if your group already lives there.

Download, tutorial pack, neighbours

PDFgui / PDFfit2

<https://www.diffpy.org/products/pdfgui.html>

Tutorial files (ILL pack)

<https://workshops.ill.fr/event/509/attachments/583/2603/PDFgui-tutorial-2022.zip>

DiffPy-CMI

<https://www.diffpy.org/products/diffpycmi>

PDFgetX3

Juhás et al., J. Appl. Cryst. 46, 560 (2013)

Farrow et al. 2007

J. Phys.: Condens. Matter 19, 335219 — the engine paper

What to take into the tutorial

- 01 Start from a cell that Bragg already believes.
- 02 Fix Qdamp from a calibrant. Do not sell it as size.
- 03 High-r is the average; low-r is the argument.
- 04 One handle per physical idea. Watch parameter correlations, bad if $|\rho| > 0.8$.
- 05 Escalate the box only when the residual has a name.