OPERA COST Action Training School
Applications and Characterization of Epitaxial Materials
June 2025

Measuring AFM wrong (and perhaps right)

David Nečas
Brno University of Technology – CEITEC
yeti@gwyddion.net

BRNO UNIVERSITY OF TECHNOLOGY
CEITEC
Central European Institute of Technology
BRNO | CZECH REPUBLIC

Scaning Probe Microscopy

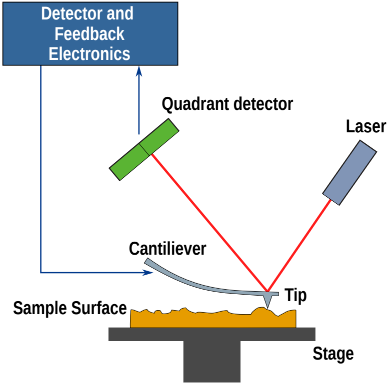
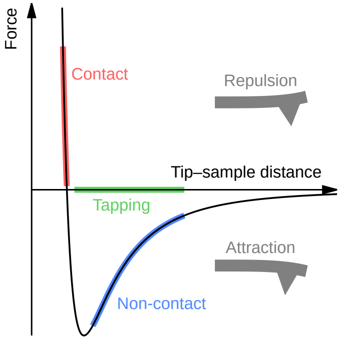
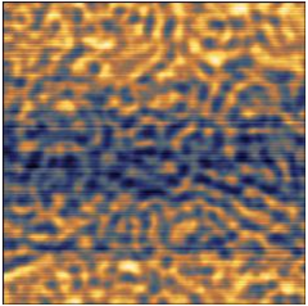


Image formation: Scanning the surface line by line





Anisotropy: Fast & slow axis

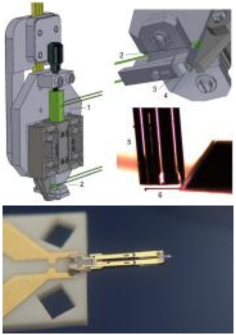
Open & closed loop

Techniques: Mechanical, electrical, magnetic, thermal, optical, . . .

More in Petr Klapetek's lecture on Friday!

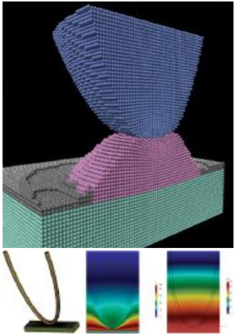
Quantitative SPM

Hardware



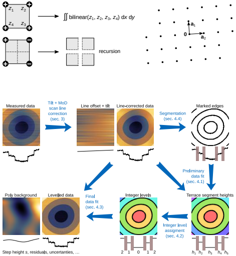
Stage & head
Probes
Electronics
Enviromental
. . .

Physics



Probe–sample interaction
Atomistic modelling
Molecular dynamics
FEM & FDTD
. . .

Algorithms



$$\hat{G}(\tau) = \frac{1}{L-\tau}\int_0^{L-\tau} \hat{z}(x)\hat{z}(x+\tau)\,\mathrm{d}x.$$

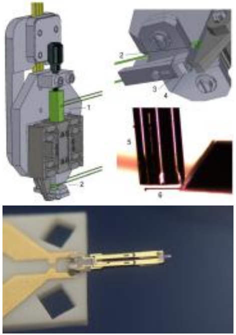
$$\frac{1}{L-\tau}\int_0^{L-\tau}\left[z(x)-\sum_j a_j\varphi_j(x)\right]\left[z(x+\tau)-\sum_k a_k\varphi_k(x+\tau)\right]\mathrm{d}x,$$

Filtering & preprocessing
Feature recognition
Model fitting
Statistical
. . .

Klapetek P., et al., Quantitative Data Processing in Scanning Probe Microscopy, 2nd edition, Elsevier (2018)

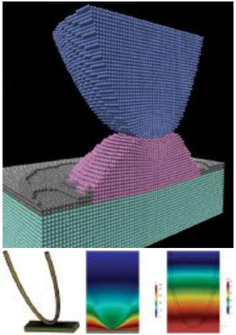
Quantitative SPM

Hardware



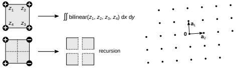
Stage & head
Probes
Electronics
Enviromental
. . .

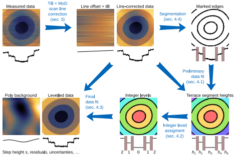
Physics



Probe–sample interaction
Atomistic modelling
Molecular dynamics
FEM & FDTD
. . .

Algorithms





$$\hat{G}(\tau) = \frac{1}{L-\tau}\int_0^{L-\tau} \hat{z}(x)\hat{z}(x+\tau)\,\mathrm{d}x.$$

$$\frac{1}{L-\tau}\int_0^{L-\tau}\left[z(x)-\sum_j a_j\varphi_j(x)\right]\left[z(x+\tau)-\sum_k a_k\varphi_k(x+\tau)\right]\mathrm{d}x,$$

Filtering & preprocessing
Feature recognition
Model fitting
Statistical
. . .

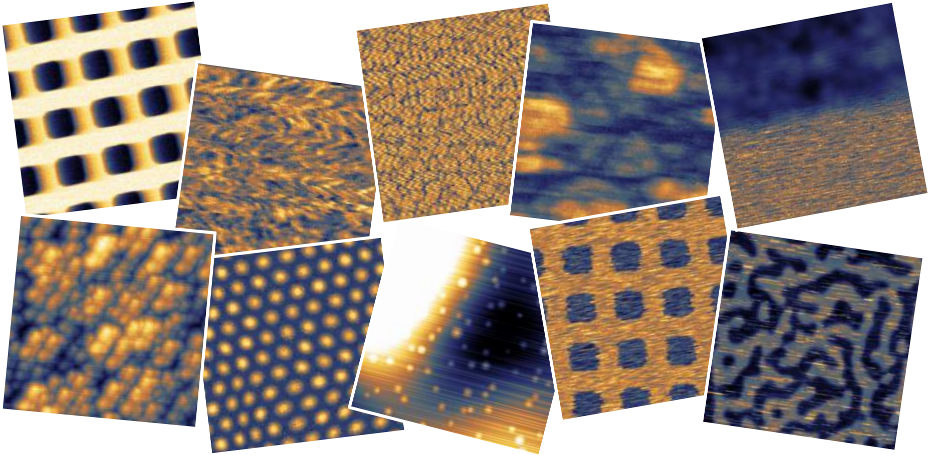
Secret ingredient

?

This talk?

Klapetek P., et al., Quantitative Data Processing in Scanning Probe Microscopy, 2nd edition, Elsevier (2018)

Bad data? Easy!

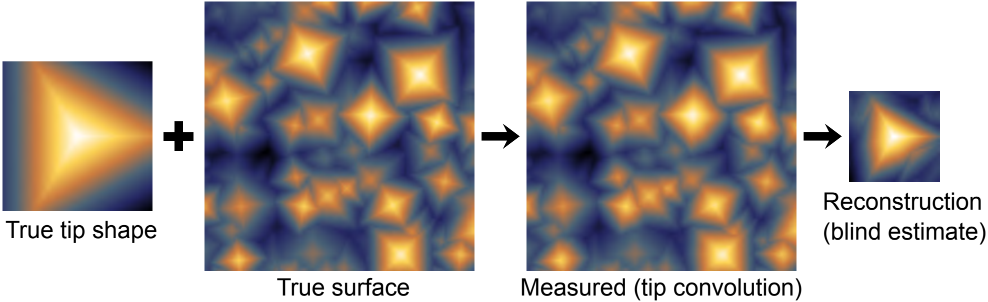


Many effects conspire to sabotage our SPM measurements
Mechanical & electromagnetic noise, laser interference, cross-talk, changes in probe properties & contact, contamination & tip convolution, bad feedback parameters, hysteresis & non-linearity, creep, aging, topography artefacts, . . .

Good data?

Tip convolution → convolution artefacts

Blind estimation
of tip shape using tip imaging
by sharp surface features

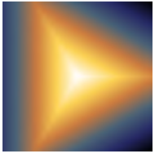


Villarrubia J.S., Algorithms for Scanned Probe Microscope Image Simulation, Surface Reconstruction, and Tip Estimation, *J. Res. Natl. Inst. Stand. Technol.* **102** (1997) 425

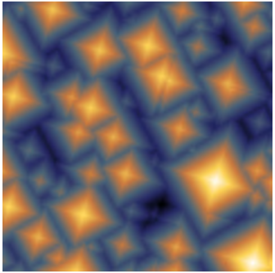
Good data?

Tip convolution → convolution artefacts

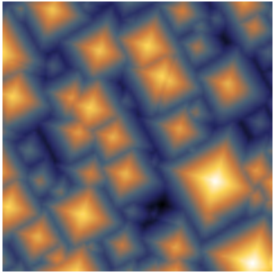
Blind estimation
of tip shape using tip imaging
by sharp surface features



True tip shape



True surface

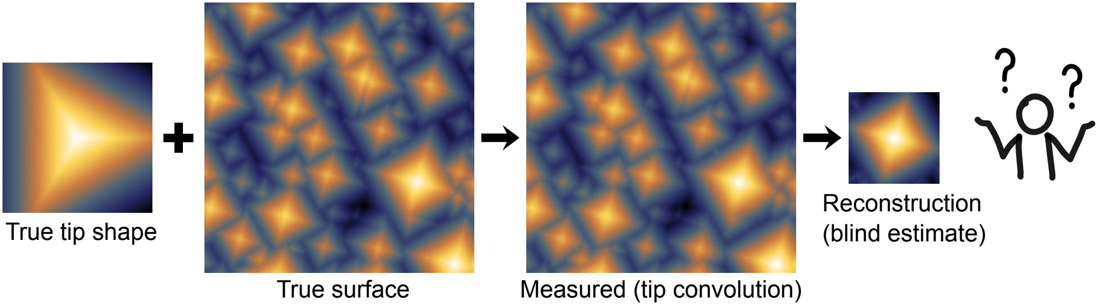


Measured (tip convolution)

Good data?

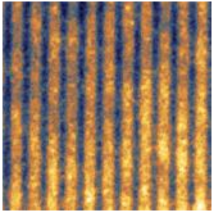
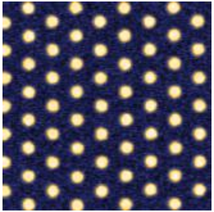
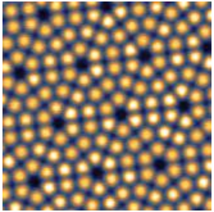
Tip convolution → convolution artefacts

Blind estimation
of tip shape using tip imaging
by sharp surface features



Similar: tip transfer/point spread function, etc.

Periodic structures

Different origins:

- litography & laser interference
- atomic lattices
- self-organised wrinkles, domains, . . .

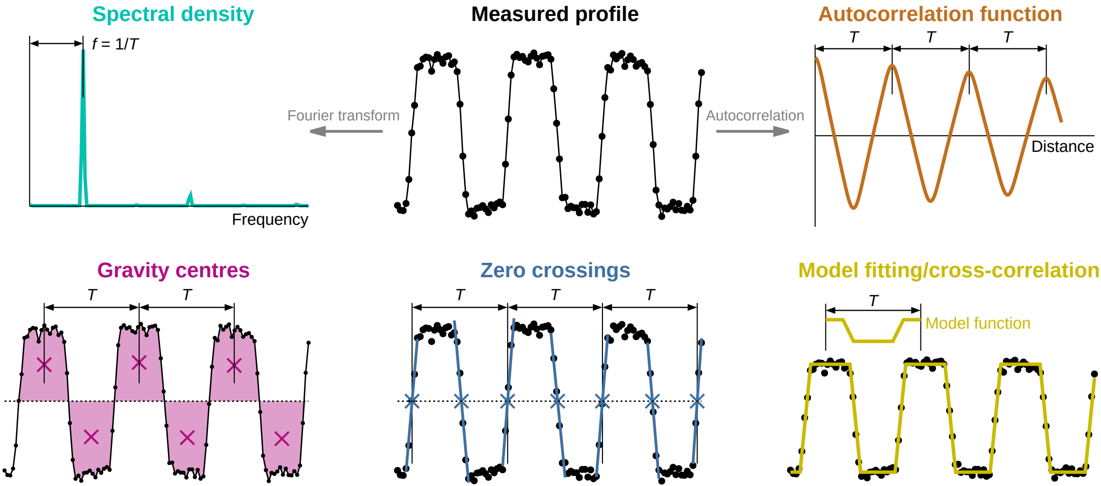
Different purposes:

- studying a structure/process
- instrument calibration
- *ex post* data correction

Data evaluation generally similar.

Pitch & height standards – but maybe not both from one measruement.

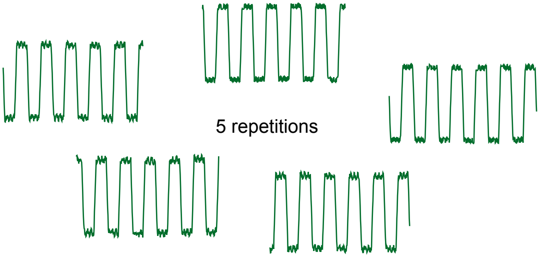
Evaluation of period/pitch



- Feature identification (direct space)
- Fourier transform
- Autocorrelation – not widely used

Nečas D., Yacoot A., Valtr M., Klapetek P., Demystifying data evaluation in the measurement of periodic structures, *Meas. Sci. Technol.* **34** (2023) 055015

More is better



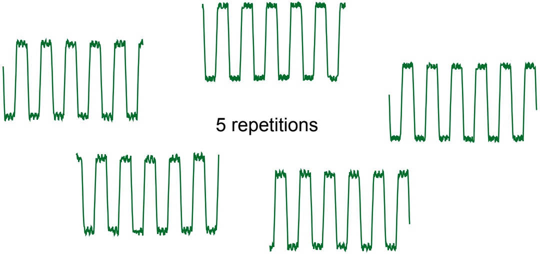
Warm up question

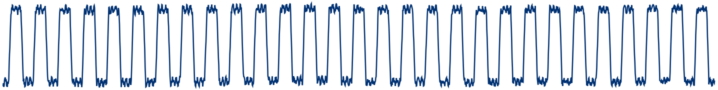
Measuring 5× is about:

(a) $\sqrt{5}\times$ worse,
(b) the same,
(c) $\sqrt{5}\times$ better,
(d) 5× better,
(e) $\infty\times$ better

than measuring once.

More is better





5× longer profile

Warm up question

Measuring 5× is about:

(a) $\sqrt{5}$× worse,
(b) the same,
(c) $\sqrt{5}$× better,
(d) 5× better,
(e) ∞× better

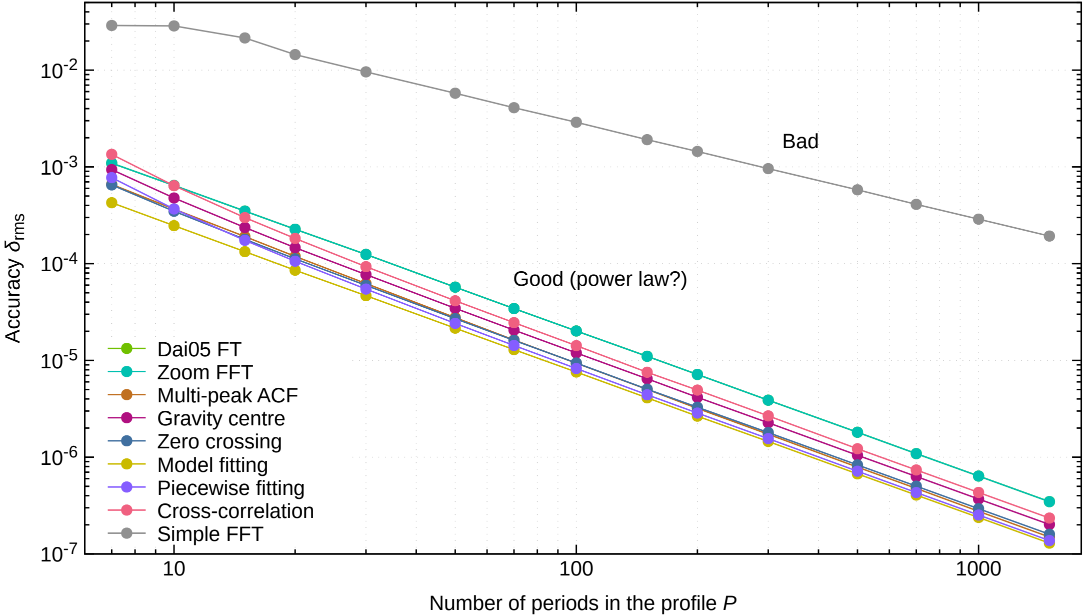
than measuring once.

One longer profile is:

(a) 5× worse,
(b) $\sqrt{5}$× worse,
(c) the same,
(d) $\sqrt{5}$× better,
(e) 5× better.

than measuring 5 shorter ones.

Scaling


Bad
Good (power law?)
Dai05 FT
Zoom FFT
Multi-peak ACF
Gravity centre
Zero crossing
Model fitting
Piecewise fitting
Cross-correlation
Simple FFT
Accuracy δ_{rms}
10^{-2}
10^{-3}
10^{-4}
10^{-5}
10^{-6}
10^{-7}
10
100
1000
Number of periods in the profile P

Scaling power

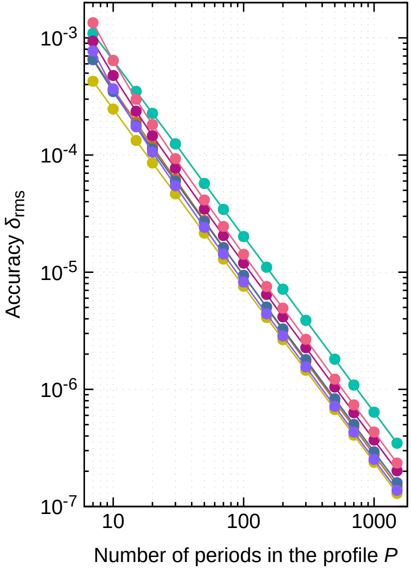
Model $x_n = nT$

Estimate $\hat{T} = \sum_{n=1}^{P} n x_n \Big/ \sum_{n=1}^{P} n^2$

Dispersion $\Delta_T^2 = \sum_{n=1}^{P} \left(\frac{\partial \hat{T}}{\partial x_n}\right)^2 \Delta_x^2 = \Delta_x^2 \Big/ \sum_{n=1}^{P} n^2 \approx \frac{3}{P^3} \Delta_x^2$

Scaling $\Delta_T \approx \frac{1}{P^{3/2}} \sqrt{3}\, \Delta_x$

T	–	period (fitted)
Δ_T	–	standard deviation of T
Δ_x	–	location error in x
P	–	number of periods



Scaling power

Model $x_n = nT$

Estimate $\hat{T} = \sum_{n=1}^{P} n x_n \Big/ \sum_{n=1}^{P} n^2$

Dispersion $\Delta_{\hat{T}}^2 = \sum_{n=1}^{P} \left(\frac{\partial \hat{T}}{\partial x_n} \right)^2 \Delta_x^2 = \Delta_x^2 \Big/ \sum_{n=1}^{P} n^2 \approx \frac{3}{P^3} \Delta_x^2$

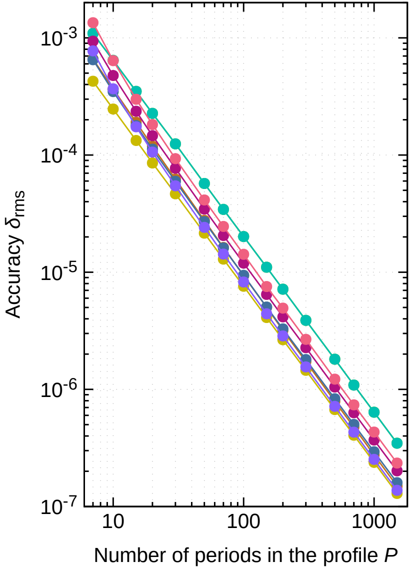
Scaling $\Delta_T \approx \frac{1}{P^{3/2}} \sqrt{3}\, \Delta_x$

T	–	period (fitted)
Δ_T	–	standard deviation of T
Δ_x	–	location error in x
P	–	number of periods

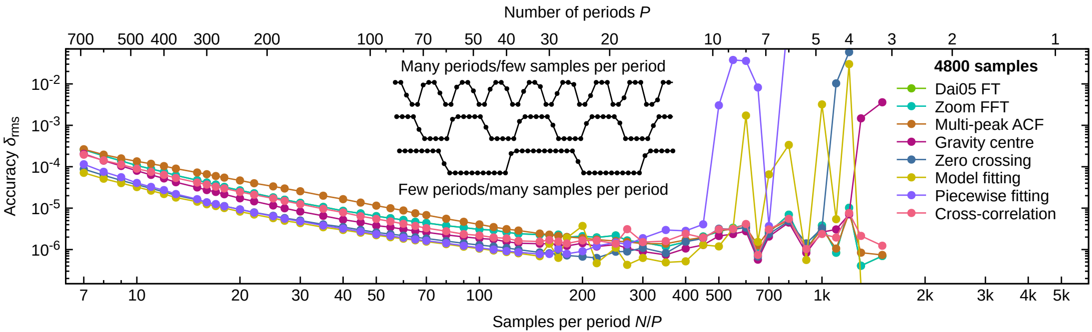
Scaling powers

- **Single long profile: –3/2**
- **Repeated measurement: –1/2**

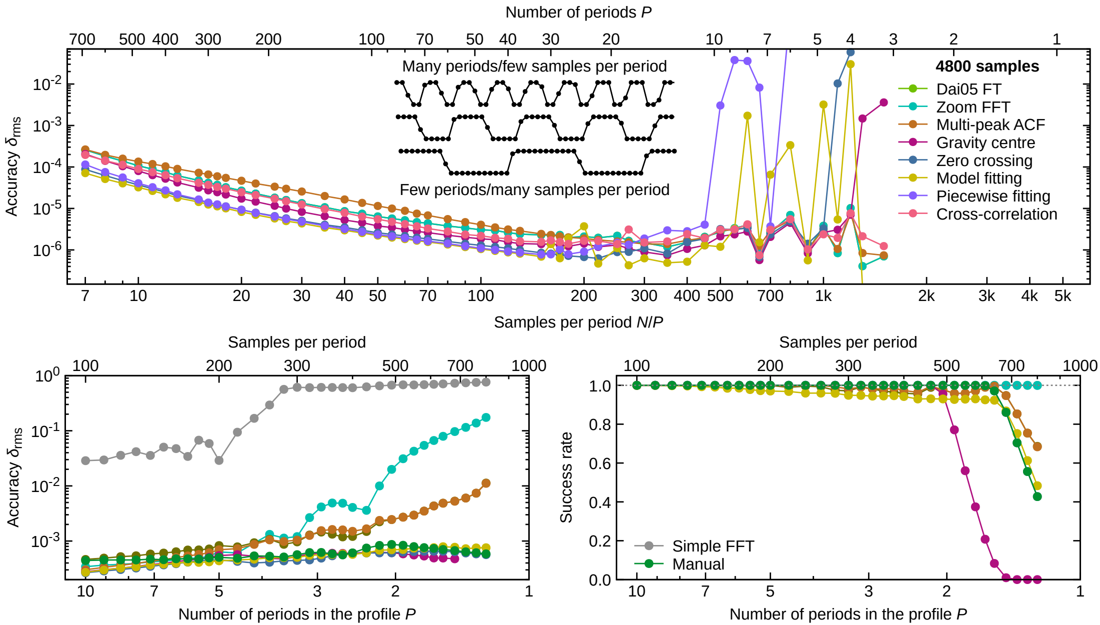
All good methods are similar – probably a theoretical limit



The extremes



The extremes



Steps on silicon

Secondary realisation of metre

Preparation of the 2018 update of the SI

Silicon lattice spacing

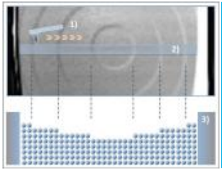
$d_{220} = 192.015\,571\,6(32) \times 10^{-12}$ m

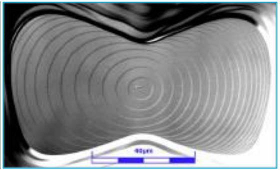
Practical SPM standard

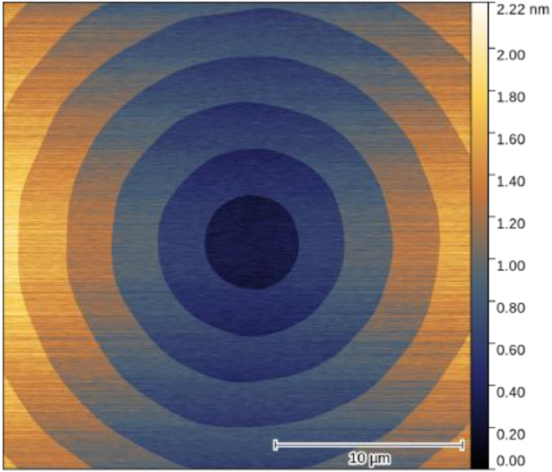
Mono atomic steps on Si (111) surface

Prepared using molecular beam epitaxy

$d_{111} = 313.560\,115\,1(53) \times 10^{-12}$ m







Consultative Committee for Length, Mise en pratique for the definition of the metre in the SI (2019)
Tiesinga E., Mohr P.J, Newell N.B. and Taylor B.N, The 2018 CODATA Recommended Values of the Fundamental Physical Constants, (2019, Web Version 8.1)
Fissel A., Kr ugener J., and Osten H.J., Preparation of large step-free mesas on Si(111) by molecular beam epitaxy, *Phys. Status Solidi C* **9** (2012) 2050

Evaluation

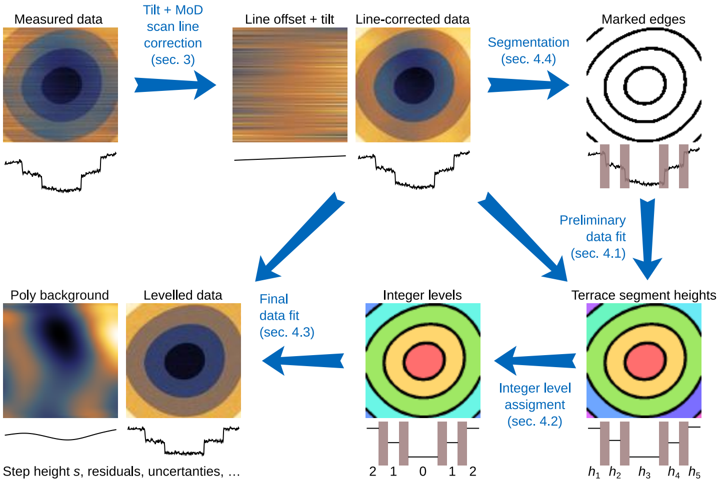
Preprocessing

- line correction
- edge detection
- terrace marking
- connectivity graph

Fitting

$$\text{Height}(x) = s \cdot \text{Level}(x) + \text{Poly}(x)$$

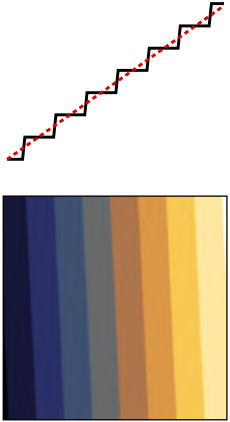
s – step height (fitted)
Level – levels (known integers)
Poly – a polynomial (fitted)



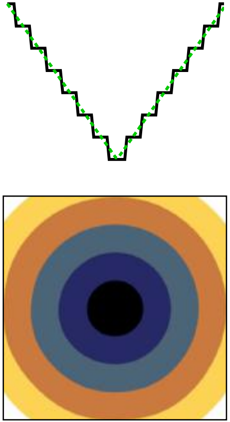
Garnæs J., Nečas D., Nielsen L., Madsen M., Torras-Rosell A., Zeng G., Klapetek P., Yacoot A., Algorithms for using silicon steps for scanning probe microscope evaluation, *Metrologia* **57** (2020) 064002

Overall shape

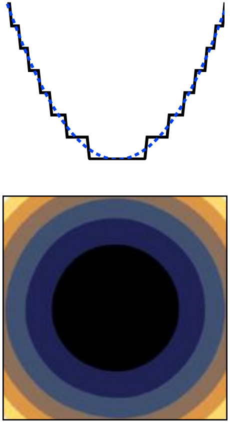
Staircase
$x/2$



Amphitheatre
$|x|$

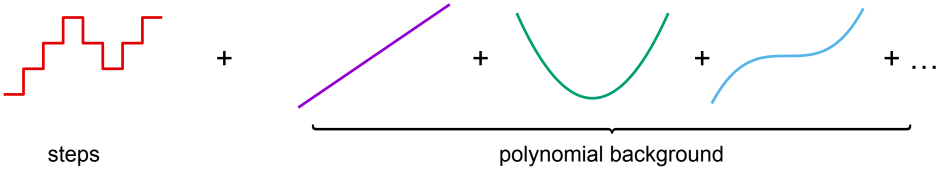


Parabolic
x^2



Which overall shape is the best? worst?

Fitting the profile

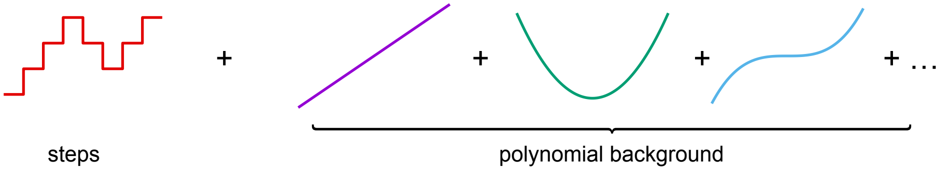


$\text{Height}(x) = s \cdot \text{Level}(x) + \text{Poly}(x)$

Distinct fitting functions – **Good**

Indistignuishable functions – **Bad**

Fitting the profile



$\text{Height}(x) = s \cdot \text{Level}(x) + \text{Poly}(x)$

Distinct fitting functions – **Good**

Indistignuishable functions – **Bad**

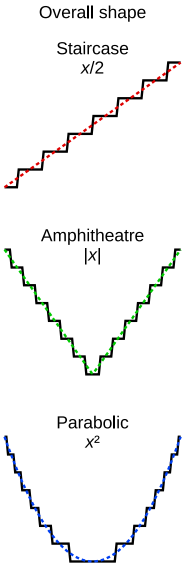
Staircase	–	looks like x	–	bad
Parabolic	–	looks like x^2	–	a bit less bad
Amphitheatre	–	does not look like any polynomial	–	good

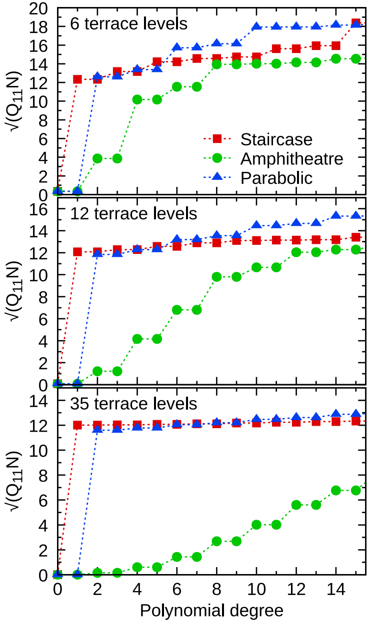
Step error $\Delta_s \propto \sqrt{Q_{11}} \cdot \text{Noise}$

Cofactor matrix $\mathbf{Q} = (\text{Normal matrix})^{-1}$

- does not depend on noise
- computed from scalar products of basis fitting functions

The matrix





Assumptions

- ideal geometry
- no around-step exclusion
- amount of data $N \to \infty$

$\sqrt{Q_{11}N}$ plotted instead of $\sqrt{Q_{11}}$ for meaningful $N \to \infty$ limit

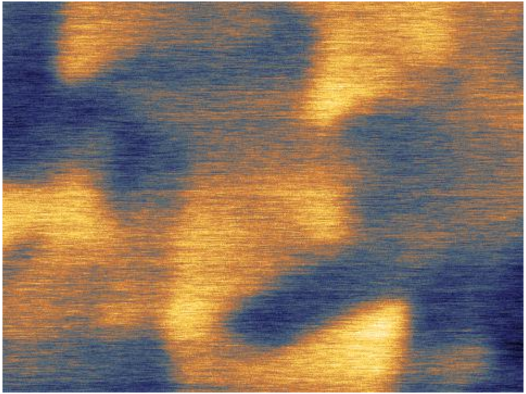
Few terraces

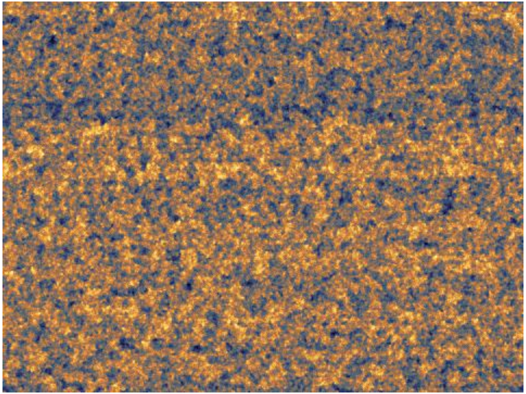
- poor measurement
- small difference

Many terraces

- good measurement
- huge difference

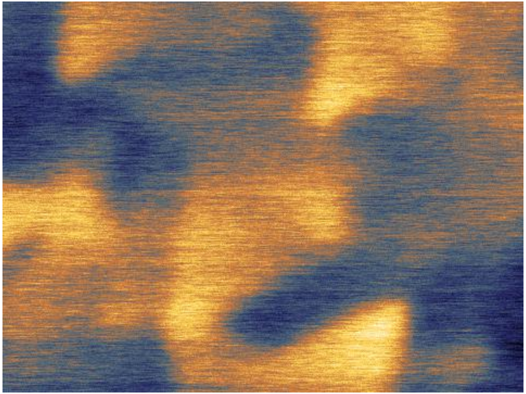
Which one is it?



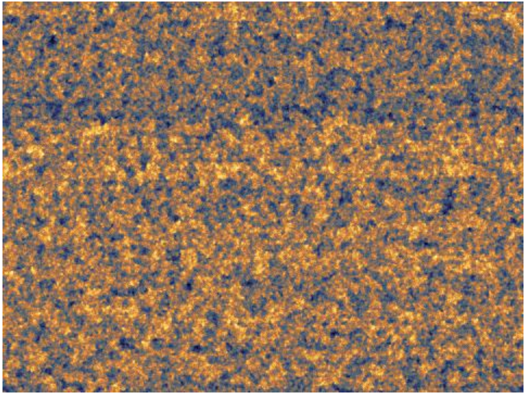


Which one is better for roughness? Left? Right? Neither?

Which one is it?



Pretty bad, $\gtrsim$ 30% bias & poor representativeness.



Probably good for evaluation.

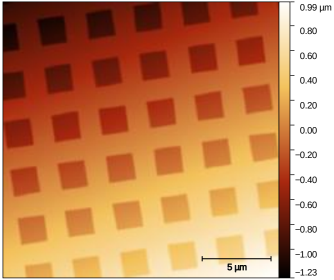
Scan line must be long to avoid losing the lower spatial frequencies.

Governed by $\alpha = T/L$. Should be $\alpha \ll 1$.

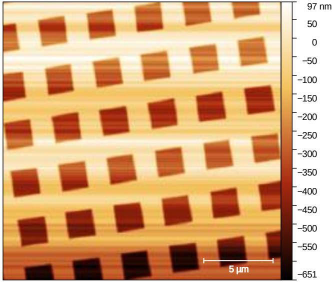
In AFM usually incompatible with ‘can nicely see features’.

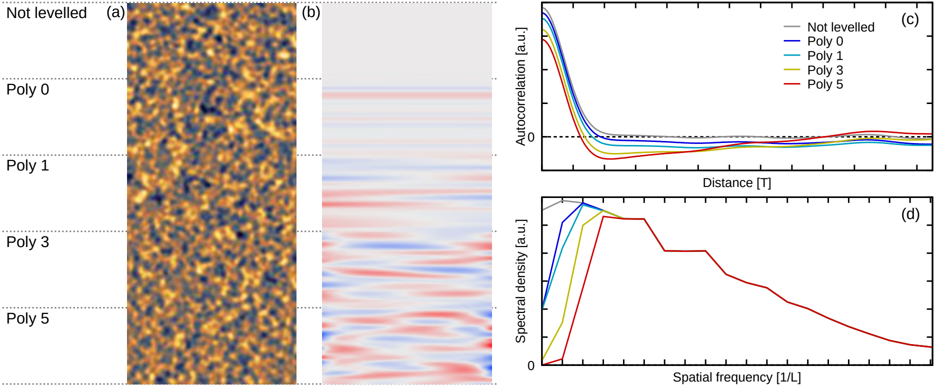
Nečas D., Klapetek P., Valtr M., Estimation of roughness measurement bias originating from background subtraction, *Meas. Sci. Technol.* **31** (2020) 094010
Nečas D., Valtr M., Klapetek P., How levelling and scan line corrections ruin roughness measurement and how to prevent it, *Sci. Rep.* **10** (2020) 15294

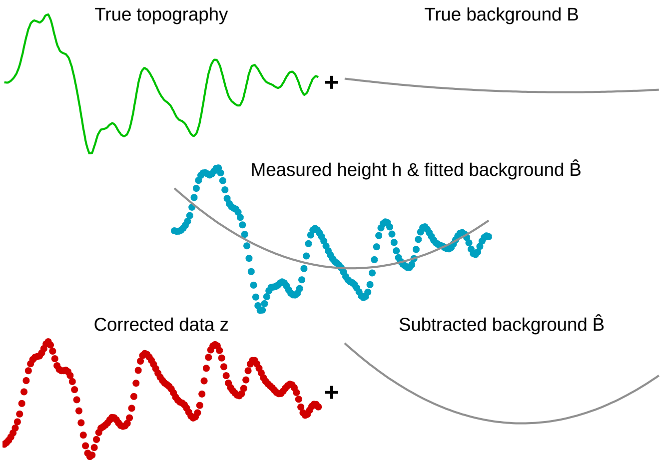
Sample tilt



Scan line misalignment



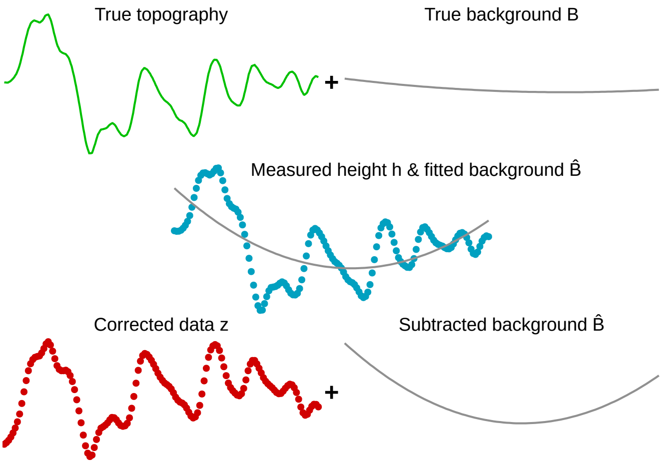




Bias of mean square roughness σ

$$E[\hat{\sigma}^2] = \sigma^2 - 2^D \int_0^1 C_n(t)\, G(tL)\, \mathrm{d}t$$

T	–	autocorrelation length
D	–	dimension (1, 2, . . .)
G	–	autocorrelation function
C_n	–	ugly function
n	–	polynomial degree



Bias of mean square roughness σ

$$\mathrm{E}[\hat{\sigma}^2] = \sigma^2 - 2^D \int_0^1 C_n(t)\, G(tL)\, \mathrm{d}t$$

T – autocorrelation length
D – dimension (1, 2, . . .)
G – autocorrelation function
C_n – ugly function
n – polynomial degree

Autocorrelation function to rescue

$$\mathrm{E}[\hat{G}] = G - R_n G$$

G – autocorrelation function
R_n – even uglier linear operator

G knows about its own bias!

Fit $G - RG$ instead of G to experimental data

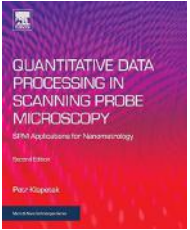
$$G_{\mathrm{Gauss}}^{\mathrm{bias}}(\tau) \approx \sigma^2 \exp\left(-\frac{\tau^2}{T^2}\right) - \sqrt{\pi} n \sigma^2 \frac{T}{L}\left(1 + \frac{\tau}{L}\right) + n^2 \sigma^2 \frac{T^2}{L^2}\left(1 + \frac{2\tau}{L}\right)$$

Or invert $G = (1 - R)^{-1} \hat{G}$ (adventurous)

Nečas D., Self-consistent autocorrelation for finite-area bias correction in roughness measurement, *Eng. Res. Express* **6** (2024) 025560

Conclusions

- Measuring wrong is easy.
- Solid ‘hardware’ part $\not\Rightarrow$ useful data.
- *What* do you measure?
- Intuition often fails us.
- Simulate!



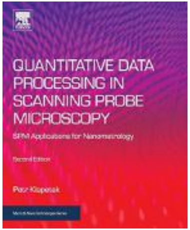
Almost everything is implemented in Gwyddion.

Nečas D., Klapetek P., Study of user influence in routine SPM data processing, Measurement Science and Technology **28** (2017) 034014
Nečas D., Klapetek P., Synthetic Data in Quantitative Scanning Probe Microscopy, *Nanomaterials* **11** (2021) 1746
Klapetek P., et al., Quantitative Data Processing in Scanning Probe Microscopy, 2nd edition, Elsevier (2018)

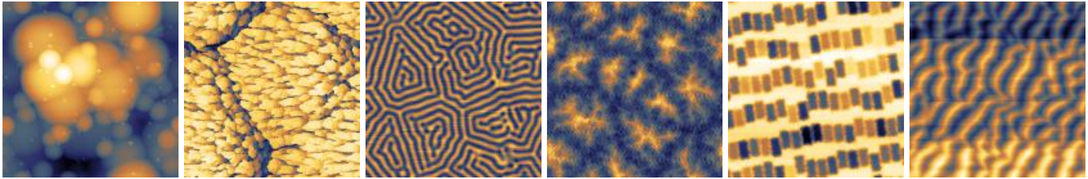
Conclusions

- Measuring wrong is easy.
- Solid ‘hardware’ part $\not\Rightarrow$ useful data.
- *What* do you measure?
- Intuition often fails us.
- Simulate!

Almost everything is implemented in Gwyddion.



Nothing you saw was real SPM data.



Nečas D., Klapetek P., Study of user influence in routine SPM data processing, Measurement Science and Technology **28** (2017) 034014
Nečas D., Klapetek P., Synthetic Data in Quantitative Scanning Probe Microscopy, *Nanomaterials* **11** (2021) 1746
Klapetek P., et al., Quantitative Data Processing in Scanning Probe Microscopy, 2nd edition, Elsevier (2018)