Remote Data Fitting @ X-ray Server

Sergey Stepanov[1] and Rebecca Forrest[2]

[1]Argonne National Laboratory, Argonne, IL 60439, USA
[2]University of Houston, Houston, TX, 77004, USA

What is X-ray Server (http://x-server.gmca.aps.anl.gov)?

Goals:

-explore simplest ways to share scientific software with community using WWW
- “do it yourself approach” (no extra layer of software engineers)
- be able to monitor usage, correct errors and prevent misuse
- avoid porting to multiple platforms and maintaining multiple versions
- users always access the latest version
- easy feedback from users and establishing collaborations
- shared X-ray parameters DB extendable by users

Technology:

- http interface (CGI wrappers) to *unchanged* scientific software

History:

- public project launched in 1997
- access to seven X-ray diffraction and scattering simulation routines
- close to 400,000 usages and about 5,500 users
- progress reported @ NOBUGS 2002, 2004, 2006.

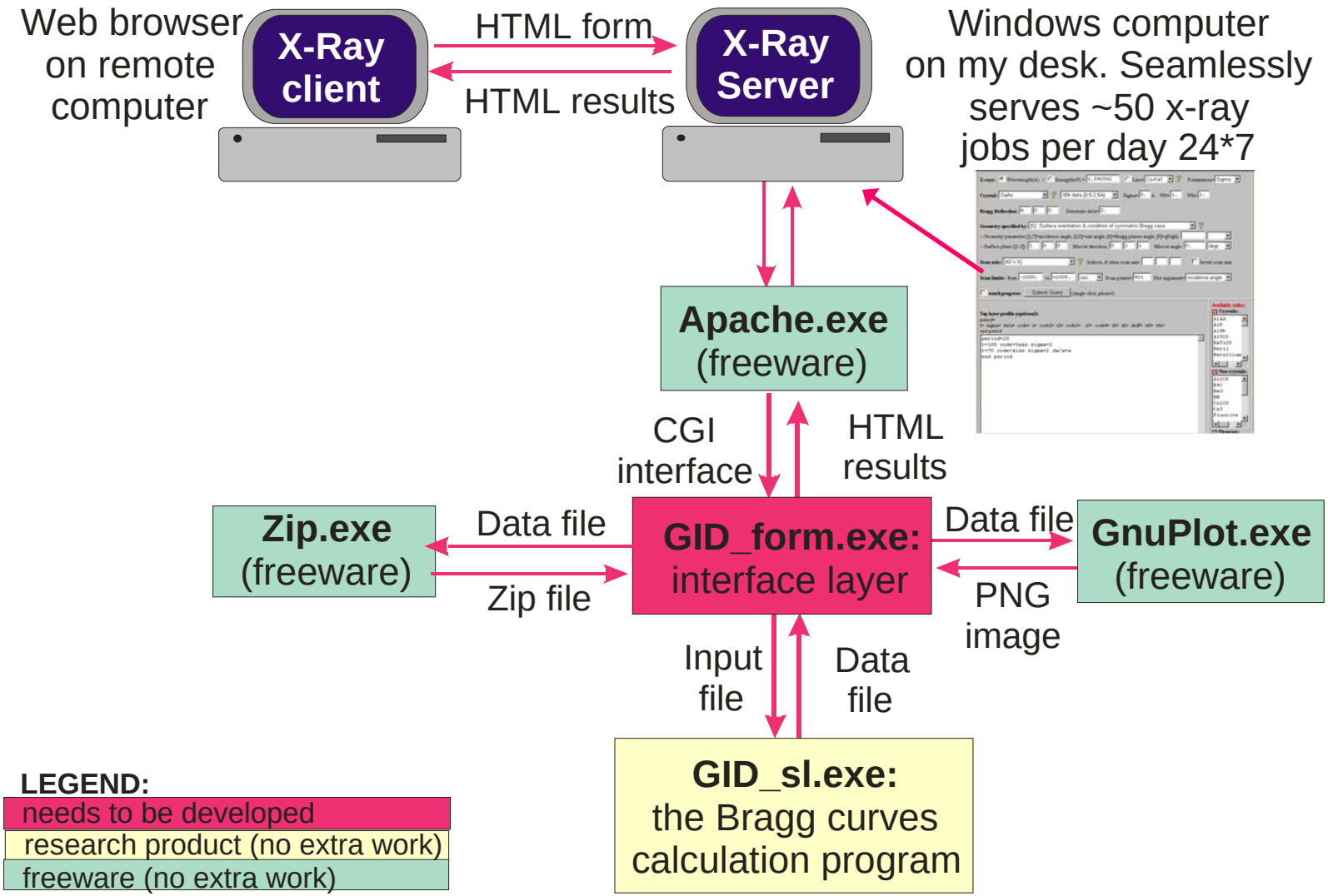
Software available through X-ray Server

This site has been online since 1997 and has served **384,225** x-ray jobs

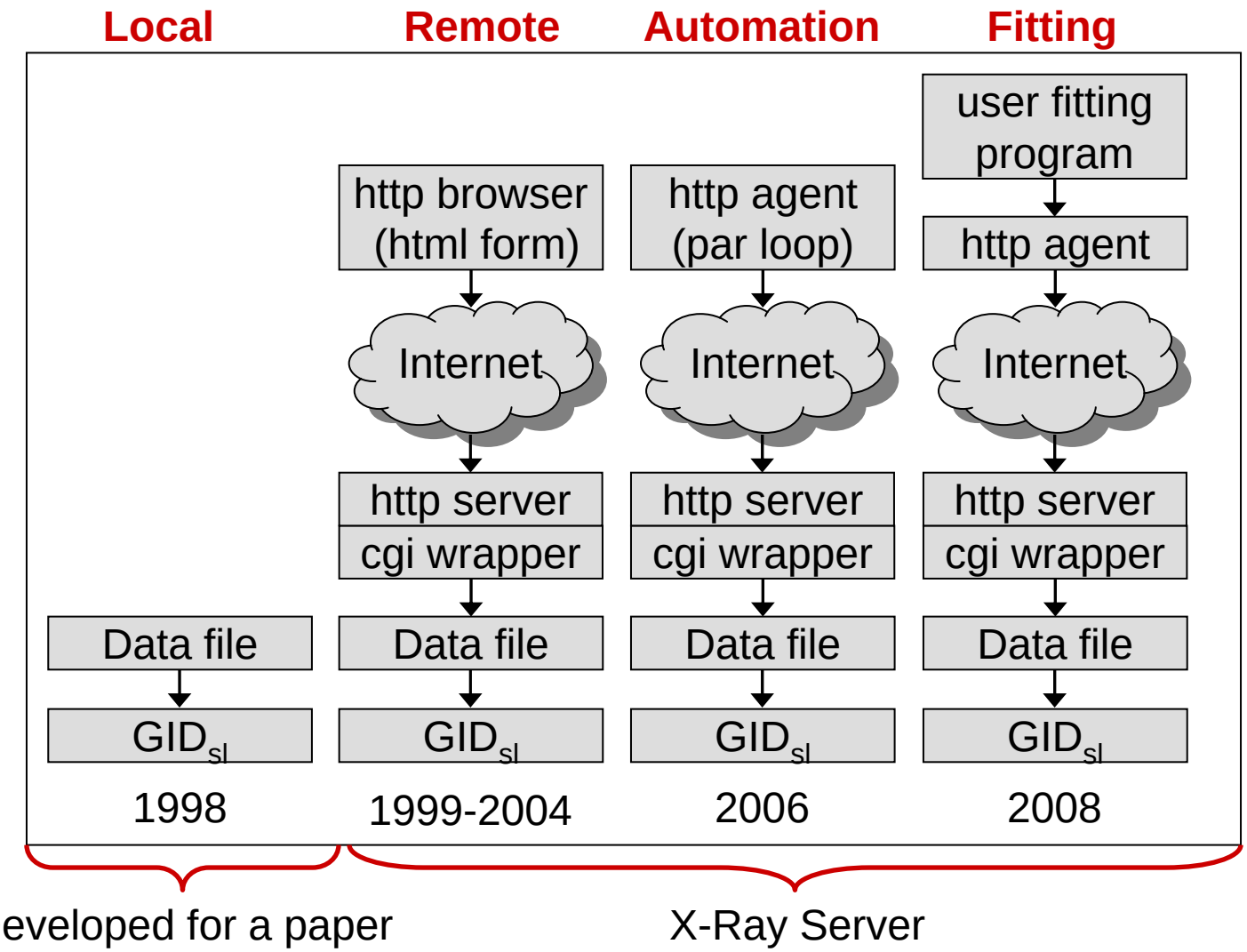
GID_SL on the Web Dynamical x-ray diffraction from strained crystals, multilayers and superlattices at usual and grazing incidence angles	***GID_SL*** on the Web calculates x-ray diffraction curves of strained crystals and multilayers for any Bragg-case diffraction with scans around arbitrary axis. It replaces Takagi-Taupin equations for extremely asymmetric and grazing incidence diffraction.	**237,963 jobs**
x_{0h} on the Web !!! x_{0h} search	***X0h*** interpolates dielectric susceptibilities x_0 and x_h for some crystals and other materials in wide range of x-ray wavelengths with the option to compare data from different databases. ***X0h-search*** is a tool to search for Bragg planes under various conditions.	**97,959 jobs**
TER_SL on the Web X-ray specular reflection from multilayers with rough interfaces at grazing incidence	***TER_SL*** on the Web calculates x-ray specular reflection and standing waves from multilayers with interface roughness and transition layers. It uses a new recursive algorithm converging faster than the recursions by Parratt.	**21,729 jobs**
TRDS_SL on the Web X-ray diffuse scattering at grazing incidence from surface and interface roughness	***TRDS_SL*** on the Web calculates x-ray diffuse scattering from interface roughness. It implements several different models of roughness and can simulate effects of skew roughness transfer, dependence of interface-interface roughness correlations on lateral size of roughness and x-ray scattering from atomic steps.	**14,769 jobs**
MAG_SL on the Web X-ray resonant specular reflection from magnetic multilayers	***MAG_SL*** on the Web calculates x-ray resonant specular reflection from magnetic multilayers with interface roughness and transition layers. It can also supply wavefields for calculations of diffuse scattering from magnetic roughness.	**8,960 jobs**
BRL on the Web X-ray multiple Bragg/Laue diffraction	***BRL*** on the Web applies a novel algorithm to calculate multiple Bragg diffraction of x-rays including the cases of x-ray waves grazing along the crystal surface and Bragg angles at 90 degr.	**2,845 jobs**

Based on ~17 research papers

How it works and what is the effort (NOBUGS'2006 slide)

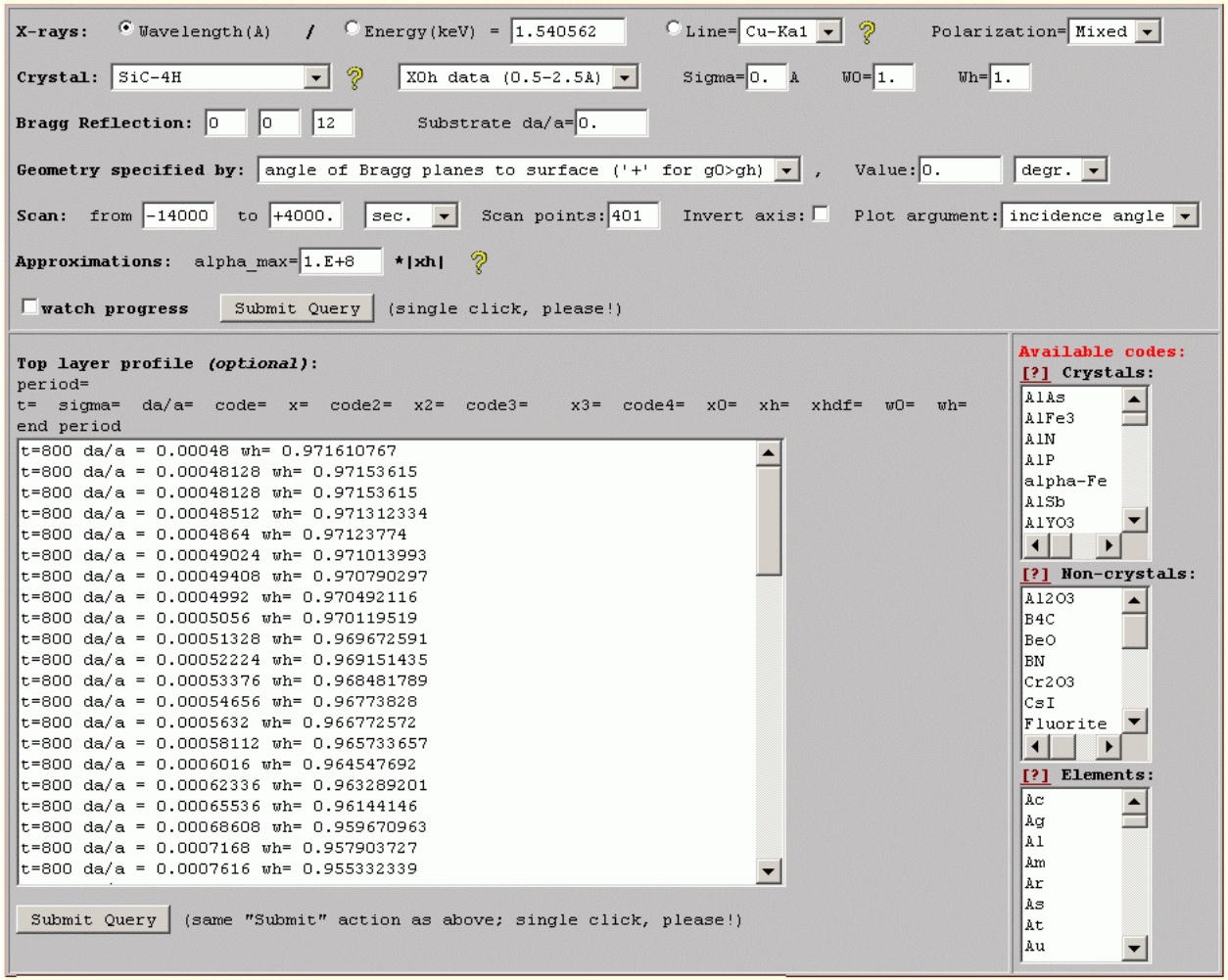


Evolution of X-Ray Server



Stepanov et.al. Phys. Rev. B, **57**, (1998)

Why client-side fitting



Providing fitting at Server for general case is non-practical:

Too many parameters!

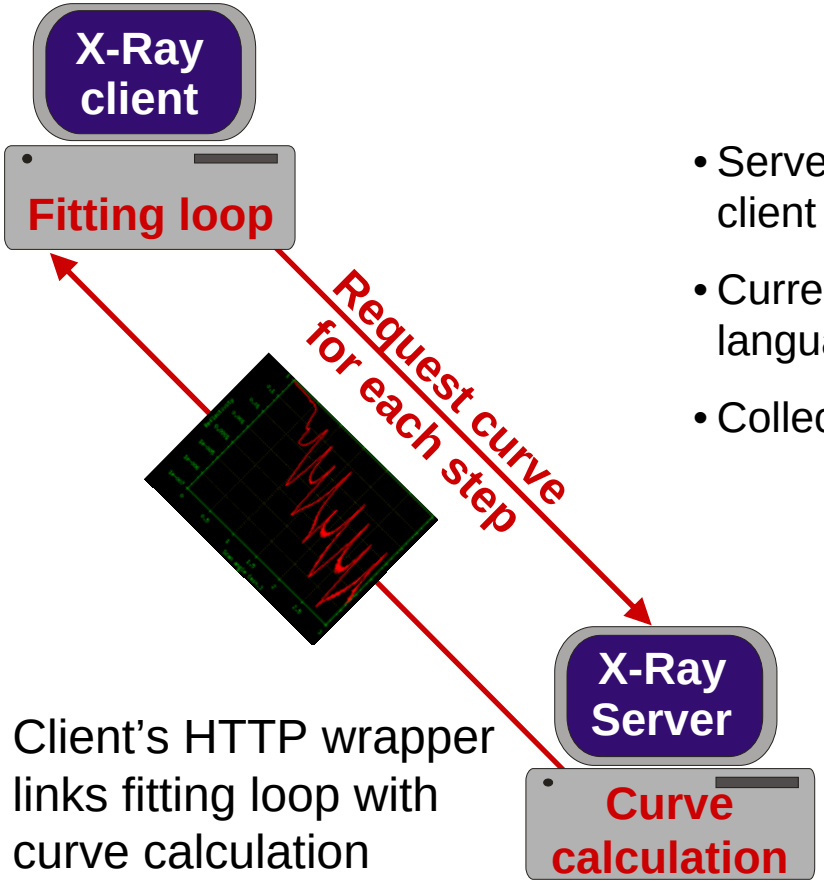
The form at the left is a typical material science application submitted from univ-poitiers.fr. The profile contains 72 lines, i.e. more than 200 parameters.

Also, account for the geometry or resolution of experiment would make the programs case-specific vs. generic.

Client-side wrappers

Client has to do some programming:

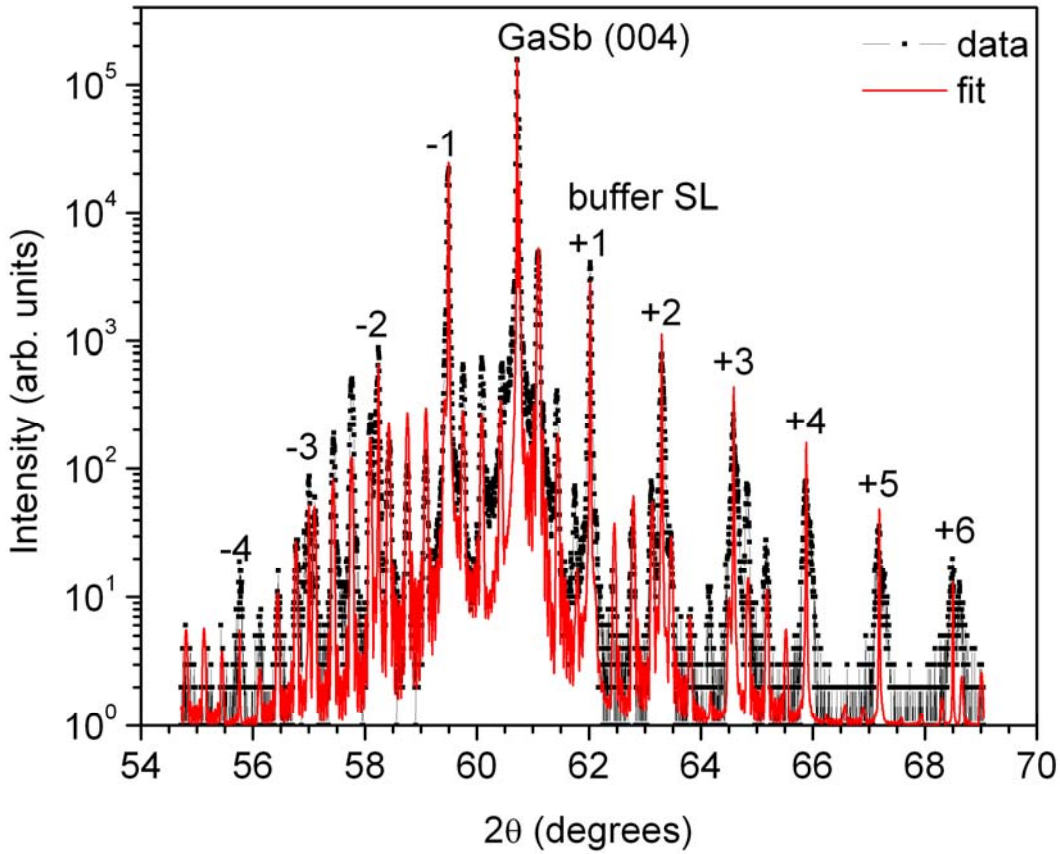
- Write a fitting loop specific to his task
- Write or adapt client-side HTTP wrapper
- Implement convolution of calculated data with his specific resolution function



Client's HTTP wrapper links fitting loop with curve calculation

- Server provides templates (documented examples) of client HTTP wrappers (typically ~100 lines of code)
- Currently in Perl and Bash, but the goal is to have more languages like Python, Tcl, Java, and etc.
- Collect wrappers from users

Example: remote fitting of diffraction from AlSb/AlAs superlattice



The structure consisted of 122 AlSb(25)/AlAs(2) "digital alloy" superlattice buffer layer periods (~10,000 Å thick), followed by five InAs(6)/$Ga_{0.7}In_{0.3}Sb$(11)/InAs(6)/$AlAs_{0.106}Sb_{0.894}$(77) "W" quantum well periods (~1524 Å thick), seven periods (~234 Å) of the same AlSb/AlAs digital-alloy barrier, and a 150 Å GaSb cap layer. The numbers in parentheses are numbers of monolayers.

Only the structure of the buffer layer was of scientific interest, therefore only parameters describing the buffer layer were varied within the fit.

- Client-side fitting and convolution programming at University of Houston (χ^2 simplex algorithm)
- Used HTTP agent examples (templates) provided on X-ray server
- No meetings or phone communications required between Argonne & Houston; only few emails
- Thousands curves calculated in short time & 2 parallel fitting jobs, but no server overload issues
- Less than time 10% overhead due to remote operation (28.9s vs 26.0s per cycle)

S.Stepanov and R. Forrest,
J. Appl. Crystallogr., v.41, No 5, p.958-962, (2008).

Frequent questions and concerns about X-server

If scientific software running on a desktop gets popular, would it cause performance issues?

Has not happened to X-server during 11 years in spite of exponentially growing usage. Scientific servers are not Google – audience is limited. If it happens to some software, then it will need another approach; ditto with highly interactive software.

Deployed solutions are nothing fancy; software engineers could do better. Why not using them?

The solutions are targeting do-it-yourself for scientists. It is important that scientist can monitor his software usage because physical models are based on multiple approximations and it is almost impossible to restrict misuse upfront. That is why software must remain in a form original author could understand. Also typical University research groups cannot afford extra engineers.

How big is overhead with all these layers of wrappers?

As we saw, the overhead was 2.9s. In our case it corresponded to less than 10%, but could be less or more depending on the duration of calculation. The overhead will shrink with faster CPU.

Is there any possibility of parallelization to improve performance?

Not in the used scheme, but performance has not been an issue. X-server frequently processes multiple tasks in parallel without noticeable degradation of productivity (dual-core Dell desktop). Still, there is a mechanism is place to restrict the number of simultaneously running tasks.

Server security: could a malformed script be used to compromise the server by a hacker?

Not happened in 11 years, although probes sometimes amounted thousands a month. Main reason for security: scientific servers have highly specialized interface, no common hacking tools would work. It is another motivation for client-side fitting (no files to upload on the server).

Conclusions

We have described the first implementation of fitting X-ray Bragg diffraction profiles from strained multilayer crystals at a remote WWW-based X-ray software server.

A practical task of fitting X-ray diffraction data from a complex semiconductor superlattice structure was solved with minimum programming on the user side, who otherwise would need to develop a complex dynamical diffraction program modelling a multilayer structure.

This approach is a simple, inexpensive and efficient way to make well established software available to small research groups, while maintaining feedback between software providers and their end users.

This approach will not be appropriate for all applications, for example those involving high demand or massive computations. We expect, however, that the experience gained in this project will be valuable for a rapidly growing number of WWW collaborations in all areas of science.

Acknowledgements

X-Ray Server uses computing resources of the GM/CA Collaborative Access Team at Argonne National Laboratory. GM/CA CAT has been financed with federal funds from the National Cancer Institute (Y1-CO-1020) and the National Institute of General Medical Science (Y1-GM-1104).

The work at the University of Houston was supported in part by the M. Hildred Blewett Scholarship of the American Physical Society, http://www.aps.org. The authors are grateful to Chadwick Canedy and Jerry Meyer at the Naval Research Laboratory for supplying the sample.