EXTENSION OF SASE BANDWIDTH UP TO 2% AS A WAY TO INCREASE NUMBER OF INDEXED IMAGES FOR PROTEIN STRUCTURE DETERMINATION BY FEMTOSECOND X-RAY NANOCRYSTALLOGRAPHY AT THE EUROPEAN XFEL

Svitozar Serkez, Vitali Kocharyan, Evgeni Saldin, Igor Zagorodnov, DESY, Hamburg, Germany
Gianluca Geloni, European XFEL GmbH, Hamburg, Germany
Oleksander Yefanov, CFEL, Hamburg, Germany

Abstract

Experiments at the LCLS confirmed the feasibility of femtosecond nanocrystallography for protein structure determination at near-atomic resolution. These experiments rely on X-ray SASE pulses with a few microradians angular spread, and about 0.2% bandwidth. By indexing individual patterns and then summing all counts in all partial reflections for each index it is possible to extract the square modulus of the structure factor. The number of indexed images and the SASE bandwidth are linked, as an increasing number of Bragg spots per individual image requires an increasing spectral bandwidth. This calls for a few percent SASE bandwidth. Based on start-to-end simulations of the European XFEL baseline, we demonstrate that it is possible to achieve up to a 10-fold increase of the electron energy chirp by strongly compressing a 0.25 nC electron bunch. This allows for data collection with a 2% SASE bandwidth, a few mJ radiation pulse energy and a few fs pulse duration, which would increase the efficiency of protein determination at the European XFEL. We prove this concept with simulations of lysozyme nanocrystals, with a size of about 300 nm.

INTRODUCTION

X-ray crystallography is currently the leading method for imaging macromolecules with atomic resolution. Third generation synchrotron sources allow for successful structure determination of proteins. The size of a typical single crystal used for conventional protein crystallography is in the order of 50μm -500μm [1]. Obtaining sufficiently large crystals is currently a serious stumbling block as regards structure determination. The new technique of femtosecond nanocrystallography is based on data collection from a stream of nanocrystals, and ideally fills the gap between conventional crystallography, which relies on the use of large, single crystals, and single-molecular x-ray diffraction.

The availability of XFELs allows for a new "diffraction before destruction" approach to overcome radiation damage due to the ultrafast and ultrabright nature of the x-ray pulses, compared to the time scale of the damage process [2]-[6]. In fact, if such time-scale is longer than the pulse duration, the diffraction pattern yields information about the undamaged material. Femtosecond nanocrystallography involves sequential illumination of many small crystals of proteins by use of an XFEL source [7]. The high number of photons incident on a specimen are expected to produce measurable diffraction patterns from nanocrystals, enabling structure determination with high resolution also for systems that can only be crystallized into very small crystals and are not suitable, therefore, for conventional crystallography. Each crystal is used for one exposure only, and the final integrated Bragg intensities must be constructed from "snapshot" diffraction patterns containing partially recorded intensities. Each pattern corresponds to a different crystal at random orientation. Experiments at the LCLS [8] confirmed the feasibility of the "diffraction before destruction" method at near atomic resolution using crystals ranging from 0.2 μm to 3 μm [7]. This method relies on x-ray SASE pulses with a few mJ energy, a few microradians angular spread, and about 0.2% bandwidth with a photon energy range between 2 keV and 9 keV.

The success of nanocrystallography depends on the robustness of the procedure for pattern determination. After acquisition, diffraction patterns are analyzed to assign indexes to Bragg peaks (indexing procedure). Each of the indexed peaks is integrated in order to obtain an intensity. Intensities of corresponding peaks are averaged within the dataset. The table of peak indexes and intensities obtained in this way is used for protein structure determination. Indexing algorithms used in crystallography enable to determine the orientation of the diffraction data from a single crystal when a relatively large number of reflections are recorded. Femtosecond nanocrystallography brings new challenges to data processing [9]. The problem is that single snapshots of crystal diffraction patterns may contain very few reflections, which are not enough for indexing. In this paper we will show how to overcome this obstacle.

The number of Bragg peaks is proportional to the bandwidth of the incident radiation pulse. Considering the baseline configuration of the European XFEL [10], and based on start-to-end simulations, we demonstrate here that it is possible to achieve a tenfold increase in bandwidth by strongly compressing electron bunches with a charge of 0.25 nC up to 45 kA. This allows data collection with a 2% bandwidth, a few mJ radiation pulse energy, a few fs pulse duration, and a photon energy range between 2 keV and 6 keV, which is the most preferable range for nanocrystallography [11].
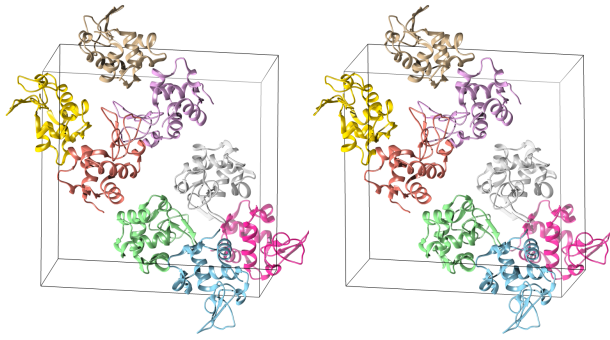
Figure 1: Stereo picture of the Lysozyme unit cell, containing 8 molecules. The cell parameters are: $a = 79.18\text{Å}$, $b = 79.18\text{Å}$, $c = 38.1\text{Å}$, data from [16].

The generation of x-ray SASE pulses at the European XFEL using strongly compressed electron bunches has many advantages, primarily because of the very high peak power and very short pulse duration that can be achieved in this way [12]. Here we demonstrate (see Section **FEL Studies** for more details) that a few TW power-level can be achieved in the SASE regime, and with 3 fs-long pulses. For electron bunches with very high peak-current, the wakefields (mainly due to the undulator vacuum pipe) have a very important effect on the lasing process. In fact, the variation in energy within the electron bunch due to wakefield effects is large, and yields, as a consequence, a tenfold increase in the SASE radiation bandwidth. However, as we demonstrate in the next section, a large bandwidth presents many advantages for nanocrystallography, and provides a route for increasing the efficiency of data processing.

POTENTIAL FOR FEMTOSECOND X-RAY NANOCRYSTALLOGRAPHY USING SASE PULSES WITH 2% BANDWIDTH

In nanocrystallography [7], crystals of a few hundred nanometers or smaller are delivered to the interaction region in a liquid jet. When a crystal in the jet is hit at random and unknown orientation, such orientation can readily be determined from the diffraction pattern itself, by indexing the Bragg peaks [9]. Several auto-indexing programs have been developed for crystallography. They search for a repeating lattice in the measured diffraction pattern, knowing the mapping of that pattern onto the Ewald sphere [13]-[15]. In the practical case of a crystal with different unit-cell spacings in each dimension, the basis vectors of the reciprocal lattice can be identified quite readily from the reciprocal lattice spacings observed in the diffraction pattern. Each Bragg peak in the pattern can be thereby indexed by its 3D Miller indexes, and properly accumulated. In this way, diffraction data are build up, averaging over all crystal shapes and sizes, and yielding complete 3D information as if one were collecting data from a single undamaged rotating crystal.
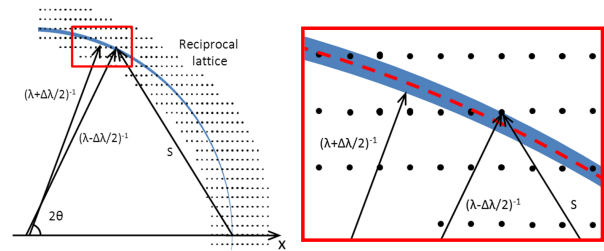
Figure 2: Cross section of the reciprocal lattice construction for imaging of single nanocrystals. The shaded portion of the reciprocal lattice laying between the two spheres of reflection indicated in the plot satisfies the diffraction condition for different values of the wavelength λ. The thickness of the reflection sphere results from the finite wavelength bandwidth.

Each x-ray pulse produces a diffraction pattern from a single crystal. The resulting data set consists of thousands of diffraction patterns from randomly oriented crystals, recorded under "snap-shot conditions". Each pattern is not angle-integrated across the Bragg reflections, and is affected by beam divergence, energy spread, broadening by the small size and possible lattice imperfections of the crystals. We assume that the pulse duration is sufficiently short, so that no radiation damage effects occur. Our aim here is to demonstrate that a tenfold increase in the SASE bandwidth results in an increase in the number of indexed images at a fixed number of hits. In this paper we test this concept with simulations of lysozyme nanocrystals, see Fig. 1. The orientation of each nanocrystal is determined from the diffraction pattern by using automated indexing software such as CrystFEL [17]. The spectral width of the x-ray beam may be simulated by summing diffracted intensities over a spread of photon energies. For simulations in this work we neglect the photon beam divergence, because this effect causes a much smaller effect than the tenfold increase in bandwidth.

Results from our simulated experiment can be represented in terms of the reciprocal lattice concept, as shown in Fig. 2. There we assume that the spectrum of the incident radiation pulse has a stepped profile between $\lambda_{\max}$ and $\lambda_{\min}$. Let us consider two Ewald spheres with radii $1/\lambda_{\min}$ and $1/\lambda_{\max}$, tangent to the origin of the reciprocal lattice, as shown in Fig. 2. The shaded region between them is accessible to the experiment, as the diffraction condition is satisfied for all reciprocal lattice points within it. We model the diffraction patterns obtained in this case as an average over the number N_p of diffraction patterns, obtained at wavelengths within the bandwidth.

A comparison between diffraction patterns obtained with a monochromatic and a polychromatic x-ray pulse is shown in Fig. 3. The radiation pulses are short enough (about 3 fs) to overcome crystal destruction. Therefore, the simulation was performed neglecting radiation damage. Patterns were calculated for crystals at random orientations. Fig. 3
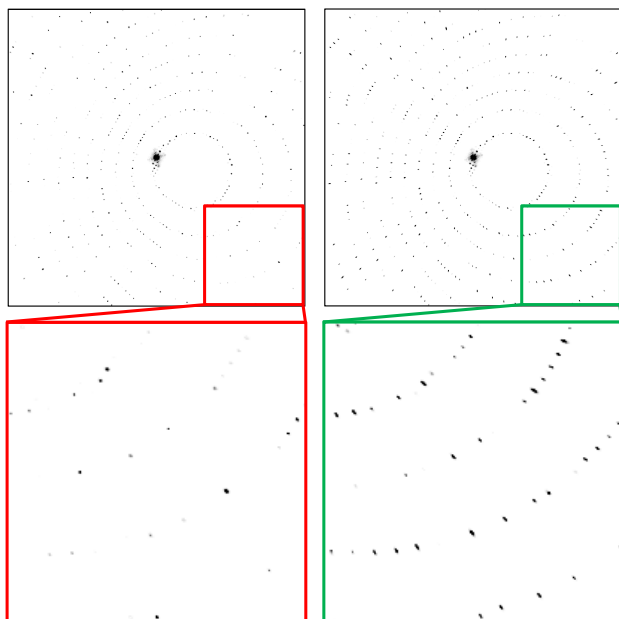
Figure 3: Left plot: Diffraction pattern simulated for a single lysozyme crystal ($20 \times 20 \times 40$ cells) exposed to an x-ray pulse with a wavelength of 2.0 $\mathring{A}$. The detector considered here is 176 mm$\times$ 176 mm in size, and is located at 80 mm distance from the sample. The simulation was performed for 2 mJ photon pulse energy and focus size of 200 nm. Right: Diffraction pattern simulated for the same crystal and detector exposed to an X-ray pulse with finite spectral width, simulated by summing diffracted intensities over a spread of wavelengths between $\lambda_{\min} = 2.00$ $\mathring{A}$ and $\lambda_{\max} = 2.03$ $\mathring{A}$. The average diffraction patterns, obtained at wavelengths within the bandwidth is shown for $N_p = 16$.

shows a typical simulation, clearly showing the significant increase in number of Bragg peaks per individual image in the case of a polychromatic x-ray beam. Our results suggest that if the bandwidth of the incident x-ray beam is tenfold increased, the number of indexed reflections will be much higher.

FEL STUDIES

We performed a feasibility study with the help of the FEL code Genesis 1.3 [18] running on a parallel machine. We will present results for the SASE3 FEL line of the European XFEL, based on a statistical analysis consisting of 100 runs. The overall undulator and electron beam parameters used in the simulations are presented in Table 1.

The expected beam parameters at the entrance of the SASE3 undulator, and the resistive wake inside the undulator are shown in Fig. 4 and Fig. 5, see [19]. Our calculations account for both wakes and quantum fluctuations in the SASE1 undulator. The nonlinear increase of the energy spread is a consequence of the quadratic superposition of the initial energy spread $\sigma_0(s)$, s being the coordinate inside the bunch, and the contribution due to quantum dif-
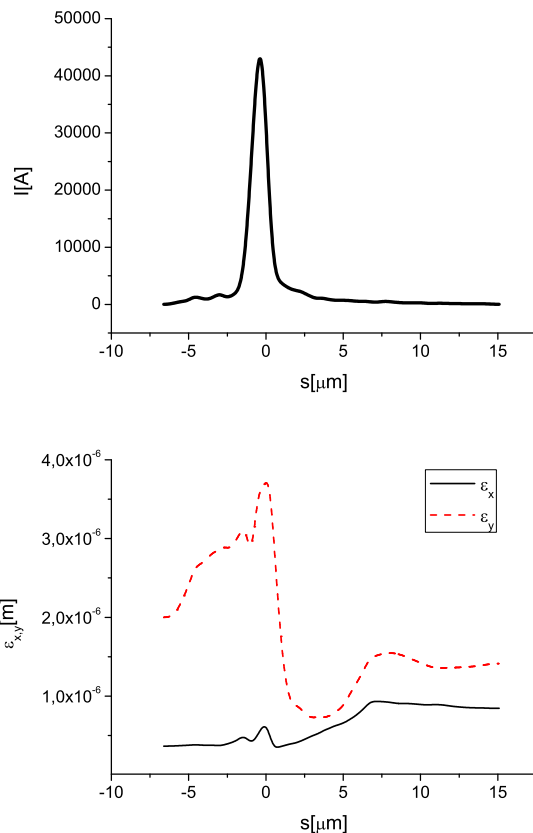

Figure 4: Results from electron beam start-to-end simulations. Top to bottom: Current profile at the entrance of SASE3 and normalized emittance as a function of the position inside the electron beam, at the entrance of SASE3.

Table 1: Parameters for the Mode of Operation at the European XFEL Used in this Paper

	Units	
Undulator period	mm	68
Periods per cell	-	73
Total number of cells	-	21
Intersection length	m	1.1
Energy	GeV	17.5
Charge	nC	0.25

fusion δ, yielding $\sigma^2(s) = \sigma_0^2(s) + \delta^2$. Note the difference in energy chirp and energy spread at the entrance of SASE1 and at the entrance of SASE3. The additional energy chirp induced during the passage through SASE1, due to the presence of resistive wakes, is very helpful in our case, because it finally leads to a larger bandwidth.

Due to collective effects in the bunch compression system, emittances in the horizontal and vertical directions are significantly different. As a result, the electron beam looks highly asymmetric in the transverse plane: in the horizontal direction $\sigma_x \sim 20\mu$m, while in the vertical direction $\sigma_y \sim 50\mu$m. The evolution of the transverse electron bunch

Figure 6: Evolution of the horizontal and vertical dimensions of the electron bunch as a function of the distance inside the SASE3 undulator. The plots refer to the longitudinal position inside the bunch corresponding to the maximum current value.
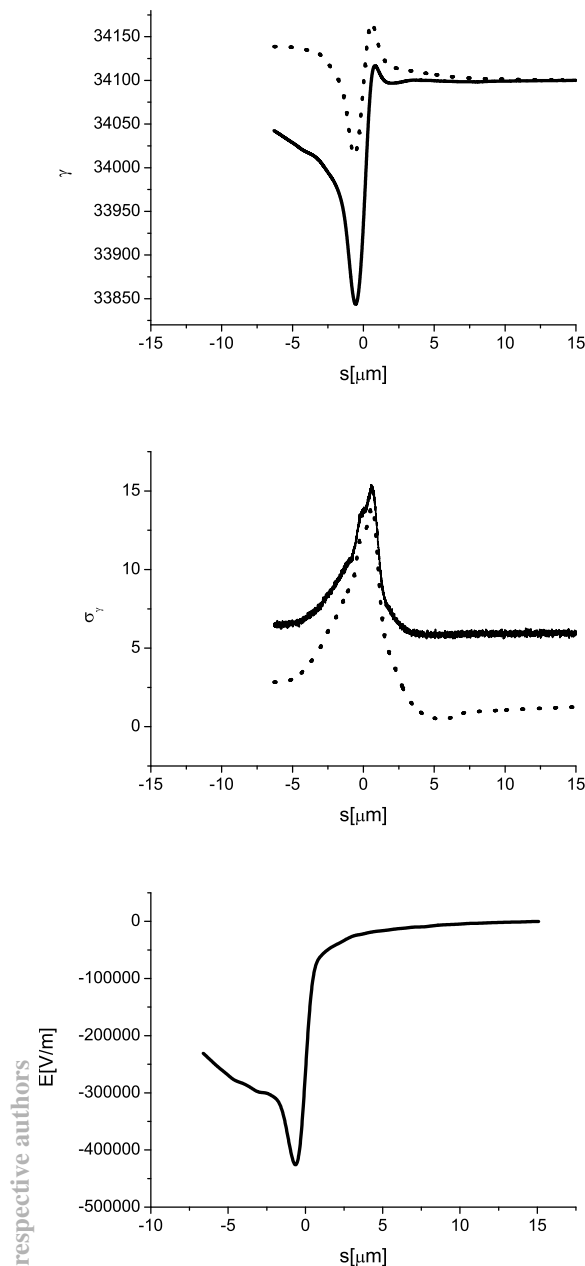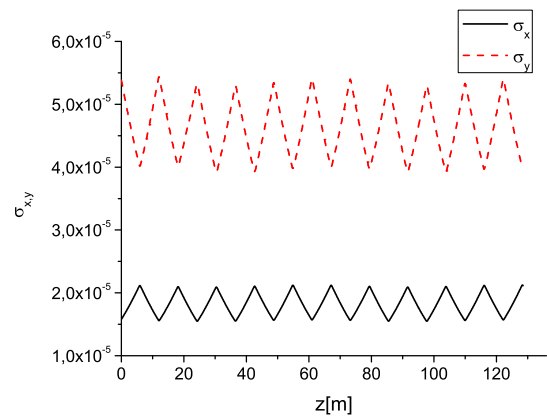

Figure 5: Results from electron beam start-to-end simulations. Top to bottom: Energy profile along the beam (Dashed line: at the entrance of SASE1. Solid line: at the entrance of SASE3). Electron beam energy spread profile (Dashed line: at the entrance of SASE1. Solid line: at the entrance of SASE3). Resistive wall wake within the undulator.

dimensions are plotted in Fig. 6.

The output SASE power distribution and spectrum after the 21 undulator segments of SASE3 are shown in Fig. 7. The evolution of the energy per pulse and of the energy fluctuations as a function of the undulator length are shown in Fig. 8. Finally, the distribution of the radiation pulse energy per unit surface and the angular distribution of the radiation pulse energy at the exit of SASE3 undulator are shown in Fig. 9.

Note that, since the electron beam is transversely asymmetric, the output radiation beam is asymmetric too. Intuition would incorrectly suggest that a larger size of the electron bunch should be responsible for a larger size of the radiation beam, but a quick comparison of Fig. 6 and Fig. 9 show that this is not the case in reality. In fact we see that, in the vertical direction, the size of the photon beam is about 50 μm, practically the same as the electron beam. This means that, within a gain length, the expansion of the radiation beam due to diffraction in the vertical direction is much less than the vertical size of the electron bunch. On the contrary, in the other direction radiation expands outside of the electron bunch due to diffraction effects, and the photon beam size grows to about 100 μm, which is 5 times larger than the electron bunch size.

Finally, it should be appreciated that the characteristics of the radiation beam produced as described in this article are very suitable for nanocrystallography. The photon energy is about 6 keV, and radiation is delivered in 3 fs pulses with an energy of about 2.5 mJ . Our study shows that for the particular mode of operation investigated here, we have straightfowardly 1 TW power in the SASE regime already, which constitutes an increase of about two orders of magnitude compared with the nominal mode of operation.
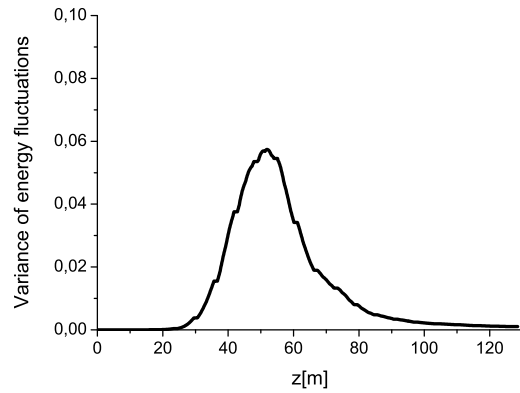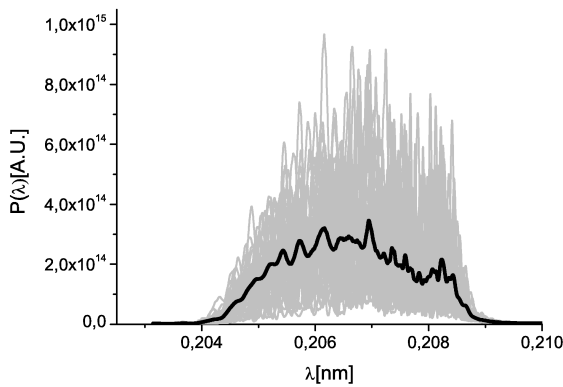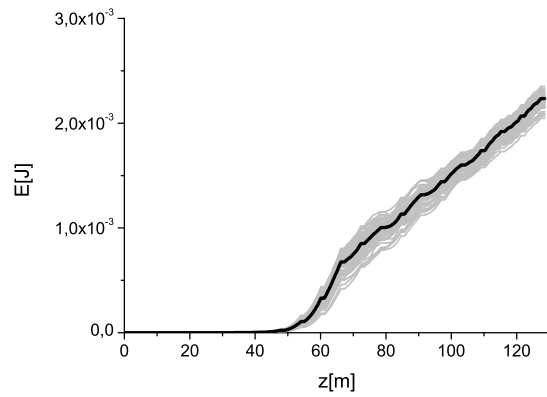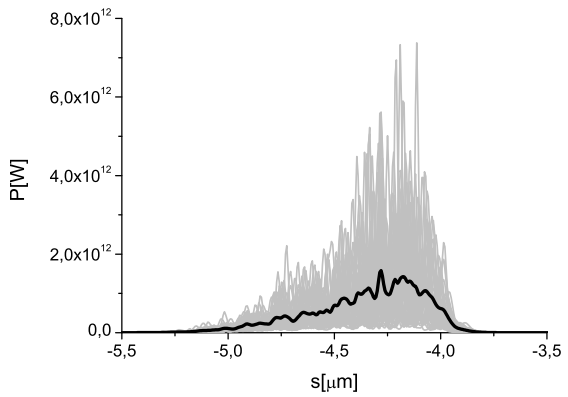
Figure 7: Power distribution and spectrum of the X-ray radiation pulse after the SASE3 undulator. Grey lines refer to single shot realizations, the black line refers to the average over a hundred realizations.

Figure 8: Evolution of the energy per pulse and of the energy fluctuations as a function of the undulator length. Grey lines refer to single shot realizations, the black line refers to the average over a hundred realizations.

CONCLUSIONS

Output characteristics of the European XFEL have been previously studied assuming an operation point at 5 kA peak current. In this case, the baseline SASE undulator sources will saturate at about 50 GW (see e.g. [10]). This power limit is very far from the multi-TW level required for bio-imaging applications. In this paper we explore the possibility to go well beyond the nominal 5 kA peak current. In order to illustrate the potential of this approach we consider a 0.25 nC bunch compressed up to 45 kA peak current. Based on start-to-end simulations it is shown here that 2 TW power could be generated in the SASE regime for a photon energy range between 2 keV and 6 keV, which is optimal for femtosecond X-ray nanocrystallography and single biomolecular imaging[1]. This example illustrates the potential for improving the performance of the European

XFEL without additional hardware[2]. This solution to generate TW power mode of operation is not without complexities. The price for using a very high peak-current is operation with a large energy chirp within the electron bunch, yielding a large SASE radiation bandwidth. However, it is shown here that there are applications like nanocrystallography, where x-ray radiation pulses with a few percent bandwidth present many advantages, compared to those produced in the nominal SASE mode of operation, which have a relative bandwidth of a fraction of a percent.

ACKNOWLEDGEMENTS

---

[1]Numerical calculations show that, using undulator tapering in the SASE mode of operation, it is possible to reach 2 TW peak power at 6 keV. However, such undulator tapering additionally reduces the bandwidth from 2% in an uniform undulator to about 1%. Since our goal here is to maximize the bandwidth for applications in nanocrystallography, in this article we consider the uniform undulator option, allowing for generation of radiation pulses of 1 TW power and 2% bandwidth.

[2]Our simulations are based on the 21 cells foreseen for the SASE3 undulator beamline [10], Table 1.

REFERENCES

[1] M. S. Hunter et al., Biophysical Journal 100, p.198 (2011).

[2] J. Hajdu, Curr. Opin. Struct. Biol. 10, 569 (2000)

[3] R. Neutze et al., Nature 406, 752 (2000)

[4] H. Chapman et al., Nat. Phys. 2, p. 839 (2006).

[5] K. J. Gaffney and H. N. Chapman, Science 316, 1444 (2007)

[6] M. M. Seibert et al., Nature 470 (7332) 78-81 (2011)

[7] H. Chapman et al., Nature 470, p.73 (2011).

[8] P. Emma et al., Nature photonics `doi:10.1038/nphoton.`
     `2010.176` (2010)

[9] Kirian et al., Opt Express 18, 5713-5723 (2010).

[10] T. Tschentscher for the European XFEL project team, "Lay-
     out of the x-ray systems at the European XFEL", XFEL.EU
     TN-11-001 (2011).

[11] S. Baradaran et al., LCLS-II New Instruments Workshops
     Report, SLAC-R-993 (2012), see Section 4.3.2. by H. Chap-
     man et al., and Section 4.3.3. by F. R. N. C. Maia et al.

[12] T. Limberg et al., "Optimized bunch compressor system
     for the European XFEL", Proceedings of 2005 Particle Ac-
     celerator Conference, Knoxville, Tennessee, 1236, `http:`
     `//jacow.org/P05/PAPERS/RPPT011.PDF` (2005)

[13] Z. Otwinowski, W. Minor , Methods Enzymol., 276, 307-
     326 (1997).

[14] I. Steller, R. Bolotovsky, M. Rossmann J. Appl. Crystallogr.,
     30 1036-1040 (1997).

[15] A. Leslie, Acta Crystallogr., D55, 1696-1702 (1999).

[16] The structure can be found in the Protein Data Bank:
     `http://www.rcsb.org/pdb/explore/explore.do?`
     `structureId=4ETD`

[17] T. A. White, et al., Applied Crystallography 45, p. 33
     (2012).

[18] S. Reiche et al., Nucl. Instr. and Meth. A 429, 243 (1999).

[19] I. Zagorodnov, "Beam Dynamics Simulations for XFEL",
     `http://www.desy.de/xfel-beam/s2e` (2011), and
     "Compression scenarios for the European XFEL",
     `http://www.desy.de/fel-beam/data/talks/files/`
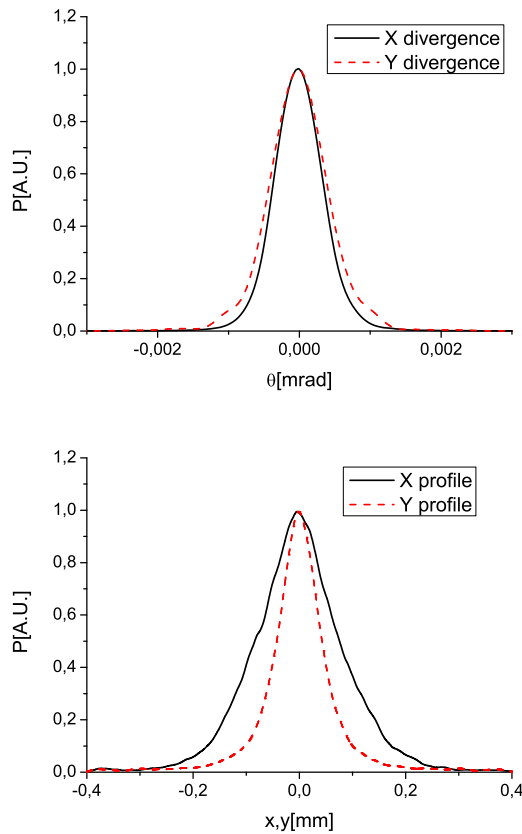   `Zagorodnov_ACC2012_ready_new.pptx`, (2012).

Figure 9: Angular distribution of the radiation pulse energy at the exit of the SASE3 undulator (top plot) and distribution of the radiation pulse energy per unit surface (bottom plot).