Computational Statistics on the Sphere

Ben Wandelt

Institut d'Astrophysique de Paris (IAP)
Lagrange Institute Paris (ILP)
UPMC
Sorbonne Université

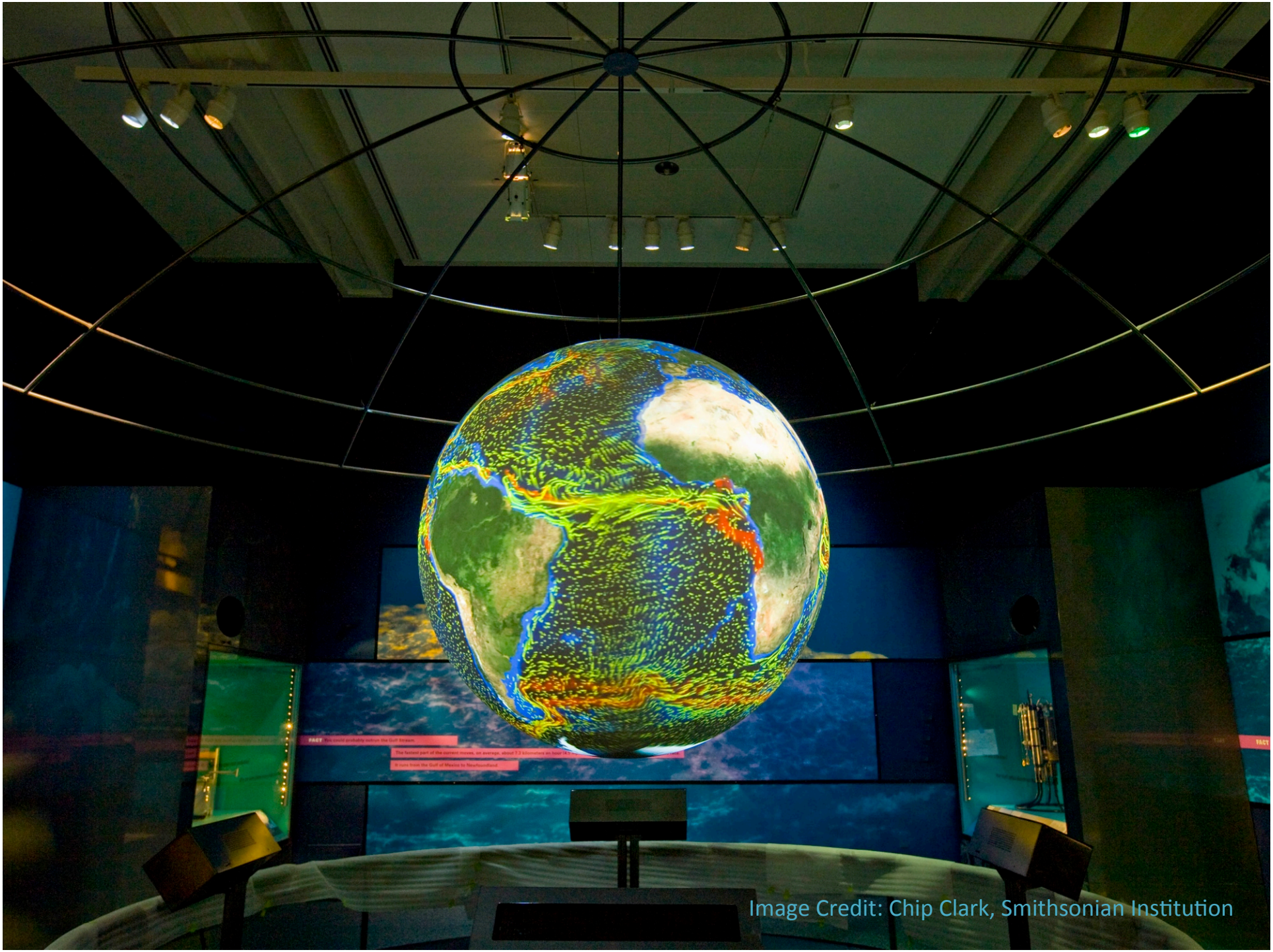
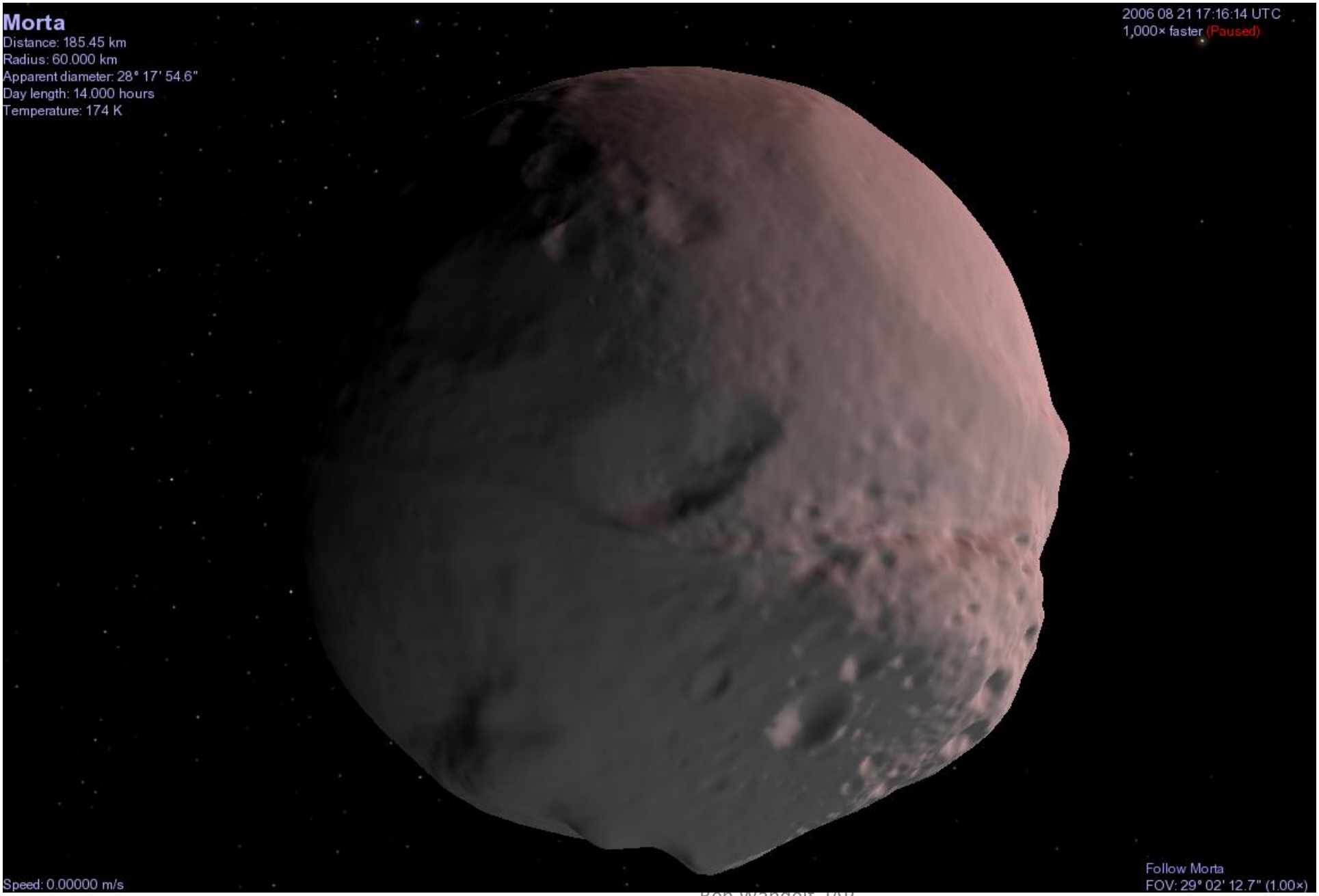
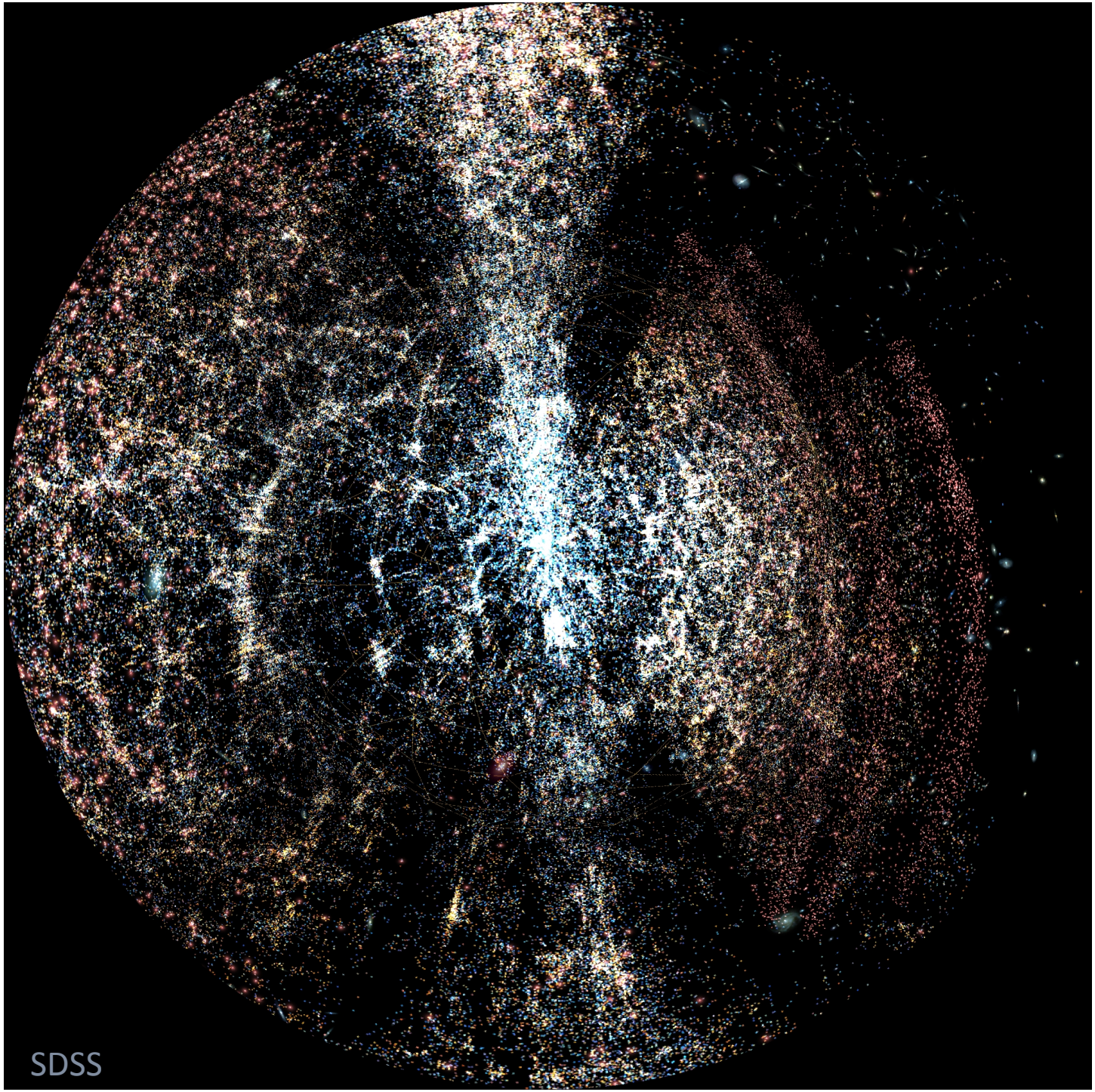
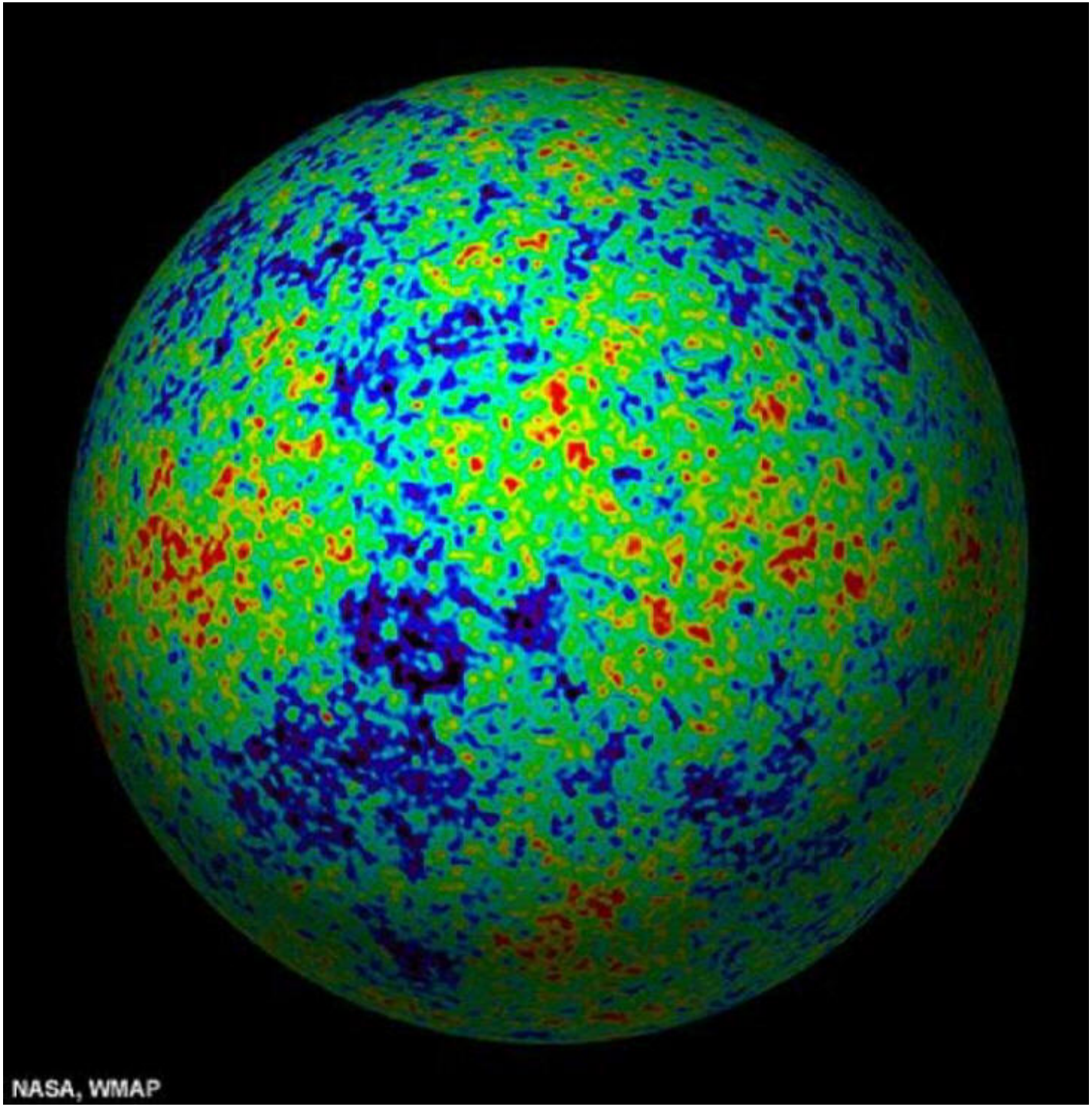


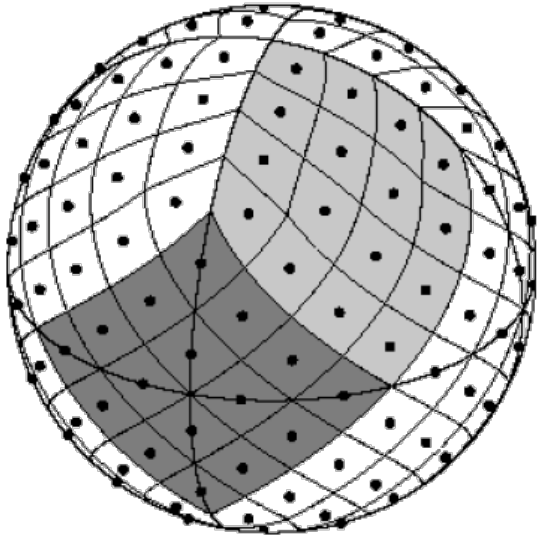
Image Credit: Chip Clark, Smithsonian Institution


Morta
Distance: 185.45 km
Radius: 60.000 km
Apparent diameter: 28° 17' 54.6"
Day length: 14.000 hours
Temperature: 174 K
2006 08 21 17:16:14 UTC
1,000× faster (Paused)
Speed: 0.00000 m/s
Follow Morta
FOV: 29° 02' 12.7" (1.00×)


SDSS


NASA, WMAP

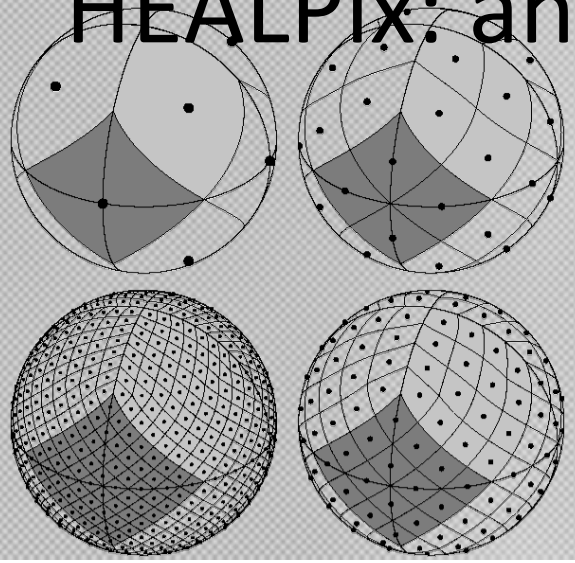
Pixelizing the sphere: HEALPix



healpix.jpl.nasa.gov

Górski, Hivon, BDW 1999; Górski et al 2005

HEALPix: an all-purpose pixelization

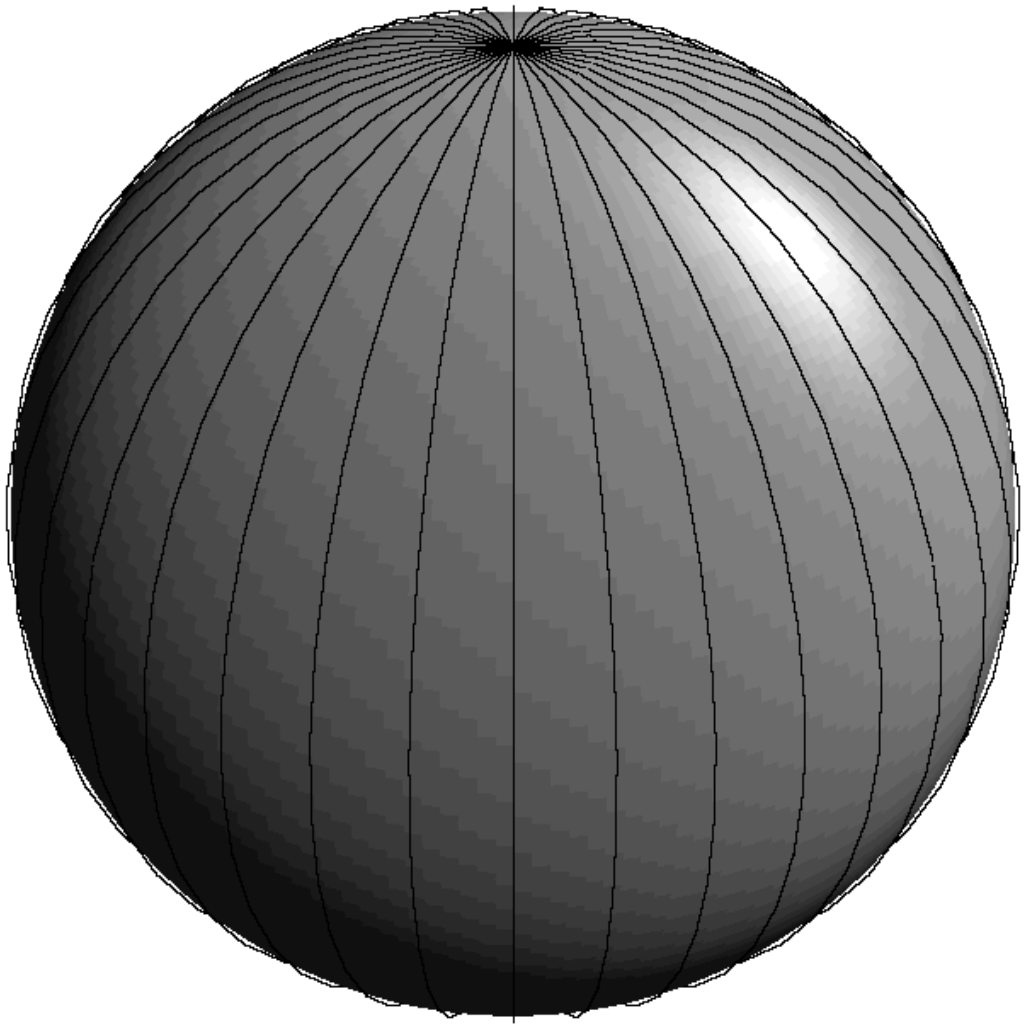


- Hierarchical grid enables adaptive refinement
- Equal Area pixels: approximately uniform coverage on the sphere, good quadrature properties
- Equal Latitude rings: fast SHTs in $O(L^3)$ ops for bandlimit L
- Symmetries: simplifies code and allows compressing distance graph between pixels
- Analytic description of pixel boundaries

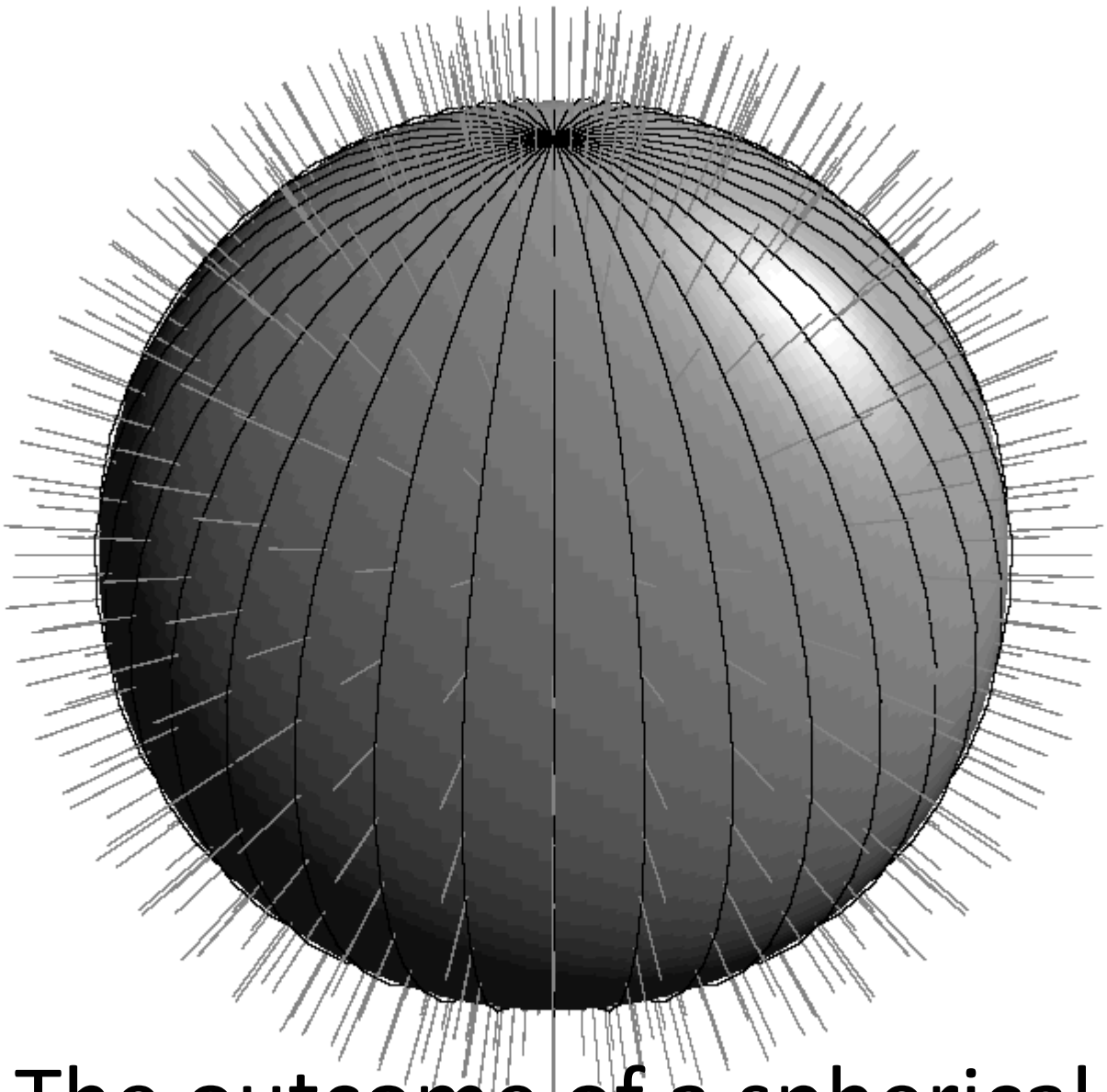
Are you spending a lot of time calculating convolutions on the sphere?

- Object detection
- PDE solver (pseudo-spectral methods)
- Optimal filtering
- Simulating random fields on the sphere
- Wavelet decompositions (including asymmetric, steerable)
- Optical smoothing in simulations
- Power spectrum inference, Wiener filtering:

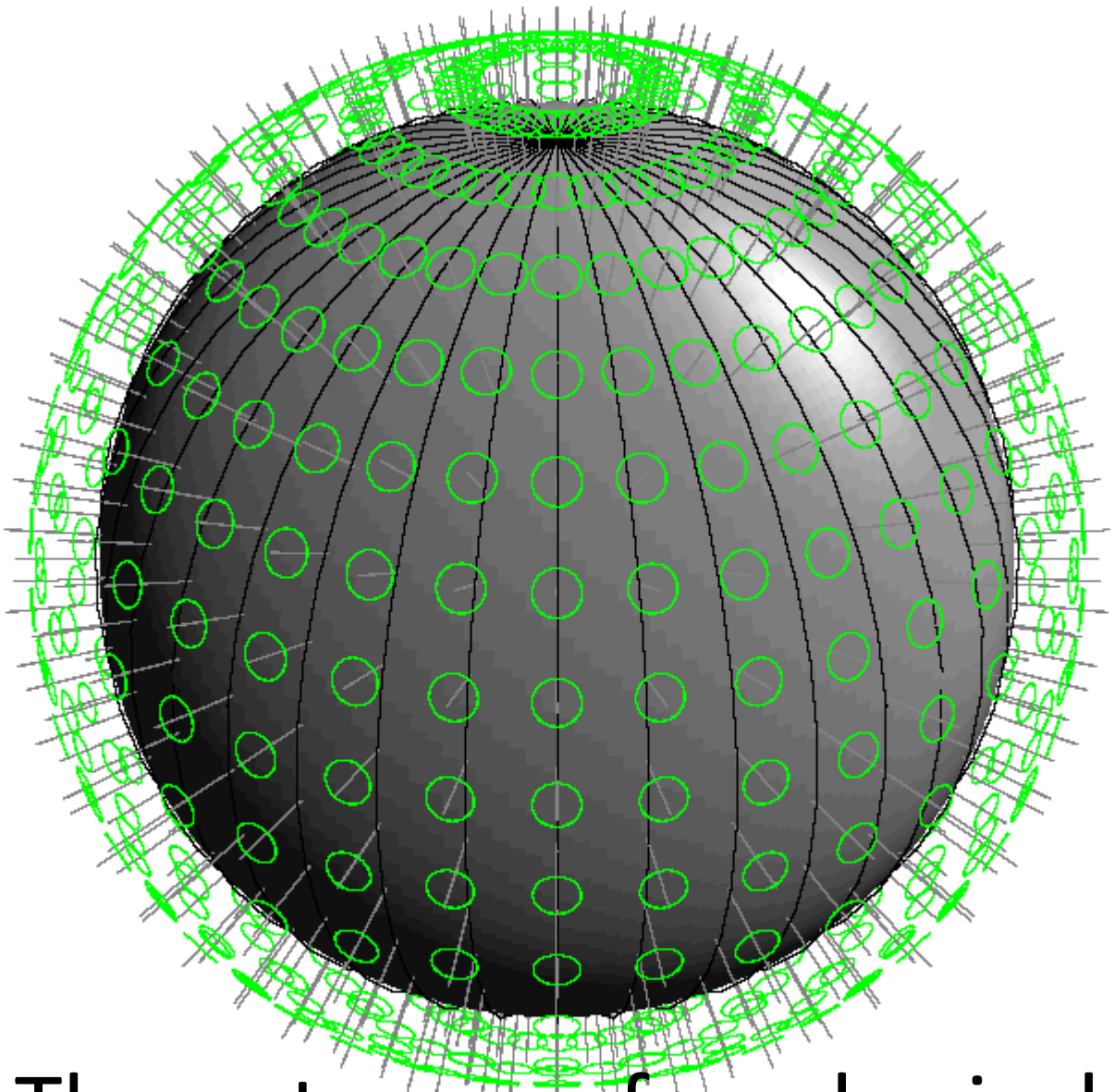
$$\left(S^{-1} + \sum_{channels} B_i N^{-1} B_i \right) m = \left(\sum_{channels} B_i N^{-1} B_i \right) d$$



The outcome of a spherical convolution is a function over SO(3)



The outcome of a spherical convolution is a function over SO(3)



The outcome of a spherical convolution is a function over SO(3)

Accelerating spherical convolution

In general, convolution on the sphere can be written as

$$T(\Phi_2, \Theta, \Phi_1) = \int d\Omega \, [\widehat{D}(\Phi_2, \Theta, \Phi_1) b](\vec{\gamma})^* s(\vec{\gamma})$$

This expression takes of order L^5 operations to compute:

They key step leading to massively accelerated convolution is the factorisation of the rotation operator (Risbo 1996)

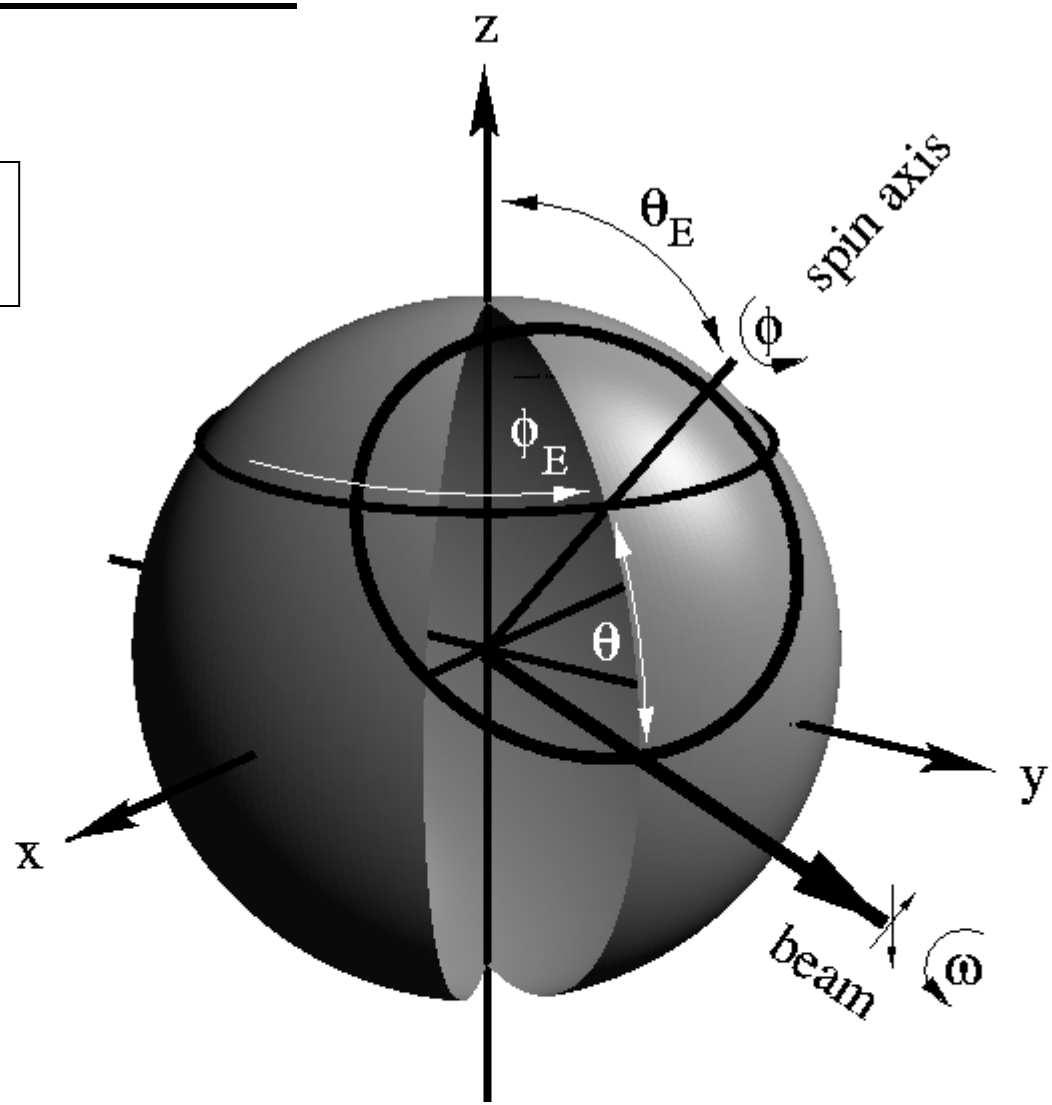
$$\widehat{D}(\Phi_2, \Theta, \Phi_1) \equiv \widehat{D}(\phi_E, \theta_E, 0) \widehat{D}(\phi, \theta, \omega)$$

$$T(\phi_E, \phi, \omega) = \int d\Omega [\widehat{D}(\phi_E, \theta_E, 0) \widehat{D}(\phi, \theta, \omega) b](\vec{\gamma})^* s(\vec{\gamma})$$

Convolution Coordinates

$$\widehat{D}(\Phi_2, \Theta, \Phi_1) \equiv \widehat{D}(\phi_E, \theta_E, 0)\widehat{D}(\phi, \theta, \omega)$$

This factorisation corresponds to moving the beam into place using two successive rotations.



Accelerating spherical convolution II

$$\widehat{D}(\Phi_2,\Theta,\Phi_1) \equiv \widehat{D}(\phi_E,\theta_E,0)\widehat{D}(\phi,\theta,\omega)$$

$$T(\phi_E,\phi,\omega) = \int d\Omega[\widehat{D}(\phi_E,\theta_E,0)\widehat{D}(\phi,\theta,\omega)b](\vec{\gamma})^* s(\vec{\gamma})$$

The special property of this factorisation is that the Fourier transform of the convolution takes only O(L^4) operations to compute!

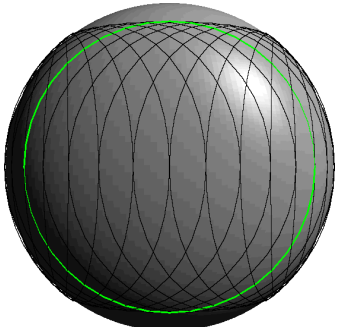
$$T_{m\,m'\,m''} = \sum_l s_{lm} d^l_{m\,M}(\theta_E) d^l_{M\,M'}(\theta) b^*_{lM'}$$

There are special cases which reduce to O(L^3):

1) Basic Scan Paths (Ring Torus)

$$T_{m\,m'}(\omega = 0) = \sum_l s_{lm} d^l_{m\,m'}(\theta_E) X_{lm'}$$

$$X_{lm} \equiv \sum_M d^l_{m\,M}(\theta) b^*_{lM}$$



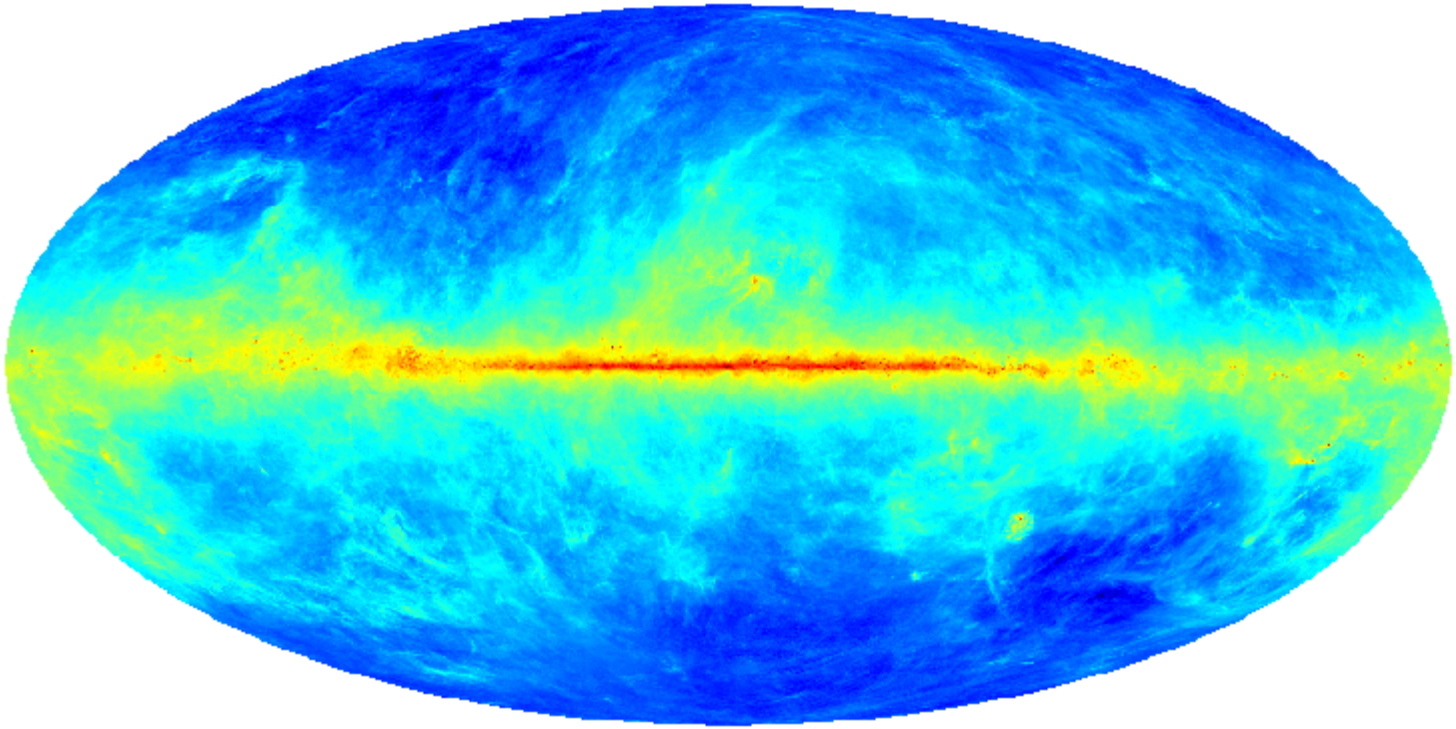
2) Azimuthally symmetric beam reduces to standard formula

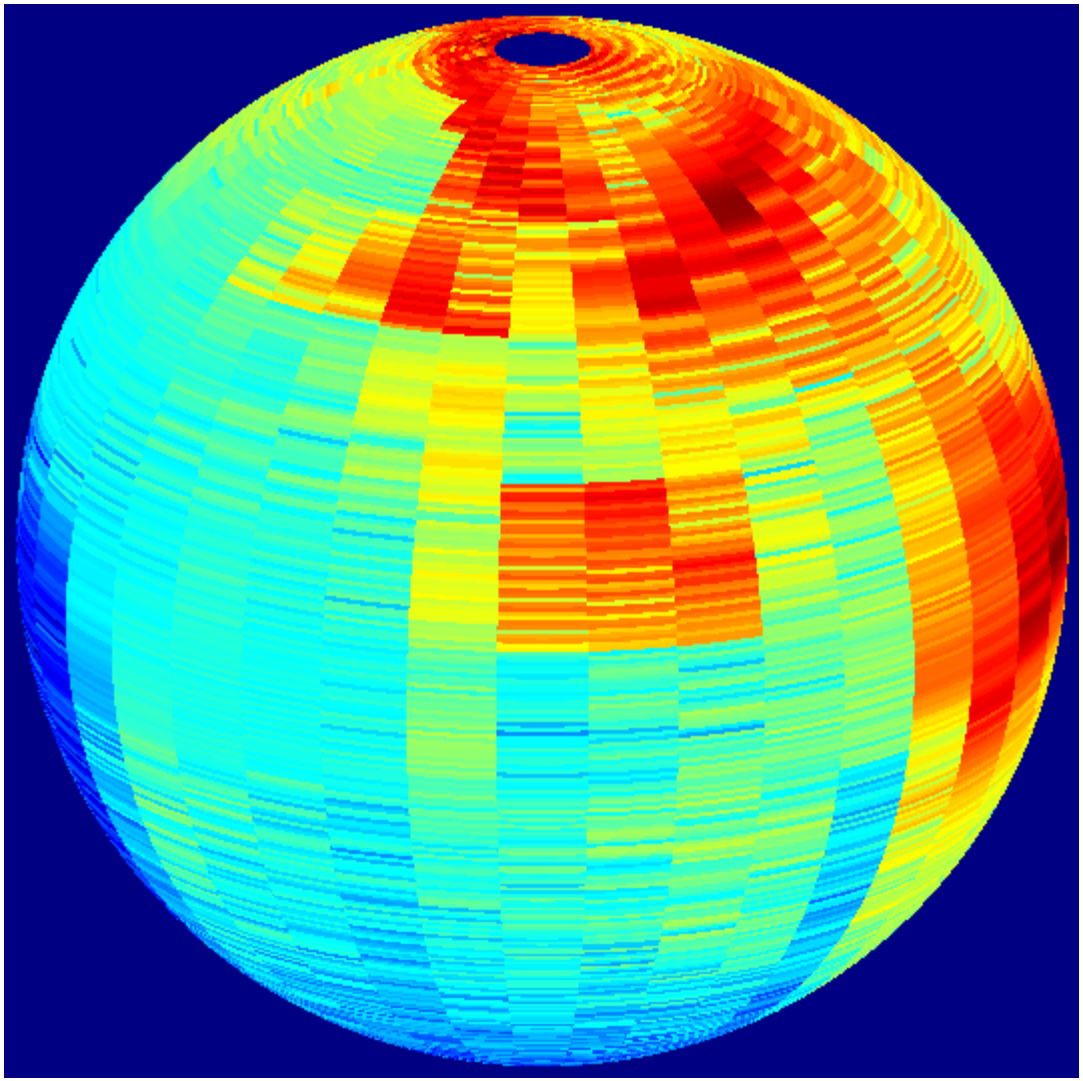
$$T(\phi_E,\phi) = \sum_{lm} Y_{lm}(\pi-\phi,\phi_E+\pi/2) b_{l0} s_{lm} \sqrt{2l+1}$$

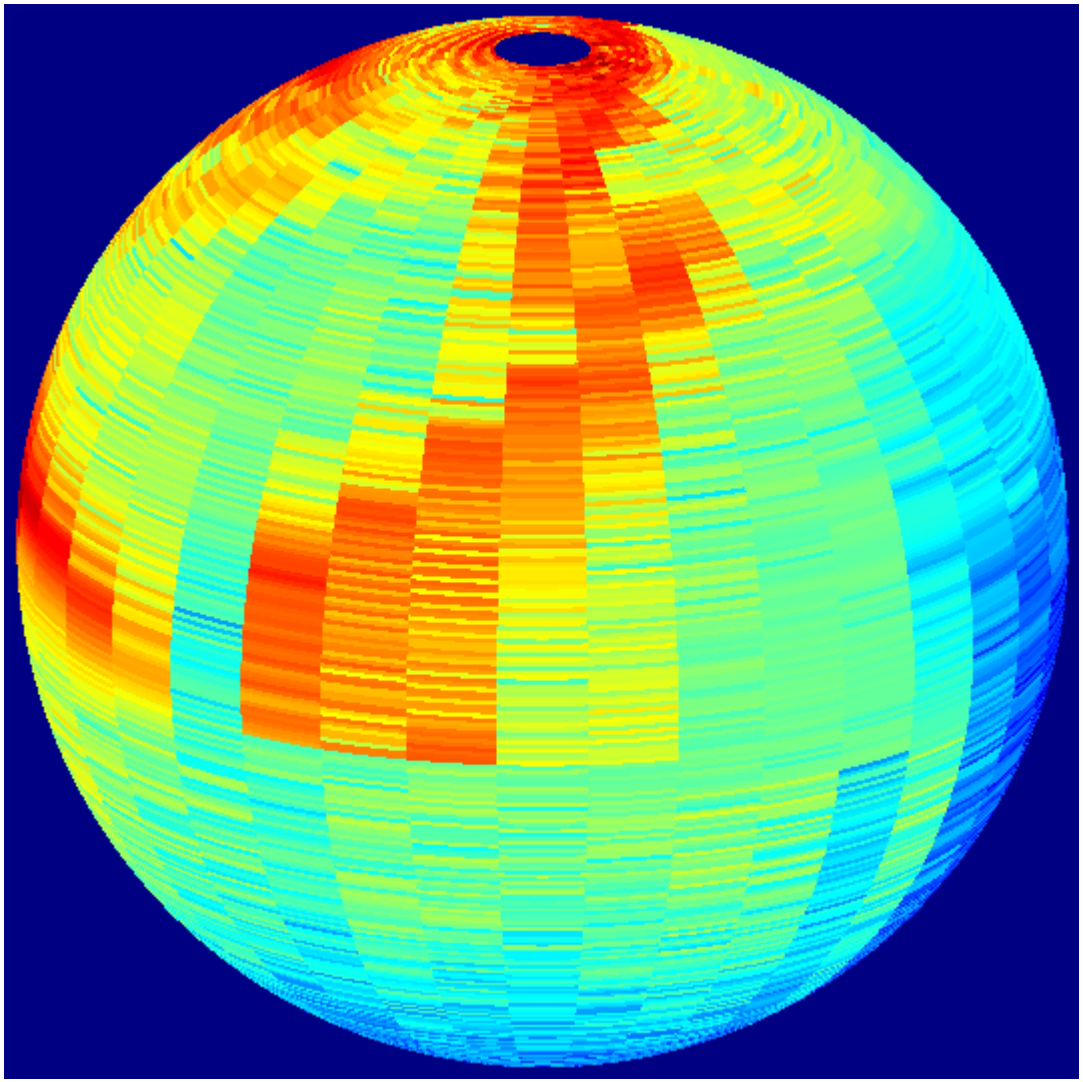
3) Low-order beam asymmetries M'<<L

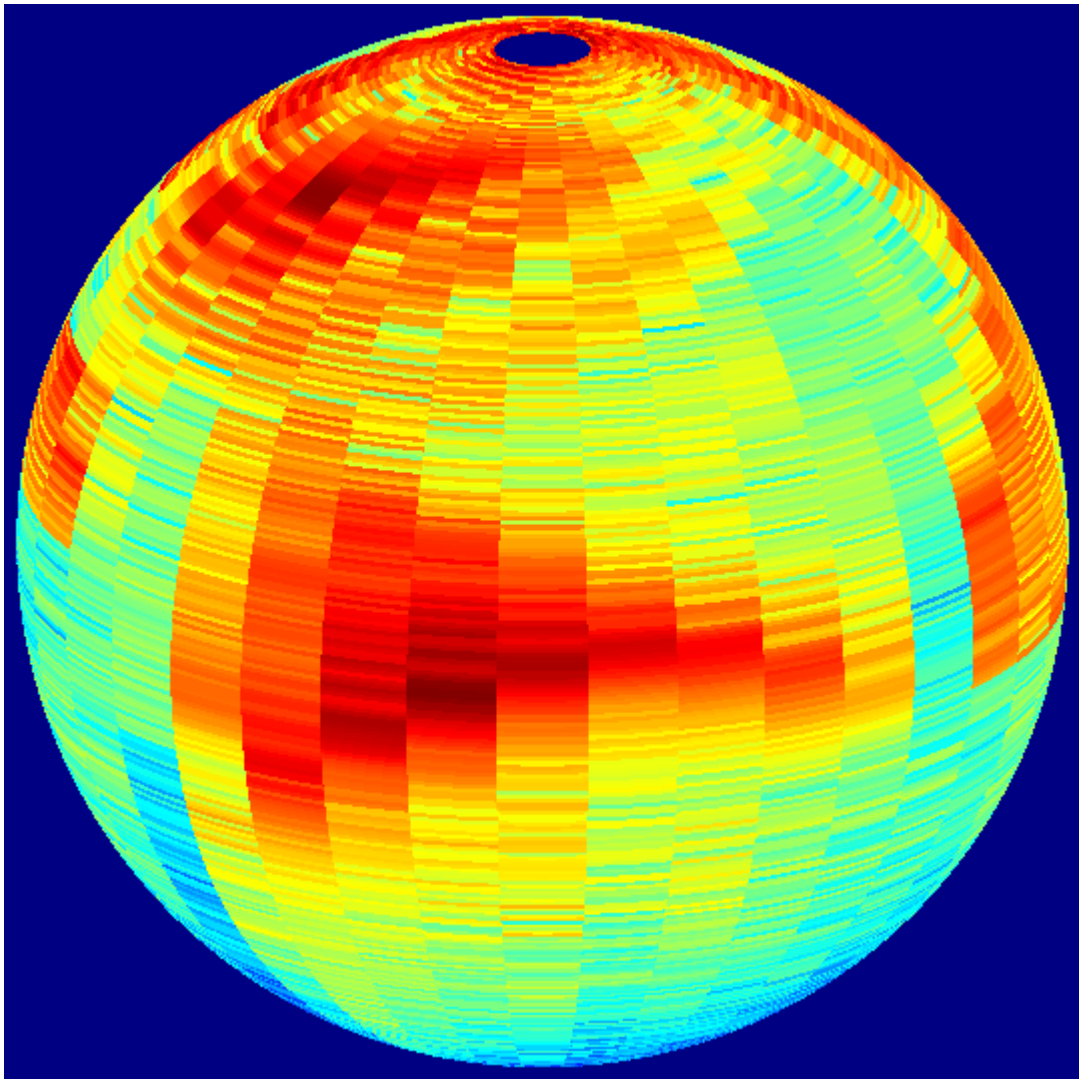
Wandelt & Gorski 2000, arXiv:astro-ph/0008227

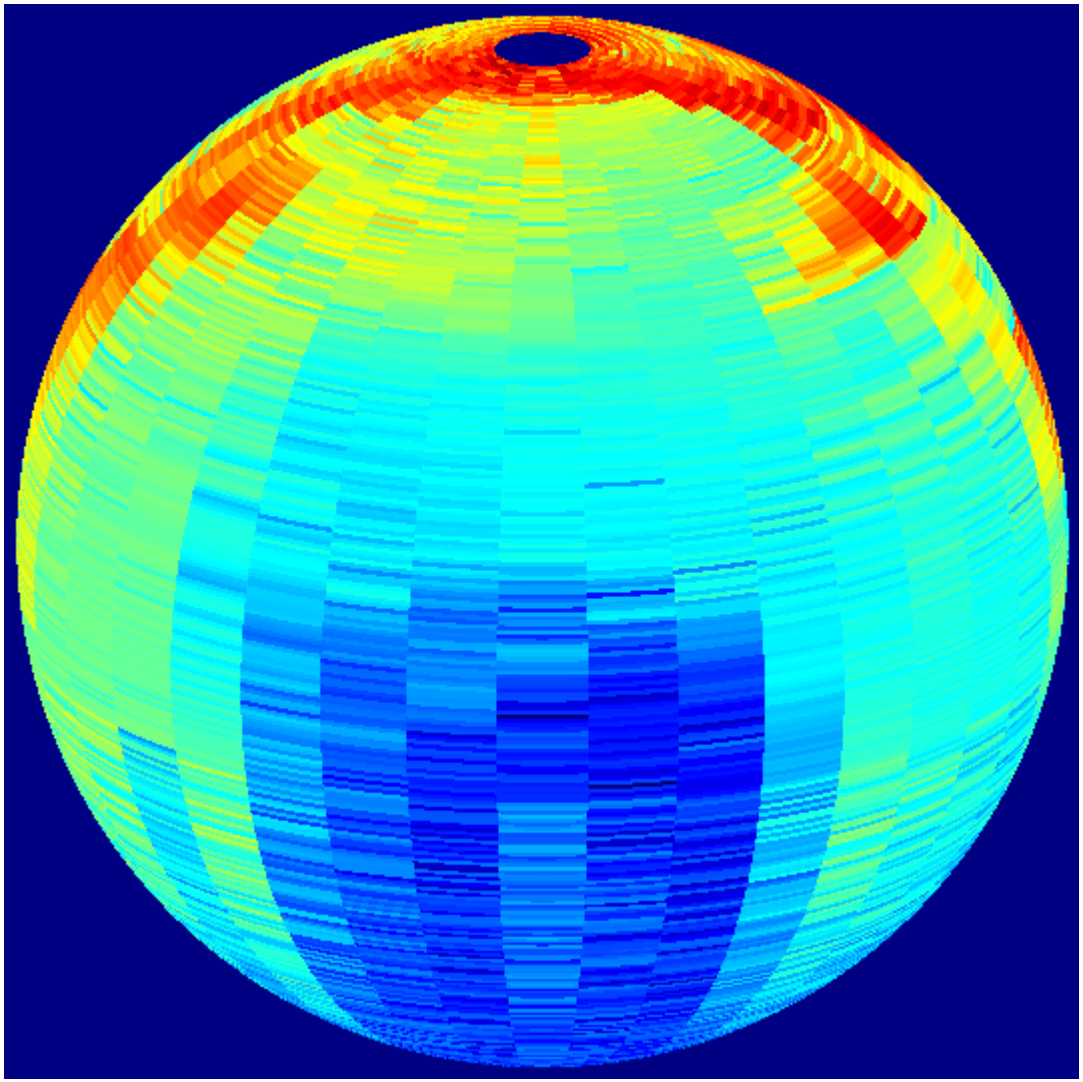
Example: Schlegel-Finkbeiner-Davis model of dust emission at 100GHz convolved with highly asymmetric physical beam model



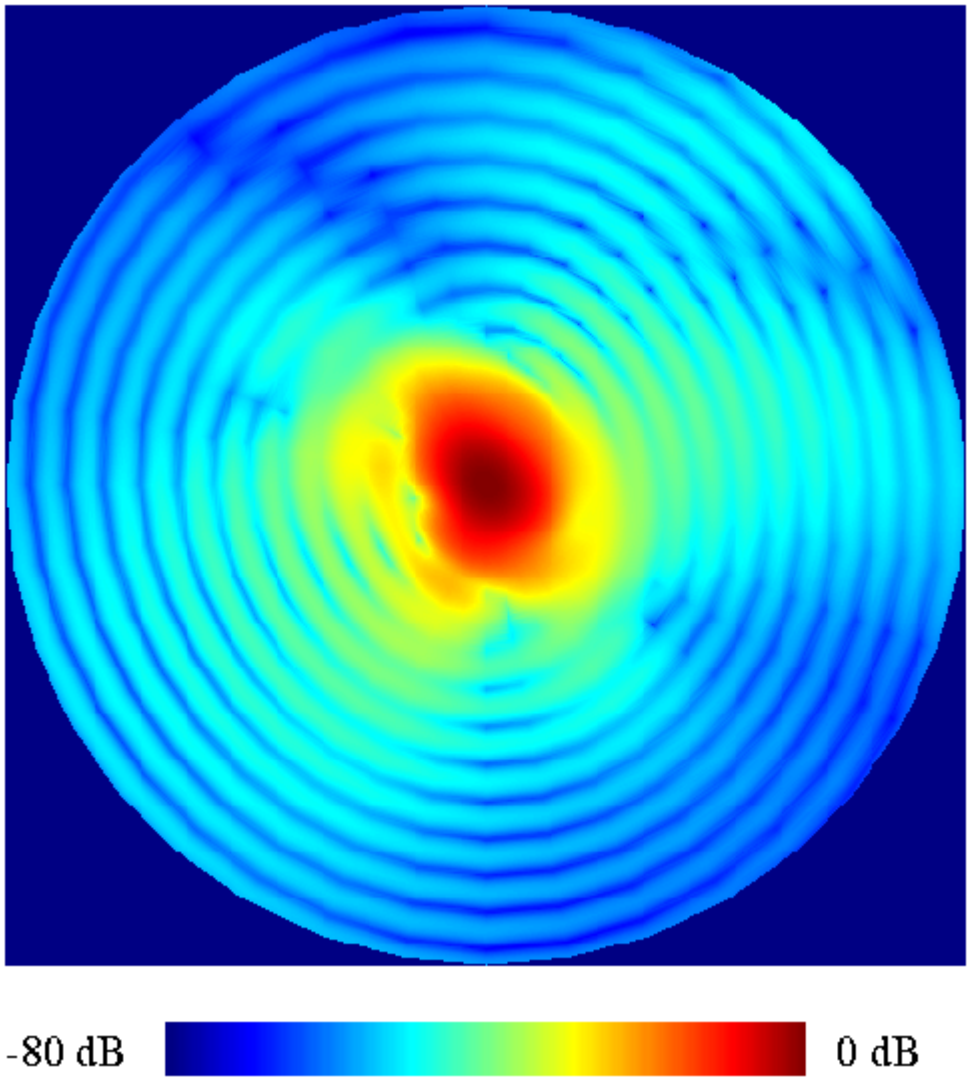




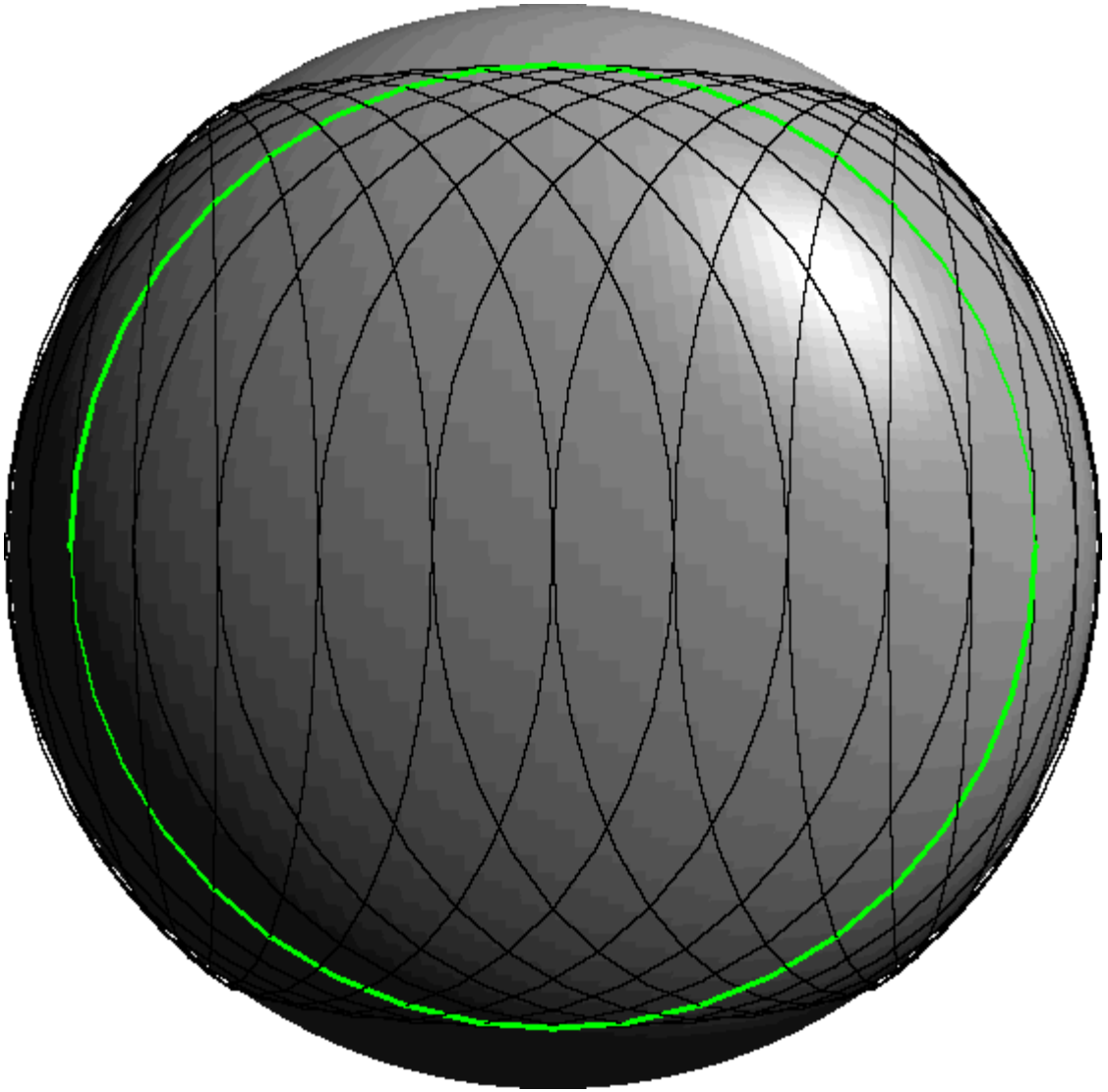


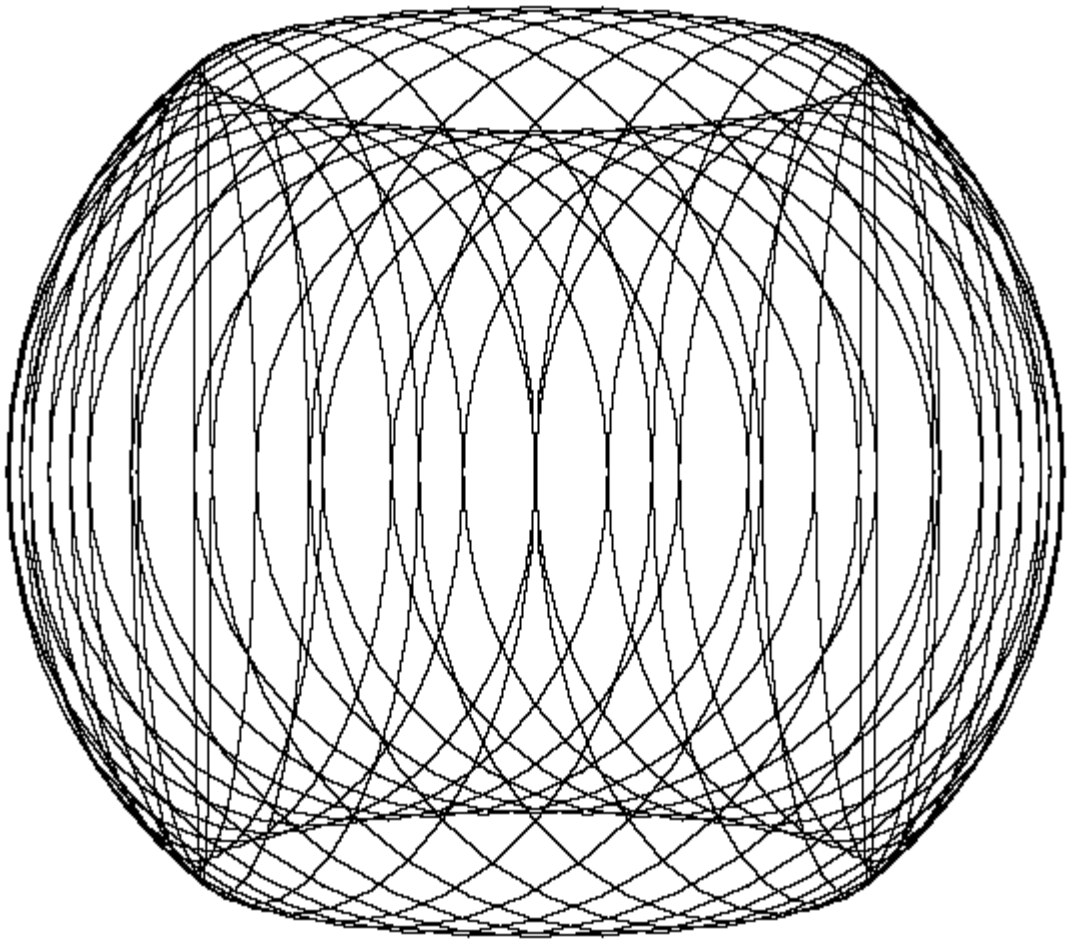


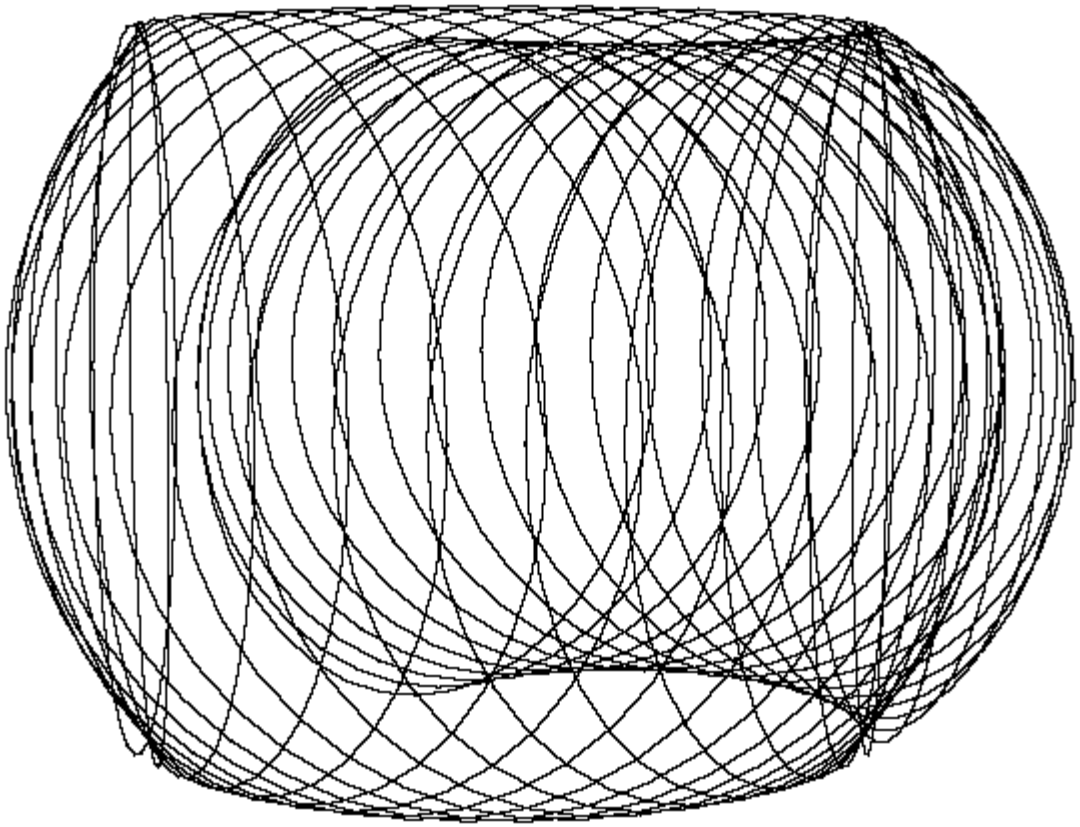
Central (main) beam

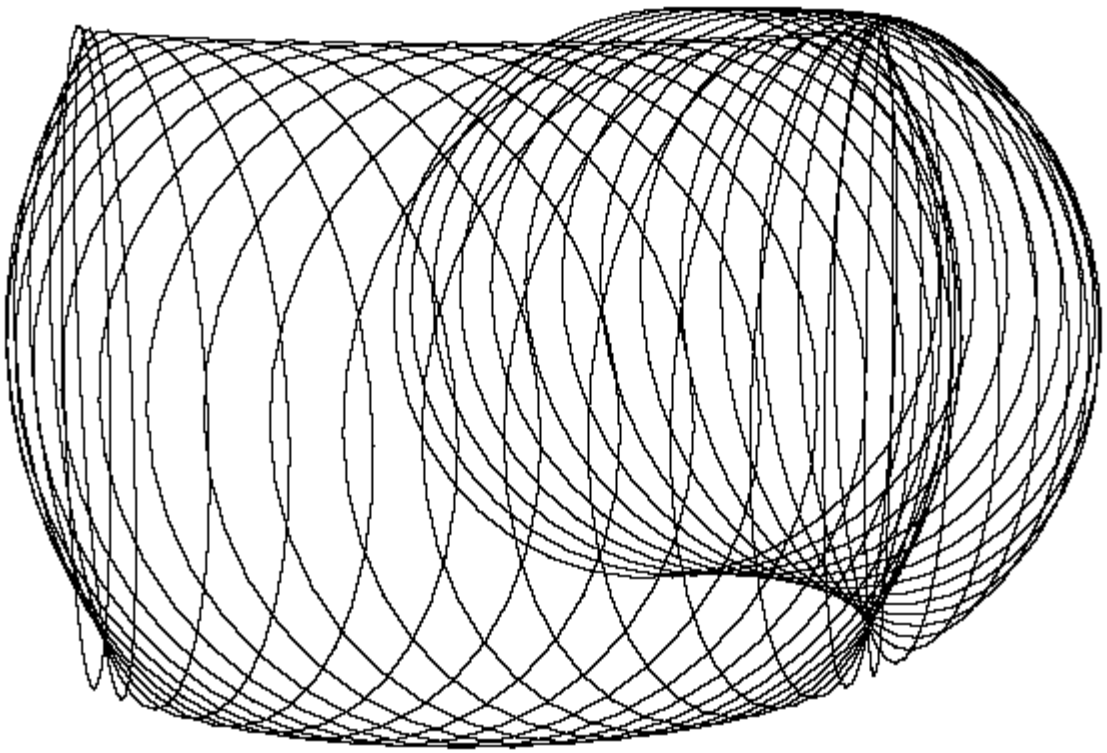


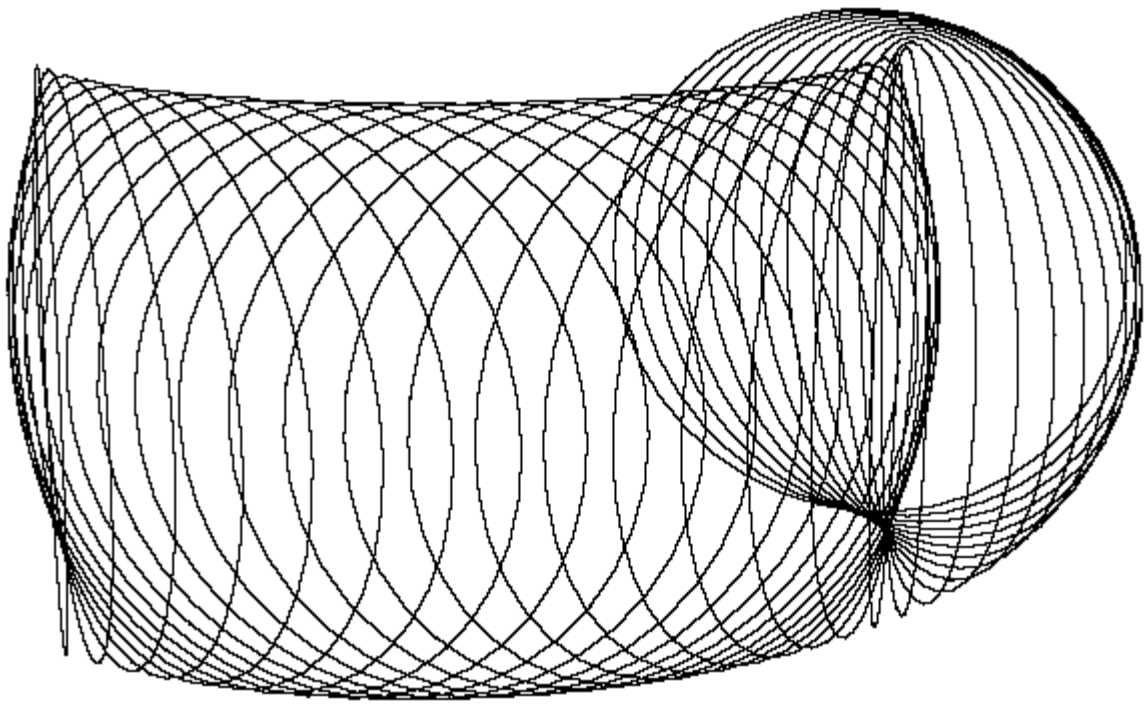
Basic Scan Path

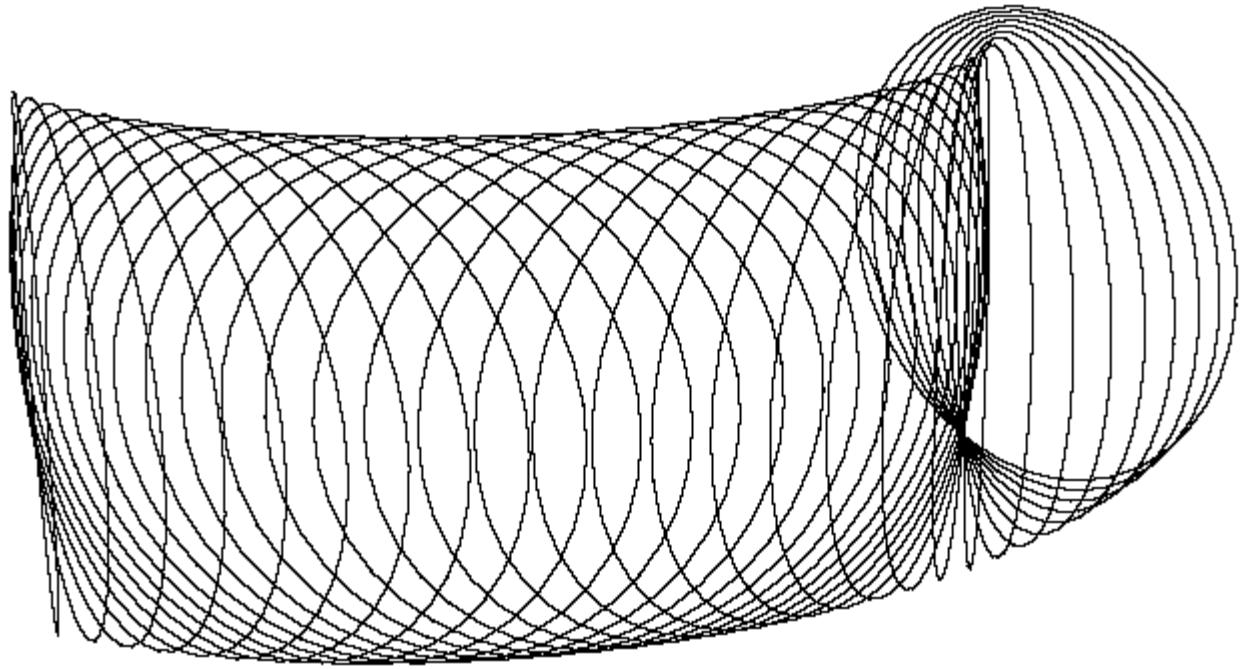


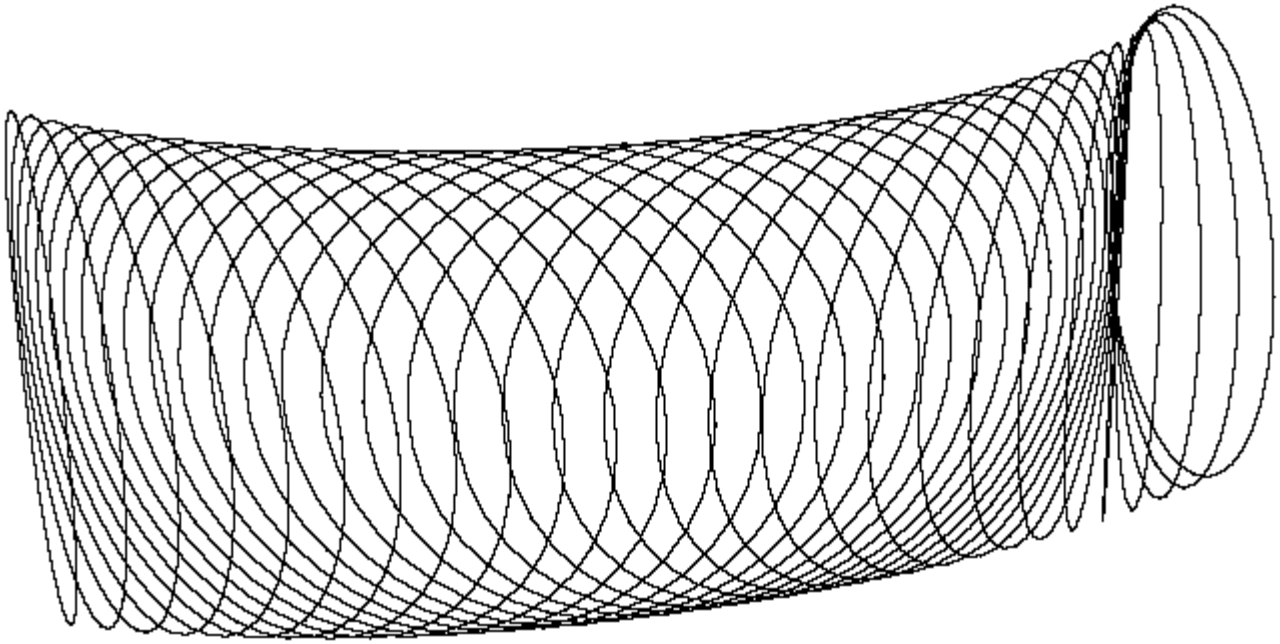


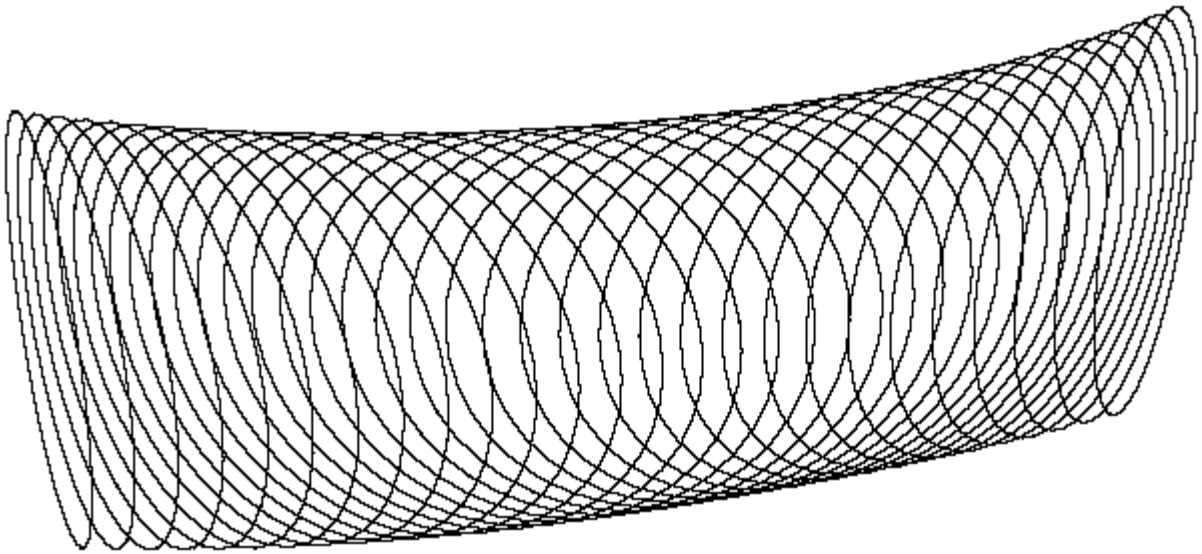


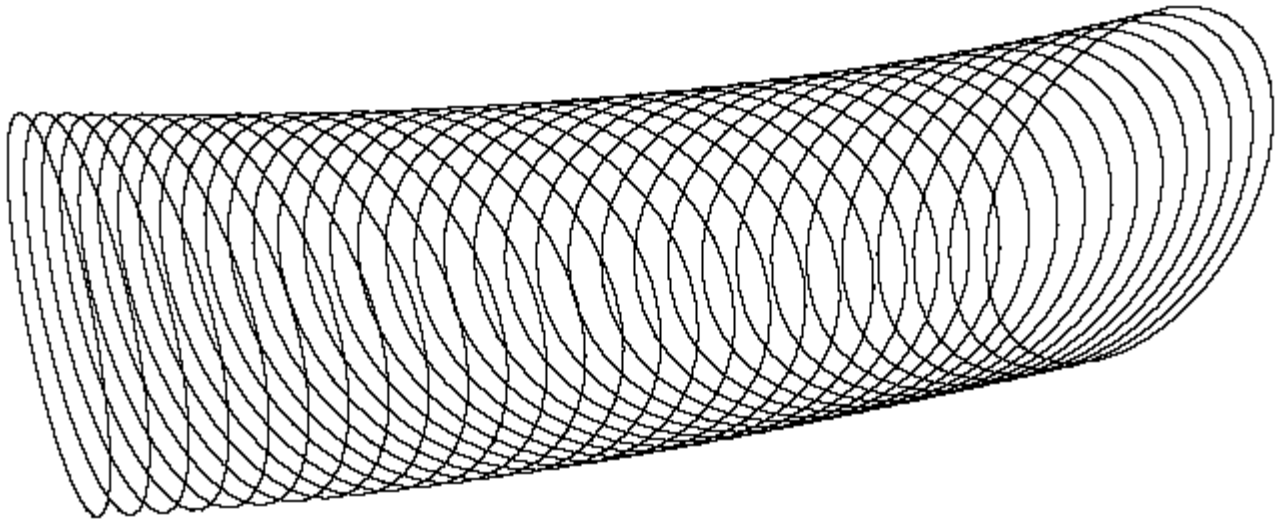


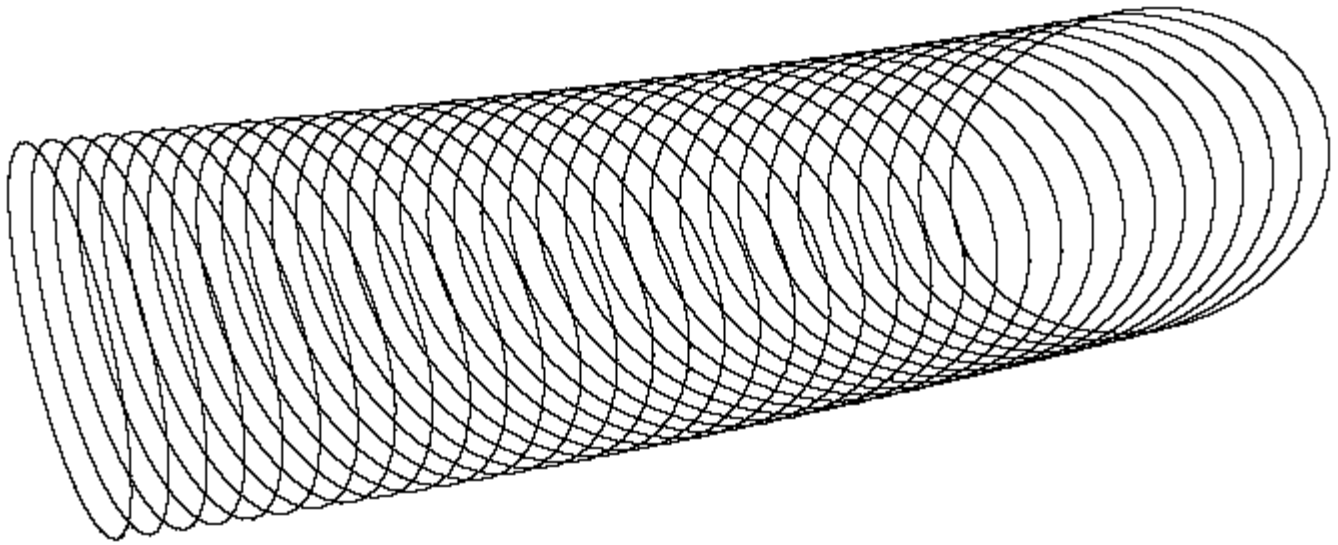


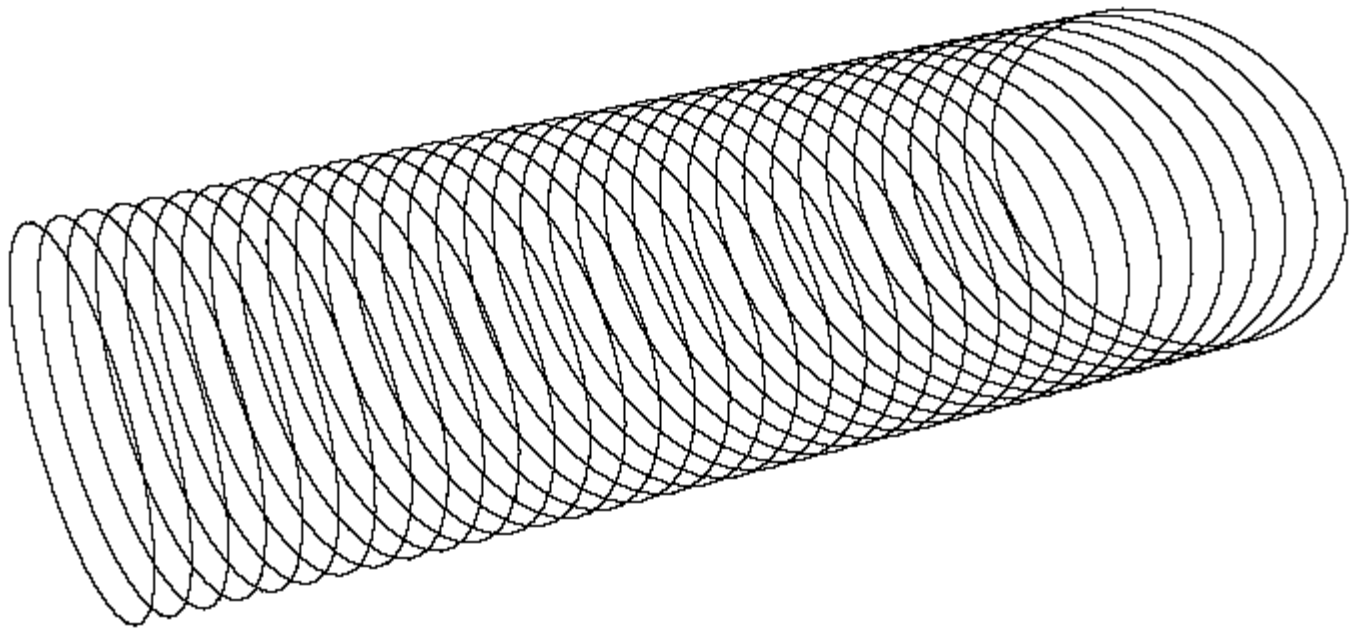


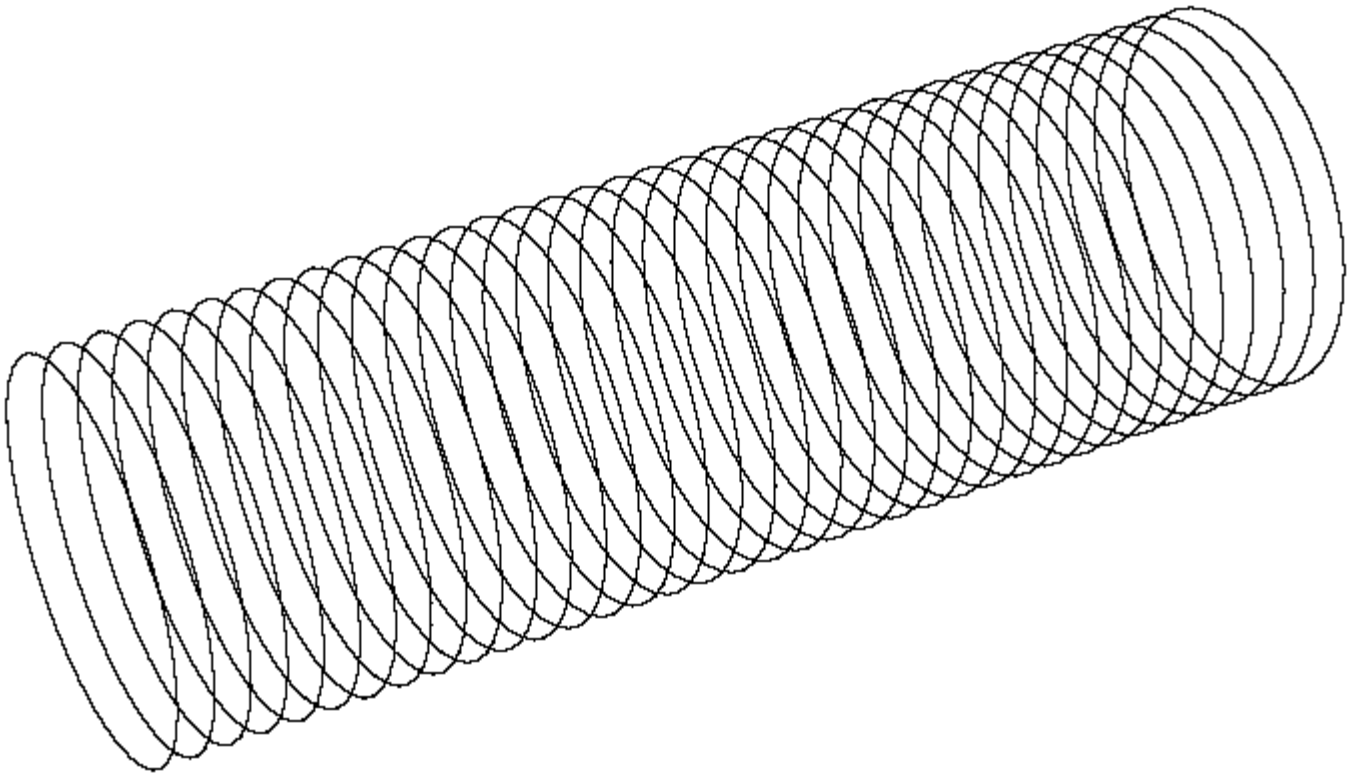


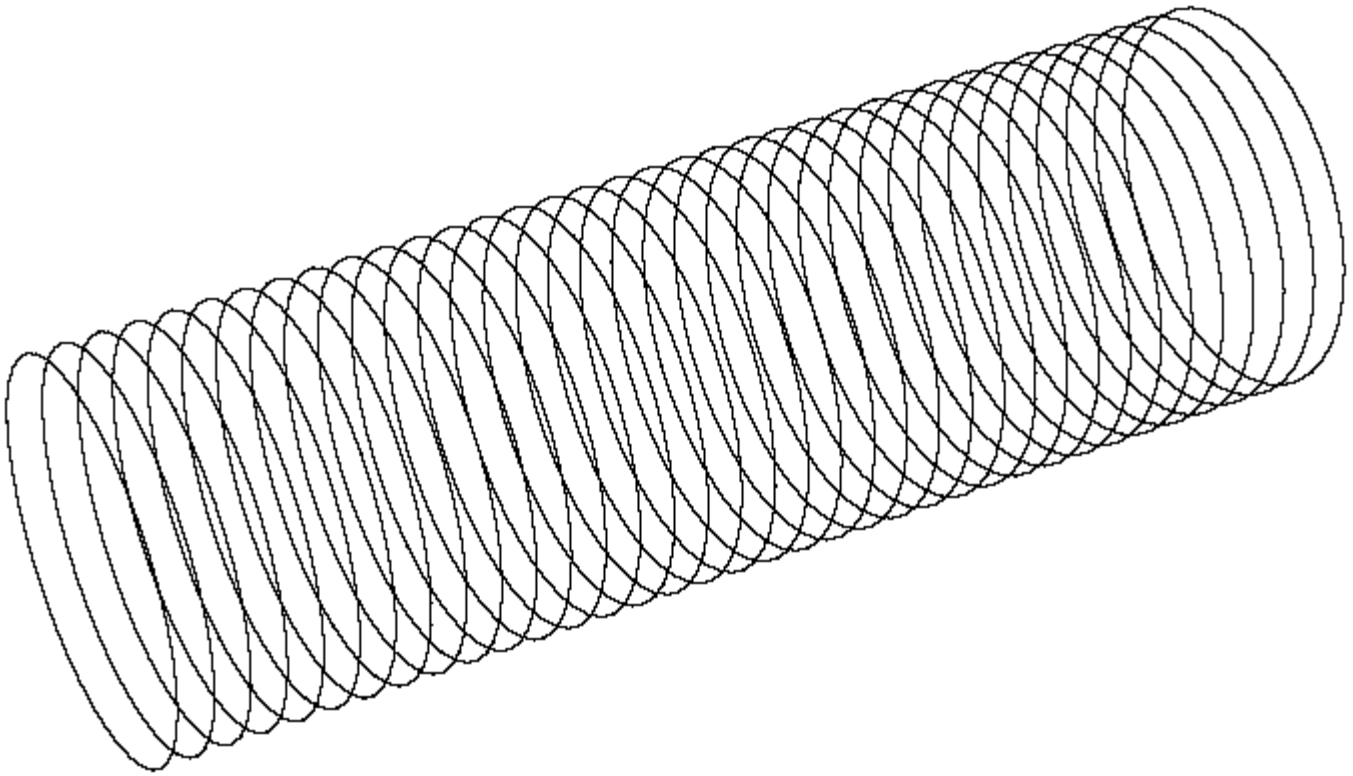


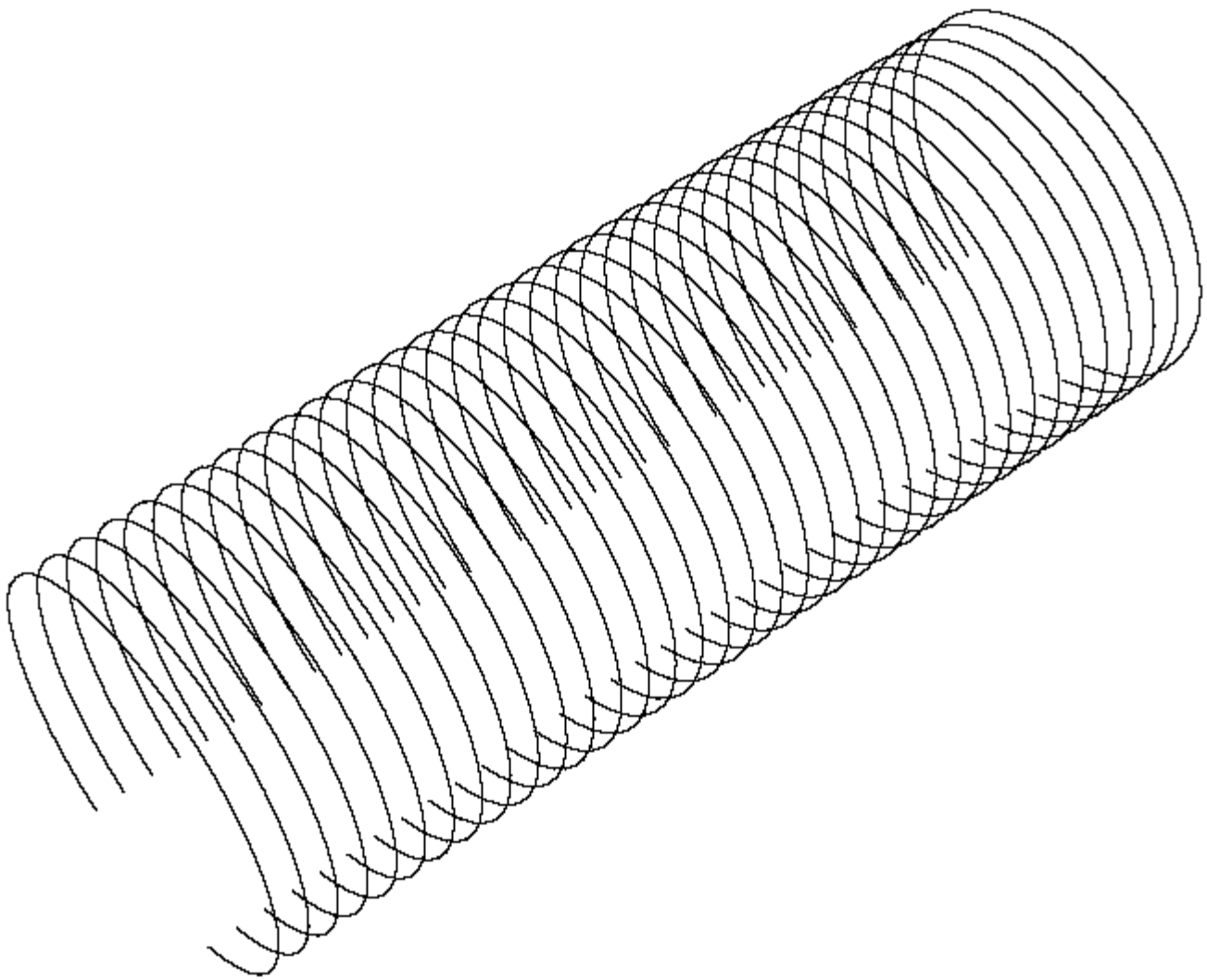


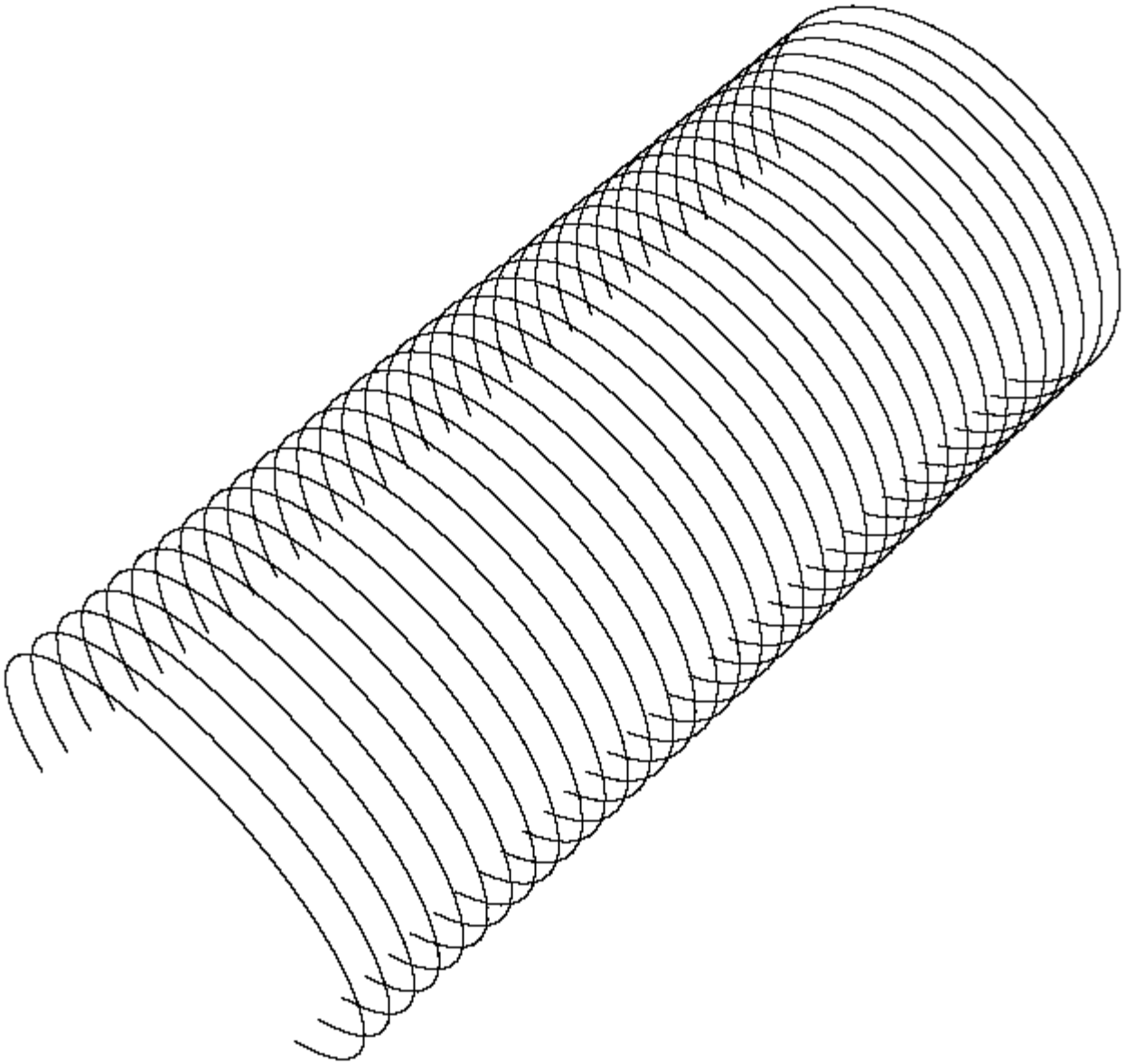


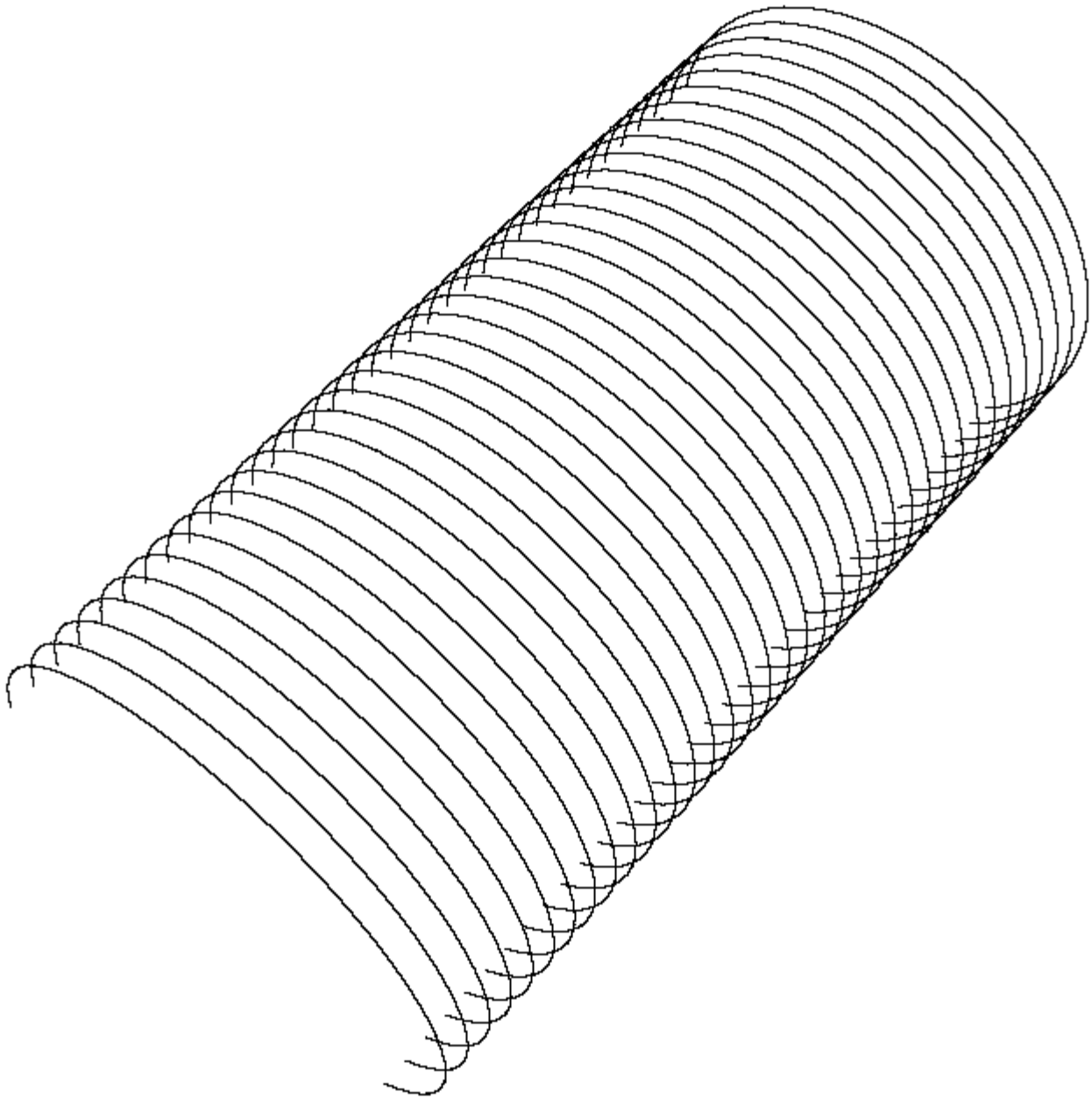


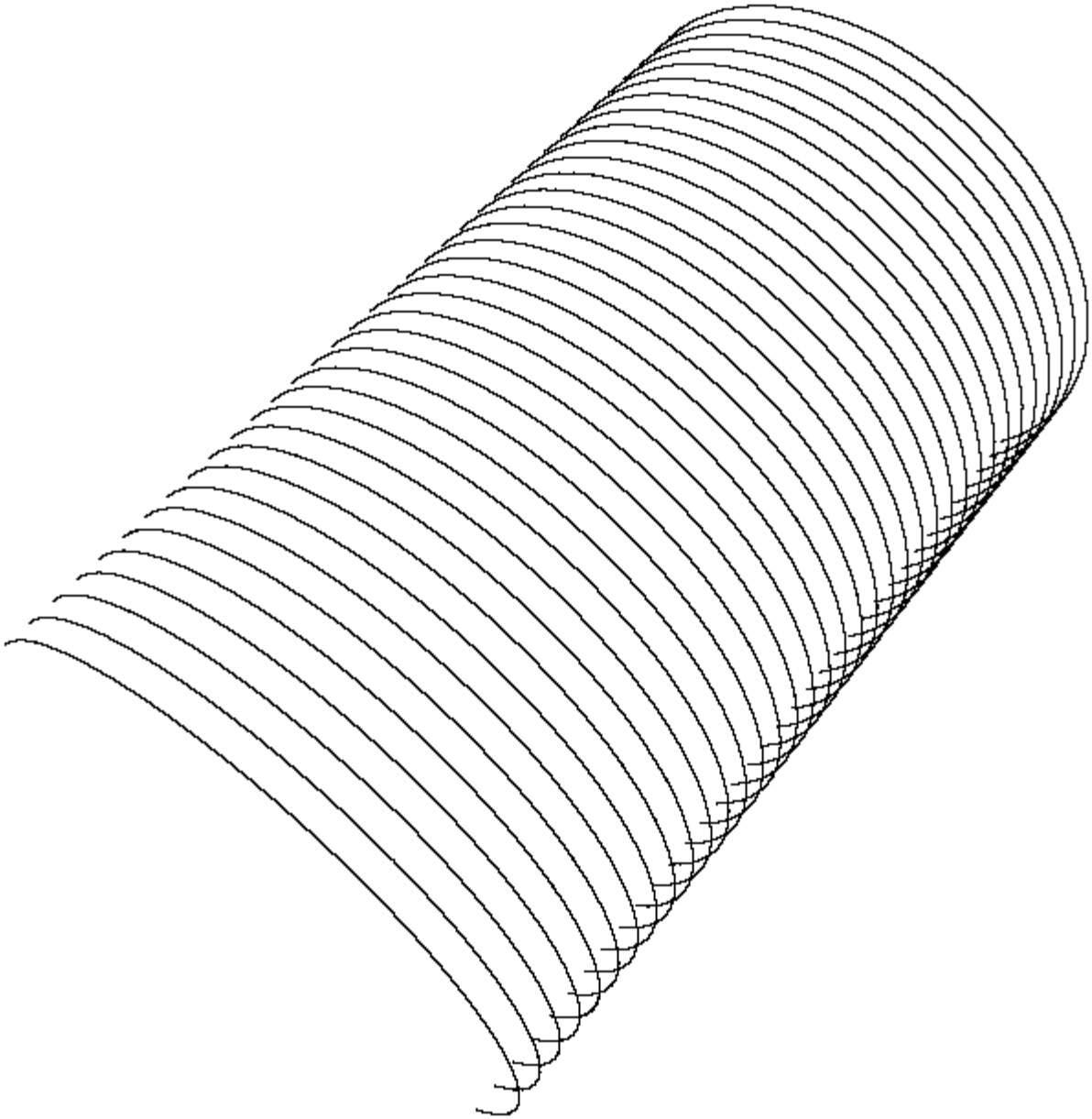


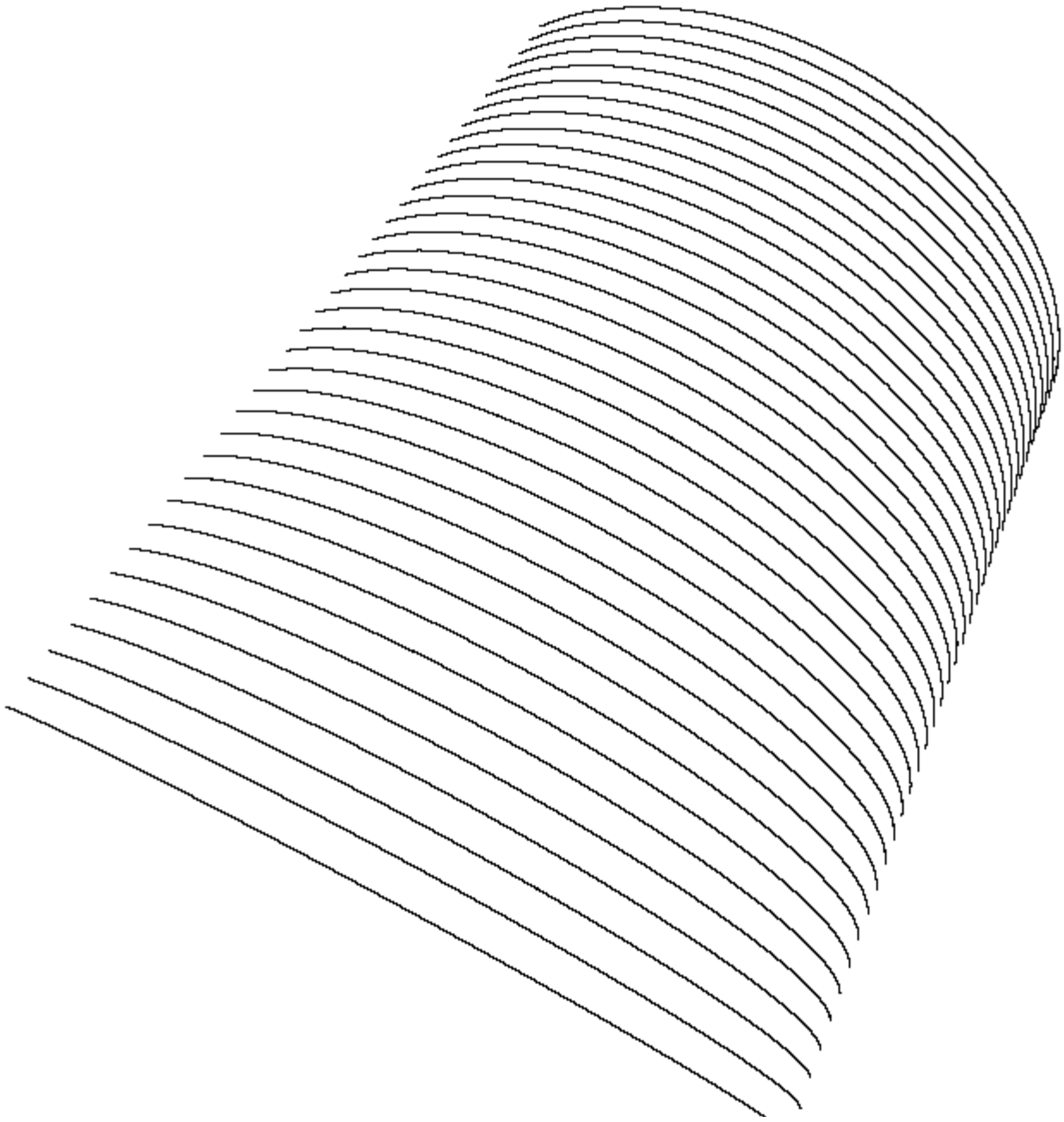


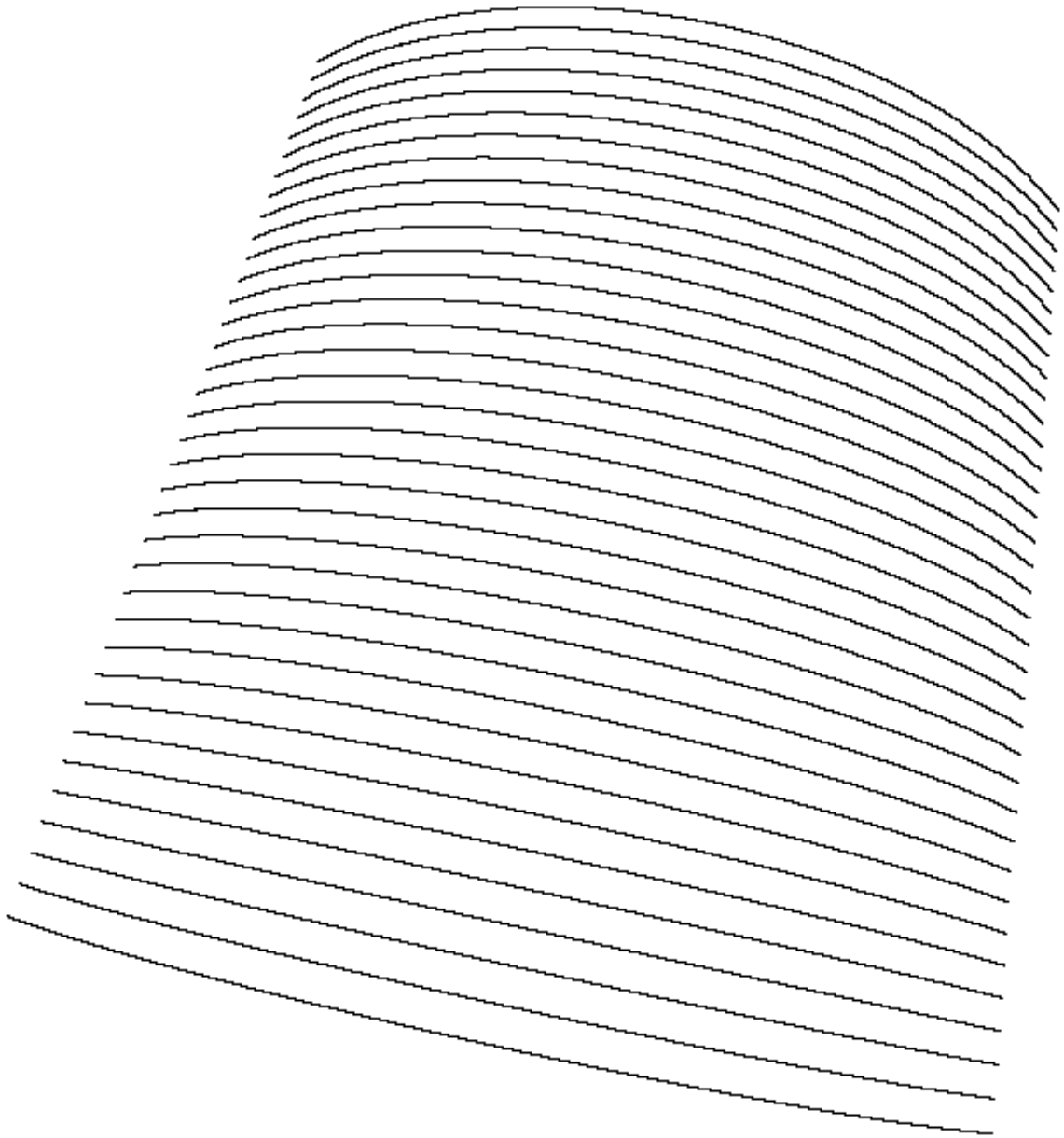


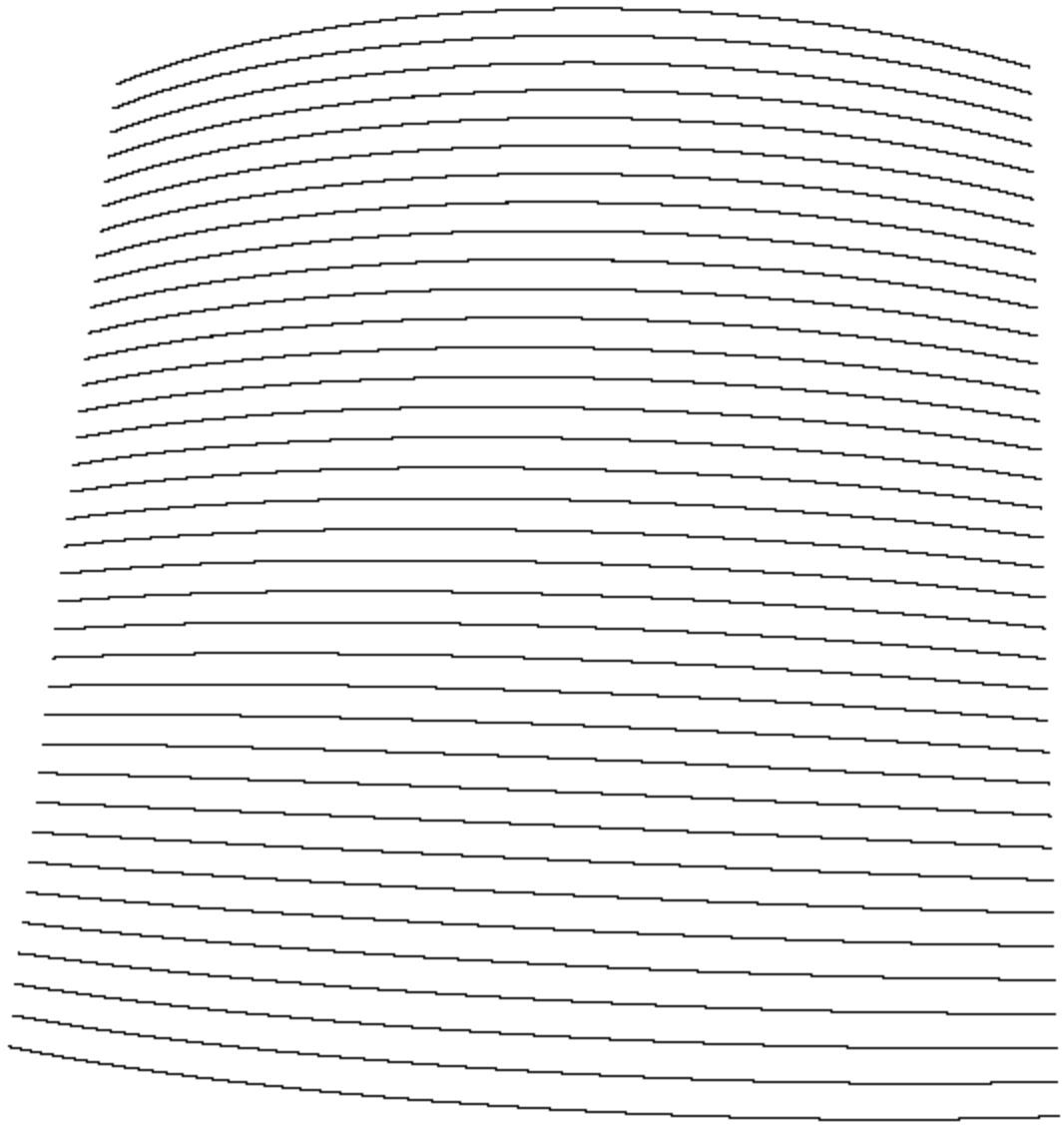


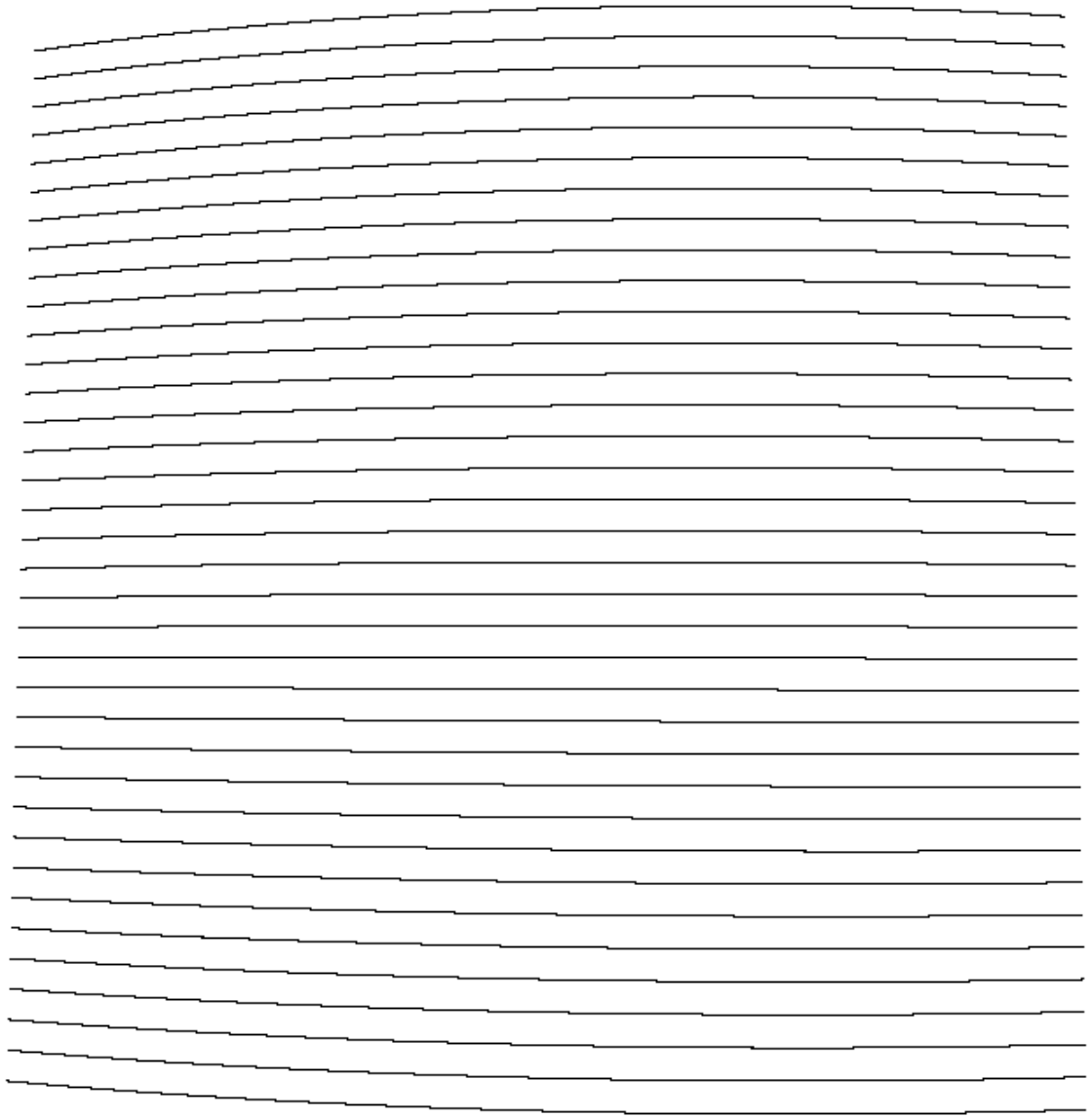


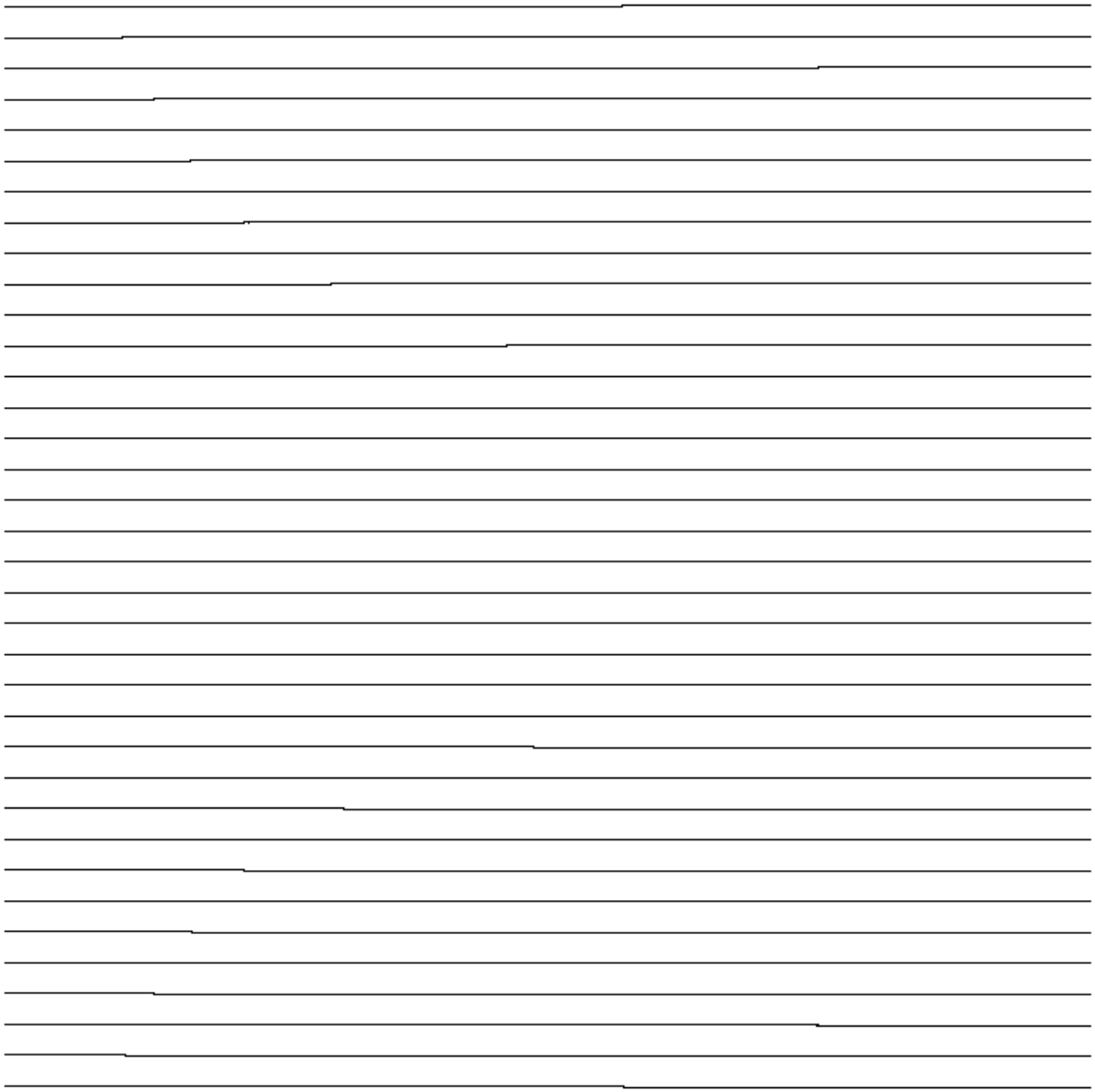




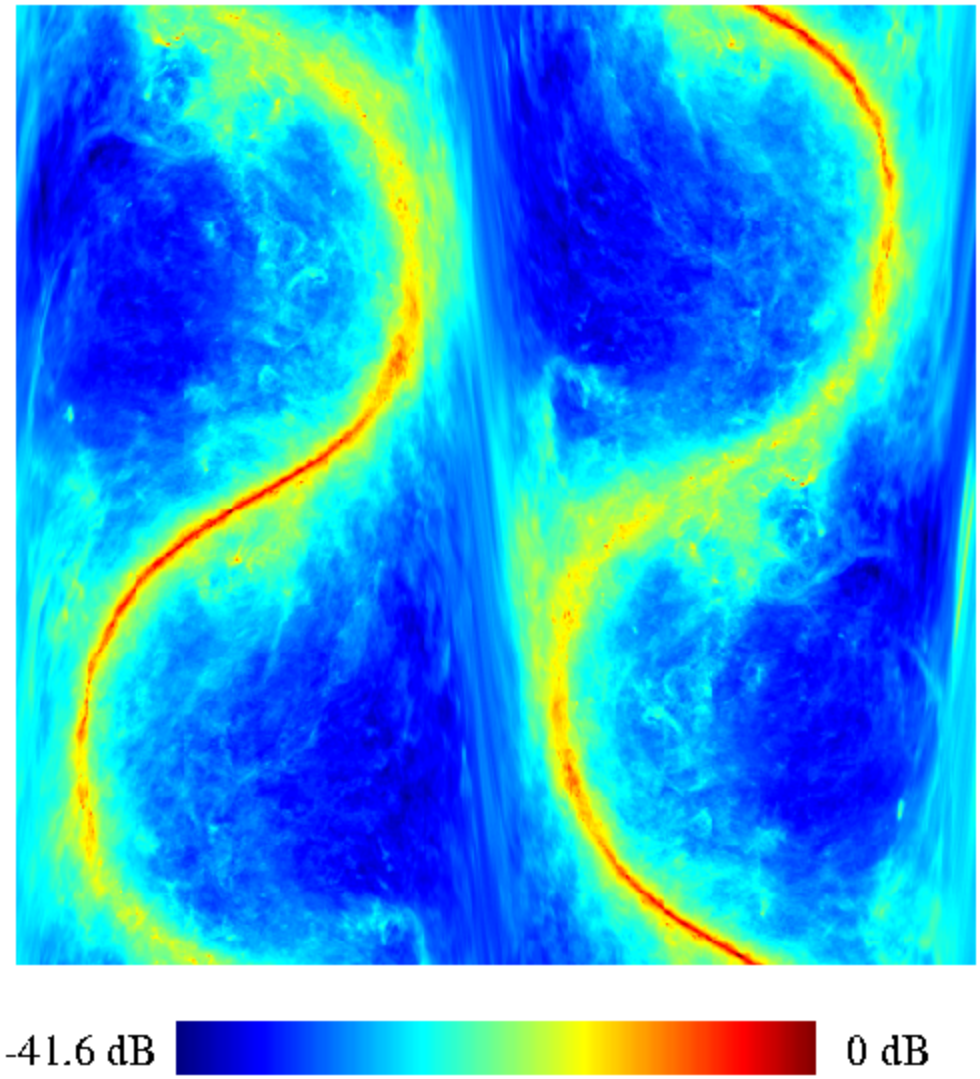




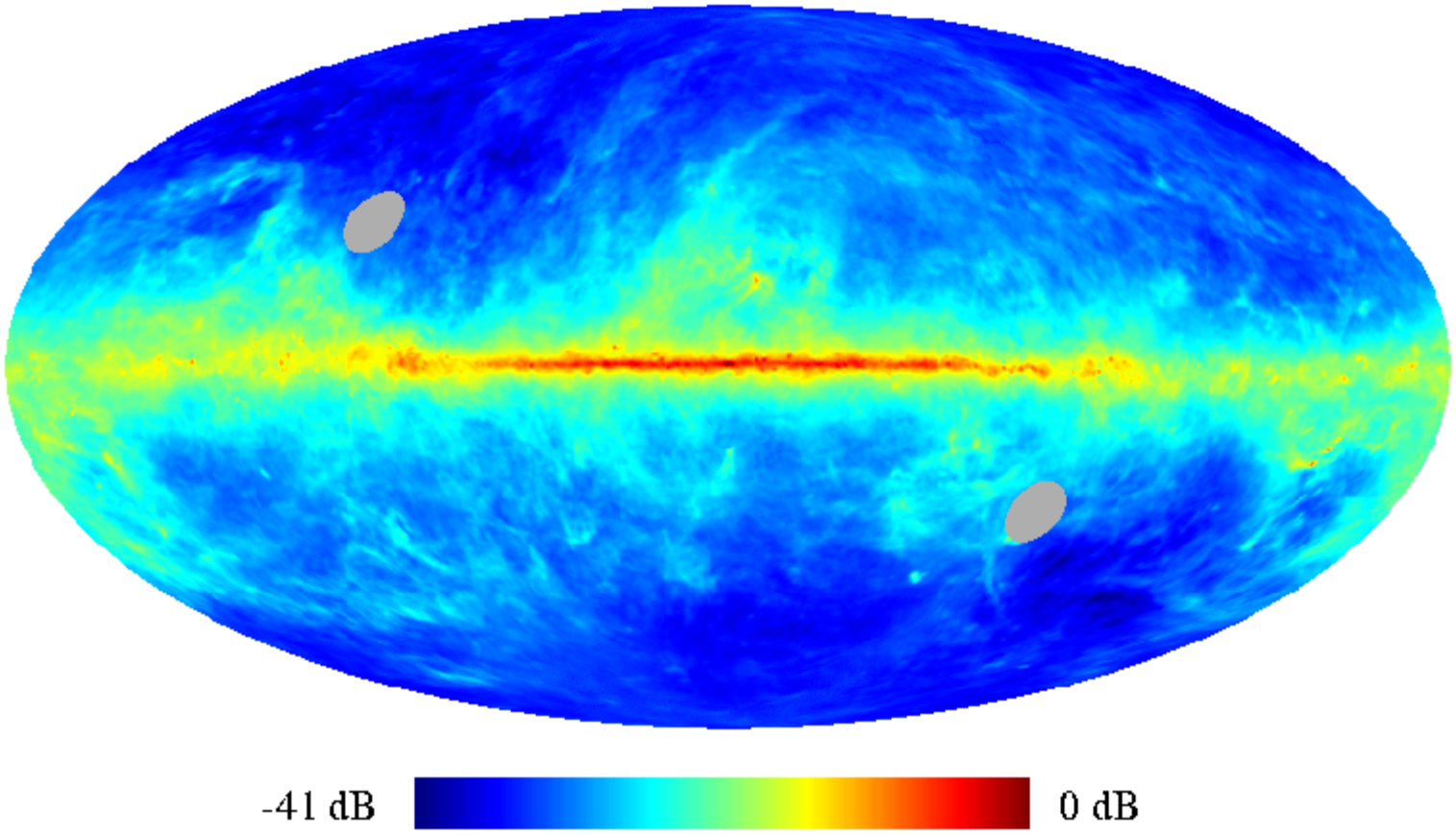




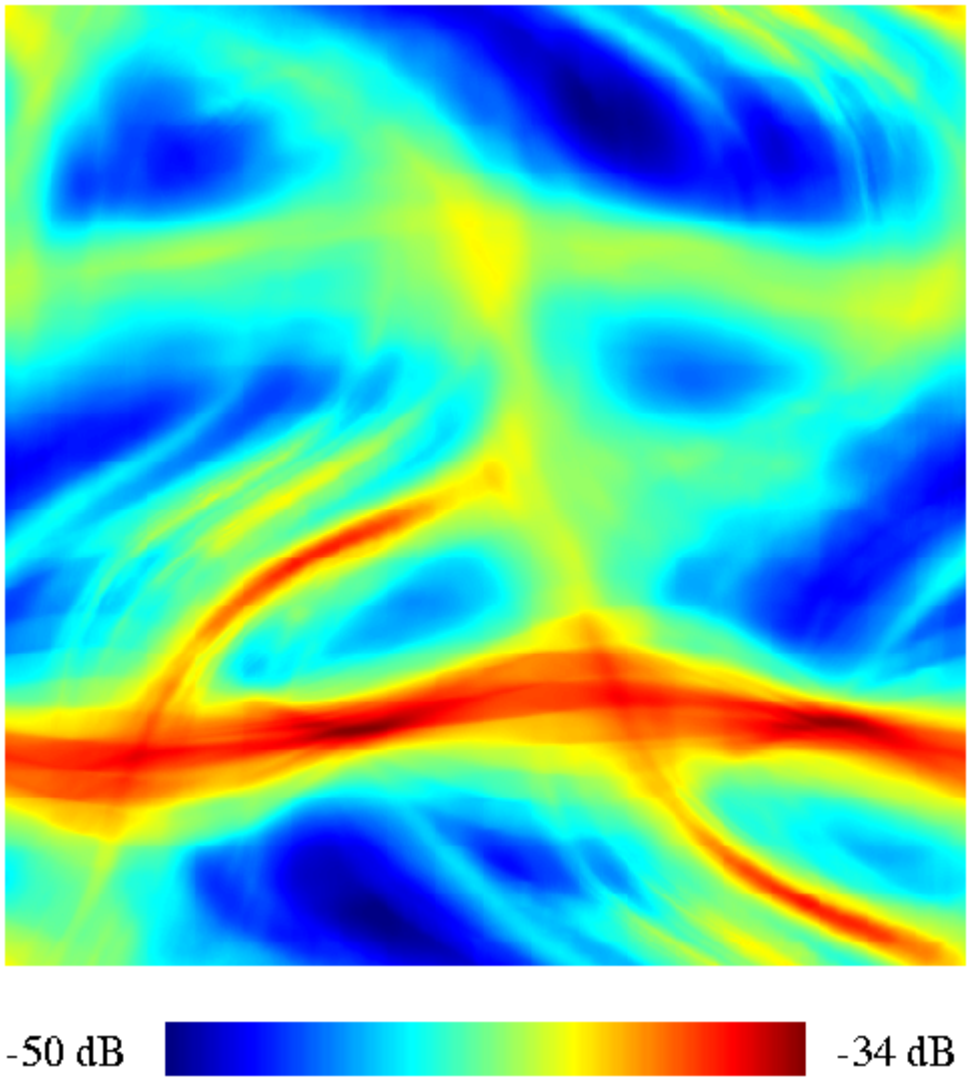
Main beam convolution



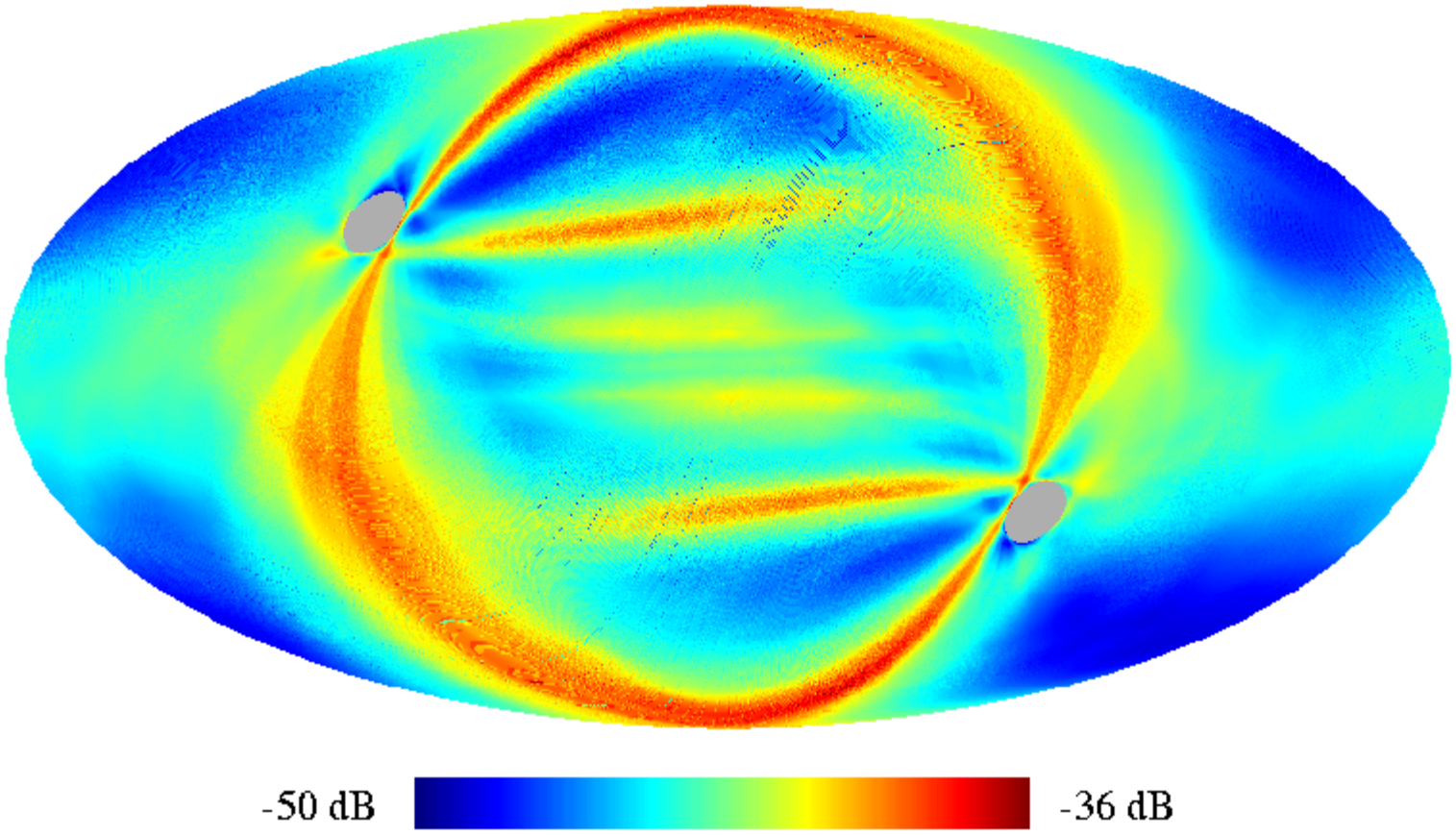
Main beam convolution



Far-side lobe convolution



Far-side lobe convolution



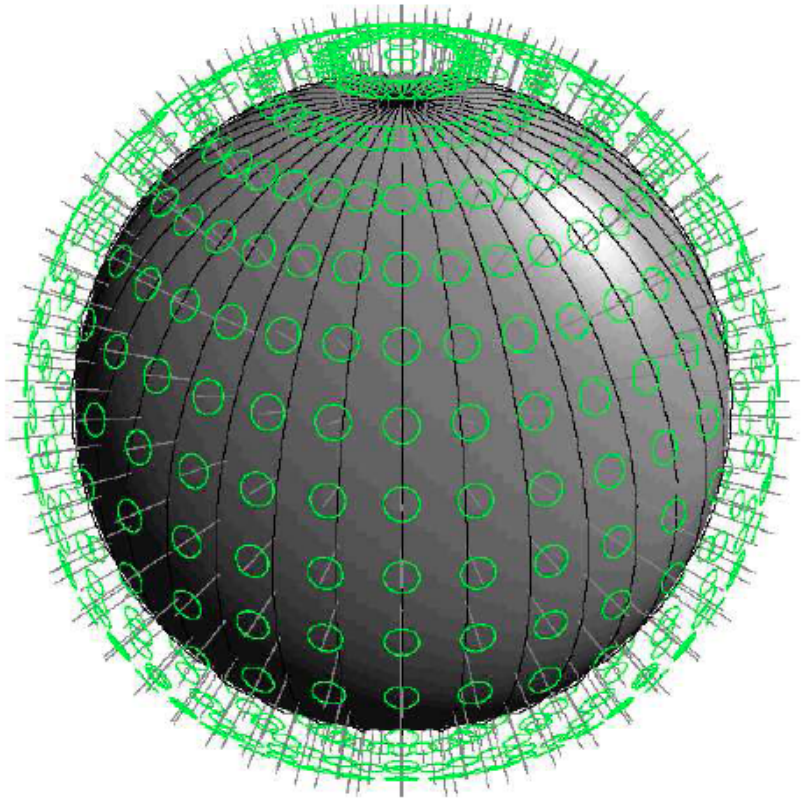
Polarization and higher spin just works^{TM}

- Everything generalizes to polarized skies and beams (and arbitrary tensor fields on the sphere).

A. Challinor, P. Fosalba, M. Ashdown,
B. D.Wandelt, K. Gorski 2000 astro-ph/0008228

Summary: SO(3) and the 3-torus

BW, astro-ph/0012416



$$T^S(\Phi_2,\Theta,\Phi_1) = \int d\Omega_\gamma \left[\widehat{D}(\Phi_2,\Theta,\Phi_1)b\right](\gamma)^* s(\gamma)$$

- For operations involving SO(3) represent in terms of mode functions on the 3-torus
- This gives
  - A fast general convolution algorithm for arbitrary functions (Wandelt & Gorski 2000, arXiv:astro-ph/0008227)
  - An exact fast transform and sampling theorem on the sphere for spin-s spherical harmonics (Huffenberger & BW 2010 ; improved by a factor of 2 by J. McEwen & Wiaux 2011)

Other convolution fun

- GPU accelerated convolution for compact kernels
- Hybrid GPU+CPU algorithm for non-compact convolution kernels
- Compressed convolution

Standard spherical convolution algorithm

$$s(\mathbf{n}_1) = \int_{S^2} K(\mathbf{n}_1, \mathbf{n}_2) r(\mathbf{n}_2) d^2\mathbf{n}_2$$ Kernel convolution

For *radial* kernels $K(\mathbf{p}, \mathbf{q}) = K(p.q)$ $p.q \equiv \mathbf{n}_p.\mathbf{n}_q$

$$K(p.q) = \sum_\ell \frac{2\ell+1}{4\pi} K_\ell P_\ell(p.q)$$ Legendre decomposition

$$K_\ell = 2\pi \int_{-1}^{1} K(z) P_\ell(z) dz$$

$$\boxed{s_{\ell m} = K_\ell \, r_{\ell m}}$$

Scales as $O(L^3)$, dominated by spherical harmonic transforms

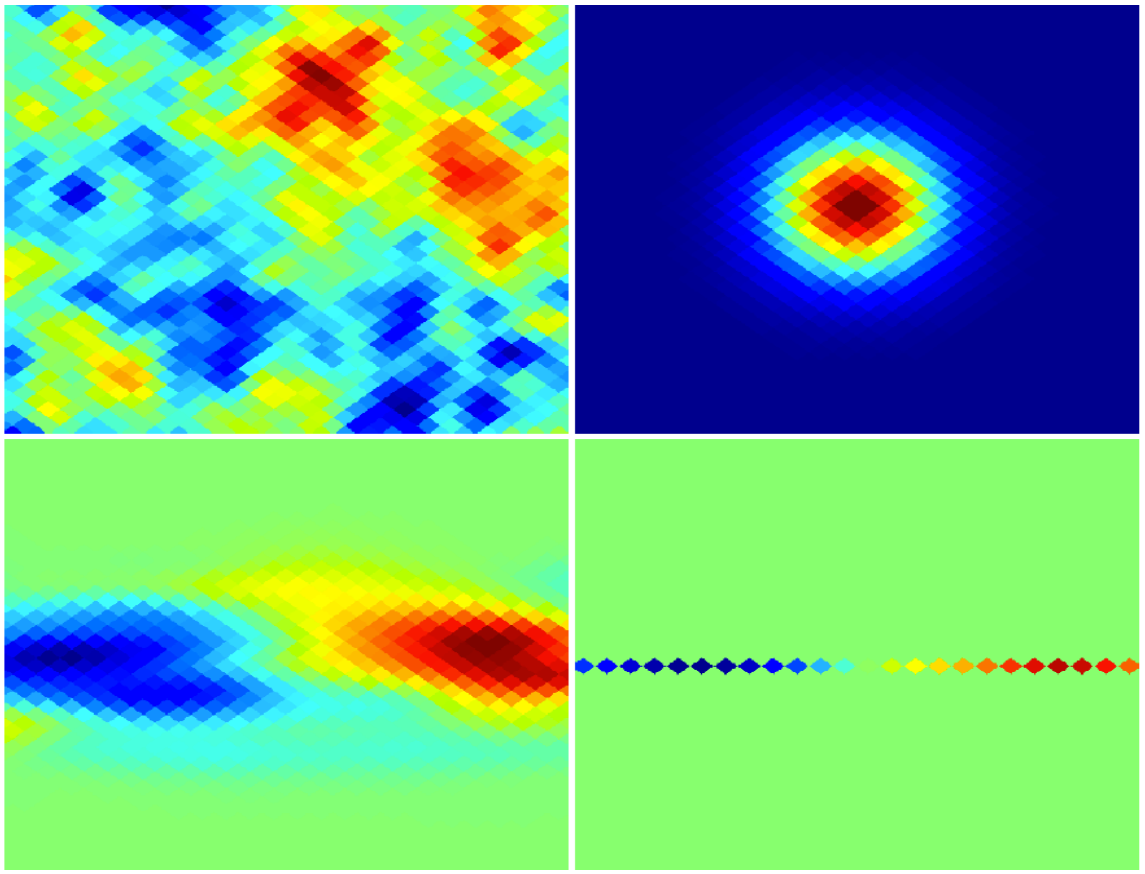
Why move beyond FSHT convolution for radial kernels?

- Spherical Harmonic Transforms are hard to parallelize
- For high resolution problems (L_{max}>1024) the recursions for the spherical harmonics $Y_{lm}(n)$ need to be computed in extended double precision to avoid underflows.
- Too general for many problems – does not exploit kernel compactness

Hybrid convolution algorithm idea: avoid spherical harmonic transform

Generate convolved map ring by ring using batch 1-D FFTs (CUDAFFT)

The figure shows the generation of one output ring.



Hybrid method allows exploiting kernel compactness

Scaling to convolve entire map is

O(kernelwidth npix log(npix))

ARKCOS: Speed-up compared to SHT convolution

Left axis:

HEALPix SHT implementation

Right axis:

`libsht` implementation (highly optimized)

Arkcos:

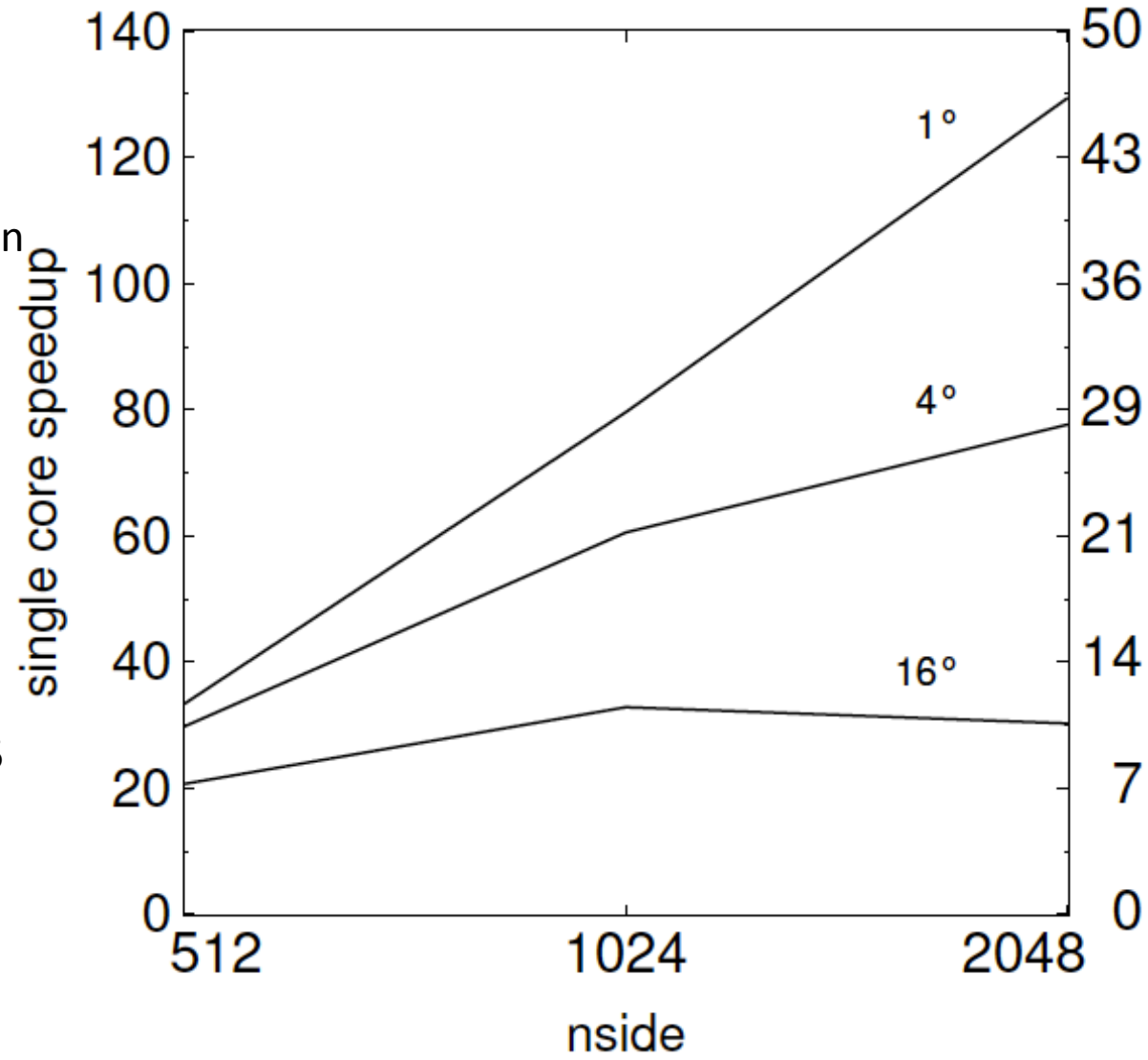
Elsner & Wandelt 2011,

Astr. & Astrophys. 532, A 35

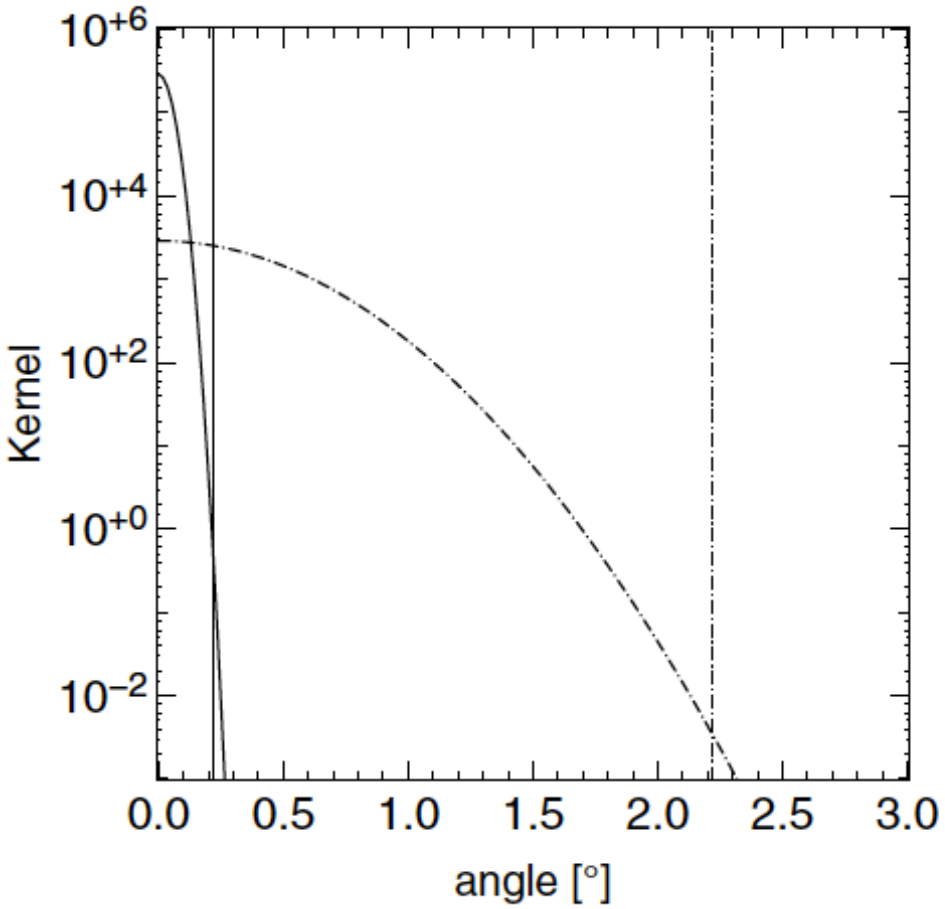
(arXiv:1104.0672)

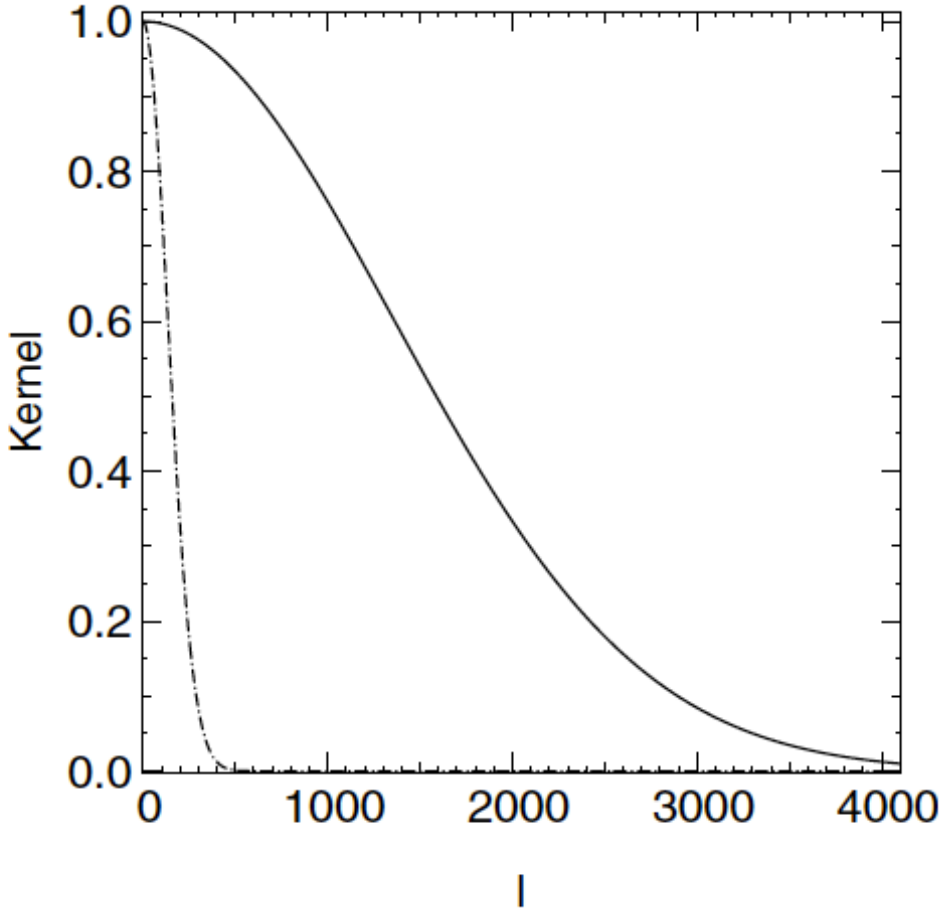
Example at nside 2048,

5 x 10^7 pixels

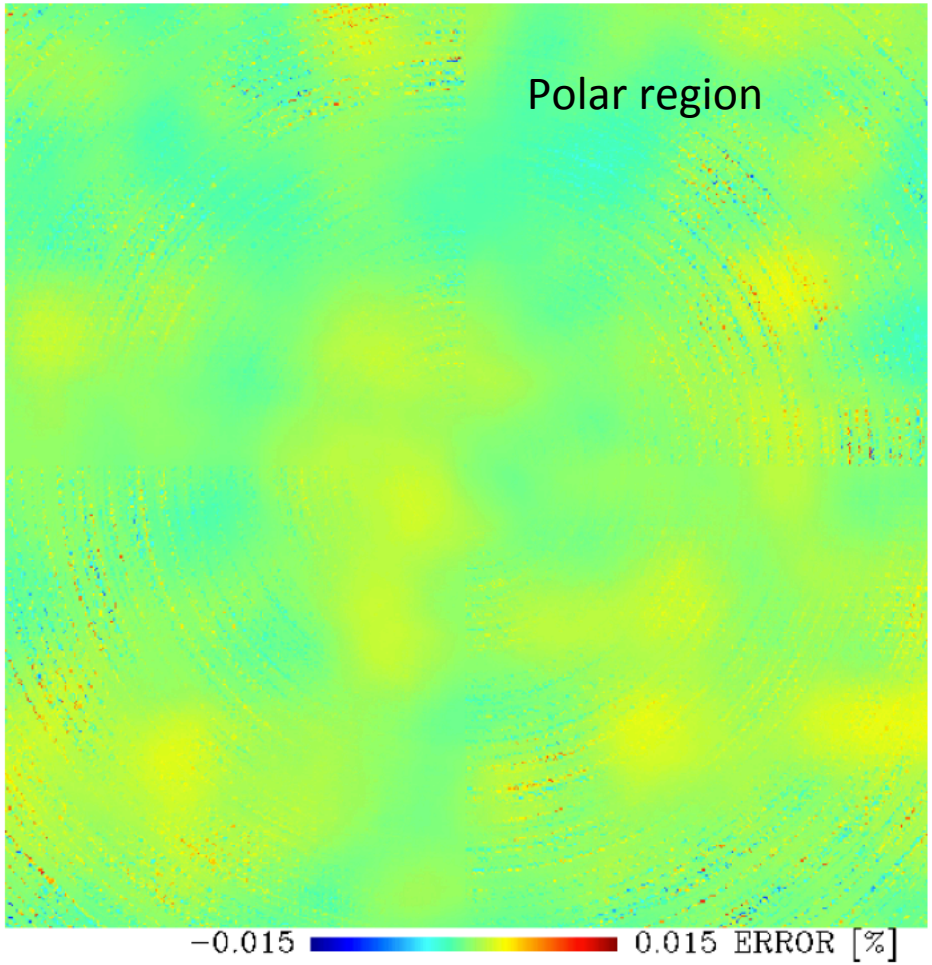


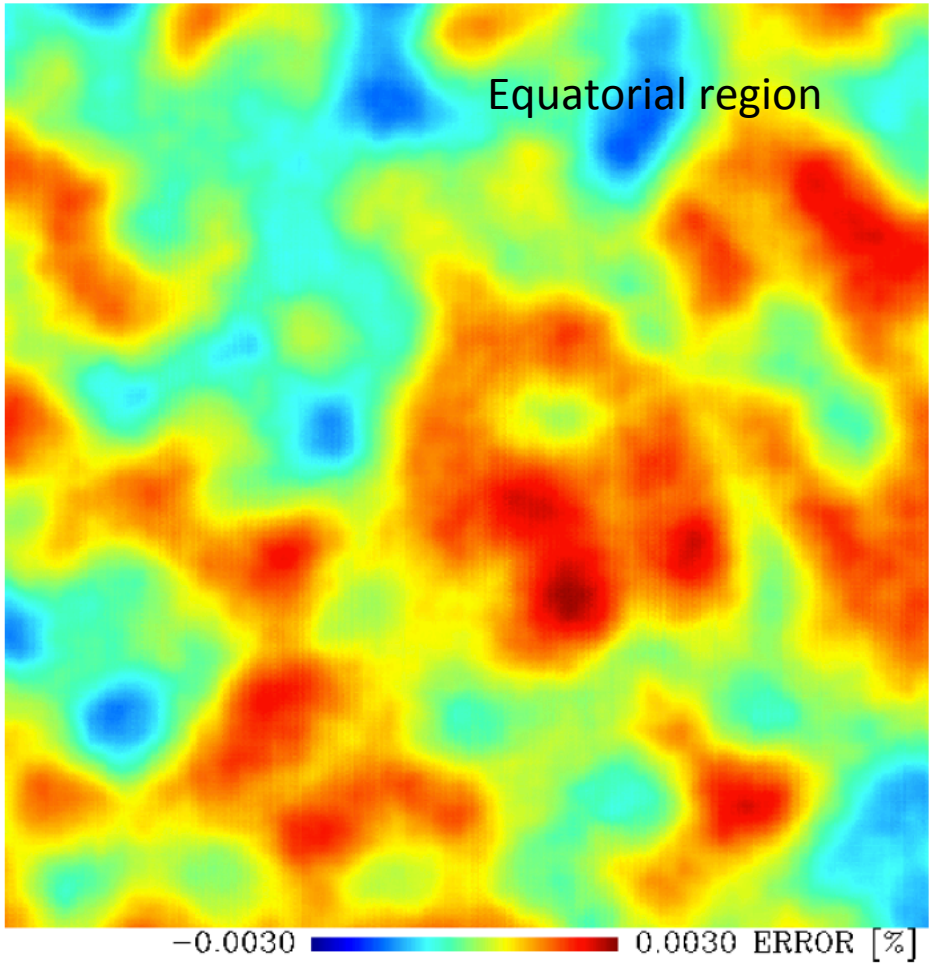
Test kernels





Accuracy: map diagnostic
RMS accuracy better than 0.0001



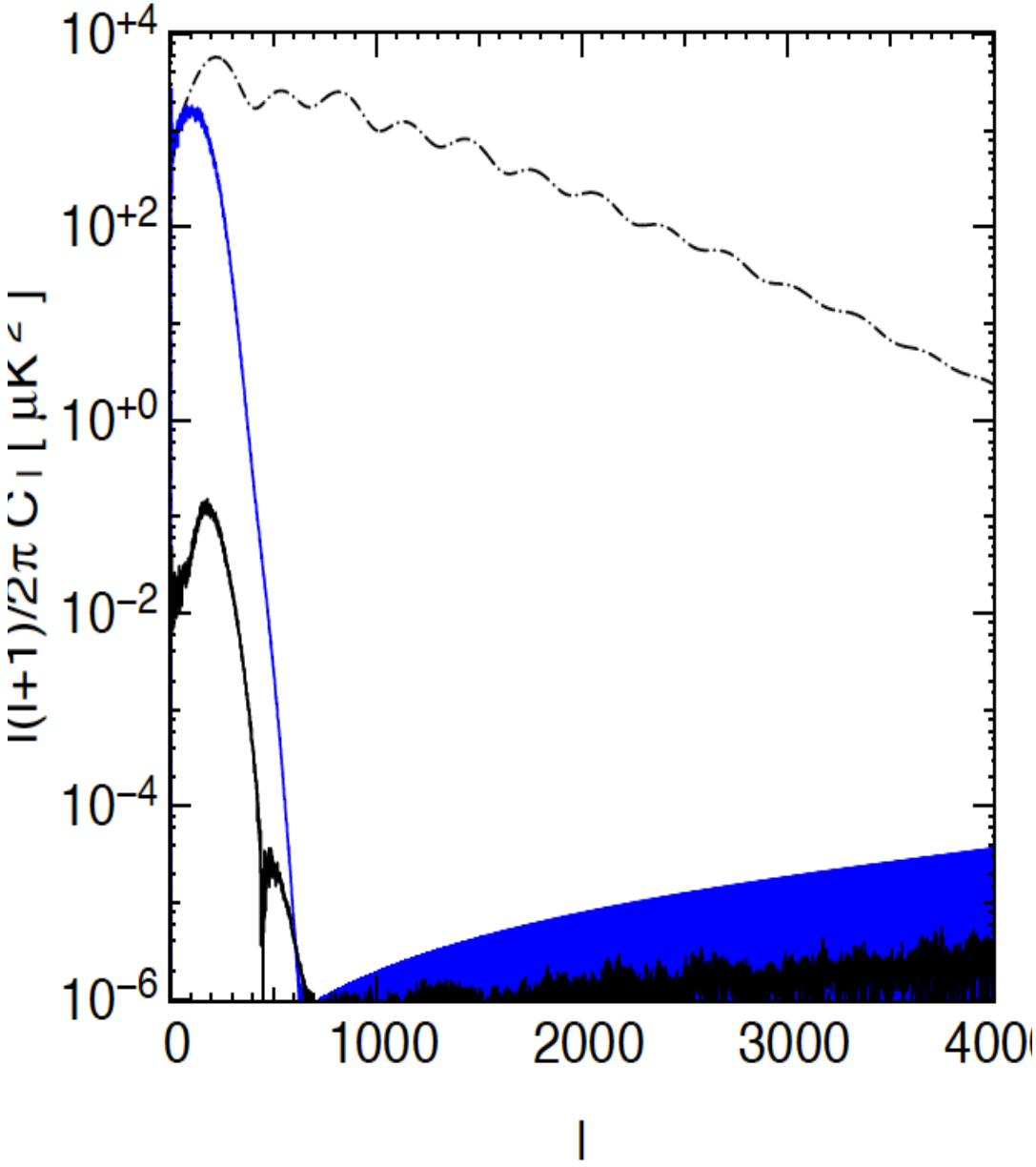


Accuracy: Power spectrum diagnostic

Smoothing with a Gaussian kernel
FWHM 1 degree

Smoothed map

Difference spectrum



GPU optimization

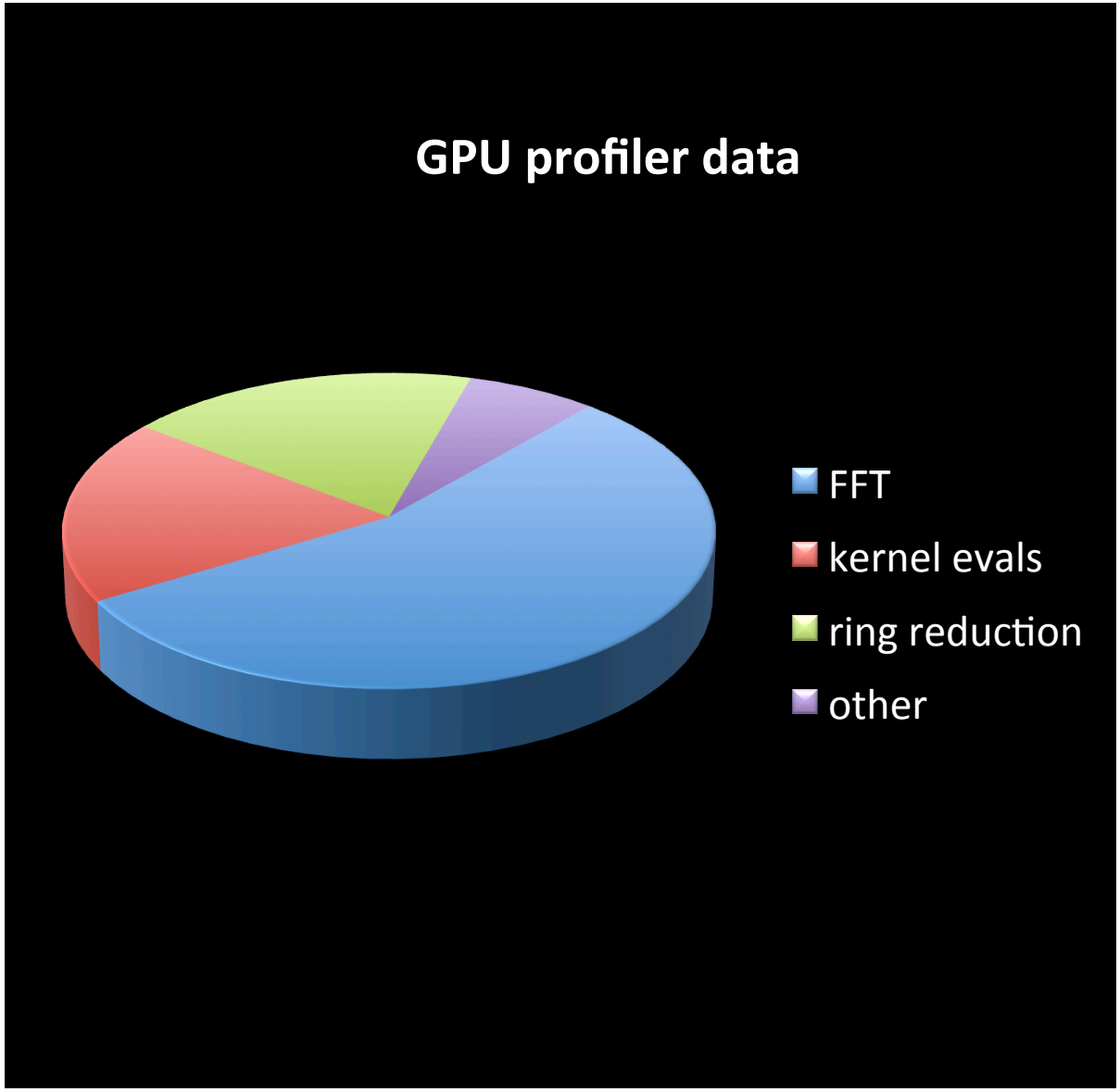
Minimize memory latency – reach over 80% of theoretical peak memory bandwidth (144GB/s out of 170 GB/s)

Exploit high level of data parallelism

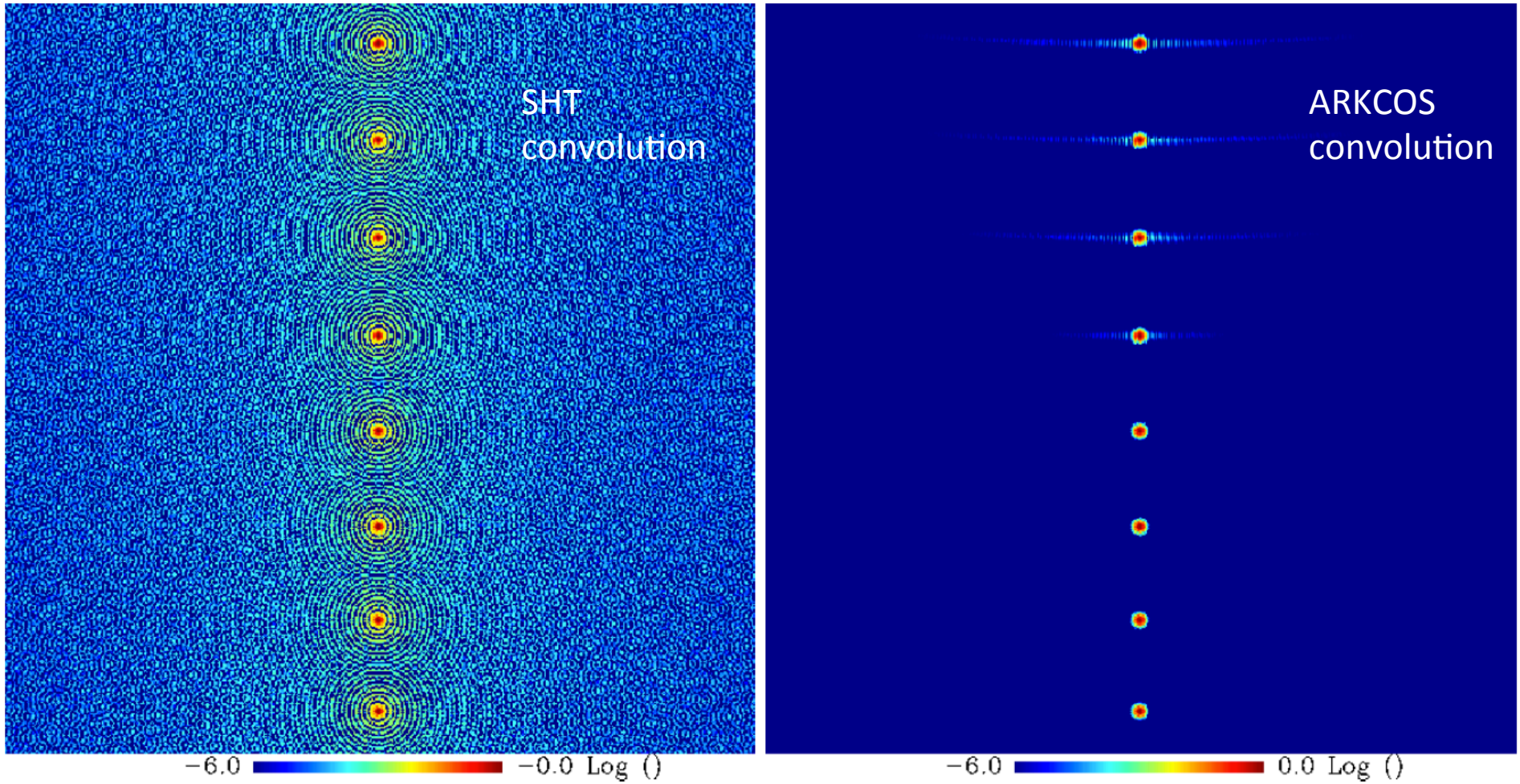
Instruction level parallelism

Use optimized library for batch FFTs (CUDAFFT)

Use single precision floats whenever possible



Side effect: suppress ringing artifacts



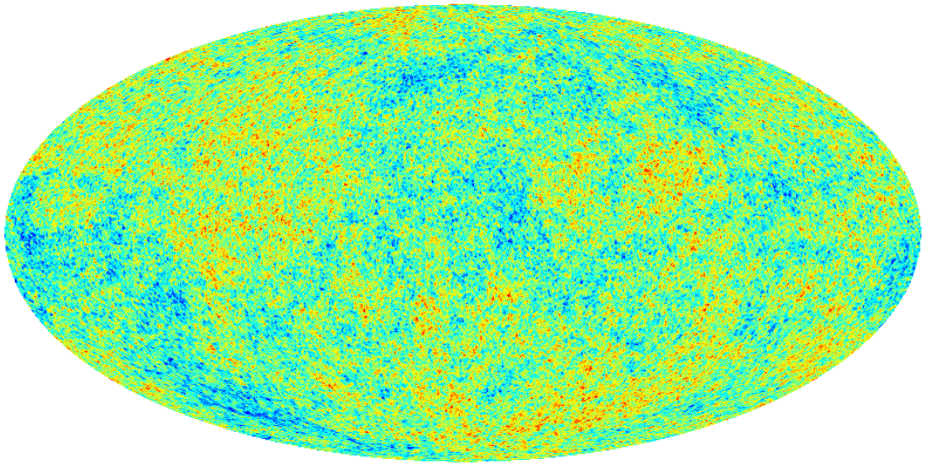
Generalization to non-compact kernels

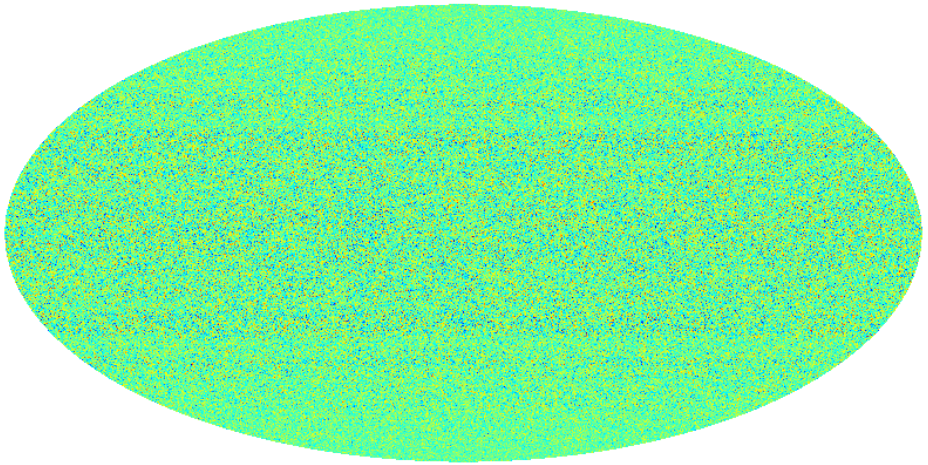
- Split kernel into two parts:
- one that is compact in real space and
- one that is compact spherical harmonic space

- Solve optimization problem to choose bandlimit and kernel footprint for the split to maximize predicted speed-up while respecting error bound

- Then perform convolution in parallel
  - SHT on CPU
  - ARKCOS on GPU

Kernel splitting results: 15x speed up

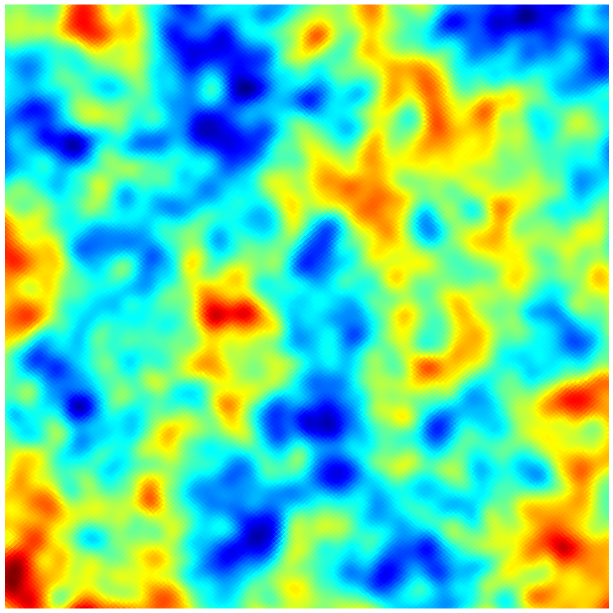


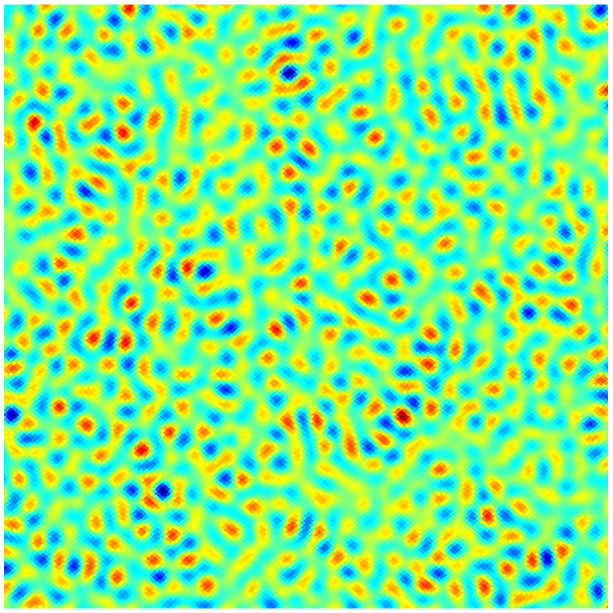


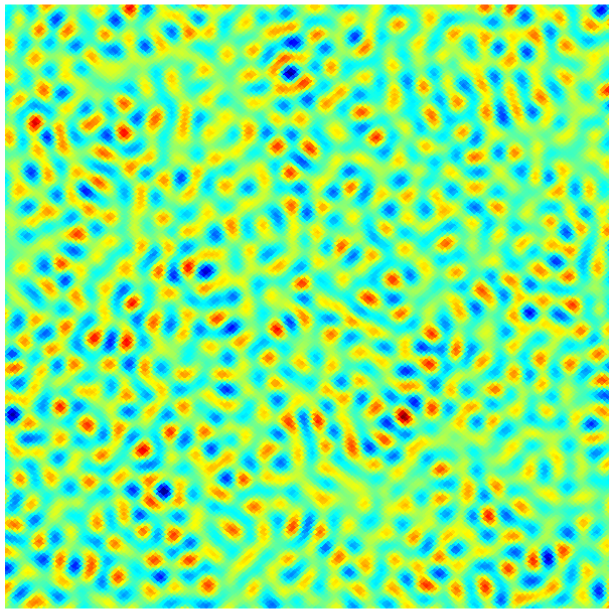
Full

l-space difference

ARKCoS







Compressed convolution reduces redundancy for large filter banks

- Problems with large filter banks:
  - "Continuous" wavelet expansion
  - Searching for many different "objects" (bubble collisions/textures in CMB or gravitational wave templates etc.) of unknown size
  - Simulations of CMB observations from $10^{2\text{-}5}$ different detectors

Elsner & Wandelt arXiv: 1312.3948

Compressed Convolution

- For an arbitrary set of kernels $s_i^{(n)} = \sum_j K_{i,j}^{(n)} d_j$
- One can expand the kernels in an small O(k) optimal basis to minimize the convolution error of all the outputs $K_{i,j} = \sum_k \lambda^k \widehat{K}(\phi^k)_{i,j}$
- Convolve with the kernel basis set and to obtain a compressed convolution output
- To get each output convolution decompress as needed in O(k) operations per output.

$$s_i^{(n)} = \sum_k \lambda^{(n),k} s_i^k$$

Constructing the basis kernels

- Minimize

$$\sigma^2 = \left\langle \sum_{(n)} \sum_{i,\,i'} s_i^{(n)} {N_{i\,i'}^{(n)}}^{-1} s_{i'}^{(n)} \right\rangle$$

$$= \sum_{n,\,i,\,i'} {N_{i\,i'}^{(n)}}^{-1} (R_{i'} k^{(n)}) C (R_i k^{(n)})^\dagger$$

for a fixed number k of basis kernels. By Schur's theorem they are the first k eigenvectors of

$$M_{nm} = \sum_{i,\,i'} \left[\left(N^{(n)\,-\frac{1}{2}} \right)^\dagger N^{(m)\,-\frac{1}{2}} \right]_{i\,i'} \left(R_{i'} k^{(m)} \right) C \left(R_i k^{(n)} \right)^\dagger$$

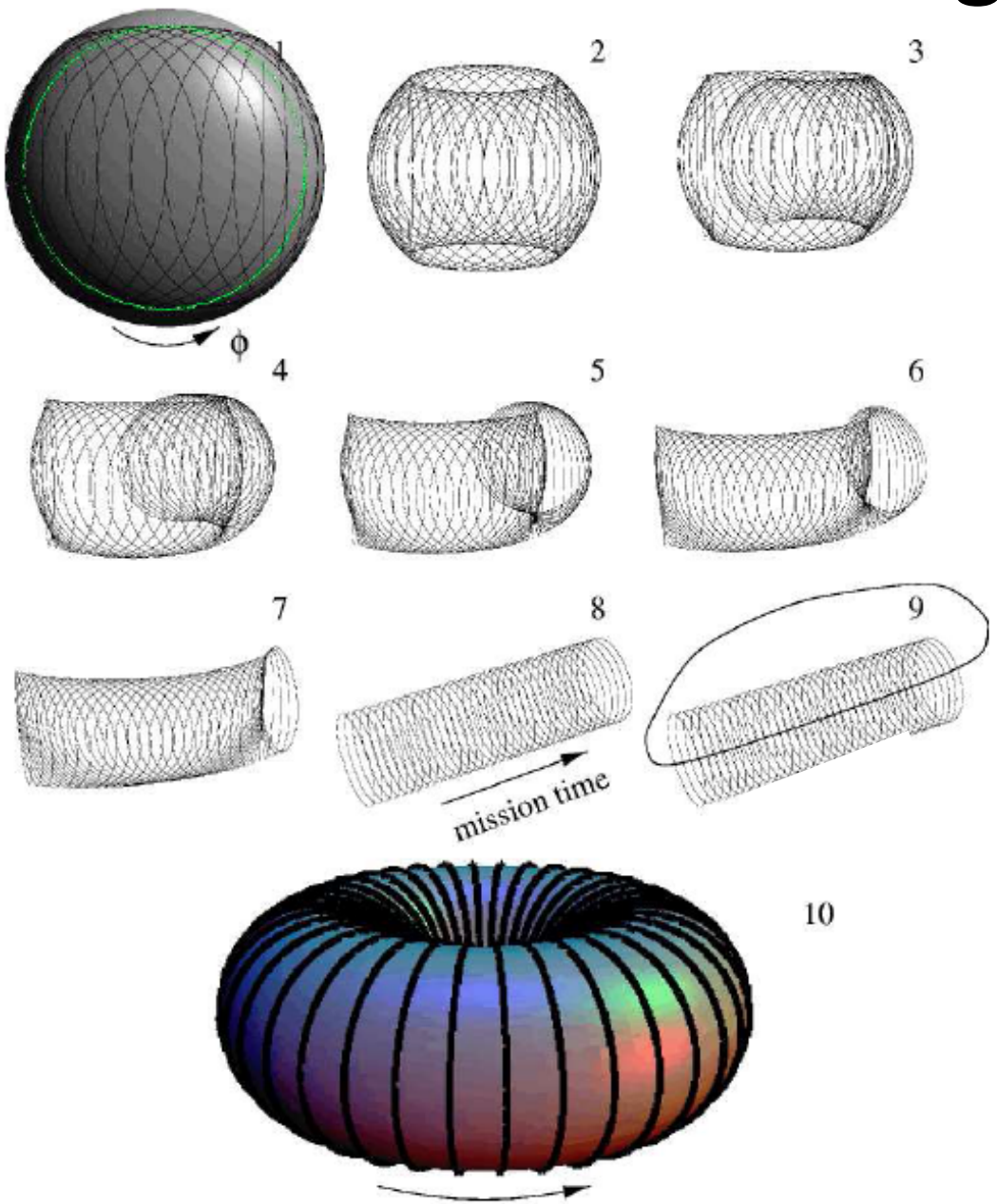
Torus representation of the sphere

- What are the best basis functions to represent functions on the sphere?
- Standard answer:

    Isotropy => Spherical Harmonics

But in the absence of fast, high-res non-Abelian transform on SO(3) (and maybe even more generally) the answer is not as clear cut.

Tools: the ring torus



- Can represent all or part of the sphere in terms of a torus double-cover.
- Then use the obvious Fourier basis.

BW, astro-ph/0012416

Fisher matrix

- For an isotropic signal on the sphere, the likelihood for the power spectrum C_l is

$$\mathcal{L}(C_\ell|d) = \frac{1}{\sqrt{|2\pi(\mathbf{S}(C_\ell)+\mathbf{N})|}} \times \exp\left[-\frac{1}{2}\, d^\dagger (\mathbf{S}(C_\ell)+\mathbf{N})^{-1} d\right]$$

- The Fisher matrix is very useful for the Cramer-Rao bound ML estimation, etc..

$$F_{\ell_1\ell_2} = -\left\langle \frac{\partial^2 \ln \mathcal{L}}{\partial C_{\ell_1} \partial C_{\ell_2}} \right\rangle = \frac{1}{2}\,\mathrm{tr}\left[\mathbf{C}^{-1}\mathbf{P}^{\ell_1}\mathbf{C}^{-1}\mathbf{P}^{\ell_2}\right]$$

- But cannot compute it! O(l_{max}^6) for bandlimit l_{max}

Tegmark arXiv:astro-ph/9611174

Signal correlations on the ring torus

$$\mathcal{T}_{mm'\,MM'} = \langle T_{mm'} T_{MM'} \rangle$$
$$= \boxed{\delta_{mM}} \sum_l C_l d^l_{mm'}(\theta_E) X_{lm'} d^l_{MM'}(\theta_E) X_{lM'}$$

Signal correlation becomes block diagonal!

Fisher matrix on the ring torus

- Go via ring torus basis – amazing property

$$\left(\frac{\partial \mathbf{C}_r}{\partial C_\ell}\right)_{p\,p'} = \left(\frac{\partial \mathbf{S}_r}{\partial C_\ell}\right)_{p\,p'}$$
$$= N^2\, d^\ell_{rp} X_{\ell p}\, d^\ell_{rp'} X^*_{\ell p'}$$
$$= q^\ell_{rp}\, q^{\ell\,*}_{rp'}\,,$$

the derivative in the torus basis is **rank 1**, as opposed to rang l in the spherical harmonic basis. So the torus basis gives a more compact representation for an isotropic signal than spherical harmonics!

Franz Elsner, BW
arXiv: 1202.4898
arXiv: 1205.0810

Bayesian power spectrum inference

- For an isotropic signal on the sphere, the likelihood for the power spectrum C_l is

$$\mathcal{L}(C_\ell|d) = \frac{1}{\sqrt{|2\pi(\mathbf{S}(C_\ell)+\mathbf{N})|}} \times \exp\left[-\frac{1}{2}\, d^\dagger (\mathbf{S}(C_\ell)+\mathbf{N})^{-1} d\right]$$

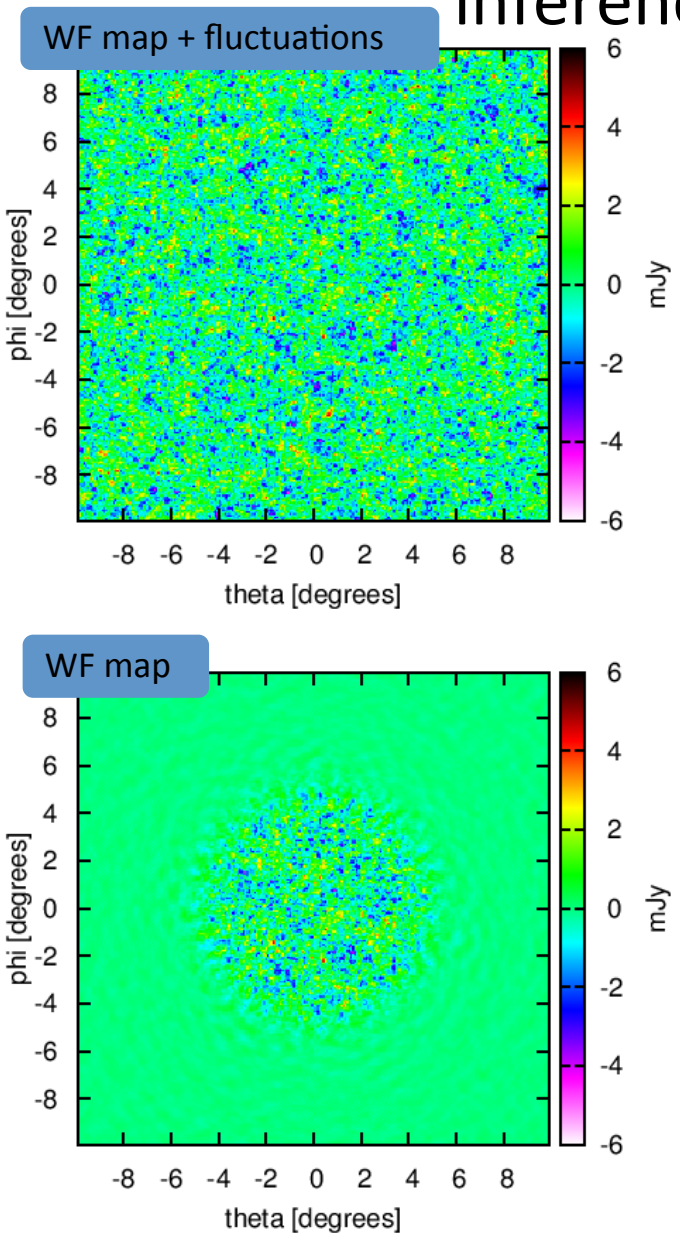
- But cannot compute it! O(l_{max}^6) for bandlimit l_{max}. *For Planck, l_{max}=2500, so t_{CPU}= 10^4 years, for a single evaluation*

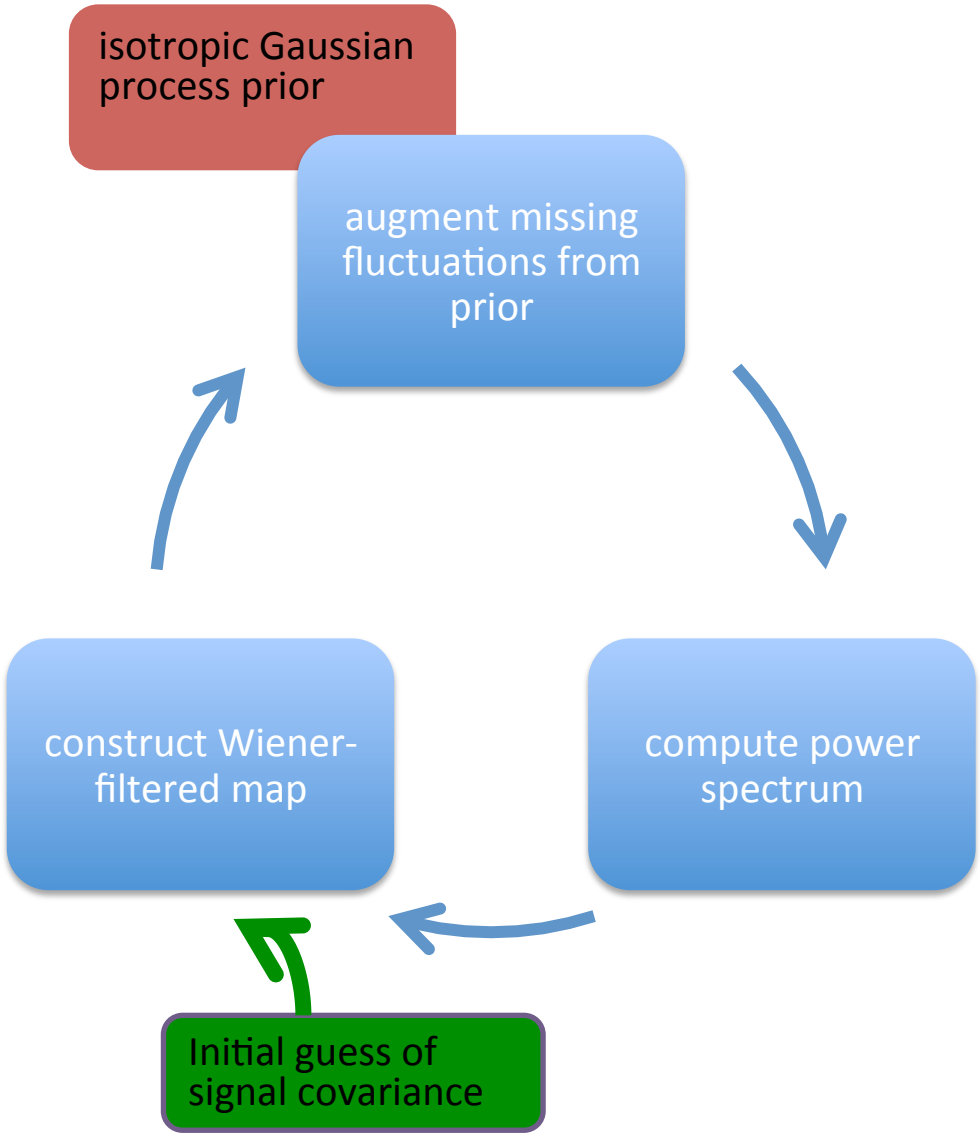
Gibbs sampling

- Augment the model with the signal map **s** – adds $O(10^7)$ parameters
- Gibbs sampling iteratively draws from the conditional densities $P(\mathbf{s}|\mathbf{d}, \mathbf{C}_l)$ and $P(\mathbf{C}_l|\mathbf{s})$ with a acceptance ratio **1**.
- This is possible in $O(l_{max}^3)$ operations, a speed-up of $\sim 10^9$).
- Used in final versions of WMAP and for low-ell Planck analysis.

Jewell, Levin, and Anderson (2004) and
Wandelt, Larson & Lakshminarayanan (2004)

Gibbs sampling is a both power spectrum inference and non-linear Wiener filter





Fundamental operation: Wiener Filter

$$d = s + n$$

Not sparse in any easily accessible basis

$$(\mathbf{S}^{-1} + \mathbf{N}^{-1})\, s_{\mathrm{WF}} = \mathbf{N}^{-1} d$$

Sparse in Fourier space

Sparse in pixel space

Usual solution strategies

- Iterative conjugate gradients (optimal for SPD matrices)
- Preconditioner (diagonal in Fourier space, multi-grid)
- Problems:
  - For Planck: ***extremely ill-conditioned***, condition number $>10^8$
  - Preconditioner not universal
  - Stability issues (Jacobi smoother in multi-grid)

Wiener Filtering without preconditioner

- Introduce auxiliary field (*messenger* field) t with covariance $\mathbf{T}$. Then

$$\left(\bar{\mathbf{N}}^{-1} + (\ \mathbf{T})^{-1}\right)\, t = \bar{\mathbf{N}}^{-1}\, d + (\ \mathbf{T})^{-1}\, s$$

$$\left(\mathbf{S}^{-1} + (\ \mathbf{T})^{-1}\right)\, s = (\ \mathbf{T})^{-1}\, t\,,$$

$$\bar{\mathbf{N}} \equiv \mathbf{N} - \mathbf{T}$$

is solved by the Wiener filter.

Can solve each of these equations *exactly.*

Iterate. Easy to show that this converges and is unconditionally stable.

Elsner & Wandelt, arXiv:1210.4931

Wiener Filtering without preconditioner

- Introduce auxiliary field (*messenger* field) t with covariance $\mathbf{T}$ and a parameter $\lambda \geq 1$. Then

$$\left(\bar{\mathbf{N}}^{-1} + (\lambda\mathbf{T})^{-1}\right) t = \bar{\mathbf{N}}^{-1} d + (\lambda\mathbf{T})^{-1} s$$

$$\left(\mathbf{S}^{-1} + (\lambda\mathbf{T})^{-1}\right) s = (\lambda\mathbf{T})^{-1} t\,,$$

is solved by the Wiener filter for $\lambda=1$.

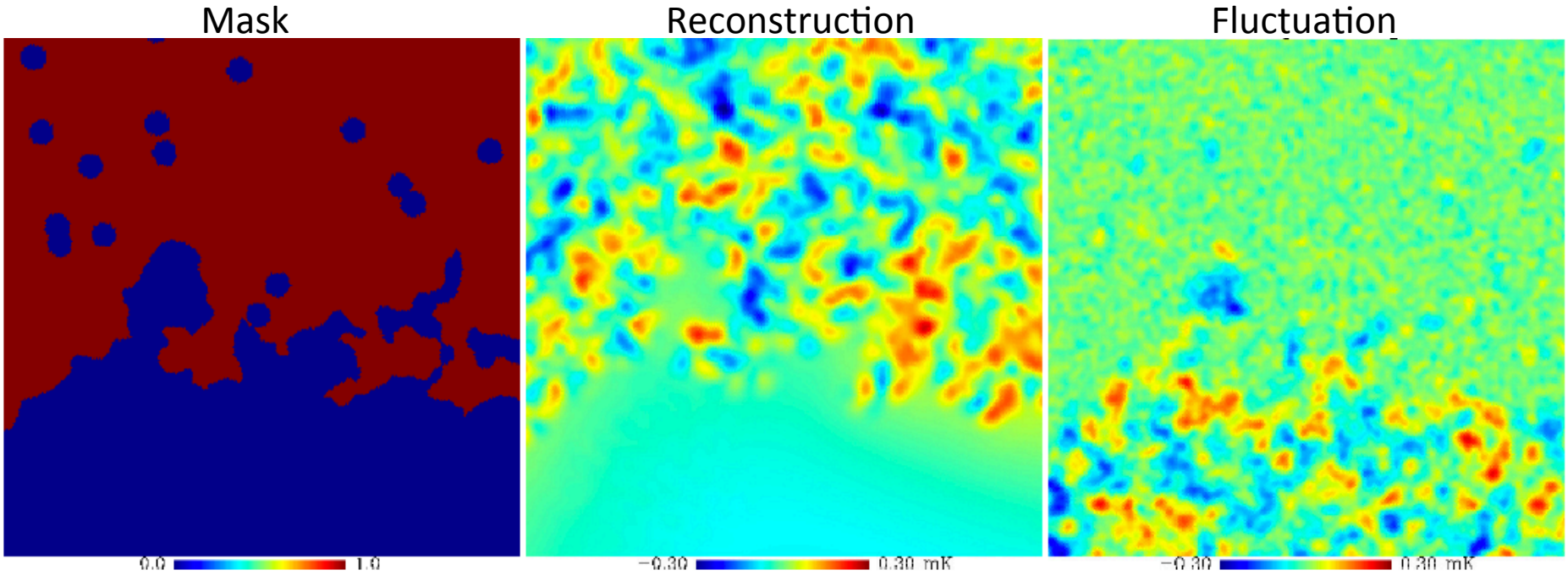
$$\bar{\mathbf{N}} \equiv \mathbf{N} - \mathbf{T}$$

Can solve each of these equations *exactly.*

Iterate. Easy to show that this converges and is unconditionally stable.

Elsner & Wandelt, arXiv:1210.4931

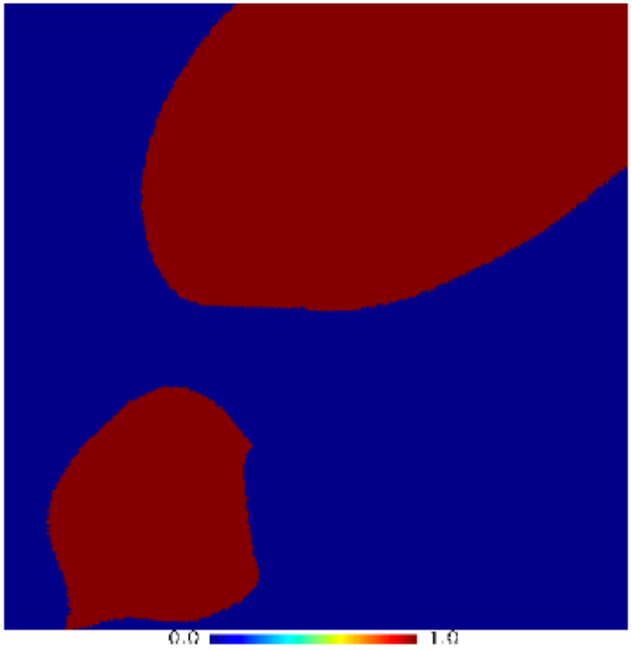
Optimal reconstruction, filtering and constrained realizations: WMAP temperature data



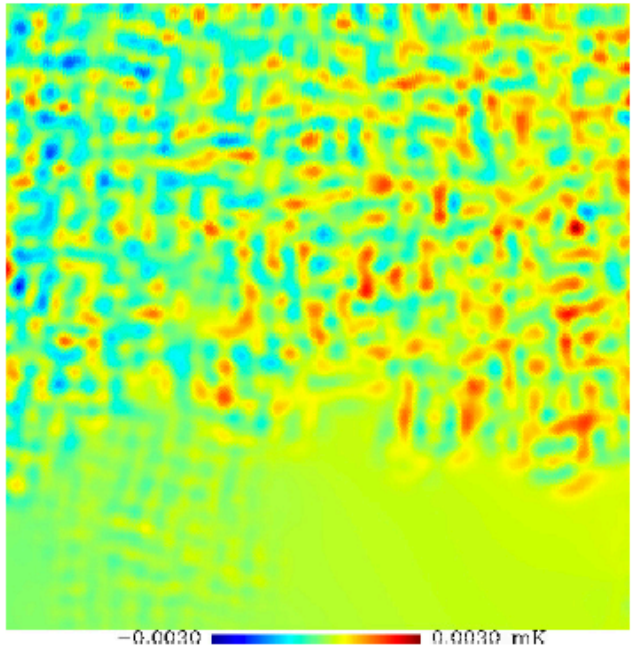
Elsner, Wandelt, arXiv:1210.4931

Polarization (WMAP data)

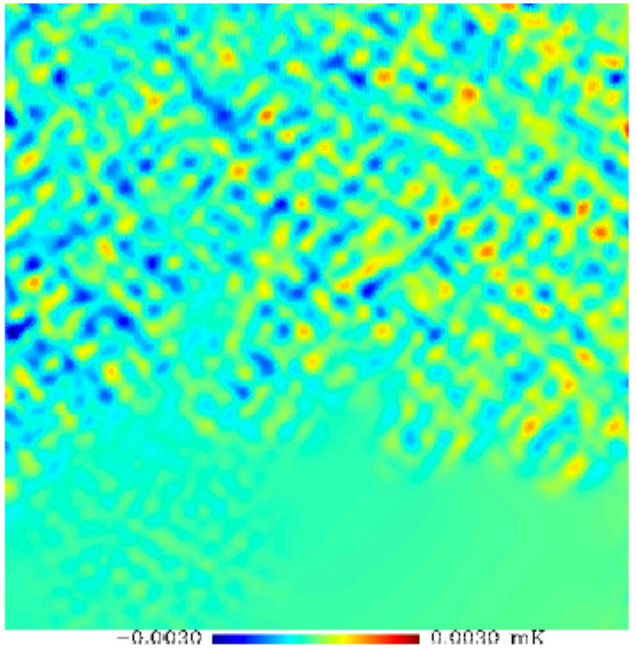
Mask



Stokes Q

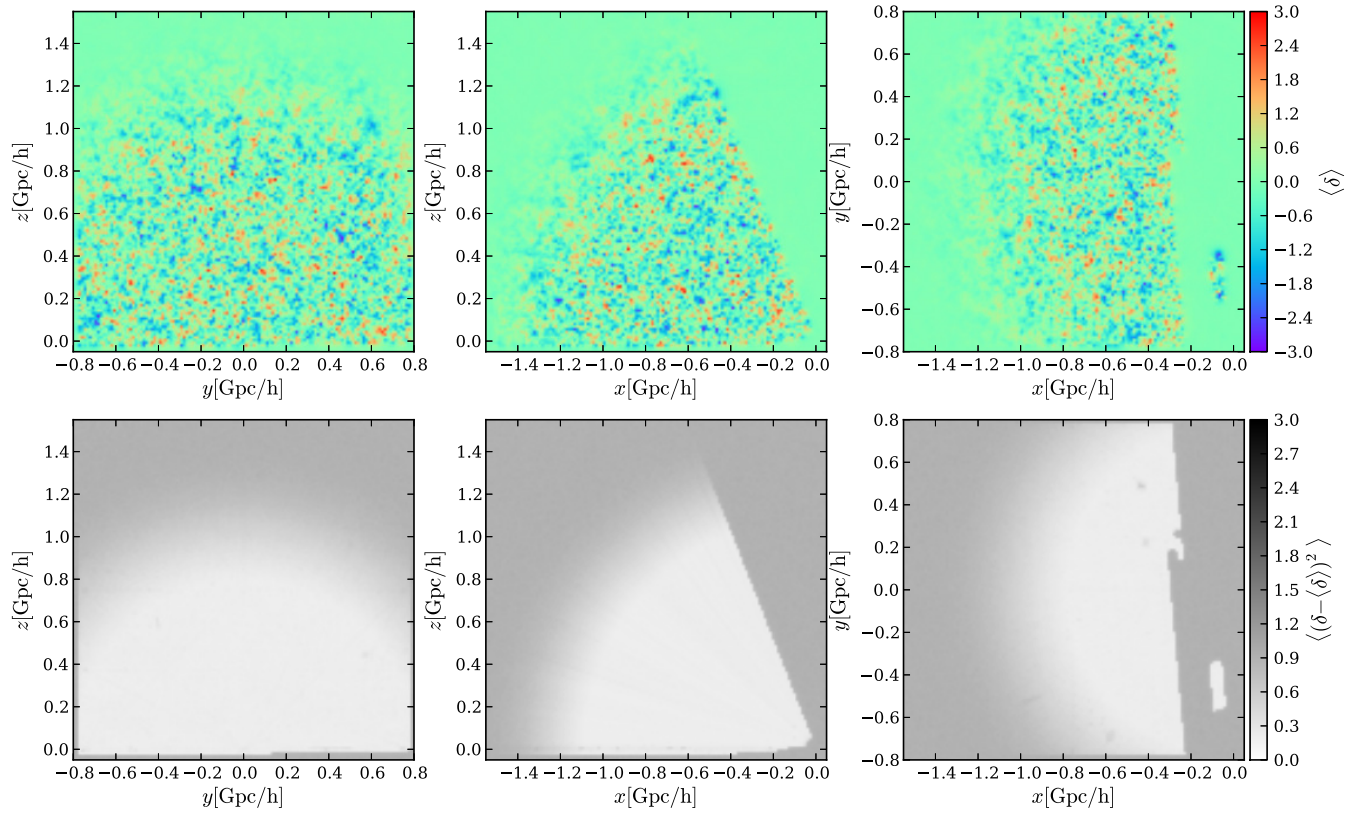


Stokes U



Elsner, Wandelt, arXiv:1210.4931

The messenger Wiener filter works even better for 3D LSS signal reconstruction and power spectrum



Sampling approaches to GRF model of large scale structure based on previous work with Jasche, Kitaura, Ensslin, Wandelt

Jasche, Lavaux, arXiv:1402.1763.

Beyond the Gaussian

- Three point correlations encode *shape* information – interest for geophysics?
- For a band-limit of L_{max} there are $O(L_{max}^2)$ pixels and $O(L_{max}^6)$ triangles on the sphere
- For an isotropic field going to spherical harmonics reduces scaling to $O(L_{max}^5)$.

Bispectrum measurement

Isotropic m-dependence

Theoretical template

$$\hat{f}_{\rm NL} = \frac{1}{N} \sum_{\ell_i, m_i} \mathcal{G}^{\ell_1\ \ell_2\ \ell_3}_{m_1 m_2 m_3} b^{\rm th}_{\ell_1 \ell_2 \ell_3}$$

$$\times \Big[C^{-1}_{\ell_1 m_1, \ell'_1 m'_1} a_{\ell'_1 m'_1}\, C^{-1}_{\ell_2 m_2, \ell'_2 m'_2} a_{\ell'_2 m'_2}\, C^{-1}_{\ell_3 m_3, \ell'_3 m'_3} a_{\ell'_3 m'_3}$$

$$- 3\, C^{-1}_{\ell_1 m_1, \ell_2 m_2} C^{-1}_{\ell_3 m_3, \ell'_3 m'_3} a_{\ell'_3 m'_3} \Big],$$

Variance-reducing linear term

Weighted data bispectrum

Meeting the computational challenge

$$\hat{f}_{\rm NL} = \frac{1}{N} \sum_{\ell_i, m_i} \mathcal{G}^{\ell_1\ \ell_2\ \ell_3}_{m_1 m_2 m_3} b^{\rm th}_{\ell_1 \ell_2 \ell_3}$$
$$\times \left[C^{-1}_{\ell_1 m_1, \ell_1' m_1'} a_{\ell_1' m_1'}\, C^{-1}_{\ell_2 m_2, \ell_2' m_2'} a_{\ell_2' m_2'}\, C^{-1}_{\ell_3 m_3, \ell_3' m_3'} a_{\ell_3' m_3'} \right.$$
$$\left. - 3\, C^{-1}_{\ell_1 m_1, \ell_2 m_2} C^{-1}_{\ell_3 m_3, \ell_3' m_3'} a_{\ell_3' m_3'} \right],$$

Scales as $O(L_{max}^5)$

- Brute force implementation unfeasible for Planck data, $L_{max} \sim 2000$
- Key idea: KSW – factorization (Komatsu, Spergel, BDW, arXiv:astro-ph/0305189)

$$b_{\ell_1 \ell_2 \ell_3} = \sum_{ijk} X^i_{\ell_1} Y^j_{\ell_2} Z^k_{\ell_3}$$

- Gives $O(L_{max}^3)$ scaling: $\sim 10^6$ for Planck
- Generalization to more general shapes using sums of separable templates (Fergusson, Liguori, Shellard arXiv: 0912.5516)

Conclusions

- Many operations on spherical data sets are convolutions
- Going to torus representations gives new ways of attacking hard problems and massive speed-ups
- GPU acceleration for radial kernel convolution
- Compressed convolution reduces redundancy for large filter banks