Slides ISSPIT 2015

Miguel Cárcamo
University of Santiago, Chile

Fernando R Rannou
University of Santiago, Chile

Pablo Enrique Roman
University of Santiago, Chile

High Performance GPU Bayesian Image Synthesis

ISSPIT 2015

M. Cárcamo[1] and F.R. Rannou[1] and P. E. Román[2] and V. Moral[3] and S. Casassus[3]

[1]Departamento de Ingeniería Informática,
Universidad de Santiago de Chile.
[2]Center of Mathematical Modeling.
Universidad de Chile.
[3]Departamento de Astronomía.
Universidad de Chile.

Partially financed by Fondecyt project 3140634 and Fondequip EQM-140101.

December, 2015

Table of contents I

Radio Interferometry

Measurements: Visibilities $V_{i,j}$

Time averaged correlation
$V_{i,j} = < \vec{E}_i^* \vec{E}_j >$

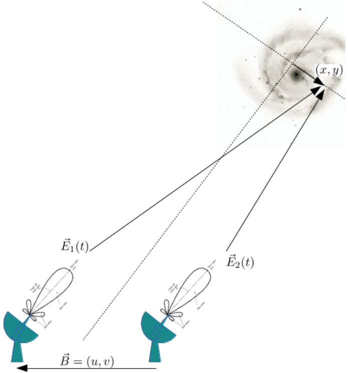
Van Cittert-Zernike theorem

$V(u,v) =$
$\int_{\mathbb{R}^2} \mathcal{A}(x,y) I(x,y) e^{-2\pi i(ux+vy)} dxdy$
Visibility $\sim$ Fourier transform

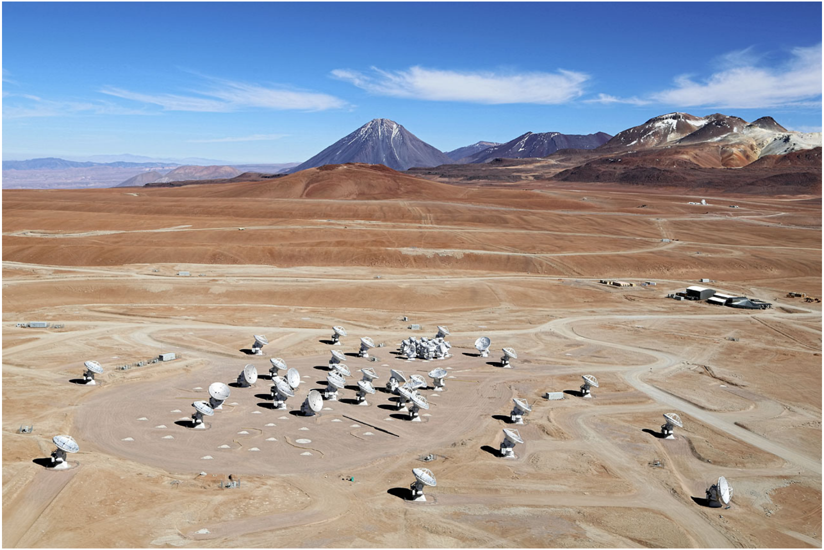
Rough estimate resolution

$\propto \lambda / B$

[Taylor et al., 1999]



Radio interferometer (ALMA)



Main issue: Non-Regular Sampling

Visibility-Image functional relationship

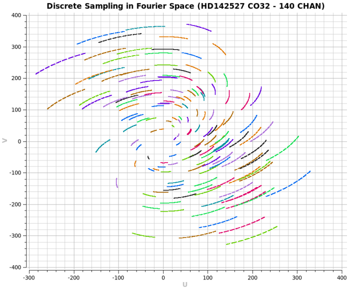
$$V(u,v)=\mathcal{F}[AI]$$

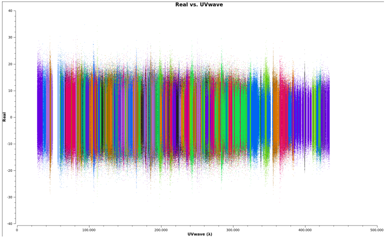
Sampling path by regular time step

Earth rotation ($\approx$phase+radial): elliptic
Frequency ($\approx$radial): rays
(uv plane in [Distance/wavelength])

Main characteristics

Far from regular: Ill posed problem.
Noisy data: $O(\sigma)\sim O(|V|)$
Real valued image:
$\bar{V}(-z)=V(z),\ z=(u,v)$





Objectives and data size

Purpose

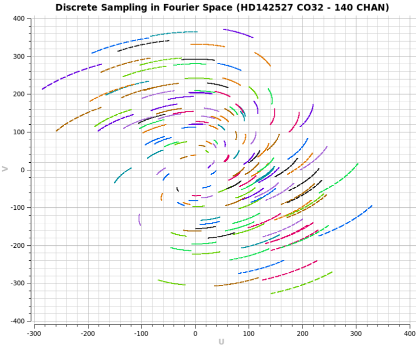
Estimating $E(I)$ from a sparse sampling of $\mathcal{F}[I]$

Data size

Number of samples location:
$N \sim 10^4 - 10^8$
Number of unknown variables (image):
$n^2 = 10^6$

Data description

Samples location: $\{z_k = (u_k, v_k)\}_{k=1}^N$
Visibilities : $\{V_k^o\}_{k=1}^N$
Variances : $\{\sigma_k\}_{k=1}^N$



CLEAN Algorithm (1974): Standard in Radio Astronomy

Sampled visibility function

$$V^s = S \cdot V$$
$$S(u,v) = \sum_{k=1}^{N} \omega_k \delta(u_k, v_k)$$

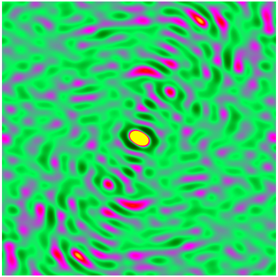
The deconvolution problem

$$I^D = \sum_{k=1}^{N} \omega_k V(u_k, v_k) e^{2\pi i (x_j u_k + y_j v_k)}$$
$$B^D = \sum_{k=1}^{N} \omega_k Cos(2\pi(xu_k + yv_k))$$
$$I^D = B^D * (I)$$

The algorithm (from 1974)

Greedy algorithm from the class of matching pursuit.
Add Gaussian components to the restored image [Rau and Cornwell, 2011].

B^D



Bayesian image synthesis: Maximum Entropy

Maximum Log-likelihood using Bayes

$$\max_I \log(\mathbb{P}(I|D)) \propto \log(\mathbb{P}(D|I)) + \log(\mathbb{P}(I))$$

Log-Prior of Gaussian measurements

$$\log(\mathbb{P}(D|I)) = -\sum_{k=1}^{N} \frac{|V_k^{obs} - V(z_k)|^2}{\sigma_k^2} = -2\chi^2$$

Log-Prior of the image: Thermal Entropy

$$\log(\mathbb{P}(I)) \propto -\sum_{l=1}^{n} I_l \log(I_l/M) = -S$$

[Narayan and Nityananda, 1986] [Levanda and Leshem, 2010]
[Sutton and Wandelt, 2006] [Donoho et al., 1991]

Maximum Entropy Method (MEM)

MEM optimization problem

$$\min_{I \geq 0} \quad \chi^2[I] + \lambda S[I]$$

MEM

- Super resolution capabilities (to 1/3)
- Positivity
- Suited for "nearly black" images.
- No user interaction
- Expensive Nonlinear problem
- Control of the noise
- better SNR

CLEAN

- Resolution to HWHM of the PSF
- Positivity is relaxed
- Allows several image scales.
- Best results with user guided deconvolution
- Efficient heuristic (equiv. to a Matching Pursuit)
- Weighting of the dirty beam

MEM in a parallel setting

New avenue for MEM

- Recovering from noisy data and higher definition/Sensitivity images is attractive
- Most astronomical objects in protoplanetary disk research are "nearly black"
- Need to compare several algorithms
- MEM is highly suitable for parallelism in gradient based method
- Current hardware facilities are becoming much more cheaper (GPU)

Artist conception of a protoplanetary disk

GPU: large number of streaming processors (SP)

Paradigm:data oriented parallelism

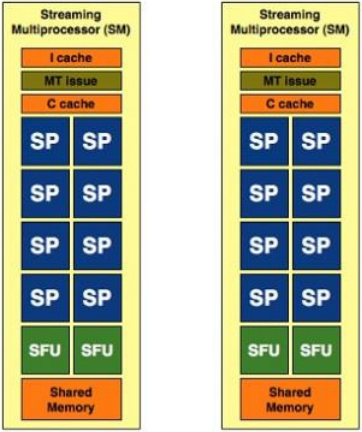
The same code (Kernel) is launched simultaneously in each SP but over distributed data.

Execution model: Thread

Each Thread in a SP has its own independent memory.

Hardware implementation: Streaming Multiprocessor

A Kernel runs on a SM accounting a Block of data.



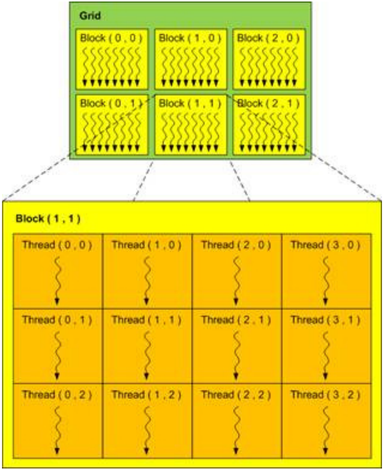
The Grid for data distribution in Threads

Organizing data for a single code

A Grid indexes the set of Block in a 1d, 2d, or 3d array.

Level of communication: In Block

Inside a Block Threads can use Shared memory, Barrier synchronization, and atomic operations.



CUDA programming

The program

Host program + Kernels

Host program

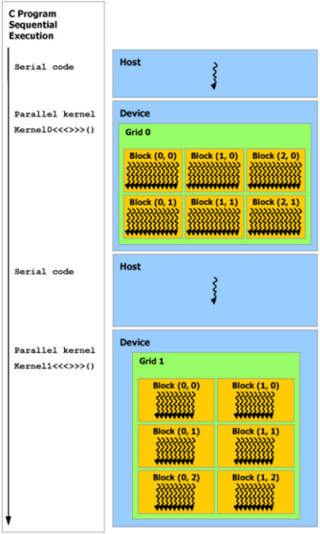
Executes in CPU

Kernel

Executes in GPU Device

Build a CUDA program

Organize execution in list of Grids, Grid of Blocks, Block of Threads.



MEM-CUDA Algorithm

The problem

$$\min_{I \geq 0} \quad \phi(I) = \chi^2[I] + \lambda S[I]$$

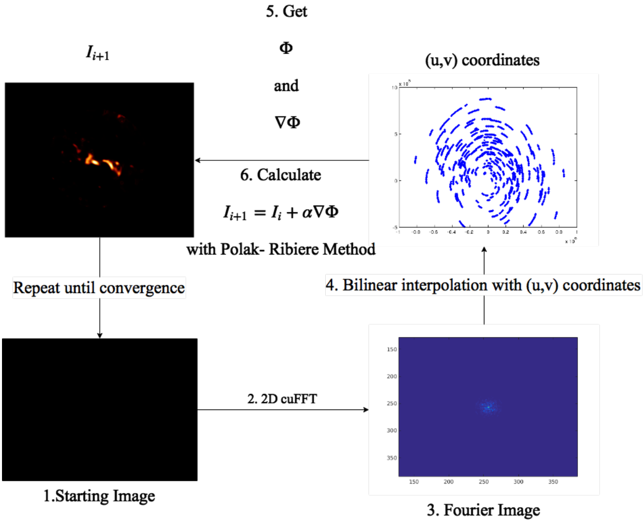
Algorithm base

Conjugate gradient : Handle high dimensionality problem
Involve computation of : $O(\phi(I)) \sim \max(N, n^2)$ and $O(\nabla\phi(I)) \sim N \cdot n^2$
Remember $N \sim 10^8..10^4$, $n^2 \sim 10^6$
But also remember, number of SP $\sim 10^3$ per device

The bottleneck

The calculation of the gradient use most of the processors resources in particular the χ^2 term.
$[\nabla\chi^2]_{k=1\ldots n^2} = \sum_j^N (\mathrm{Re}(V_j^R)\cos(2\pi < X_k, Z_j >) - \mathrm{Im}(V_j^R)\sin(2\pi < X_k, Z_j >)\frac{1}{\sigma_j^2}$
But each term in summation does not depends on other term!

The process


5. Get
Φ
and
$\nabla\Phi$
I_{i+1}
(u,v) coordinates
6. Calculate
$I_{i+1} = I_i + \alpha\nabla\Phi$
with Polak- Ribiere Method
Repeat until convergence
4. Bilinear interpolation with (u,v) coordinates
2. 2D cuFFT
1.Starting Image
3. Fourier Image

The conjugate gradient

Polak Ribiere Algorithm

1: Select $I_0 \in \Re^n$
2: $d_0 = -\nabla\Phi(I_0)$
3: **for** i=0 **to** MAX ITERATIONS **do** {For loop in host}
Compute α_i such that $\Phi(I_i + \alpha_i d_i) = \min_\alpha \Phi(I_i + \alpha_i d_i)$
5: Set $I_{i+1} = I_i + \alpha_i d_i$
6: **if** $||\nabla\Phi(I_{i+1})||^2 \leq \epsilon$ **then** {Sum reduction}
7: Stop.
8: **else**
9: $g_{i+1} = -\nabla\Phi(I_{i+1})$
10: $\beta_i = \frac{g_{i+1}^T(g_{i+1} - d_i)}{d_i^T d_i}$ {Sum reduction}
11: $d_{i+1} = g_{i+1} + \beta_i d_i$
12: **end if**
13: **end for**

line 4: The line minimization is done in host calling the Φ kernel.

line 10: The numerator and denominator are calculated separately.

Gradient computation per thread

The basic operation

$$\nabla\phi_{k,j} = (\mathrm{Re}(V_j^R)\cos(2\pi < X_k, Z_j >) - \mathrm{Im}(V_j^R)\sin(2\pi < X_k, Z_j >)\frac{1}{\sigma_j^2}$$

The thread

$\nabla\phi_k = 0$
for $(j = 1 \rightarrow N)$ **do**
    $\nabla\phi_k {+}{=} \nabla\phi_{k,j}$
end for

Experimental Setting

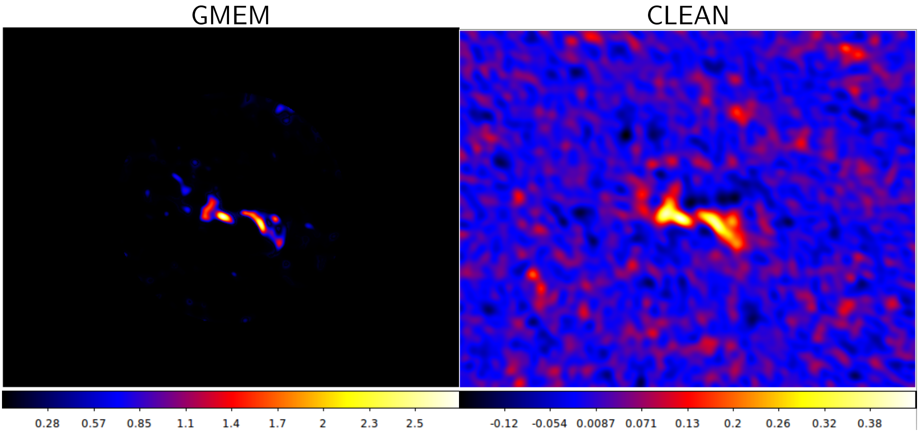
Harware

2x 10 Core Intel Xeon E5-2640 V2 2GHz
256 GB RAM
4 NVIDIA Tesla K80 (We used only the half of one = NVIDIA K40)

Data

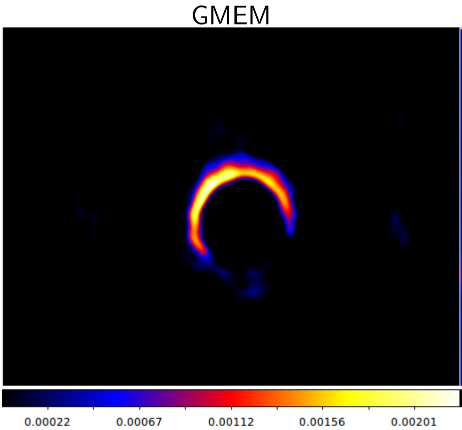
Cycle-0 ALMA Data ADS/JAO.ALMA#2011.0.00465.S
Protoplanetary disk object HD142527
Frequency setting on the emission line CO 6-5 and HCO+
Frequency setting on the continuum emission (band9)

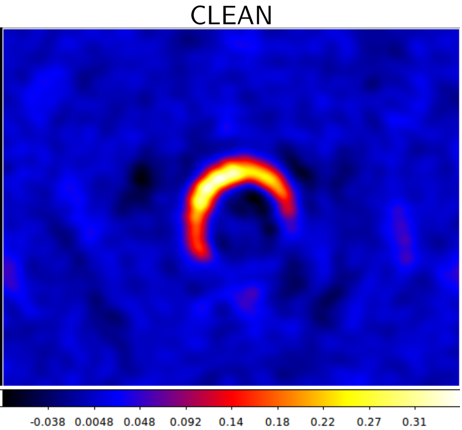
CO 6-5 Line emission (691GHz)



18.27 s, 8 it
78762 Vs
[Casassus et al., 2015]

Continuum emission at 700GHz



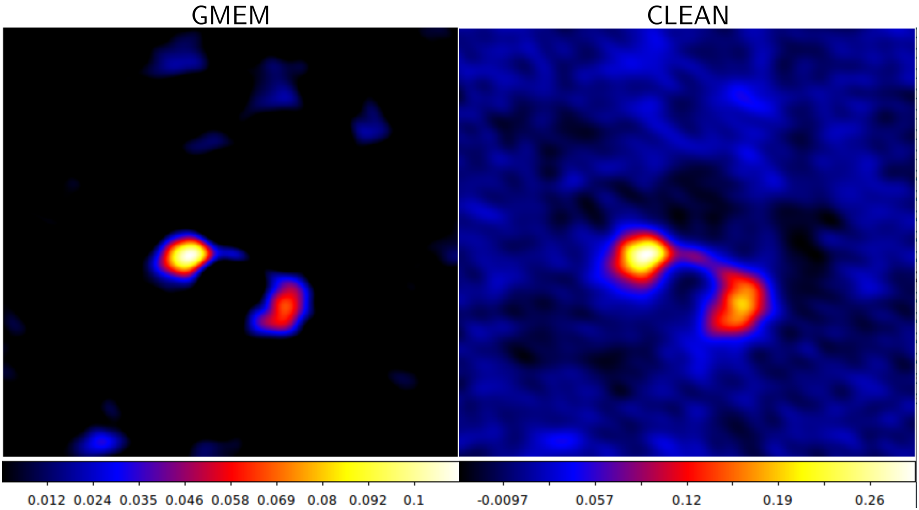


6.65 s, 6 it

26125 Vs

[Casassus et al., 2013]

HCO+ line emission 356.7GHz



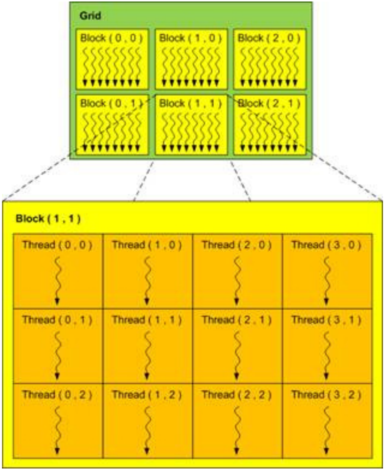
4.27 s, 6 it
16065 Vs

Computational performance

- Test set: co65_ext_chan64
- Number of Visibilities: 78762
- CPU Cores: 20
- GPU Cores: 2496

Image Size	**Time (s)**		**Speedup**
	CPU	**GPU**	
128x128	49	2.79	17.56
256x256	189	11.87	15.92
512x512	1127	13.13	85.83
1024x1024	6721	18.27	367.87

Block size performance (512x512 image)



(N_x, N_y)	Time (s)
4, 8	28.32
8, 4	28.32
8, 8	16.44
4, 16	16.44
16, 4	16.52
8, 16	12.41
16, 16	12.49
32, 4	12.52
8, 32	12.38
16, 32	12.29
32, 32	12.48

Conclusions

GPU sucessfully speedup MEM

Experiment shows drastic speedup in MEM

Impact in scientific research

The MEM algorithm becomes practical for interferometry research.

Montecarlo processing is now feasible

Boosting enable obtaining error bars in images.

Inviting further research on MEM

MEM assesment and comparizon with compressed sensing methods reveals to be interesting.

Future work

Future work will tackle the use of multiple GPUs

Bibliography I

S. Casassus, G. van der Plas, W. R. F. Dent, E. Fomalont, J. Hagelberg, A. Hales, A. Jordan, D. Mawet, F. Menard, A. Wootten, D. Wilner, A. M. Hughes, M. R. Schreiber, J. H. Girard, B. Ercolano, H. Canovas, P. E. Roman, and V. Salinas. Flows of gas through a protoplanetary gap. *Nature*, 493:191–194, 2013.

S. Casassus, S. Marino, S. Perez, P. Roman, A. Dunhill, P. J. Armitage, J. Cuadra, A. Wootten, G. van der Plas, L. Cieza, V. Moral, V. Christiaens, and M. Montesinos. Accretion kinematics through the warped transition disk in hd142527 from resolved co(65) observations. *The Astrophysical Journal*, 811 (2), 2015.

D. L. Donoho, I. M. Jhonston, J. C. Hoch, and A. S. Stern. Maximum entropy and the nearly black object. *Journal of the royal statistical society, Serie B, Methodological*, 24(1), 1991. doi: 10.2307/2345948.

R. Levanda and A. Leshem. Synthetic aperture radio telescopes. *IEEE Signal Processing Magazine*, 27:14–29, Jan. 2010. doi: 10.1109/MSP.2009.934719.

Bibliography II

R. Narayan and R. Nityananda. Maximum entropy image restoration in astronomy. *Annual review of astronomy and astrophysics*, 24:127–170, 1986. doi: 10.1146/annurev.aa.24.090186.001015.

U. Rau and T. J. Cornwell. A multi-scale multi-frequency deconvolution algorithm for synthesis imaging in radio interferometry. *Astronomy and Astrophysics*, 532, 2011.

E. C. Sutton and B. D. Wandelt. Optimal Image Reconstruction in Radio Interferometry. *The Astrophysical Journal Supplement Series*, 162:401–416, Feb. 2006. doi: 10.1086/498571.

G. B. Taylor, C. L. Carilli, and R. A. Perley. *Synthesis Imaging in Radio Astronomy II*. Number 180 in Astronomical Society of the Pacific Conference Series. ASP, 1999. ISBN 1-58381-005-6.