Principles of Echolocation
in Bats and Machines

Note 11

Why Random Arrays?

Grating lobes, bandwidth and the layout of the eRTIS microphones

Jan Steckel
Cosys-Lab, University of Antwerp, and Flanders Make, Belgium jan.steckel@uantwerpen.be

Version 2026-11-18 16:40

1 Few Microphones Over a Large Aperture

Every 3D sonar of our lab has one emitter and an array of MEMS microphones, and every one of them had to answer the same question first. Where do the microphones go? Figure 1 shows four answers. HiRIS fills a large plate with 1024 microphones on a regular grid. The μRTIS puts 30 microphones on a small regular grid, and the nRTIS puts 16 on a circle around its emitter. The eRTIS spreads 32 microphones over a plate of about 66 by 38 mm, and if you look at its front plate, the holes seem to be scattered at random. That is on purpose.

The reason is a trade-off between the number of channels and the size of the array. The aperture, the size of the array, sets how sharp the images are. The spacing between the microphones sets whether the images can be trusted. With 32 channels, a regular grid can have one of the two, but not both. In this note we show why, and how a random layout with a broadband call gets most of both.

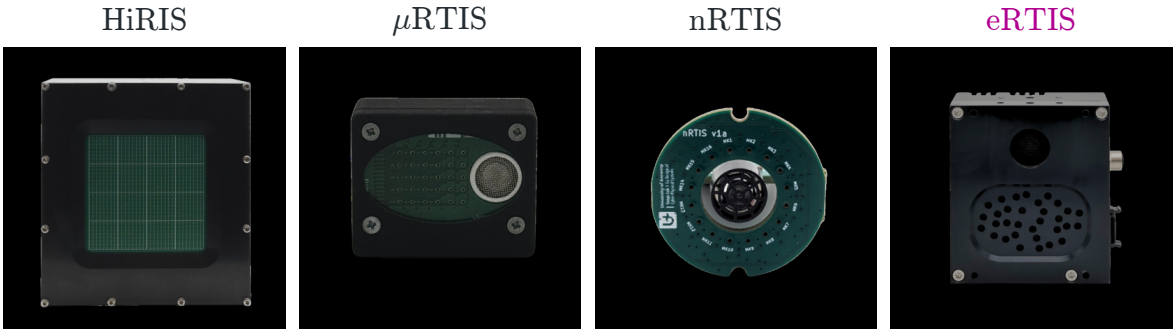


Figure 1: Four Cosys-Lab 3D sonar sensors (photos from www.3dsonar.eu). HiRIS and the μRTIS use dense regular grids, the nRTIS a circle, and the eRTIS a sparse pseudo-random layout with a large aperture.

2 Direction Becomes Delay

We use the conventions of Note 5, *Beyond Delay-and-Sum*, which derives the signal model in full. The sensor looks along x, y is its left and z is up. The microphones lie in the y-z plane,

and $\boldsymbol{r}_m$ is the position of microphone m relative to the centre of the array, our reference point (Fig. 2a). A far source sends a plane wave, and we describe its direction with the unit vector $\boldsymbol{u}$, which points from the array towards the source. Microphone m then receives the waveform with the delay

$$\tau_m = -\frac{\boldsymbol{u}^T \boldsymbol{r}_m}{c} \tag{1}$$

relative to the reference point, with $c \approx 343$ m/s. The sign is easy to check. A microphone with $\boldsymbol{u}^T \boldsymbol{r}_m > 0$ lies closer to the source than the reference point, by exactly that distance (amber in Fig. 2a). It hears the wave earlier, so its delay is negative.

At one frequency f, a delay τ_m is a phase factor $e^{-j2\pi f \tau_m}$. Delay-and-sum beamforming steered to a look direction $\boldsymbol{u}_0$ removes the phases that a source in $\boldsymbol{u}_0$ would cause and averages the N microphones. For a source in direction $\boldsymbol{u}$, the output relative to the source amplitude is the beampattern

$$B(\boldsymbol{u}_0) = \left| \frac{1}{N} \sum_{m=1}^{N} e^{\,j2\pi f\,(\boldsymbol{u}-\boldsymbol{u}_0)^T \boldsymbol{r}_m / c} \right| . \tag{2}$$

When $\boldsymbol{u}_0 = \boldsymbol{u}$, all N terms are 1 and $B = 1$. In every other look direction the terms have different phases and partly cancel. The beampattern is the image of a single point source, the point spread function of the array. Its main lobe tells how sharp the image is, and everything outside it leaks a strong echo into the rest of the image.

3 Spatial Aliasing

Take a line of M microphones along y, a spacing d apart, and a source at bearing θ in the horizontal plane (measured from x, positive to the left). Then $\boldsymbol{u} = [\cos\theta, \sin\theta, 0]^T$, and $\boldsymbol{u}^T \boldsymbol{r}_m = y_m \sin\theta$. Looking in direction θ_0, neighbouring microphones differ in phase by $\psi = 2\pi d(\sin\theta - \sin\theta_0)/\lambda$, with $\lambda = c/f$ the wavelength, and the sum in equation (2) becomes a geometric series,

$$B = \left| \frac{\sin(M\psi/2)}{M \sin(\psi/2)} \right| . \tag{3}$$

The main lobe sits at $\psi = 0$. But the same value $B = 1$ comes back at every $\psi = 2\pi k$, whenever d times $\sin\theta - \sin\theta_0$ is a whole number of wavelengths. There all microphones are in phase again, and the array cannot tell the look direction from the true one. These copies of the main lobe are the grating lobes, at

$$\sin\theta_0 = \sin\theta - k\,\frac{\lambda}{d}, \qquad k = \pm 1, \pm 2, \ldots \tag{4}$$

They stay out of the frontal half-space for every source direction only if $d < \lambda/2$.

Figure 2b shows the example from the lecture. Eight microphones listen at 40 kHz ($\lambda = 8.6$ mm) to a source at 20°. At half-wavelength spacing, the beampattern has one main lobe. At $d = 10$ mm, equation (4) with $k = 1$ gives $\sin\theta_0 = 0.34 - 0.86$, and the array shows a ghost at $-31°$ that is exactly as strong as the true source.

This is the sampling theorem again, now in space. A microphone samples the sound in time, and an array samples it in space, at the positions of its microphones. Table 1 lists the parts we need here, and Note 5 has the full comparison. The last row is the subject of the rest of this note. In time, sampling at irregular instants turns the alias of a tone into broadband noise. In space, irregular positions do the same to the grating lobes.

What does that mean for the eRTIS? Its call sweeps from 75 to 20 kHz, so the wavelength runs from 4.6 to 17.1 mm, and half a wavelength at the top of the band is 2.3 mm. Filling the 66 by 38 mm aperture at that spacing would take about 510 microphones. A regular 8×4 grid of 32 microphones over the same aperture has spacings of 9.5 and 12.5 mm. Looking straight ahead,

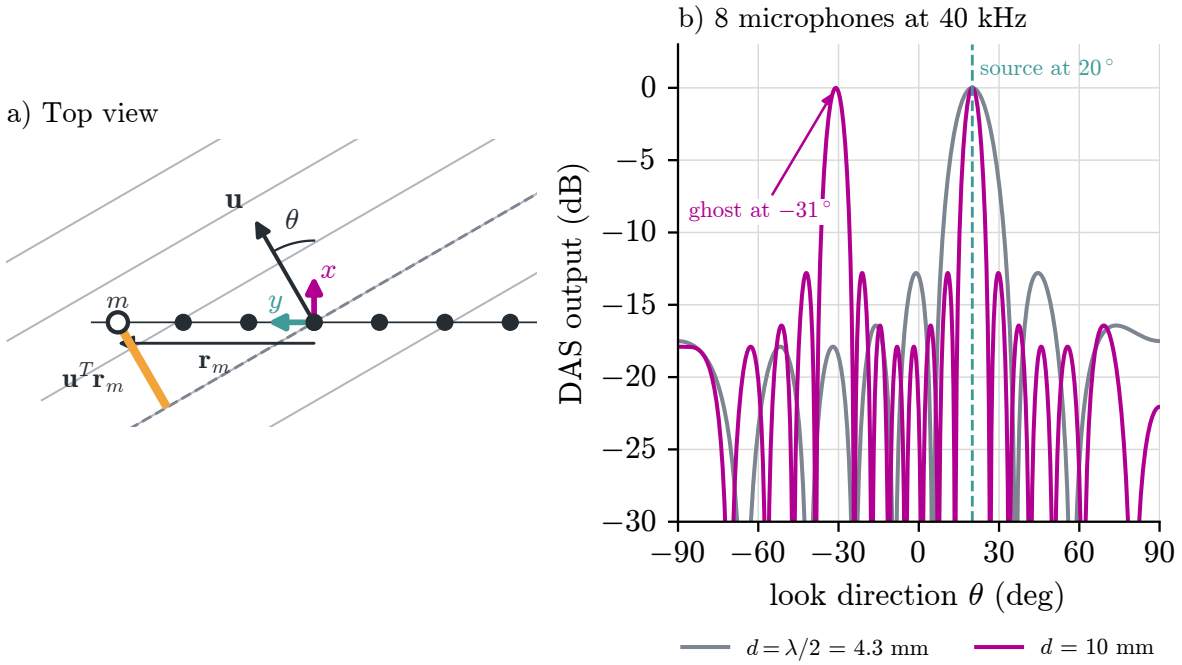


Figure 2: a) Top view of a line array, with the sensor drawn as a triad at the reference point (dot, x forward in magenta, y to the left in teal). The plane wave arrives from $\boldsymbol{u}$ (grey: wavefronts, dashed: the one through the reference point). Microphone m at $\boldsymbol{r}_m$ is closer to the source by $\boldsymbol{u}^T\boldsymbol{r}_m$ (amber). b) Delay-and-sum output of 8 microphones at 40 kHz for a source at 20°. At $d = 10$ mm, more than $\lambda/2$, a ghost as strong as the source appears at $-31°$.

Table 1: Sampling in time and sampling in space.

	time (one microphone)	space (the array)
sample spacing	$T_s = 1/f_s$	microphone spacing d
Nyquist	$f_s > 2f_{\max}$	$d < \lambda_{\min}/2$
aliasing	a tone at a wrong frequency	a grating lobe in a wrong direction
resolution	$\Delta f \approx 1/T$ (record length)	$\Delta\theta \approx \lambda/D$ (aperture D)
irregular sampling	aliases spread into a noise floor	grating lobes spread into a sidelobe floor

its grating lobes enter the frontal hemisphere above 27 kHz vertically and 36 kHz horizontally, and at 75 kHz they sit only 29° to the side and 21° up and down.

Key idea 1

An array samples the wavefield in space. A regular spacing larger than half a wavelength creates grating lobes, ghost copies of the main lobe that are as strong as the main lobe itself.

4 Three Ways to Place 32 Microphones

Figure 3 compares three layouts with the same 32 microphones. We draw the beampatterns as Lambert azimuthal equal-area (LAEAP) discs of the frontal hemisphere, seen from behind the sensor; Note 10, *Why LAEAP?*, explains why. The layouts are drawn from behind as well, so left is on the left in every panel.

The dense regular grid uses a spacing of 2.3 mm, half a wavelength at 75 kHz, and has no grating lobes anywhere in the call band. Its aperture is only 16 by 7 mm, though, and at 40 kHz

its main lobe is 24.1° wide. The sparse regular grid spreads the same microphones over the eRTIS aperture. Its main lobe shrinks to 5.8°, but at 40 kHz the grating lobes are already inside the hemisphere, at ±65° azimuth and ±43° elevation, at full height. The eRTIS layout has the same aperture and a main lobe almost as narrow (6.7°), because the width depends on the size of the aperture alone. Its grating lobes are gone. The energy they held is spread over the whole hemisphere as a speckled floor of sidelobes.

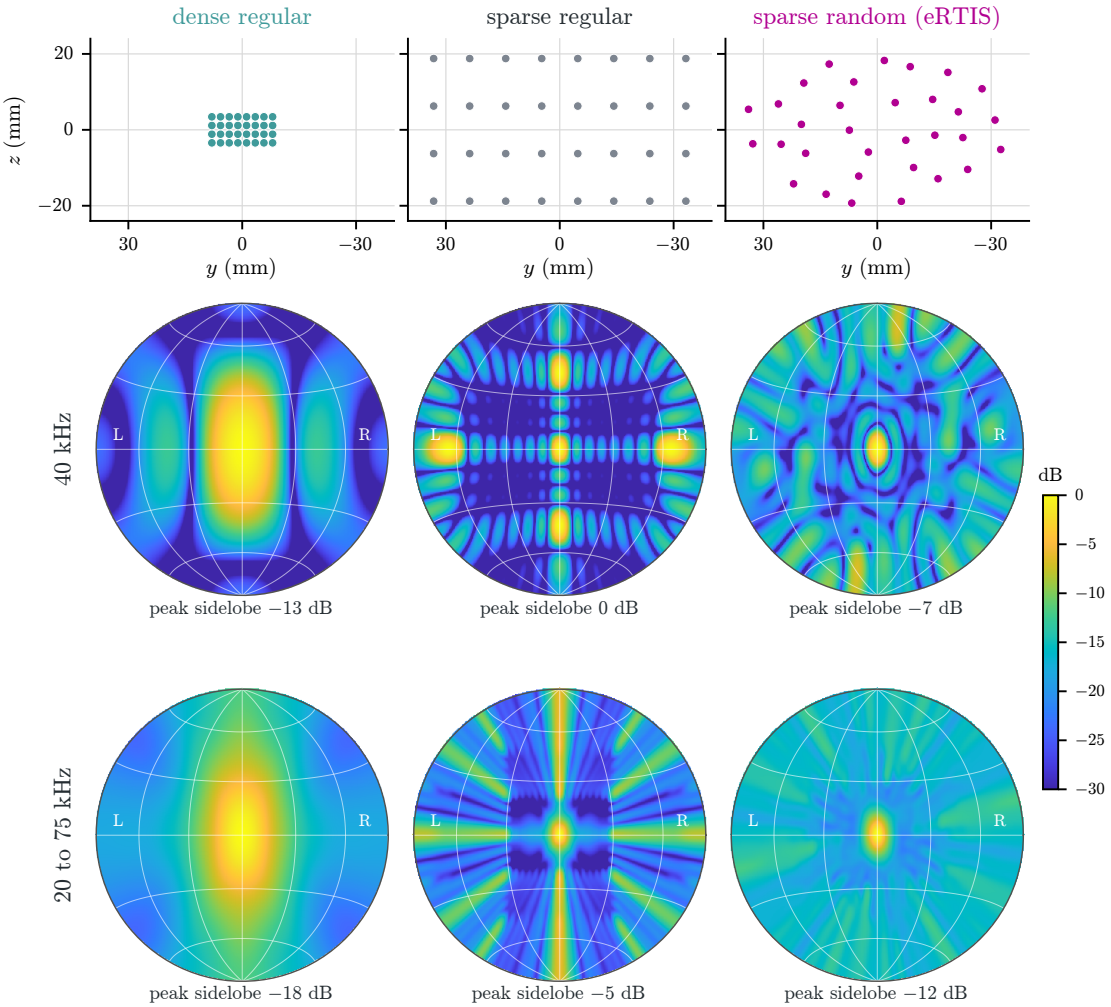


Figure 3: The same 32 microphones in three layouts (top, array plane seen from behind), and their delay-and-sum beampatterns straight ahead as LAEAP discs (grid lines every 30°, L and R mark the sensor's left and right). Middle row: one frequency, 40 kHz. Bottom row: the incoherent sum over the call band, 20 to 75 kHz. The dense grid has a wide main lobe, the sparse grid has grating lobes, and the eRTIS layout has a narrow main lobe on a sidelobe floor that the band smooths out.

How high is that floor? Away from the main lobe, the phases $\phi_m = 2\pi f(\boldsymbol{u} - \boldsymbol{u}_0)^T \boldsymbol{r}_m / c$ of a random layout span many cycles and behave like independent random numbers. The expected power of the sum in equation (2) is then

$$\mathrm{E}\big[B^2\big] = \frac{1}{N^2} \sum_{m=1}^{N} \sum_{n=1}^{N} \mathrm{E}\big[e^{j(\phi_m - \phi_n)}\big] = \frac{1}{N^2}\, N = \frac{1}{N}, \tag{5}$$

because only the N terms with $m = n$ survive the average. The phasors add up in power instead of in amplitude. For 32 microphones the floor sits at −15 dB, and the mean sidelobe level of the eRTIS layout at 40 kHz is −16 dB. Single speckles reach higher. At one frequency, the highest sidelobe of the eRTIS layout lies between −13 and −5 dB over the call band (Fig. 4a).

The eRTIS layout is not purely random. Its microphones keep a minimum distance from each other, here 6.8 mm. That keeps them from clustering, which would waste aperture, and the randomness of the positions removes the periodicity that grating lobes need.

5 Bandwidth Rescues the Random Array

A sonar call is broadband. The energy of a beamformed echo, summed over time, is by Parseval's theorem the sum over frequency of $|S(f)|^2 B^2(\boldsymbol{u}_0, f)$, with $S(f)$ the spectrum of the echo. The broadband image is therefore an incoherent sum of the narrowband beampatterns over the band, and the two sparse layouts react to that sum very differently.

The grating lobes of the regular grid move with frequency, following equation (4). Over the band they smear into ridges, but every frequency still contributes a full-height lobe somewhere, and the ridges survive at −5 dB (Fig. 3, bottom). The speckles of the random layout change place from one frequency to the next, so the sum averages them down. Its broadband peak sidelobe drops to −12 dB on a smooth floor of about −16 dB. Figure 4b shows how this improves with the fractional bandwidth, the bandwidth divided by the centre frequency. The full eRTIS sweep has a fractional bandwidth of 116%. The default sweep of the eRTIS software, 25 to 50 kHz, still brings the random layout to −11 dB, against −4 dB for the sparse grid. For a pulse-echo sonar, this bandwidth comes for free with the call.

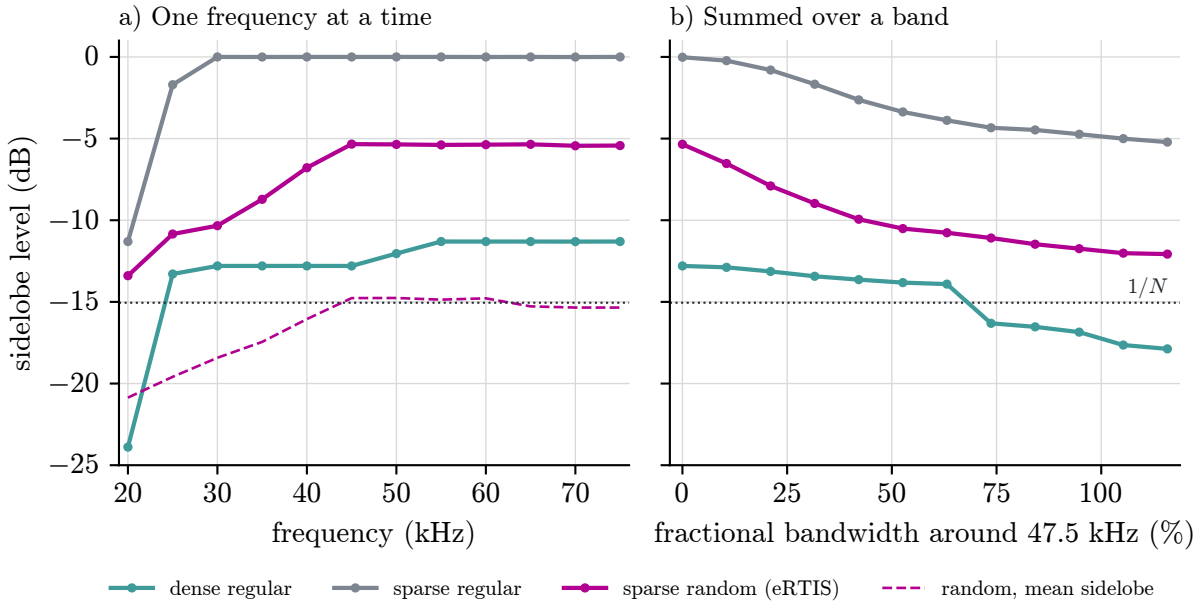


Figure 4: Peak sidelobe levels over the whole frontal hemisphere, steered straight ahead. a) At one frequency at a time, with the mean sidelobe level of the eRTIS layout (dashed) and the $1/N$ level for 32 microphones (dotted). b) For the incoherent sum over a band centred on 47.5 kHz, against its fractional bandwidth; the right end is the full call band, 20 to 75 kHz.

What does this mean for an image? Sidelobes only compete between echoes that arrive at the same time, because the matched filter already separates echoes in range. Take a wall at 2 m straight ahead and a small target in the same range cell, 25° to the side, with an echo 15 dB weaker than that of the wall. With the sparse grid, the grating ridges of the wall reach −5 dB, 10 dB above the target, so wherever a ridge crosses the target's direction, the wall paints a ghost stronger than the target. With the eRTIS layout, the floor of the wall lies near −16 dB and its highest speckle at −12 dB. The target stands out against the floor, but a speckle can come close. As a rule, delay-and-sum on this array shows an echo reliably when it is stronger than the highest broadband sidelobe of the strongest echo in its range cell.

Key idea 2

A random layout trades grating lobes for a floor of weak sidelobes near $1/N$. Grating lobes only smear over a band, while the random sidelobes average down, so a broadband sonar gets a large aperture from few microphones.

6 Regular or Random?

Random layouts are not better in every respect. Table 2 sums up the trade-off. A random array buys aperture with bandwidth, while a regular array buys structure with channel count. That structure matters for fast beamforming, where a regular grid allows an FFT, and for the high-resolution methods of Note 5. MVDR and MUSIC need spatial smoothing to cope with the coherent echoes of a single ping, and spatial smoothing averages over shifted, identical copies of a subarray. A random layout has no such copies, so it rules smoothing out. Note 5 shows this on the eRTIS layout, together with sparse reconstruction, which works on any layout. Nonlinear beamformers such as delay-multiply-and-sum (Note 6, *Multiplying Instead of Adding*) are another way to push the sidelobe floor down.

Table 2: Regular or random? The trade-off, after the CPS lecture *Array Signal Processing*.

	regular grid	sparse random
spacing rule	$d \leq \lambda_{\min}/2$, or grating lobes	none, mean spacing $\gg \lambda/2$ works
aperture for N microphones	grows with $\sqrt{N}$ at fixed spacing	set by the board
sidelobes	structured, -13 dB for uniform weights	noise-like floor near $1/N$, lower with bandwidth
narrowband use	yes	poor
fast beamforming	FFT, separable	full table of delays
spatial smoothing	yes (shifted subarrays)	no
our sensors	HiRIS, μRTIS	eRTIS

Random arrays have a history in our lab. Steckel and Peremans (2014) used a random, sparse microphone array to find air leaks in 2D from their broadband ultrasound, with an algorithm inspired by compressed sensing. From the raw signals they first built an intermediate acoustic image, the energyscape, and then found several leaks of different strengths at once, including weak leaks next to a strong one. The eRTIS uses the same idea for 3D in-air sonar. A more recent paper attacks spatial aliasing from the hardware side, with 3D-printed waveguides for broadband MEMS arrays that beat spatial aliasing up to 95 kHz.

Key idea 3

Choose the layout and the beamformer together. Sparse random layouts suit broadband delay-and-sum and sparse reconstruction. Dense regular grids suit FFT beamforming and the subspace methods that need spatial smoothing.

Exercise

1. For the line of Fig. 2b with $d = 10$ mm, above which frequency does a grating lobe enter the frontal half-space when you look straight ahead? And when you look 30° to the left?
2. How many microphones would a half-wavelength grid at 75 kHz need over the eRTIS aperture? What data rate would that give at 4.5 Mbit/s per PDM microphone?
3. Derive the $1/N$ floor yourself. Why does it need the phases to span many cycles, and why does it fail near the main lobe?
4. What would go wrong with a purely random layout without a minimum distance between microphones?
5. In the wall example, how much weaker may the target be with 128 microphones, if the floor follows $1/N$?

Further Reading

- Y. T. Lo, "A mathematical theory of antenna arrays with randomly spaced elements," IEEE Trans. Antennas Propag., 12(3), 257–268, 1964.
- H. L. Van Trees, *Optimum Array Processing*, Wiley, 2002.
- J. Steckel and H. Peremans, "Ultrasound-based air leak detection using a random microphone array and sparse representations," IEEE SENSORS, 1026–1029, 2014.
- R. Kerstens, D. Laurijssen and J. Steckel, "eRTIS: A fully embedded real time 3D imaging sonar sensor for robotic applications," IEEE ICRA, 2019.
- 3D-printed waveguides for broadband MEMS arrays, IEEE Xplore, https://ieeexplore.ieee.org/document/10784819.
- In this series: Note 5, *Beyond Delay-and-Sum* (signal model, MVDR, MUSIC, spatial smoothing, sparse reconstruction); Note 10, *Why LAEAP?* (the map projection); Note 6, *Multiplying Instead of Adding* (DMAS).