PERFORMANCE TESTS OF SPECTRAL KURTOSIS-BASED RFI REAL-TIME DETECTION AND FLAGGING STRATEGIES FOR MULTIPLE-ANTENNA SYSTEMS WITH THE EXPANDED OWENS VALLEY SOLAR ARRAY

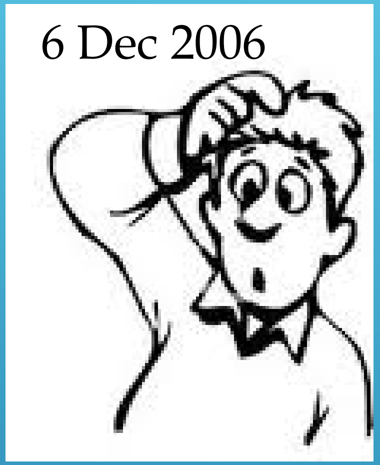
Gelu M. Nita*(1) , Dale E. Gary(1) , and Gregory Hellbourg(2)

(1) New Jersey Institute of Technology, Newark, NJ, 08536, USA

(2) California Institute of Technology, Pasadena, CA, USA

RFI2022, 14- 18 Feb. 2022 , Virtual Event hosted by ECMWF, Reading, UK

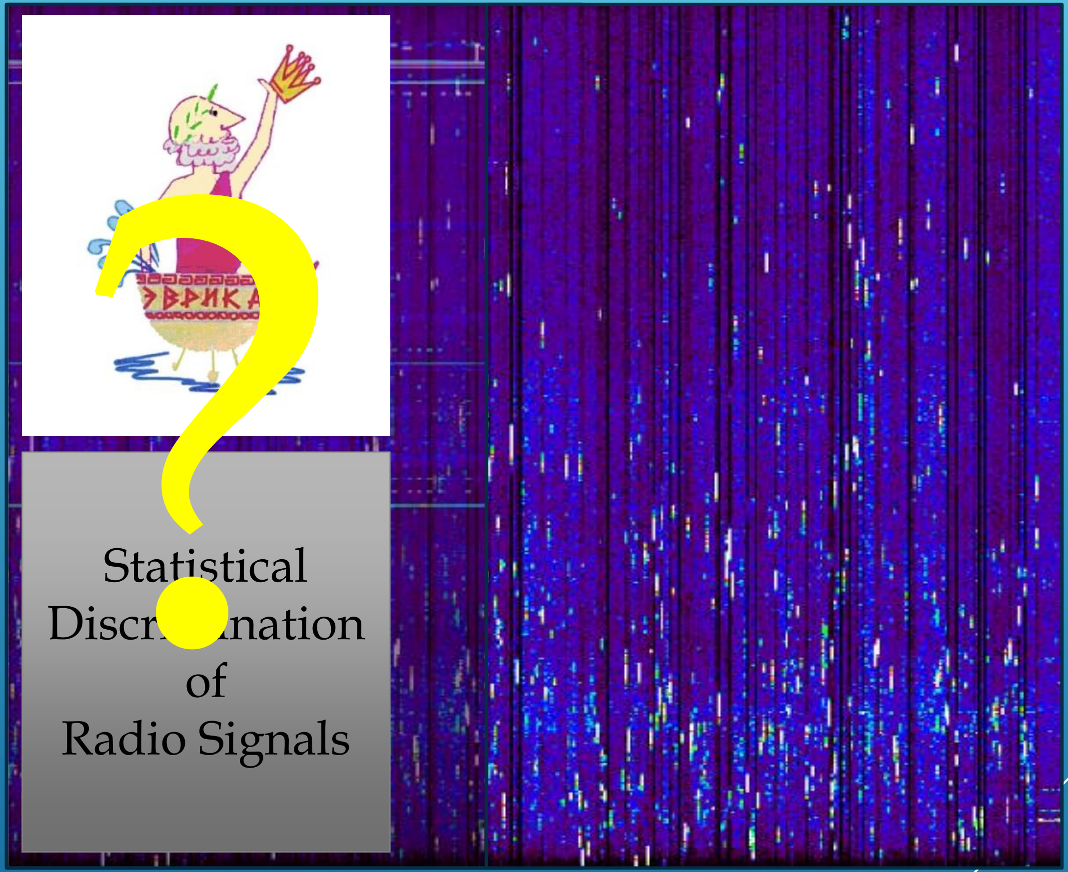
A TALE ABOUT A WELL-KNOWN RADIO DATA ANALYSIS CHALLENGE



AND ITS NOT SO WIDELY-KNOWN STATISTICAL SOLUTION…

Radio Frequency Interference

Astrophysical Signal



TIME DOMAIN (Full Bandwidth)		FREQUENCY DOMAIN (Sub-bands)	
Statistical Distribution	Signal	Signal	Statistical Distribution
Gaussian	Real Voltages: x_i	$FFT(x_{0\ldots N}) = \{R, I\}_{1\ldots N/2}$	{Gaussian, Gaussian}
ChiSqr $\chi^2_{\nu=1} \stackrel{\text{def}}{=} \Gamma_{d=1/2}$	Powers: $P_i = x_i^2$	$P_k = \{R^2 + I^2\}_{k=1\cdots N/2}$	Exponential $\chi^2_{\nu=2} \stackrel{\text{def}}{=} \Gamma_{d=1}$
Gamma $\Gamma_{d=M/2}$	$S_1 = \sum_{i=1}^{M} P_i$	$S_1(k) = \sum_{i=1}^{M} P_k$	Gamma $\Gamma_{d=M}$

Distribution	Kurtosis
	$K[X] = \frac{E[(X-\mu)^4]}{(E[(X-\mu)^2])^2}$
Gaussian	$E[K] = 3$
Gamma Γ_d	$E[K] = 3 + \frac{6}{d}$

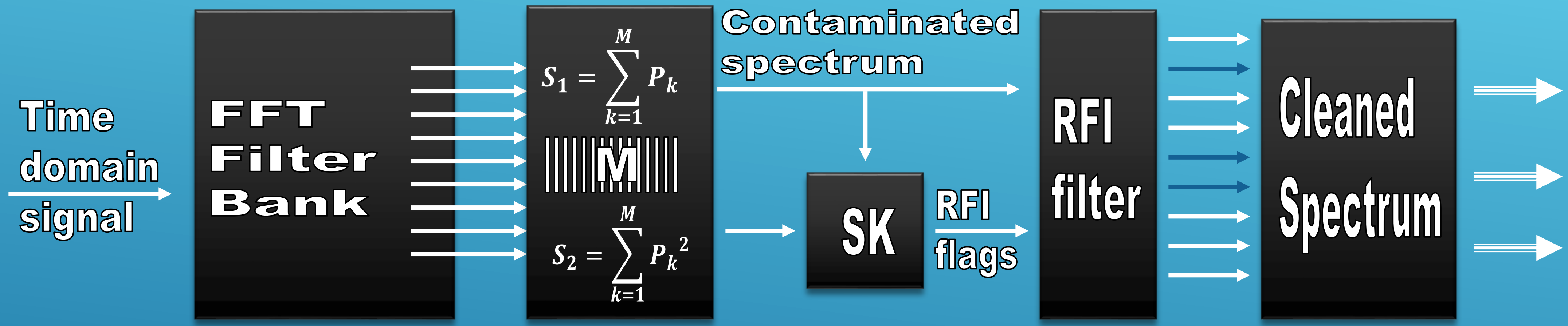
Spectral Variability
$SV[X] = \frac{E[(X-\mu)^2]}{\mu^2}$
$E[SV] = \sigma^2/\mu^2$
$E[SV] = 1/d$

SPECTRAL KURTOSIS
Gamma Γ_d
$SK \stackrel{\text{def}}{=} d \times SV[X]$
$E[SK] = 1$

SIGNAL STATISTICS 101: KURTOSIS VS SPECTRAL KURTOSIS

THE SPECTRAL KURTOSIS SPECTROMETER

NITA ET AL. 2007 PASP, 119, 805
GARY, LIU & NITA 2010 PASP, 122, 560



The unbiased Spectral Kurtosis Estimator

$$SK \equiv \frac{M+1}{M-1}\left(\frac{MS_2}{S_1^2} - 1\right)$$

$$E(SK) = 1$$

$$\sigma^2(SK) \approx \frac{4}{M}$$

THE GENERALIZED SPECTRAL KURTOSIS ESTIMATOR

Nita & Gary 2010, MNRAS 406 L60-L64

Theorem: Given that, for a particular signal, the set of its power estimates P_k obeys a gamma distribution characterized by the shape parameter ***d***, the infinite series of statistical moments $MS_2/S_1{}^2$, were $S_1 = \sum_{k=1}^{M} P_k$ and $S_2 = \sum_{k=1}^{M} P_k{}^2$, is given by:

$$E\left[\left(\frac{MS_2}{S_1^2}\right)^n\right] = \frac{M^n\Gamma(Md)}{\Gamma(d)^M\Gamma(Md+2n)} \times \frac{\partial^n}{\partial t^n}\left[\sum_{r=0}^{n}\frac{1}{r!}\Gamma(2r+d)t^r\right]^M\Bigg|_{t=0}$$

Corollary: The Generalized Spectral Kurtosis Estimator defined by

$$SK = \frac{Md+1}{M-1}\left(\frac{MS_2}{S_1{}^2} - 1\right)$$

- Has an unbiased unity expectation $E[SK] = 1$, independent of the integrated power S_1
- The infinite series of statistical moments of its PDF are analytically defined only in terms of M and d

The SK estimator is well suited for detecting mixed signals not obeying the same gamma probability distribution: Detection thresholds characterized by analytically defined probabilities of false alarm (PFA)

PEARSON TYPE IV PDF INVOLVING THE FIRST FOUR EXACT SK MOMENTS

Pearson Type IV PDF

(Pearson 1895; Nagahara 1999)

$$p(x)=\frac{1}{a\sqrt{\pi}}\frac{\Gamma\left(m+i\frac{\nu}{2}\right)\Gamma\left(m-i\frac{\nu}{2}\right)}{\Gamma\left(m-\frac{1}{2}\right)\Gamma(m)}\left[1+\left(\frac{x-\lambda}{a}\right)^2\right]Exp\left[-\nu\,ArcTan\left(\frac{x-\lambda}{a}\right)\right]$$

(Heinrich 2004)

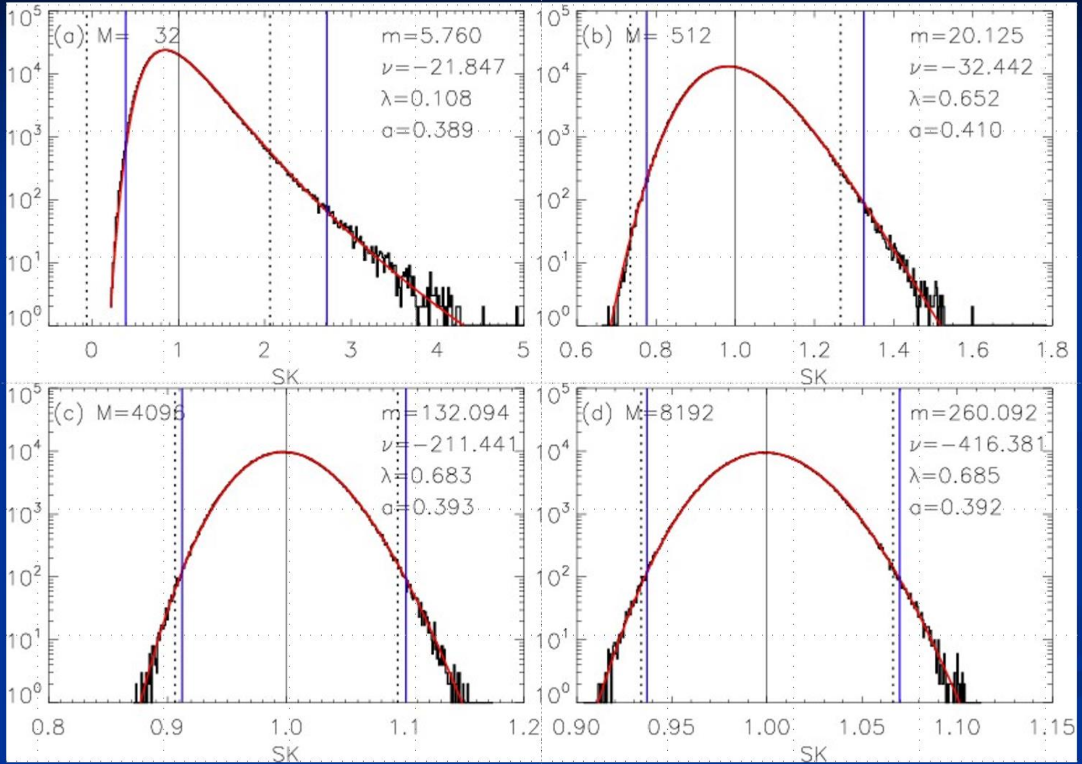
$$\nu=-\frac{r(r-2)\sqrt{\beta_1}}{\sqrt{16(r-1)-\beta_1(r-2)^2}}$$

$$r=\frac{6(\beta_2-\beta_1-1)}{2\beta_2-3\beta_1-6}\qquad a=\frac{1}{4}\sqrt{\mu_2\left[6(r-1)-\beta_1(r-2)^2\right]}$$

$$m=\frac{r+2}{2}\qquad \lambda=\mu_1'-\frac{1}{4}(r-2)\sqrt{\mu_2\beta_1}$$

March 29, 2010 RFI Mitigation Workshop, Groningen The Netherlands 15

Pearson IV PDF vs. Monte Carlo Simulations



March 29, 2010 RFI Mitigation Workshop, Groningen The Netherlands 18

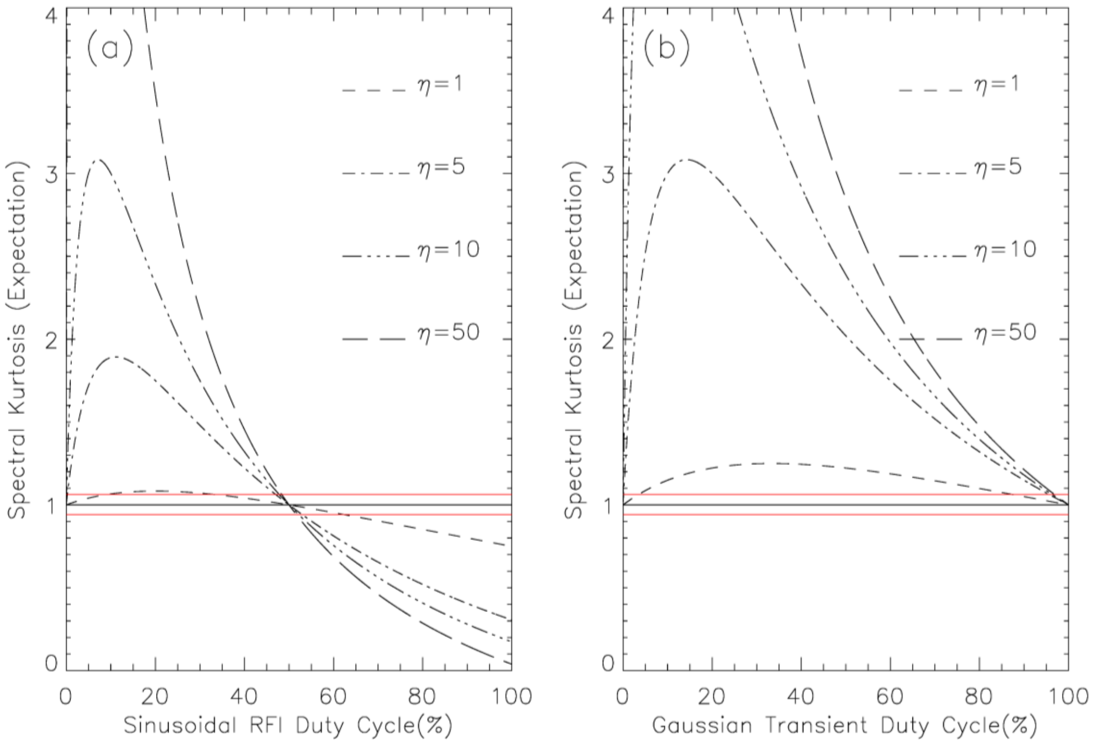
IDL code for computing GSK thresholds may be found at: https://github.com/Gelu-Nita/GSK

PRACTICAL CASES WELL SUITED FOR SK ANALYSIS

- Raw power estimates based on time domain real signals
  - Gamma distribution of shape factor $d=0.5$ (Chi-Square distribution)
- Raw power estimates based on time or frequency domain complex signals
  - Gamma distribution of shape factor $d=1$ (Exponential distribution)
- Accumulations of N raw power estimates of shape factor δ
  - Gamma distribution of shape factor $d=N\delta$
- Power estimates based on quantized time domain signals or quantized frequency domain power estimates (Nita, Gary, and Hellbourg 2017, IEEE; Nita, Keimpema, & Paragi 2019, Journal of astronomical instrumentation)
  - Gamma distribution having an instrument-dependent shape factor d

SK DEPENDENCE ON THE INTEGRATION–RELATIVE DUTY-CYCLE
RFI AND GAUSSIAN TRANSIENT SIGNALS
(NITA ET AL. 2007, PASP, 119; NITA 2016, MNRAS, 458)

The SK values associated with transient RFI may lay above or below unity as the transient duty-cycle relative to the integration time is smaller or larger than 50%. **The SK estimator is blind to 50% RFI duty-cycle!**

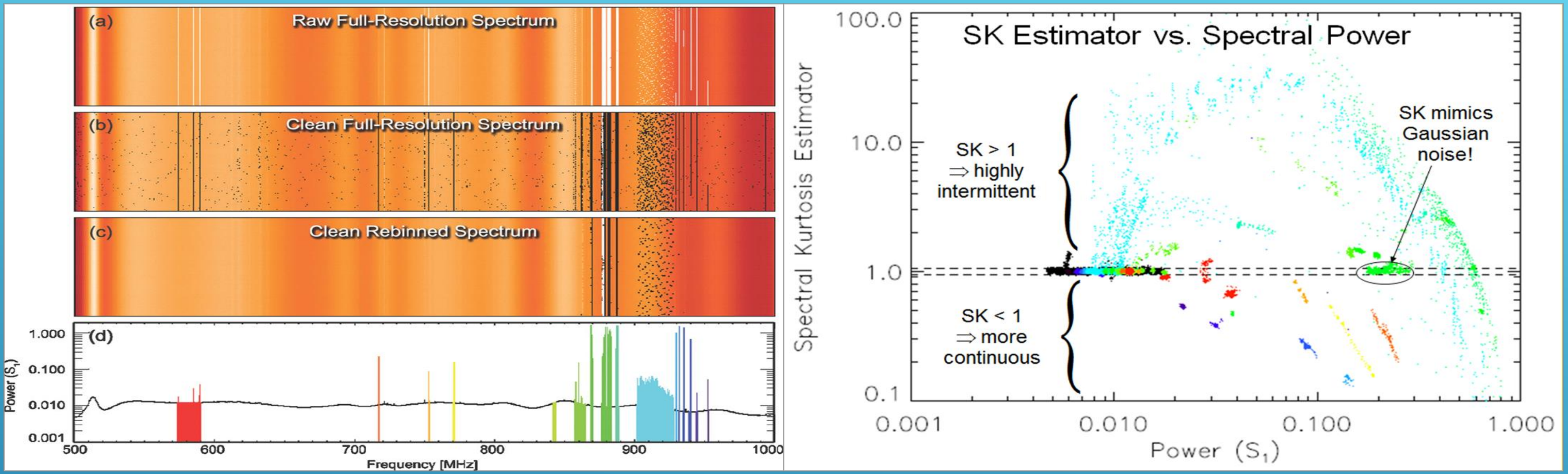


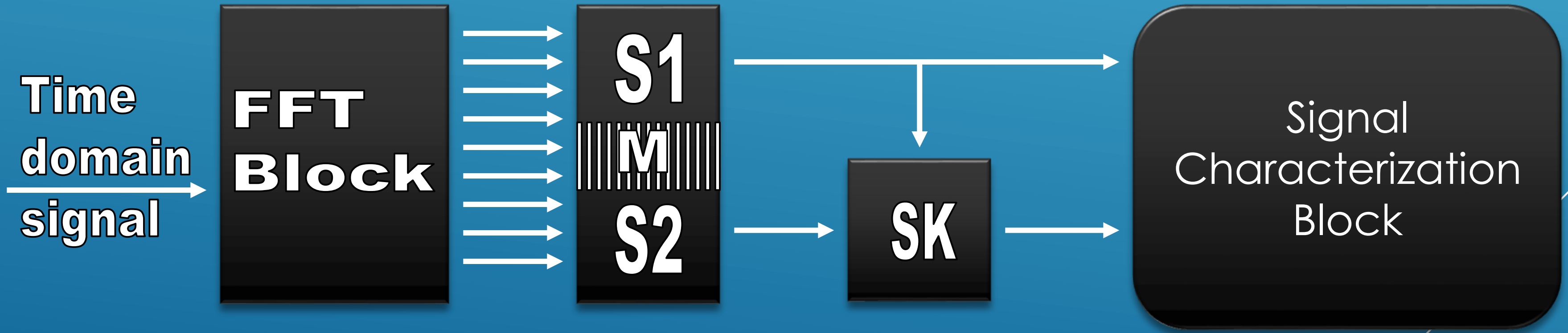
Gaussian transients have SK values larger than unity for any duty-cycle!

For both types of transients, the SNR-dependent SK estimator sensitivity peaks somewhere below 50% duty-cycle

SPECTRAL KURTOSIS: A POWERFUL SIGNAL CLASIFICATION TOOL

EOVSA Testbed Single Antenna Dynamic Spectrum





THEORETICAL BACKGROUND

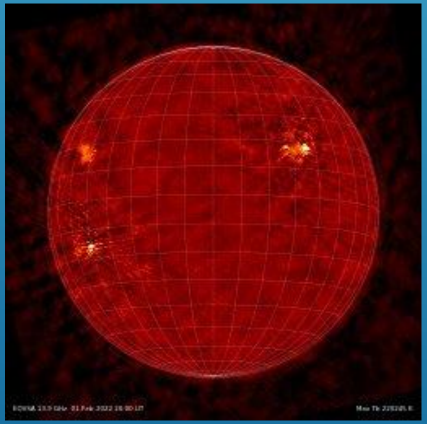
Scan me

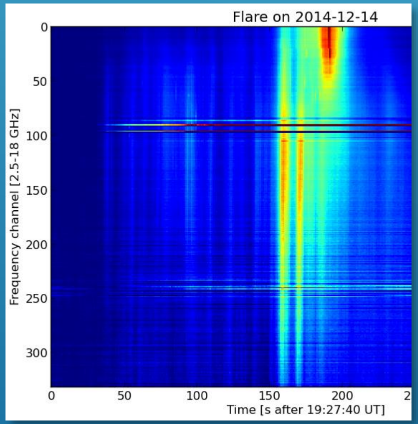
A series of 10 papers on Spectral Kurtosis Theory and Applications

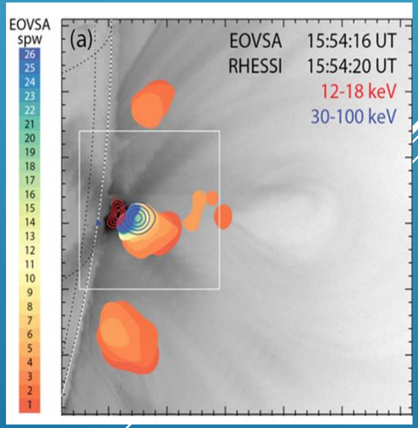
EXPANDED OWENS VALLEY SOLAR ARRAY

World-first frequency agile interferometer equipped with a hardware embedded SK real-time computation engine

Table 1: EOVSA Specifications	
Frequency range	1 – 18 GHz
Number of data channels/antenna	2 (dual polarization)
IF bandwidth	500 MHz single sideband
Frequency resolution	4096 spectral channels per 600 MHz band) 500 science channels variable ~1-40 MHz
Time resolution	Sample time: 20 ms Full Sweep: 1 s
Polarization	Full Stokes (IQUV)
Number correlator inputs per poln	16
Number and type of antennas	Thirteen 2.1-m One 27-m equatorial (cal. only)
System Temperature	570 K (2 m); 35 K (27 m)
Baselines for imaging	78
Angular resolution	$56/n_{GHz} \times 51/n_{GHz}$ arcsec
Array size	1.1 km EW x 1.2 km NS







The EOVSA correlator outputs integrated power and squared power for all 14 antennas and R and L circular polarizations with 20ms-0.125MHz time-frequency resolution

EOVSA CORRELATOR HIGH BIT RESOLUTION POWER AND SQUARED POWER OUTPUTS

NITA, HICKISH, MACMAHON, AND GARY 2016, J. ASTRONOMICAL INSTRUMENTATION 5(4)

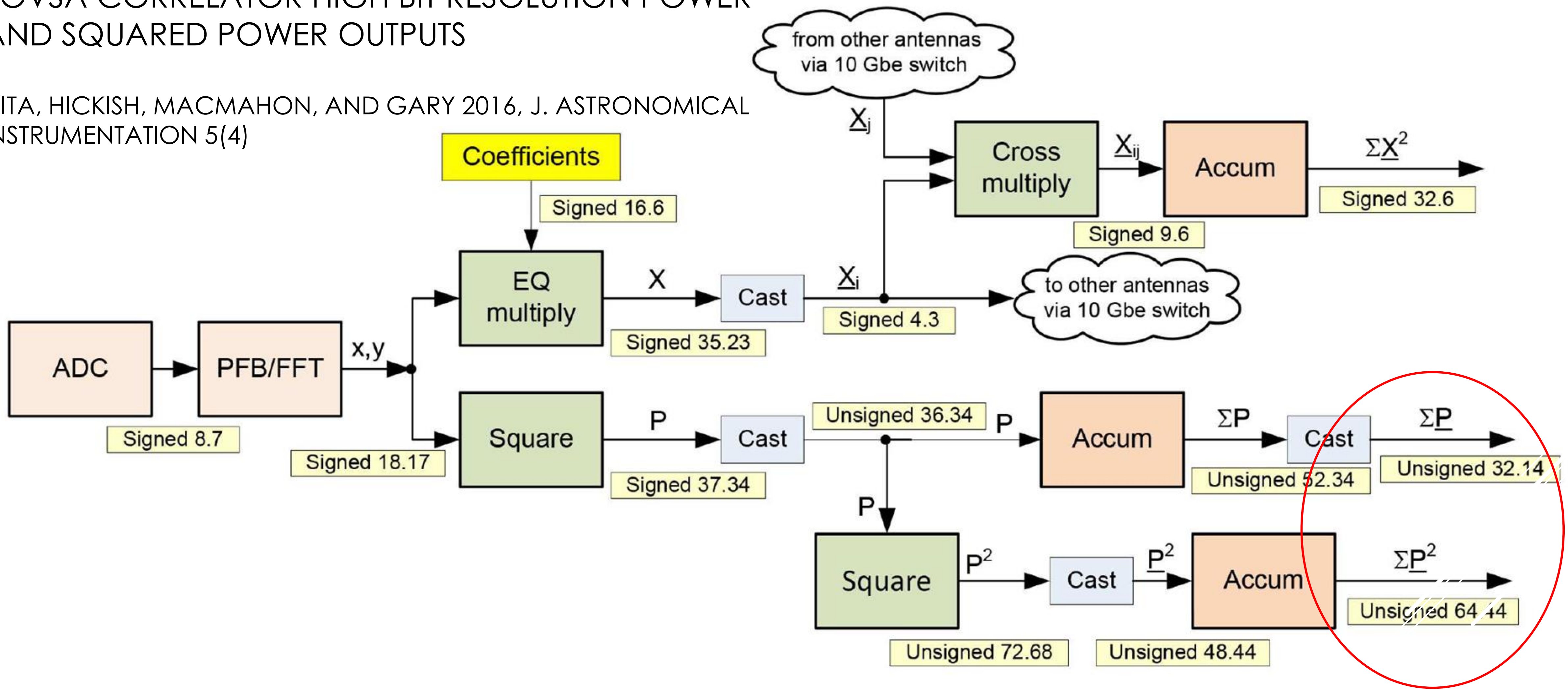
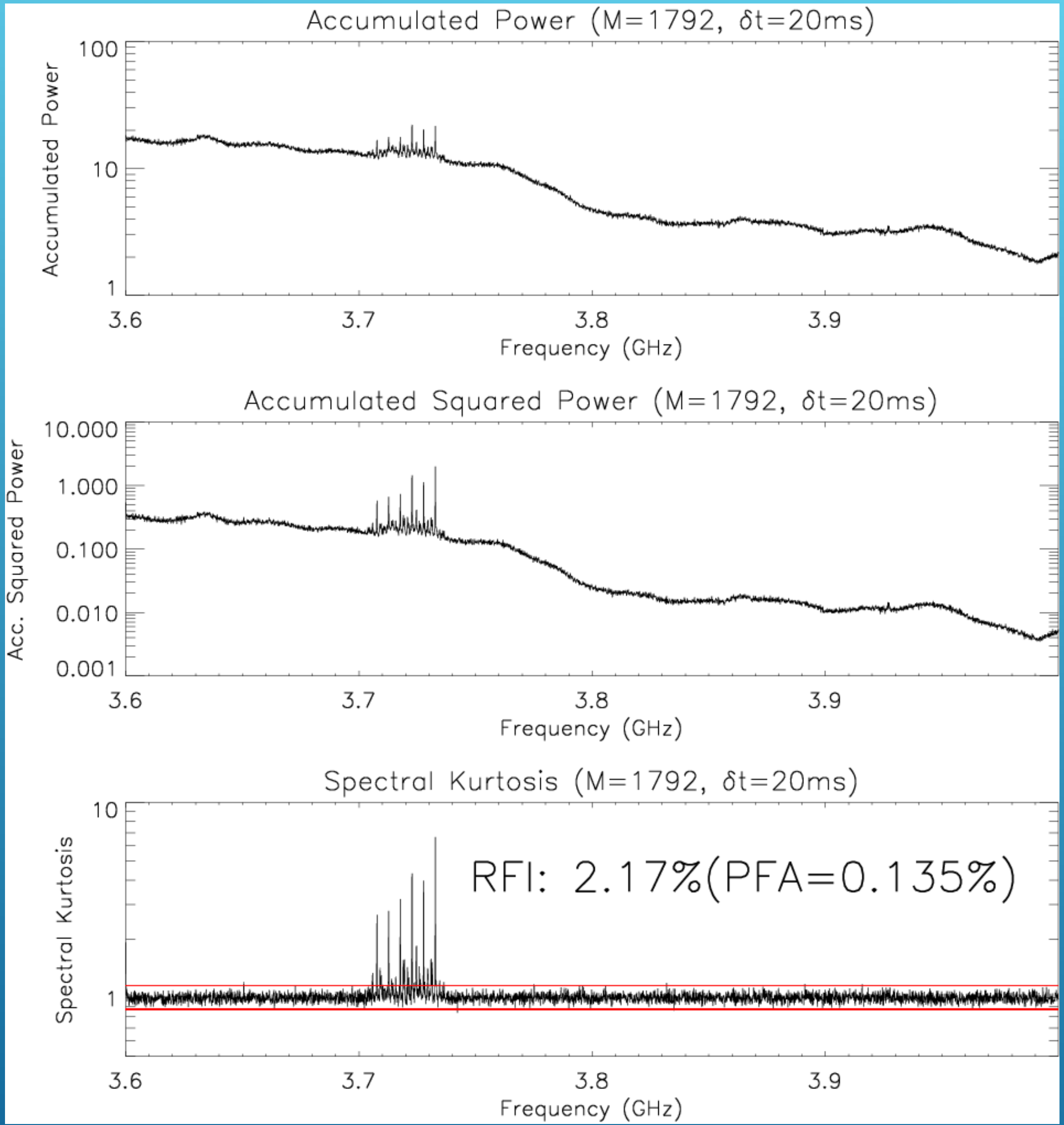


Fig. 1. *Block diagram of EOVSA correlator. Quantities with no bit-truncation are shown without an underbar, while quantities with potential truncation are shown with an underbar. Truncation is performed via the "Cast" blocks, which change the output bit-width for practical purposes of data volume and compatibility with standard computational data types, e.g. 32-bit data for accumulated power* ($\sum \underline{P}$) *and cross-power* ($\sum \underline{X}^2$), *and 64-bit data for power-squared* ($\sum \underline{P}^2$).

SPECTRAL KURTOSIS RFI FLAGING OF SINGLE-ANTENNA AUTO-CORRELATION SPECTRA

Expanded Owens Valley Solar Array(EOVSA) Implementation and Performance Tests

EOVSA SK RFI DETECTION EXAMPLE



$$S_1 = \sum_{k=1}^{M} P_k$$

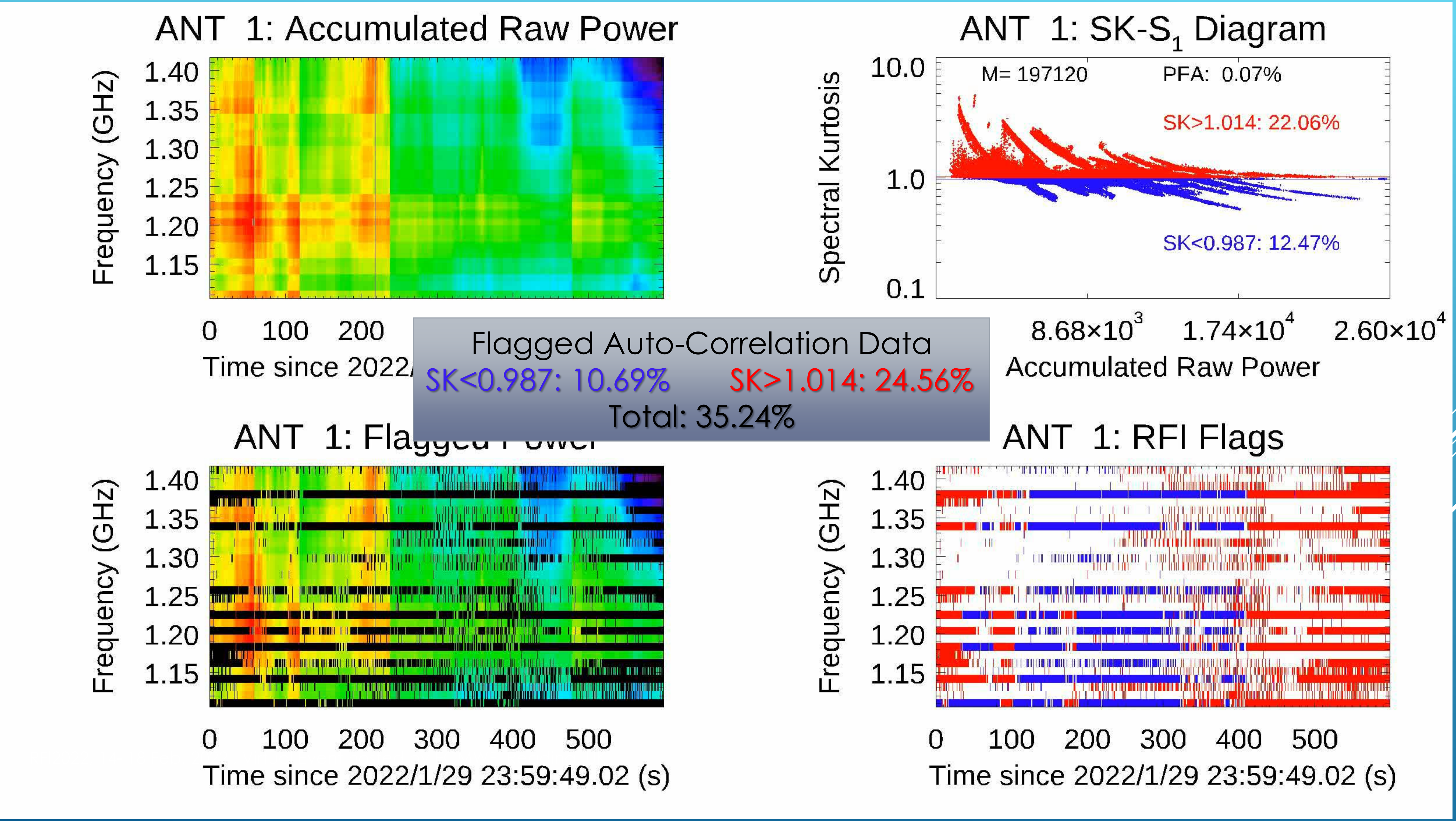
$$S_2 = \sum_{k=1}^{M} {P_k}^2$$

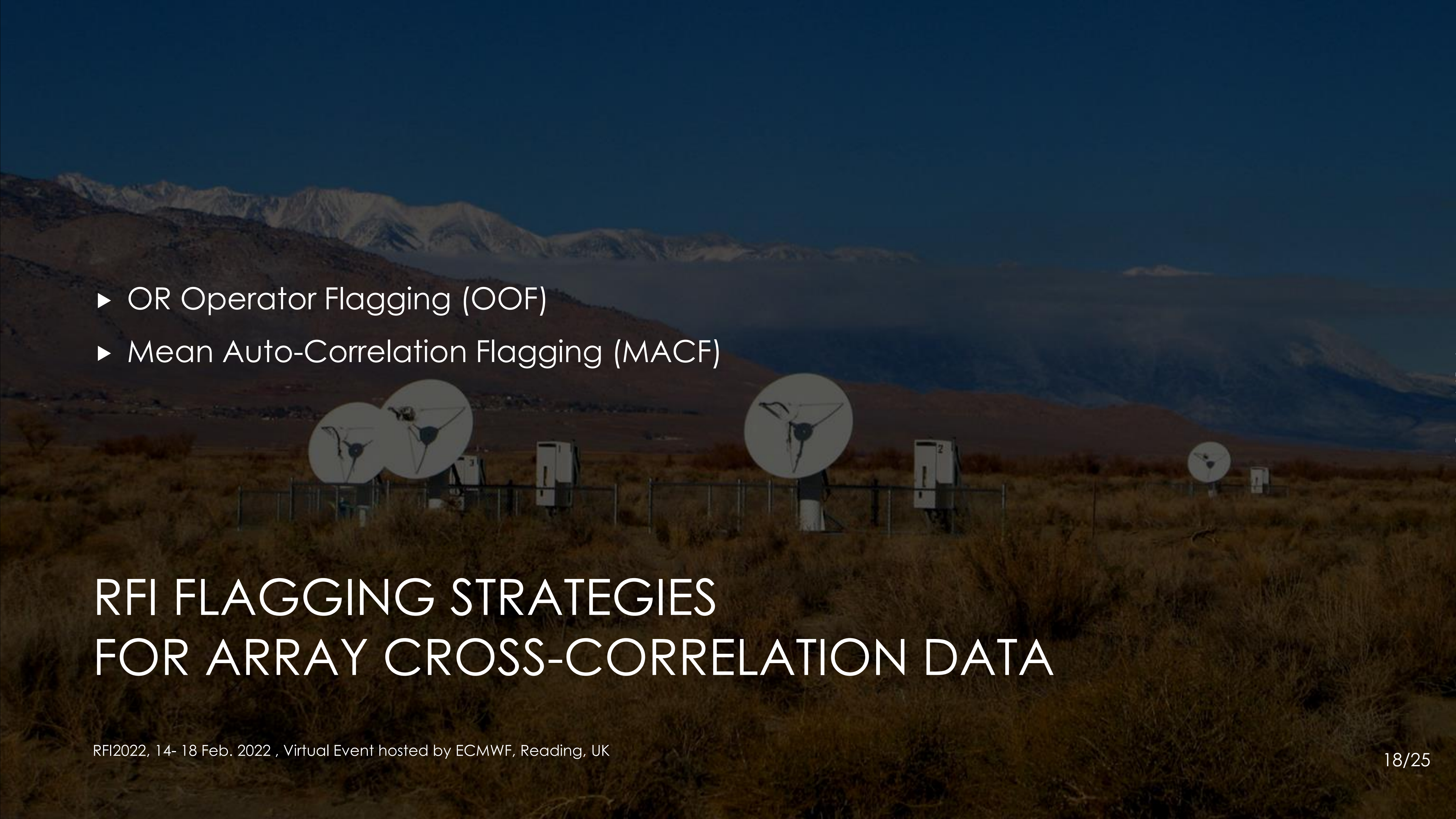
$$SK = \frac{M+1}{M-1}\left(\frac{MS_2}{{S_1}^2} - 1\right)$$

EOVSA FREQUENCY AND TIME RESOLUTIONS

- EOVSA correlator raw data
  - 4096 spectral channels per 600 MHz band, 20ms, M=1792 accumulation length
- Normal Mode of Operation:
  - 1-18 GHz RF band swept in 1 second
  - 500 science bands having variable 1-40MHz spectral width
- High Time Resolution Mode (recently implemented, used for this presentation)
  - 1.01-1.42GHz RF bandwith
  - 30 science bands having 10.74MHz frequency resolution
  - 20ms time resolution
  - M=197120=1792x110 accumulation length (combined time and frequency integration)

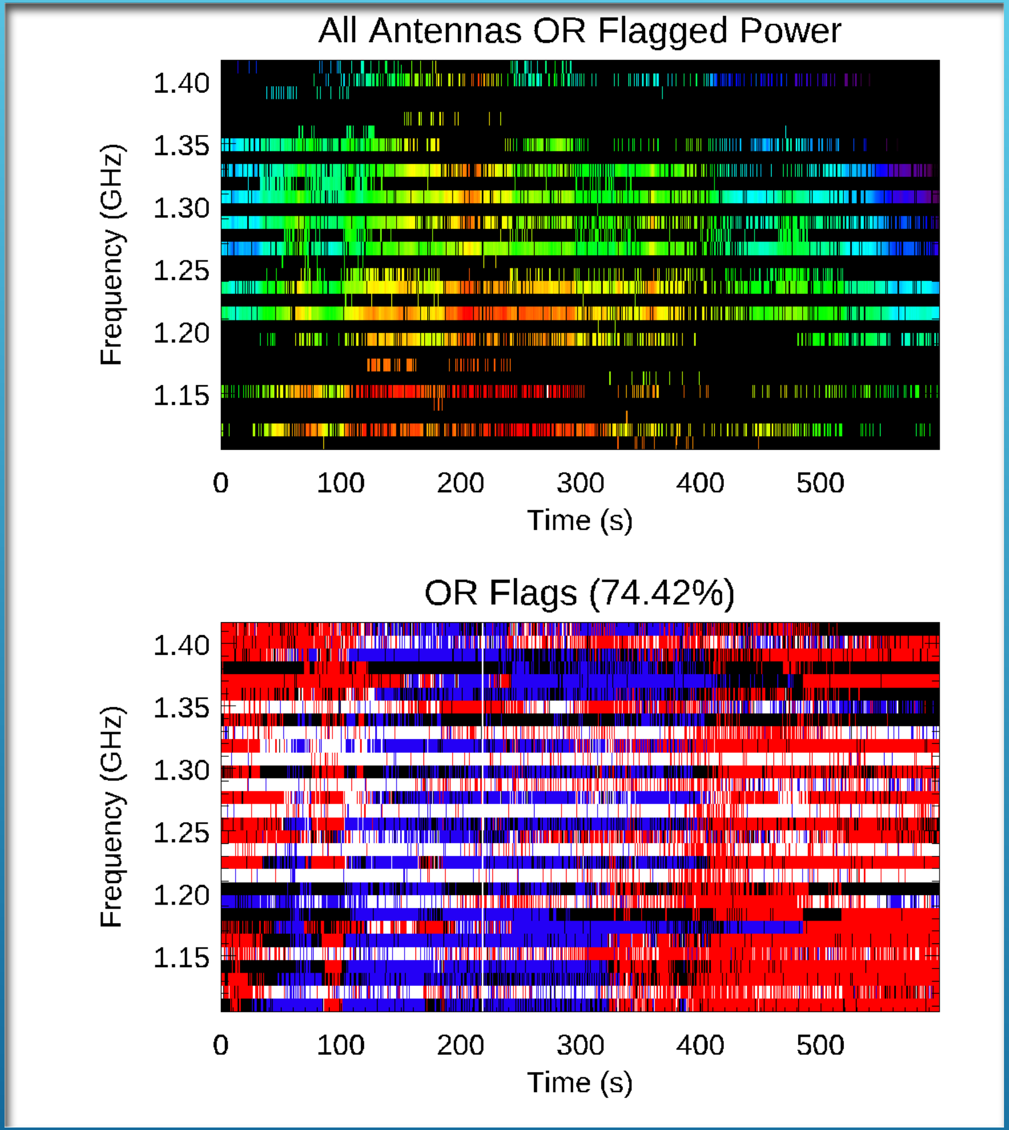
EOVSA SOLAR RADIO BURST OBSERVATION: ANTENNA SK FLAGS





- OR Operator Flagging (OOF)
- Mean Auto-Correlation Flagging (MACF)

RFI FLAGGING STRATEGIES FOR ARRAY CROSS-CORRELATION DATA



- Using array-wide OR autocorrelation flagging based on the autocorrelation RFI flags generated for each antenna may result in excessive data loss (74% in this example), while some baselines may not be affected by local interference.
- However, the OR flagging approach may be implemented for each individual baseline, case in in which the cross-correlation data is flagged only where at least one of the autocorrelation SK flags is set.

A CAUTIONARY NOTE ON ARRAY-WIDE OR AUTOCORRELATION FLAGGING

BASELINE OR OPERATOR FLAGGING (OOF)

NITA AND HELLBOURG, URSI GASS 2020

This strategy flags antenna cross-correlations depending on the OR operation between the flags evaluated on the independently computed antenna auto-correlation Spectral Kurtosis RFI flags.

For any given pair of antennas, the probability of false alarm RFI flagging, PFA_{OR}, may be computed in terms of the individual antenna PFA_{auto} as

$$PFA_{OR} = (2 - PFA_{auto})PFA_{auto}$$

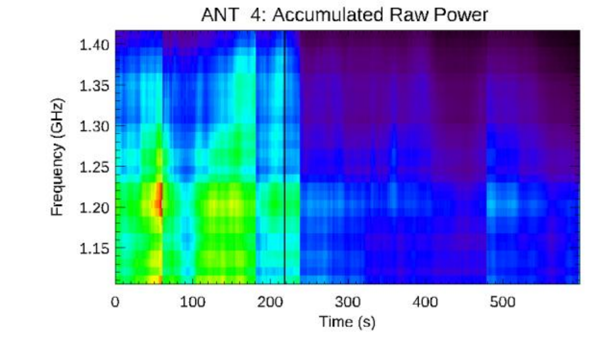
where PFA_{auto} may be computed using one of the analytical formulae provided by Nita & Gary 2010, MNRAS, 406, L60-L64

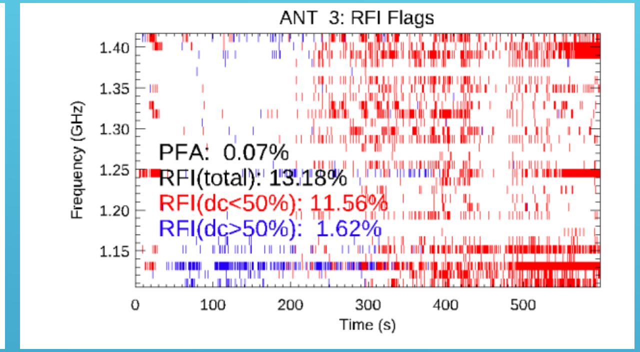
Hence, to achieve a desired PFA_{OR} for a given baseline, the individual antennas auto-correlation flags should be computed for

$$PFA_{auto} = 1 - \sqrt{1 - PFA_{OR}}$$

$$PFA_{OR} = 0.135\% \xrightarrow{requires} PFA_{auto} = 0.07\%$$

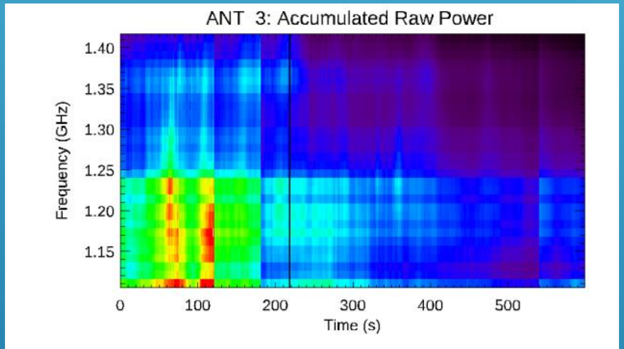
OOF CROSS-CORRELATION FLAGGING EXAMPLE

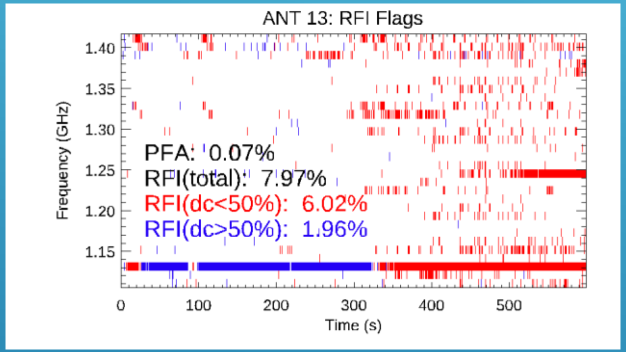




Antenna 4

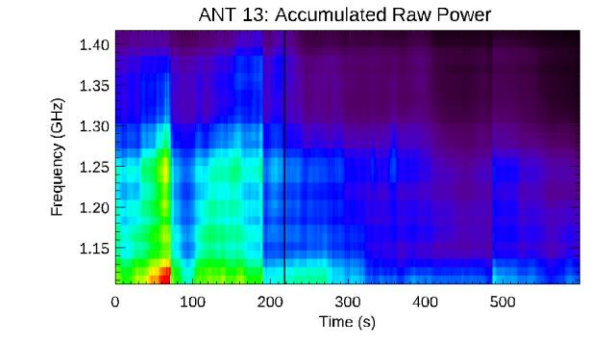
PFA	0.07%
Total	13.18%
DC<50%	11.56%
DC>50%	1.62%

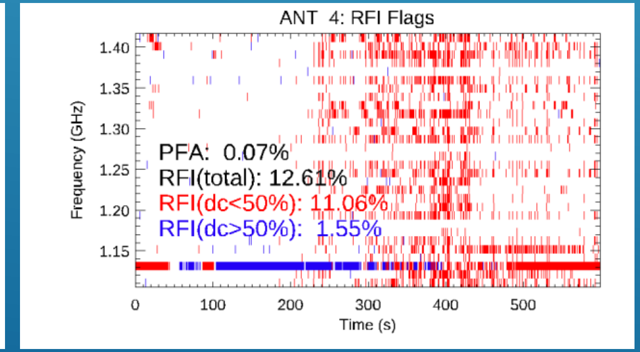




Antenna 3

PFA	0.07%
Total	7.97%
DC<50%	6.02%
DC>50%	1.96%



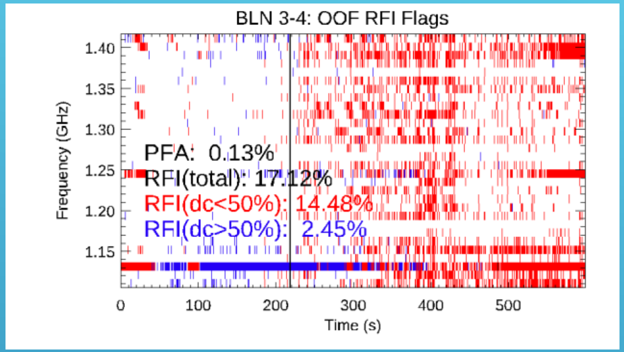


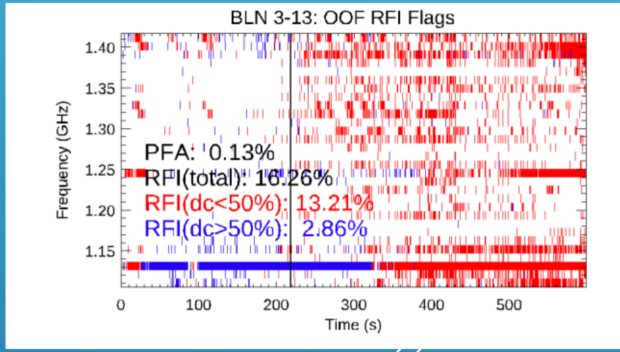
Antenna 13

PFA	0.07%
Total	12.61%
DC<50%	11.06%
DC>50%	1.55%

OOF A3-A4 (35 m)

PFA	0.135%
Total	17.12%
DC<50%	14.48%
DC>50%	2.45%





OOF A3-A13 (937 m)

PFA	0.135%
Total	16.25%
DC<50%	13.21%
DC>50%	2.86%

MEAN AUTO-CORRELATION FLAGGING (MACF)

The MACF strategy, proposed by Taylor et al. 2019, JAI, 08 (01) ("Spectral Kurtosis-Based RFI Mitigation for CHIME") consists in averaging *N* individual, antenna-based SK estimators,

$$\langle SK \rangle = \frac{1}{N}\sum_{i=1}^{N} SK_i$$

The $\langle \boldsymbol{SK} \rangle$ estimator has unity expectation and a variance that is N times smaller than that of the individual antenna-based estimators,

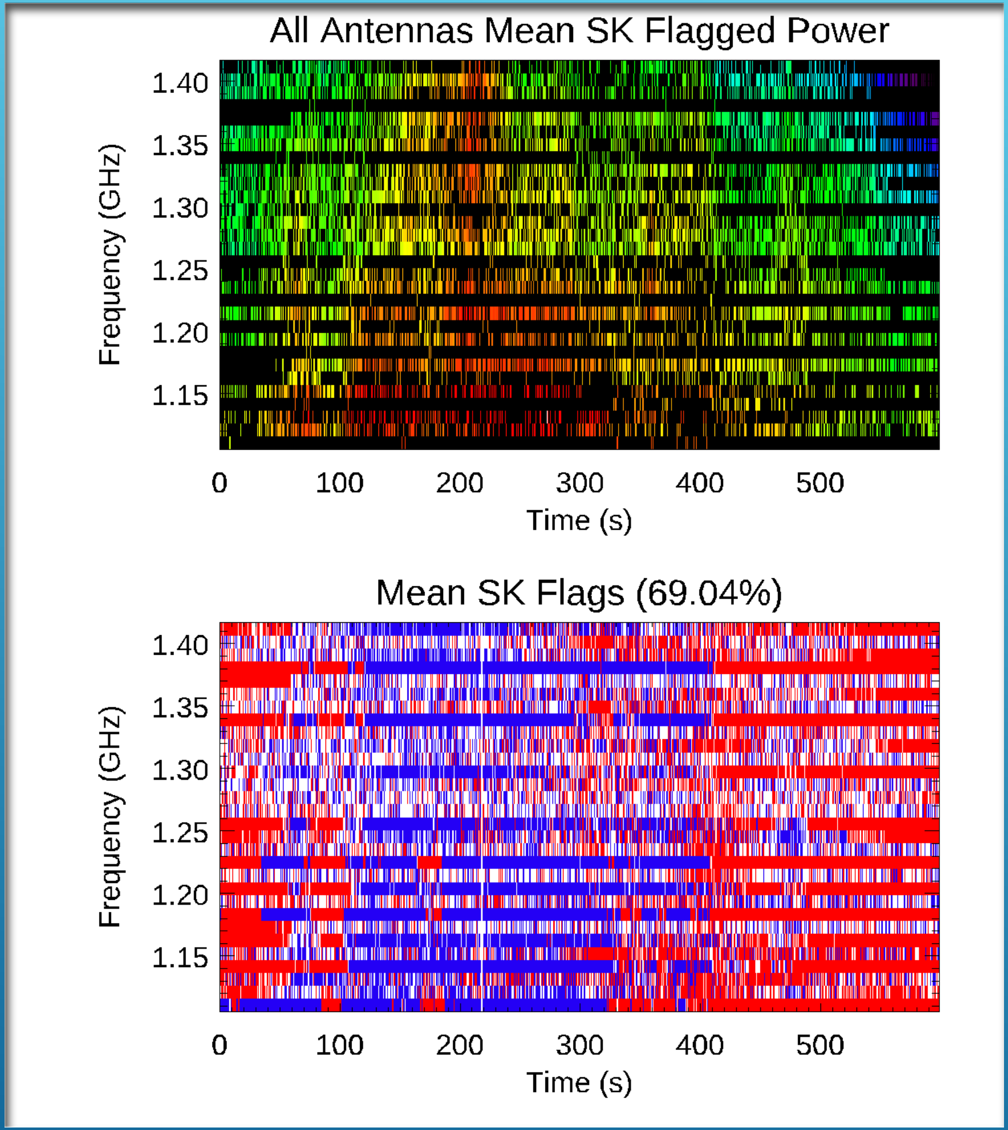
$$\sigma_{SK}^2 = \approx 2\left(1+\frac{1}{d}\right)\frac{1}{M}$$

$$\sigma_{\langle SK \rangle}^2 = \frac{1}{N}\sigma_{SK}^2 \approx 2\left(1+\frac{1}{d}\right)\frac{1}{N \times M} + O\left[\frac{1}{(N \times M)^2}\right]$$

The first order approximation of the series expansion allows one to approximate the PFA associated with a chosen pair of RFI flagging thresholds using the simple substitution

$$M \to N \times M$$

in the analytical PFA expressions provided by Nita & Gary 2010, MNRAS, 406, L60-L64



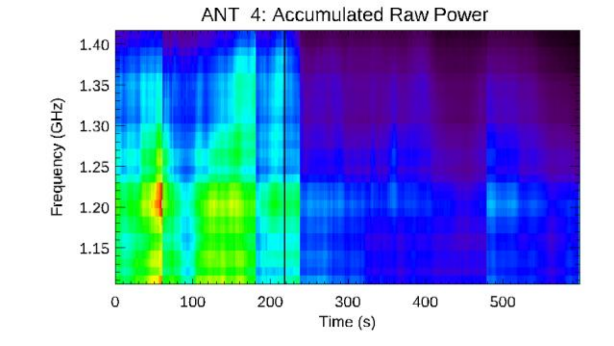
- Using an array-wide MACF approach may result in applying excessive RFI flags (69% in this example) for those baselines that may not be affected by local interference
- Nevertheless, the MACF approach may be implemented for each individual baseline, case in which the baseline estimators may be expressed in term of the autocorrelation estimators as

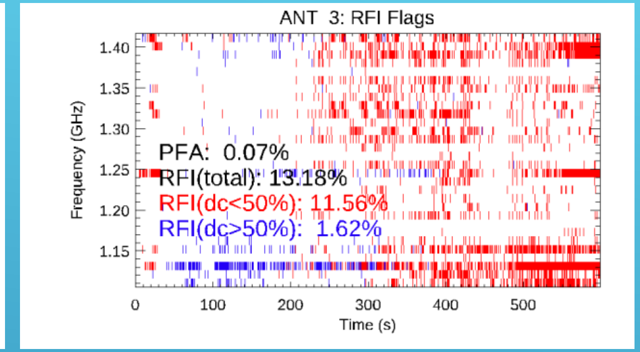
$$\langle SK \rangle_{ij} = \frac{1}{2}\left(SK_i + SK_j\right)$$

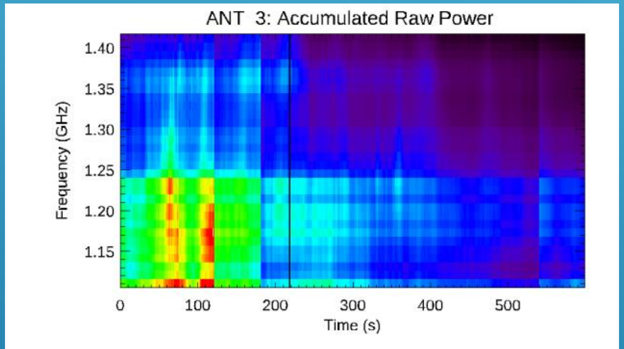
and the probabilities of false alarm may be computed using the substitution $M \rightarrow 2M$

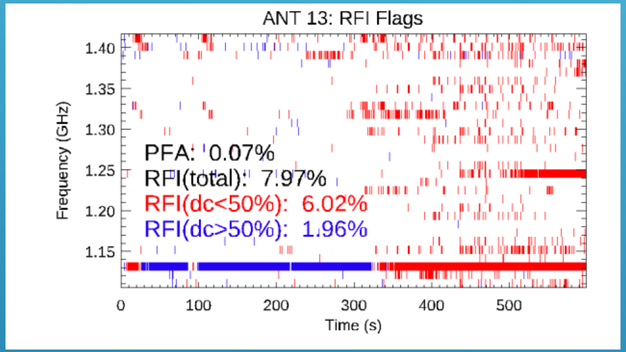
A CAUTIONARY NOTE ON ARRAY-WIDE MACF FLAGGING

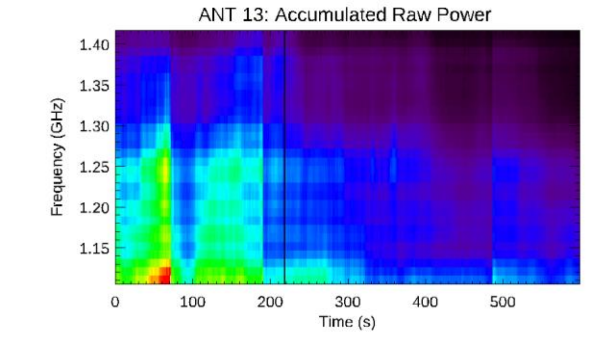
MACF CROSS-CORRELATION FLAGGING EXAMPLE

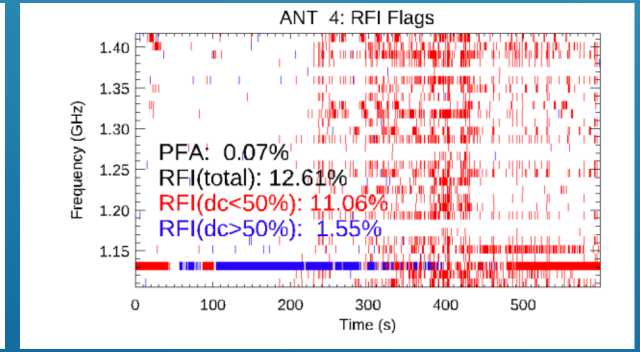










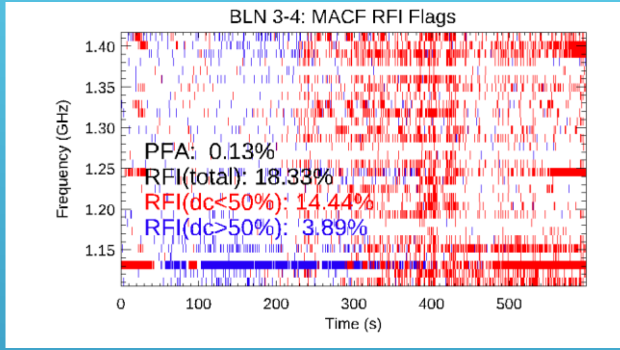


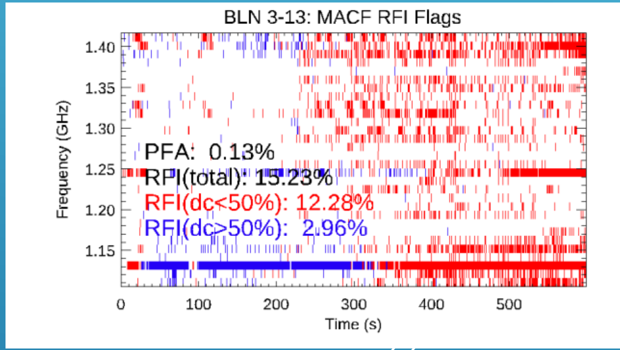
Antenna 4	
PFA	0.07%
Total	13.18%
DC<50%	11.56%
DC>50%	1.62%

Antenna 3	
PFA	0.07%
Total	7.97%
DC<50%	6.02%
DC>50%	1.96%

Antenna 13	
PFA	0.07%
Total	12.61%
DC<50%	11.06%
DC>50%	1.55%

MACF A3-A4 (35 m)	
PFA	0.135%
Total	18.33%
DC<50%	14.44%
DC>50%	3.89%





MACF A3-A13 (937 m)	
PFA	0.135%
Total	15.23%
DC<50%	12.28%
DC>50%	2.96%

13-Antenna Array	Auto-Correlation Flags	OOF Baseline Flags	MACF Baseline Flags
Duty Cycle <50%	24.56%	18.95%	18.23%
Duty Cycle>50%	10.69%	10.12%	9.60%
Total	35.24%	29.32%	27.82%

We tested the performance of the Spectral Kurtosis estimator using RFI-contaminated data obtained with the Expanded Owens Valley Solar Array during the evolution of a solar radio burst, showing that:

- The autocorrelation SK estimator is a statistical tool capable of automatic, real-time discrimination of artificial and natural signals
- The two autocorrelation SK estimators corresponding to the baseline antennas may be combined to compute OOF or MACF baseline estimators that perform comparably well in selectively detecting RFI, while ensuring minimal astrophysical data loss, provided that the astronomical signal of interest varies at a temporal scale larger than the accumulation length of the correlator
- Nevertheless, given the fact that the natural transient signals may only produce statistically significant SK values larger than unity, while the RFI SK may deviate in both directions, the RFI flags corresponding to SK values lower than unity may be always safely employed to selectively remove RFI signals having relative duty cycles larger than 50%, even in the presence of astrophysical transient signals such as solar radio spikes, pulsars, or FRBs.

CONCLUSION