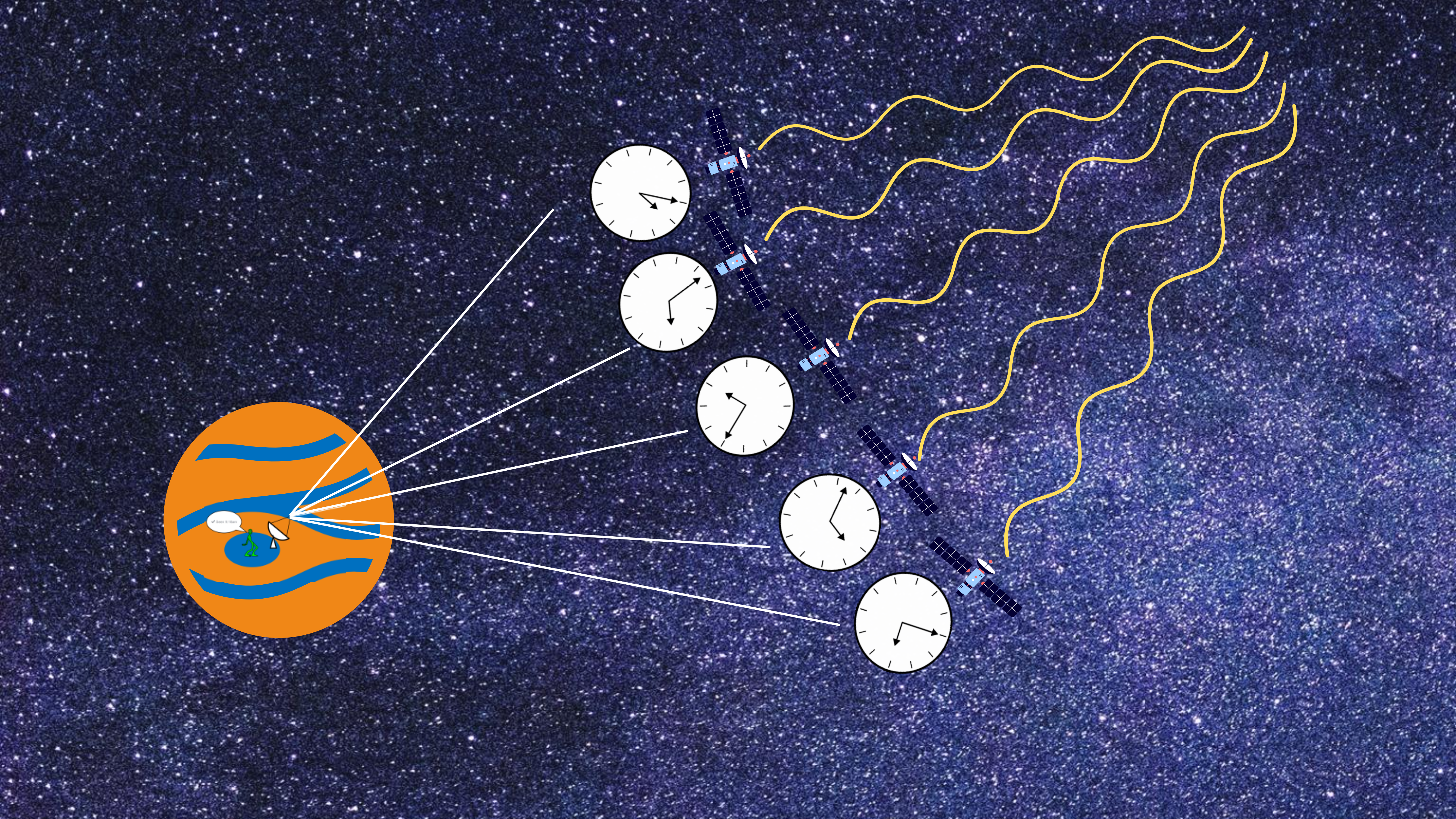
Robust and Misspecified Estimation for Time Scale Generation in a Swarm of Satellites

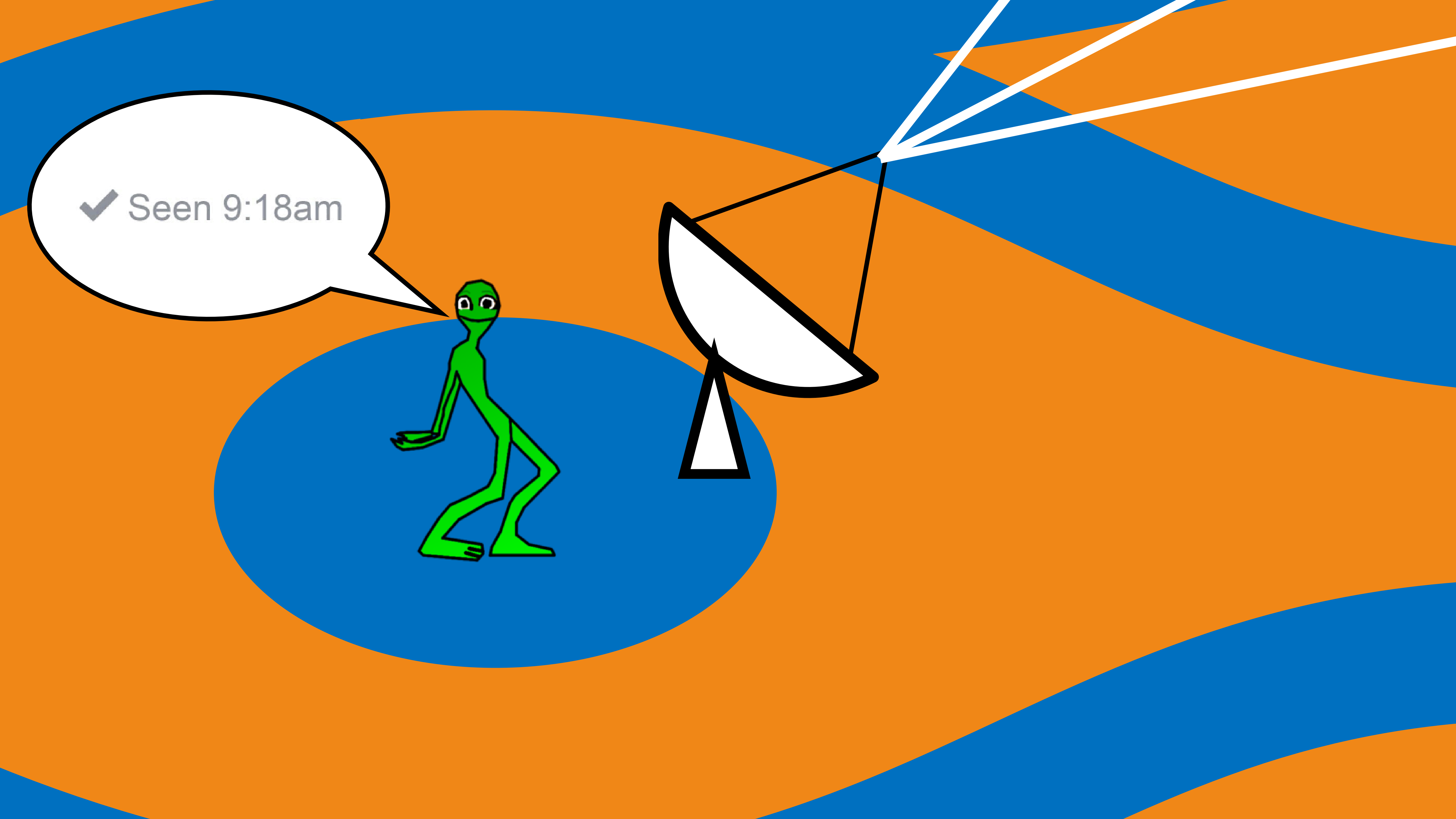
26/02/2026

Hamish McPhee

cnes ThalesAlenia Space a Thales / Leonardo company TéSA


Is there anybody out there?!!




✔ Seen 9:18am

Robust and Misspecified Estimation for Time Scale Generation in a Swarm of Satellites

26/02/2026

Hamish McPhee

cnes ThalesAlenia Space a Thales / Leonardo company TéSA

Contents

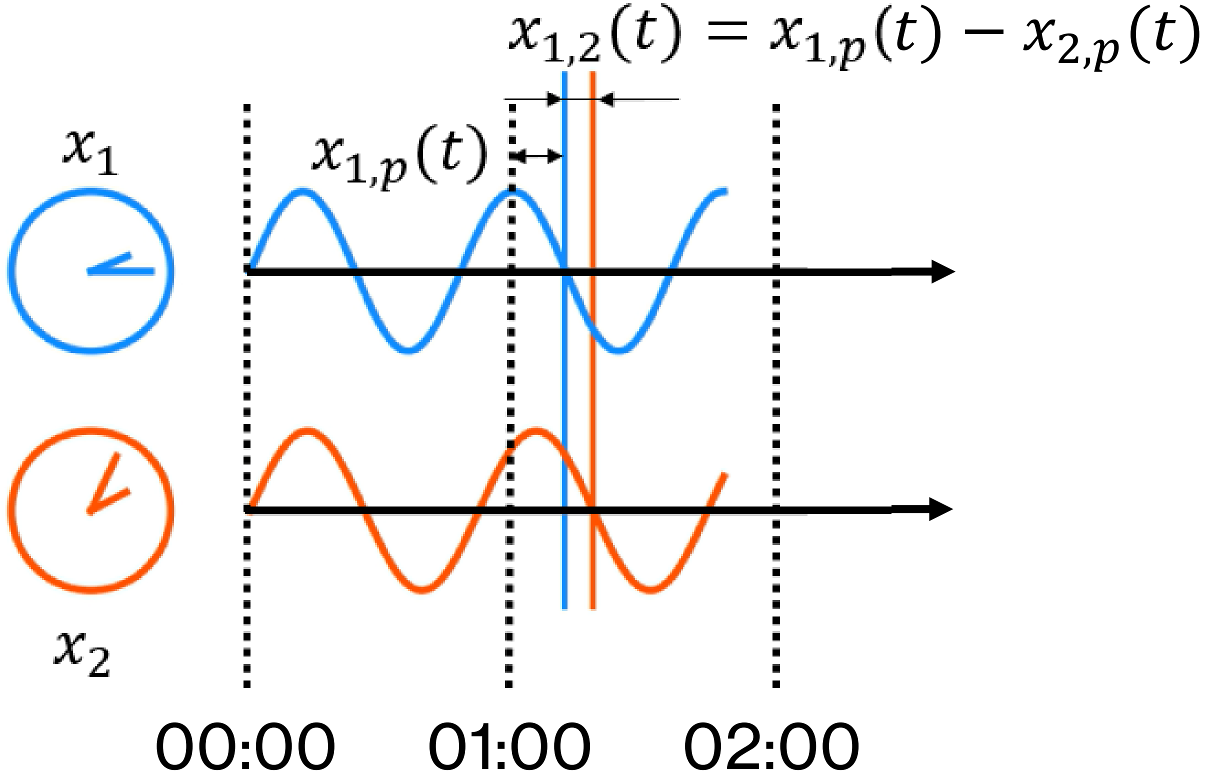
- Time Scales — 10:35 – 10:40
- Misspecified Estimation — 10:40 – 10:55
- Robust Estimators — 10:55 – 11:10
- Real Data — 11:10 – 11:15
- Questions — 11:15 –

10:35 – 10:40

Time Scales

What is a clock?

$$V(t) = V_0[1 + \alpha(t)]\cos\big(t + x(t)\big)$$



What is a Time Scale?

Time Scale

Time Scale Generation

- For an ensemble of N clocks:

$$\mathbf{z}(t) = \mathbf{H}\mathbf{x}, \qquad rank(\mathbf{H}) = N - 1,$$

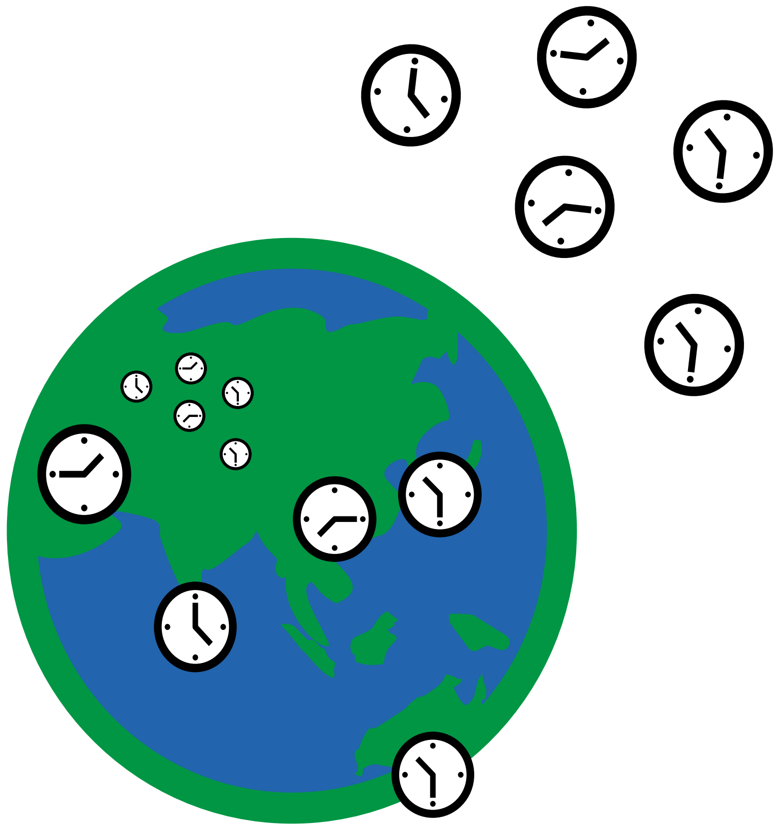
- Introduce constraint

$$\mathbf{w}^\top(\hat{\mathbf{x}} - \mathbf{x}) = 0$$

- Time Scale Equation

$$x_{i,E}(t) = \sum_{j=1}^{N} w_j(t)\left(\hat{x}_{j,E}(t) - x_{j,i}(t)\right)$$

$$x_{i,E}(t) = x_{i,p}(t) + \mathbf{w}^\top(\hat{\mathbf{x}} - \mathbf{x})$$



Time Scale Equation

- **Prediction**:

$$\hat{x}_{j,E}(t) = x_{j,E}(t-\tau) + \tau y_j(t-\tau), \qquad y(t) = \frac{dx(t)}{dt}$$

- **Time Scale Residuals**

$$r_{j,i}(t) = \hat{x}_{j,E}(t) - x_{j,i}(t), \qquad x_{i,E}(t) = \sum_{j=1}^{N} w_j\,(t) r_{j,i}(t),$$

$$r_{j,i}(t) = \ x_{i,p}(t) + \left(\hat{x}_{j,E}(t) - x_{j,p}(t)\right) = x_{i,p}(t) + e_j(t)$$

$$r_{j,i}(t) \sim \mathcal{P}\left(x_{i,p}(t), \sigma^2, \boldsymbol{\eta}\right)$$

10:40 – 10:55

Misspecified Estimation

Model Specification

- Want to choose the correct model

$$r_{j,i}(t) \sim \mathcal{P}\left(x_{i,p}(t), \sigma^2, \boldsymbol{\eta}\right) \text{ or } \mathcal{P}(\mu, \sigma^2, \boldsymbol{\eta})$$

- Under nominal conditions, we assume

$$r_{j,i}(t) \sim \mathcal{N}(\mu, \sigma^2)$$

- When faced with anomalies, $\mathcal{P} \neq \mathcal{N}$

Heavy-tailed Distributions

- Student's t-distribution $\boldsymbol{\psi} = [\mu_T, \sigma_T^2, \nu]$

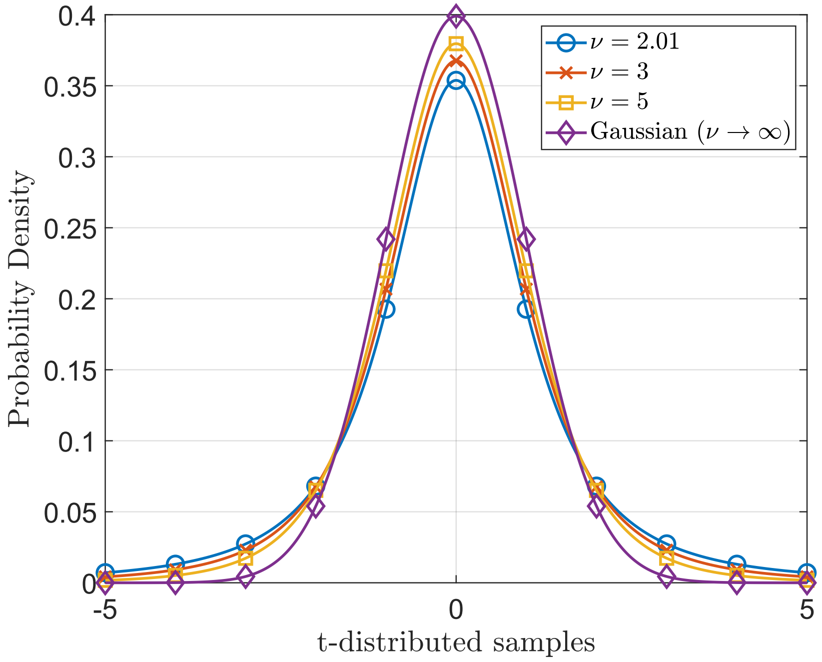
$$p_T(r; \boldsymbol{\psi}) = (\pi \nu \sigma_T^2)^{-\frac{1}{2}} \frac{\Gamma\left(\frac{\nu + 1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \left(1 + \frac{(r - \mu_T)^2}{\nu \sigma_T^2}\right)^{-\frac{\nu+1}{2}}$$

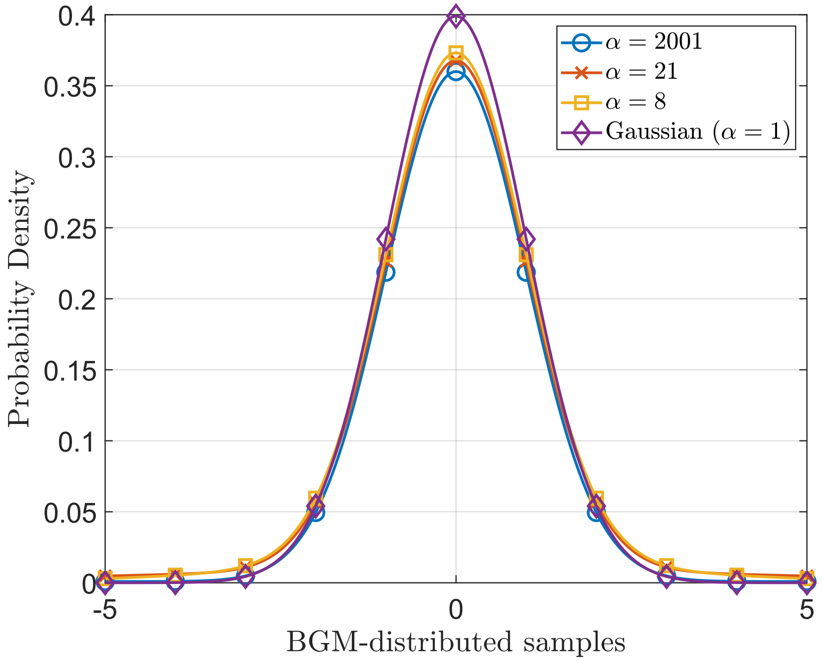
- Bivariance Gaussian Mixture $\boldsymbol{\phi} = [\mu_{GM}, \sigma_{GM}^2, \alpha, \varepsilon]$

$$p_{GM}(r; \boldsymbol{\phi}) = (1 - \varepsilon) p_G(r; \mu_{GM}, \sigma_{GM}^2) + \varepsilon p_G(r; \mu_{GM}, \alpha \sigma_{GM}^2)$$

Heavy-tailed Distributions

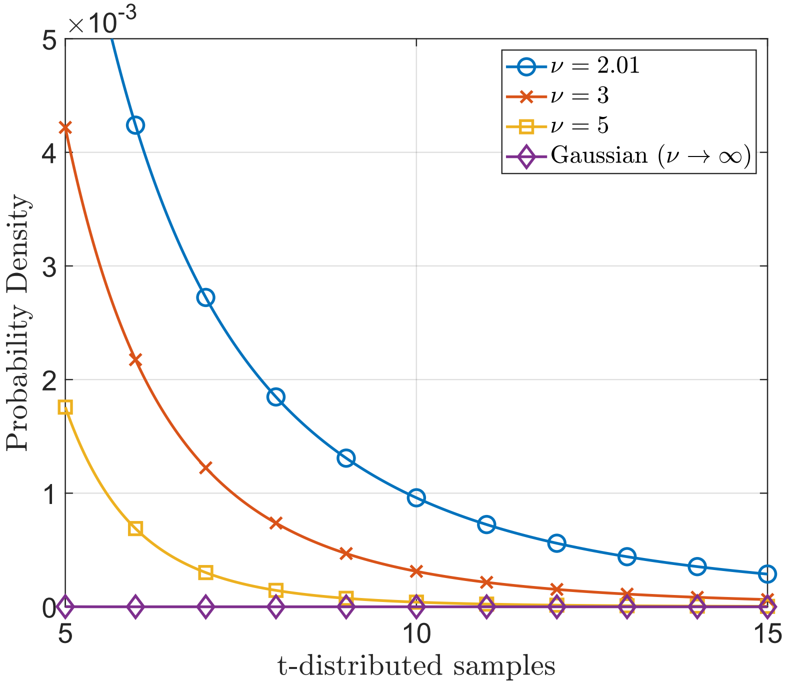
- Form of the PDFs for different nuisance parameters
- $\varepsilon = 0.1$, and α calculated to ensure equivalent 2 order moment

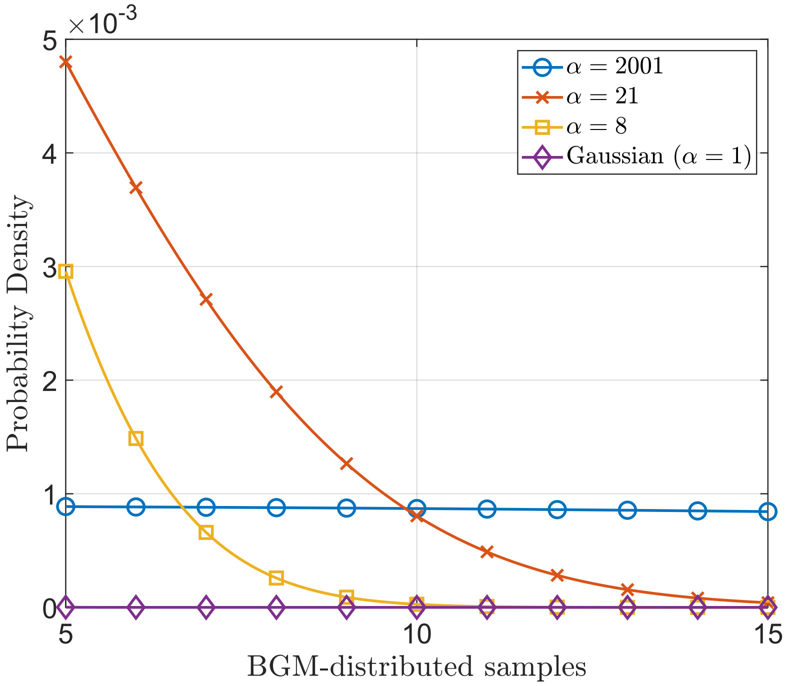




Heavy-tailed Distributions

- How do the tails differ?
- BGM tails flatten faster for extreme outliers

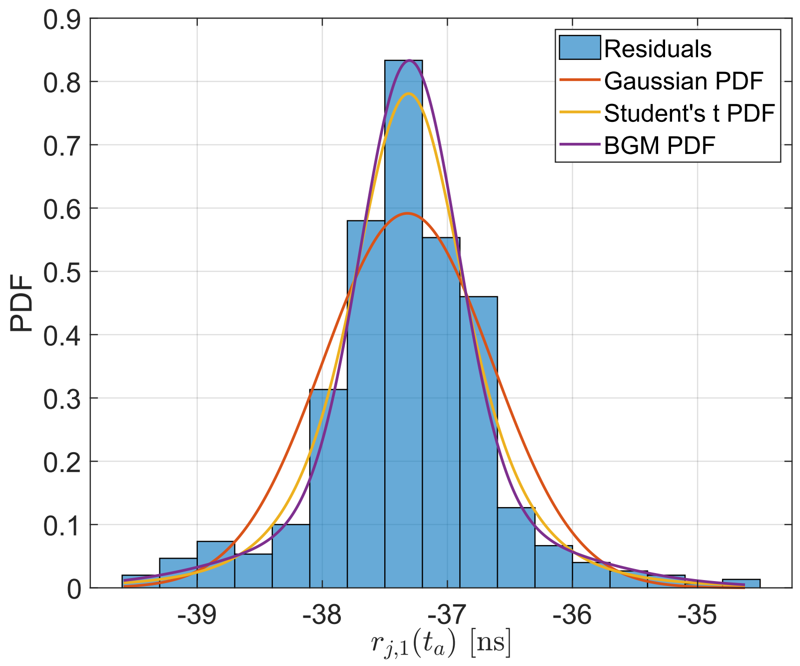




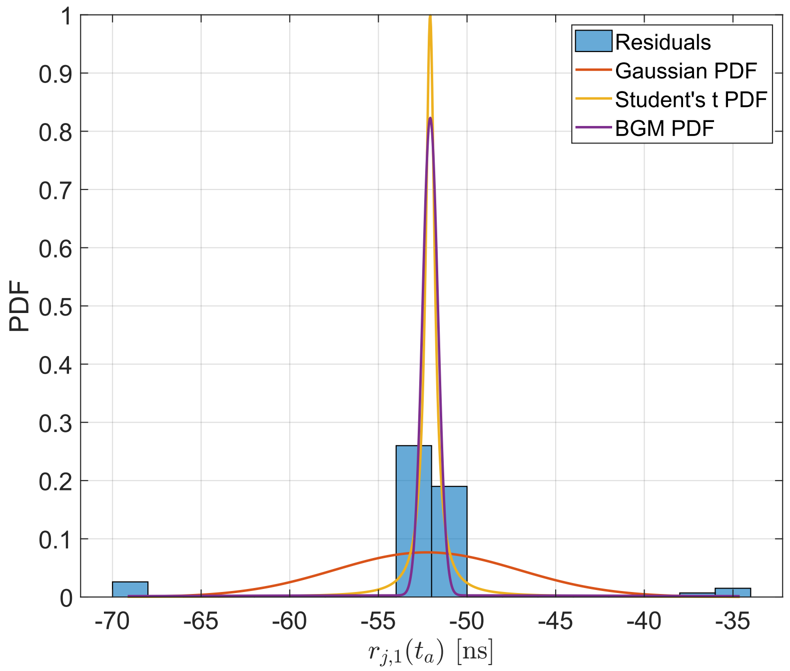
Heavy-tailed Distributions

- Data generation: 500 satellite clocks, 10% with phase jumps Δx at $t = t_a$

$$r_{j,i}(t) \sim \mathcal{P}\left(x_{i,p}(t), \sigma^2, \boldsymbol{\eta}\right)$$



$\Delta x \pm 1.4$ ns



$\Delta x = \pm 14$ ns

Misspecified Cramér-Rao Bound

- Cramér-Rao Bound for estimation of true parameters $\boldsymbol{\phi}_0 = \left[x_{i,p}(t), \sigma_0^2, \boldsymbol{\eta}\right]^\mathsf{T}$

$$E\left[\left\|\boldsymbol{\phi}_0 - \widehat{\boldsymbol{\phi}}_{\mathrm{MLE}}\right\|^2\right] \geq \mathbf{CRB}_{\boldsymbol{\phi}},$$

- If we instead base our estimator on a misspecified model (MMLE)

$$E\left[\left\|\boldsymbol{\theta}_0 - \widehat{\boldsymbol{\theta}}_{\mathrm{MMLE}}\right\|^2\right] \geq \mathbf{MCRB}_{\boldsymbol{\theta}}(p||q) + \left(\boldsymbol{\theta}_0 - \boldsymbol{\theta}_{p||q}\right)^2,$$

Where $\boldsymbol{\theta}_0 = \left[x_{i,p}(t), \sigma_0^2\right]^\mathsf{T}$, and $\boldsymbol{\theta}_{p||q}$ is the **pseudo-true** parameter vector

Deriving the MCRB

$$\boldsymbol{\theta}_{p||q} = \min_{\boldsymbol{\theta}} D\big(p(r;\boldsymbol{\phi})||q(r;\boldsymbol{\theta})\big) = \min_{\boldsymbol{\theta}} E_p\big[\log\big(p(r;\boldsymbol{\phi})\big) - \log\big(q(r;\boldsymbol{\theta})\big)\big],$$

$$\mathbf{MCRB}_{\boldsymbol{\theta}} = \mathbf{A}\big(\boldsymbol{\theta}_{p||q}\big)^{-1}\mathbf{B}\big(\boldsymbol{\theta}_{p||q}\big)\mathbf{A}\big(\boldsymbol{\theta}_{p||q}\big)^{-1},$$

$$\mathbf{A}\big(\boldsymbol{\theta}_{p||q}\big) = E_p\left[\frac{\partial \log\big(q(r;\boldsymbol{\theta})\big)}{\partial \boldsymbol{\theta}}\frac{\partial \log\big(q(r;\boldsymbol{\theta})\big)}{\partial \boldsymbol{\theta}}^{\top}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_{p||q}}, \mathbf{B}\big(\boldsymbol{\theta}_{p||q}\big) = E_p\left[\frac{\partial^2 \log\big(q(r;\boldsymbol{\theta})\big)}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}^{\top}}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_{p||q}}$$

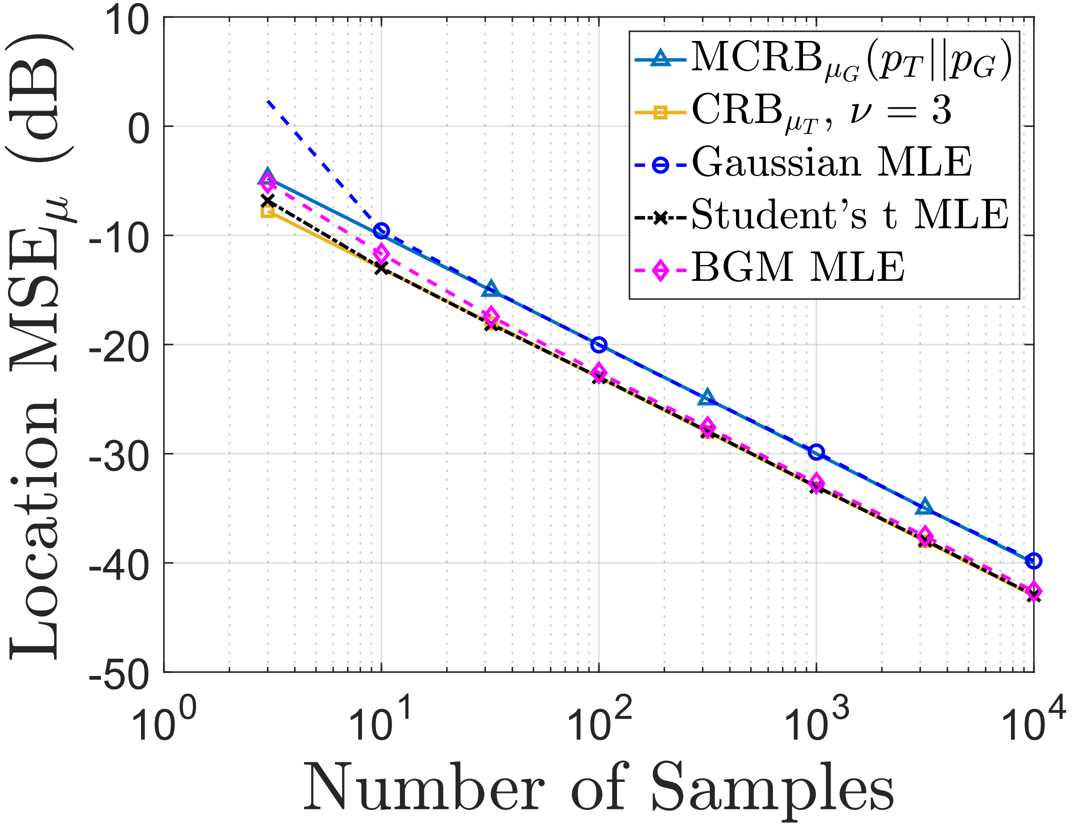
$$\mathrm{MCRB}_{\mu} = \frac{\sigma^2_{p||p_G}}{N}$$

Misspecification Student’s t data

- Data generation: t-dist.

$$\boldsymbol{\psi} = [\mu_T, \sigma_T^2, \nu]$$

  - Assuming a Gaussian
  - Assuming a Student’s t-distribution
  - Assuming a BGM

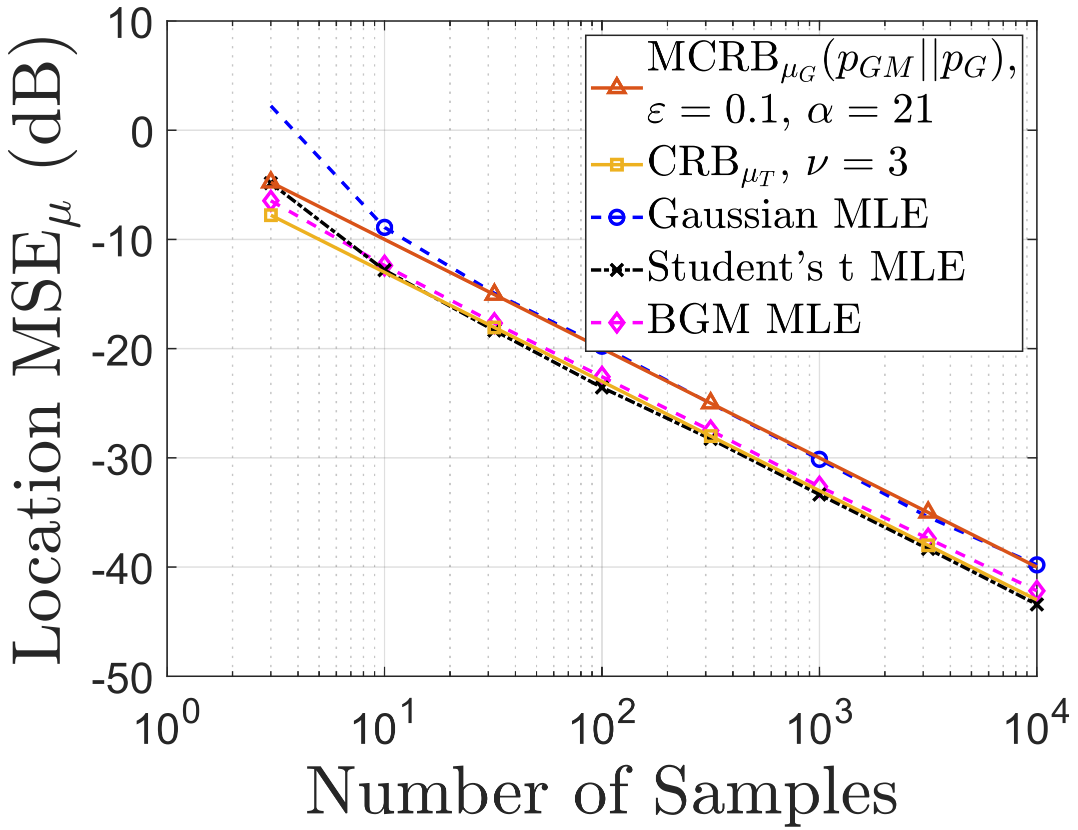


Misspecification BGM data

- Data generation: BGM

$$\boldsymbol{\phi} = [\mu_{GM}, \sigma_{GM}^2, \alpha, \epsilon]$$

  - Assuming a Gaussian
  - Assuming a BGM
  - Assuming a Student's t-distribution



10:55 – 11:10

Robust Estimators

Expectation Maximization

- MLE maximizes likelihood of true distribution $p(r;\mu)$

$$\frac{\partial}{\partial\mu}\left[\sum_{j=1}^{N} -\log\big(p(r;\mu)\big)\right] = 0, \text{ solve for } \mu \text{ to obtain } \hat{\mu}_{MLE}$$

- For both heavy-tailed distributions, $\hat{\mu}_{MLE}$ has no closed form

- Minimize the complete log-likelihood

$$\hat{\mu}_{EM} = \arg\min_{\mu} \sum_{j=1}^{N} l_c\big(r_j, v_j; \mu\big)$$

Expectation Maximization

- Representation of Heavy-tailed distributions with latent variables v_j

  – Student's t distribution: $r_j \sim T(\mu_T, \sigma_T^2, \nu) \rightarrow r_j | v_j \sim \mathcal{N}\left(\mu_T, \frac{\sigma_T^2}{v_j}\right), v_j \sim \mathcal{G}\left(\frac{\nu}{2}, \frac{\nu}{2}\right)$

  – BGM: $r_j \sim GM(\mu_{GM}, \sigma_{GM}^2, \alpha, \varepsilon) \rightarrow r_j | v_j \sim \begin{cases} \mathcal{N}(\mu_T, \sigma_{GM}^2), v_j = 0 \\ \mathcal{N}(\mu_T, \alpha\sigma_{GM}^2), v_j = 1 \end{cases}, \quad v_i \sim \text{Binomial}(1, \varepsilon)$

Expectation Maximization

- E-step: Calculate $E\left[v_j|r_j\right]$, $E_{v_j|r_j}\left[l_c\left(r_j, v_j\right)\right] = Q(r;\mu)$

$$E_T\left[v_j|r_j\right] = \frac{\hat{v}_{k-1}+1}{\hat{v}_{k-1}+\frac{\left(r_j-\hat{\mu}_{k-1}\right)^2}{\hat{\sigma}_{k-1}^2}}, \qquad E_{GM}\left[v_j|r_j\right] = \frac{\hat{\varepsilon}_{k-1} p_G\left(r_j; \hat{\mu}_{k-1}, \hat{\alpha}_{k-1}\hat{\sigma}_{k-1}^2\right)}{p_{GM}\left(r_j; \hat{\mu}_{k-1}, \hat{\sigma}_{k-1}^2, \hat{\alpha}_{k-1}, \hat{\varepsilon}_{k-1}\right)}$$

- M-step: Minimize cost function $\hat{\mu}_k = \arg\min_{\mu} Q(r;\mu) = \sum_{j=1}^{N} w(r_j - \hat{\mu}_{k-1}) r_j$

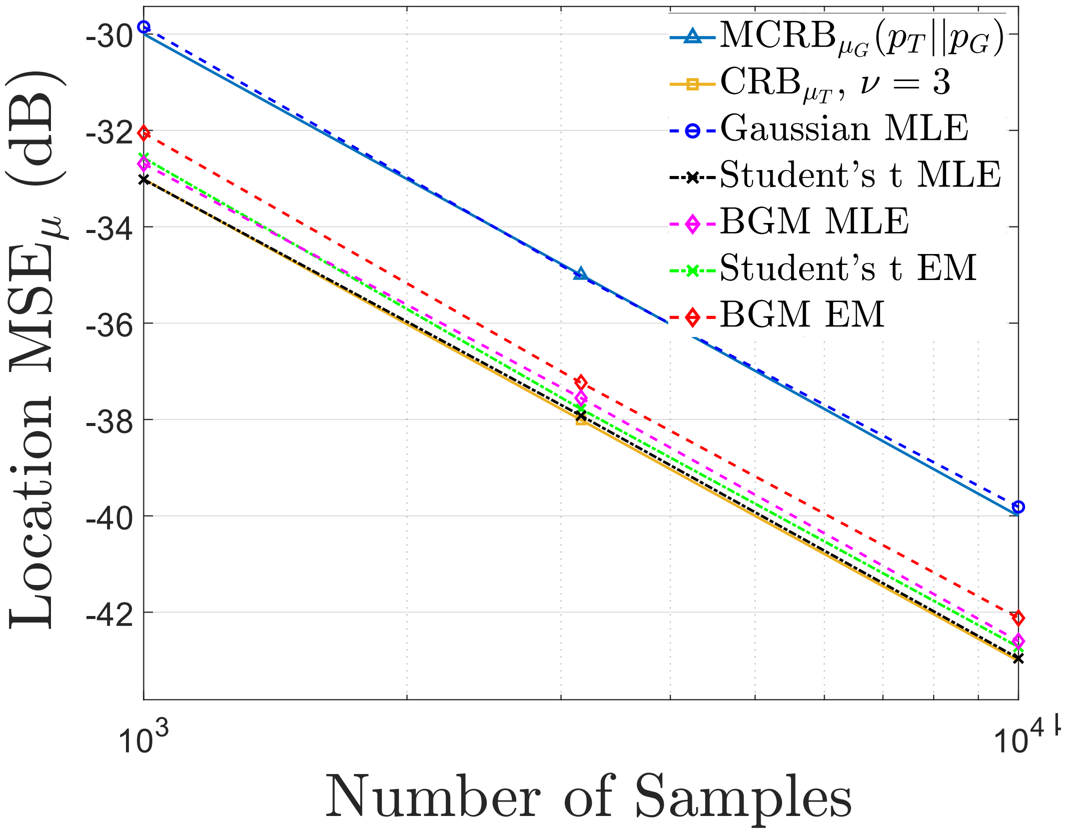
$$w_T\left(r_j - \hat{\mu}_{k-1}\right) \propto E_T\left[v_j|r_j\right], \qquad w_{GM}\left(r_j - \hat{\mu}_{k-1}\right) \propto 1 - E_{GM}\left[v_j|r_j\right] + \frac{E_{GM}\left[v_j|r_j\right]}{\hat{\alpha}_{k-1}}$$

Misspecification Student's t data

- Data generation: t-dist.

$$\boldsymbol{\psi} = [\mu_T, \sigma_T^2, \nu]$$

  - Assuming a Gaussian
  - Assuming a Student's t-distribution
  - Assuming a BGM

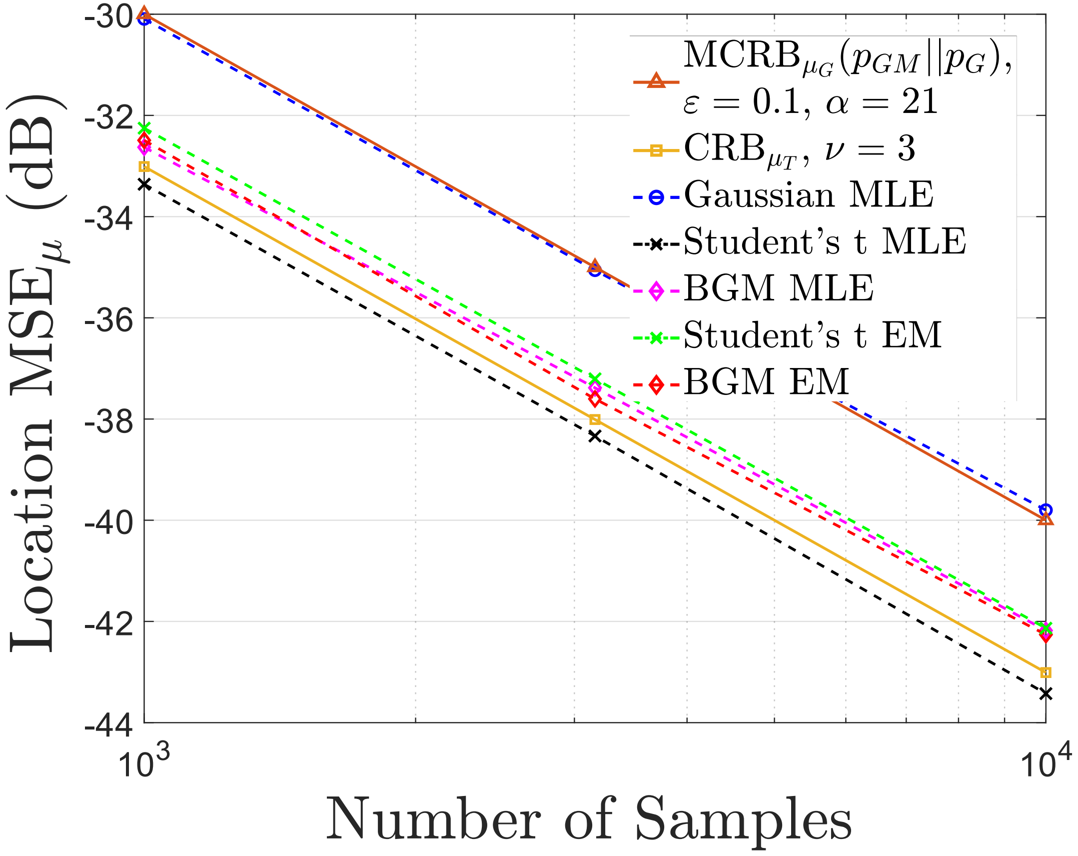


Misspecification BGM data

- Data generation: BGM

$$\boldsymbol{\phi} = [\mu_{GM}, \sigma_{GM}^2, \alpha, \epsilon]$$

  - Assuming a Gaussian
  - Assuming a BGM
  - Assuming a Student's t-distribution



M-estimators of Location

- MLE maximizes likelihood of true distribution $p(r;\mu)$

$$\hat{\mu}_{MLE} = \arg\max_{\mu} \prod_{j=1}^{N} p(r_j;\mu) = \arg\min_{\mu} \sum_{j=1}^{N} \rho(r_j - \mu)$$

- Where $\rho(r_j - \mu) = -\log\left(p(r_j;\mu)\right) = c(r_j - \mu)^2$ for Gaussian $p(r_j;\mu)$

$$\hat{\mu}_{MLE} = \arg\min_{\mu} \sum_{j=1}^{N} (r_j - \mu)^2 \quad \rightarrow \quad \hat{\mu}_{MLE} = \frac{1}{N} \sum_{j=1}^{N} r_j$$

M-estimators of Location

- M-estimators minimize different loss functions

$$\hat{\mu}_M = \arg\min_{\mu} \sum_{j=1}^{N} \rho(r_j - \mu) \;\rightarrow\; \hat{\mu}_M = \sum_{j=1}^{N} w(r_j - \hat{\mu}_M) r_j$$

$$w(x) = \begin{cases} \dfrac{\rho'(x)}{x}, & x \neq 0 \\ \rho''(0), & x = 0 \end{cases}$$

- Iteratively Reweighted Least Squares (IRLS)

Robust Loss and Weights

Tukey's Bisquare Loss

$$\rho(x) = \begin{cases} 1 - \left(1 - \left(\frac{x}{b}\right)^2\right)^3, & |x| \le b \\ 1, & |x| > b \end{cases}$$

Tukey's Bisquare Weights

$$w(x) = \begin{cases} \left(1 - \left(\frac{x}{b}\right)^2\right)^2, & |x| \le b \\ 0, & |x| > b \end{cases}$$

$b = 4.685 \rightarrow 95\%$ efficiency

Student's Loss

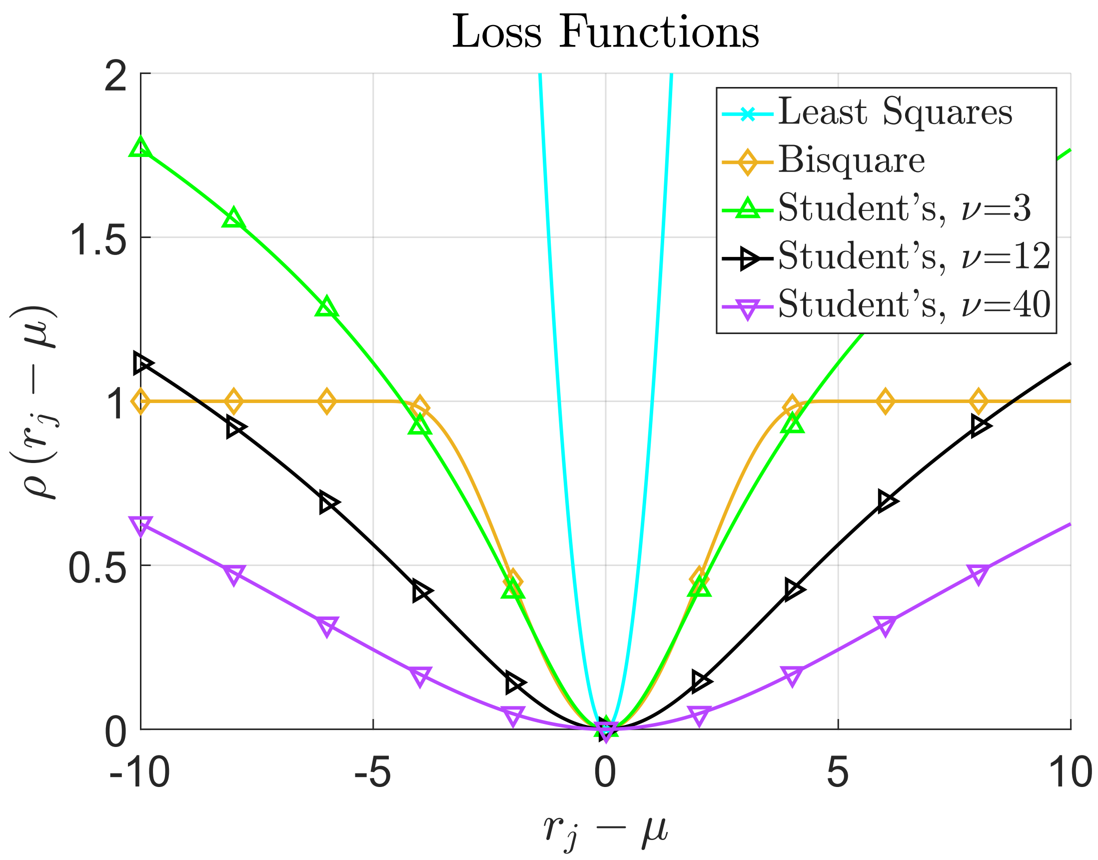
$$\rho(x) = \frac{1}{2}\log(x^2 + \nu) - \frac{1}{2}\log(\nu)$$

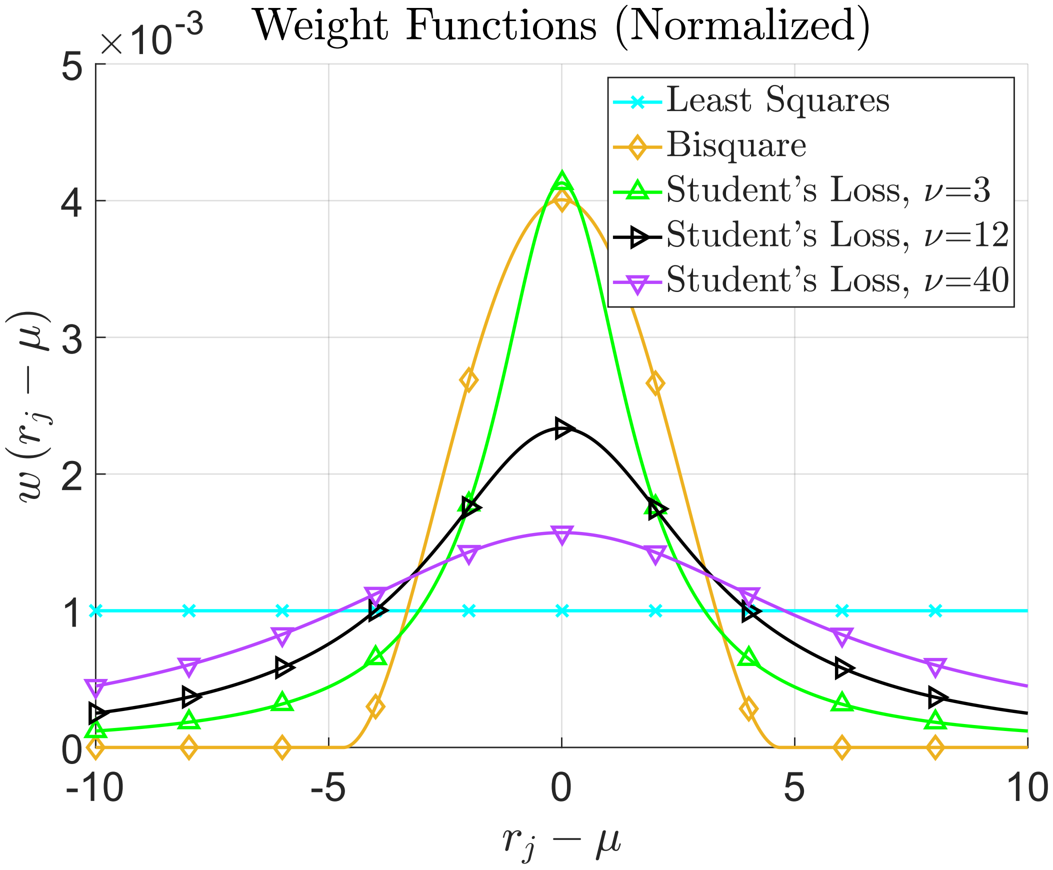
Student's Weights

$$w(x) = \frac{1}{x^2 + v}$$

$\nu = ? \rightarrow 95\%$ efficiency

Robust Loss and Weights



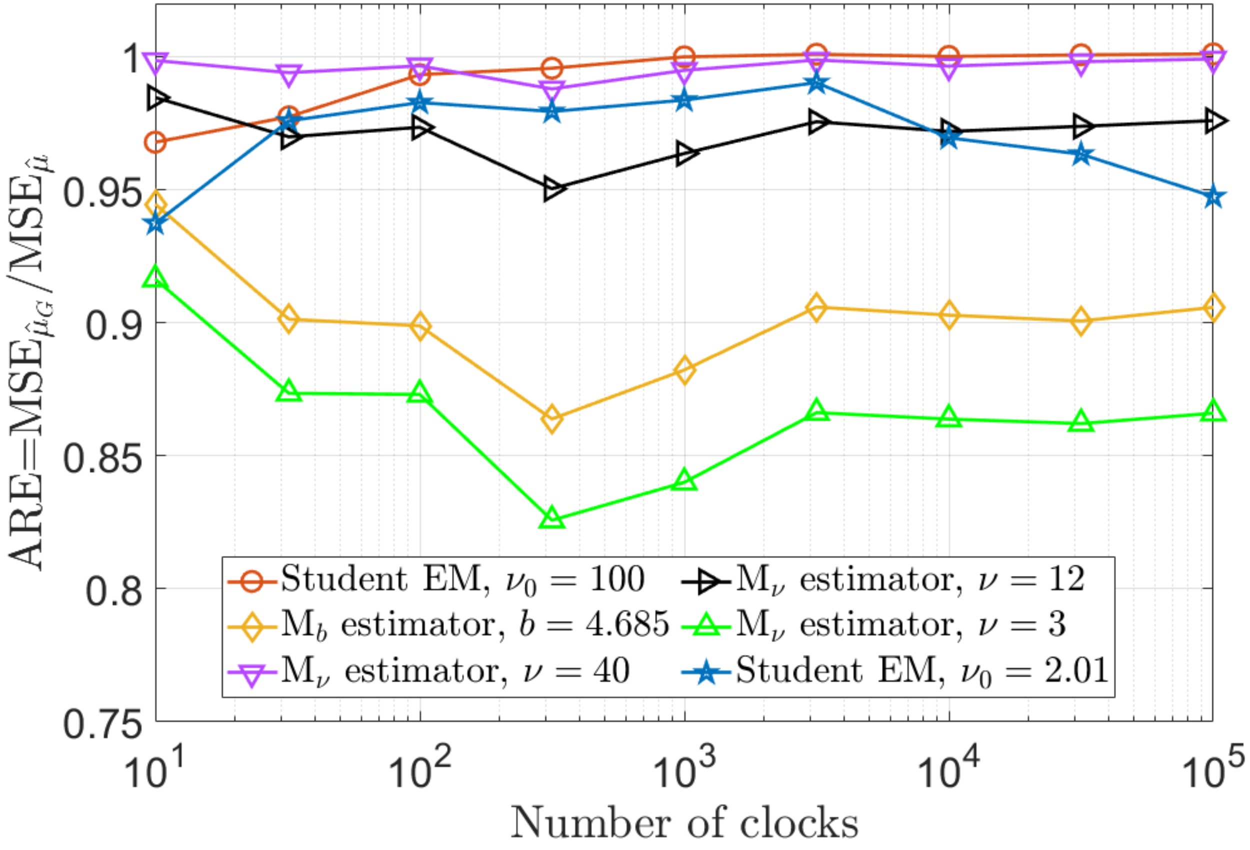


$$\hat{\mu} = \arg\min_{\mu} \sum_{j=1}^{N} \rho(r_j - \mu)$$

$$\hat{\mu}_k = \sum_{j=1}^{N} w(r_j - \hat{\mu}_{k-1}) r_j$$

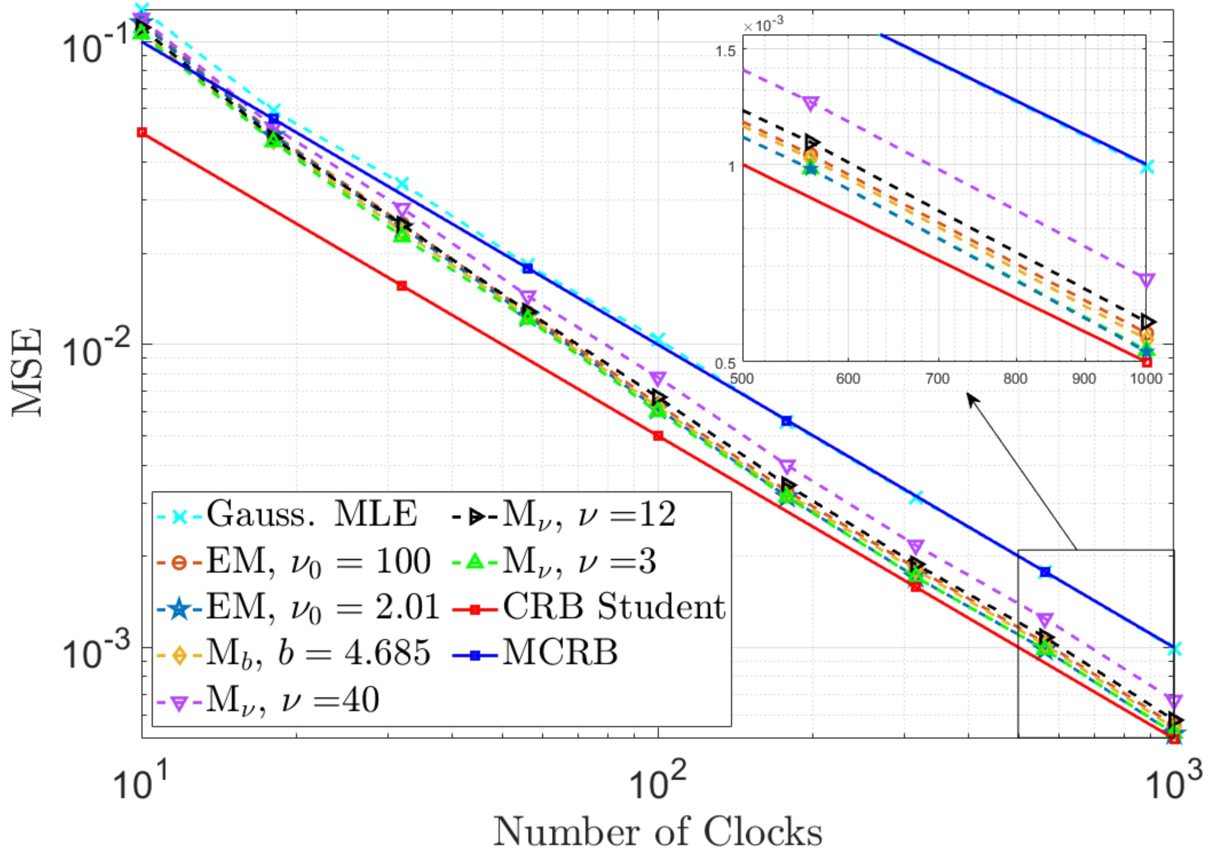
Comparison M-estimators vs. EM

- Nominal Efficiency
- Robustness
- Complexity



Comparison M-estimators vs. EM

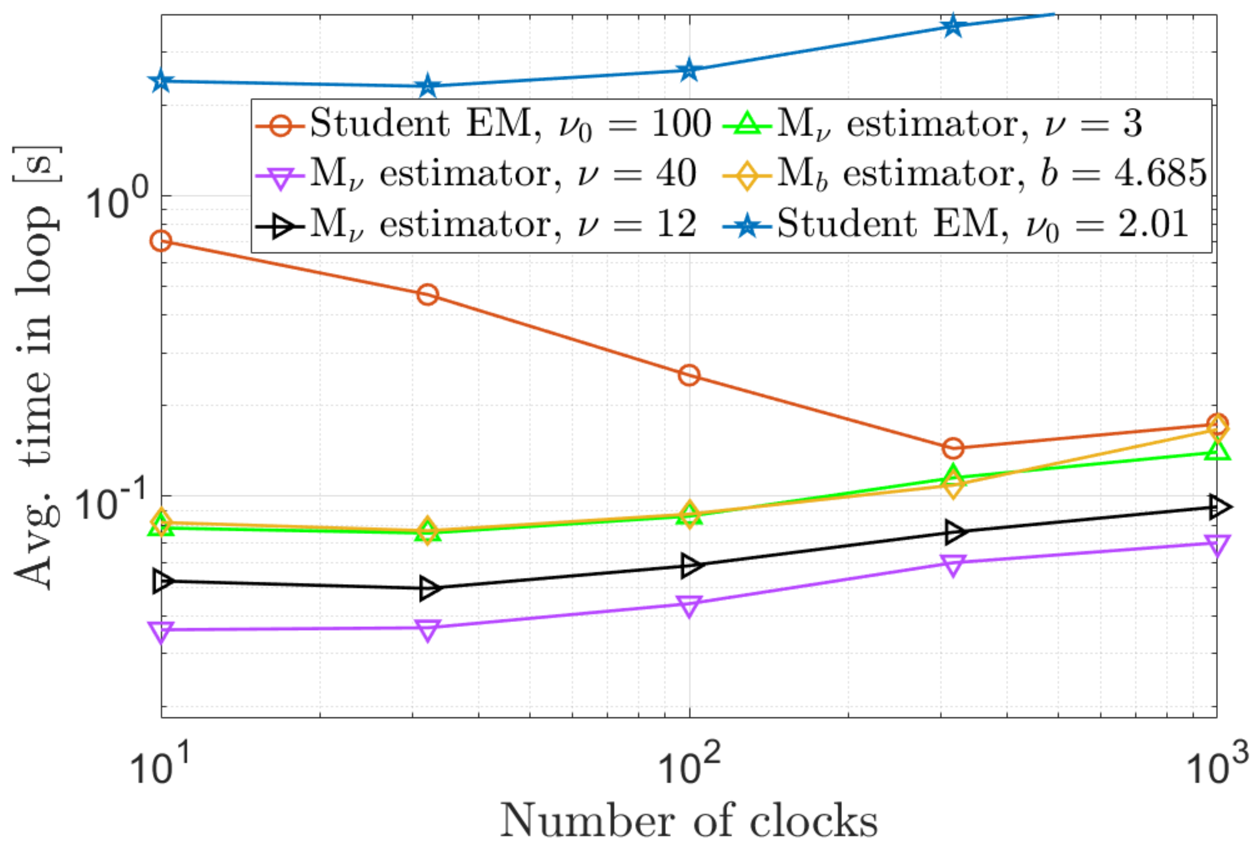
- Nominal Efficiency
- **Robustness**
- Complexity



Comparison M-estimators vs. EM

- Nominal Efficiency
- Robustness
- Complexity

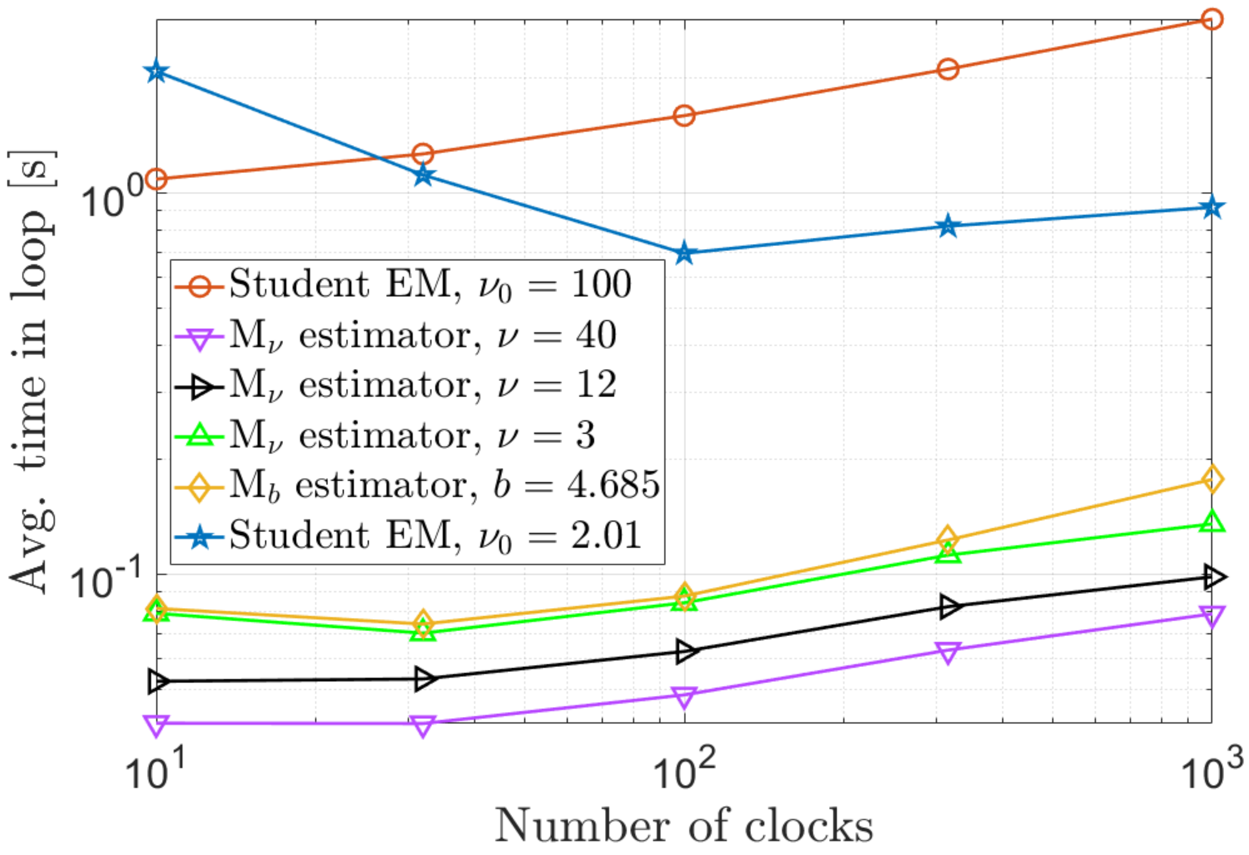
Without anomalies



Comparison M-estimators vs. EM

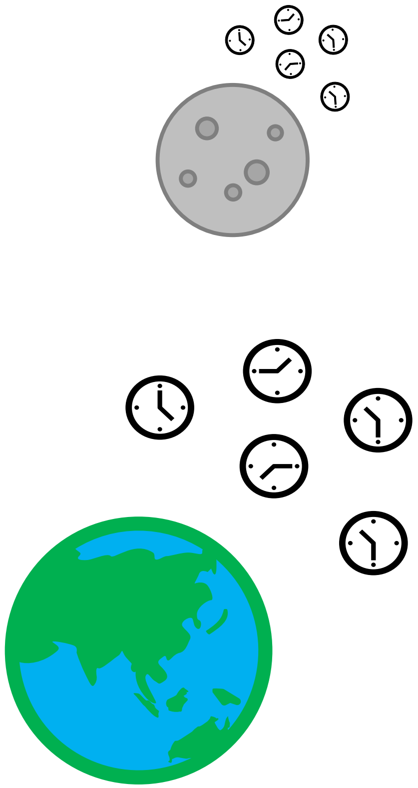
- Nominal Efficiency
- Robustness
- Complexity

With anomalies



Comparison M-estimators vs. EM

Metric	EM	M estimator
Efficiency	High $\hat{\nu}_0$ - Nominal	$\nu = 12$ for 95% efficiency
Robustness	Low $\hat{\nu}_0$ - Anomalies	Greatest robustness loses efficiency
Complexity	Autonomous High comp. budget	Autonomous Constrained comp.

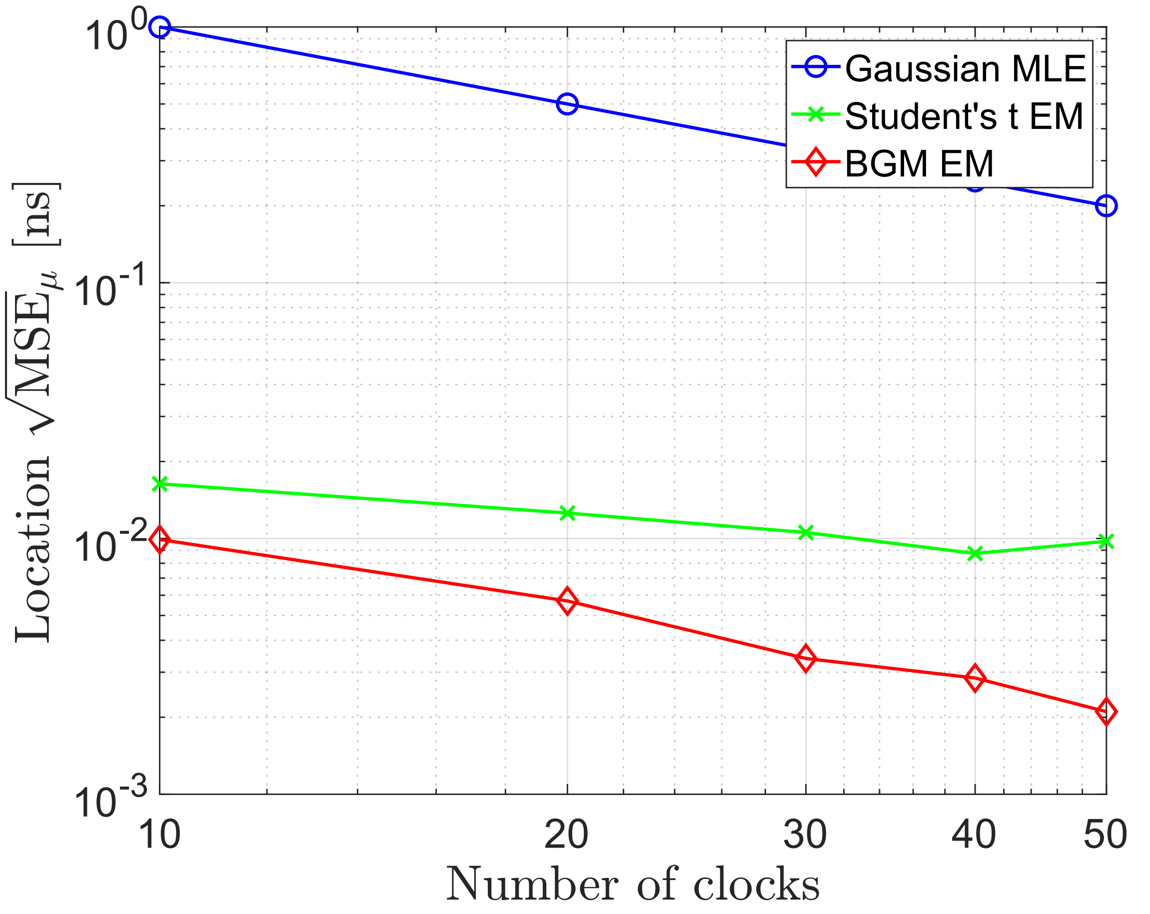


11:10 – 11:20

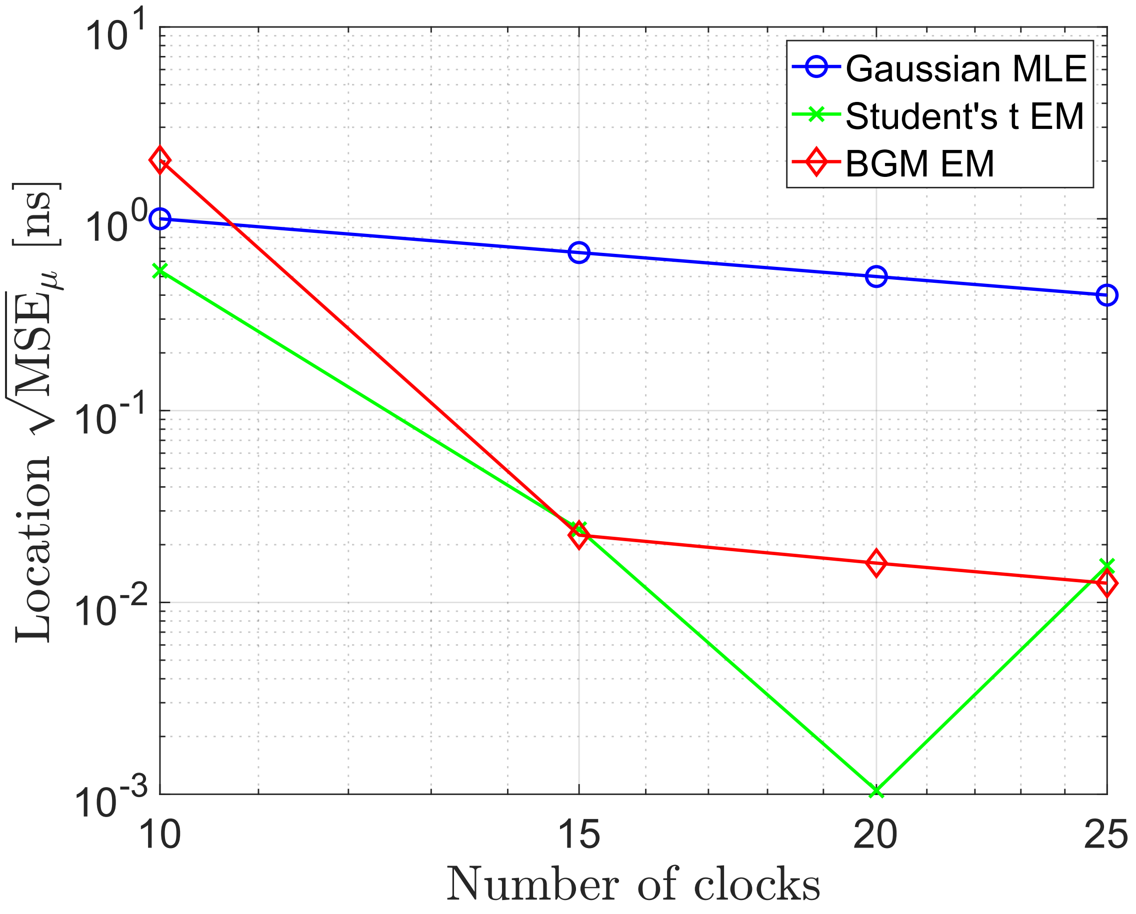
Real Data

Clock Location Estimate – 10 ns jump

Synthetic Chip Scale Atomic Clocks

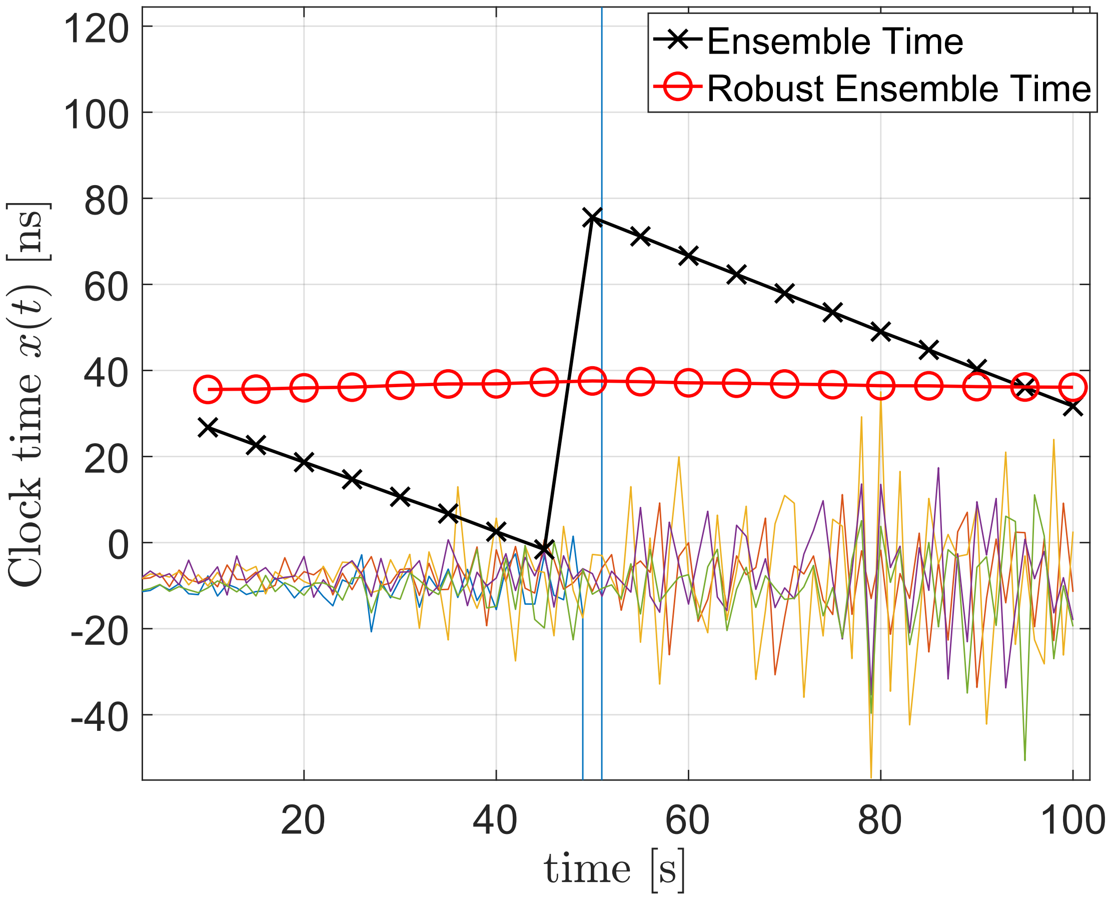


Real Chip Scale Atomic Clocks

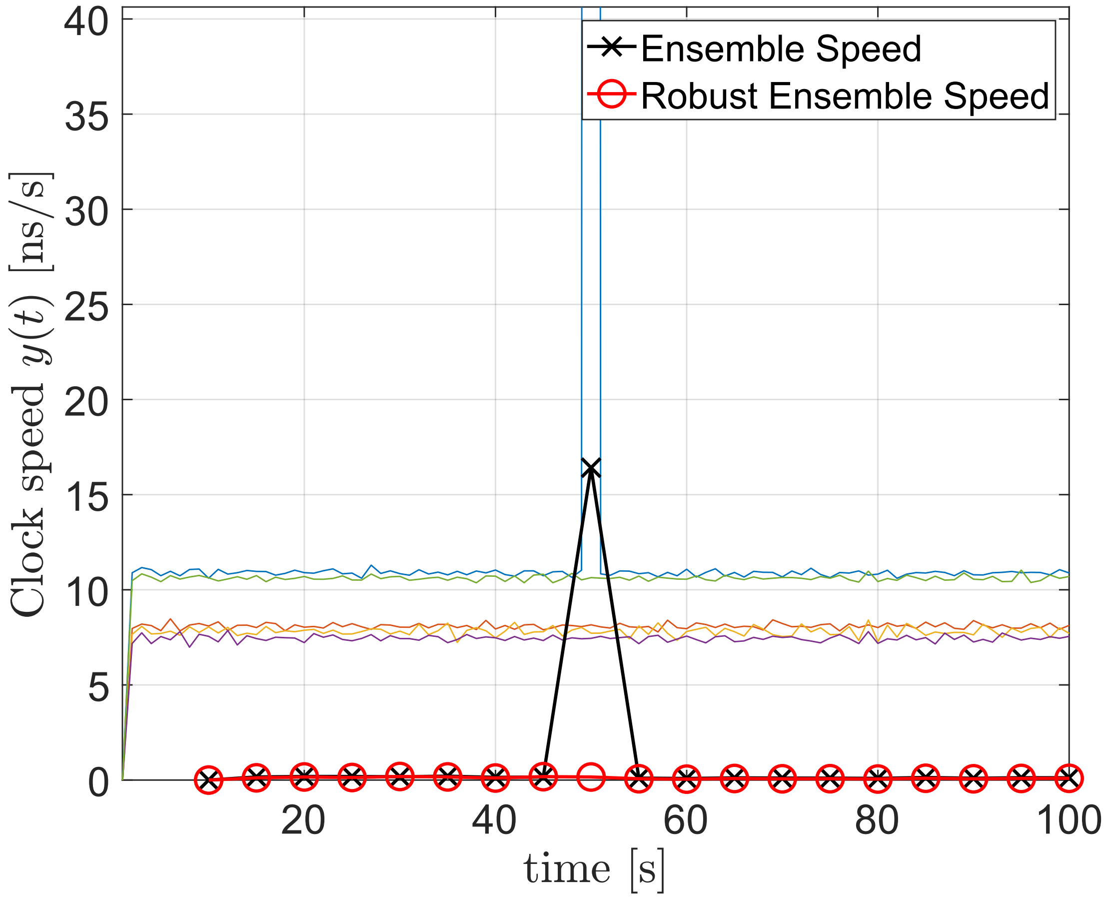


Time Scale performance – 1 μs jump

Absolute Phase



Absolute Frequency



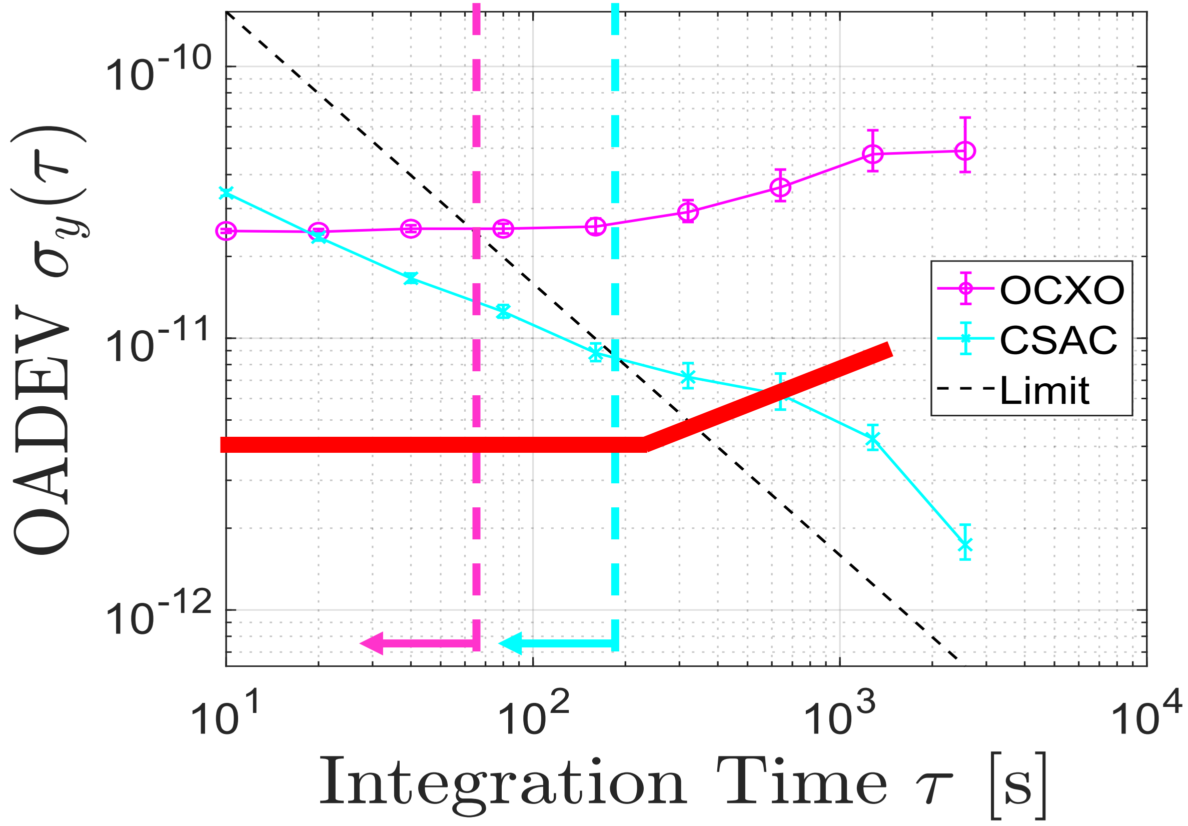
Time Scale performance - Stability

- Overlapping Allan Deviation (OADEV)
- Stability constraint:

$$f_{obs} = 100 \text{ MHz}$$

$$\tau\sigma_y(\tau) \leq \frac{1}{2\pi f_{obs}}$$

- Time scale improves range



Conclusion

- Time scales for distributed systems
- Misspecified Estimation
- Robust estimators

Time Scale Exercise

Scan the QR code with your phone or go to
https://perso.tesa.prd.fr/hmphee/timer.html

Enter your name and click **submit**,

Your task is to time the appearances of the newly discovered alien!

After pressing **Start**, watch the video on your screen.

Press **Time** as soon as you see a new Alien

Time Scale Exercise

Let’s look at your timing performance!
https://perso.tesa.prd.fr/hmphee/all_times.html

11:20 –

Questions

Estimating ν

To update the estimate of the number of degrees of freedom, find the zero of:

$$f(\hat{\nu}_k) = N\psi\left(\frac{\hat{\nu}_k}{2}\right) - N\log\left(\frac{\hat{\nu}_k}{2}\right) - N\psi\left(\frac{\hat{\nu}_{k-1}+1}{2}\right) - N + \sum_{j=1}^{N} u_j + \log\left(\frac{1}{2}\left(\hat{\nu}_{k-1} + \frac{\left(z_j - \hat{\mu}_{k-1}\right)^2}{\hat{\sigma}_{k-1}^2}\right)\right) \quad \text{(A1)}$$

$$\psi(x) = \frac{\Gamma'(x)}{\Gamma(x)}$$

Newton’s method is used to iteratively converge to the zero:

$$\hat{\nu}_{k,n+1} = \hat{\nu}_{k,n} - \frac{f\left(\hat{\nu}_{k,n}\right)}{f'\left(\hat{\nu}_{k,n}\right)}$$

Closed Form MCRBs

General derivation of $\boldsymbol{\theta}_{p||p_G} = \left[\mu_{p|p_G}, \sigma^2_{p|p_G}\right]^T = \left[E_p[z_i], \mathrm{var}_p[z_i]\right]^T$

$$\mathbf{MCRB}_{\boldsymbol{\theta}}(p_T||p_G) = \begin{bmatrix} \frac{\sigma^2_{p_T}}{N} & 0 \\ 0 & \left(\frac{\nu-1}{\nu-4}\right)\frac{2\sigma^4_{p_T}}{N} \end{bmatrix}$$

$$\mathbf{MCRB}_{\boldsymbol{\theta}}(p_{GM}||p_G) = \begin{bmatrix} \frac{\sigma^2_{p_{GM}}}{N} & 0 \\ 0 & \left(\frac{Q(\varepsilon(k-1))}{2(\varepsilon(k-1)+1)^2}\right)\frac{2\sigma^4_{p_{GM}}}{N} \end{bmatrix}$$

Misspecified Heavy-tails

Pseudo-true location parameter is equal to true location

$$\mu_{p_T||p_{GM}} = E_{p_T}[z], \qquad \mu_{p_{GM}||p_T} = E_{p_{GM}}[z],$$

Pseudo-true scale parameter requires numerical methods to solve minimum KLD

MCRB not available in closed form

Joint estimation of Student's t parameters

Jointly estimating μ, σ^2, ν:

$$\mathrm{CRB}_{\sigma^2} = 2\sigma^4 \frac{(\nu + 3)}{\nu}\left(1 - \frac{4\nu}{f(\nu) + 4\nu}\right)$$

$$f(\nu) = 2\nu^2(\nu + 1)^2(\nu + 3)\xi(\nu) + (\nu + 5)(\nu + 1)\nu$$

$$\xi(\nu) = \frac{1}{4}\left(\psi'\left(\frac{\nu + 1}{2}\right) - \psi'\left(\frac{\nu}{2}\right)\right), \qquad \psi'(x) = \frac{\partial^2}{\partial x^2}\log\bigl(\Gamma(x)\bigr)$$

Fixing ν, and estimating μ, σ^2

$$\mathrm{CRB}_{\sigma^2} = 2\sigma^4 \frac{(\nu + 3)}{\nu}$$