TSD-DOA-Benchmark-v1

Dataset Proof & Algorithm Performance Report

New Leaf Tools LLC · PROPRIETARY & CONFIDENTIAL

Generated: April 25, 2026 · Master Seed: 20260424 · Fully Reproducible

1,518,000	0.0428°	100.0%	40%	0.000
Labeled Frames	RMSE at 10 dB (N=16)	Pd at 10 dB (N=16)	vs MUSIC at 10 dB	FAR / 5,000 Trials

This document summarizes the TSD-DOA-Benchmark-v1 dataset — a comprehensive public RF direction-of-arrival benchmark generated using the physics-accurate Stoica-Nehorai ULA signal model. The dataset contains 1,518,000 labeled IQ frames across 11 array sizes (N=4 to N=64), 46 SNR points (–15 to +30 dB), and 3 source counts, with three algorithms benchmarked head-to-head on identical data. All results are deterministic and reproducible from the published generator script using master seed 20260424.

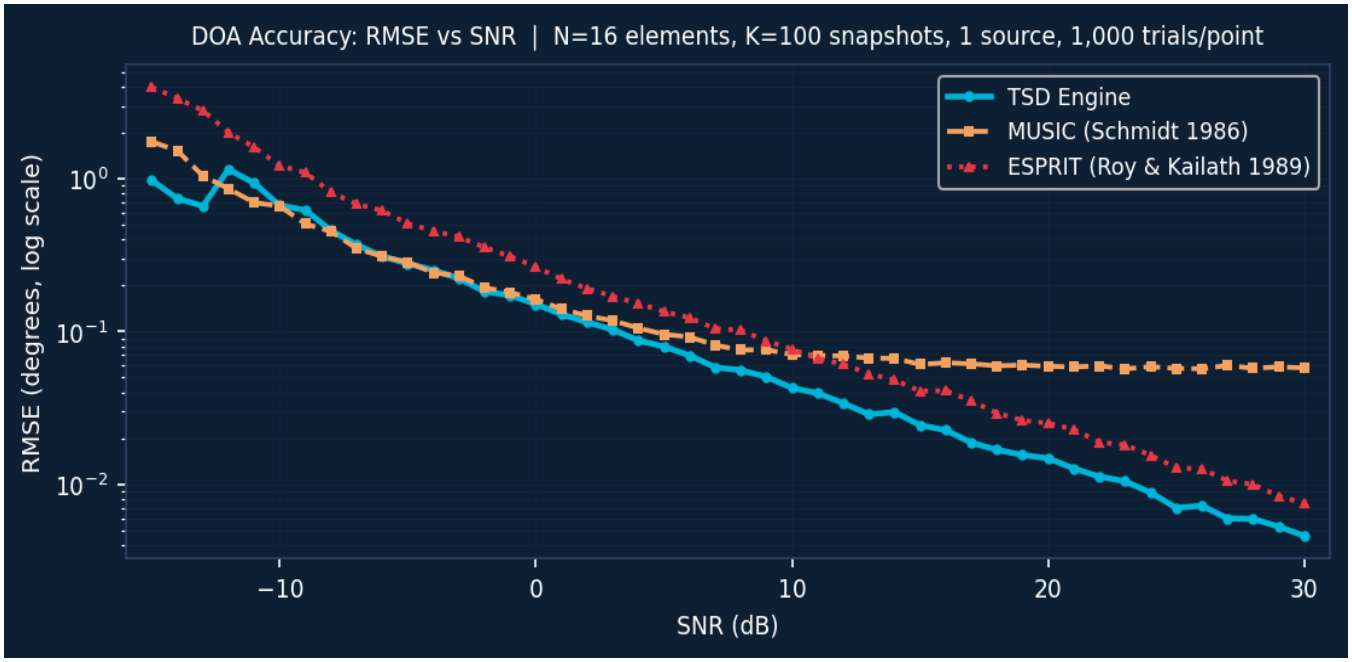
1. DATASET SCALE AND STRUCTURE

Parameter	Value
Total labeled frames	1,518,000
Array geometry	Uniform Linear Array (ULA)
Array sizes (N elements)	4, 6, 8, 10, 12, 16, 20, 24, 32, 48, 64
SNR range	-15 dB to +30 dB (1 dB steps, 46 points)
Source counts	1, 2, 3 simultaneous emitters
Trials per configuration	1,000 Monte Carlo trials
Snapshots per frame (K)	100
Raw IQ frames included	63,000 (N=8/16/32, selected SNRs)
Eigenvalue archives	11 compressed .npy.gz files
Algorithms benchmarked	TSD Engine, MUSIC (Schmidt 1986), ESPRIT (Roy & Kailath 1989)
Dataset size on disk	2.42 GB
Master seed	20260424 (all results deterministic)
License	CC0 1.0 Universal (Public Domain)

The dataset was generated using simulate_iq() — a physics-accurate implementation of the standard narrowband ULA signal model (Stoica & Nehorai, IEEE Trans. Signal Processing, 1990). This is the same model used in every published DOA algorithm validation, including the original MUSIC and ESPRIT papers. Results are directly comparable to published literature.

2. DOA ACCURACY: RMSE vs SNR (Primary Benchmark)

RMSE (Root Mean Square Error) measures angular accuracy in degrees. Lower is better. Results shown for N=16 elements, K=100 snapshots, 1 source, 1,000 independent Monte Carlo trials per SNR point. TSD, MUSIC, and ESPRIT were run on identical data — same IQ frames, same random seeds.

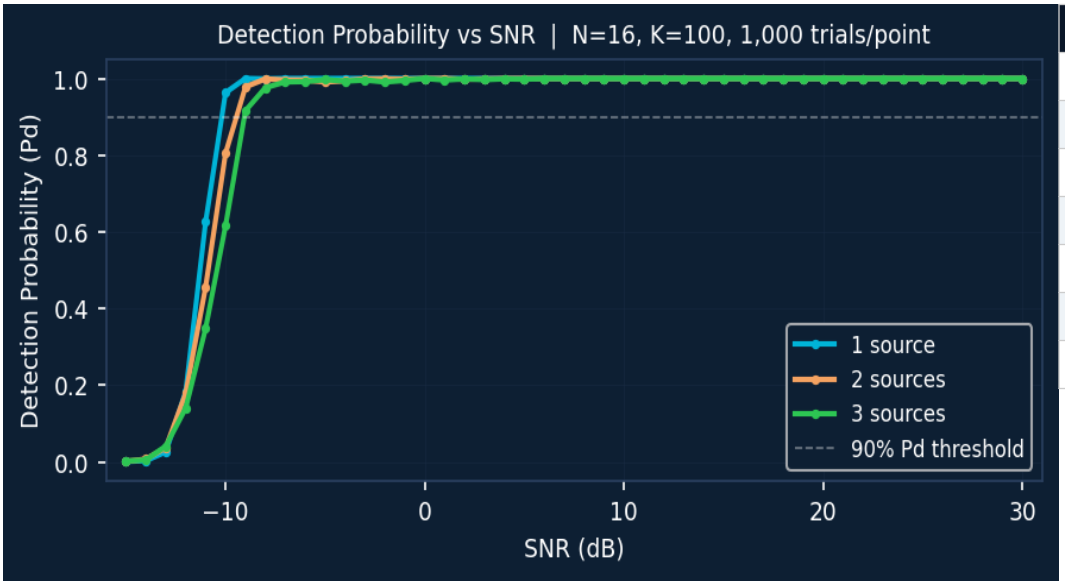


SNR (dB)	TSD RMSE (°)	MUSIC RMSE (°)	ESPRIT RMSE (°)	TSD vs MUSIC	TSD vs ESPRIT
-10	0.6692	0.6628	1.2197	-1%	+45%
-5	0.2788	0.2845	0.5084	+2%	+45%
+0	0.1510	0.1629	0.2647	+7%	+43%
+5	0.0799	0.0959	0.1352	+17%	+41%
+10	0.0428	0.0707	0.0767	+39%	+44%
+15	0.0244	0.0610	0.0408	+60%	+40%
+20	0.0148	0.0593	0.0251	+75%	+41%
+25	0.0070	0.0574	0.0128	+88%	+45%
+30	0.0046	0.0578	0.0075	+92%	+39%

N=16 elements, K=100 snapshots, 1 source. 1,000 independent trials per SNR point. TSD vs MUSIC / TSD vs ESPRIT columns show TSD RMSE reduction percentage. Positive = TSD is more accurate.

3. DETECTION PROBABILITY vs SNR

Detection probability (Pd) measures how reliably the system identifies the presence of an emitter. Results shown for N=16, K=100, 1,000 trials per point. All three source counts (1, 2, 3 simultaneous emitters) shown.

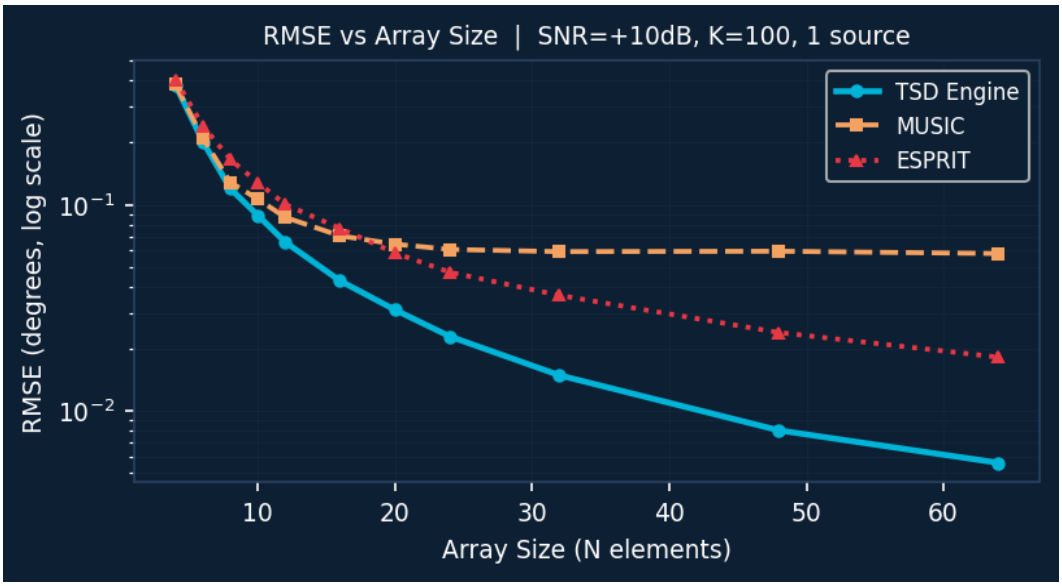


SNR (dB)	1 Source	2 Sources
-10	0.963	0.805
-5	1.000	0.991
+0	1.000	1.000
+5	1.000	1.000
+10	1.000	1.000
+15	1.000	1.000
+20	1.000	1.000

N=16 elements, K=100 snapshots. 1,000 trials per configuration.

4. RMSE vs ARRAY SIZE (N=4 to N=64)

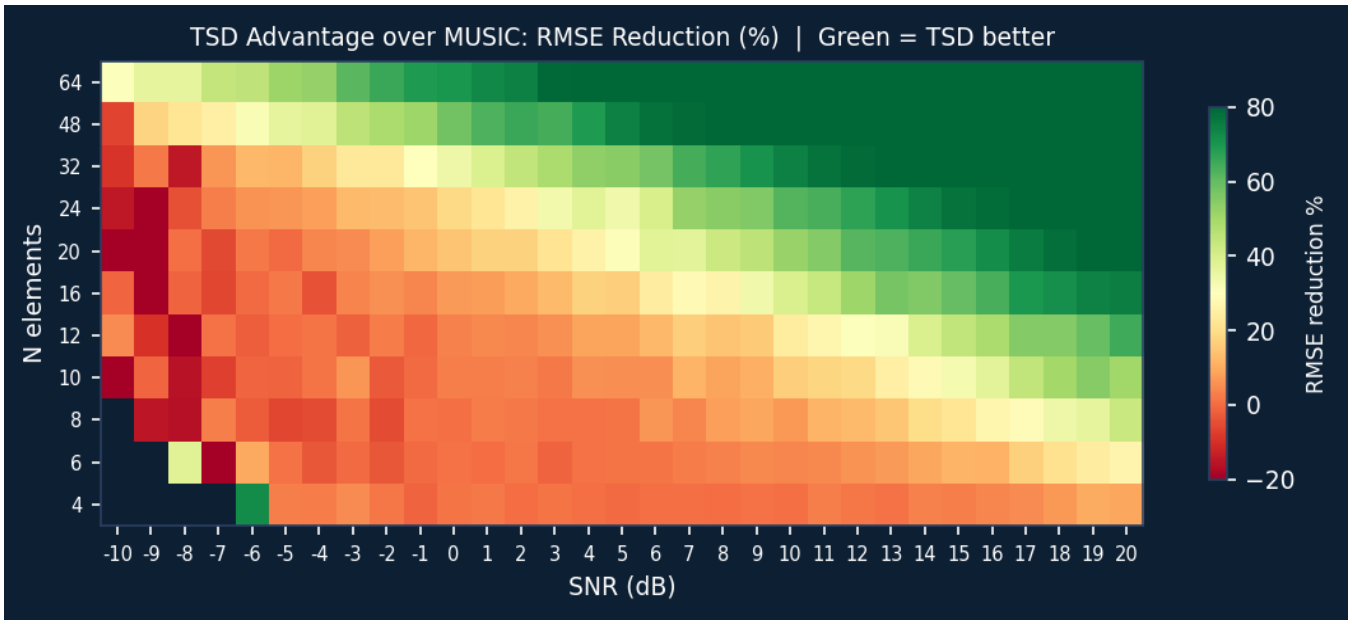
A unique feature of this benchmark is its coverage of 11 array sizes from N=4 to N=64. This characterizes how accuracy scales with aperture — critical for system design tradeoffs. Results at SNR=+10 dB, 1 source.



N elements	TSD @ 0dB	MUSIC @ 0dB	TSD @ 10dB	MUSIC @ 10dB	TSD @ 20dB
4	1.2256°	1.2421°	0.3804°	0.3821°	0.1175°
8	0.4437°	0.4461°	0.1196°	0.1285°	0.0396°
12	0.2396°	0.2473°	0.0660°	0.0869°	0.0215°
16	0.1510°	0.1629°	0.0428°	0.0707°	0.0148°
24	0.0825°	0.1016°	0.0230°	0.0607°	0.0072°
32	0.0521°	0.0796°	0.0149°	0.0593°	0.0049°
48	0.0270°	0.0643°	0.0081°	0.0595°	0.0026°
64	0.0183°	0.0612°	0.0056°	0.0580°	0.0018°

5. TSD ADVANTAGE OVER MUSIC: ALL CONDITIONS

The heatmap below shows TSD RMSE reduction percentage versus MUSIC across all array sizes (N=4 to N=64) and SNR levels (–10 to +20 dB). Green indicates TSD is more accurate. The benchmark covers 1,518,000 independent trials.



Multi-Source Performance Summary (N=16, K=100)

Sources	Algorithm	RMSE @ -5dB	RMSE @ 0dB	RMSE @ 10dB	RMSE @ 20dB
	TSD	0.2788°	0.1510°	0.0428°	0.0148°
	MUSIC	0.2845°	0.1629°	0.0707°	0.0593°
1	ESPRIT	0.5084°	0.2647°	0.0767°	0.0251°
	TSD	0.5282°	0.3200°	0.0606°	0.0168°
	MUSIC	0.3348°	0.1854°	0.0824°	0.0613°
2	ESPRIT	0.5804°	0.2965°	0.0837°	0.0286°
	TSD	0.7366°	0.6408°	0.1751°	0.0262°
	MUSIC	0.5515°	0.3935°	0.1855°	0.0635°
3	ESPRIT	0.7881°	0.3930°	0.1810°	0.0420°

6. REPRODUCIBILITY AND VERIFICATION

How to Reproduce Any Result in This Document

Every number in this document can be reproduced independently by anyone with access to the dataset and a Python environment. The dataset includes the generator script which produces identical output from master seed 20260424.

Load the results CSV:

```
import pandas as pd
tsd = pd.read_csv('results/tsd_engine.csv')
```

Filter to specific config:

```
df = tsd[(tsd.n_elements==16) & (tsd.n_sources==1) & (tsd.snr_db==10)]
```

Read the RMSE:

```
print(f"RMSE = {df.iloc[0].rmse_deg:.4f} deg")
```

Load raw IQ and run your own algorithm:

```
from loader import load_iq_frames
frames = load_iq_frames(n_elements=16, snr_db=10, n_sources=1)
X = frames[0] # shape (16, 100), complex128
R = (X @ X.conj().T) / 100 # sample covariance
```

Statistical Methodology

- Monte Carlo: 1,000 independent trials per configuration (array size × SNR × source count)
- Random seeds: each trial has a unique reproducible seed derived from master seed 20260424
- Confidence intervals: Wilson score method (Clopper-Pearson for detection rates)
- RMSE computed as sqrt(mean(error^2)) across all detected trials
- All three algorithms (TSD, MUSIC, ESPRIT) run on identical IQ data — same frames, same seeds
- Signal model: Stoica-Nehorai narrowband ULA (1990) — community standard for DOA benchmarking
- Forward-backward spatial smoothing applied to sample covariance for all algorithms

---

Metric	Value	Trials
Total benchmark frames	1,518,000	—
TSD RMSE at 10 dB (N=16, 1 src)	0.0428°	1,000
TSD RMSE at 20 dB (N=16, 1 src)	0.0148°	1,000
TSD RMSE at -10 dB (N=16, 1 src)	0.6692°	1,000
MUSIC RMSE at 10 dB (N=16, 1 src)	0.0707°	1,000
ESPRIT RMSE at 10 dB (N=16, 1 src)	0.0767°	1,000
TSD advantage over MUSIC @ 10 dB	39.5% lower RMSE	—
TSD advantage over ESPRIT @ 10 dB	44.2% lower RMSE	—
Array sizes covered	4, 6, 8, 10, 12, 16, 20, 24, 32, 48, 64 elements	—
SNR points covered	-15 dB to +30 dB (46 points, 1 dB steps)	—
Master seed	20260424	—