High Resolution Vector Processing and IOT in PostgreSQL

Jun 6, 2018, Dallas, Texas, USA

John Scott
Founder, SetSpace, Inc, 1998-now

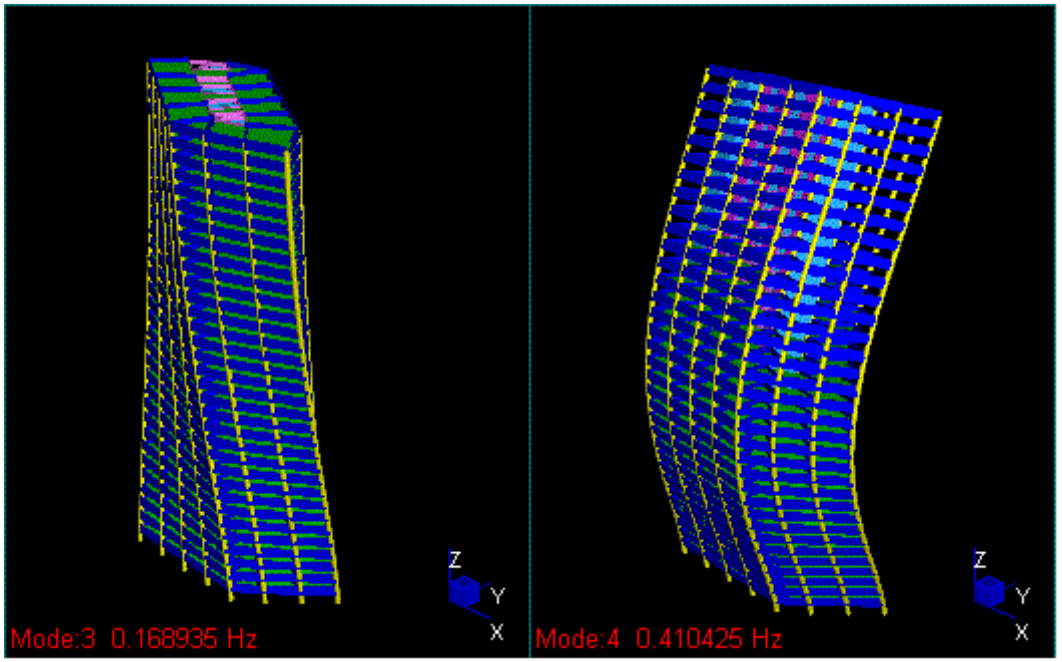
June 6, 2018
Dallas, Texas, USA

Tacoma Bridge Collapse in Low Winds

Tacoma Narrows Bridge Collapse "Gallopin' Gertie"

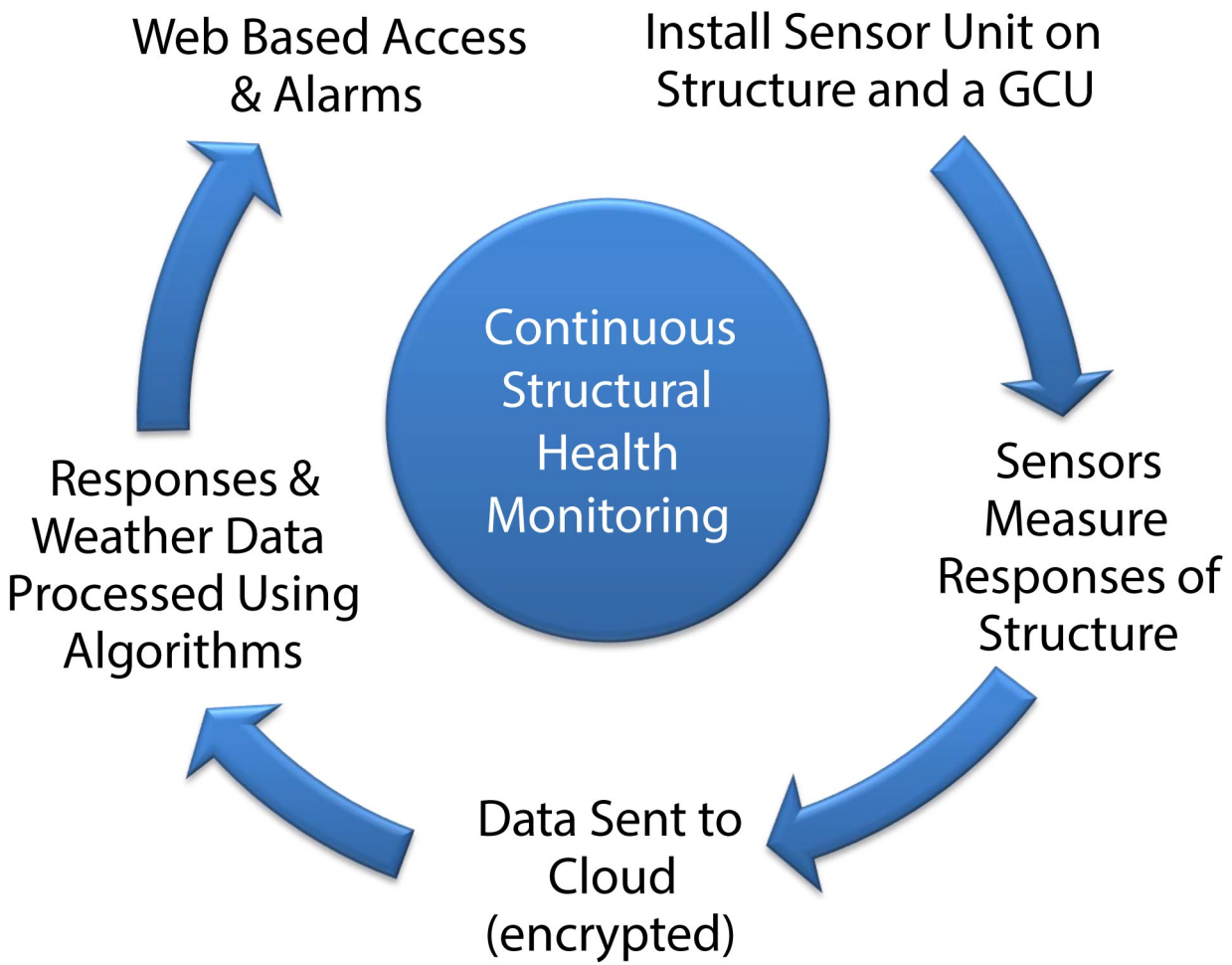
Video of Tacoma Bridge Collapse (https://youtu.be/lXyG68_caV4?t=61)

Flexible Structures Vibrate at Peculiar Frequencies

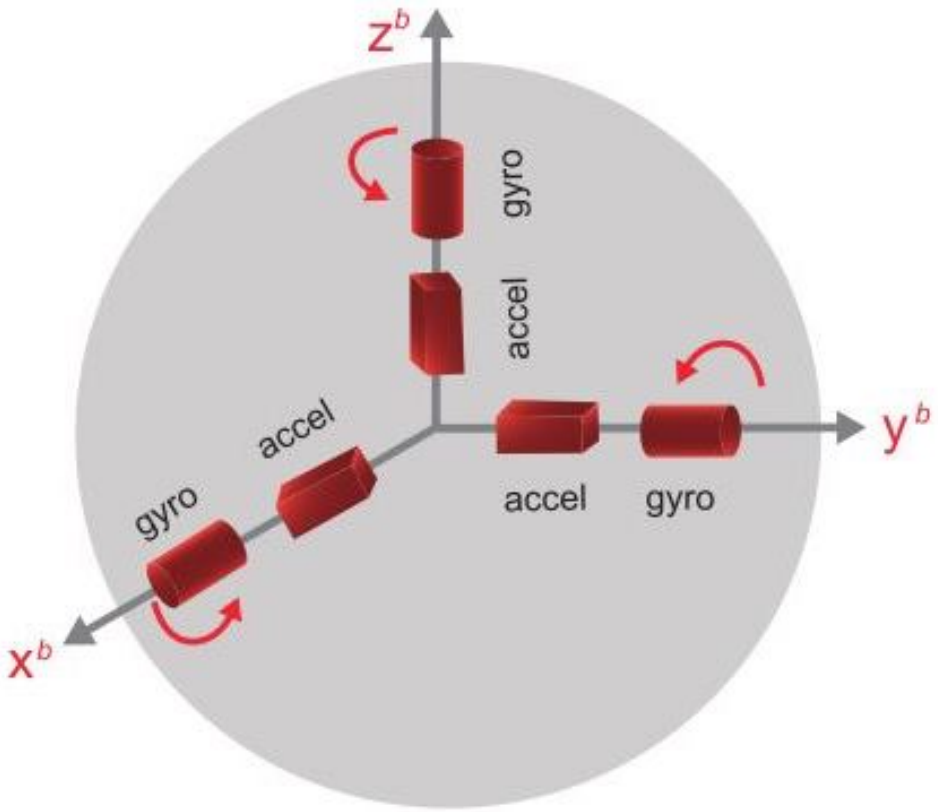


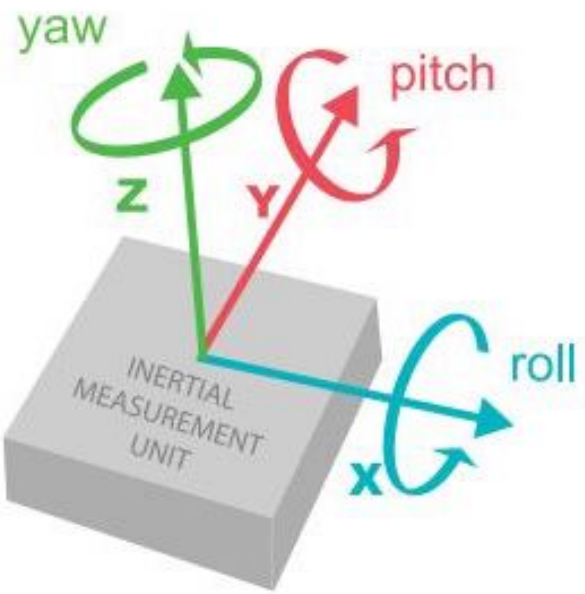
Small Energy Stress Certain Points, Like a Coat Hanger

Cycle of Monitoring of Flexible Structure



Cheap Inertial Measurement Units (IMU) Changed the World

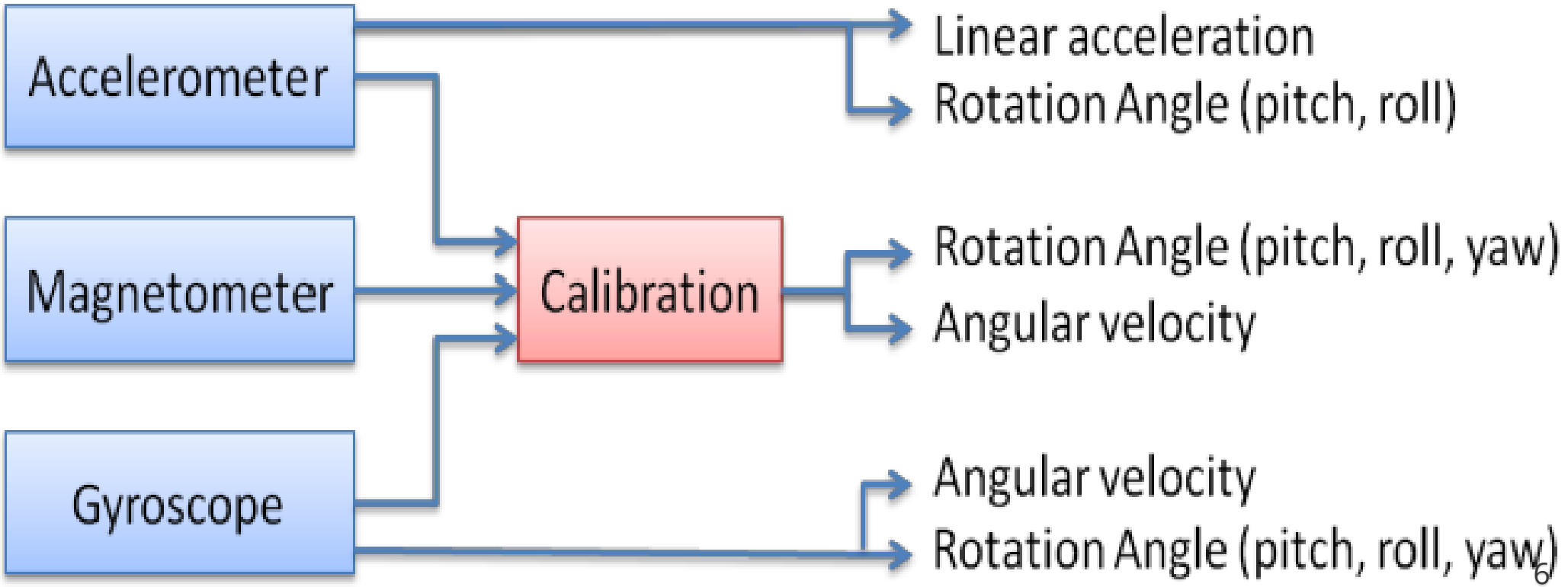




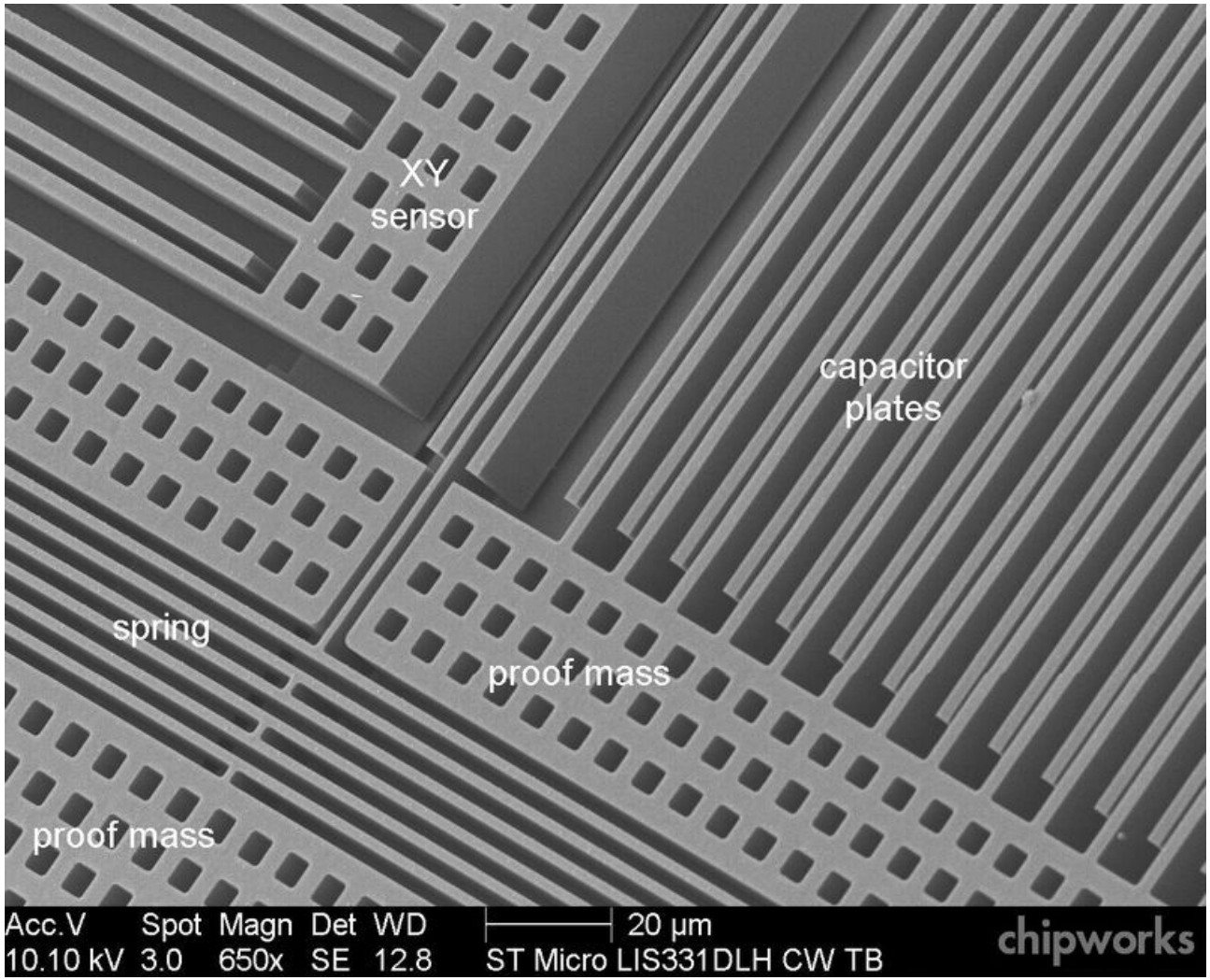
Inertial Measurement unit
3 accelerometers, 3 gyroscopes

Micro-Electro Mechanics Sensor (MEMS) In Smartphones, FitBit, Drones, Aircraft

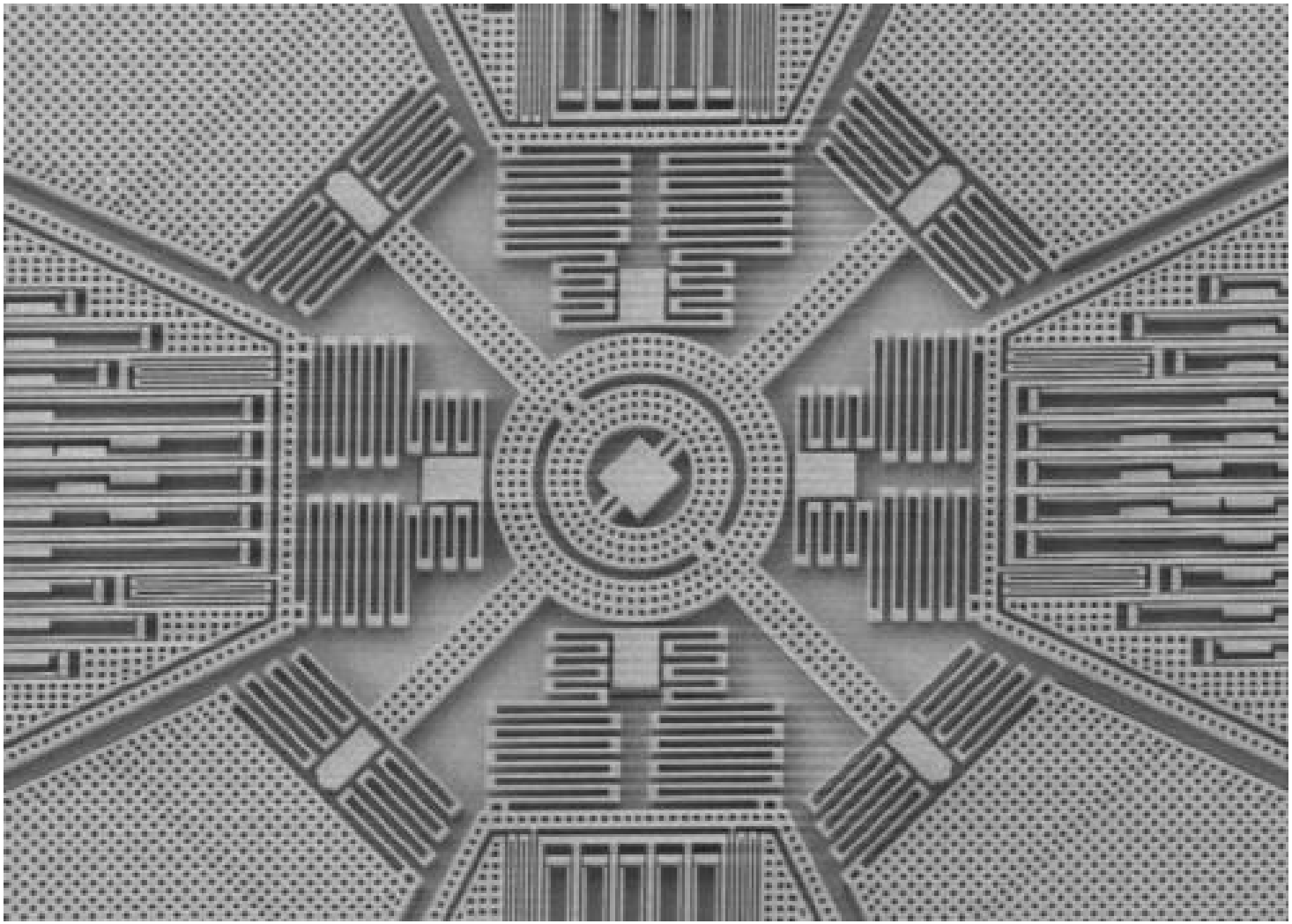
Inside an IMU - Flow of Micro Electronic Mechanical Sensors (MEMS)



Linear Accelometer (MEMS) Viewed in Electron Microscope!



Angular Accelometer (MEMS) Viewed in Electron Microscope!



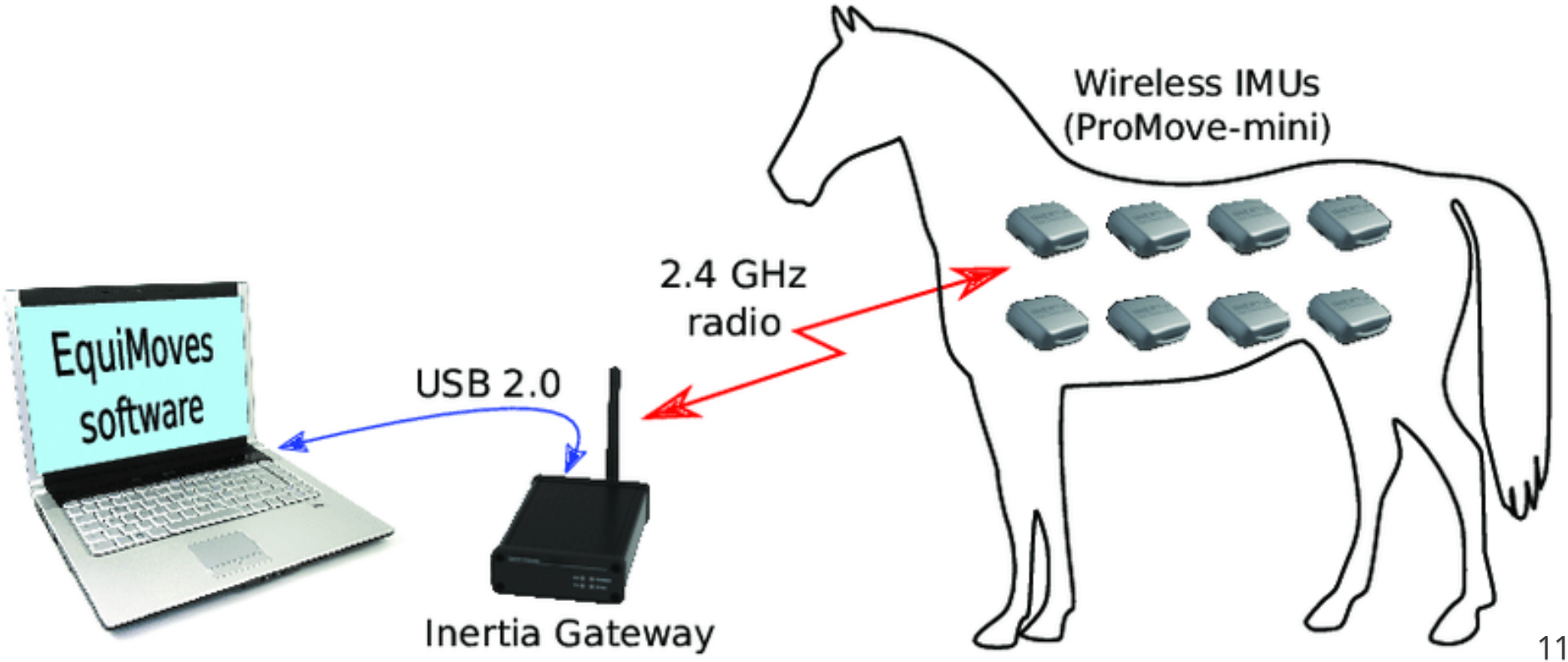
VectorNav 300 with Gravitational and Magnetic Map of Earth

Ruggedized with 2 GPS External Antenna for Recalibrating the Integrated Kalman Filter

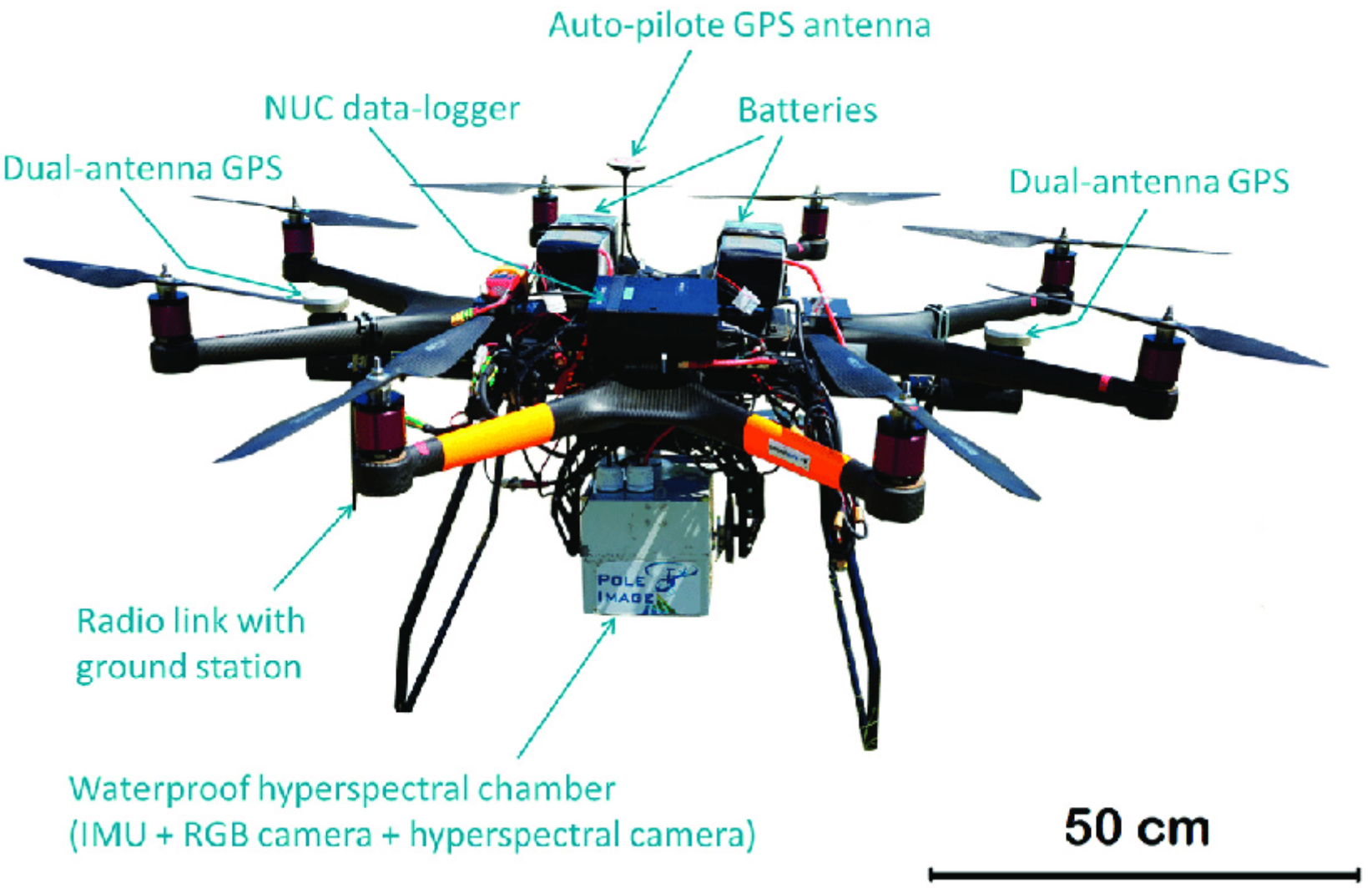
VectorNac 300 Fits in Your Hand - $800

Made in Dallas, Texas!

Inertial Measurement Units in Everything!



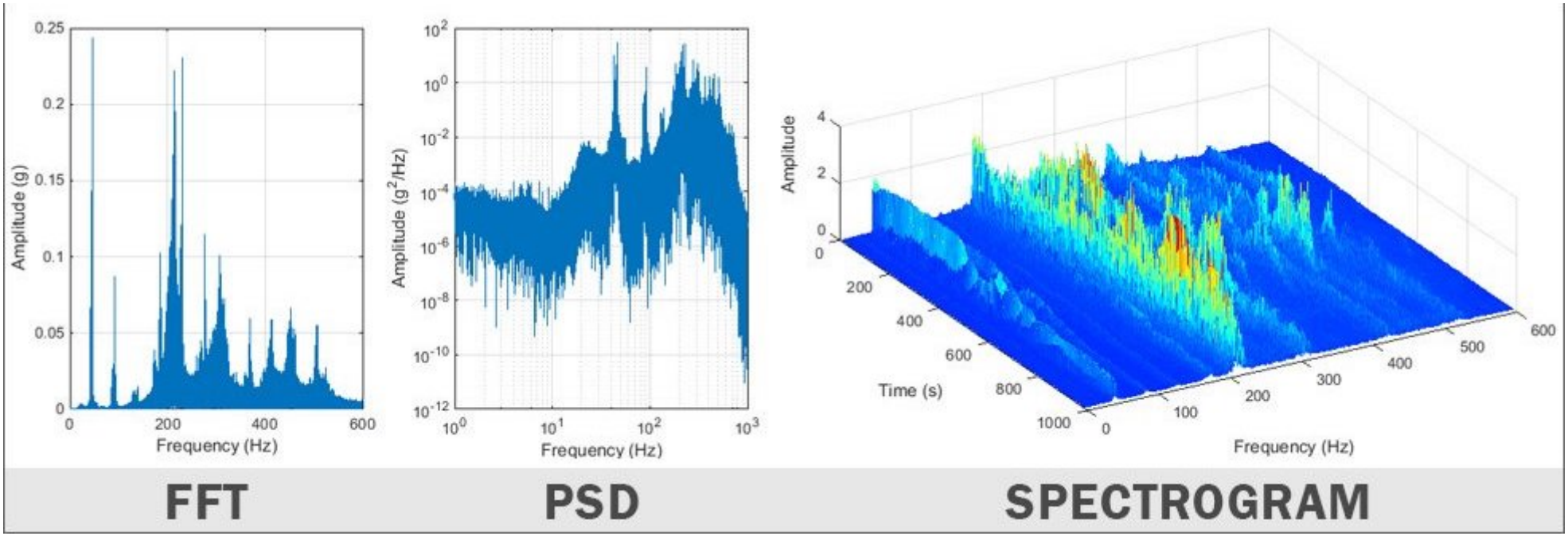
Drones has Many IMUs + GPS



Why Monitor Flexible Structures?

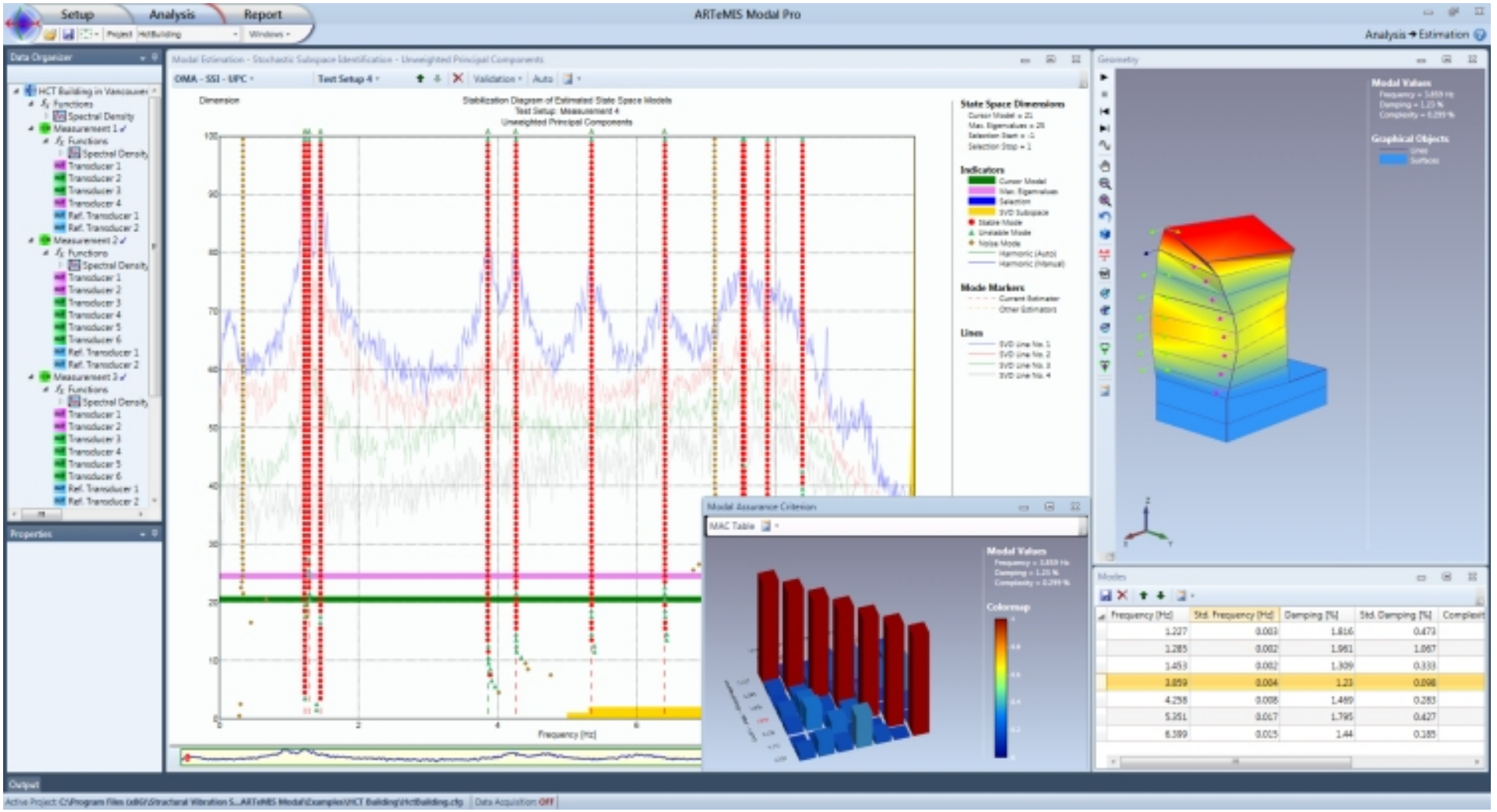
- Alert Owner of Dramatic Movement of Bridge
- Twist, Sway and Motion Stress on Structure
- Slow Moving Oscillations - Coat Hanger Bending
- Prioritize Maintence After Extreme Weather
- Detect Permanent Deformations
- Insurance Benefits

Modal Analysis and Movement Done with Fast Fourier Transform



Important - All Analysis Done on 1024 Sample Vectors !!

Bread and Butter of Vibrational/Modal Analysis



Case Study - High Resolution Samples of 10 Vectors/Second/Sensor

- Our Customer is a Market Maker
- Oscillations up to 5 Times/Second (Nyquist Theorem)
- 1.5GB/month per sensor, uncompressed
- At Least Two Sensors per Structure
- Speced to 100,000 sensors -> 145 TB/month, uncompressed

YMR Samples Arrive from Bridge as a CSV File (via rsync)

```
ARCHIVE=archive/2017/10/08
BRIDGE=bridge/gcu01.st.lab.rj2tech.com
SENSOR=vn100-001
DAY=20171008_054649

FILE=$ARCHIVE/$BRIDGE/$SENSOR-$DAY.row.bz2

bzcat $FILE
```

```
sample_time yaw pitch roll ax ay az gx gy gz mx my mz

2017-10-08T05:46:49.171969067-05:00 -93.155 -4.702 1.571 -.775 -.259 -9.517 .000151 -.000056 -.000678 .0419 .2295 .4134

2017-10-08T05:46:49.268936247-05:00 -93.158 -4.703 1.570 -.777 -.252 -9.499 -.000214 .000166 -.000028 .0430 .2306 .4134

2017-10-08T05:46:49.371055098-05:00 -93.160 -4.703 1.569 -.794 -.248 -9.504 .001882 .001253 .000470 .0396 .2317 .4135

2017-10-08T05:46:49.473195095-05:00 -93.160 -4.701 1.571 -.774 -.254 -9.515 -.000170 -.000501 .000594 .0419 .2295 .4146

2017-10-08T05:46:49.568933484-05:00 -93.155 -4.704 1.567 -.794 -.249 -9.515 -.000689 -.000223 -.000024 .0431 .2340 .4160

...
```

1024 * 3 samples (rows) == 5.1min in single log segment, about 64kbytes

Why Use PostgreSQL Instead of MongoDB/Spark?

- SQL Best for Complex Historical Patterns in Data
- PostgreSQL Easy to Extend to File Store for Samples (FDW)
- PostgreSQL Allow New Data Types - Vector Can be Hidden in File Store
- Materialized Views of MATLAB Calculations
- A/B Comparison (Profiles) of Motion and Mode Algorithms

```
We Always Have CitusDB!
```

Vector Sample Log Table - bridge_log_vnymr_10hz

- Stores 5.1min @ 10hz == 3072 Tuples
- Key is Surrogate Sensor Id and Start Time for Each of 3, 1024 Segments
- Single PG Vector/Array for Sample Times
- Single PG Vector/Array for 12 Dimensional Floating Point Samples
- Compress Very Well

SQL for Sample Times in Table bridge_log_vnymr_10hz

```
CREATE TABLE bridge_log_vnymr_10hz

...

    start_time    timestamptz            --  sample_time[1]
                NOT NULL,

    sample_time    timestamptz[3072] CHECK (
                sample_time[1] = start_time
                AND
                array_length(sample_time, 1) = 3072
                AND

                --  Note:  need to verify all times increasing!

                sample_time[1] < sample_time[3072]

             ),    --  null means samples offline
...
```

SQL for 1024 * 3 * 12 YMR Samples in Table bridge_log_vnymr_10hz

```
/*
 *  samples[][] stores 3072 vectors in 12 dimensions:
 *
 *    [1] =    yaw
 *    [2] =    pitch
 *    [3] =    roll
 *    [4] =    acc_x
 *    [5] =    acc_y
 *    [6] =    acc_z
 *    [7] =    gyro_x
 *    [8] =    gyro_y
 *    [9] =    gyro_z
 *    [10] =    mag_x
 *    [11] =    mag_y
 *    [12] =    mag_z
 */
ymr    real[3072][12] CHECK (
            array_length(ymr, 1) = 3072
            AND
            array_length(ymr, 2) = 12
        ),        --  null means samples may be offline
```

Vectors Compress Well but Tricky in SQL So Unpack with a View

```
CREATE VIEW vnymr_10hz_acc AS
  SELECT
        st.sensor_id,
        st.start_time,   ymr_st.sample_time,
        ymr_x.acc_x,     ymr_y.acc_y,       ymr_z.acc_z
    FROM
        tower_log_vnymr_10hz_sample_time st
          JOIN tower_log_vnymr_10hz_sensor sen ON (
                sen.sensor_id = st.sensor_id
                AND
                sen.start_time = st.log_start_time
          )
          JOIN tower_log_vnymr_10hz ymr ON (
                ymr.tower = sen.tower
                AND
                ymr.drop_id = sen.drop_id
                AND
                ymr.start_time = sen.start_time
          )

          --  sample time projection
          JOIN LATERAL unnest(
                ymr.sample_time[st.sample_offset:st.sample_offset + 1023]
            ) WITH ORDINALITY AS ymr_st(sample_time, row) ON (
                TRUE
            )

          --  x acceleration projection
          JOIN LATERAL unnest(
                ymr.ymr[st.sample_offset:st.sample_offset + 1023][4:4]
            ) WITH ORDINALITY AS ymr_x(acc_x, row) ON (
                ymr_x.row = ymr_st.row
            )
```

Unpacked View is Easy to Query

```
/*
 *  When is sensor #12 on the bridge accelerating more than one meter/sec
 *  during the last 7 days?
 */

SELECT
    MIN(sample_time) AS alarm_start_time,
    MAX(sample_time) AS alarm_end_time
  FROM
      vnymr_10hz_acc
  WHERE
    sensor_id = 12
    AND
    sample_time >= now() + '-7 days'
    AND

    --  Sum of Squares - meters/sec^2

    SQRT(acc_x*acc_x  +  acc_y*acc_y  +  acc_z*acc_z) > 1
;
```

Materialize Views Store MATLAB Results

```
/*
 *  X/Y/Twist/Sway/SwayAz Displacement by the VectorNav KF for VNYMR 10hz
 */
DROP TABLE IF EXISTS displace_vnkf_10hz CASCADE;
CREATE TABLE displace_vnkf_10hz
(
        sensor_id         sensor_id,
        profile           formal_name
                                REFERENCES vnymr_kalman_profile(name) ON DELETE CASCADE,
        start_time        timestamptz,

        x                       real[] CHECK (array_length(x, 1) = 1024),
        y                       real[] CHECK (array_length(y, 1) = 1024),
        twist                   real[] CHECK (array_length(twist, 1) = 1024),
        sway                    real[] CHECK (array_length(sway, 1) = 1024),
        swayaz                  real[] CHECK (array_length(swayaz, 1) = 1024),

        PRIMARY KEY       (sensor_id, profile, start_time),

        FOREIGN KEY       (sensor_id, start_time)
                          REFERENCES vnymr_10hz_order (
                                sensor_id,
                                start_time
                          ) ON DELETE CASCADE
);
```

Vectors Can Reference Raw CSV Files Using Foreign Data Wrappers

- Use Custom Data Type - Can Not Override array[] (JavaScript has same problem)
- Vectors < 3 Months in PostgreSQL
- Vectors > 3 and < Year in CSV File
- Null Otherwise

Storing Dense Vector Data Can be Done in PostgreSQL

Thank you

Jun 6, 2018, Dallas, Texas, USA

John Scott
Founder, SetSpace, Inc, 1998-now

jmscott@setspace.com (mailto:jmscott@setspace.com)

June 6, 2018
Dallas, Texas, USA