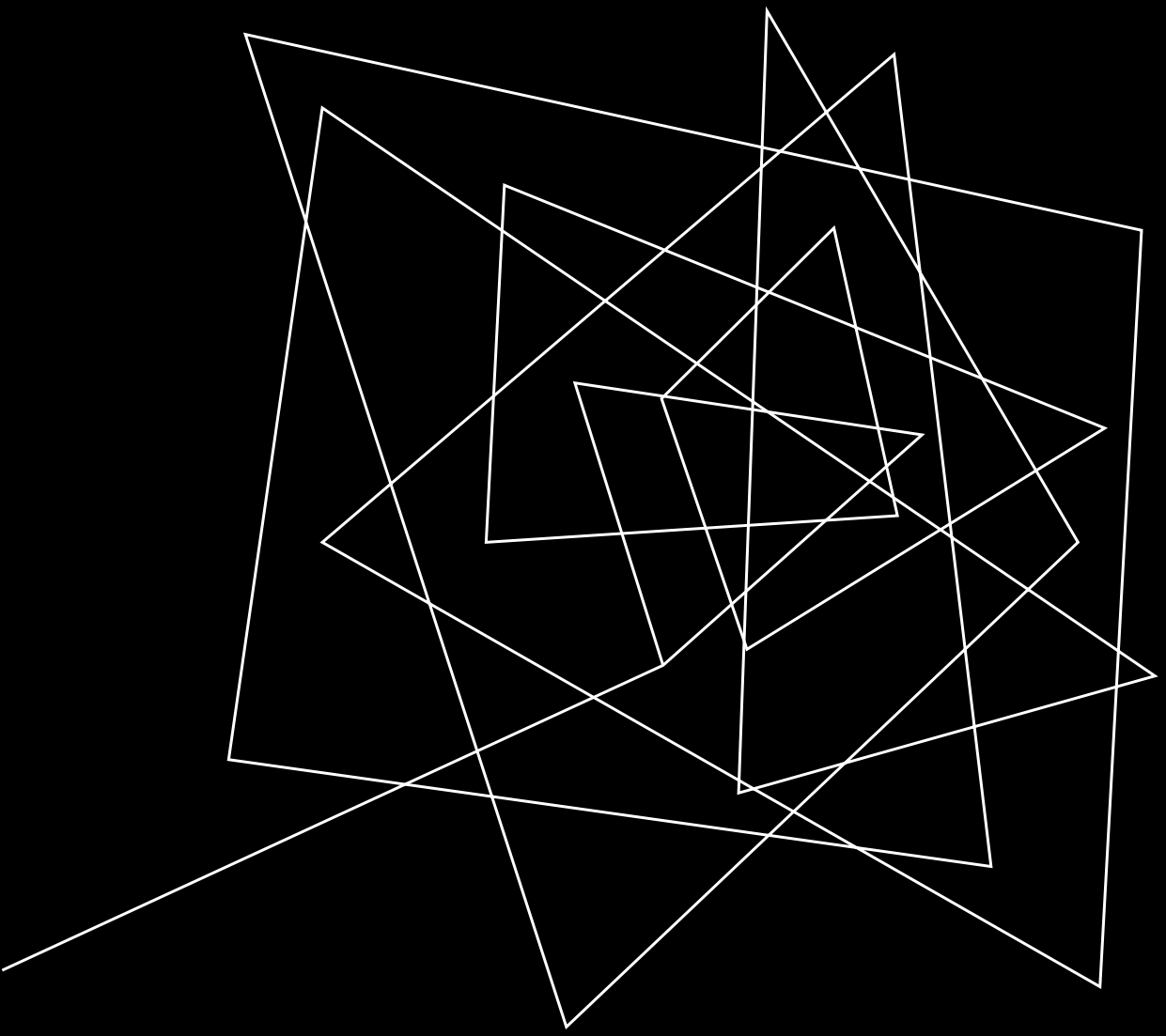
DATA SCIENCE PORTFOLIO

John Lingo

INTRODUCTION

I have been performing data mining, data analysis, and data visualization tasks for over 15 years as part of my role leading a contract engineering team that is responsible for developing, optimizing, and maintaining embedded firmware designs for high volume commercial products.



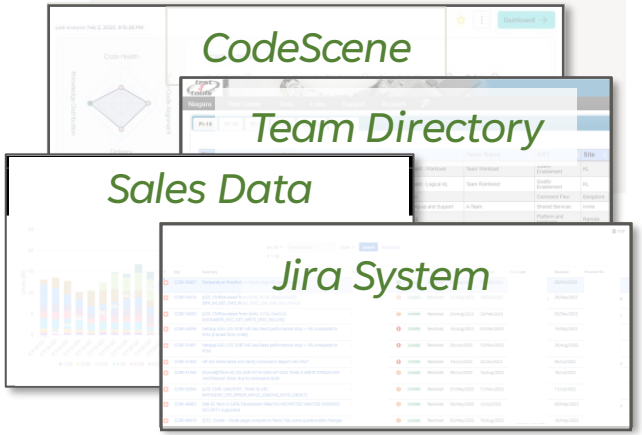
SAMPLE PROJECT – DATA MINING / ANALYSIS

Study of Relative Cost of Firmware Development at Three “Lower Cost” development sites

STUDY OF FIRMWARE DEVELOPMENT COST – METHODOLOGY

Data Mining

All data was extracted from available client databases/systems, to help assure client of study's objectivity



Data Cleaning

Numerous "filters" were utilized to eliminate "outlier" data points

Multiple actions were taken to account for "missing" information as accurately as possible

Data Analysis

Data was aggregated and processed via both Excel and Python

pandas

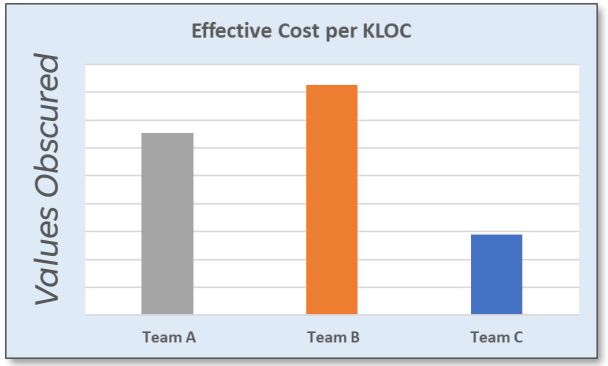
seaborn

Data Visualization

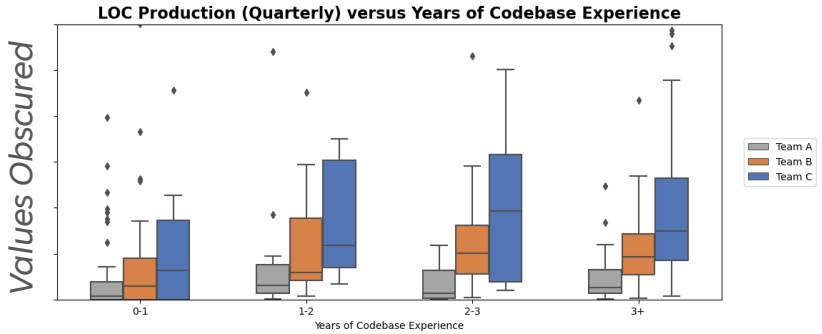
Results were graphed, then presented to the client in person using a PowerPoint slideshow

Note: Some information above has been intentionally obscured to ensure protection of client intellectual property

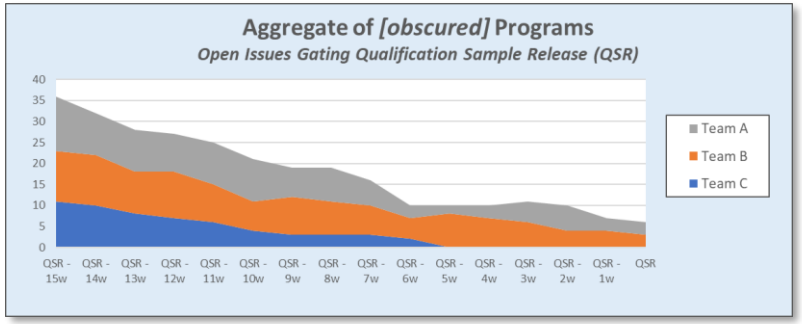
STUDY OF FIRMWARE DEVELOPMENT COST – RESULTS



Study demonstrated that "Team C" was producing code at a significantly lower overall cost to the client (even though the "Team C" designers had the highest average yearly cost to the client on an *individual* salary basis).

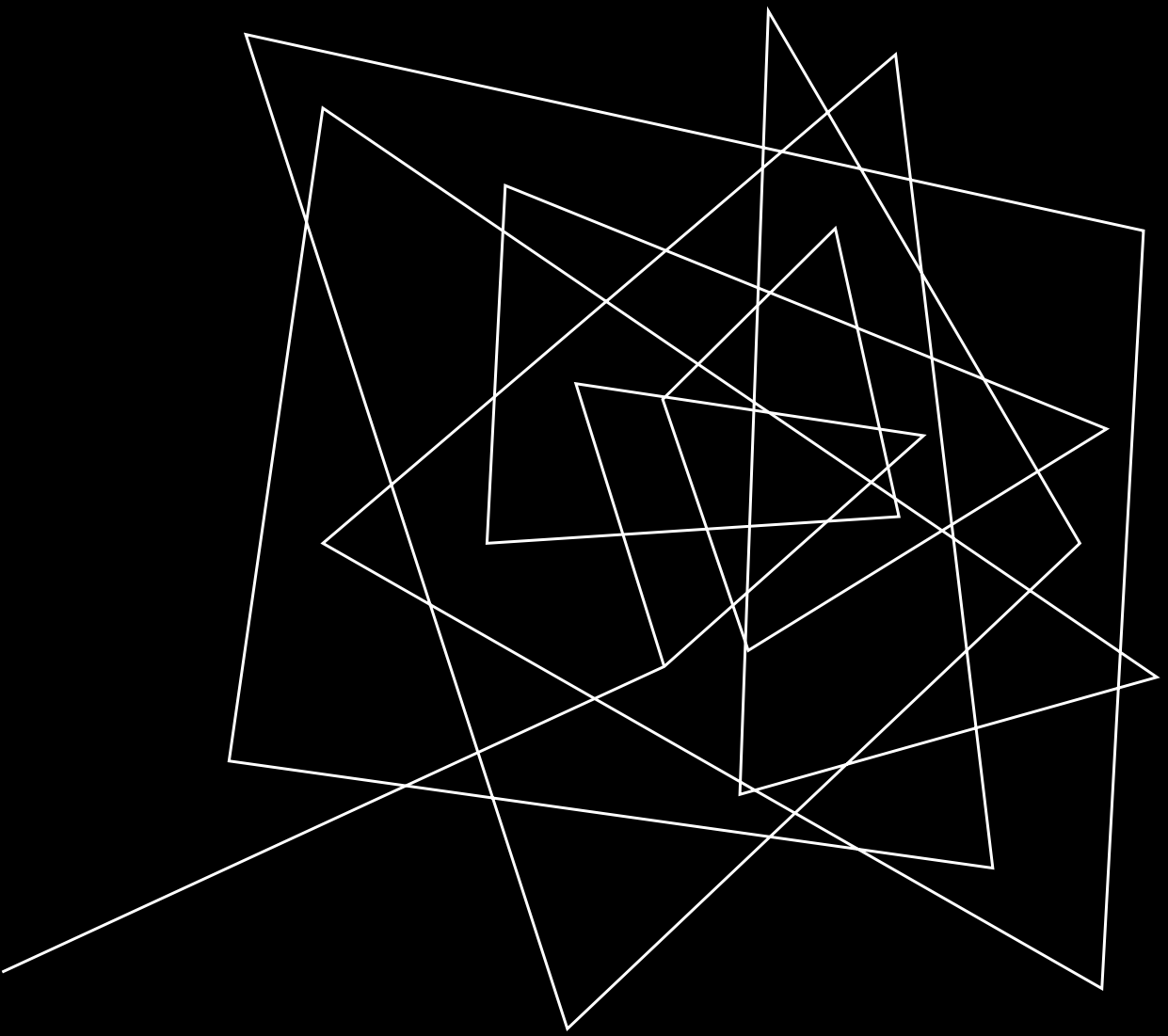


Study demonstrated that the higher productivity of "Team C" members existed even shortly after the time a designer joined the team, and this productivity advantage persisted over time.



Study identified that across multiple recent client products, "Team C" had resolved all of their issues that were gating customer qualification samples significantly sooner than had the other two teams.

The study also translated this "earlier delivery" into a potential cost savings figure for the client ("if all teams were performing at the level of Team C").

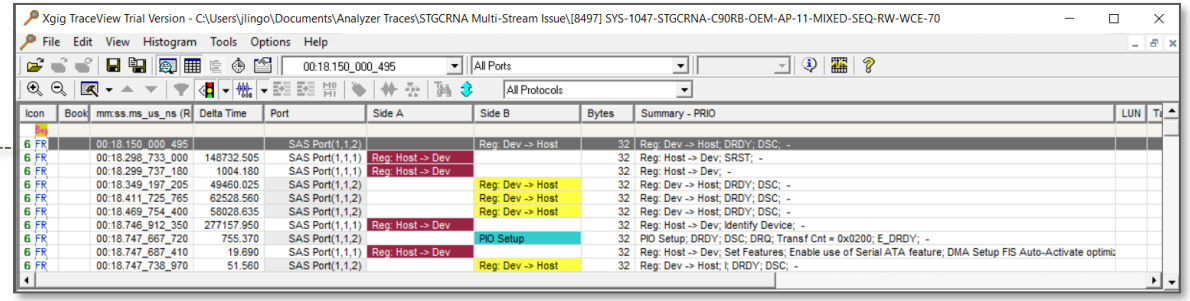


SAMPLE PROJECTS – DATA VISUALIZATION

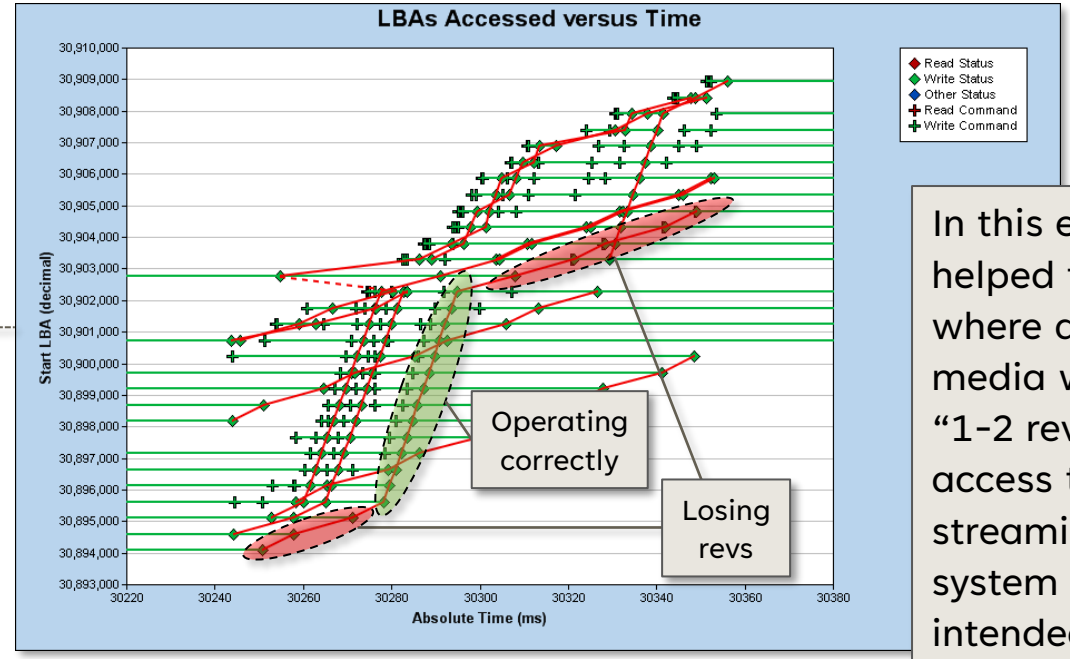
The following projects utilized MS Visual Basic and the ChartDirector library to support visual analysis of data related to storage device performance, for the purpose of issue resolution and product optimization

VISUALIZATION PROJECT 1: BUS TRACE VIEWER / ANALYZER

Problem: 3rd party bus trace viewer did not provide sufficient insight to allow designers to identify / understand storage device performance issues



Solution: Created a tool to process the trace to display *time* and *media location* (LBA = Logical Block Address) data for key (Read / Write) commands, including identification of streams of sequential media accesses (see red lines at right)



In this example, the tool helped to identify cases where accesses of rotating media were experiencing "1-2 revolution gaps" in access time, instead of streaming data to the host system continuously as intended

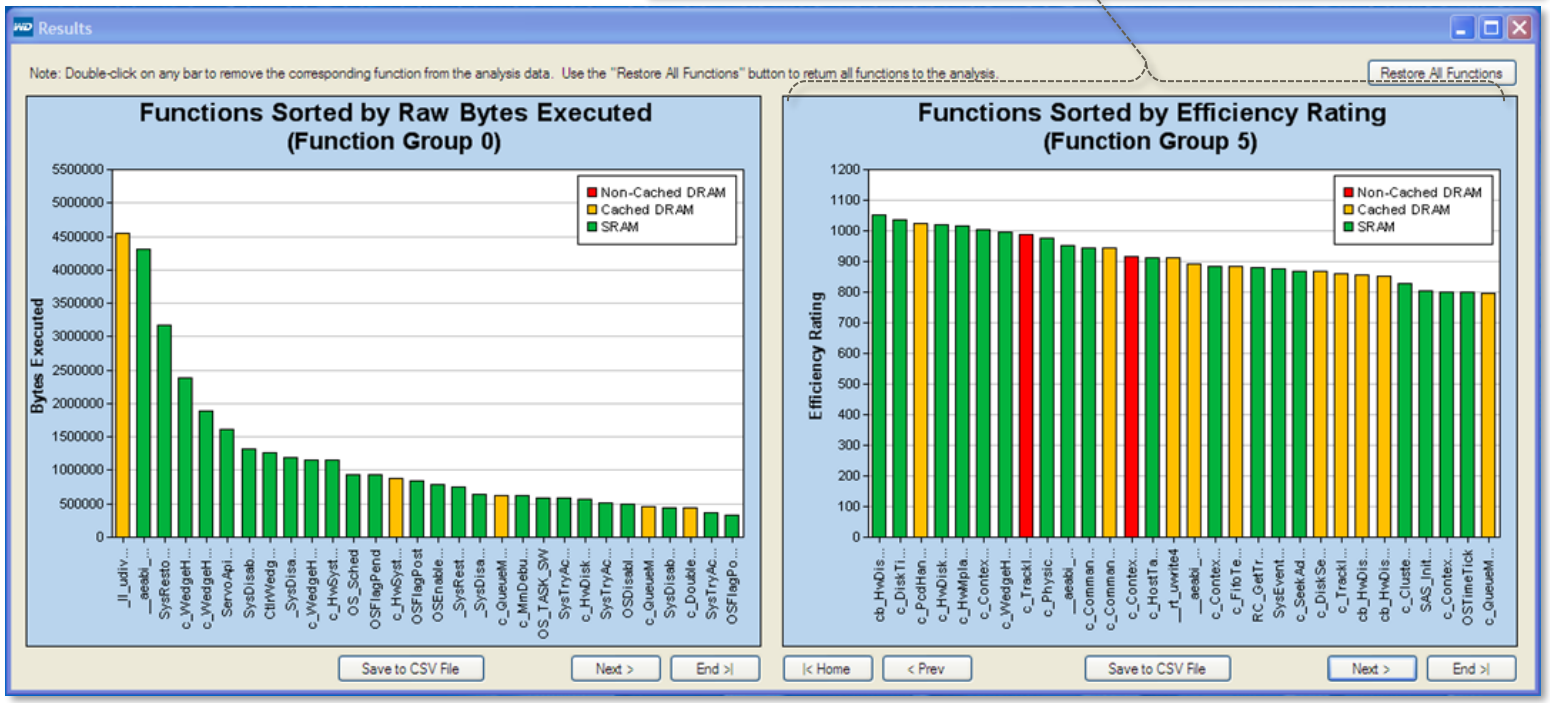
VISUALIZATION PROJECT 2: MEMORY MAP OPTIMIZATION

Problem: Embedded systems typically operate in "memory-constrained" environments, where it is important to ensure that the software functions that have the greatest impact on device performance are located in the highest speed memory (in this case, SRAM)

Solution: Created a tool to process an emulator trace of a critical code sequence, then use the memory map from the compile / link process to map this code to its associated memory type – this allows designers to identify opportunities for improving performance by changing the memory assignments of key functions

The "Efficiency Rating" is based on the cumulative number of bytes executed in a function, relative to that function's size

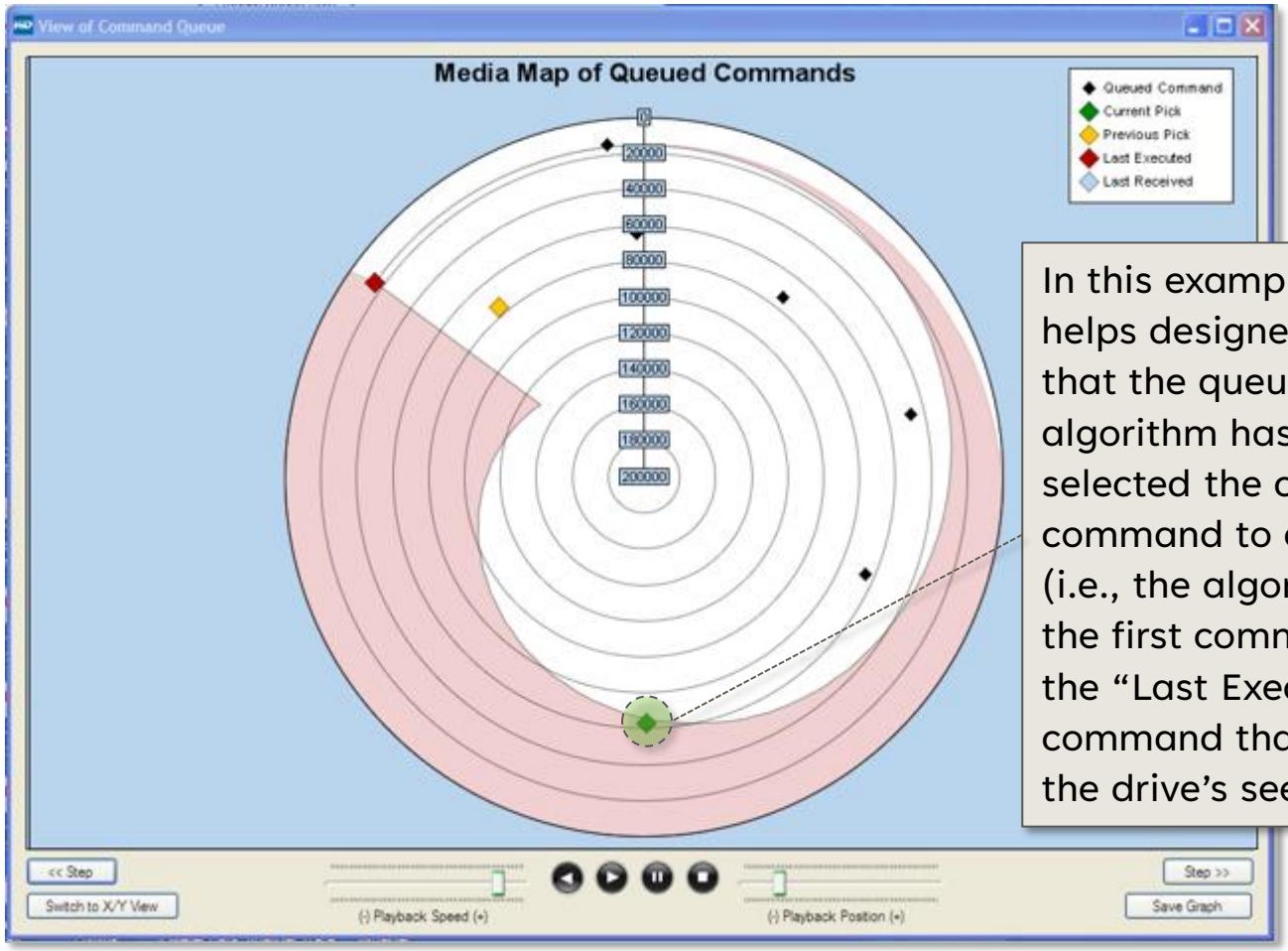
- Based on this metric, the red and yellow lines below indicate potential opportunities for further memory map optimization



VISUALIZATION PROJECT 3: QUEUE SORT PLAYBACK

Problem: Disk drives select their "next access" from a queue of hundreds of pending commands, and it is often difficult for firmware designers to visualize the operation of this 2-dimensional process

Solution: Created a tool to play back a trace containing information on the state of the drive's command queue over time, including the drive's "seek profile" (see red envelope at right, which represents how far the drive's read/write heads can move as the media rotates away from the "Last Executed" position)

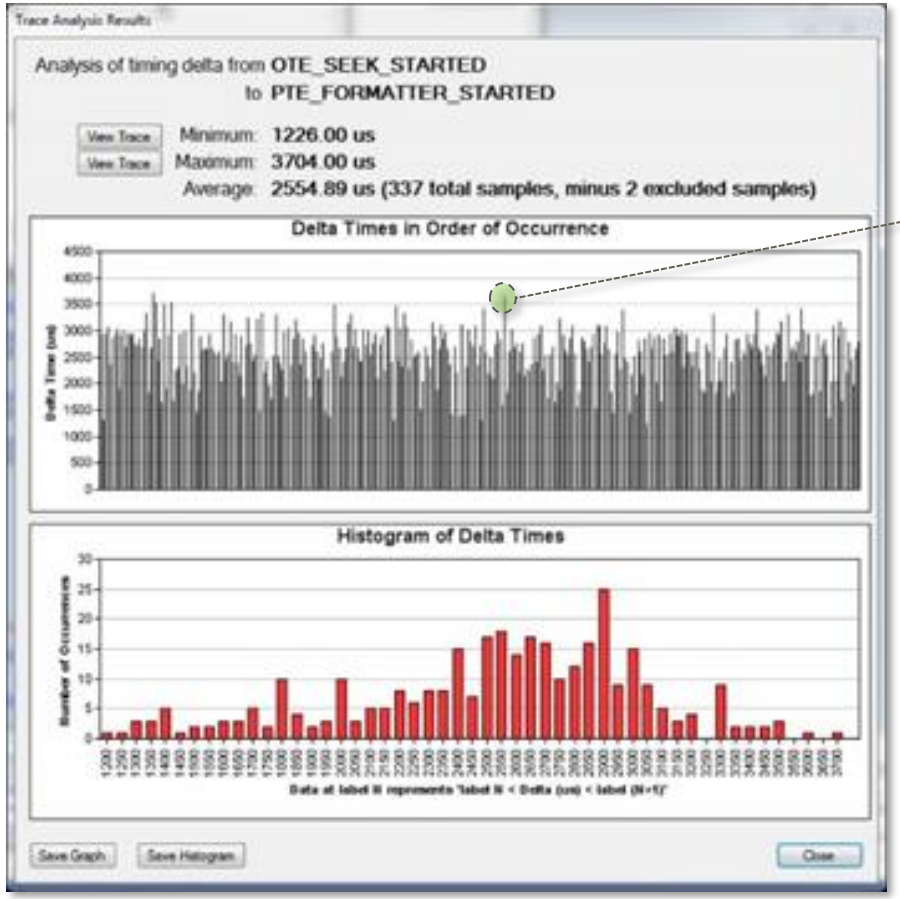


In this example, the tool helps designers recognize that the queue sort algorithm has indeed selected the correct command to execute next (i.e., the algorithm selected the first command after the "Last Executed" command that falls within the drive's seek profile)

VISUALIZATION PROJECT 4: TIMING OF KEY SEQUENCES

Problem: Embedded systems often have key code sequences that must be executed within a short period of time. While the firmware design may satisfy this requirement the majority of the time, designers must be able to identify the cause for cases in which the timing requirement is NOT met.

Solution: Created a tool to process an emulator trace, create a dictionary of all "timing points" contained within that trace, and allow users to select two timing points in order to receive information on the time delta between all occurrences of those two points in the trace



By clicking on a given bar on the "Order of Occurrence" chart, users can navigate directly to the start of the associated code segment in the trace

The "Histogram" portion of the display helps designers identify any modalities or other anomalies / outliers in the data

SUMMARY

The common theme of my projects over the past 15 years has been the improvement of business outcomes by:

- Extracting key information from complex data sets, and
- Building a supporting narrative around the resulting findings, and
- Conveying that information in an accessible (typically graphical) manner to my peers and management

THANKS FOR TAKING TIME TO VIEW THESE SAMPLES OF MY WORK

John Lingo

john.g.lingo@gmail.com

EVERY GREAT DEVELOPER YOU KNOW GOT THERE BY SOLVING PROBLEMS THEY WERE UNQUALIFIED TO SOLVE UNTIL THEY ACTUALLY DID IT.

– Patrick McKenzie