CENTRILLION

Reference (2)

Jeff (CHI-HSUAN HO)

Intensity Extraction and Data collecting Software – Summit Grid Review, Improve & Evaluate

Jeff (CHI-HSUAN HO)

Gridding Development

- Summit.Grid
  - Chip Spec Parameters Setup

```
{
    "name": "banff",
    "w": 2480,
    "h": 2480,
    "w_cl": 496,
    "h_cl": 496,
    "cell_w_um": 4,
    "cell_h_um": 4,
    "space_um": 1,
    "location_marker": {
        "template": "resource/banff/pat_white.tif",
        "mask"    : "resource/banff/pat_white_mask.tif",
        "w": 60,
        "h": 60
    },
    "shooting_marker": {
        "origin_desc": "center of the chip",
        "type": "regular_matrix",
        "mk_pats": [
            {
                "filter": 0,
                "w_um": 60,
                "h_um": 60,
                "path": "resource/banff/pat_white.tif",
                "mask": "resource/banff/pat_white_mask.t
            },
            {
                "um2px_r": 2.68,
                "path": "resource/banff/pat_2_68.tif"
            },
            {
                "um2px_r": 2.41,
                "path": "resource/banff/pat_2_41.tif"
```

```
, "chip": "banff33"
, "view": {
      "offset": [ 0, 0 ]
    , "layout": [ 1, 1 ]
    , "stride": [ 810, 810 ]
  }

// override default parameters
, "init_autofocus": {
      "range_step": 2000
    , "epsilon": 5.0
  }
, "scan_channel": {
      "name": "White8"
    , "pixel_format": "Mono8"
    , "gain": 0
    , "exposure_time_abs": 500
    , "camera_delay_time": 1
    , "filter": 0
  }

// module parameters
, "sweep_distance": 320
, "sweep_stepsize": 20
, "boxgrid": {
      "layout": [   3,   3 ]
    , "stride": [ 405, 405 ]
    , "cellsz": [ 120, 120 ]
    , "enable": [ 0, 2, 6, 8 ]
  }
```

```
{
    "name": "banff33",
    "spec": "banff",
    "fov": {
        "mask_src_type": "au
        "mask": "resource/ba
        "x_i" : -810,
        "y_i" : -810,
        "rows": 3,
        "cols": 3,
        "w_d" : 810,
        "h_d" : 810,
        "seq": [
            [[0, 0]], [[1, 0]], [[2, 0]],
            [[2, 1]], [[1, 1]], [[0, 1]],
            [[0, 2]], [[1, 2]], [[2, 2]]
        ]
    },
    "origin_infer": {
        "algo": "aruco_detection",
        "pyramid_level": 3,
        "nms_count": 9,
        "cell_size_px": 5,
        "loc_mk_layout_pt": [3,3]
    },
    "af": {
        "search_range"    : 120
      , "step_size"       : 30
      , "margin"          : 0
      , "kernel_scale"    : 0.0
      , "length_scale"    : 45.0
      , "noise_level"     : 1e-3
      , "maxiter"         : 10000
```

```
        "gain": 0,
        "exposure_time_abs": 250000,
        "camera_delay_time": 1,
        "filter": 2,
        "marker_type": "AM3"
    },
    {
        "name": "green-AM5B-250ms",
        "pixel_format": "Mono14",
        "gain": 0,
        "exposure_time_abs": 250000,
        "camera_delay_time": 1,
        "filter": 2,
        "marker_type": "AM5B"
    },
    {
        "name": "red-AM1-250ms",
        "pixel_format": "Mono14",
        "gain": 0,
        "exposure_time_abs": 250000,
        "camera_delay_time": 1,
        "filter": 4,
        "marker_type": "AM1"
    },
```

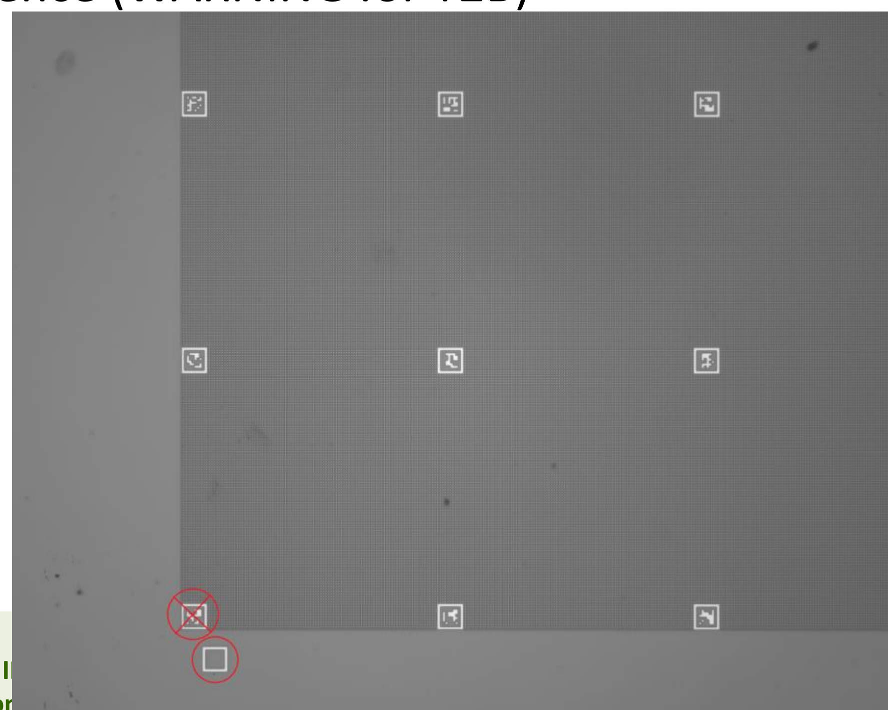
```
LAB     Merge branch 'master' into LAB
FAB     Merge branch 'master' into FAB
master  [Feature] Update the scanning parameter for S
        QAQC Shasta chip (S2B) support.
        Merge branch 'master' into LAB
        Merge branch 'master' into FAB
        Support for scanning and gridding Shasta (S2B
        QAQC Jasper chip (J1C, J2C) support.
        Merge branch 'master' into LAB
        Merge branch 'master' into FAB
        Support for scanning and gridding Jasper (J1C
        QAQC Jasper chip (J1C) support.
        Merge branch 'master' into FAB
```

Gridding Development

- Summit.Grid
  - Bugfix for wrong nms_count (WARNING for S1C)
    - Original

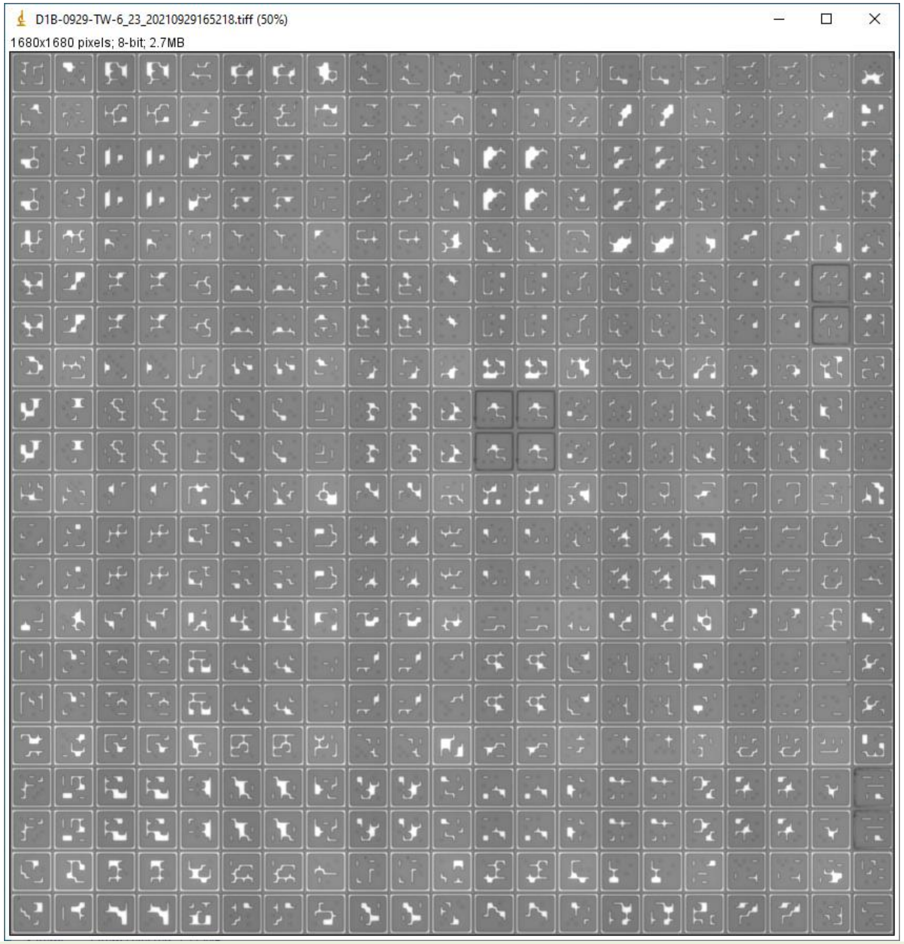
```
// detection parameters
nms_count_       = (fov_wd_ / mk_wd_cl_ + 1) * (fov_hd_ / mk_hd_cl_ + 1);      Alex, 2 years ago • support new aruco recognition …
nms_radius_      = aruco_marker_->at("nms_radius");
```

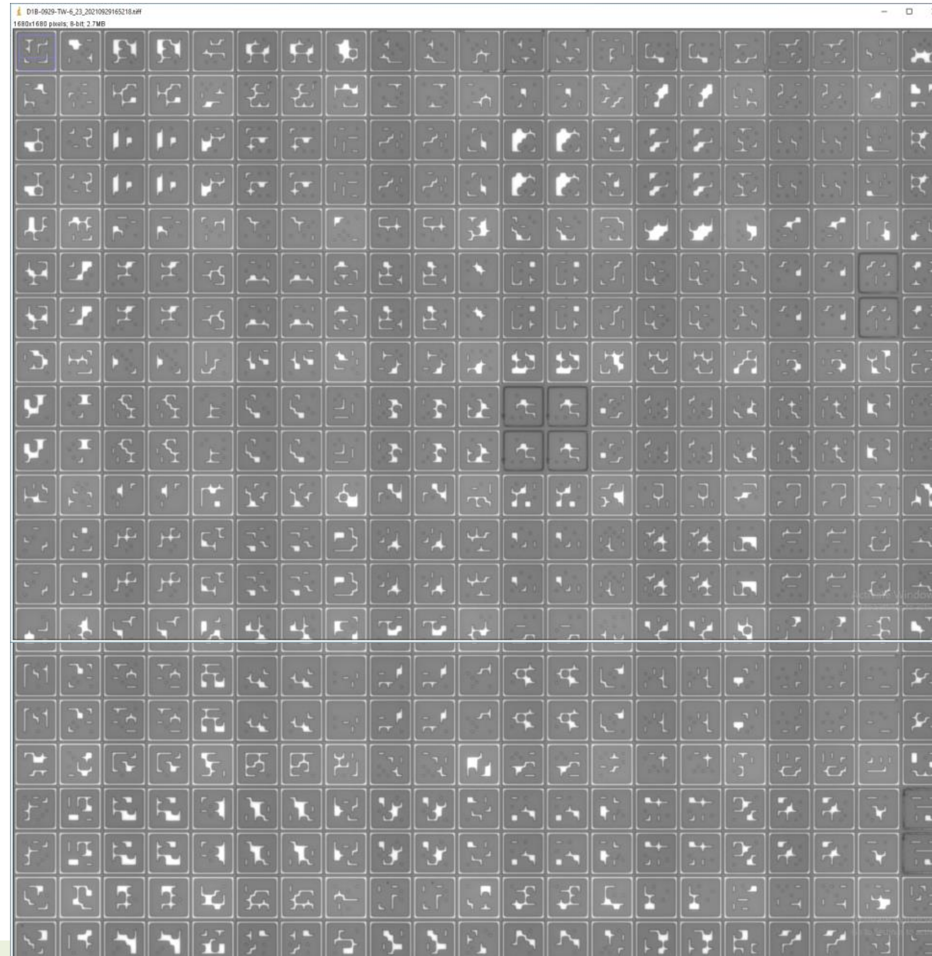
  - Noise influence (WARNING for Y2B)



Gridding Development

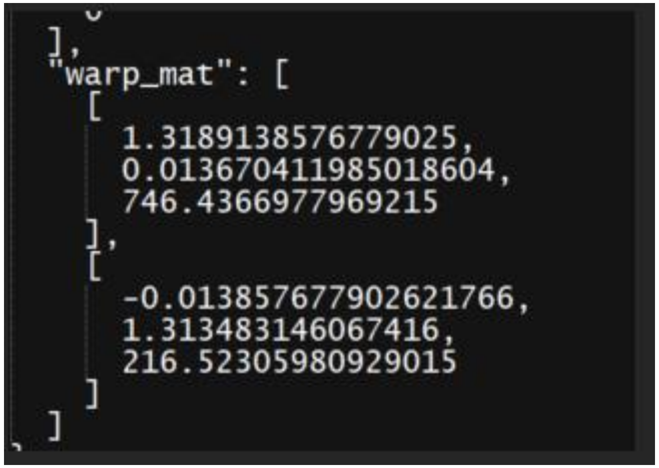
- Summit.Grid
  - Rescue mechanism for gridding bad fov (for erosion).

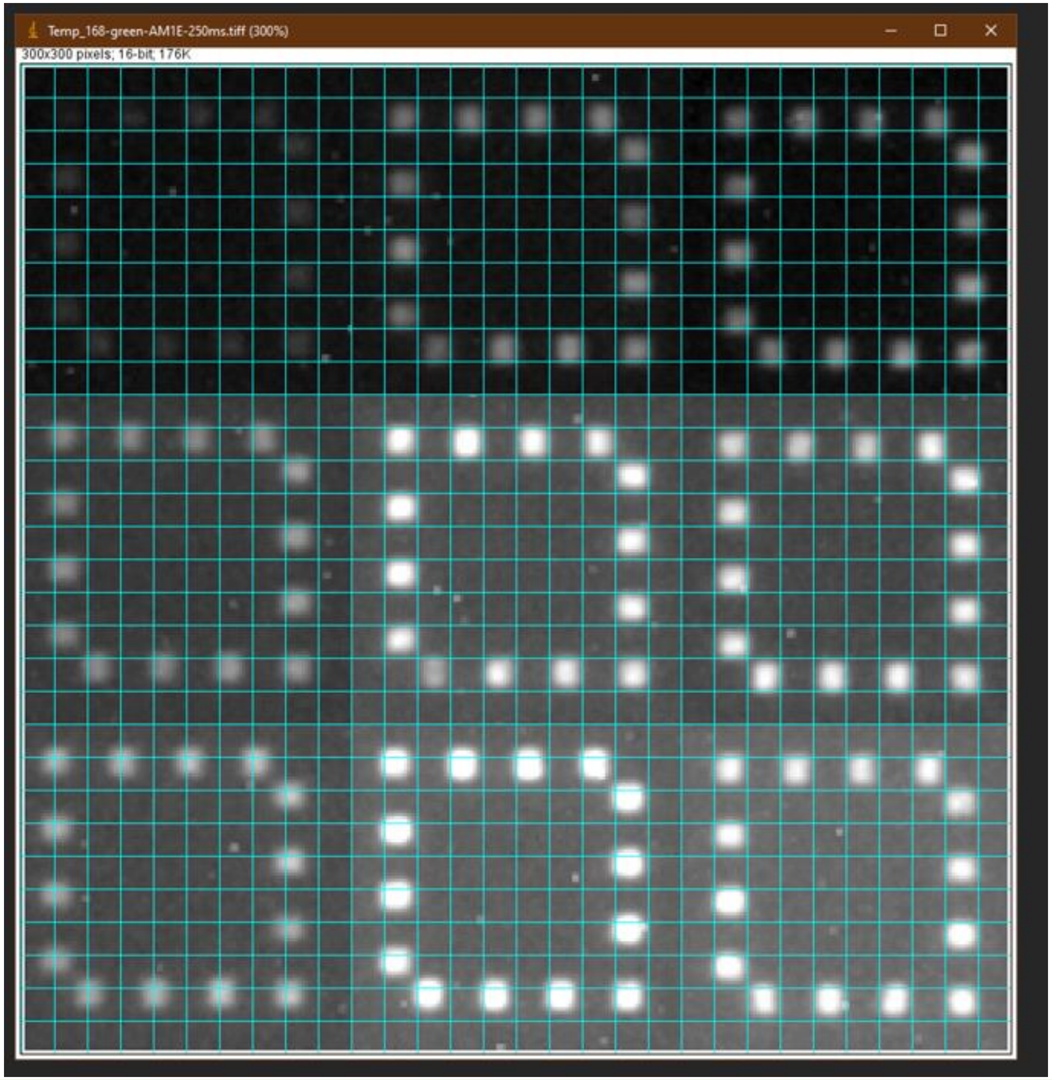




Gridding Development

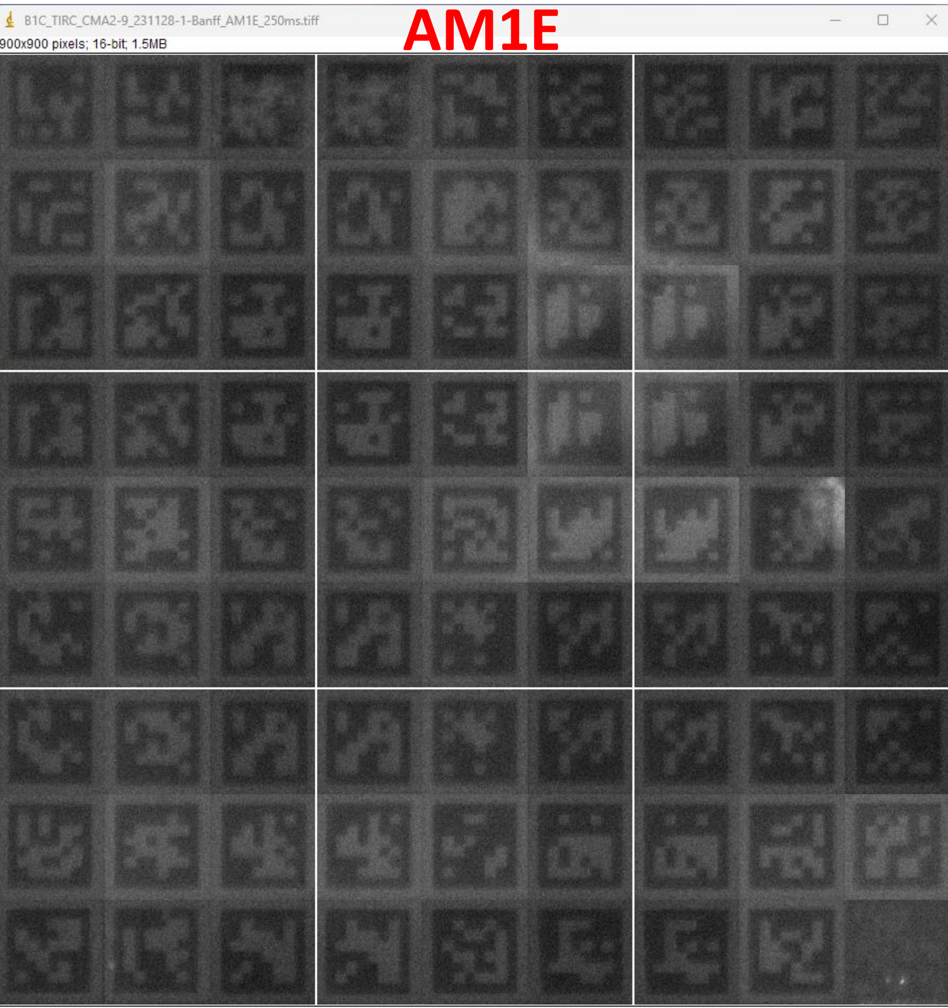
- Summit.Grid
  - PGD Images Processing.

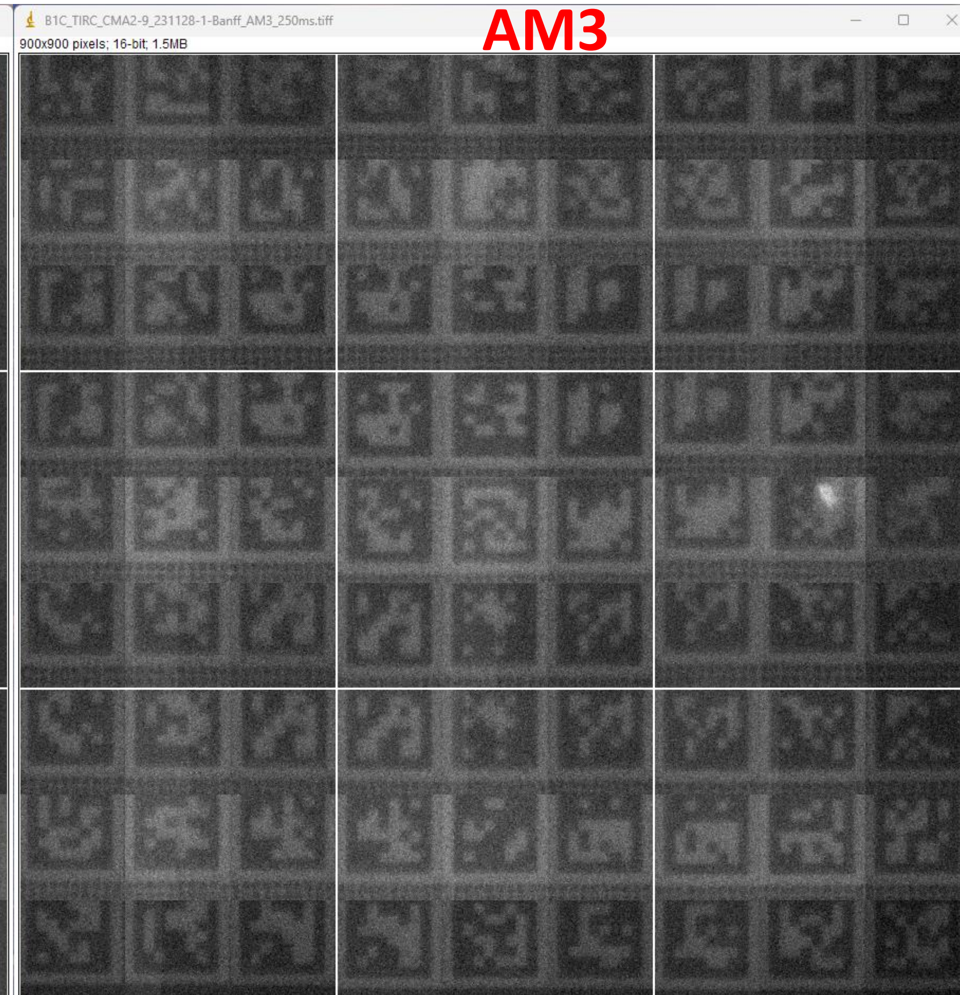

```
],
"warp_mat": [
  [
    1.3189138576779025,
    0.013670411985018604,
    746.4366977969215
  ],
  [
    -0.013857677902621766,
    1.313483146067416,
    216.52305980929015
  ]
]
```



Gridding Development

- Summit Grid checking support.



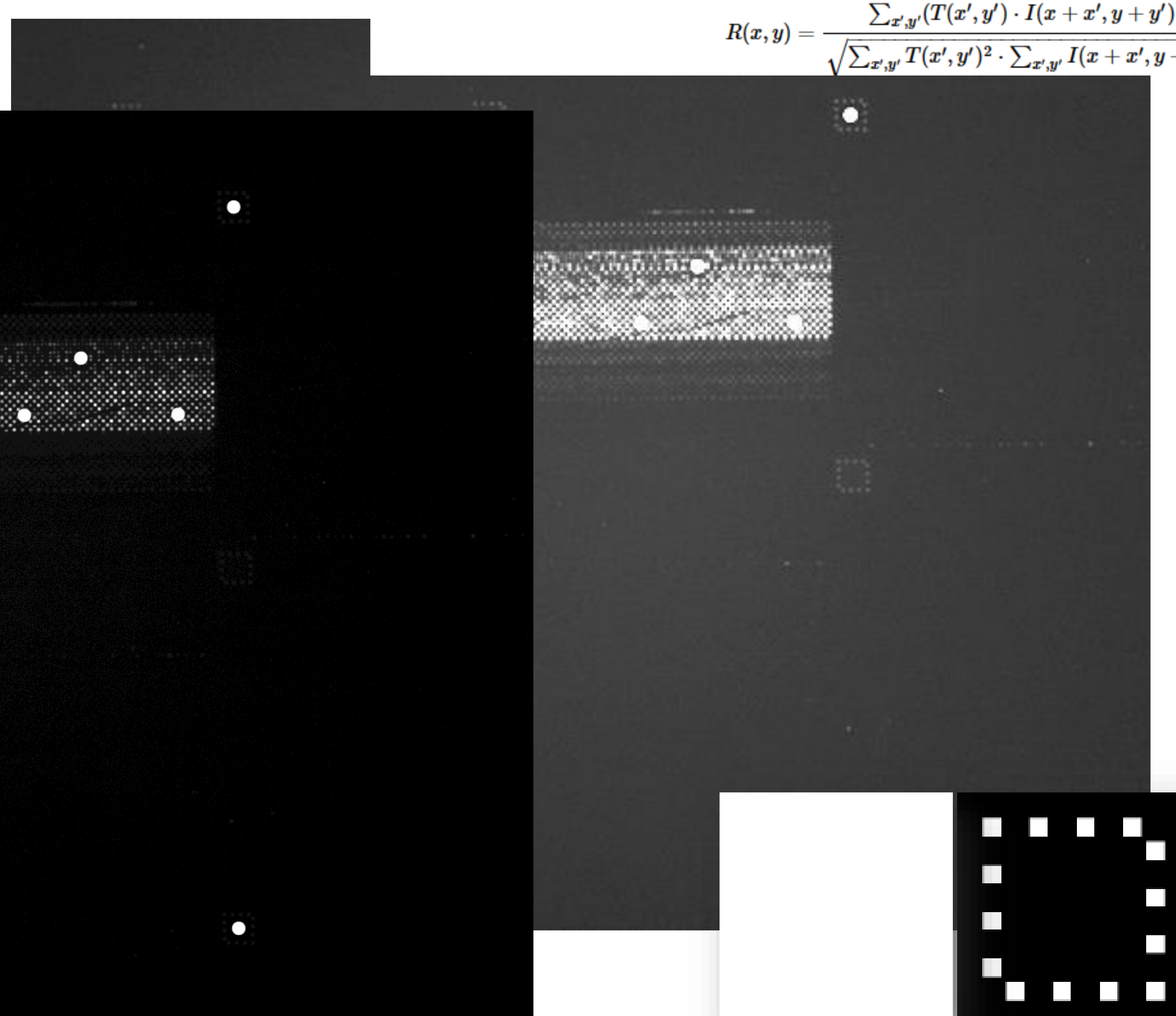


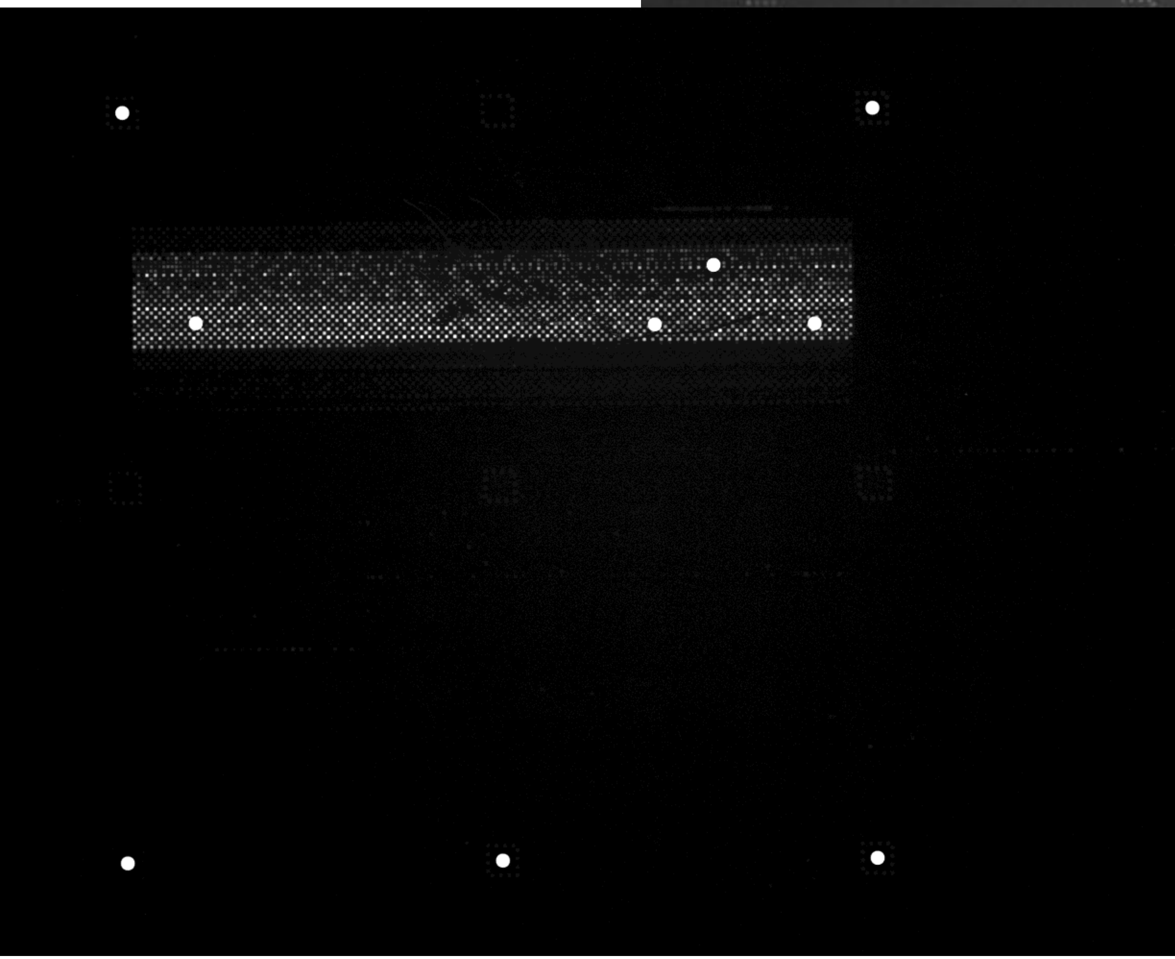
Performance for New Gridding Software

- Summit.Grid
  - Old statistical method

method=TM_CCORR_NORMED

$$R(x,y) = \frac{\sum_{x',y'}(T(x',y') \cdot I(x+x',y+y'))}{\sqrt{\sum_{x',y'} T(x',y')^2 \cdot \sum_{x',y'} I(x+x',y+y')^2}}$$





Performance for New Gridding Software

- Summit.Grid
  - New statistical method

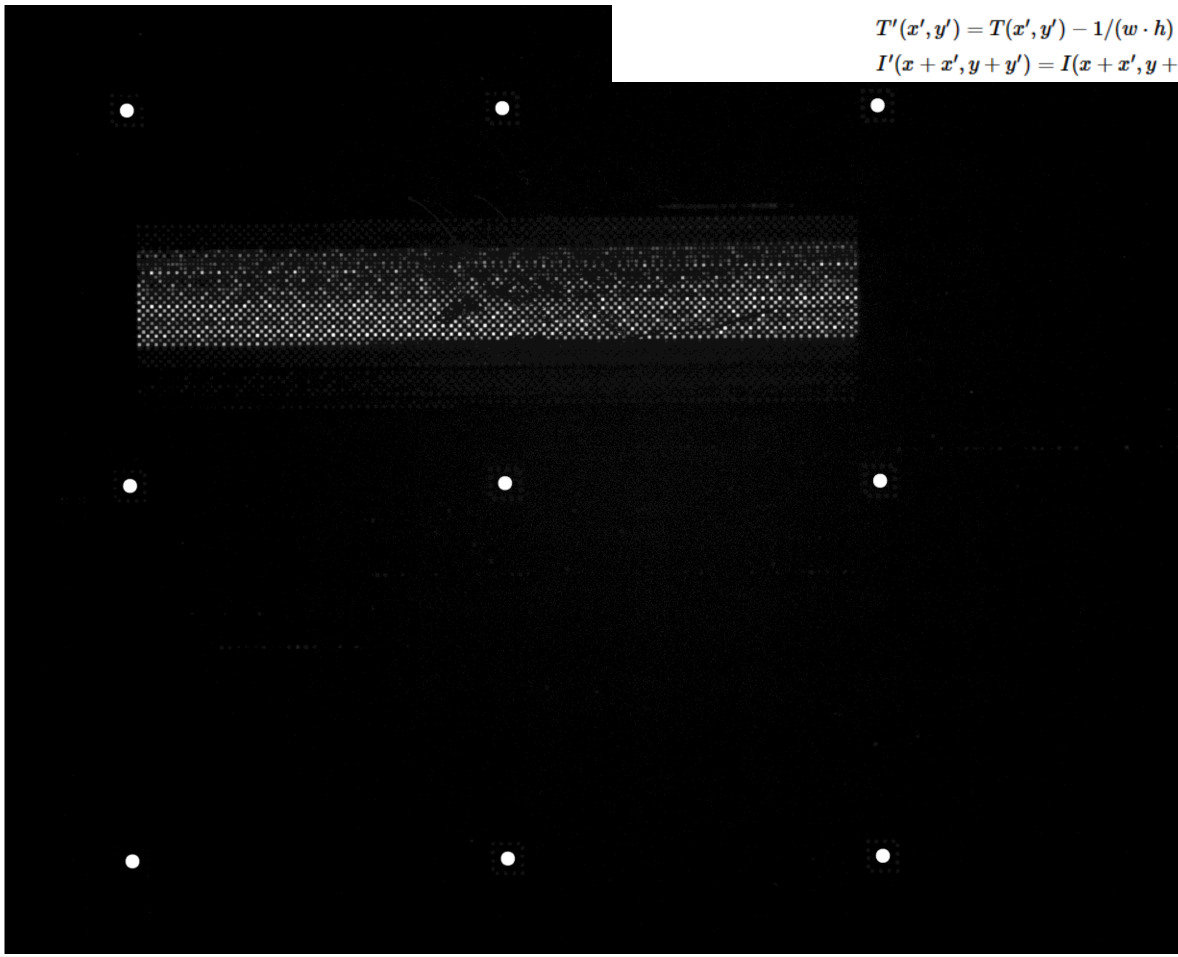
method=TM_CCOEFF_NORMED

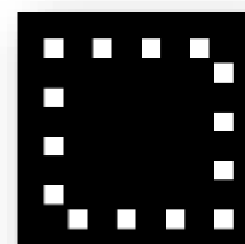
$$R(x,y) = \frac{\sum_{x',y'}(T'(x',y') \cdot I'(x+x',y+y'))}{\sqrt{\sum_{x',y'} T'(x',y')^2 \cdot \sum_{x',y'} I'(x+x',y+y')^2}}$$

where

$$T'(x',y') = T(x',y') - 1/(w \cdot h) \cdot \sum_{x'',y''} T(x'',y'')$$

$$I'(x+x',y+y') = I(x+x',y+y') - 1/(w \cdot h) \cdot \sum_{x'',y''} I(x+x'',y+y'')$$



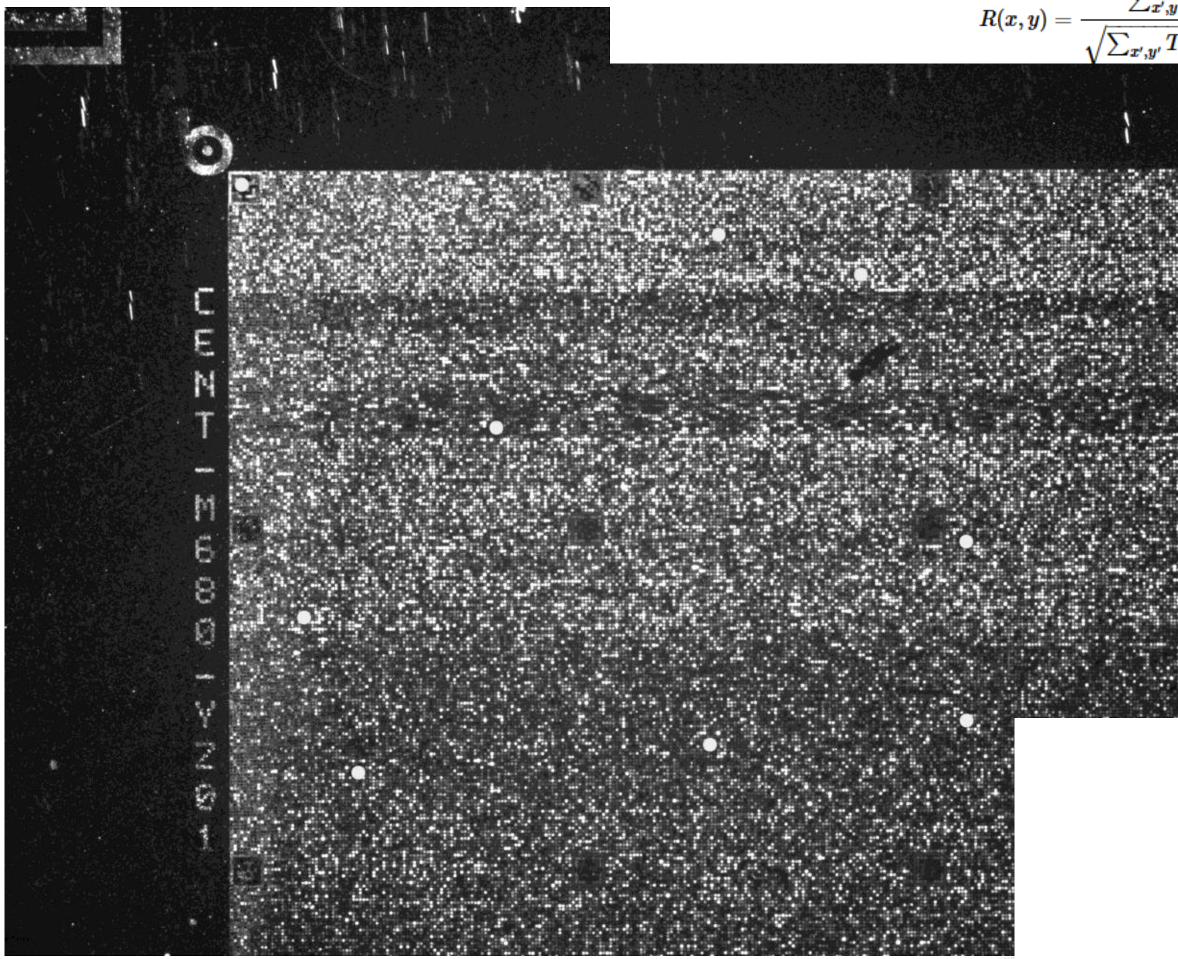


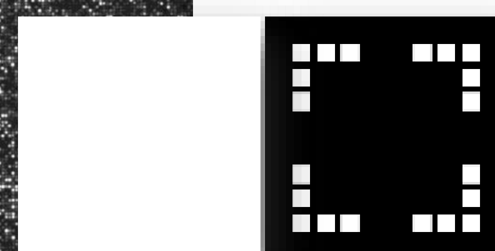
Performance for New Gridding Software

- Summit.Grid
  - Old statistical method

method=TM_CCORR_NORMED

$$R(x,y) = \frac{\sum_{x',y'}(T(x',y') \cdot I(x+x',y+y'))}{\sqrt{\sum_{x',y'} T(x',y')^2 \cdot \sum_{x',y'} I(x+x',y+y')^2}}$$





Performance for New Gridding Software

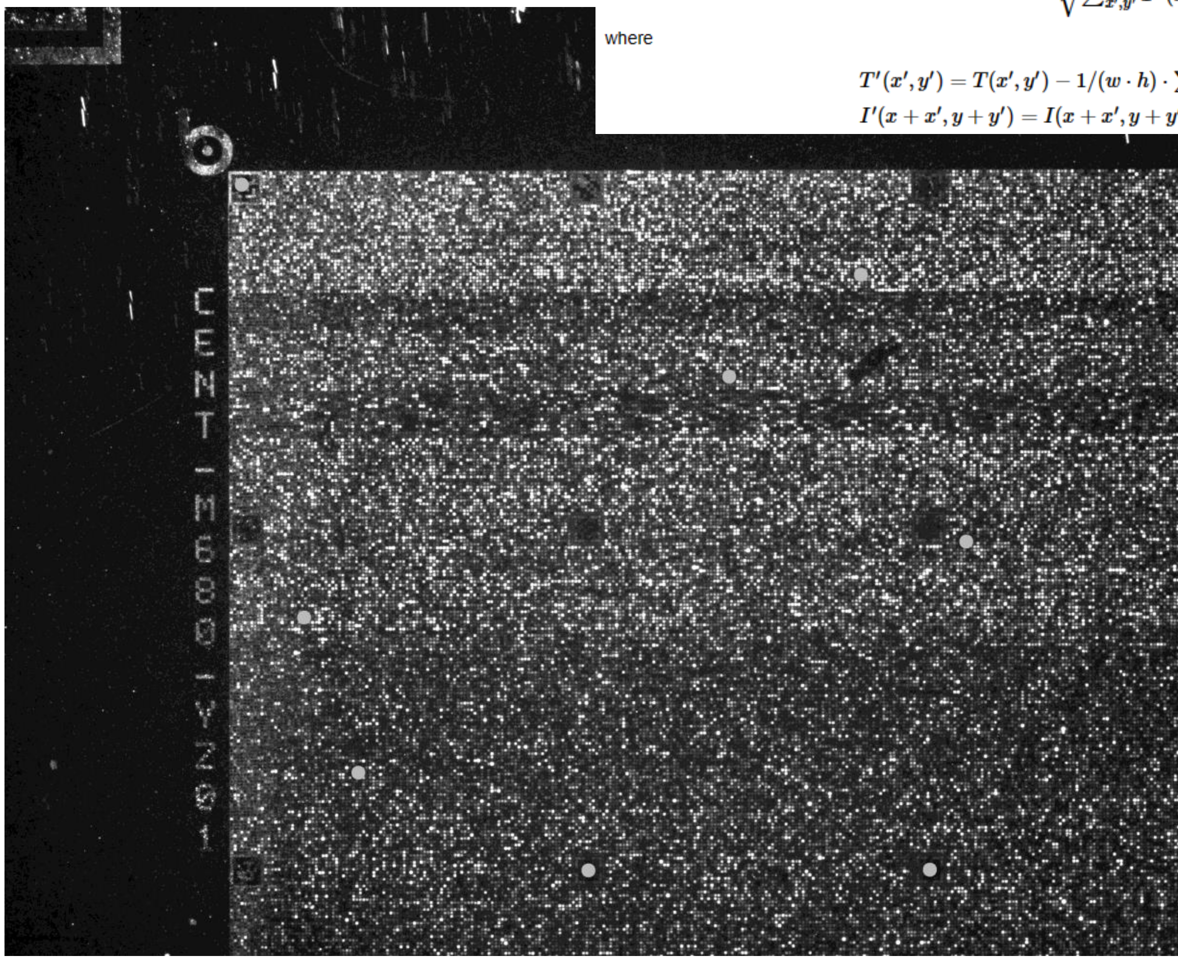
- Summit.Grid
  - New statistical method

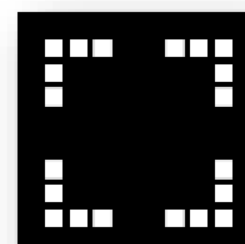
method=TM_CCOEFF_NORMED

$$R(x,y)=\frac{\sum_{x',y'}(T'(x',y')\cdot I'(x+x',y+y'))}{\sqrt{\sum_{x',y'}T'(x',y')^2\cdot\sum_{x',y'}I'(x+x',y+y')^2}}$$

where

$$T'(x',y')=T(x',y')-1/(w\cdot h)\cdot\sum_{x'',y''}T(x'',y'')$$
$$I'(x+x',y+y')=I(x+x',y+y')-1/(w\cdot h)\cdot\sum_{x'',y''}I(x+x'',y+y'')$$

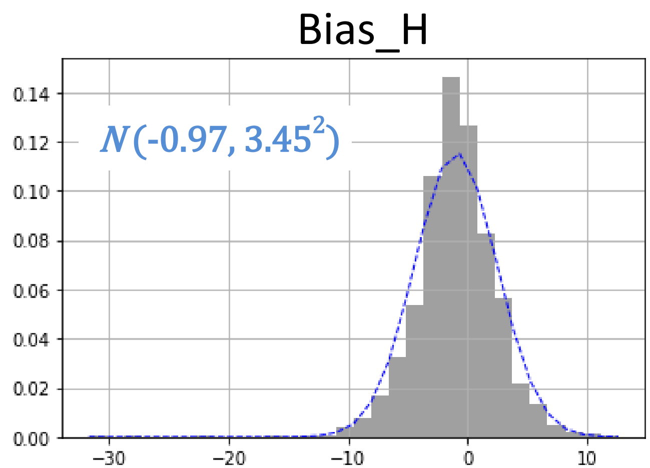


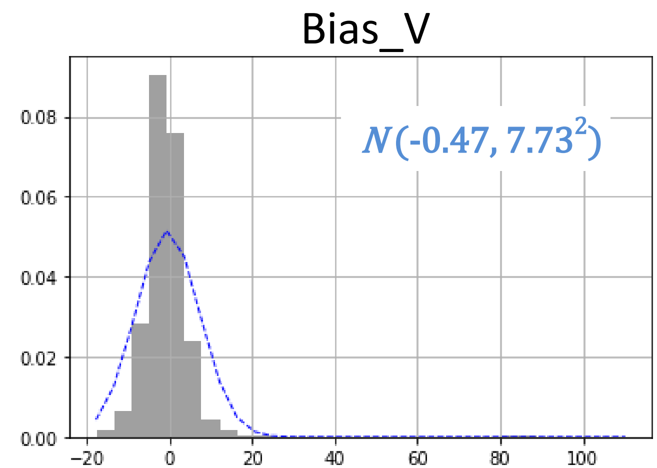


Sampling Distribution for Different Chip Scan Mode

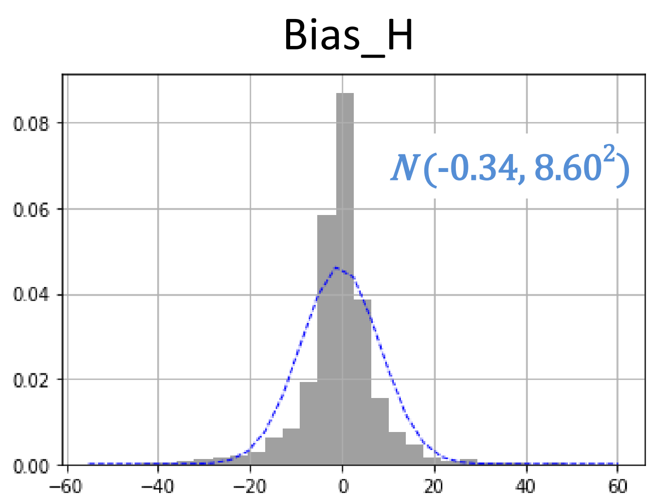
- Quick scan mode experiments
  - Sample: 5 YZ01 chips (7x7 FOVs) x 10 runs => 2450 FOVs
  - Estimation: Var(X+Y)

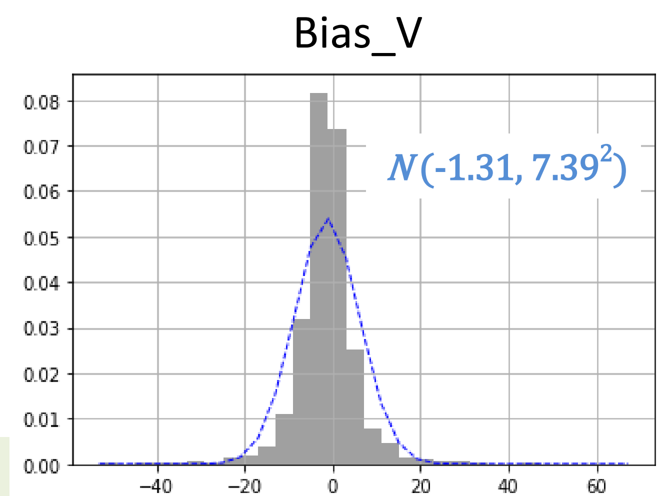
SUMMIT Test 2





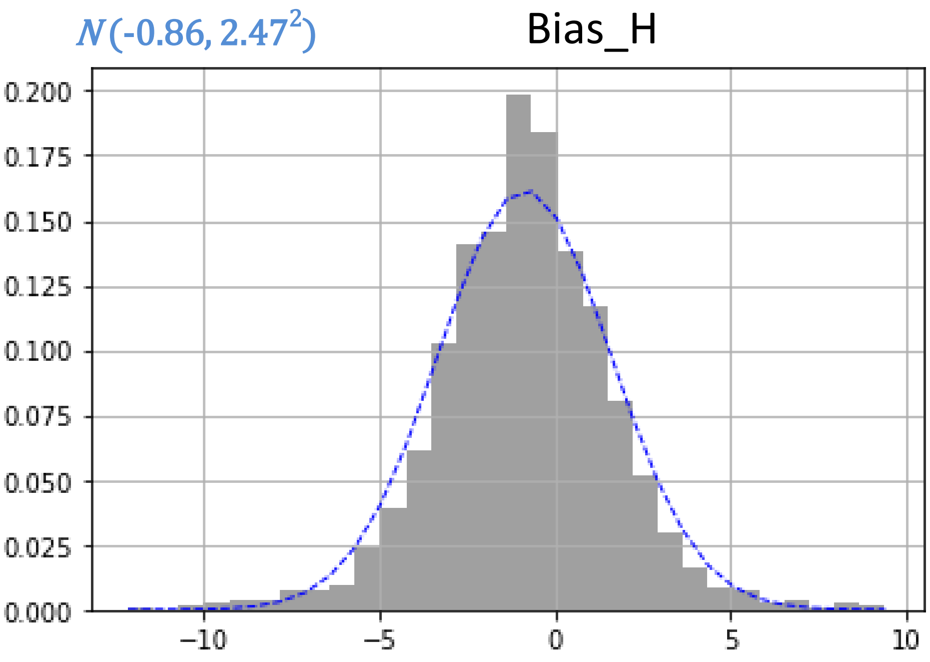
SUMMIT Test 3

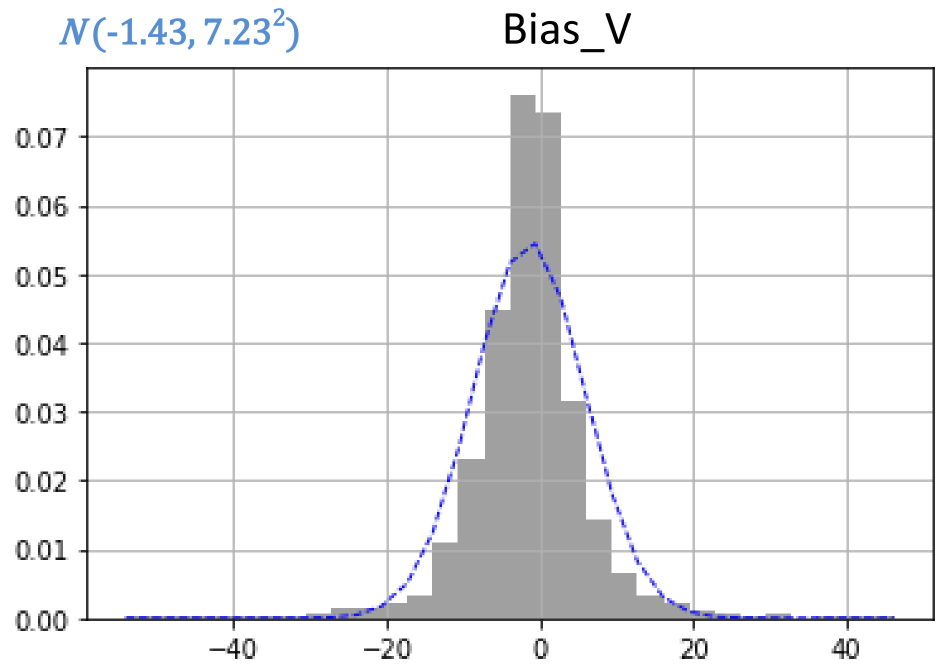




Sampling Distribution for Different Chip Scan Mode

- Quick scan mode experiments
  - Sample: 5 YZ01 chips (7x7 FOVs) x 10 runs => 2450 FOVs
  - Estimation: Var(X+Y)
  - SUMMIT with precise sliding



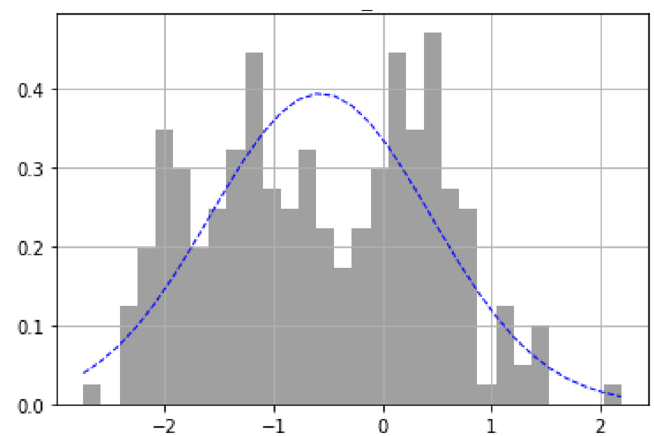


Sampling Distribution for Different Chip Scan Mode

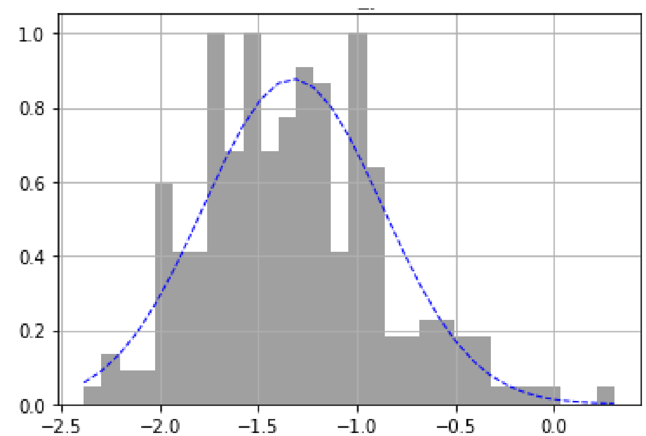
- Regular high precision mode experiments
  - Sample: 5 YZ01 chips (7x7 FOVs) x 1 runs => 245 FOVs
  - Estimation: Var(X)

SUMMIT Test 2

$\mathcal{N}(-0.57, 1.01^2)$ Bias_H

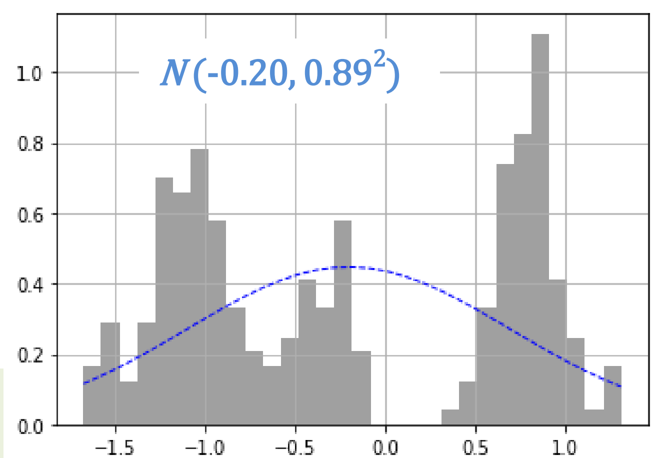


$\mathcal{N}(-1.32, 0.45^2)$ Bias_V

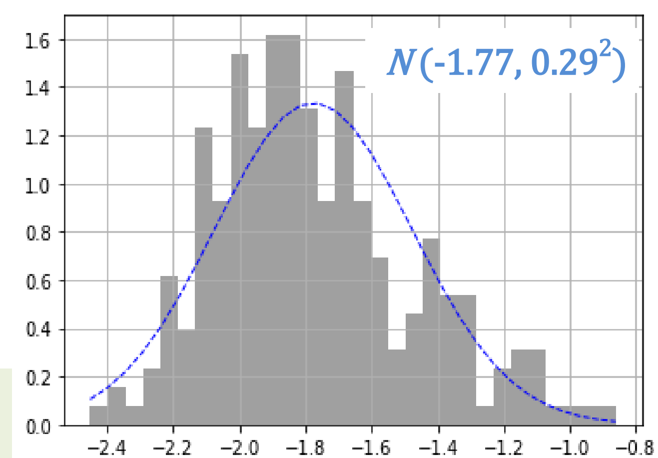


SUMMIT Test 3

Bias_H

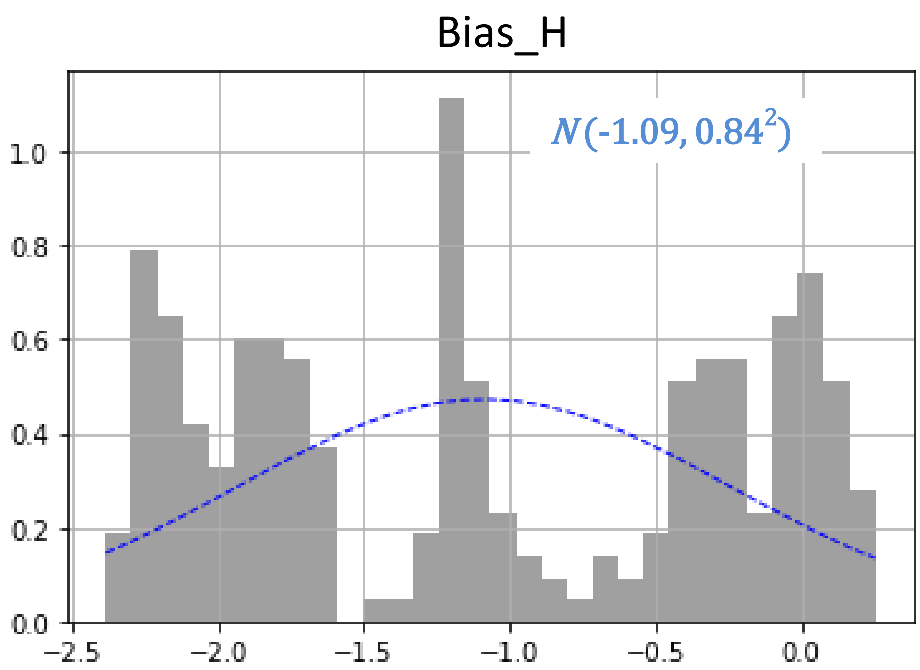


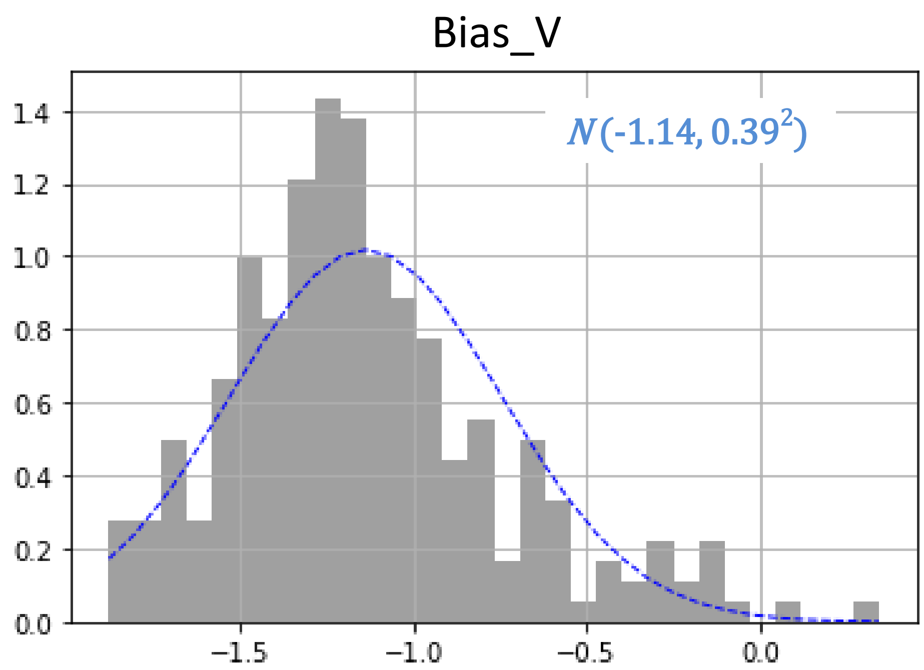
Bias_V



Sampling Distribution for Different Chip Scan Mode

- Regular high precision mode experiments
  - Sample: 5 YZ01 chips (7x7 FOVs) x 1 runs => 245 FOVs
  - Estimation: Var(X)
  - SUMMIT with Precise Sliding



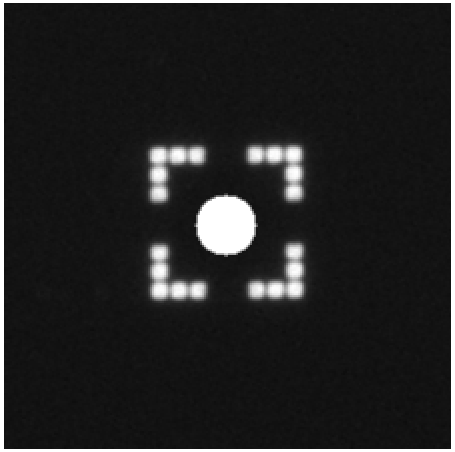


Statistical Conclusion from Experiments Results

- Gridding
  - SUMMIT Parameters Estimation
    - Let $X \equiv r.v.$ of the displacement from changing the filter (BF -> fluorescent).
    - Let $Y \equiv r.v.$ of the displacement from relocating the plate to the same position.
    - In the quick scan mode,

      Estimate $Var(X+Y) = Var(X) + Var(Y)$
    - In the regular high precision mode,

      Estimate $Var(X)$ Only



Statistical Conclusion from Experiments Results

- Gridding
  - Chebyshev's Inequality
    - $P(|Z-\mu| \geq k \cdot \sigma) \leq \frac{1}{k^2}$
    - A. Quick scan for estimating $Var(X+Y)$
      - $k = 14.3, \sigma = 7.74,$ Cover radius: 110.7 (pixels)
      - 99.5% ↑ confidence that the true marker center lies in the cover.
      - 2.3 x BF_mark_size, Lower bound: 110.7 (pixels)
    - B. Regular scan for estimating $Var(X)$
      - $k = 7.2, \sigma = 1.01,$ Cover radius: 7.272 (pixels)
      - 98.06% ↑ confidence that the true marker center lies in the cover.
      - 0.15 x BF_mark_size
  - Normal Distribution
    - $P(|T-\mu| < Z_{0.0025} \cdot \sigma) =$ 99.5%, $Z_{0.0025} = 2.807$, 99.5% confidence that the true marker center lies in the cover.
    - A. Quick scan for estimating $Var(X+Y)$
      - $\sigma = 7.74$, Cover radius: 21.7 (pixels)
    - B. Regular scan for estimating $Var(X)$
      - $\sigma = 1.01$, Cover radius: 2.8 (pixels)
  - Overall Performance - successfully recognized rate: nearly 100%.

Signal Intensity Extraction Techniques Comparison

- Raw Data (NPcall Analyzer)

Grid2 (current)	No.	Data	Grid1	Grid2_subpix	Grid2_subpix_cvfix	Ranking		
85.245%	1	85_46_20210324123447	85.944%	86.131%	86.131%	3	1	1
86.113%	2	85_68_20210303140400	87.582%	87.506%	87.506%	1	2	2
84.400%	3	85_69_20210303142117	87.605%	87.261%	87.261%	1	2	2
85.746%	4	85_76_20210303143808	86.719%	87.244%	87.238%	3	1	2
87.552%	5	85_77_20210303145506	87.751%	88.974%	88.817%	3	1	2
93.293%	6	90_21_20210325132601	93.916%	93.893%	93.945%	2	3	1
92.110%	7	90_54_20210324125308	92.512%	92.587%	92.634%	3	2	1
92.255%	8	90_62_20210324131130	93.036%	92.978%	92.984%	1	3	2
91.748%	9	90_70_20210324132953	92.657%	92.611%	92.611%	1	2	2
90.653%	10	90_77_20210325142116	91.492%	91.533%	91.550%	3	2	1
93.467%	11	95_38_20210324121430	94.015%	94.079%	94.161%	3	2	1
95.798%	12	95_76_20210324164230	96.096%	96.294%	96.311%	3	2	1
94.837%	13	95_77_20210221134715_94_8	95.402%	95.367%	95.379%	1	3	2
95.868%	14	95_77_20210324165927	96.230%	96.317%	96.270%	3	1	2
94.965%	15	95_78_20210221140401	95.688%	95.641%	95.624%	1	2	3

Signal Intensity Extraction Techniques Comparison

CENTRILLION

- Raw Data (GT Caller)

No.	Data	Grid1	Grid2_subpix	Grid2_subpix_cvfix	Ranking		
1	85_46_20210324123447	92.225%	93.225%	93.358%	3	2	1
2	85_68_20210303140400	93.333%	92.650%	92.450%	1	2	3
3	85_69_20210303142117	93.325%	92.991%	92.958%	1	2	3
4	85_76_20210303143808	92.492%	93.266%	93.050%	3	1	2
5	85_77_20210303145506	93.058%	94.208%	94.192%	3	1	2
6	90_21_20210325132601	98.667%	99.600%	98.825%	3	1	2
7	90_54_20210324125308	98.175%	98.369%	98.183%	3	1	2
8	90_62_20210324131130	98.500%	98.392%	98.458%	1	3	2
9	90_70_20210324132953	98.108%	97.942%	97.950%	1	3	2
10	90_77_20210325142116	97.950%	98.158%	98.158%	3	1	1
11	95_38_20210324121430	99.092%	99.008%	99.025%	1	3	2
12	95_76_20210324164230	99.666%	99.666%	99.683%	2	2	1
13	95_77_20210221134715_94_8	99.392%	99.416%	99.408%	3	1	2
14	95_77_20210324165927	99.583%	99.608%	99.608%	3	1	1
15	95_78_20210221140401	99.367%	99.316%	99.333%	1	3	2

Signal Intensity Extraction Techniques Comparison

CENTRILLION

- Multiple data NP call results comparison (from finally 38 results)
  - H_0: Grid1 intensity $\geq$ Intensity extracted from new developed process (Grid2)
  - H_1: Grid1 intensity $<$ Intensity extracted from new developed process (Grid2)

t-Test: Paired Two sample for Means (NP call Analyzer)

	Grid1	Grid2_subpix
Mean	15307.82185	15322.24077
Variance	2060861.144	2050166.456
Observations	38	38
Pearson Correlation	0.999513654	
Hypothesized Mean Difference	0	
Df	37	
t Stat	**-1.980941596**	
P(T<=t) one-tail	0.027533779	
t Critical one-tail	1.68709362	
P(T<=t) two-tail	0.055067559	
t Critical two-tail	2.026192463	

	Grid1	Grid2_subpix_cvfix
Mean	15307.82185	15322.34463
Variance	2060861.144	2047607.206
Observations	38	38
Pearson Correlation	0.999572526	
Hypothesized Mean Difference	0	
df	37	
t Stat	**-2.123345038**	
P(T<=t) one-tail	0.020238129	
t Critical one-tail	1.68709362	
P(T<=t) two-tail	0.040476259	
t Critical two-tail	2.026192463	

$\Rightarrow$ Reject H_0

t-Test: Paired Two sample for Means (GT Caller)

	Grid1	Grid2_subpix
Mean	16100.97195	16126.25089
Variance	2272377.702	2210252.078
Observations	38	38
Pearson Correlation	0.998958931	
Hypothesized Mean Difference	0	
df	37	
t Stat	**-2.182738079**	
P(T<=t) one-tail	0.017738267	
t Critical one-tail	1.68709362	
P(T<=t) two-tail	0.035476534	
t Critical two-tail	2.026192463	

	Grid1	Grid2_subpix_cvfix
Mean	16100.97195	16120.53029
Variance	2272377.702	2195827.937
Observations	38	38
Pearson Correlation	0.99898407	
Hypothesized Mean Difference	0	
df	37	
t Stat	-1.672830702	
P(T<=t) one-tail	0.051398391	
t Critical one-tail	1.68709362	
P(T<=t) two-tail	0.102796783	
t Critical two-tail	2.026192463	

$\Rightarrow$ Reject H_0

CENTRILLION

Normal Gamma Background Correction & Data Preprocess

Jeff (CHI-HSUAN HO)

CENTRILLION

- **Model Assumption**

  - For each single array:

$$X_j = S_j + B_j$$

  - $BgC: X_j \Rightarrow S_j$ Enhance the biological validity of the results.

Improving background correction for Illumina BeadArrays: the normal-gamma model.

Sandra Plancade [1]*, Yves Rozenholc [2], Eiliv Lund [1]

[1]Department of Community Medicine, Faculty of Health Sciences, University of Tromsø, 9037 Tromsø, Norway.
[2]Department of Applied Mathematics, MAP5, 45 rue des Saints-Pères, University Paris Descartes, 75006 Paris.

ABSTRACT
Motivation: Illumina beadarray technology provides high quality data, including non specific negative control features which allow a precise estimation of the background noise. As reported in many studies, the traditional background subtraction proposed in BeadStudio leads

Namely, let X be the observed intensity of a given probe, we assume that

$$X = S + B \quad (1)$$

where S is the true signal which counts for the abundance of

CENTRILLION

- **Models and Notations**
  - For each single array j:

$$\boldsymbol{X_j} \;=\; \boldsymbol{S_j} \;+\; \boldsymbol{B_j}$$

  - $X_j = \begin{cases} S_j + B_j, & j \in J \Rightarrow regular\ probes\ set \\ 0 + B_j = B_j, & j \in J_0 \Rightarrow negative\ control\ probes\ set \end{cases}$
  - $X_j \sim f_x(x)$, $S_j \sim f_s(s)$, $B_j \sim f_B(b)$, S_j and B_j are independent.
  - $N(\mu, \sigma^2) \Rightarrow f_{\mu,\sigma}^{norm}(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$
  - $\phi(x) \Rightarrow N(0,1)$, $\Phi(t) = \int_{-\infty}^{t} \phi(x)dx$
  - $Gamma(k,\theta) \Rightarrow f_{k,\theta}^{gam}(x) = \frac{\left(\frac{1}{\theta}\right)^k}{\Gamma(k)} x^{k-1} \exp\left\{-\frac{x}{\theta}\right\}$, k: shape parameter, θ: scale parameter
    $\xRightarrow[k=1,\theta=\alpha]{} Exp(\alpha) \Rightarrow f_{\alpha}^{exp}(x) = \frac{1}{\alpha} \exp\left\{\frac{x}{\alpha}\right\}$

CENTRILLION

- **Models and Notations**
  - $\boldsymbol{X_j} = \boldsymbol{S_j} + \boldsymbol{B_j}$, $X_j \sim f_x(x)$, $S_j \sim f_s(s)$, $B_j \sim f_B(b)$
  - By the convolution formula, $x_j = s_j + b_j \Rightarrow b_j = x_j - s_j \Rightarrow |J| = \left|\frac{db_j}{dx_j}\right| = 1$

$$\Rightarrow X_j \sim f_x(x) = \int_{-\infty}^{\infty} f_{X,S}(x,s)ds = \int_{-\infty}^{\infty} f_{S,B}(s, x-s)|J|ds = \int_{-\infty}^{\infty} f_{S,B}(s, x-s)ds = \int_{-\infty}^{\infty} f_S(s) f_B(x-s)ds$$

$\Rightarrow$ Estimated Signal: $\hat{S}(x) = E[S|X=x] = \int_{-\infty}^{\infty} S f_{S|X=x}(s)ds = \int_{-\infty}^{\infty} S \frac{f_{S,X}(s,x)}{f_X(x)} ds$

$$= \frac{\int_{-\infty}^{\infty} S f_{S,X}(s,x)ds}{\int_{-\infty}^{\infty} f_{S,X}(s,x)ds} = \frac{\int_{-\infty}^{\infty} S f_S(s) f_B(x-s)ds}{\int_{-\infty}^{\infty} f_S(s) f_B(x-s)ds}$$

  - Thus, if $f_x(x)$ is known $\Rightarrow \hat{S}(x)$ is known.
  - No analytic expression $\Rightarrow$ Fast Fourier Transformation-based (fft) approximation.

CENTRILLION

- **The normexp Model**
  - $S_j \sim f_s(s) = \begin{cases} Exp(\alpha), & j \in J \\ 0, & j \in J_0 \end{cases}, \; B_j \sim f_B(b) \Rightarrow N(\mu, \sigma^2)$

    $\Rightarrow X_j \sim f_X(x) \equiv f_{\mu,\sigma,\alpha}^{nexp}(x) = \frac{1}{\alpha} \exp\left\{\frac{\sigma^2}{2\alpha^2} - \frac{x-\mu}{\alpha}\right\} \Phi(\bar{x}), \quad where\ \bar{x} = \frac{\left(\mathrm{x} - \mu - \frac{\sigma^2}{\alpha}\right)}{\sigma}$

    $\Rightarrow \hat{S}^{nexp}(x|\Theta) = \sigma\left(\bar{x} + \frac{\phi(\bar{x})}{\Phi(\bar{x})}\right), \;\; \Theta = (\mu, \sigma, \alpha)$
  - If we know $(\hat{\mu}, \hat{\sigma}, \hat{\alpha}) \;\Rightarrow$ we know $\hat{S}^{nexp}(x)$

- **The Parameter Estimation of normexp Model**

A. MLE
B. Adapted RMA
C. Non-parametric estimation (NP)
D. Bayesian estimation

Methods and Materials

- **The normal-gamma Model**
  - $S_j \sim f_s(s) = \begin{cases} Gamma(k,\theta), & j \in J \\ 0, & j \in J_0 \end{cases}, \; B_j \sim f_B(b) \Rightarrow N(\mu, \sigma^2)$

    $\Rightarrow X_j \sim f_X(x) \equiv f^{ng}_{\mu,\sigma,k,\theta}(x) = \int f^{gam}_{k,\theta}(t) f^{norm}_{\mu,\sigma}(x-t)dt \Rightarrow fft - based\ approximation$

    $\Rightarrow \hat{S}^{ng}(x|\Theta) = \frac{\int s f^{gam}_{k,\theta}(s) f^{norm}_{\mu,\sigma}(x-s)ds}{f^{ng}_{\mu,\sigma,k,\theta}(x)} = \frac{k\theta\left(\int f^{gam}_{k+1,\theta}(s) f^{norm}_{\mu,\sigma}(x-s)ds\right)}{f^{ng}_{\mu,\sigma,k,\theta}(x)}$

    $= \frac{k\theta f^{ng}_{\mu,\sigma,k+1,\theta}(x)}{f^{ng}_{\mu,\sigma,k,\theta}(x)} \Rightarrow fft - based\ approximation$
  - If we know $(\hat{\mu}, \hat{\sigma}, \hat{k}, \hat{\theta}) \Rightarrow$ we know $\hat{S}^{ng}(x) \Rightarrow \hat{S}_j = \hat{S}^{ng}(x_j)$
- **The Parameter Estimation of normal-gamma Model**

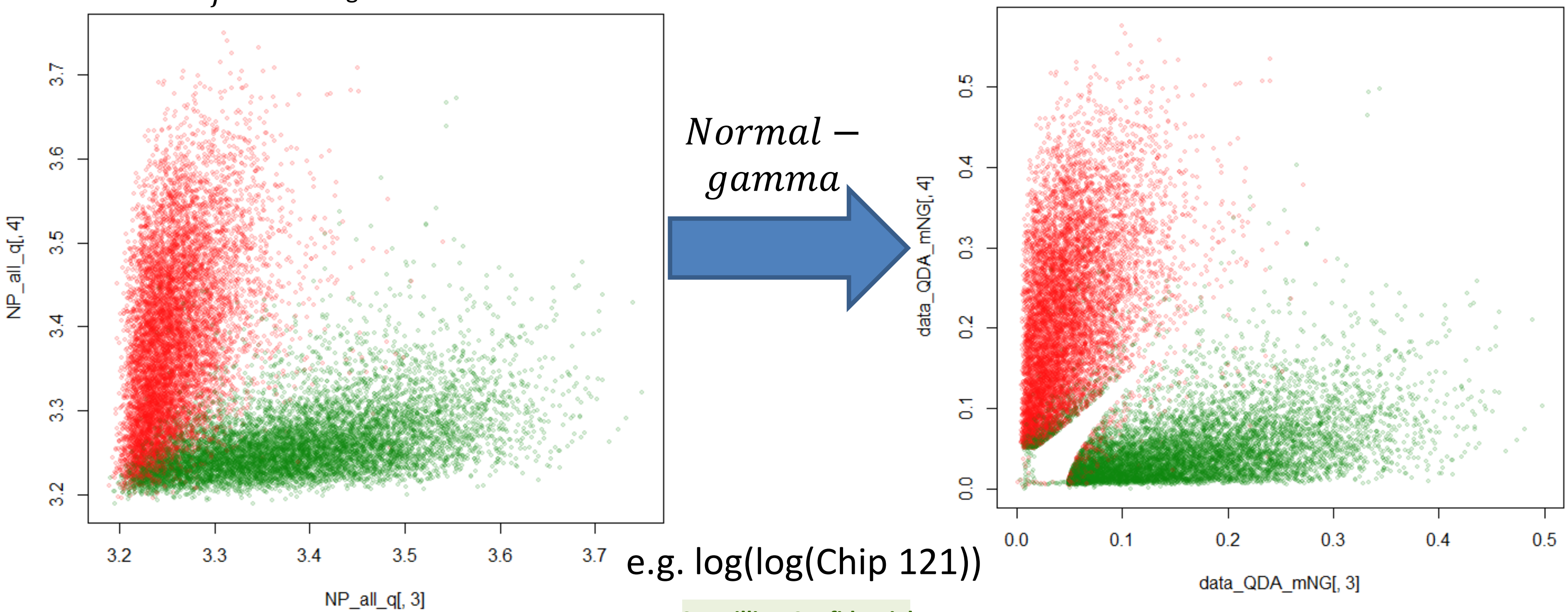
  A. MLE with classical minimization algorithms (L-BFGS-B)

Performance on the Real Data

- Normal-gamma BgC.
  - For each single probe in each cluster (channel):

$$X_j \;=\; S_j \;+\; N_j$$

  - $Normal - gamma$ Correction: $X_j \Rightarrow S_j$. Enhance the biological validity of the results.
  - N_j repre sents the noise. ⇒ Assumption: Normal distribution.
  - Remove N_j and u se Gamma distribution to estimate S_j.
  - X_j ⇒ Normal-gamma distribution.



e.g. log(log(Chip 121))

Performance on the Real Data

CENTRILLION

- GMM-EM + Normal-gamma BgC.
    - Set the related environment in R. (Data Preprocess, NP Probes, QN & log, QDA)
    - Set the corresponding evaluation tools in R. (NP call rate, No call rate)
    - Run the GMM-EM in R. (10 times => max NP call.)
    - Run the normal-gamma correction (all together) before running QDA.
    - Debug for the no call rate in the NP call analyzer.
    - An example: NP call Analyzer: NP call: 94.4%, call: 66.9%

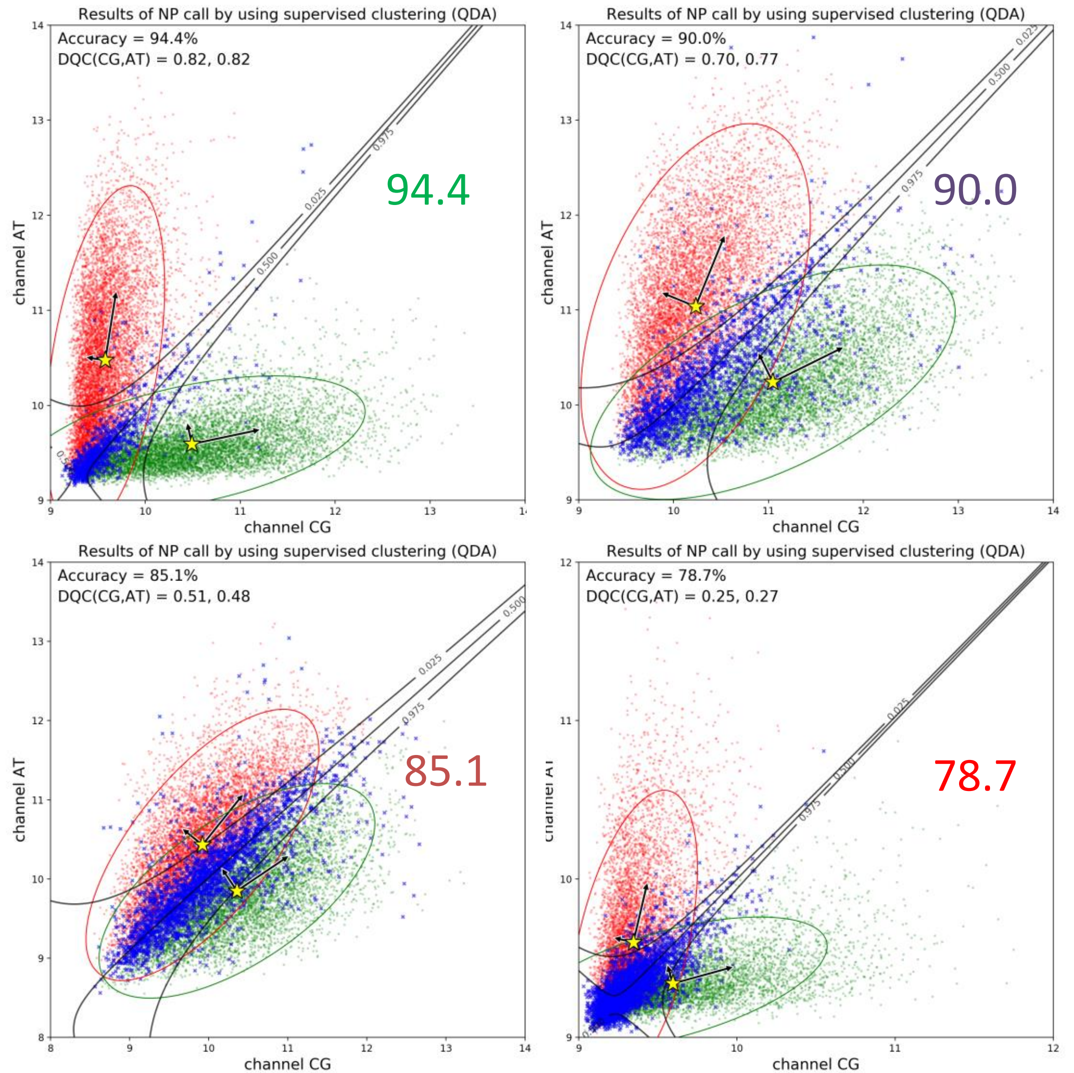
NP call (log(log(·)))	
GMM + EM	94.2308%
QDA	94.5863%
ALL_NG + QDA	94.2774%

Call rate (log(log(·)))	
GMM + EM	66.8298%
QDA	67.8205%
ALL_NG + QDA	67.7681%

Data Preprocess

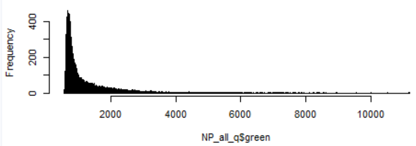
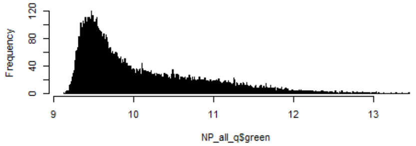
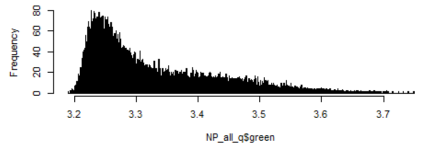
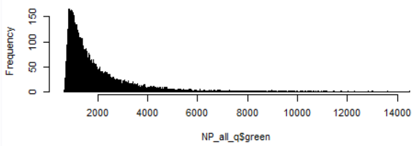
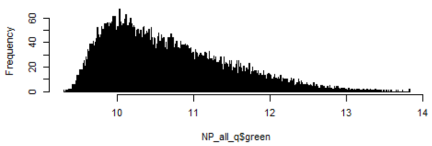
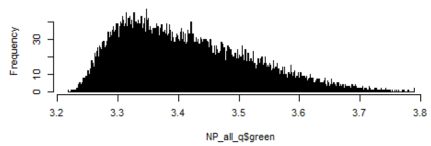
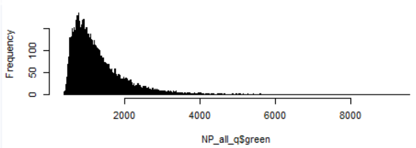
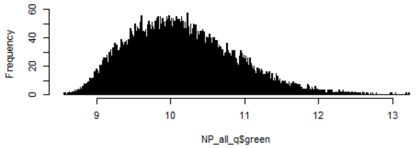
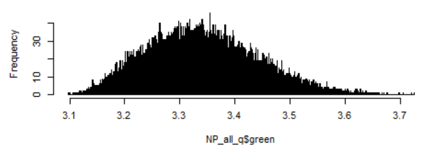
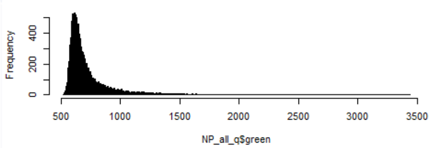
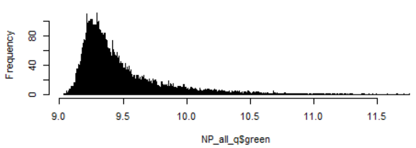
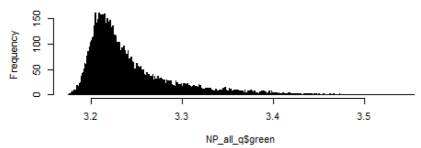
- Data Example

Wafer	Chips	Np call rate (%) (NP call Analyzer)
198-04	121	94.4
197-02	230	90.0
198-15	277	85.1
197-02	233	78.7



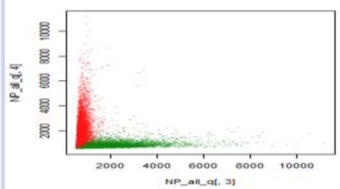
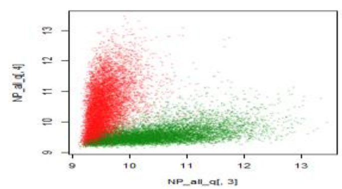
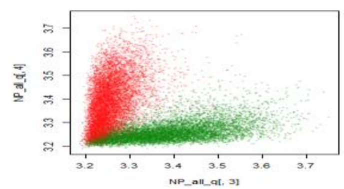
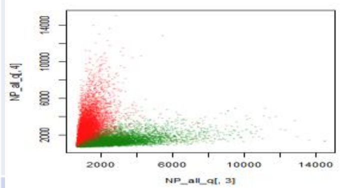
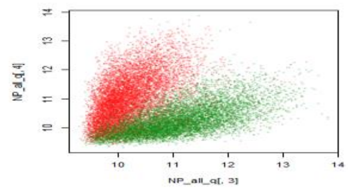
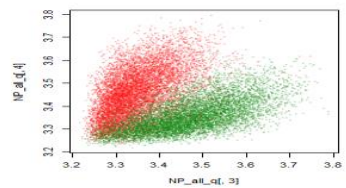
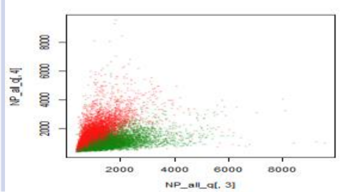
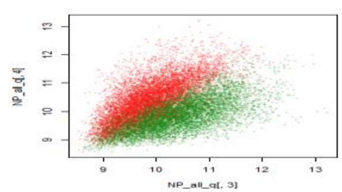
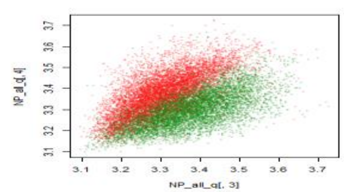
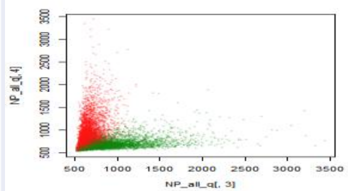
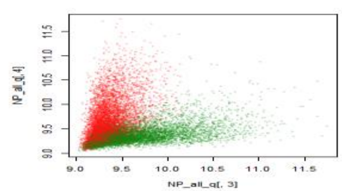
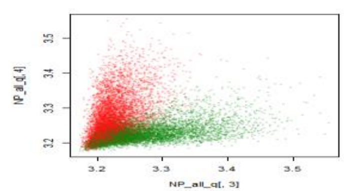
Data Preprocess

- Logarithm Effect

Chips	Original Raw Data	Log(·)	Log(Log(·))
121			
230			
277			
233			

Data Preprocess

- Logarithm Effect

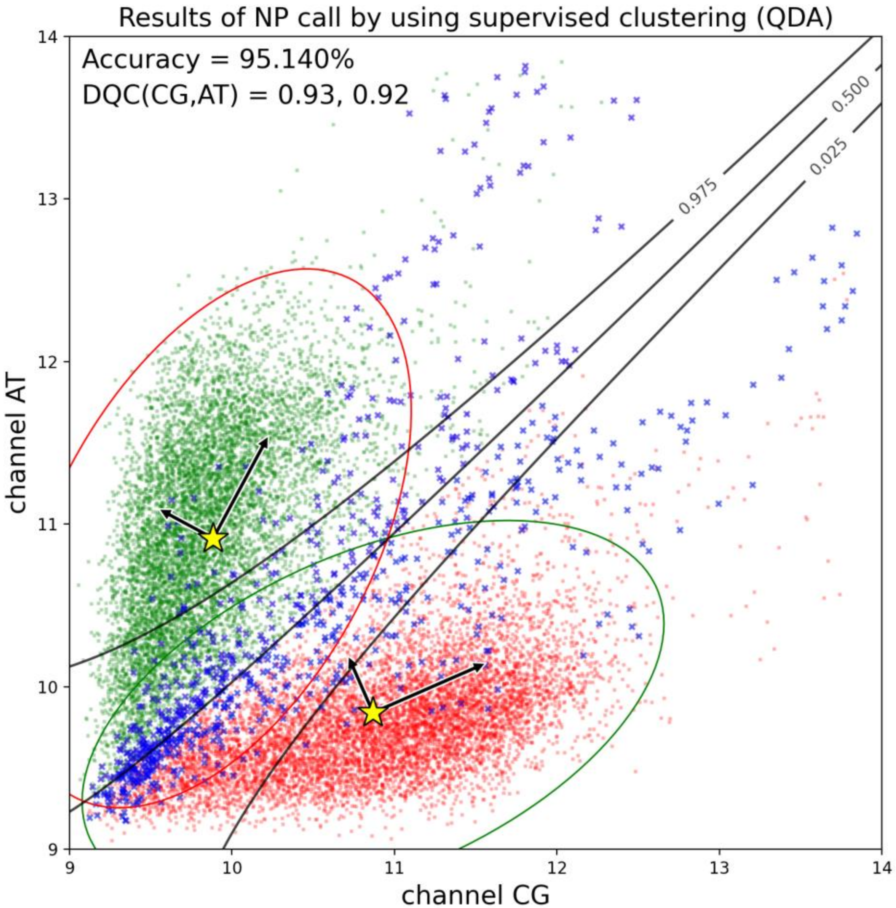
Chips	Original Raw Data	Log(·)	Log(Log(·))
121			
230			
277			
233			

Chip QC and NPcall Analyzer

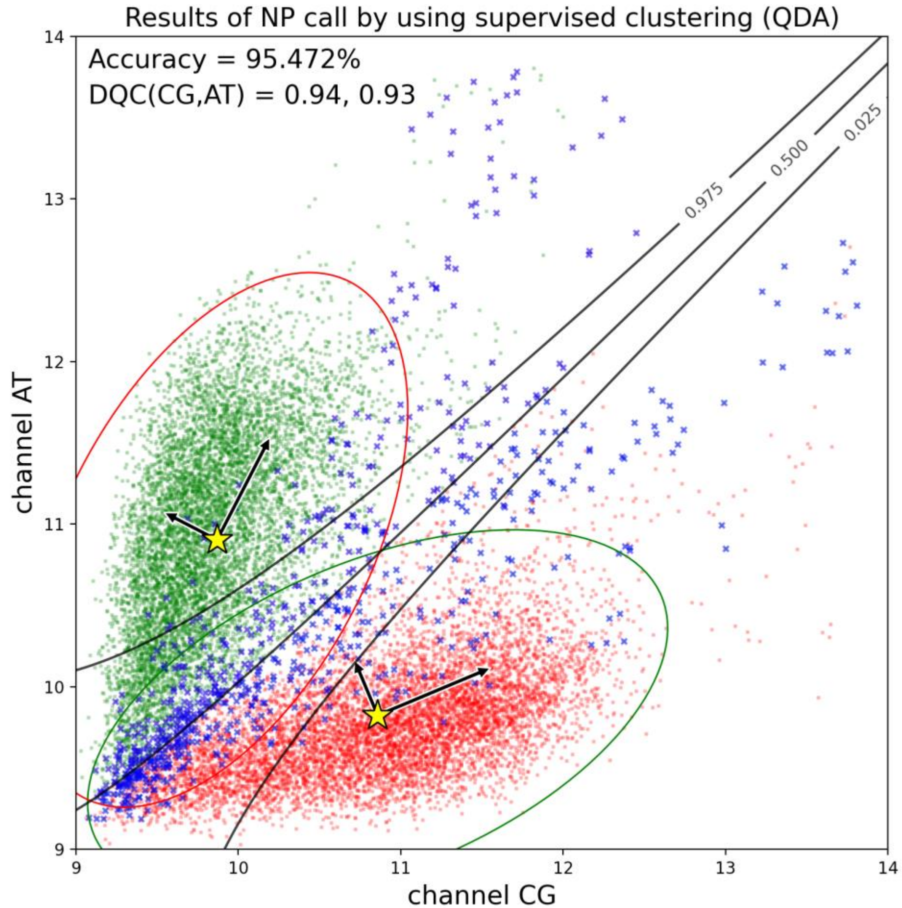
Jeff (CHI-HSUAN HO)

Summit.NPCallAnalyzer

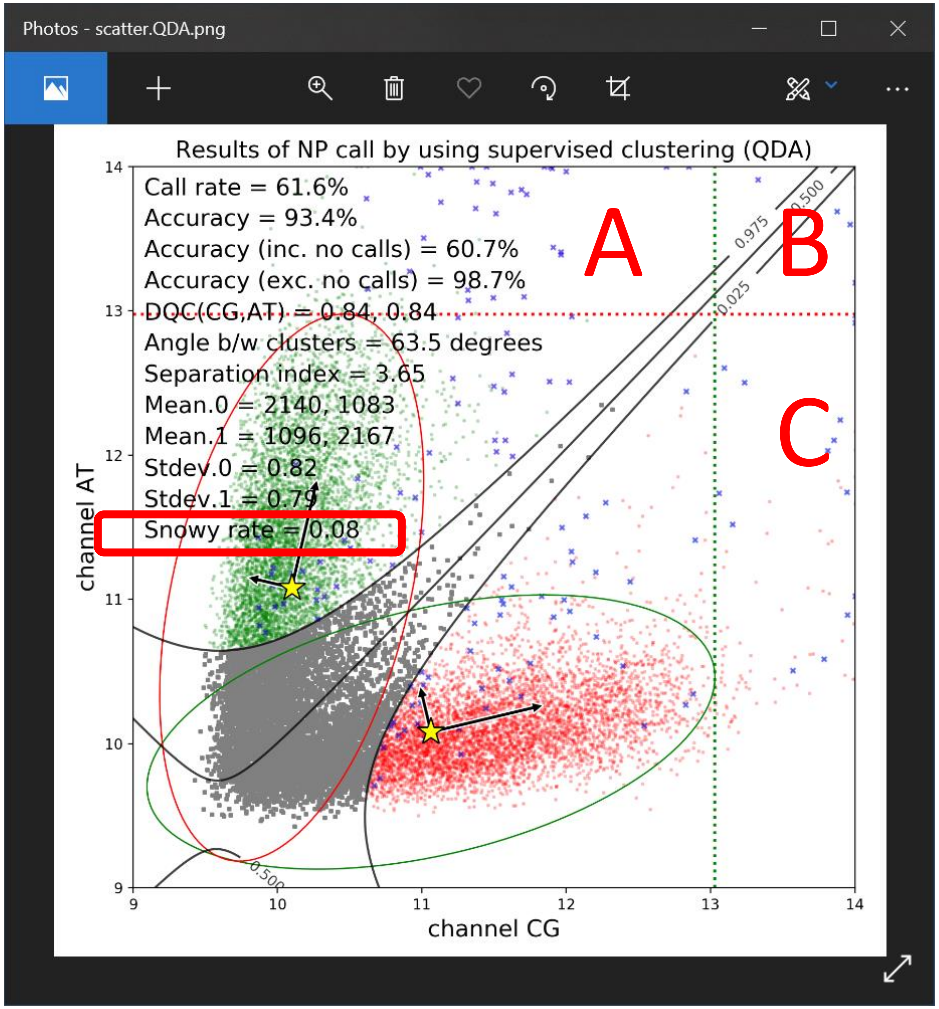
Chip No.54 Quality Control



Chip NO.62 Quality Control



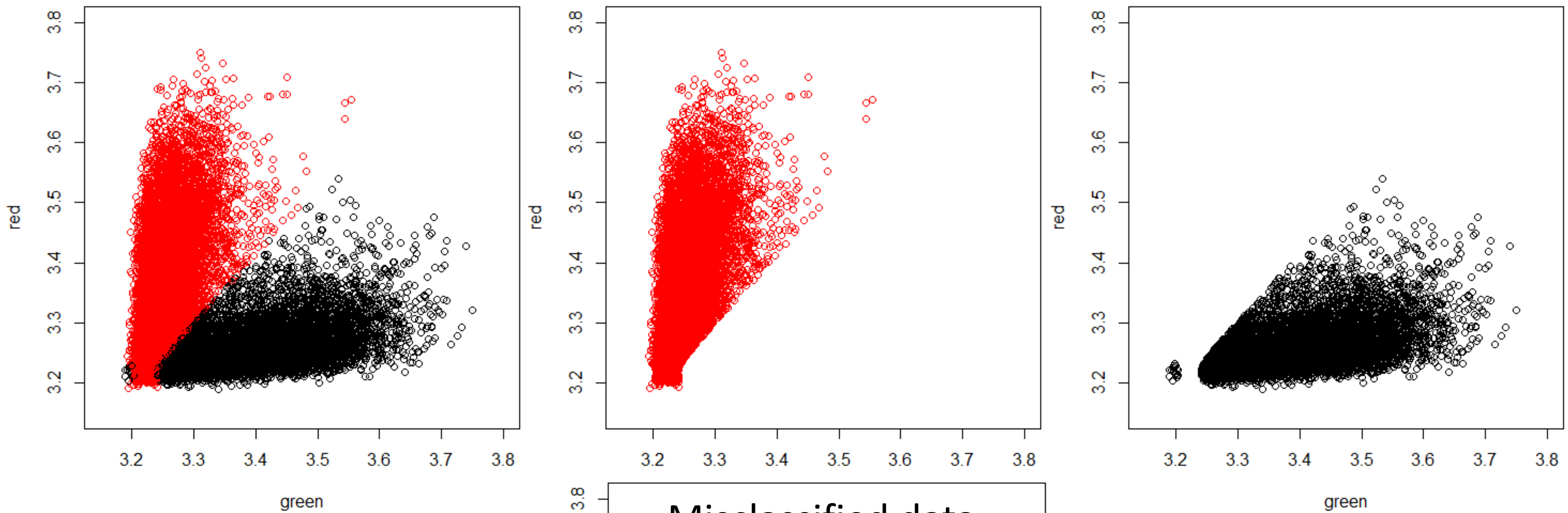
Summit.NPCallAnalyzer



- Snowy rate $= \dfrac{x \text{ in } B}{\text{all points in } A+B+C}$

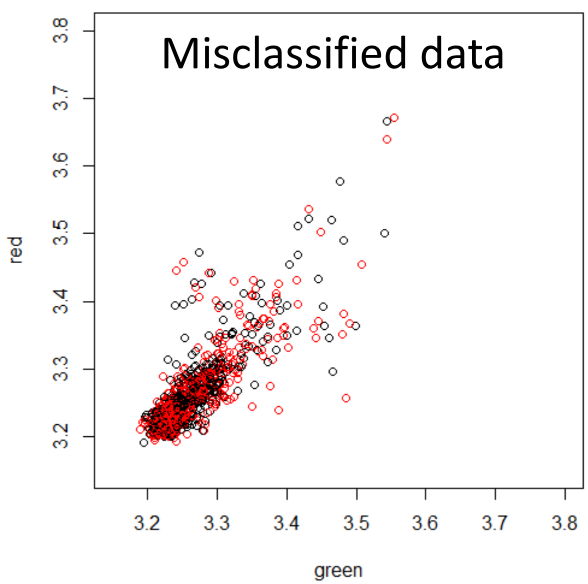
Summit.NPCallAnalyzer

- GMM-EM Results (log(log))

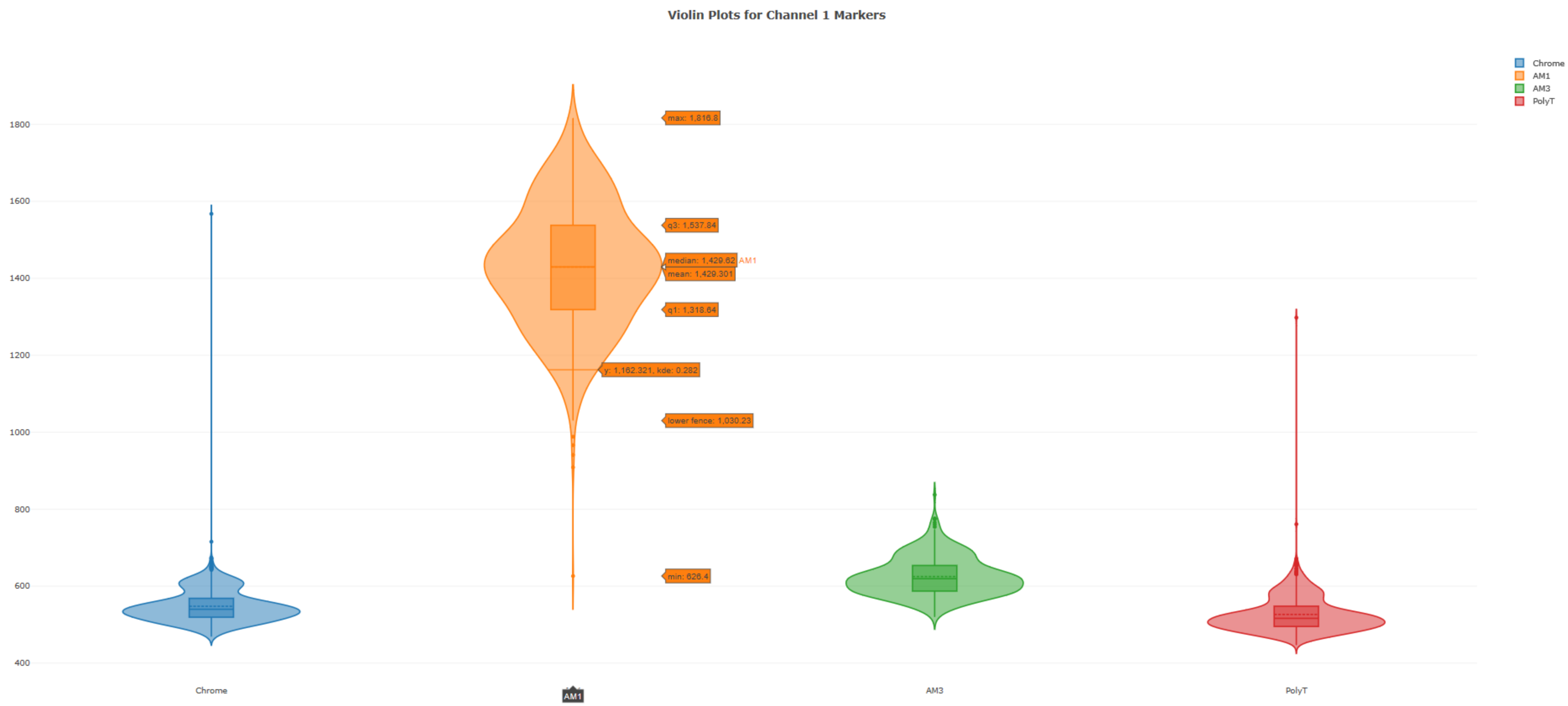


NP call: 94.2308%

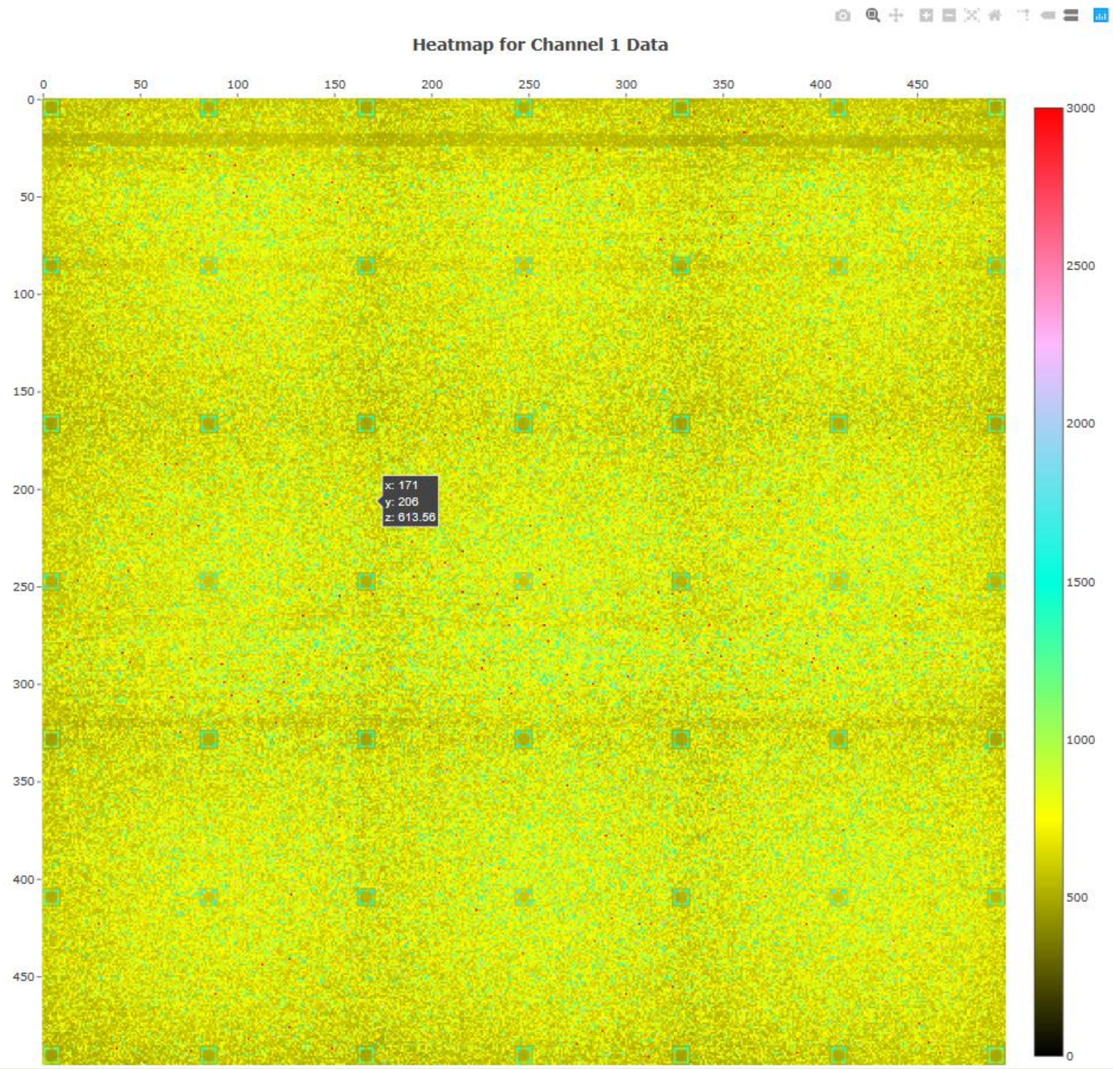
No Call Rate: 66.8298%



Banff chip QC – Violin Plot



Banff chip QC – Heatmap



Banff chip QC – Marker Raw Data

CENTRILLION

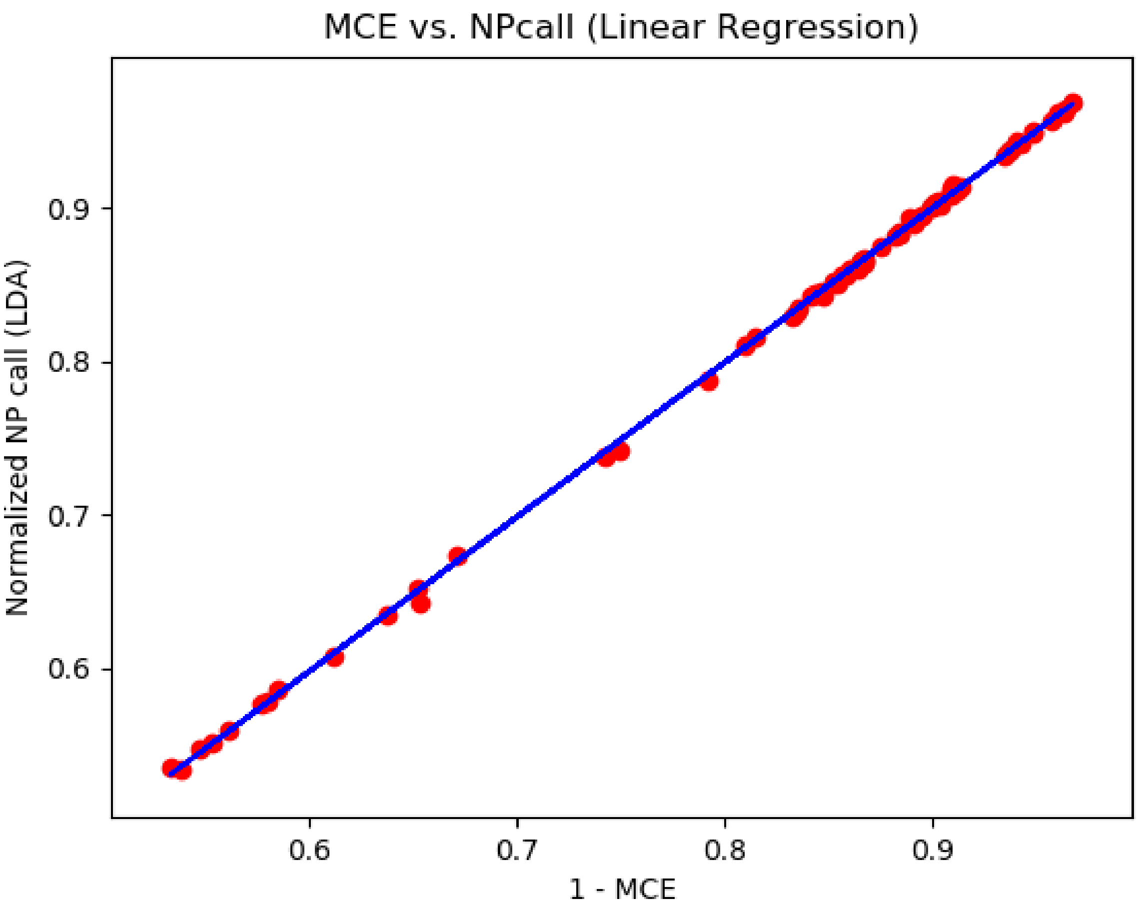
marker_ch1.csv - Excel

A1 487.24

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	
1	487.24	488.16	485.8	492.36	502.04	482.16	503.44	491.17	505.72	500.8	nan	nan	nan	nan	nan	nan	nan	nan	nan	na
2	492.84	908.88	1176.56	1107.6	604.4	610.56	1087.88	1232.97	1247.84	512.56	nan	nan	nan	nan	nan	nan	nan	nan	nan	na
3	505.84	941	503.44	487.92	471.76	481.48	501.96	520.93	1232.72	521.52	nan	nan	nan	nan	nan	nan	nan	nan	nan	na
4	512.08	626.4	499.52	493.76	471.12	500.76	500.08	515.4	1231.2	503.76	nan	nan	nan	nan	nan	nan	nan	nan	nan	na
5	496.16	561.28	493.92	488.36	489.32	492.52	486.2	503.97	542.04	496.28	nan	nan	nan	nan	nan	nan	nan	nan	nan	na
6	501.4	569.76	504.92	534.2	480.52	498.84	485.92	516.13	621.84	524.84	nan	nan	nan	nan	nan	nan	nan	nan	nan	na
7	499.88	1113.76	501.52	489.4	500.6	477.44	506.6	490.5	1230.76	521	nan	nan	nan	nan	nan	nan	nan	nan	nan	na
8	514.33	1297.9	525.1	479.67	493.3	478.57	508.27	502.22	1192.9	547.2	nan	nan	nan	nan	nan	nan	nan	nan	nan	na
9	520.96	1349.44	1247.16	1273.52	683.08	607	1319.56	1300.5	1266.76	526.32	nan	nan	nan	nan	nan	nan	nan	nan	nan	na
10	502.52	514.36	528.24	507.64	513.48	503.16	524.92	520.63	537.84	509.72	nan	nan	nan	nan	nan	nan	nan	nan	nan	na
11	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	na
12	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	na
13	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	na
14	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	na
15	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	na
16	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	na
17	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	na
18	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	na

marker_ch1

Correlation and Regression analysis (MCE vs. NP call)



Training data: 80% data
Model: $NPcall = -0.005 + 1.005 \cdot (1 - MCE)$
R^2: 99.96%

Testing data: 20% data
MSE: 5.999574703240224e-06

It still need NP data to calculate the MCE.