# シミュレーションを併用した重イオンビームトランスポートラインのベイズ最適化に関する研究

# Study for Bayesian optimization of heavy ion beam transport using simulation

2023 11/27

理研仁科加速器科学研究センター 西　隆博

共同研究者

岩井 瑛人[a,c], 内山 暁仁[b], 清水 陽平[b], 杉本 崇[a], 炭竃 聡之[b],
福田 直樹[b], 藤井 洋樹[b], 前坂 比呂和[c]

a. 公益財団法人高輝度光科学研究センター
b. 理研仁科加速器科学研究センター
c. 理研放射光科学研究センター

# アウトライン

・理研 RIBF と機械学習を用いたビーム輸送最適化

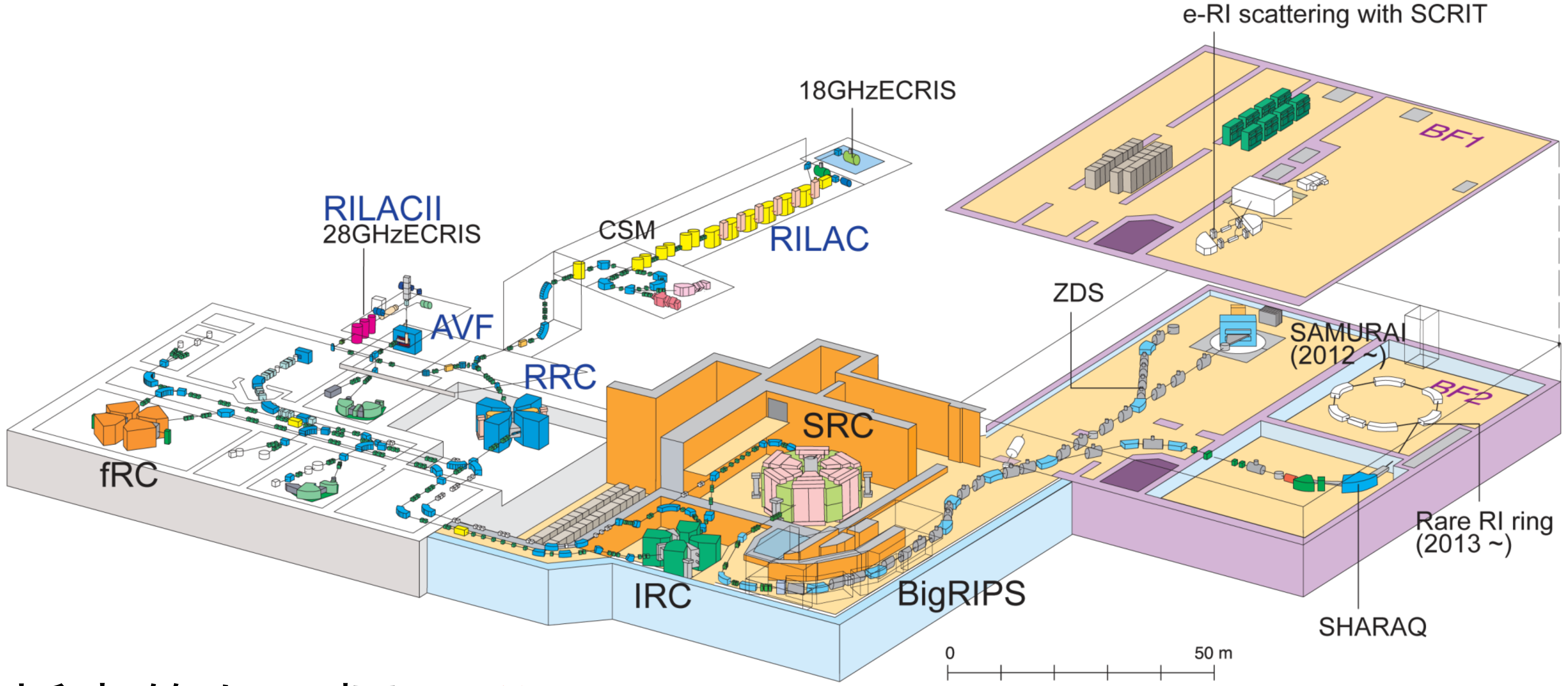
・ベイズ最適化のためのシミュレーションの概要

・ベイズ最適化のハイパーパラメータ
(acquisition function / objective function / phase ellipse)

・各ハイパーパラメータでのシミュレーション結果の比較

・解析の現状まとめと今後の予定

# 理研 RIBF と機械学習を用いたビーム輸送最適化




将来的な目標: **> 1 pμA** $^{86+}$U beam
(×10 times upgrade)

⇔ 局所的なビームロス ~ **0.2%** で機器を壊しうる。

目標の高強度ウランビーム ( > 1puA / 100 kW)
の達成には、

- **局所的なビームロスを 0.1 ~ 1% に抑える**
- **環境などの変化に追随する常時最適化**

が必要となる

**ビーム輸送パラメータの常時自動最適化**

- 600 以上のコントロールパラメータ
- その他環境変数

**Control by ML**

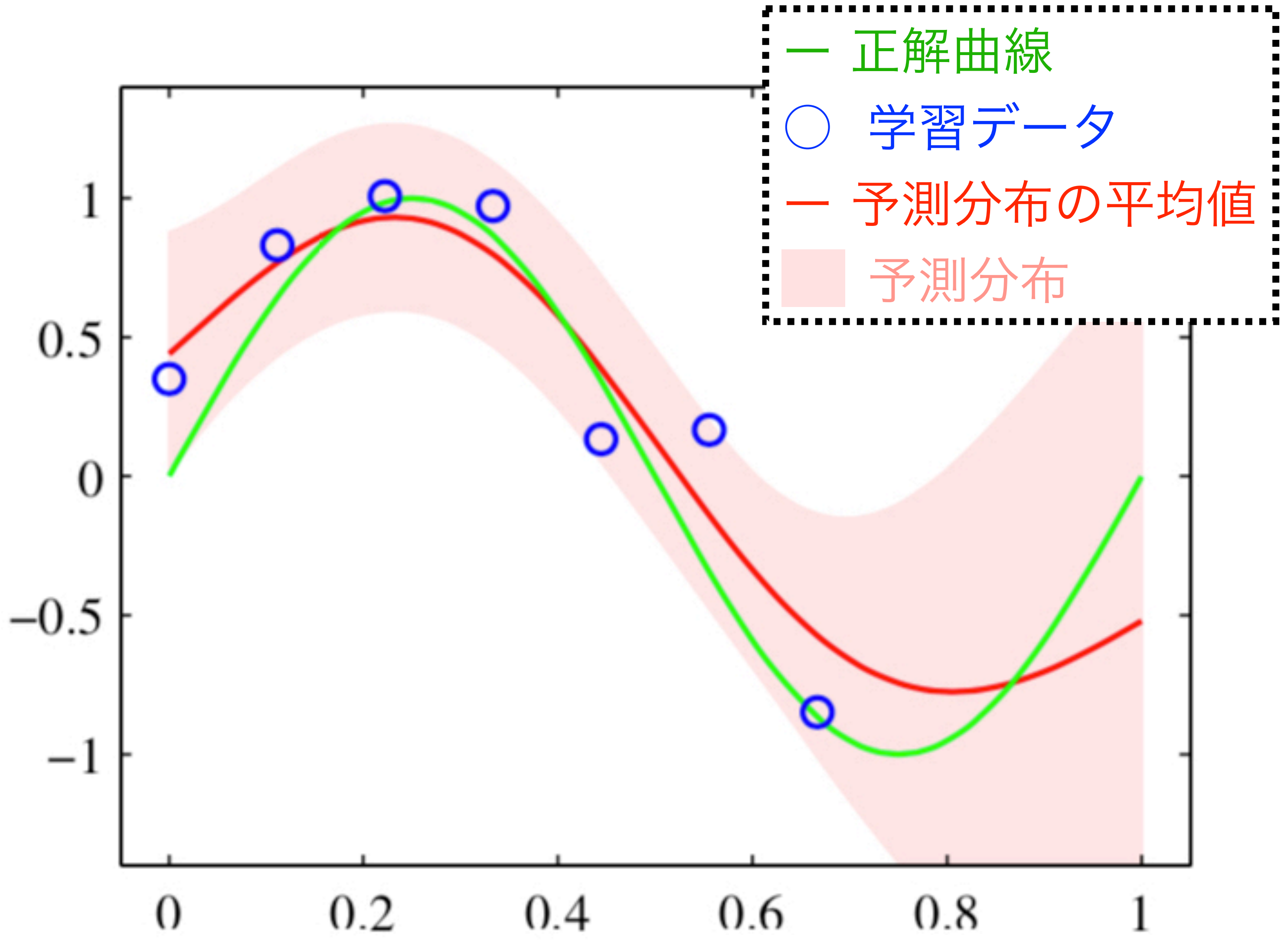
**シミュレーションと比べて理解可能なモデルの構築**

- 非破壊ビーム検出器
- ビームロスモニター
- 下流検出器

# パラメータの最適化手法: ガウス過程回帰を用いた逐次学習

**SACLA@Spring-8 で実用化されている機械学習による自動調整**
**→ RIBF に応用し、将来的な高精度のビーム調整に繋げたい**

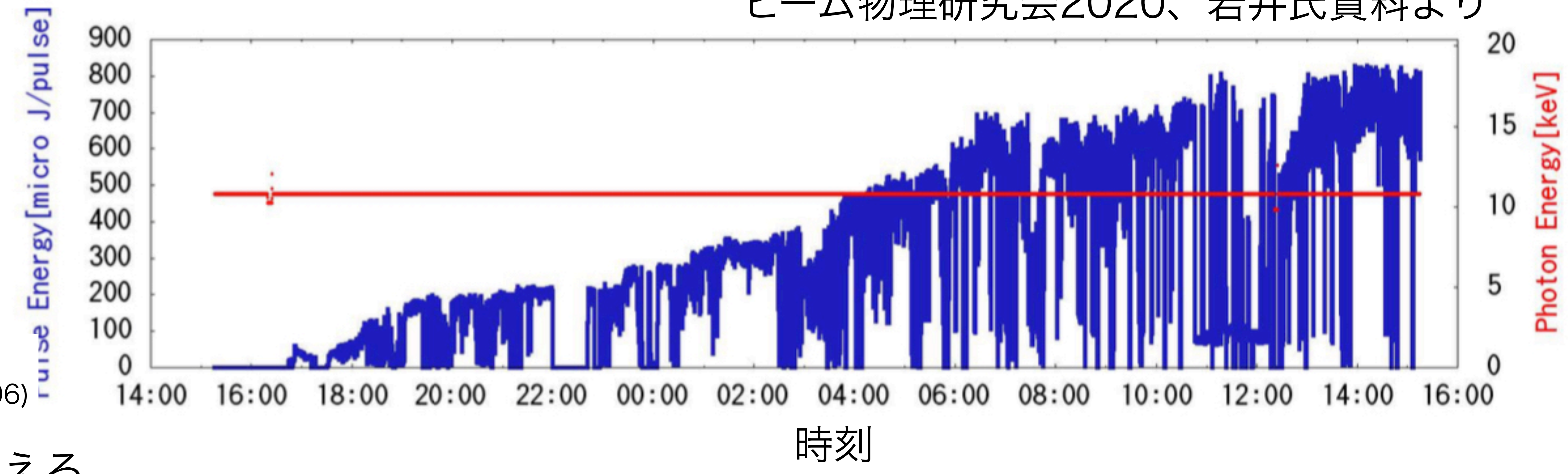
ガウス過程回帰




C. M. Bishop, Pattern Recognition and Machine Learning, Springer (2006)

Gaussian Process を用いた逐次学習による自動調整
XFEL の強度を最適化した例 (~15 パラメータ同時最適化)

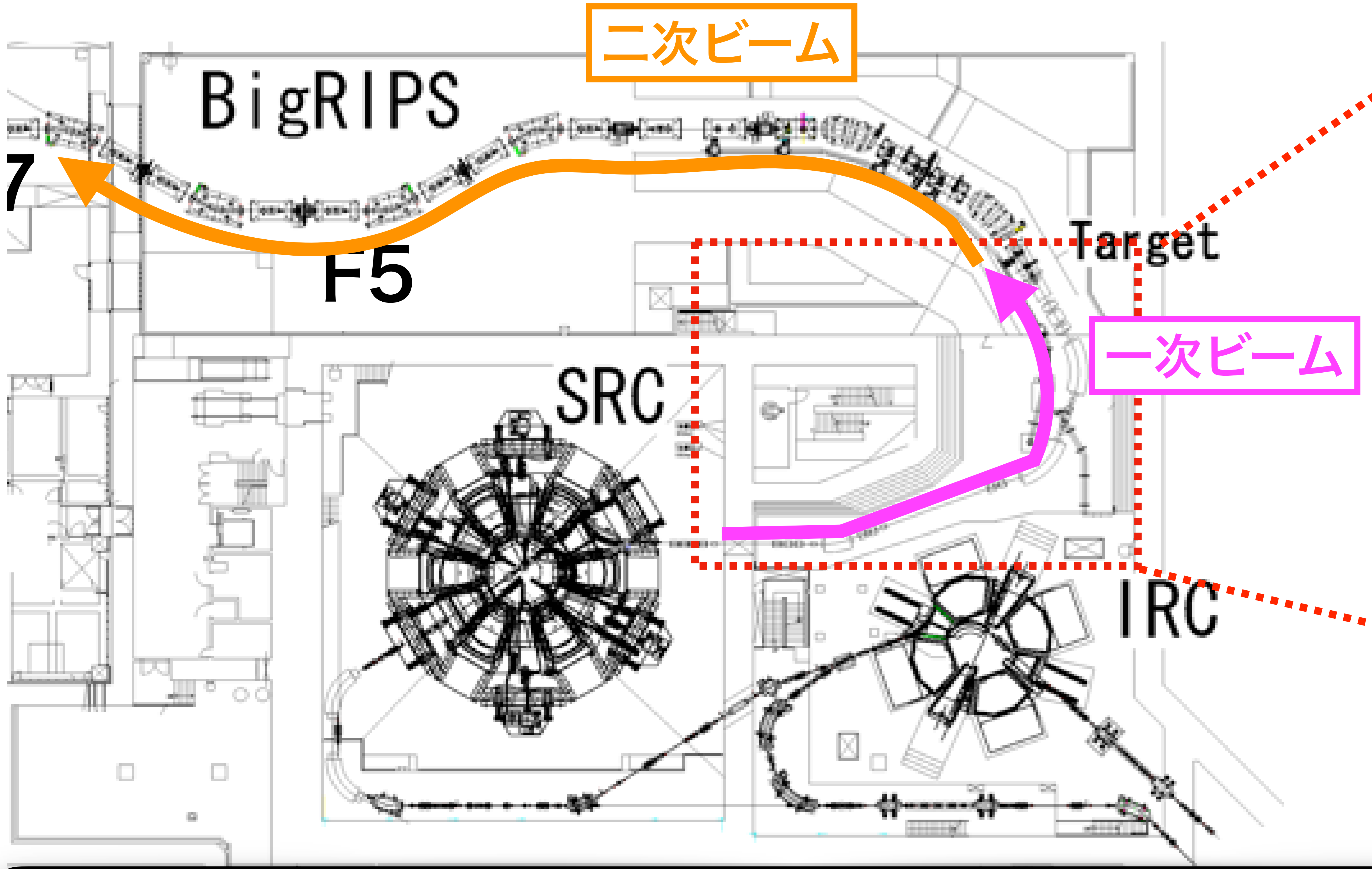
ビーム物理研究会2020、岩井氏資料より

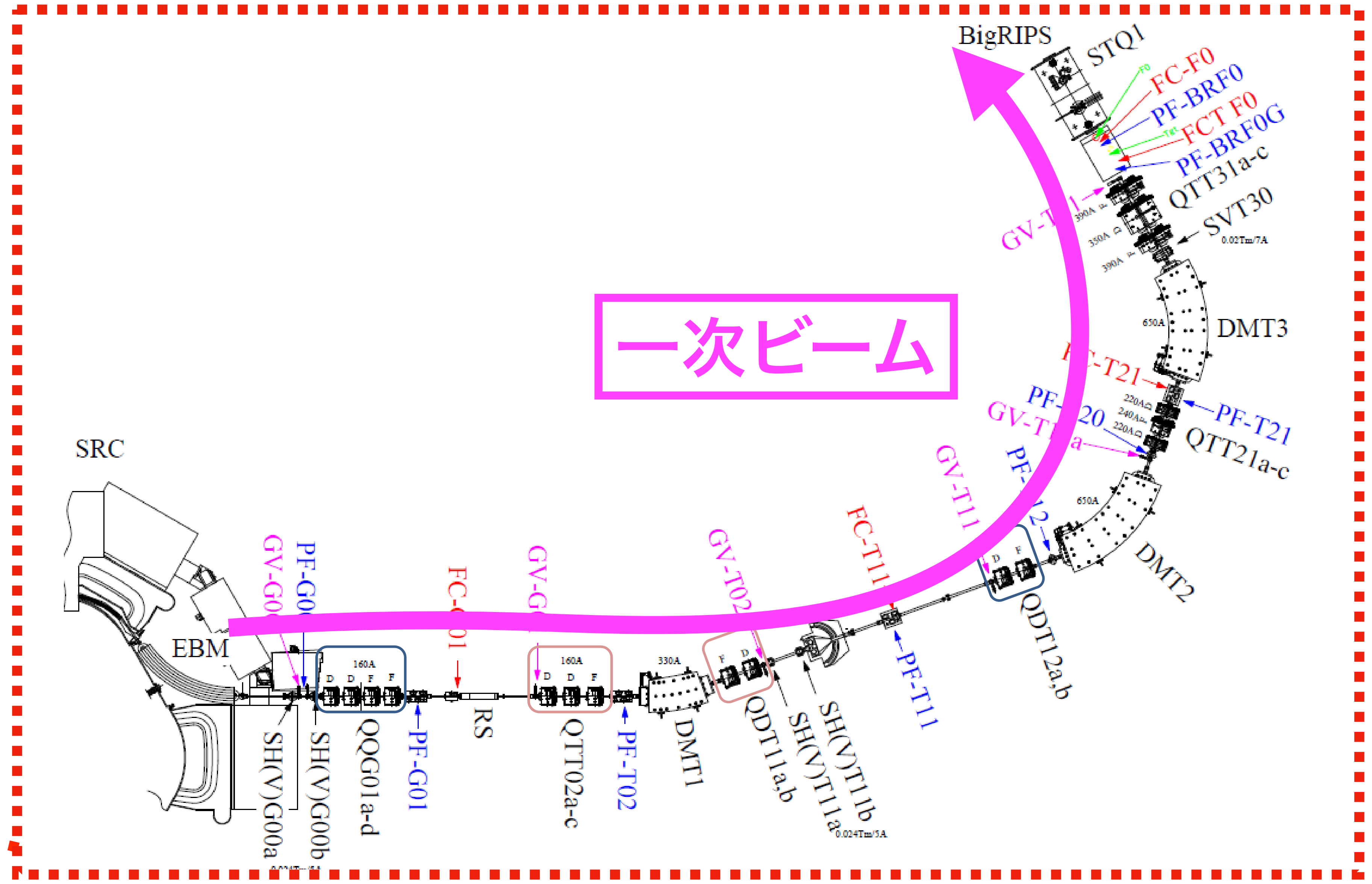



→ JASRI 岩井さん
E. Iwai et al, Proc. of PASJ 2021 WEOB02

- 誤差、不定性を持つ多次元データを取り扱える
- local min. にハマりにくい
- **比較的少ない試行回数で最適化が可能**

# RIBF への機械学習の導入の第一歩: 一次ビーム最下流の光学系調整



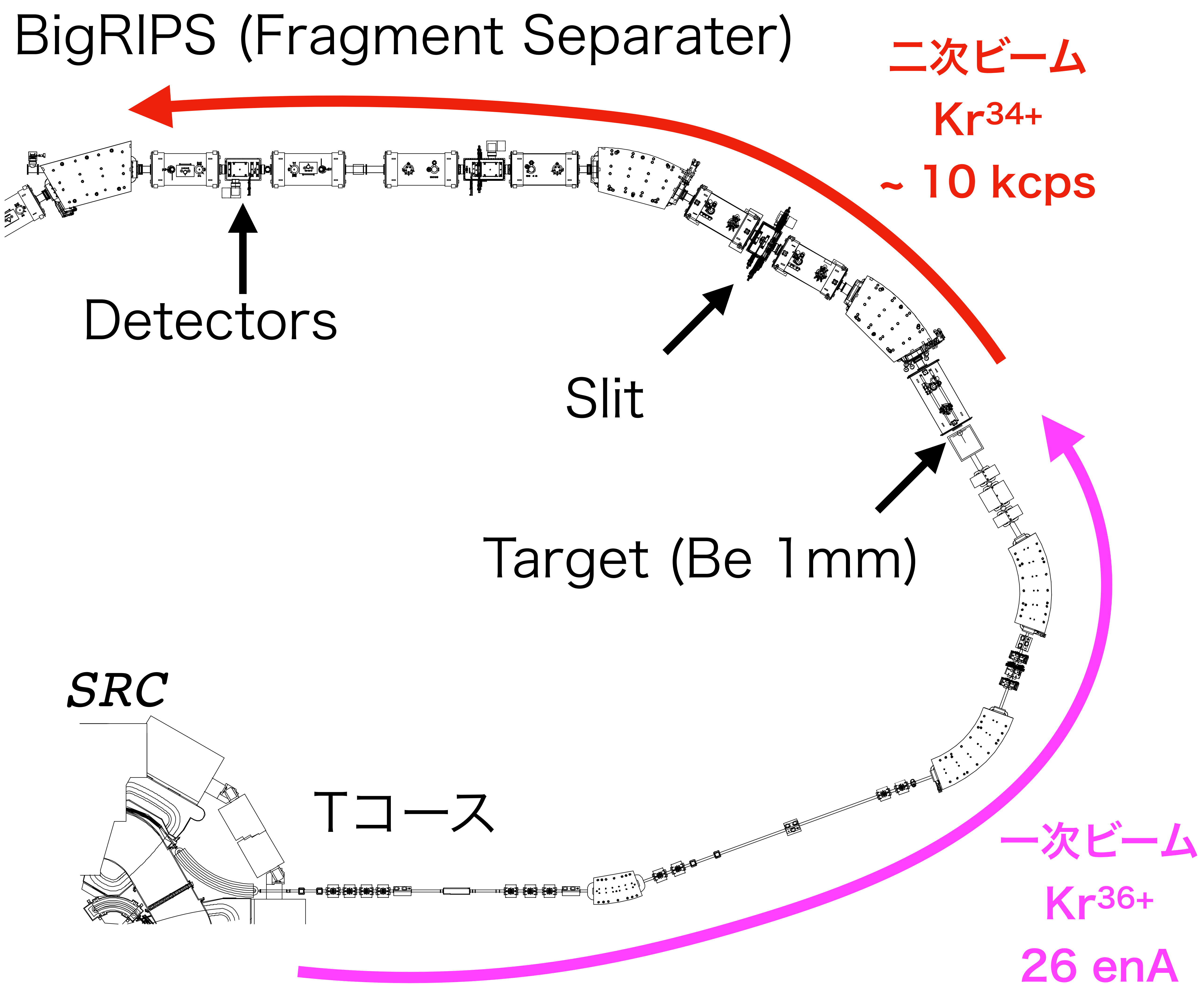

Tコース: SRC から BigRIPS までをつなぐ一次ビームライン
Quadrupole × 17, Steerer × 5, Dipole × 3

- 下流に Fragment Separator である BigRIPS
  豊富な検出器などの情報からデータを集めやすい
- 通常オペレーションの場合、最もロスの大きい
  ウランビームでは通過効率 ~ 90% 程度と見積もられる
- 通過効率と小さいスポットを同時に実現することを目指す

将来的な目標である **1 pμA** の
ウランビーム実現には、
- **ビームロスを 1 ~ 0.1 % 程度に抑える**
- **像は小さく($\sigma_h$ ~ < 1 mm) 保つ**

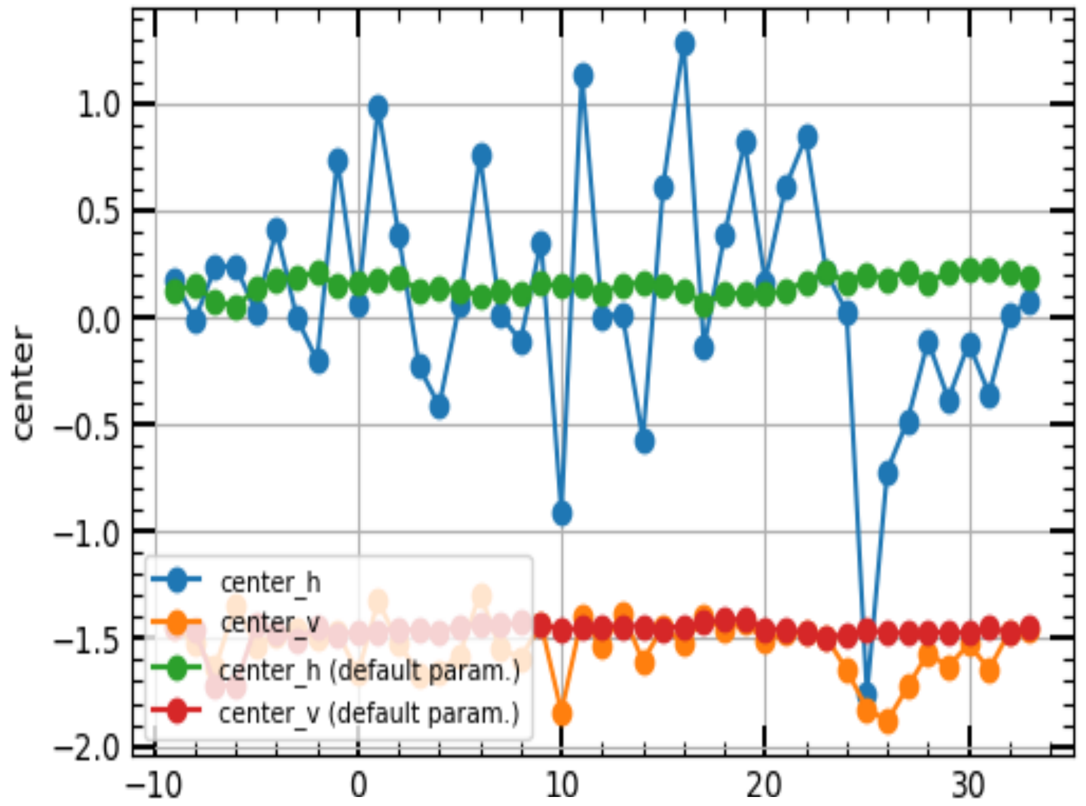
ことが必要になる。

# “高強度”一次ビームで荷電変換粒子の測定による光学系最適化

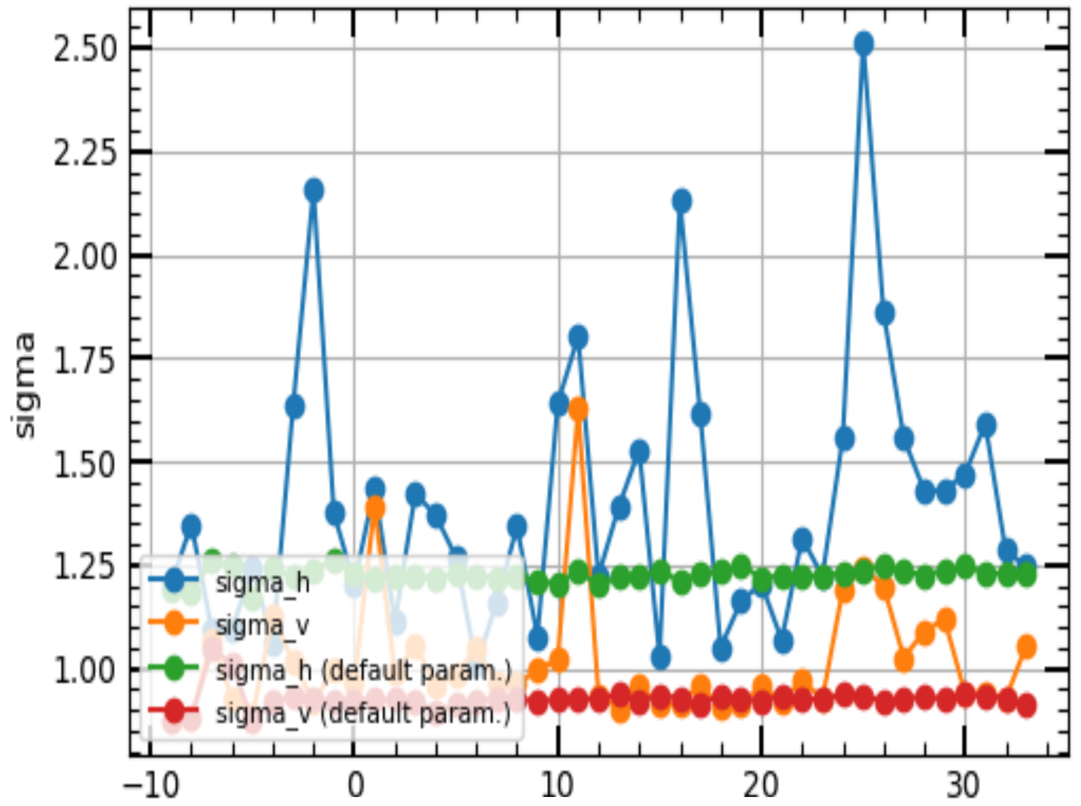



下流検出器
・シンチレータ (for 通過効率)
・飛跡検出器 (for スポット)
で $Kr^{34+}$ **10 kcps** / 0.0001 enA を測定
これを指標に、**26 enA** の $Kr^{36+}$ ビームの最適化を行う。

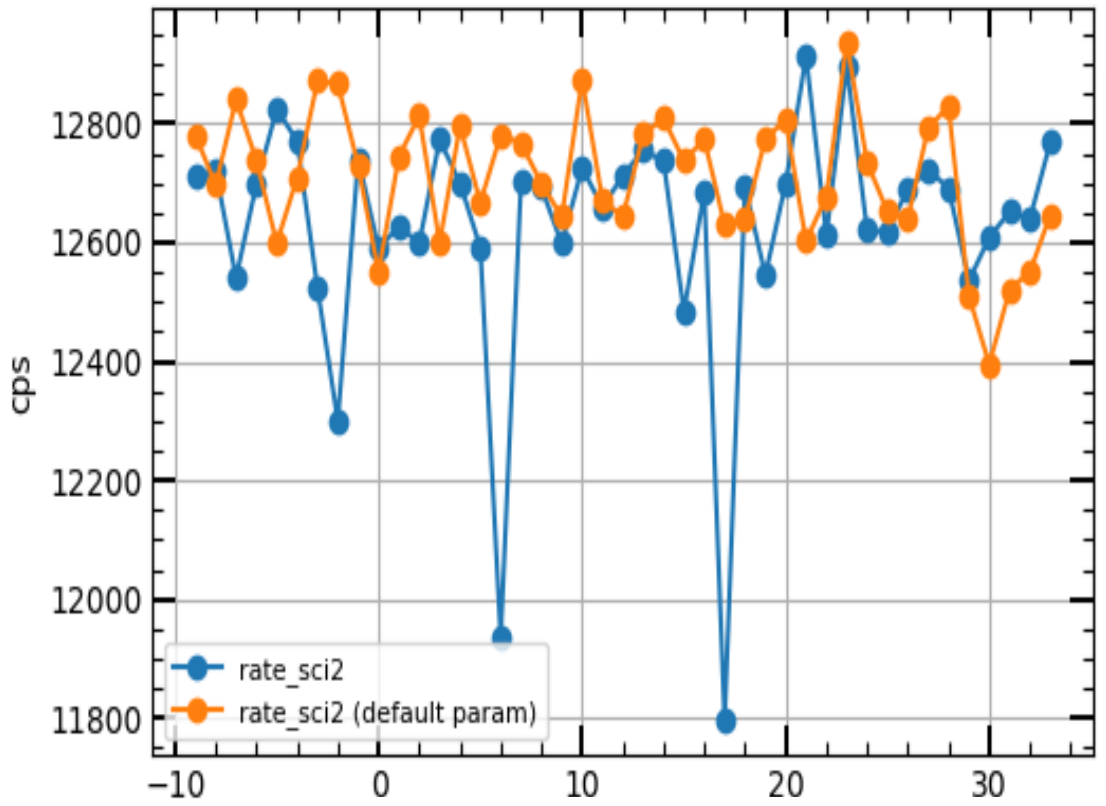
# “高強度”一次ビームで荷電変換粒子の測定による光学系最適化



$Pos_{h,\,v}$ vs Epoch



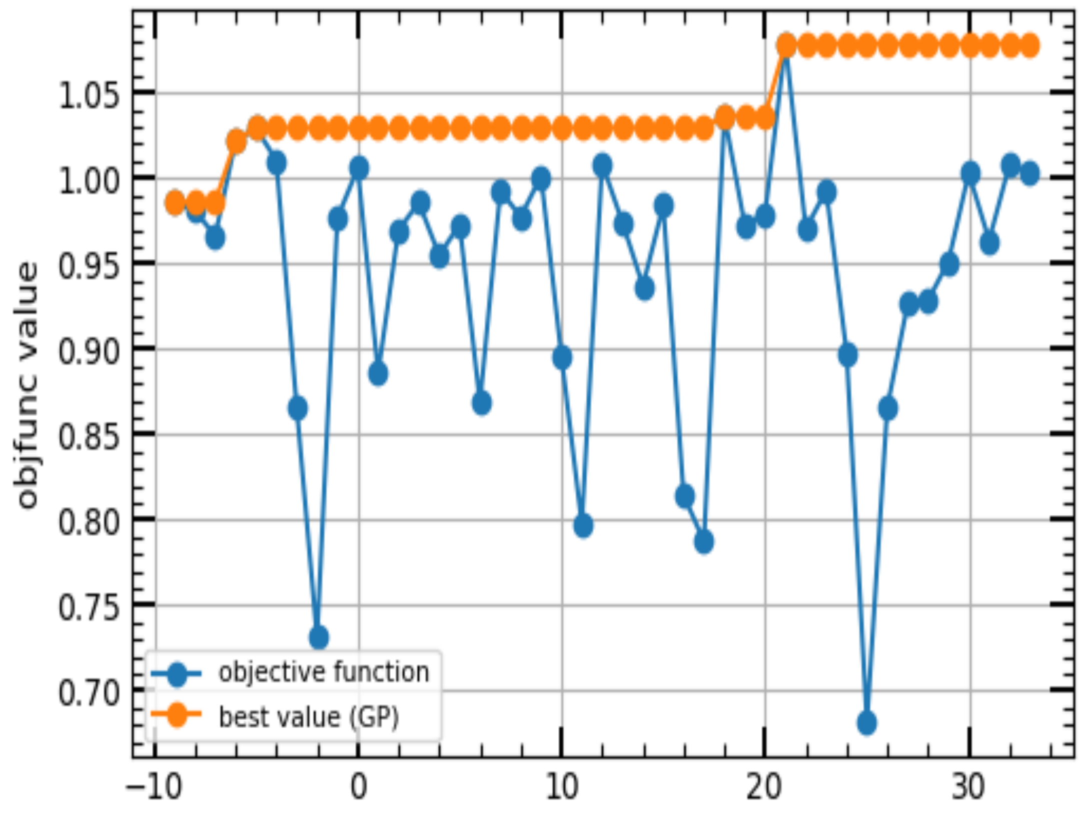
$\sigma_{h,\,v}$ vs Epoch



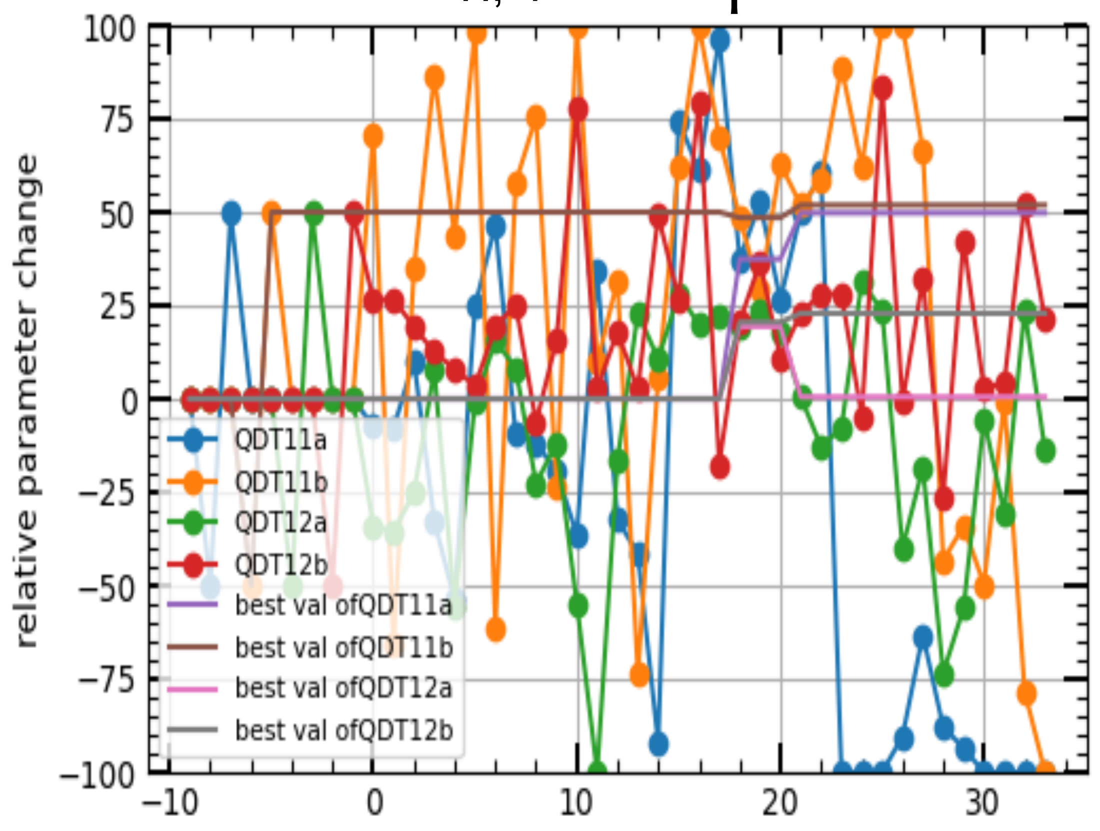
Count rate vs Epoch

- 4 変数 (四重極電磁石)
- 1 epoch ~ 25 s
- 一度の最適化 ~ 30 分程度
- 一次ビーム ~ 26 enA

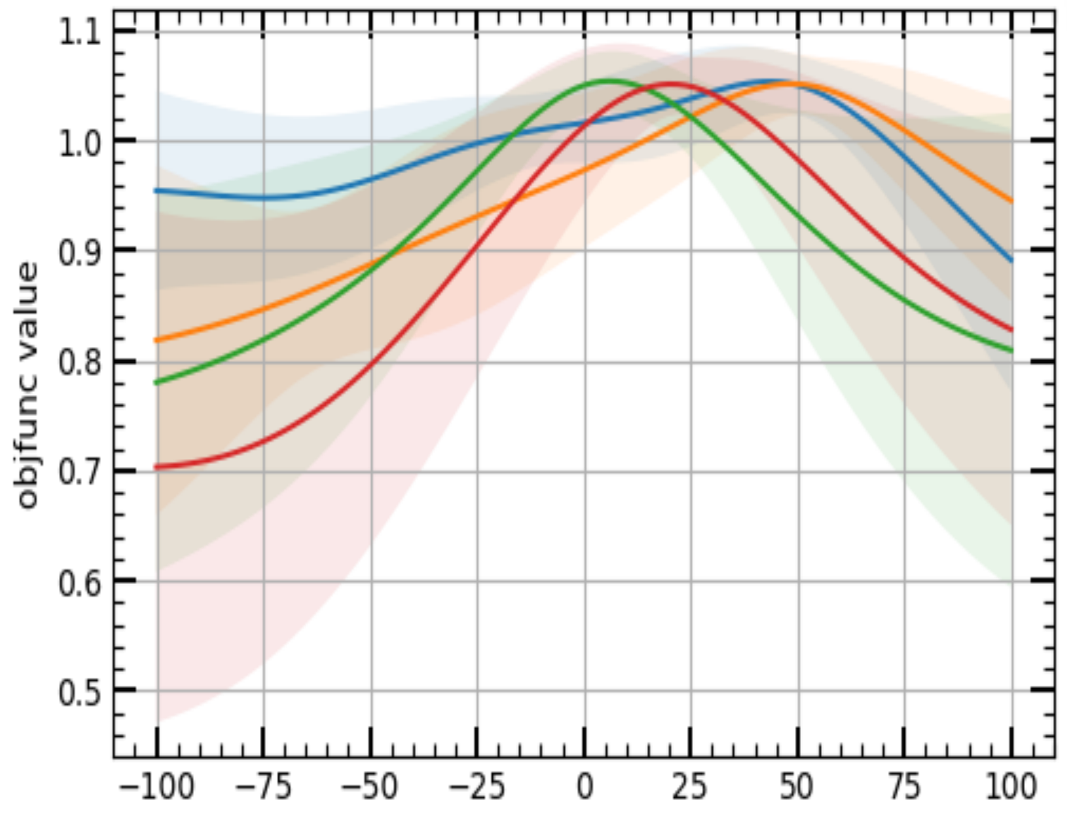
で最適化



評価関数 vs Epoch



電流値 (相対値) vs Epoch

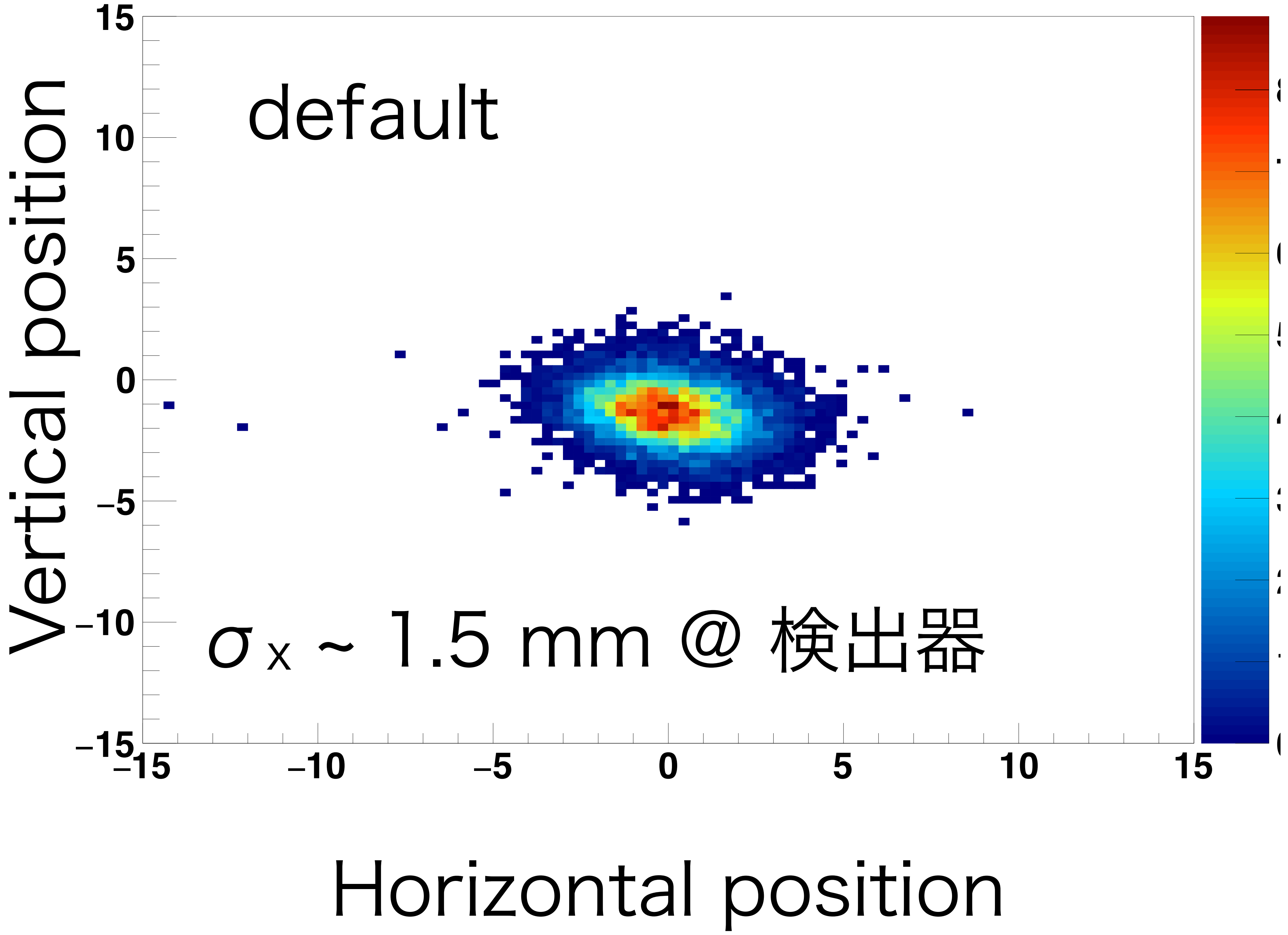


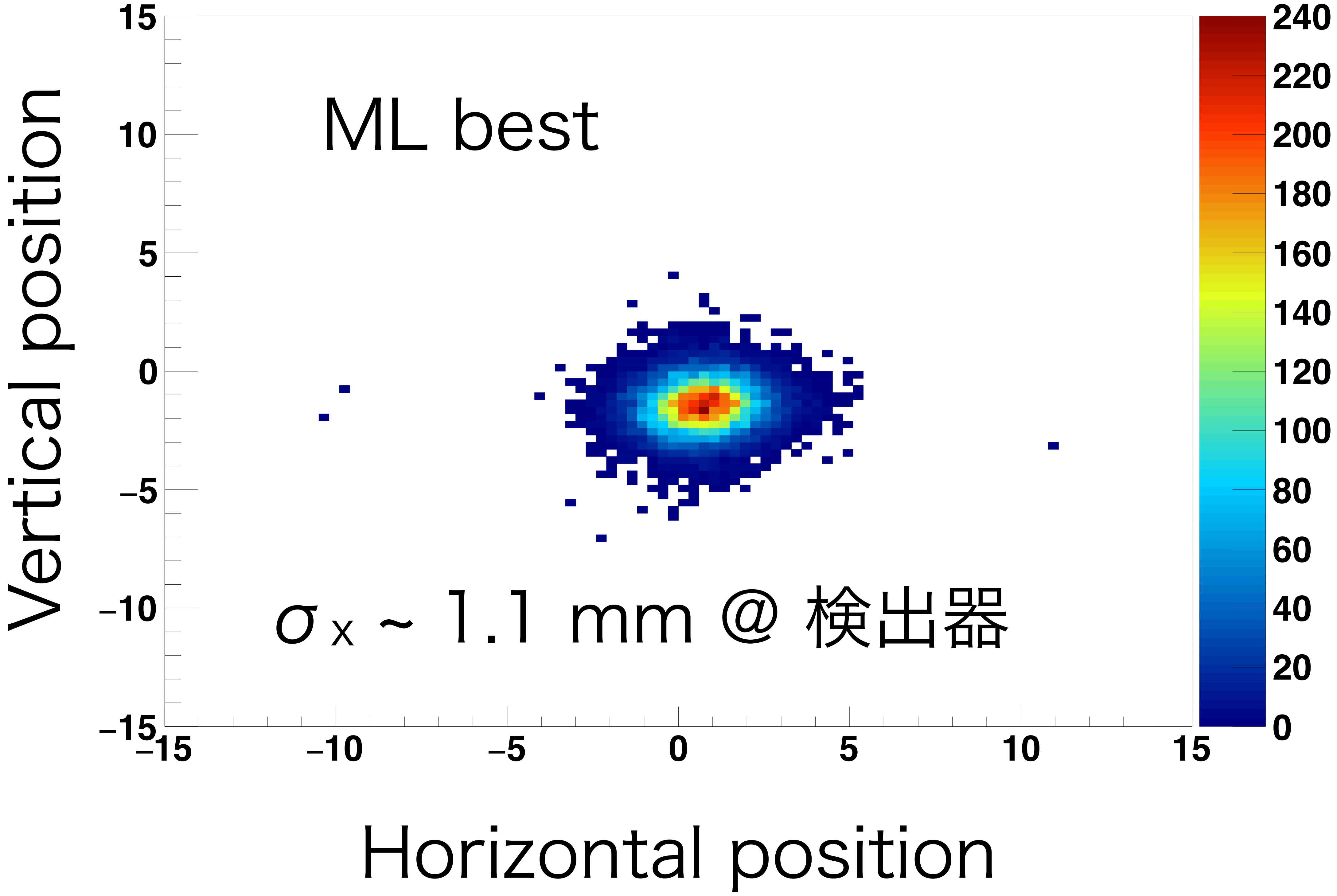
評価関数 vs 電流値

※1 epoch の時間は、
・電流値のセット
・検出器の測定
などで律速

# “高強度”一次ビームで荷電変換粒子の測定による光学系最適化

初期光学系 (人間の手で最適化) と ML で得られた光学系を比較
→ **通過効率はほぼ変わらず。スポットサイズは ~30% 程度減少に成功。**



# RIBF におけるベイズ最適化を使った光学系調整の現状

**達成**

・ベイズ最適化 (ガウス過程回帰) プログラムが期待通りに動くことを確認
・~ 26 enA までの"高強度" 重イオンビームに対し、スポットサイズ / 通過効率の同時最適化に成功。

**課題**

・安全に自動調整を行うシステム (Safe optimization using LineBO)
・より効率的な最適化手法へ改善
(現状: 4 パラメータで ~ 30 分 / 50 epoch)

# RIBF におけるベイズ最適化を使った光学系調整の現状

**<u>達成</u>**

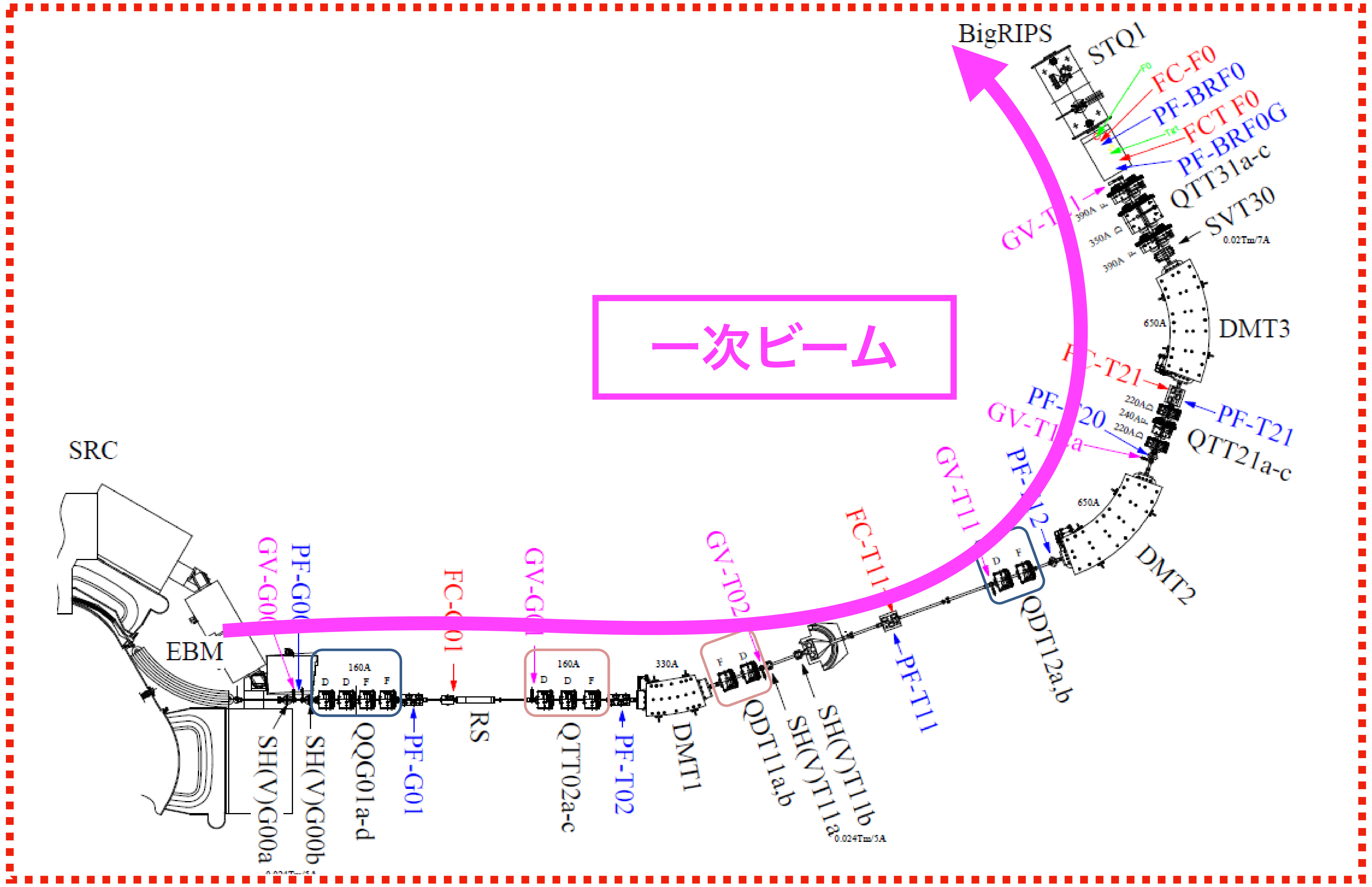
・ベイズ最適化 (ガウス過程回帰) プログラムが期待通りに動くことを確認
・~ 26 enA までの”高強度” 重イオンビームに対し、スポットサイズ / 通過効率
　の同時最適化に成功。

**<u>課題</u>**

・安全に自動調整を行うシステム (Safe optimization using LineBO)
<u>・より効率的な最適化手法へ改善</u>
<u>　(現状: 4 パラメータで ~ 30 分 / 50 epoch)</u>
　VAE を使った変数の自由度の削減 (次の森田氏の講演)
　シミュレーションを基にした事前学習 ( physics informed Gaussian Process など)
　**objective / acquisition function の改善**

**ビーム輸送:**
**ほとんどシミュレーション通り**
**になるので相性が良い**

# なぜシミュレーションできるビーム輸送に機械学習手法が必要か？




① ビームの分布の”裾” の問題
”ほとんど” はガウス分布を仮定して問題ない。
→ シミュレーション通り
しかし、ビームロスは残りのビームの裾がきめる。
このビームロスの量が施設が耐えられるビーム強度を決める

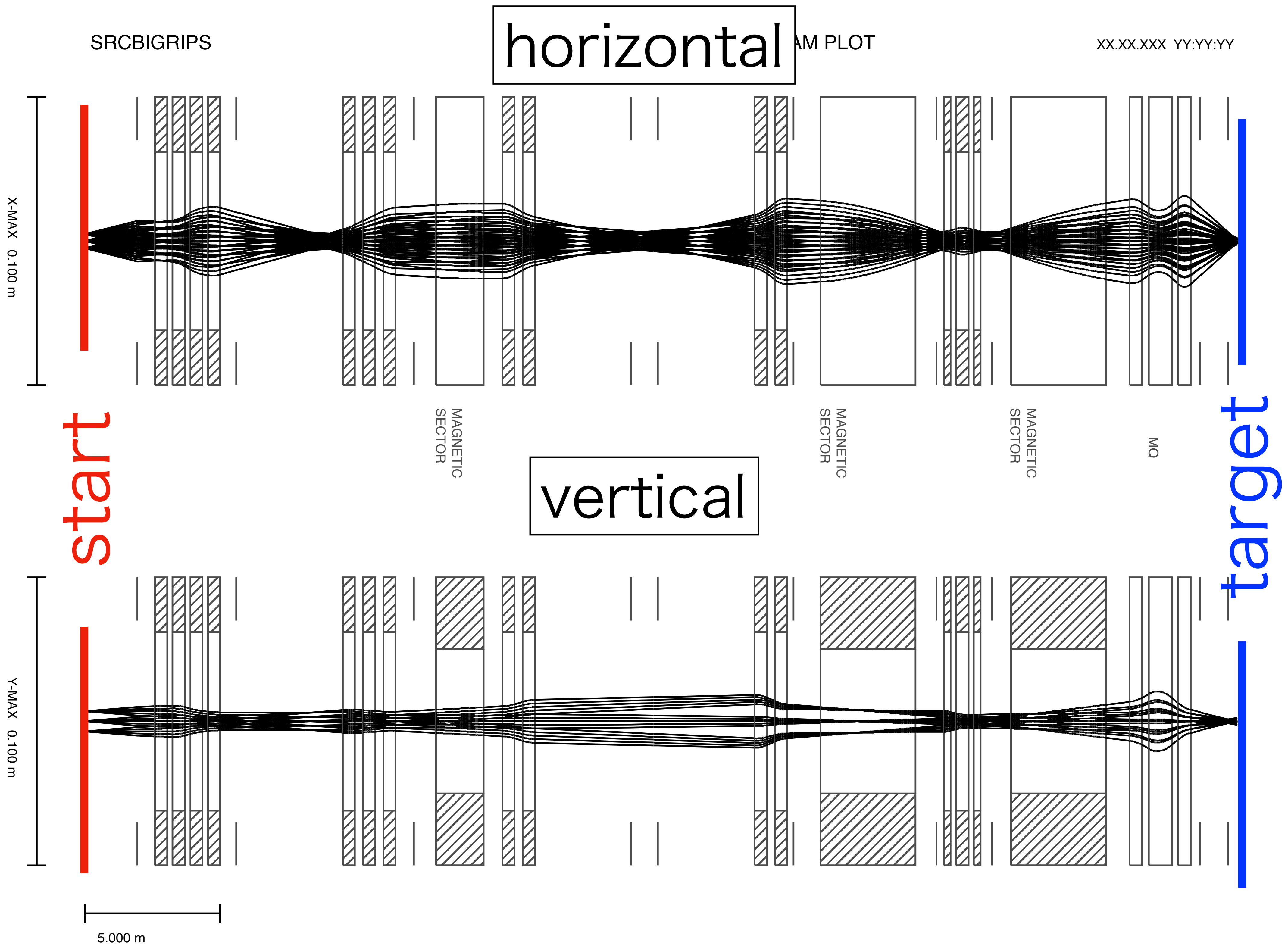
② ビームの初期状態の問題
一次ビームの位相楕円を正確に測定するのは難しい。
特に大強度化によって変化したビームの位相楕円を測定して計算するのは困難。

③ 磁石の配置 / 磁場の誤差
磁石の位置や磁場などに誤差があった場合にも、ベイズ最適化であれば対応できる。

# ベイズ最適化のためのシミュレーションの概要

光学系計算 (gicosy)

http://web-docs.gsi.de/~weick/gicosy/

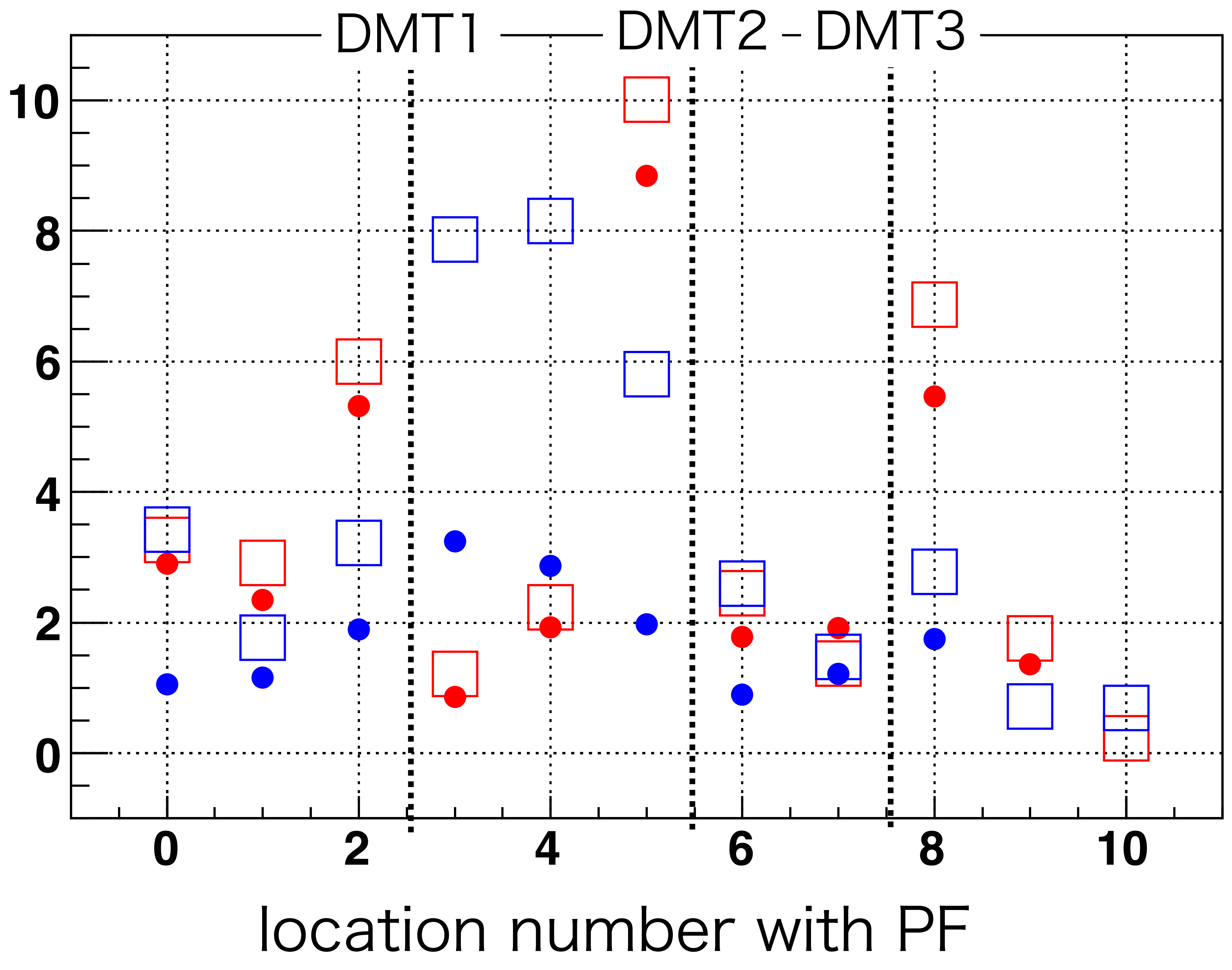



磁石の geometry と磁場

→ 3 次までの転送行列

光学計算と wire scanner の比較

特に horizontal はよく一致している。

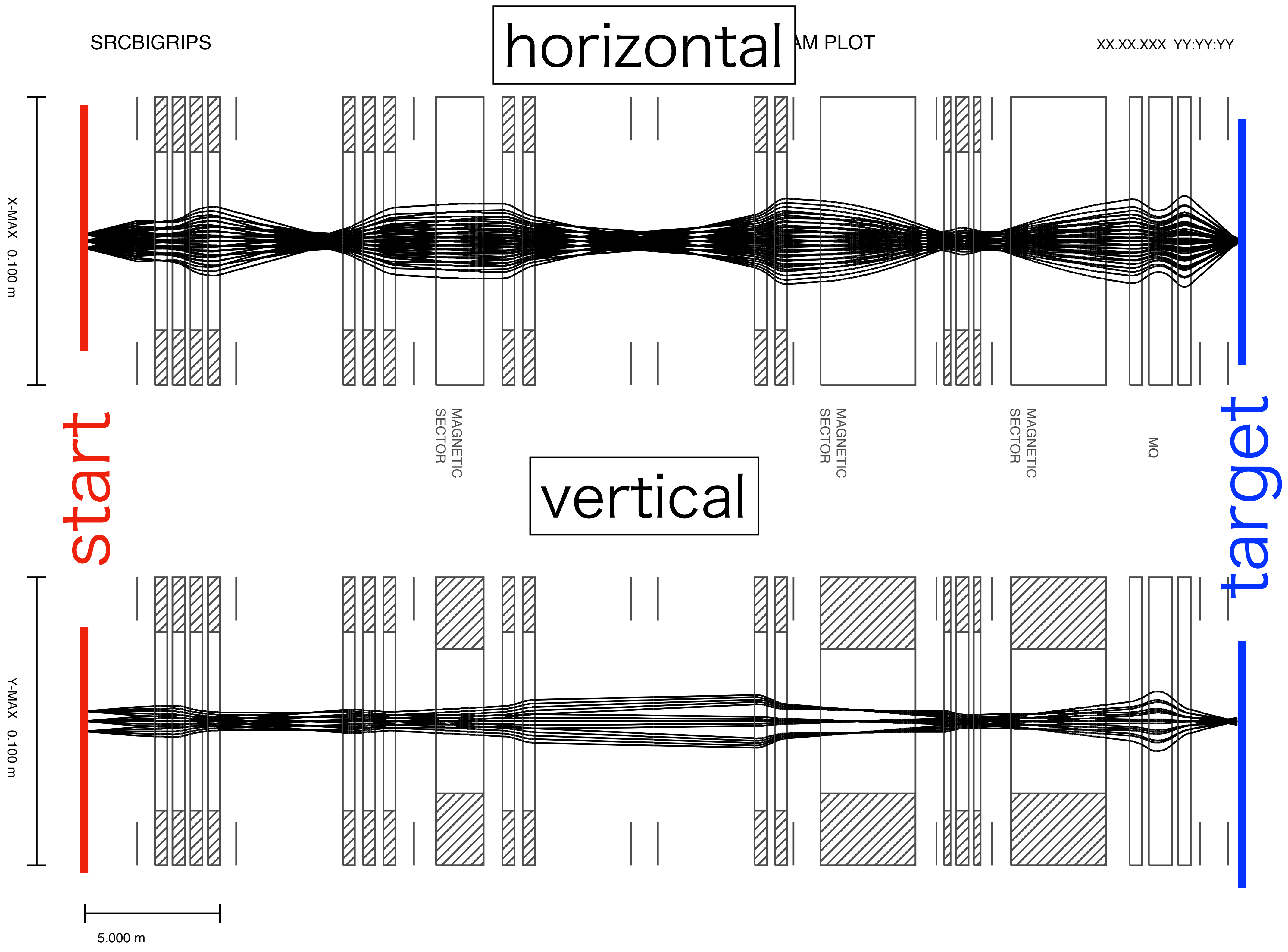



horizontal / vertical

● wire scanner / □ gicosy calc.

# ベイズ最適化のためのシミュレーションの概要

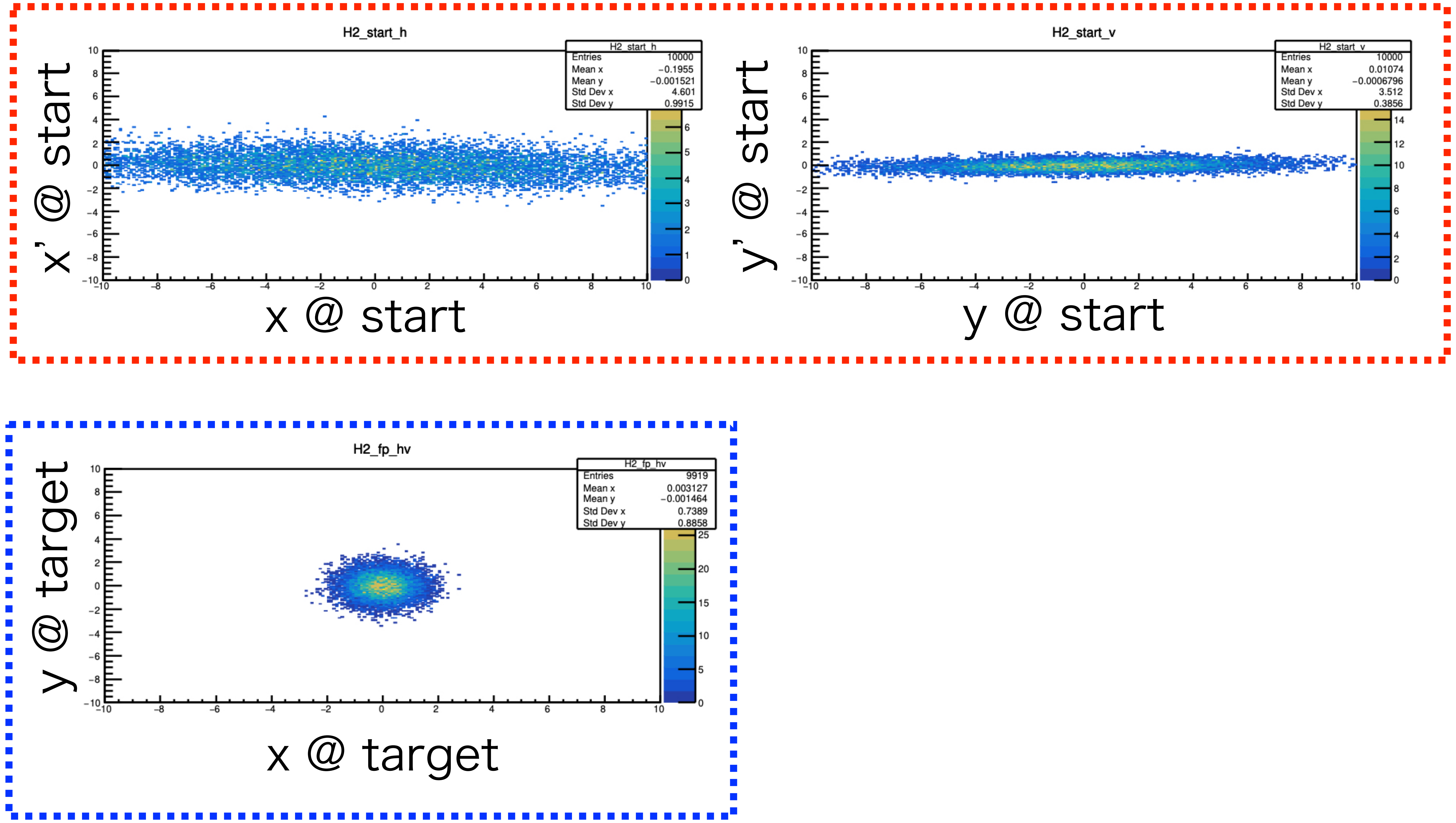
光学系計算 (gicosy)
http://web-docs.gsi.de/~weick/gicosy/




磁石の geometry と磁場

→ 3 次までの転送行列

モンテカルロシミュレーション(MOCADI※)




start の位相楕円 + 転送行列

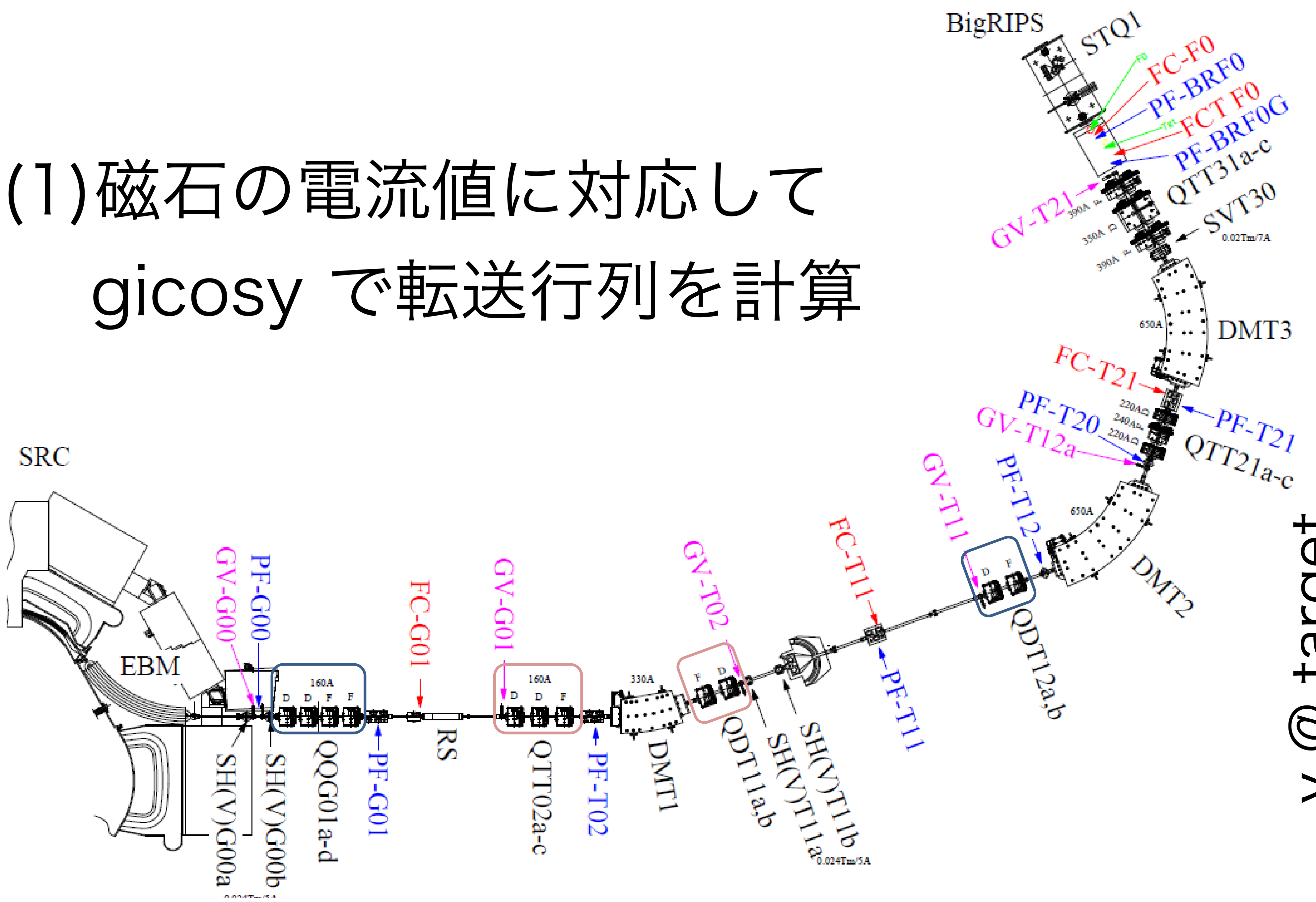
→ 各光学要素での粒子の位置を計算 / Apperture を判定。

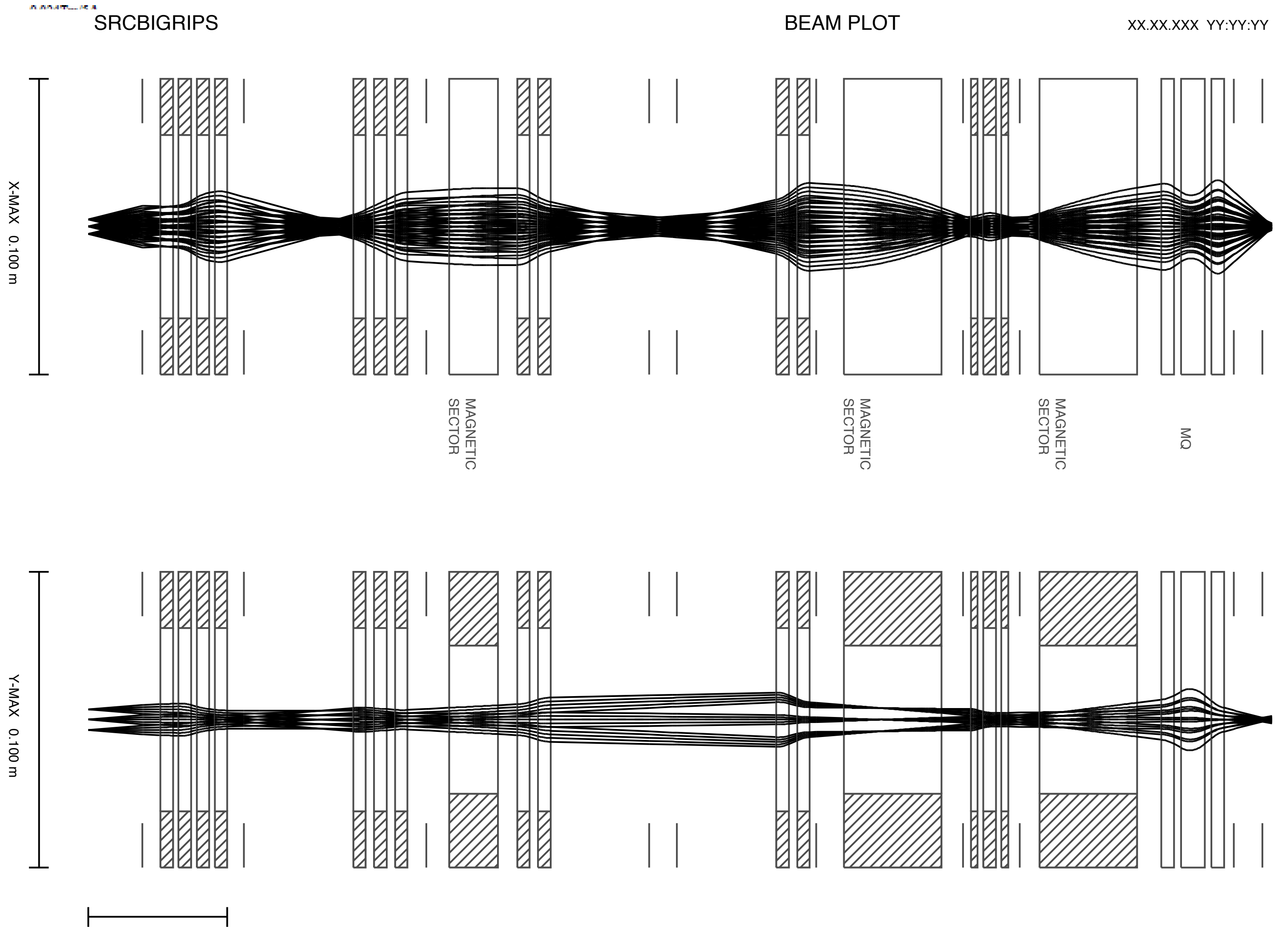
target で"生きている" 粒子の数と分布の幅を測定。

※[1] N.Iwasa et al., Nucl. Instr. Meth. B126, 284 (1997); Th.Schwab, GSI Report No. 91-10 (1991).
[2] C.Scheidenberge et al., Phys. Rev. Lett. 73, 50 (1994); C.Scheidenberge et al., Phys. Rev. Lett. 77, 3987 (1996).

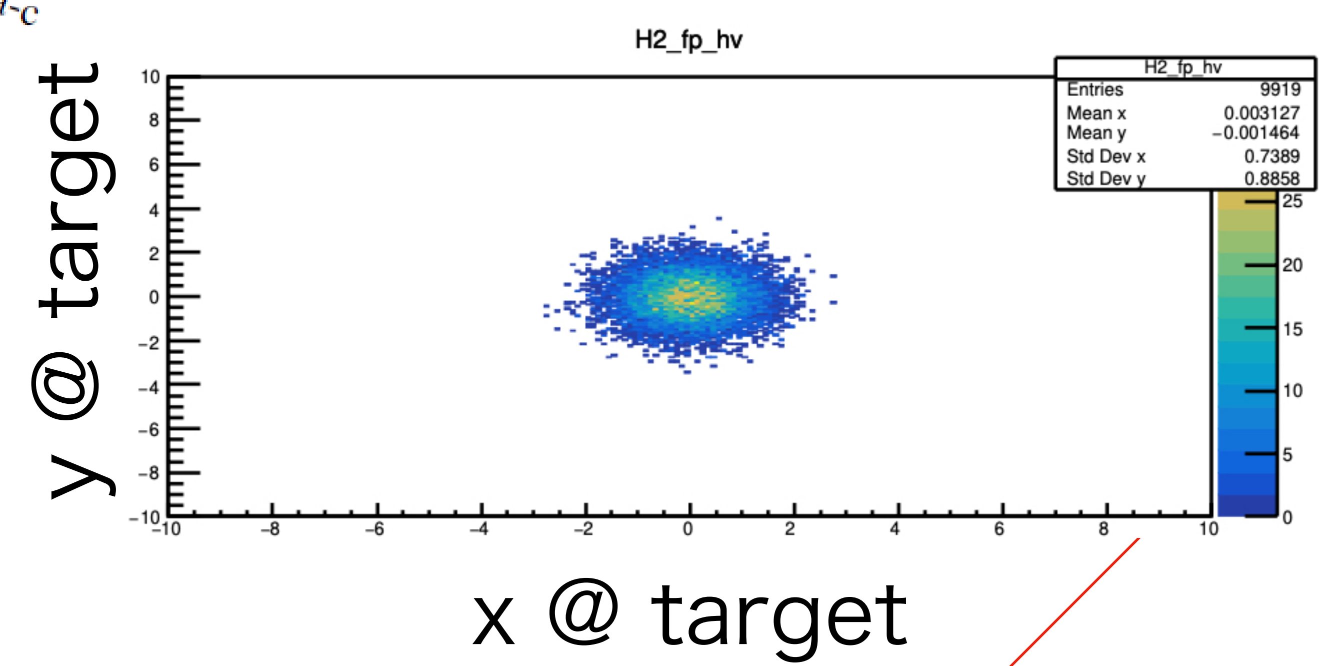
# ベイズ最適化のためのシミュレーションの概要

(1)磁石の電流値に対応して
gicosy で転送行列を計算



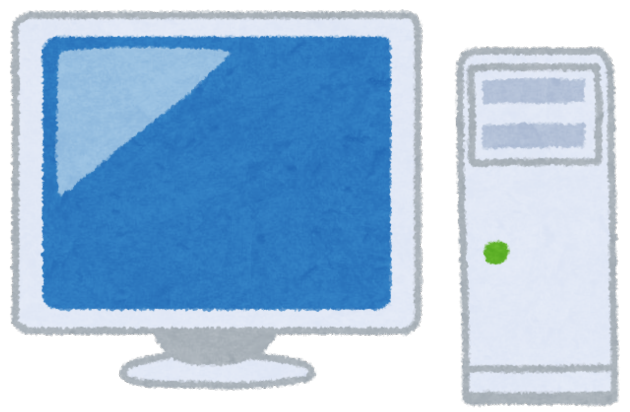




(2) MOCADI で target の像を
シミュレート
通過率と像幅を測定




(3) objective function を計算
acquisition function に
従って、新しい候補を選定

**ガウス過程回帰**
**(逐次学習)**



(1)~ (3) を繰り返して
最適な電流値を探す

# ベイズ最適化のためのシミュレーションの概要

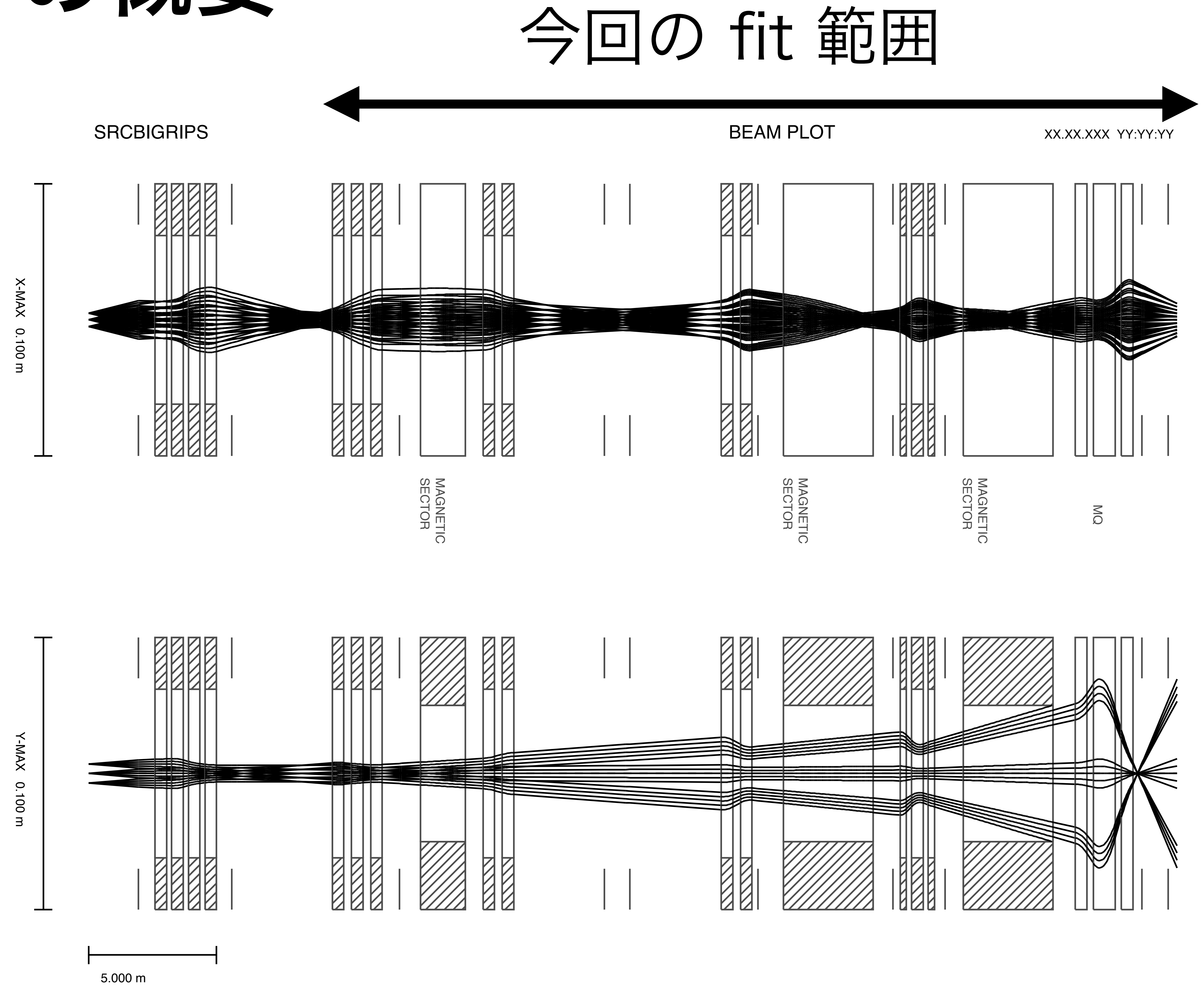
## 今回のシミュレーションの条件

Variable　　: Quadrupole magnet × 13

電流値の範囲: 0 ~ 100 A (最後のTQのみ ~ 250 A)

初期電流値　: 過去の試験データから少しずらしたもの

(通過率 ~ 73%, (σx, σy) = (4.2, 5.8))




## 変化させるもの

・acquisition function

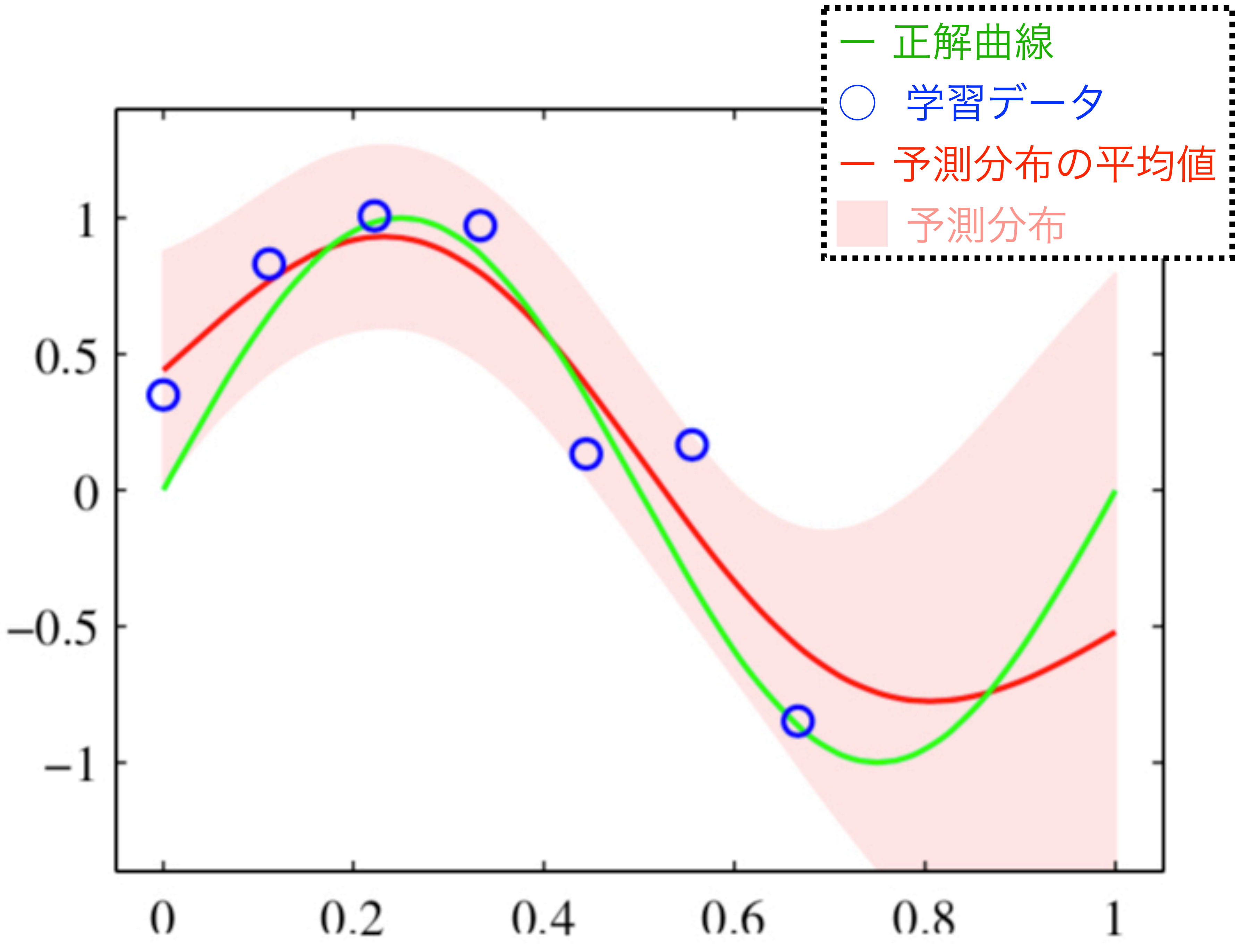
・objective function

・初期位相楕円 (結果の robustness の確認のため)

最大 ~ 200 epoch (⇔ 現状だと 2 時間) 程度で、効率的に良い解を見つける条件を確認する。
または、その手法を探す。

# ベイズ最適化のハイパーパラメータ: ① acquisition function

acquisition function / 獲得関数 :
これまでのデータを基に作った
予測分布から、次の探索パラメータ
をどこにするか決める指標
**活用と探索のバランスを決める。**




## Probability of Improvement (PI)

**活用 >> 探索**

$$\vec{x}_{next} = \mathrm{argmax}_{\vec{x}} P(f(\vec{x}) > y_{max})$$

これまでの最大値 $y_{max}$ よりも得られる値が
大きくなる確率を最大にする x

## Eexpected Improvement (EI)

これまで採用

**活用 > 探索 (?)**

$$\vec{x}_{next} = \mathrm{argmax}_{\vec{x}} \mathbb{E}[\max(f(\vec{x}) - y_{max}, 0)]$$

これまでの最大値 $y_{max}$ よりも得られる値が
大きくなる期待値 (悪化は無視)を最大にする x

## Lower Confidence Bound (LCB)

**活用 ~ 探索 (?)**
***β* 次第…**

$$\vec{x}_{next} = \mathrm{argmax}_{\vec{x}} (\mu_{\vec{x}} + \beta^{1/2}\sigma(\vec{x}))$$

$\beta^{1/2}\sigma$ まで考慮したときに最も良い値を得る
可能性のある x

# ベイズ最適化のハイパーパラメータ: ② objective function

最大化 (最小化) したいもの: 粒子数 N, 像幅 $\sigma_x$, $\sigma_y$ @ target

従来の獲得関数:

$$f(N, \sigma_x, \sigma_y) = -c_0 \frac{N}{N_0} + \left(\frac{\sigma_x}{\sigma_x^\star}\right)^2 + \left(\frac{\sigma_y}{\sigma_y^\star}\right)^2$$

$c_0$ : 任意の係数 (~1)
$N_0$ : 初期値での標的上での粒子数
$\sigma^\star_{x,y}$ : 像幅の目安 ~ 1 mm

問題点: 符号が反転する場合がある。各項が同じ程度の寄与になってない。

# ベイズ最適化のハイパーパラメータ: ② objective function

最大化 (最小化) したいもの: 粒子数 N, 像幅 $\sigma_x$, $\sigma_y$ @ target

**常に符号を変えず、かつ各要素が同程度の寄与になるように考案**

sum 型:

$$f_{sum}(N, \sigma_x, \sigma_y) = -\frac{N}{N_0} - \frac{\sigma_x^\star}{\sigma_x^\star + \sigma_x} - \frac{\sigma_y^\star}{\sigma_y^\star + \sigma_y}$$

multilied 型:

$$f_{multi}(N, \sigma_x, \sigma_y) = -\frac{N}{N_0} \cdot \frac{\sigma_x^\star}{\sigma_x^\star + \sigma_x} \cdot \frac{\sigma_y^\star}{\sigma_y^\star + \sigma_y}$$

$N_0$ : 初期値での標的上での粒子数

$\sigma^\star_{x,y}$ : 像幅の目安 ~ 1 mm

各 transmission, σ での $f_{sum}$ / $f_{multi}$ の絶対値

| σ\Transmission | 0.75 | 0.85 | 0.95 |
|---|---|---|---|
| 1.5 | 1.8 / 0.16 | 2.0 / 0.19 | 2.1 / 0.21 |
| 1.0 | 2.0 / 0.26 | 2.2 / 0.29 | 2.3 / 0.33 |
| 0.5 | 2.4 / 0.46 | 2.5 / 0.52 | 2.6 / 0.58 |

# ベイズ最適化のハイパーパラメータ: ② objective function

最大化 (最小化) したいもの: 粒子数 N, 像幅 $\sigma_x$, $\sigma_y$ @ target

**常に符号を変えず、かつ各要素が同程度の寄与になるように考案**

sum 型:

$$f_{sum}(N, \sigma_x, \sigma_y) = -\frac{N}{N_0} - \frac{\sigma_x^\star}{\sigma_x^\star + \sigma_x} - \frac{\sigma_y^\star}{\sigma_y^\star + \sigma_y}$$

multilied 型:

$$f_{multi}(N, \sigma_x, \sigma_y) = -\frac{N}{N_0} \cdot \frac{\sigma_x^\star}{\sigma_x^\star + \sigma_x} \cdot \frac{\sigma_y^\star}{\sigma_y^\star + \sigma_y}$$

$N_0$ : 初期値での標的上での粒子数
$\sigma^\star_{x,y}$ : 像幅の目安 ~ 1 mm

各 transmission, σ での $f_{sum}$ / $f_{multi}$ の絶対値

**transmission 85% / $\sigma_{x,y}$ ~ 1mm を基準とした時**

| σ\Transmission | 0.75 | 0.85 | 0.95 |
|---|---|---|---|
| 1.5 | 0.8 / 0.6 | 0.9 / 0.6 | 1.0 / 0.7 |
| 1.0 | 0.9 / 0.9 | 1.0 / 1.0 | 1.1 / 1.1 |
| 0.5 | 1.1 / 1.6 | 1.2 / 1.8 | 1.2 / 2.0 |

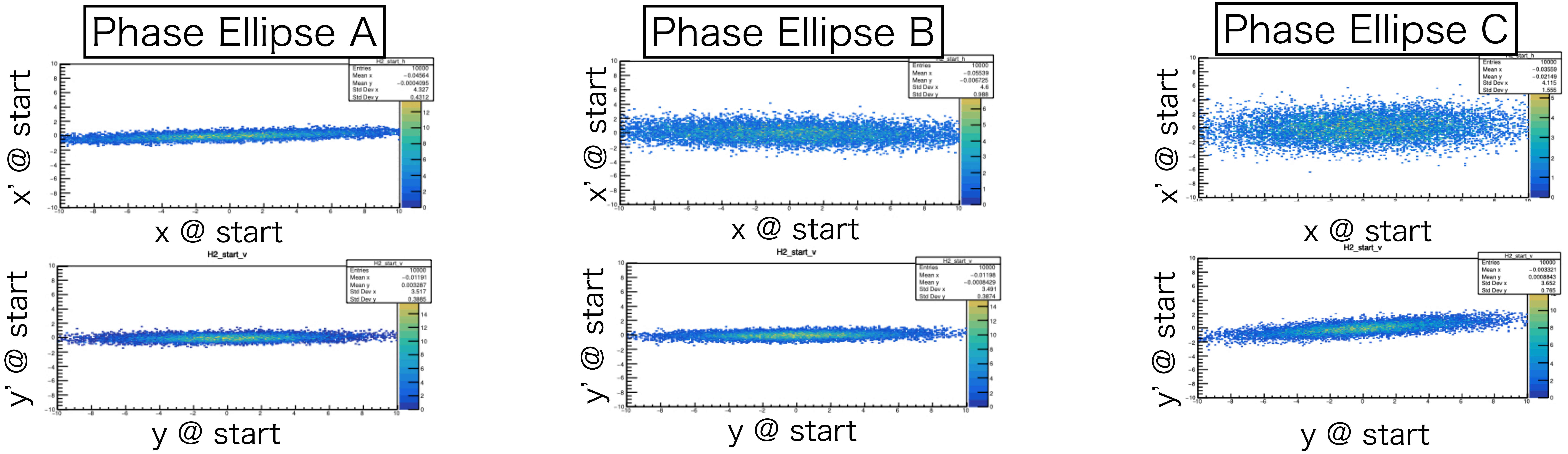
**$f_{multi}$ では**

- **より σ の影響が大きくなる。**
- **transmission / σ を同時に良くすると特に値が伸びる**

これが一番いい形？他に経験や提案があればぜひ！

# ベイズ最適化のハイパーパラメータ(?): ③ phase ellipse

結果の robustness を確認するため、複数の phase ellipse についてシミュレーションを行う。




EDC を抜けるために角度方向が抑えられているはずなので、
ある程度扁平な楕円を仮定した上で ε、α、δ との相関などを変更。

# 各ハイパーパラメータでのシミュレーション結果の比較

**比較するもの**

・300 epoch 以内で最も”良い” 光学系

・Transmission 90% 以上 / $\sigma_{x,y} < 1$ mm 以下の解を最初に見つけた epoch

・最適化の過程

・得られた解の variety

# 各ハイパーパラメータでのシミュレーション結果の比較

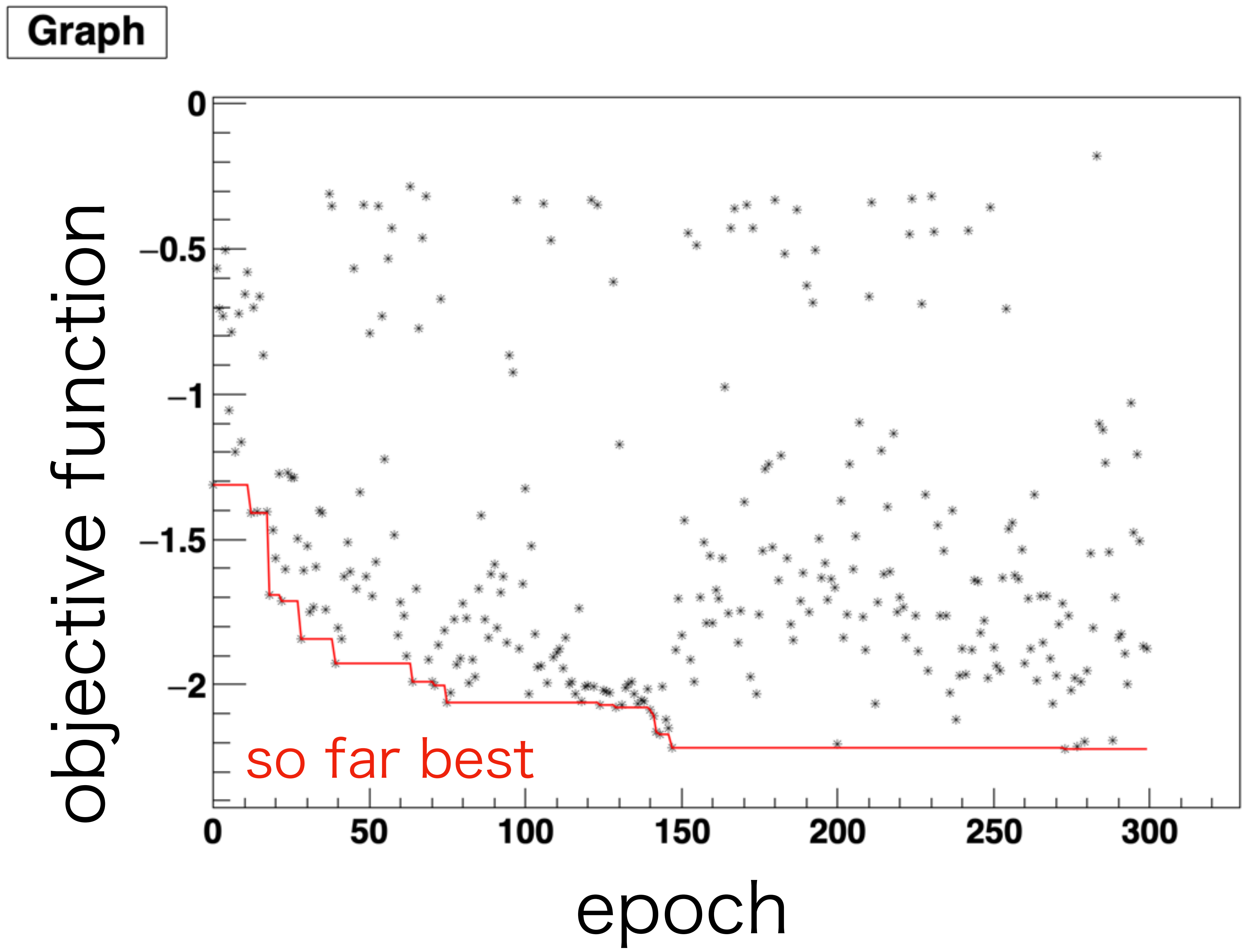
**比較するもの**

・300 epoch 以内で最も"良い" 光学系

・Transmission 90% 以上 / $\sigma_{x,y}$ < 1 mm 以下の解を最初に見つけた epoch

**・最適化の過程**

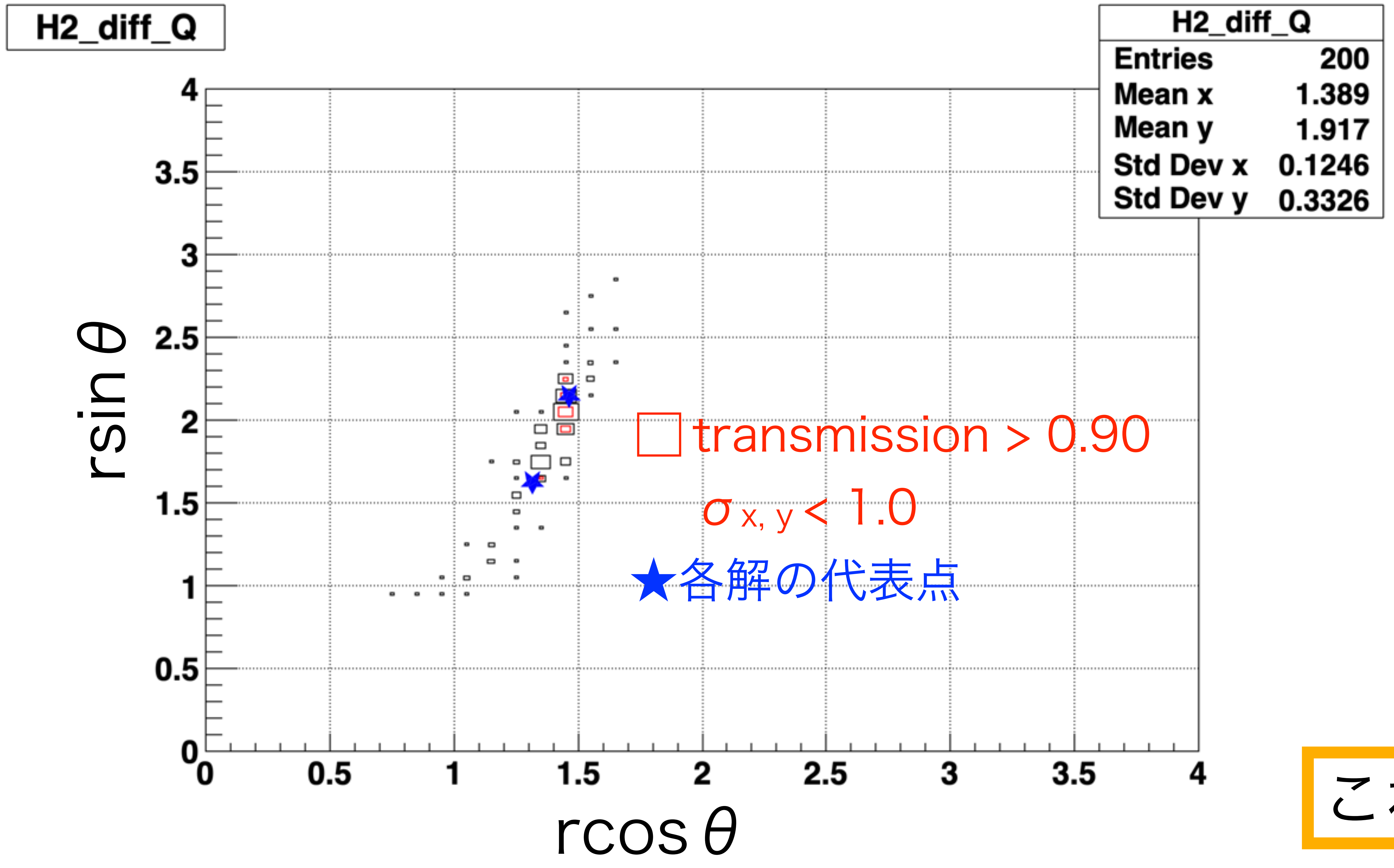
・得られた解の variety



# 各ハイパーパラメータでのシミュレーション結果の比較

**比較するもの**

・300 epoch 以内で最も”良い” 光学系
・Transmission 90% 以上 / $\sigma_{x,y}$ < 1 mm 以下の解を最初に見つけた epoch
・最適化の過程
**・得られた解の variety**




元の電流値に対する”内積” とノルムで定義

$$\vec{Q}_{norm} \equiv \{Q_i/Q_i^0\}$$

$$r \equiv |\vec{Q}_{norm}|/|\vec{Q^0}_{norm}|$$

$$\theta \equiv \arccos \frac{\vec{Q}_{norm} \cdot \vec{Q^0}_{norm}}{|\vec{Q}_{norm}||\vec{Q^0}_{norm}|}$$

Qi ：磁石の電流値
Qi⁰：電流値の初期値

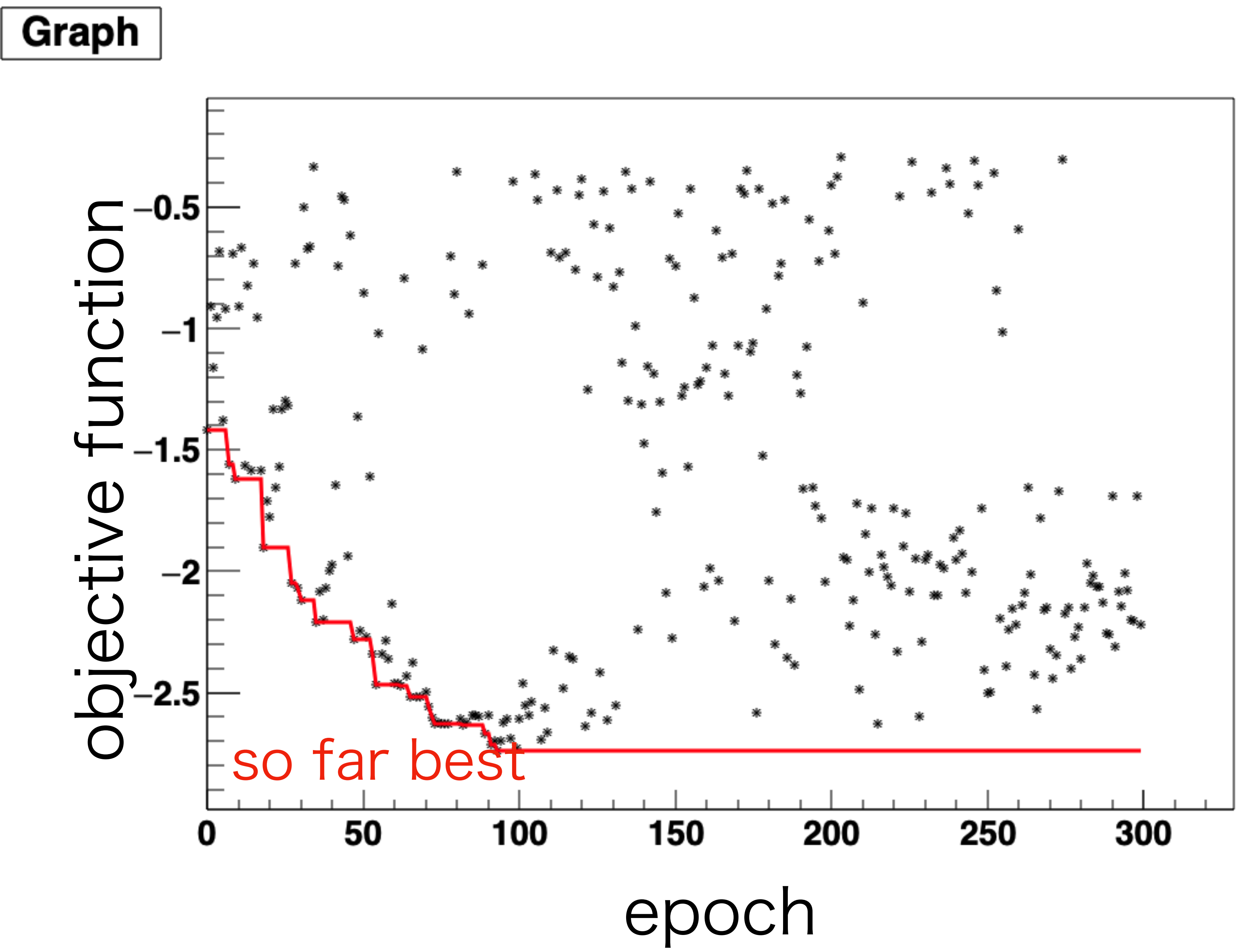
(rcos θ , rsin θ ) で距離が 0.5 以上
→異なる解と判定

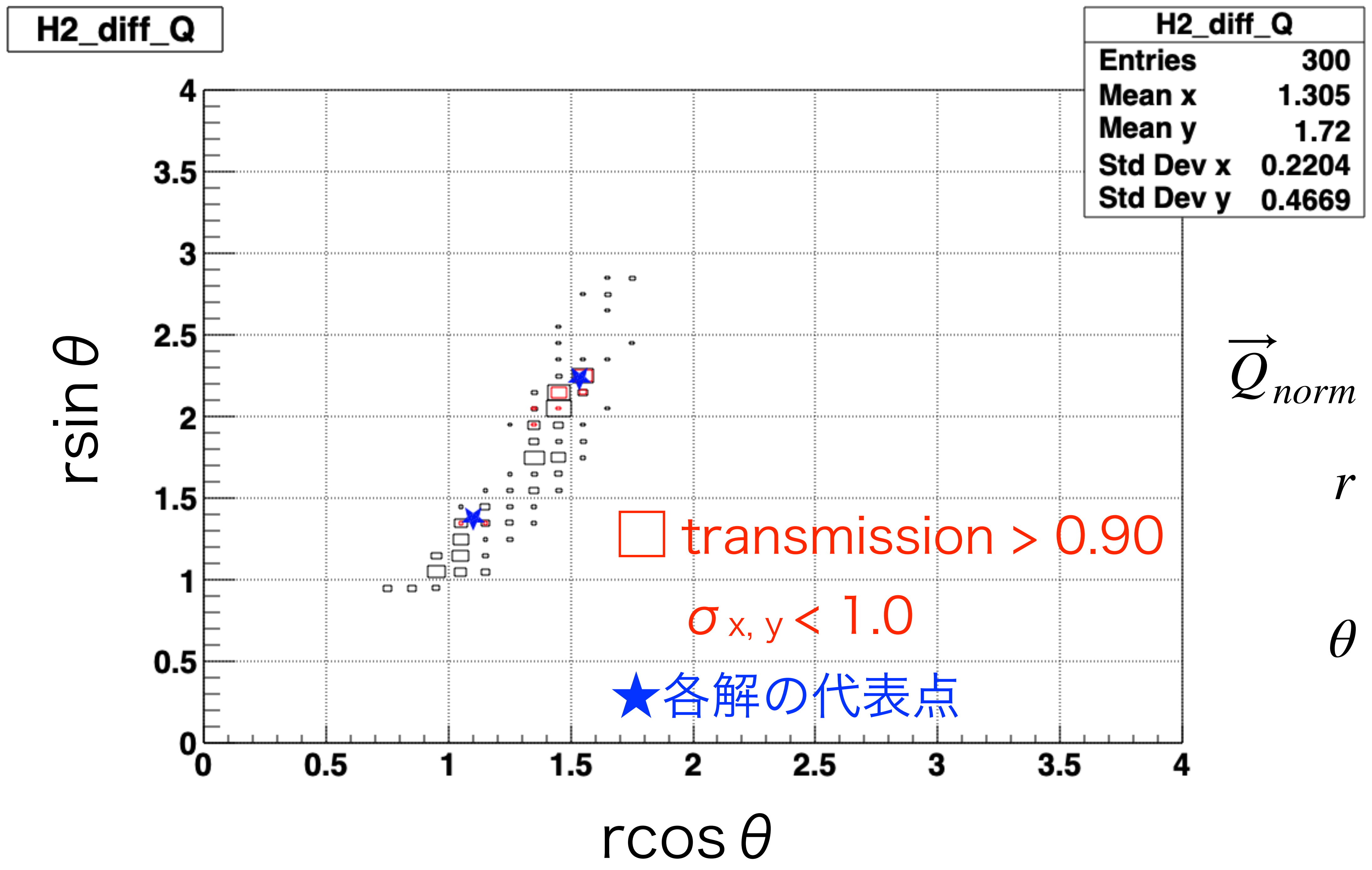
これが一番いい形？他に経験や提案があればぜひ！

# 各ハイパーパラメータでのシミュレーション結果の比較

異なる Acquisition function の比較 (objective function: $f_{sum}$/ phase ellipse A)

Probability of Improvement (PI)




$$\vec{Q}_{norm} \equiv \{Q_i/Q_i^0\}$$

$$r \equiv |\vec{Q}_{norm}|/|\vec{Q^0}_{norm}|$$

$$\theta \equiv \arccos \frac{\vec{Q}_{norm} \cdot \vec{Q^0}_{norm}}{|\vec{Q}_{norm}||\vec{Q^0}_{norm}|}$$

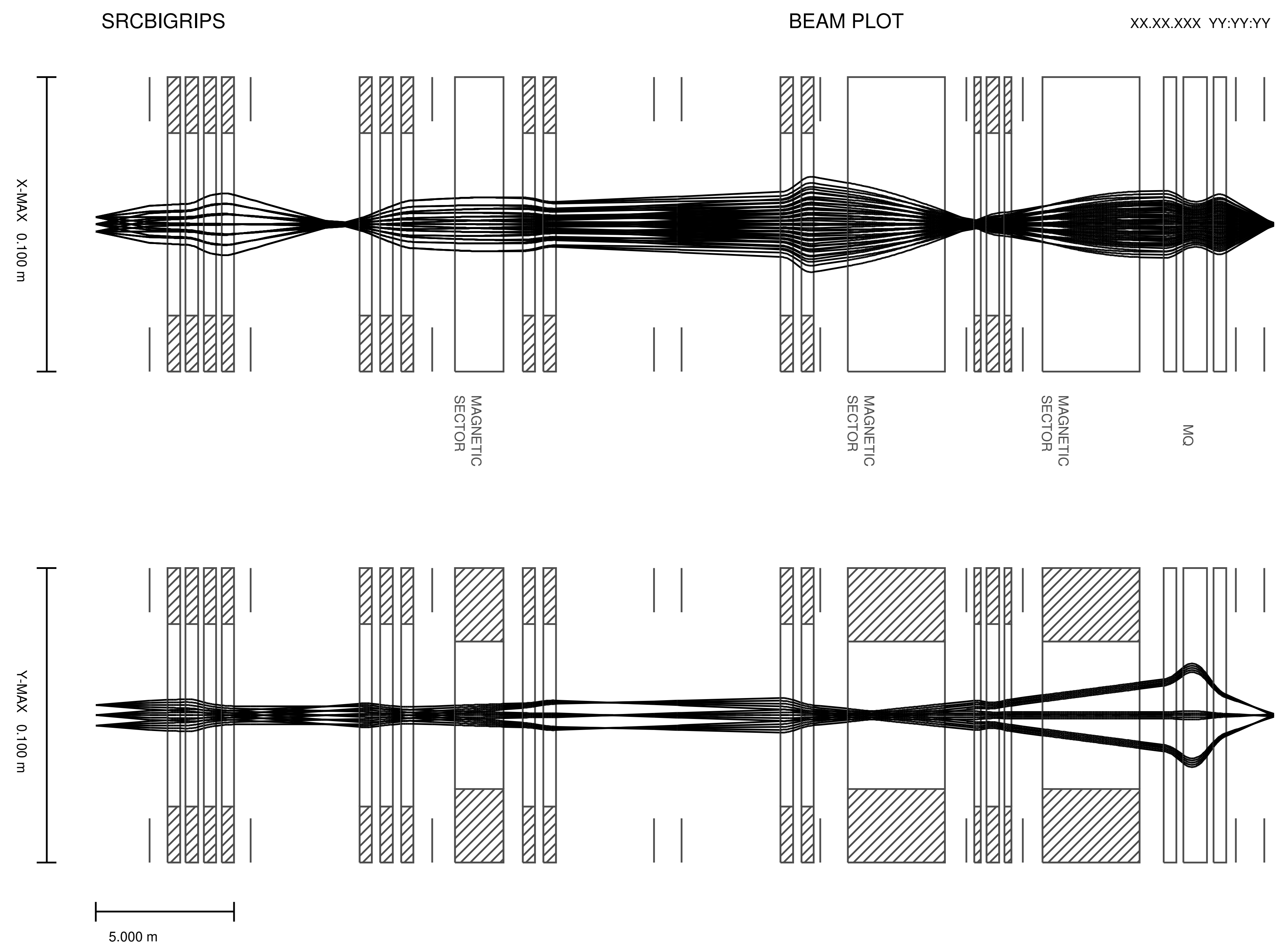
| | epoch | obj. | transmission | $\sigma_x$ | $\sigma_y$ |
|---|---|---|---|---|---|
| **最初に閾値を超えたもの** | 73 | -2.63 | 0.91 | 0.34 | 0.53 |
| **best** | 93 | -2.74 | 0.98 | 0.45 | 0.36 |

# 各ハイパーパラメータでのシミュレーション結果の比較

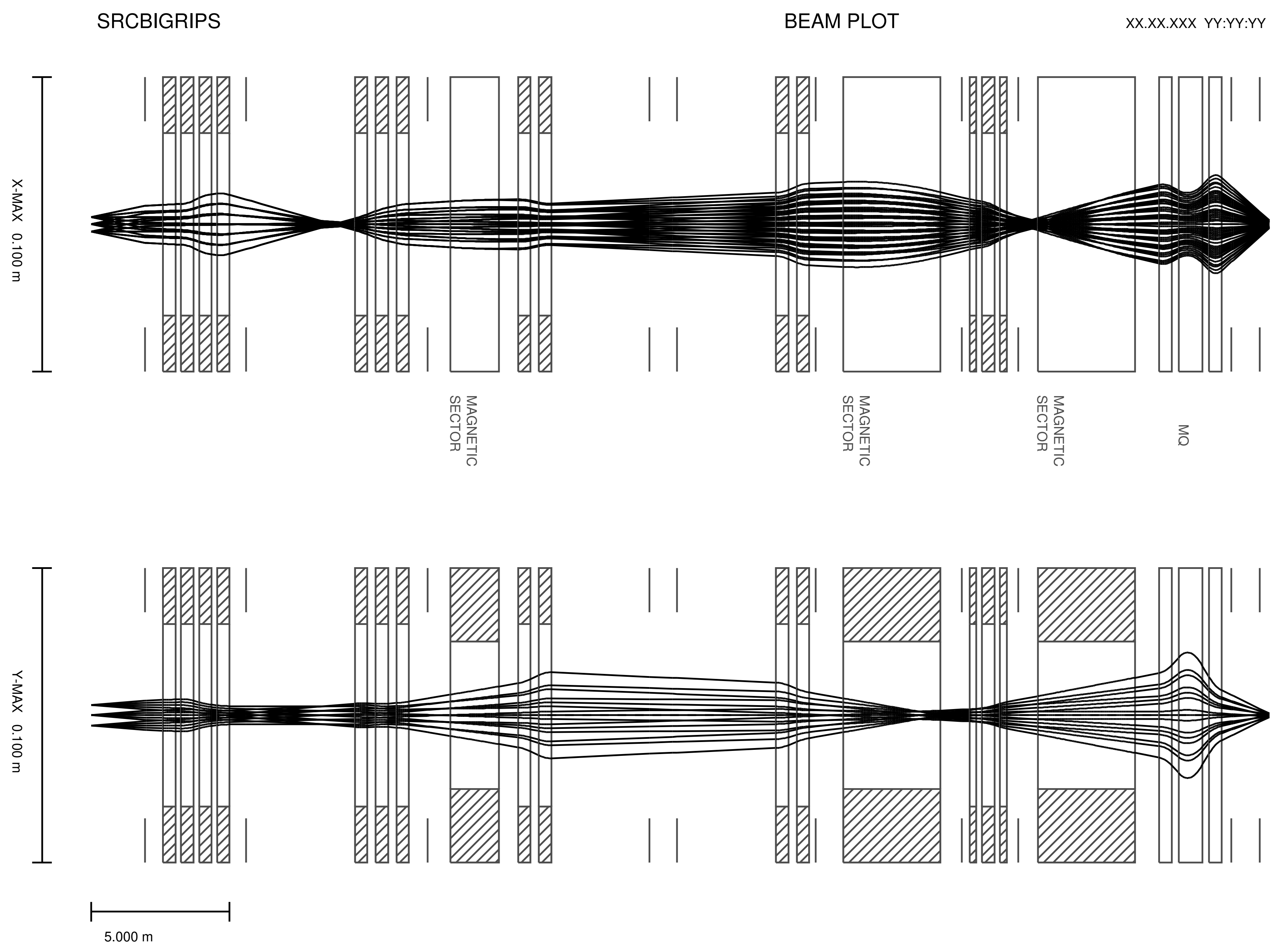
異なる Acquisition function の比較 (objective function: $f_{sum}$/ phase ellipse A)

Probability of Improvement (PI)

代表点 1 (93 epoch)




代表点 2(266 epoch)

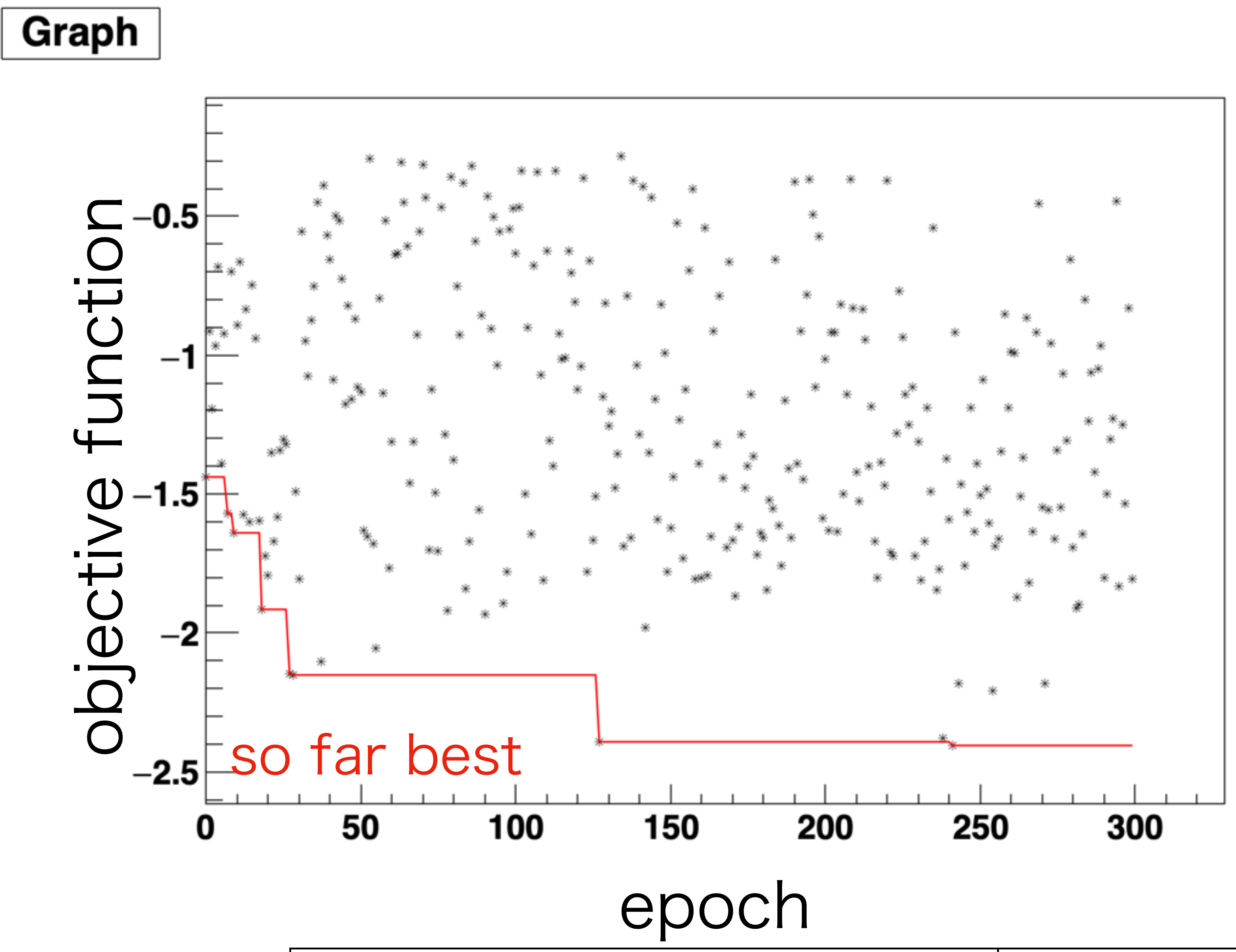


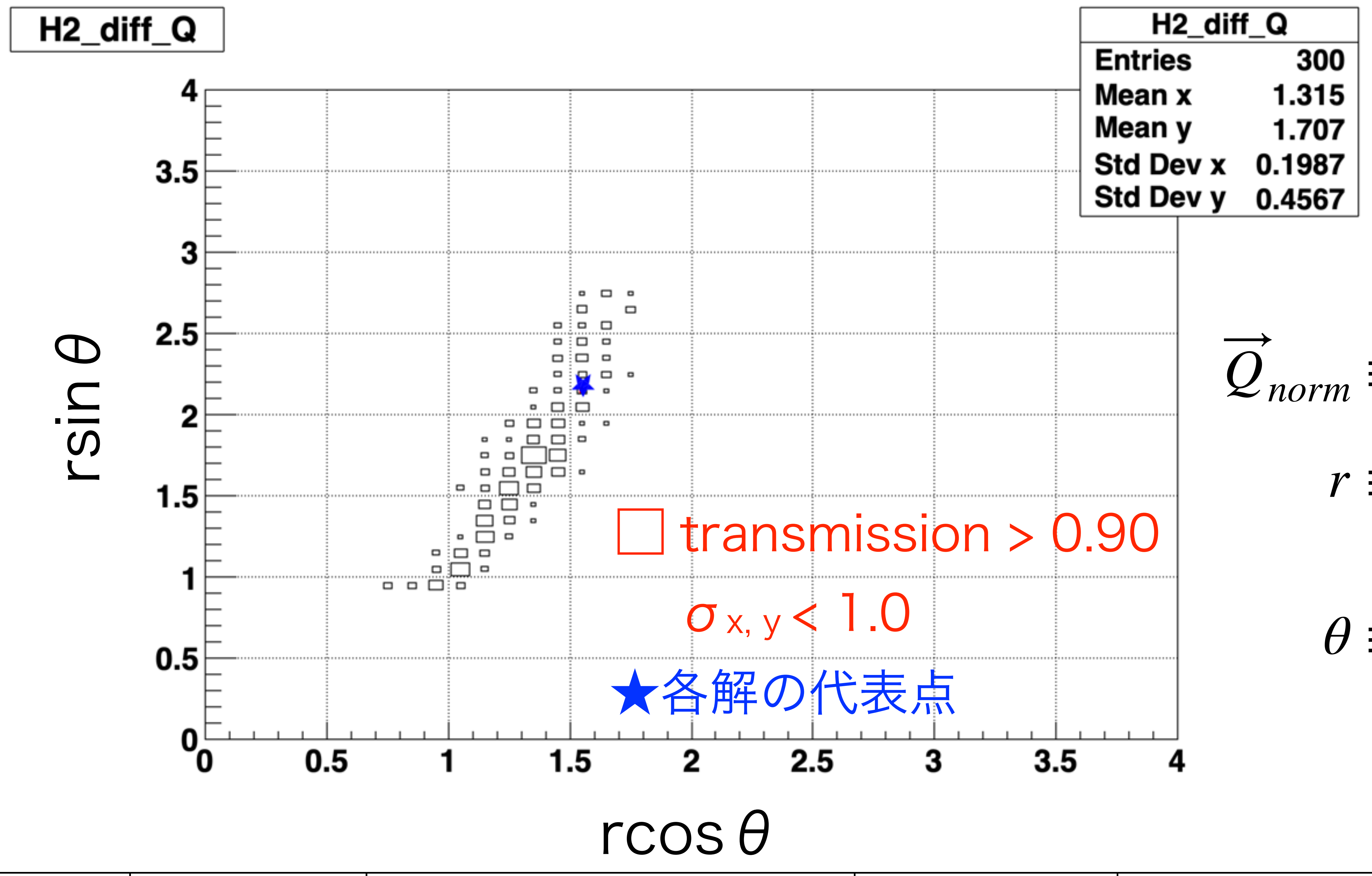

~ 最大 40 A 程度電流が変化している。

# 各ハイパーパラメータでのシミュレーション結果の比較

異なる Acquisition function の比較 (objective function: $f_{sum}$/ phase ellipse A)

Expected Improvement (EI)




$$\vec{Q}_{norm} \equiv \{Q_i/Q_i^0\}$$

$$r \equiv |\vec{Q}_{norm}|/|\vec{Q^0}_{norm}|$$

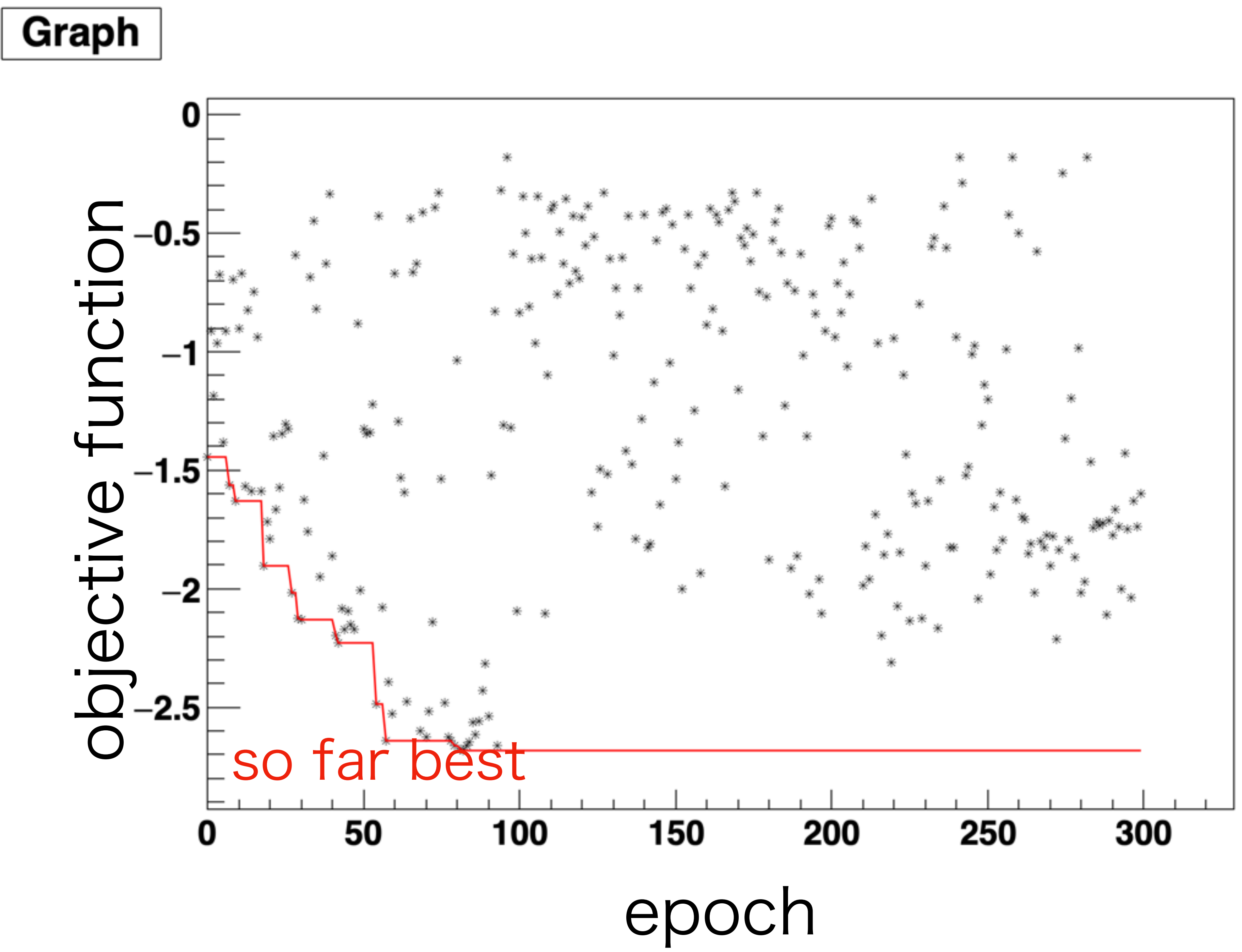
$$\theta \equiv \arccos \frac{\vec{Q}_{norm} \cdot \vec{Q^0}_{norm}}{|\vec{Q}_{norm}||\vec{Q^0}_{norm}|}$$

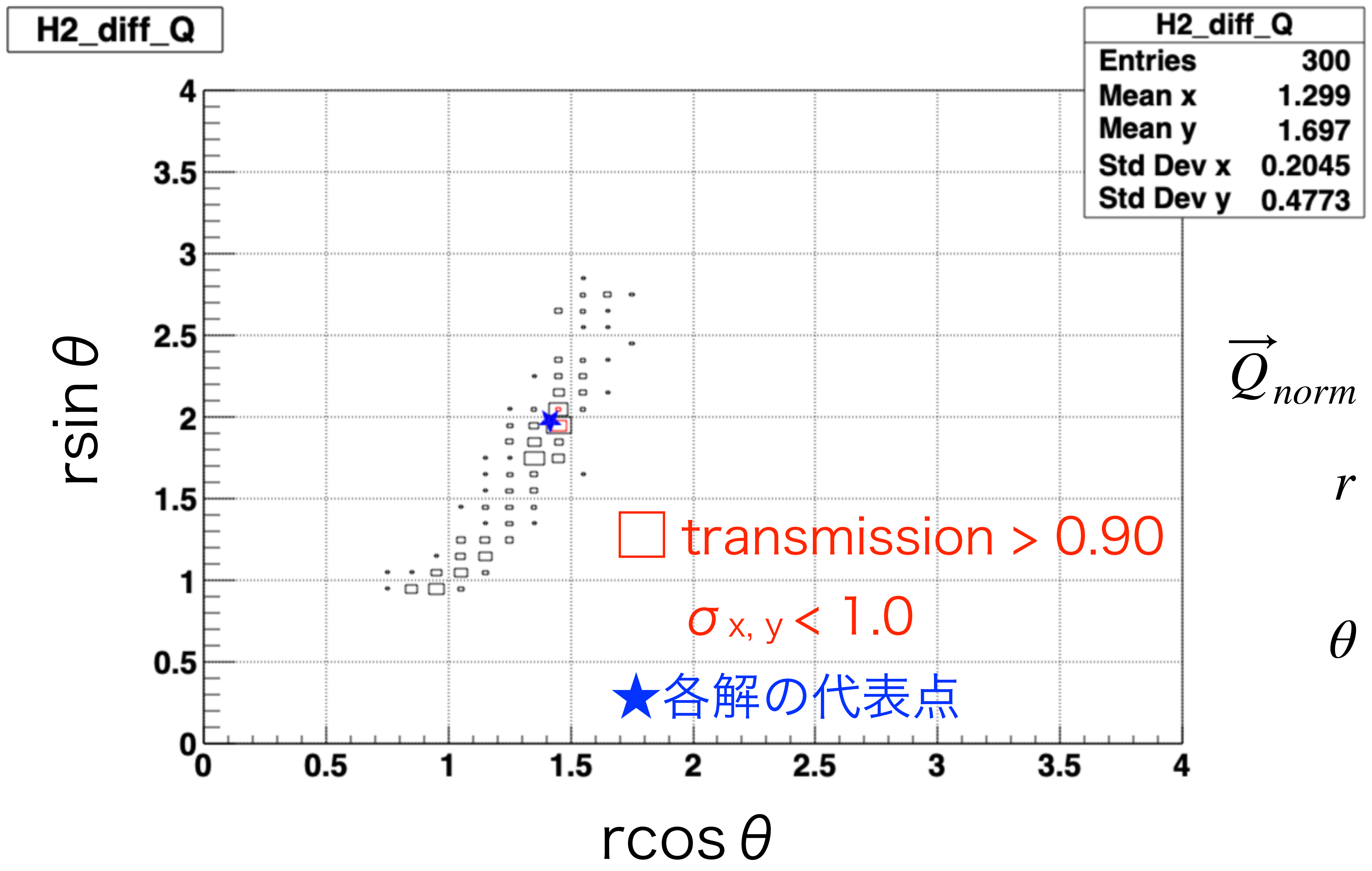
| | epoch | obj. | transmission | σ x | σ y |
|---|---|---|---|---|---|
| 最初に閾値を超えたもの | 241 | -2.40 | 0.98 | 0.98 | 0.78 |
| best | 241 | -2.40 | 0.98 | 0.98 | 0.78 |

# 各ハイパーパラメータでのシミュレーション結果の比較

異なる Acquisition function の比較 (objective function: $f_{sum}$/ phase ellipse A)

Lower Confidence Bound (LCB) / β = 0.2 (default)




$$\vec{Q}_{norm} \equiv \{Q_i/Q_i^0\}$$

$$r \equiv |\vec{Q}_{norm}|/|\vec{Q^0}_{norm}|$$

$$\theta \equiv \arccos \frac{\vec{Q}_{norm} \cdot \vec{Q^0}_{norm}}{|\vec{Q}_{norm}||\vec{Q^0}_{norm}|}$$

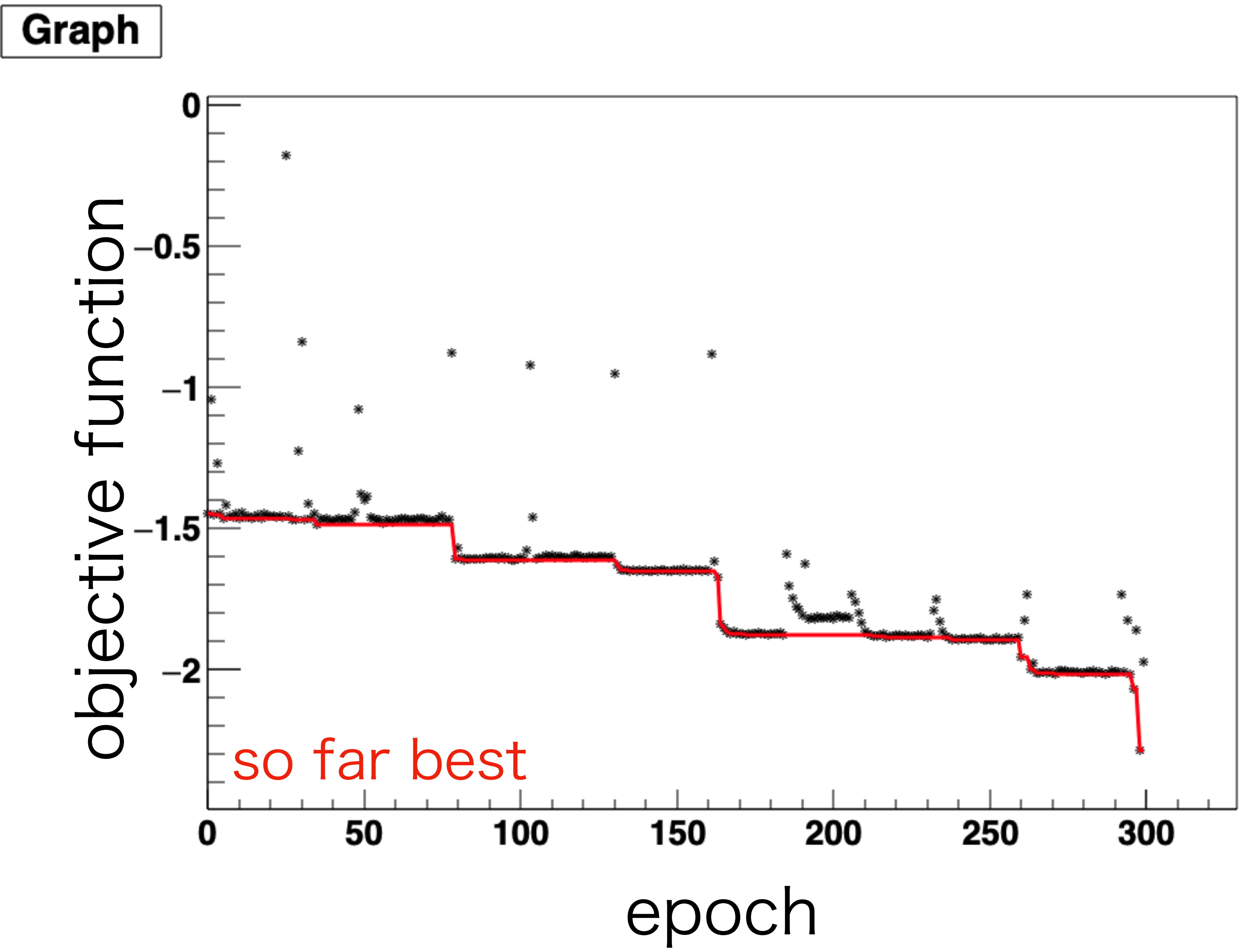
| | epoch | obj. | transmission | $\sigma_x$ | $\sigma_y$ |
|---|---|---|---|---|---|
| **最初に閾値を超えたもの** | 57 | -2.64 | 0.95 | 0.35 | 0.63 |
| **best** | 82 | -2.68 | 0.99 | 0.49 | 0.50 |

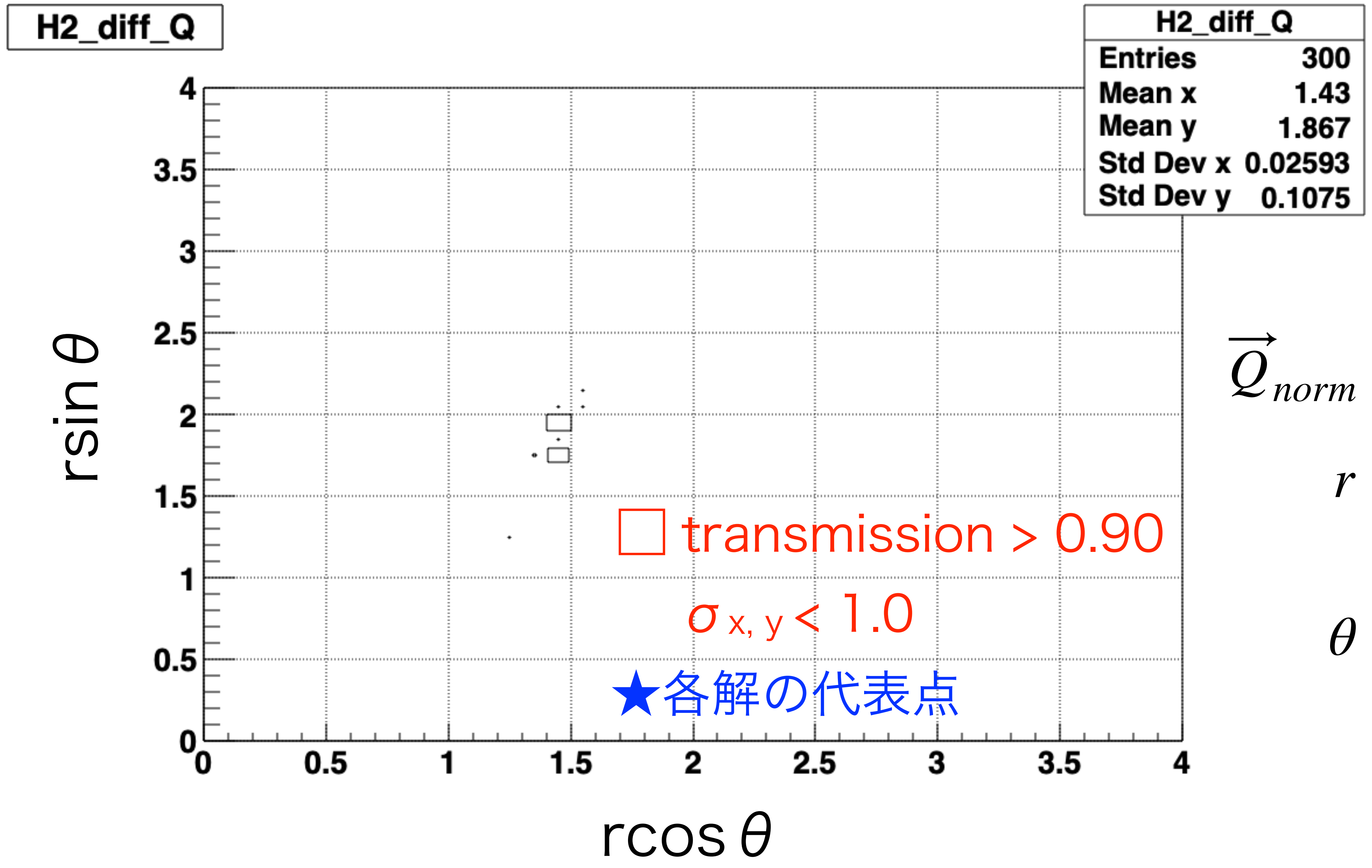
# 各ハイパーパラメータでのシミュレーション結果の比較

異なる Acquisition function の比較 (objective function: $f_{sum}$/ phase ellipse A)

Minimize (scipy.optimize / Powell)

やはりこれは全くダメそう




$$\overrightarrow{Q}_{norm} \equiv \{Q_i/Q_i^0\}$$

$$r \equiv |\overrightarrow{Q}_{norm}|/|\overrightarrow{Q^0}_{norm}|$$

$$\theta \equiv \arccos \frac{\overrightarrow{Q}_{norm} \cdot \overrightarrow{Q^0}_{norm}}{|\overrightarrow{Q}_{norm}||\overrightarrow{Q^0}_{norm}|}$$

| | epoch | obj. | transmission | $\sigma_x$ | $\sigma_y$ |
|---|---|---|---|---|---|
| **最初に閾値を超えたもの** | — | — | — | — | — |
| **best** | 298 | -2.28 | 0.97 | 1.27 | 0.90 |

# 各ハイパーパラメータでのシミュレーション結果の比較

異なる Acquisition function の比較 (objective function: $f_{sum}$)

| phase ellipse A | PI | EI | LCB |
|---|---|---|---|
| 最初に閾値を超えた epoch | 73 | 241 | 57 |
| best obj. | -2.74 | -2.40 | -2.68 |

| phase ellipse B | PI | EI | LCB |
|---|---|---|---|
| 最初に閾値を超えた epoch | 127 | — | 51 |
| best obj. | -2.46 | -2.31 | -2.65 |

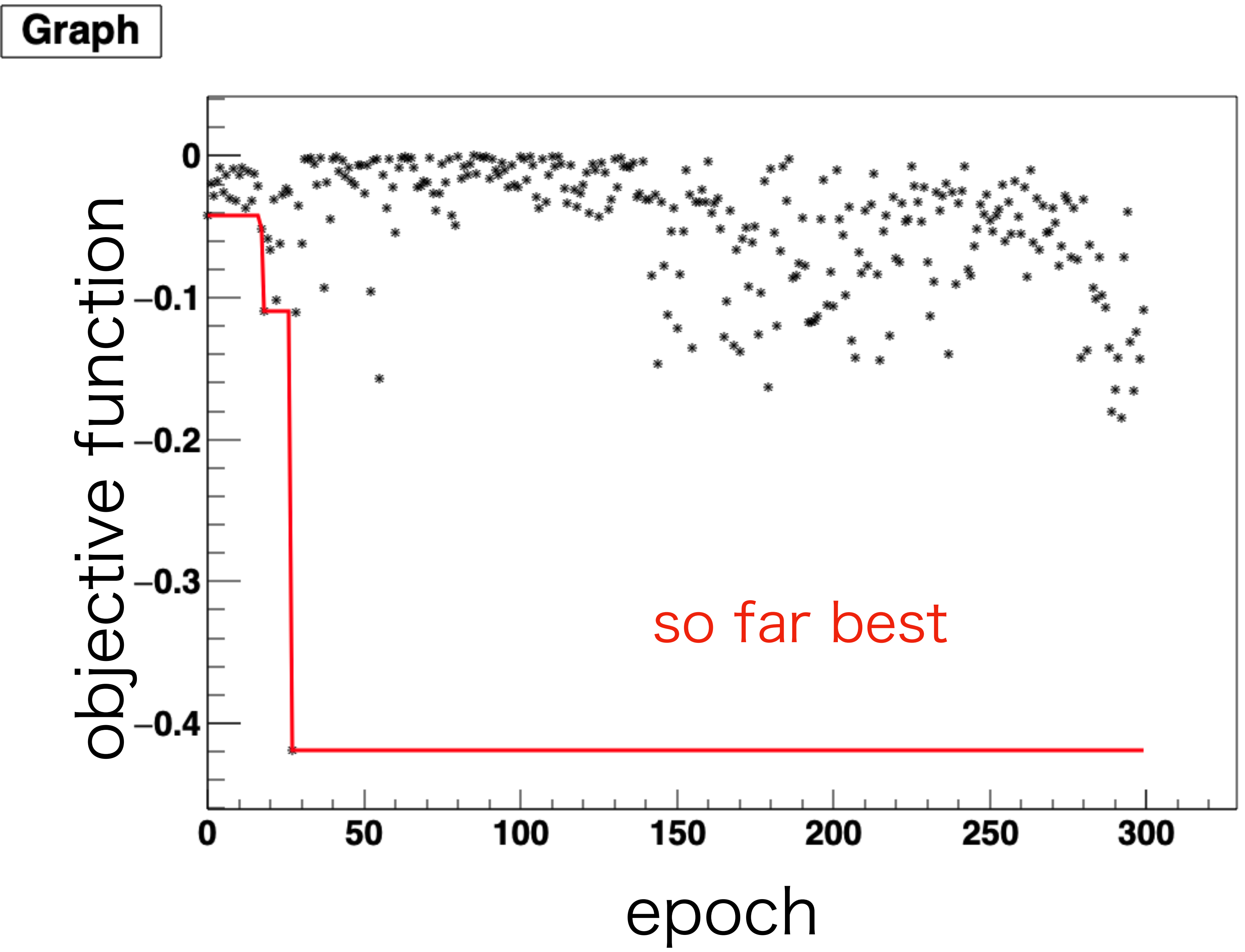
| phase ellipse C | PI | EI | LCB |
|---|---|---|---|
| 最初に閾値を超えた epoch | 283 | — | 143 |
| best obj. | -2.28 | -2.04 | -2.22 |

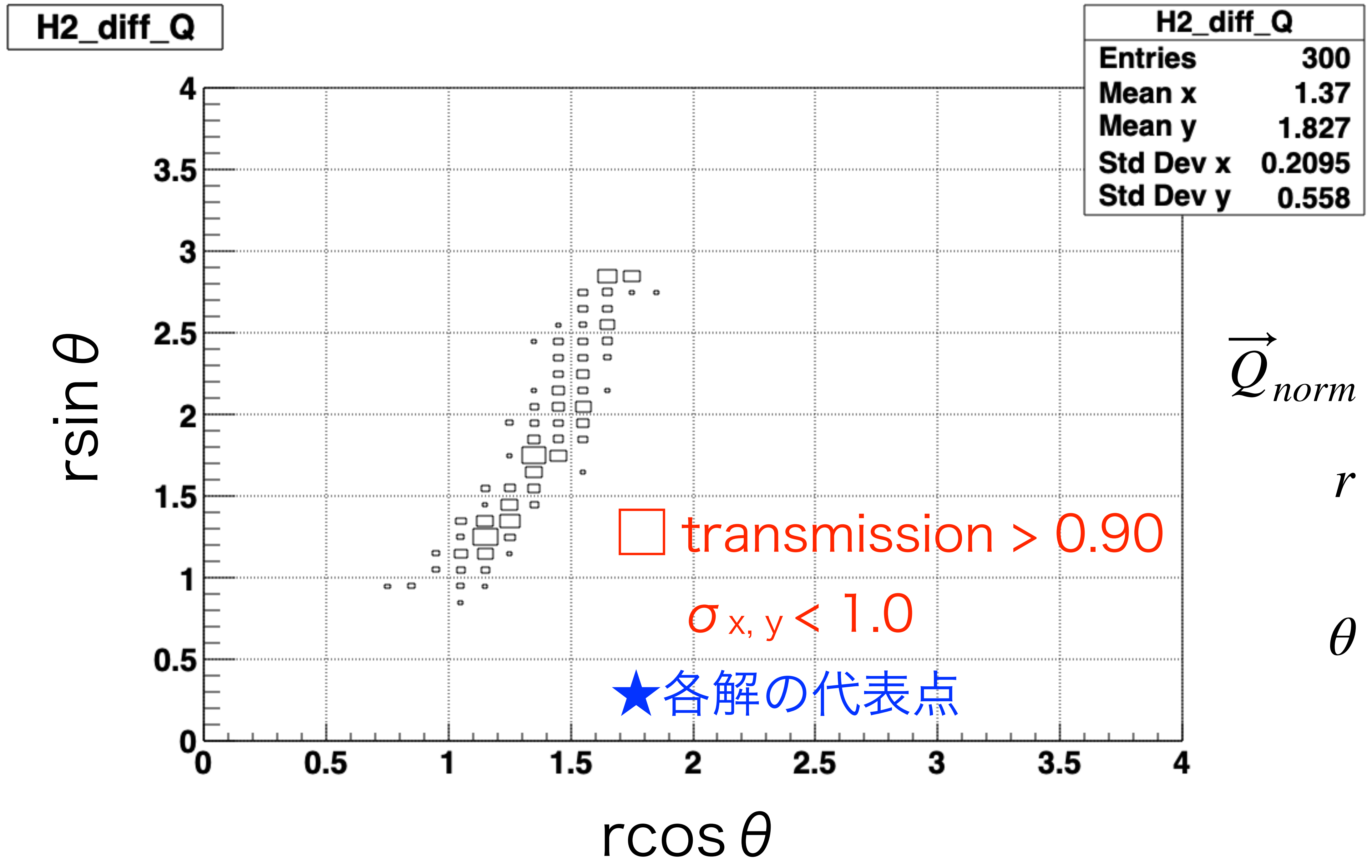
**LCB がもっともバランスが良さそうに見える。**

# 各ハイパーパラメータでのシミュレーション結果の比較

**異なる Objective function の比較 (f_multi)**

Lower Confidence Bound (LCB) / $\beta$ = 0.2 (default)




$$\overrightarrow{Q}_{norm} \equiv \{Q_i/Q_i^0\}$$

$$r \equiv |\overrightarrow{Q}_{norm}|/|\overrightarrow{Q^0}_{norm}|$$

$$\theta \equiv \arccos \frac{\overrightarrow{Q}_{norm} \cdot \overrightarrow{Q^0}_{norm}}{|\overrightarrow{Q}_{norm}||\overrightarrow{Q^0}_{norm}|}$$

| | epoch | obj. | transmission | $\sigma_x$ | $\sigma_y$ |
|---|---|---|---|---|---|
| **最初に閾値を超えたもの** | — | — | — | — | — |
| **best** | 27 | -0.42 | 0.94 | 1.20 | 0.38 |

# 各ハイパーパラメータでのシミュレーション結果の比較

それぞれの Acquisition function での $f_{multi}$ の結果

| phase ellipse A | PI | EI | LCB |
|---|---|---|---|
| 最初に閾値を<br>超えた epoch | — | — | — |
| best obj. | -0.30 | -0.20 | -0.42 |

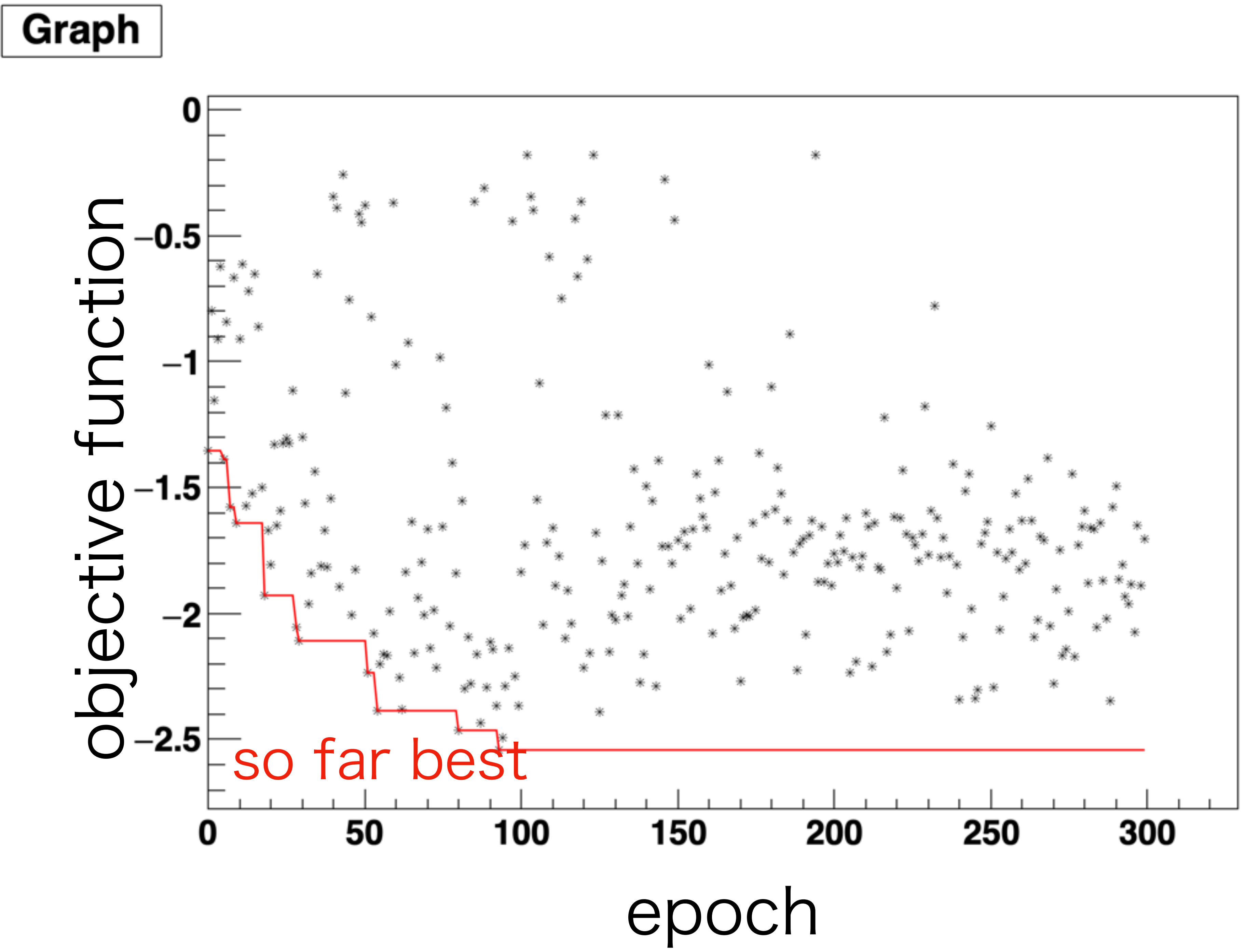
| phase ellipse B | PI | EI | LCB |
|---|---|---|---|
| | obj. | obj. | obj. |
| 最初に閾値を<br>超えたもの | — | — | 130 |
| best | -0.23 | -0.20 | -0.51 |

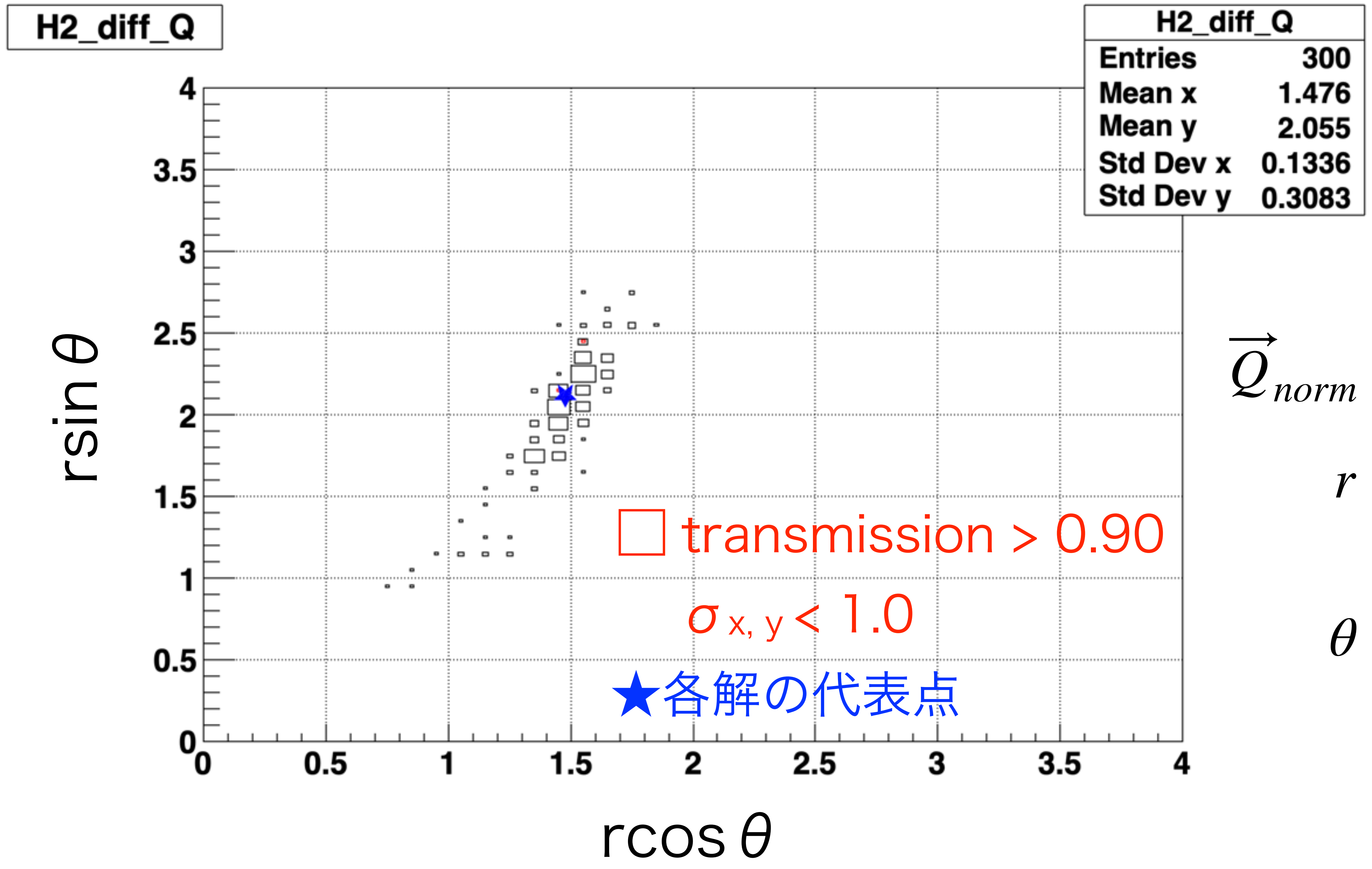
$f_{multi}$ は $f_{sum}$ に比べて機能しなさそう

# 各ハイパーパラメータでのシミュレーション結果の比較

LCB における $\beta$ の違い (objective function: $f_{sum}$/ phase ellipse B)

Lower Confidence Bound (LCB) / $\beta$ = 0.01




$$\vec{Q}_{norm} \equiv \{Q_i/Q_i^0\}$$

$$r \equiv |\vec{Q}_{norm}|/|\vec{Q^0}_{norm}|$$

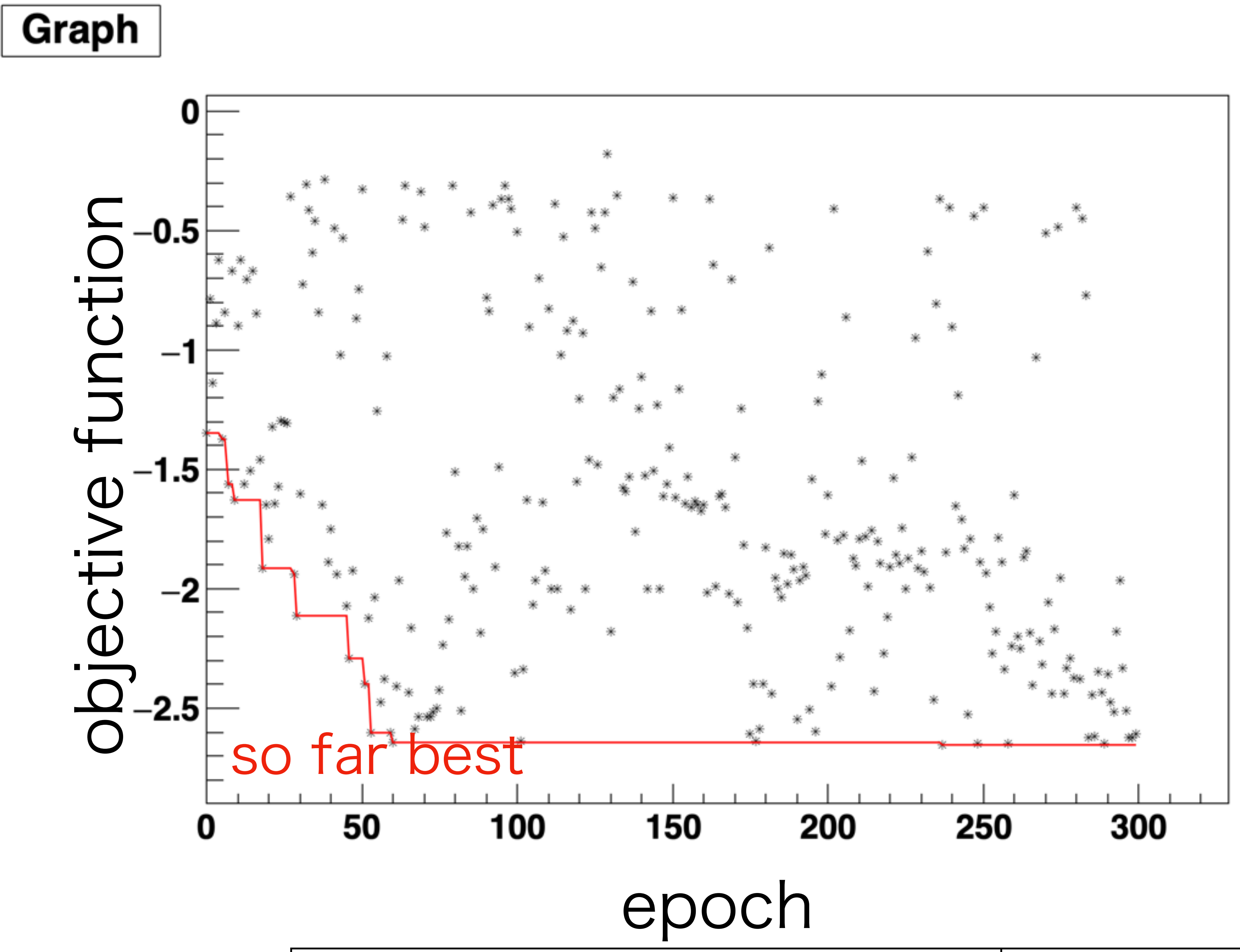
$$\theta \equiv \arccos \frac{\vec{Q}_{norm} \cdot \vec{Q^0}_{norm}}{|\vec{Q}_{norm}||\vec{Q^0}_{norm}|}$$

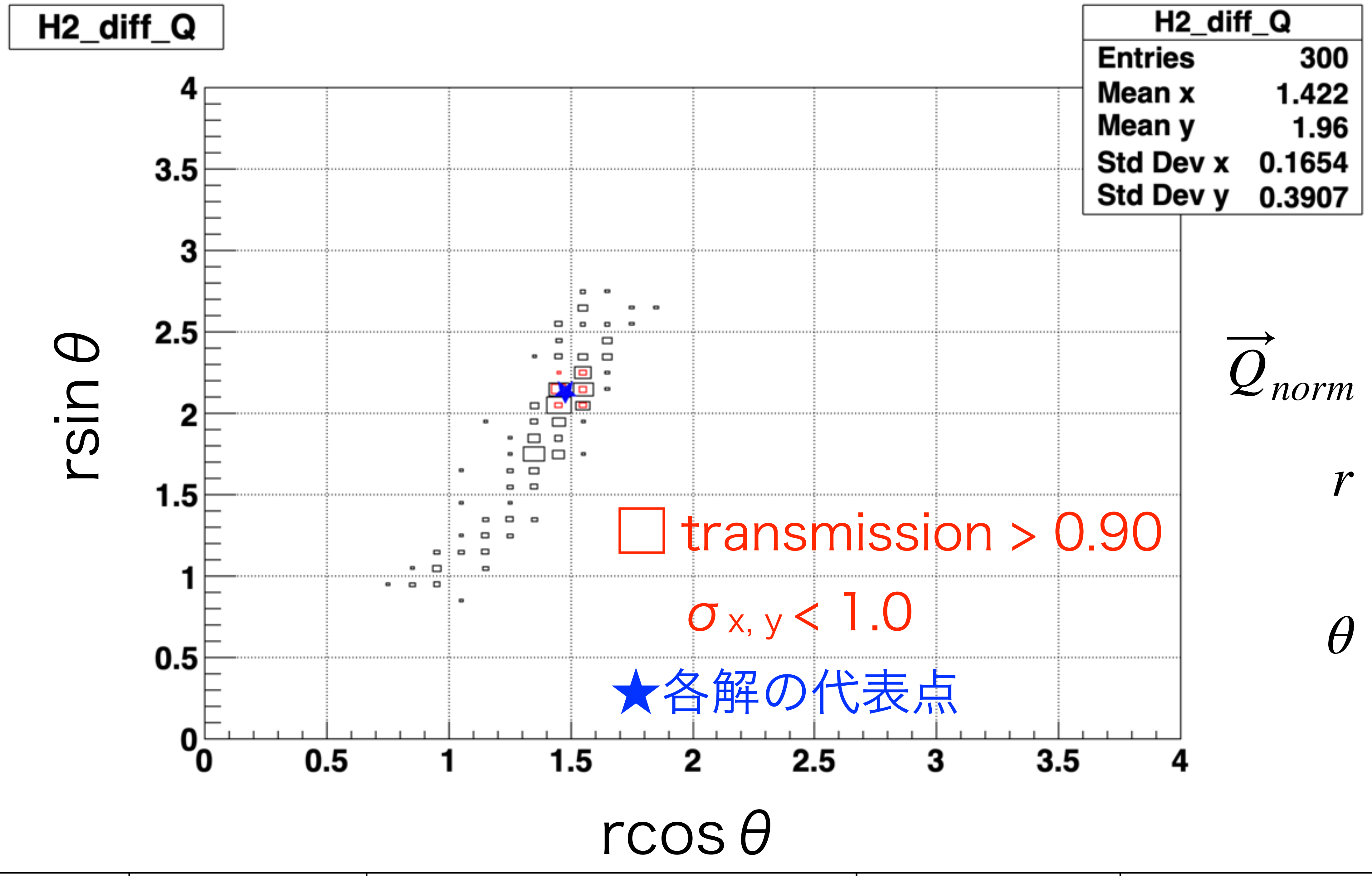
| | epoch | obj. | transmission | $\sigma_x$ | $\sigma_y$ |
|---|---|---|---|---|---|
| **最初に閾値を超えたもの** | 93 | -2.54 | 0.996 | 0.93 | 0.52 |
| **best** | 290 | -2.32 | 0.960 | 0.62 | 0.76 |

# 各ハイパーパラメータでのシミュレーション結果の比較

LCB における β の違い (objective function: $f_{sum}$/ phase ellipse B)

Lower Confidence Bound (LCB) / β = 0.2




$$\vec{Q}_{norm} \equiv \{Q_i/Q_i^0\}$$

$$r \equiv |\vec{Q}_{norm}|/|\vec{Q^0}_{norm}|$$

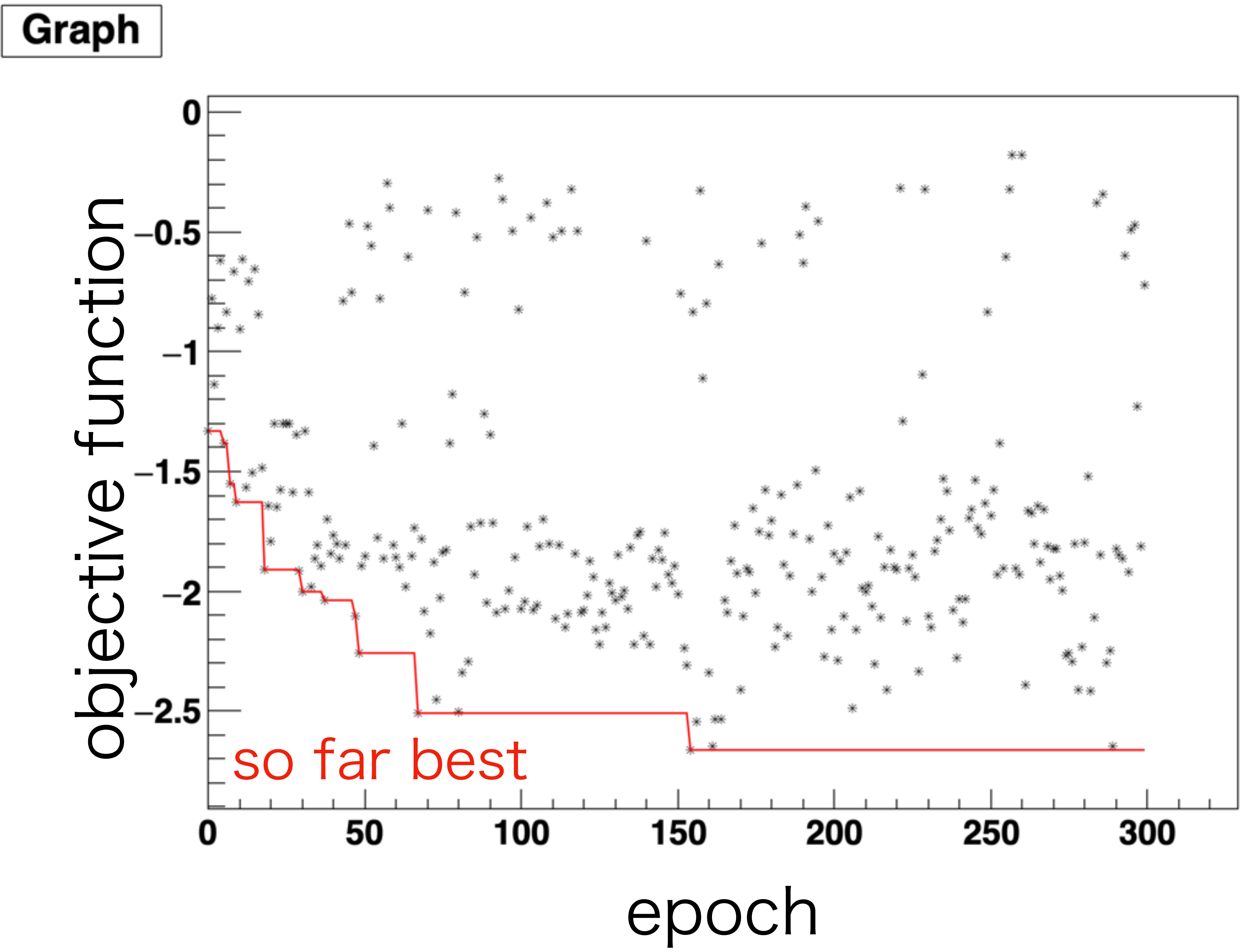
$$\theta \equiv \arccos \frac{\vec{Q}_{norm} \cdot \vec{Q^0}_{norm}}{|\vec{Q}_{norm}||\vec{Q^0}_{norm}|}$$

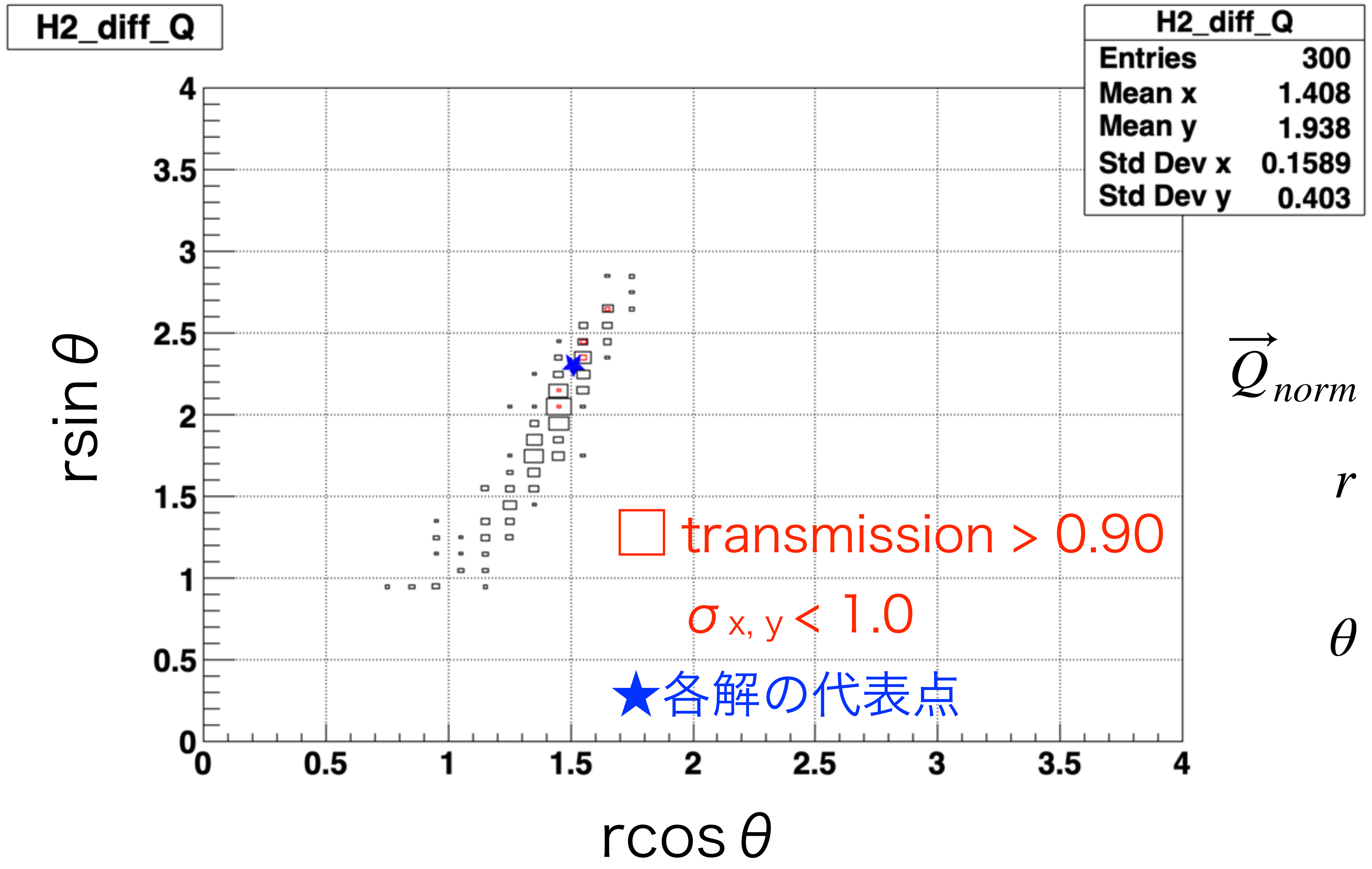
| | epoch | obj. | transmission | σx | σy |
|---|---|---|---|---|---|
| 最初に閾値を超えたもの | 89 | -2.25 | 0.939 | 0.96 | 0.59 |
| best | 237 | -2.65 | 0.920 | 0.49 | 0.37 |

# 各ハイパーパラメータでのシミュレーションの結果の比較

LCB における β の違い (objective function: $f_{sum}$/ phase ellipse B)

Lower Confidence Bound (LCB) / β = 1.0




$$\vec{Q}_{norm} \equiv \{Q_i/Q_i^0\}$$

$$r \equiv |\vec{Q}_{norm}|/|\vec{Q^0}_{norm}|$$

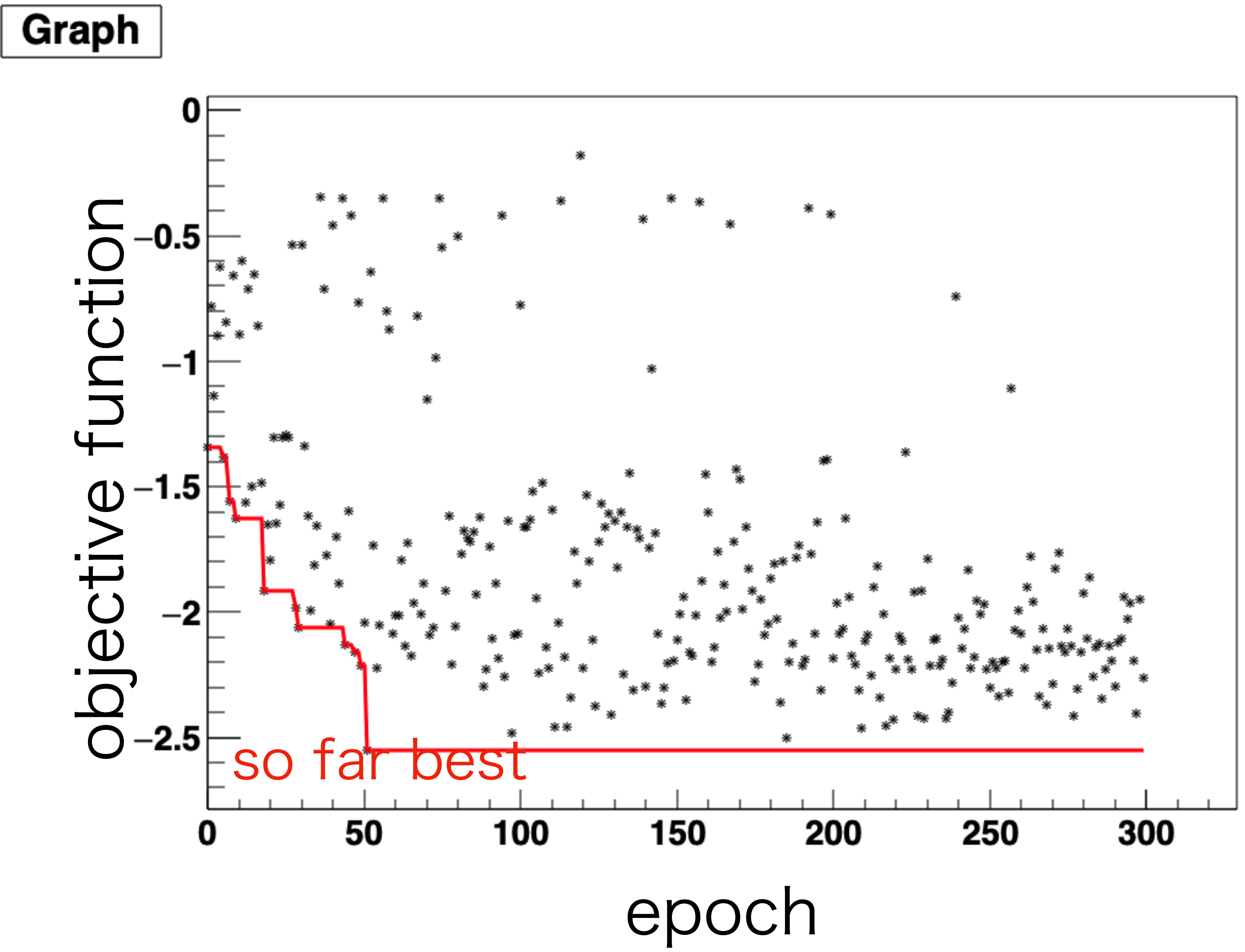
$$\theta \equiv \arccos \frac{\vec{Q}_{norm} \cdot \vec{Q^0}_{norm}}{|\vec{Q}_{norm}||\vec{Q^0}_{norm}|}$$

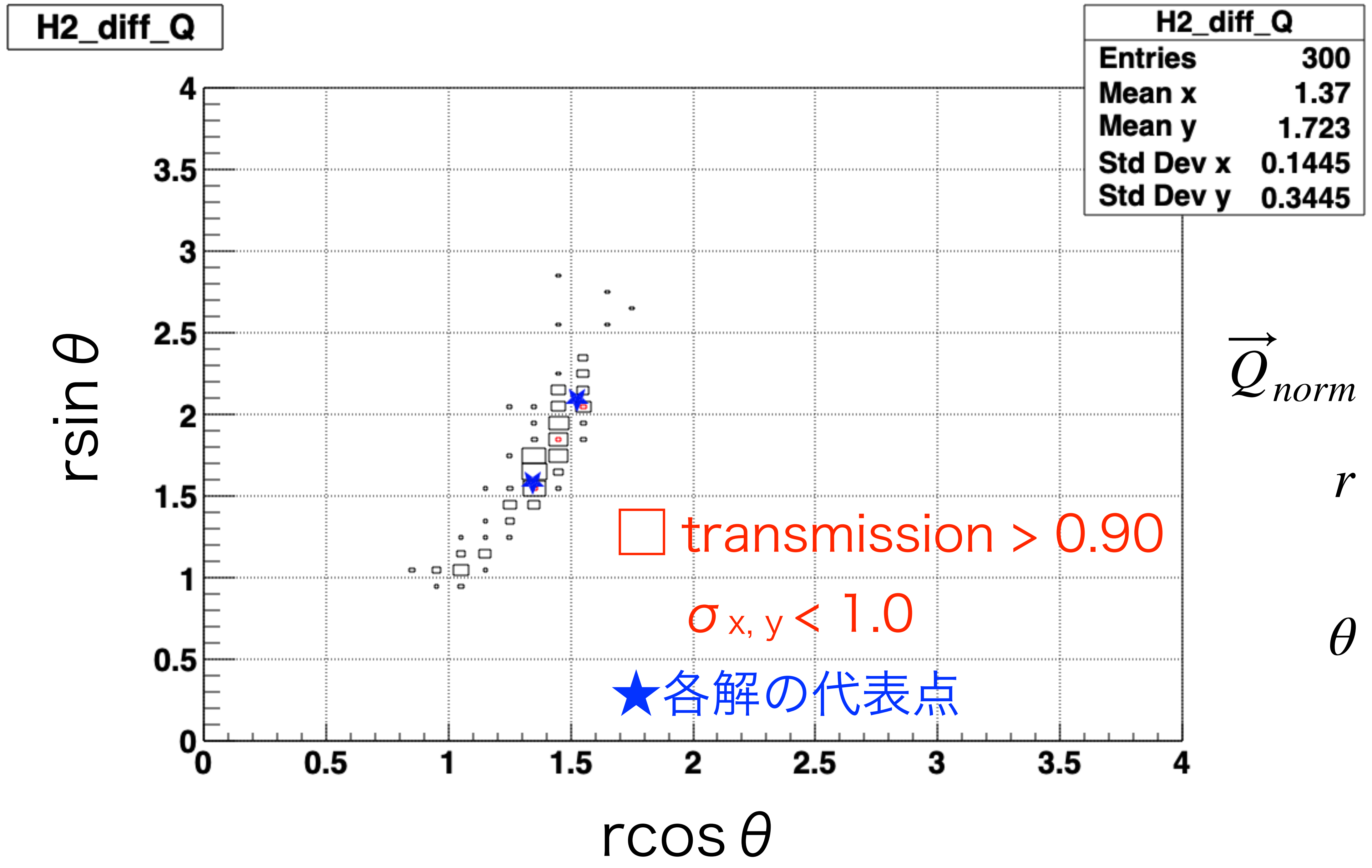
| | epoch | obj. | transmission | $\sigma_x$ | $\sigma_y$ |
|---|---|---|---|---|---|
| **最初に閾値を超えたもの** | 67 | -2.51 | 0.918 | 0.52 | 0.64 |
| **best** | 154 | -2.66 | 0.915 | 0.31 | 0.52 |

# 各ハイパーパラメータでのシミュレーション結果の比較

LCB における $\beta$ の違い (objective function: $f_{sum}$/ phase ellipse B)

Lower Confidence Bound (LCB) / $\beta$ = 9.0




$$\vec{Q}_{norm} \equiv \{Q_i/Q_i^0\}$$

$$r \equiv |\vec{Q}_{norm}|/|\vec{Q^0}_{norm}|$$

$$\theta \equiv \arccos \frac{\vec{Q}_{norm} \cdot \vec{Q^0}_{norm}}{|\vec{Q}_{norm}||\vec{Q^0}_{norm}|}$$

| | epoch | obj. | transmission | $\sigma_x$ | $\sigma_y$ |
|---|---|---|---|---|---|
| 最初に閾値を超えたもの | 111 | -2.46 | 0.966 | 0.59 | 0.92 |
| best | 51 | -2.55 | 0.960 | 1.05 | 0.31 |

# 各ハイパーパラメータでのシミュレーション結果の比較

LCB における $\beta$ の違い (objective function: $f_{sum}$/ phase ellipse B)

| phase ellipse B | 0.01 | 0.2 | 0.5 | 1.0 | 9.0 | 100.0 |
|---|---|---|---|---|---|---|
| 最初に閾値を超えた epoch | 93 | 51 | 157 | 67 | 111 | 82 |
| best obj. | -2.54 | -2.65 | -2.75 | -2.66 | -2.55 | -2.57 |

| phase ellipse C | 0.01 | 0.2 | 0.5 | 1.0 | 9.0 | 100.0 |
|---|---|---|---|---|---|---|
| 最初に閾値を超えた epoch | 262 | 143 | 216 | 151 | — | 89 |
| best obj. | -2.3 | -2.22 | -2.16 | -2.28 | -2.15 | -2.32 |

系統的な $\beta$ への依存性は今回の比較では不明。。。

# 解析の現状まとめと今後の予定

・理研RIBF において、将来的な高強度ビームをより高精度で取り扱うために
機械学習を用いたビーム制御のための研究を始めている。
・シミュレーションを用いて、ビームトランスポートに対するベイズ最適化について
ハイパーパラメータの最適化手法の検討を行った。
・Acquisition function については、今回の範囲では LCB (UCB) が最も優秀な結果を示した。
・Objective function については、sum の形での定義した場合にうまく機能した。
・LCB のパラメータ $\beta$ の効果は今回の検証では確認できなかった。

・今後、ある程度の候補が見つかった時点で**より近傍探索の手法に切り替える**ことを考える。
(ただし、近傍から通常の最急降下法は最適化までの 試行回数の短縮には繋がらなかった)
・また、シミュレーションを基にした事前学習を取り入れていく。
(VAE / physics informed Gaussian process など)
・RIBF の他のビームライン (RILAC/LEBTなど) にもシミュレーション含め適応していく予定。

# Detail of Algorithm: Bayesian Optimization with GPR

1. Measurement of data as “initial values”
   for N parameters, measurement of 2N + 1 data point : default ± $\sigma_i$ ( i = 1, 2, ⋯, N)

2. Create a model of the distribution of the objective function based measured data using GPR
   In Bayesian estimation, objective function has distribution for any x, i.e.,

$$p(t|x) = \frac{1}{\sqrt{2\pi}\sigma(x)} \exp\left(-\frac{(x-m(x))^2}{2\sigma(x)^2}\right)$$

Create a model means estimate m(x) and $\sigma$(x) for any x.
In linear estimation model,

# Detail of Algorithm: Bayesian Optimization with GPR

1. Measurement of data as “initial values”
   for N parameters, measurement of 2N + 1 data point : default ± $\sigma_i$ ( i = 1, 2, ⋯, N)

2. Create a model of the distribution of the objective function based measured data using GPR
   In Bayesian estimation, objective function has distribution for any x, i.e.,

$$p(t|x) = \frac{1}{\sqrt{2\pi}\sigma(x)} \exp\left(-\frac{(x - m(x))^2}{2\sigma(x)^2}\right)$$

Create a model means estimate m(x) and $\sigma$(x) for any x.
In linear estimation model,

$$m(x) = \Sigma w_m \phi_m(x)$$

w: model parameter $\phi$: basis function

C. M. Bishop, Pattern Recognition and Machine Learning, Springer (2006)

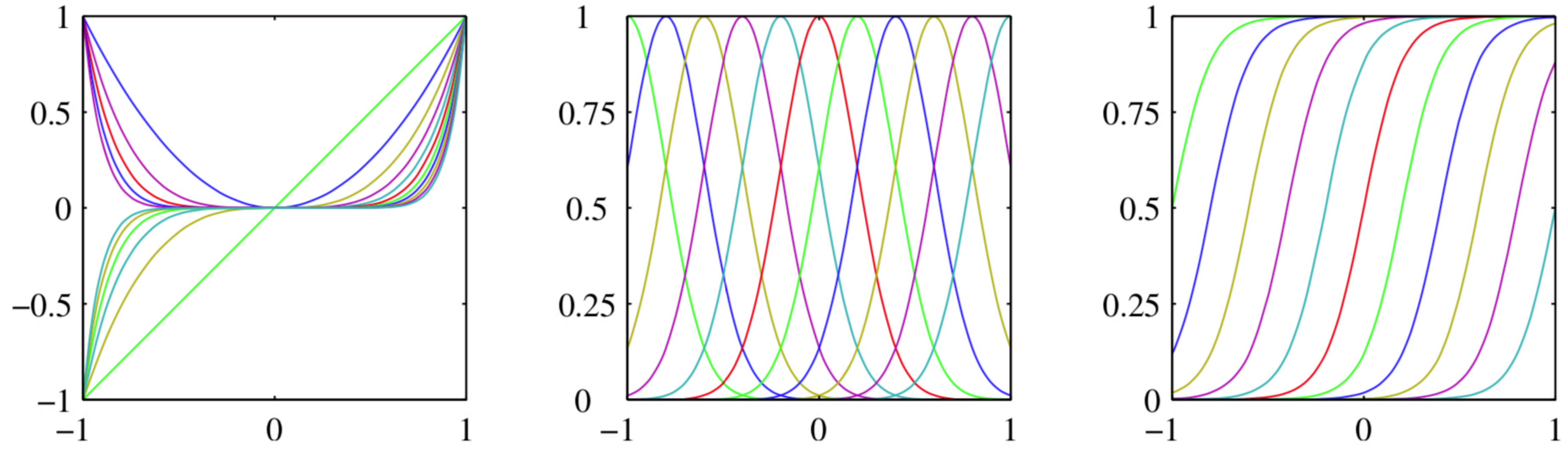


**Figure 3.1** Examples of basis functions, showing polynomials on the left, Gaussians of the form (3.4) in the centre, and sigmoidal of the form (3.5) on the right.

# Detail of Algorithm: Bayesian Optimization with GPR

1. Measurement of data as “initial values”
   for N parameters, measurement of 2N + 1 data point : default ± $\sigma_i$ ( i = 1, 2, ⋯, N)

2. Create a model of the distribution of the objective function based measured data using GPR
   In Bayesian estimation, objective function has distribution for any x, i.e.,

$$p(t|x) = \frac{1}{\sqrt{2\pi}\sigma(x)} \exp\left(-\frac{(x - m(x))^2}{2\sigma(x)^2}\right)$$

Create a model means estimate m(x) and $\sigma$(x) for any x.
In linear estimation model,

$$m(x) = \Sigma w_m \phi_m(x)$$

w: model parameter $\phi$: basis function
Instead of determine w from data set, estimate probability distribution of w and integral it as

$$p(t|x) = \int p(t|x, \overrightarrow{w})p(w|\underbrace{\{x_i, t_i\}}_{\text{measured data set so far}})d\overrightarrow{w}\ .$$

# Detail of Algorithm: Bayesian Optimization with GPR

1. Measurement of data as "initial values"
   for N parameters, measurement of 2N + 1 data point : default ± $\sigma_i$ ( i = 1, 2, ⋯, N)

2. Create a model of the distribution of the objective function based measured data using GPR
   In Bayesian estimation, objective function has distribution for any x, i.e.,

$$p(t|x) = \frac{1}{\sqrt{2\pi}\sigma(x)} \exp\left( -\frac{(x - m(x))^2}{2\sigma(x)^2} \right)$$

m(x) and $\sigma$(x) are calculated with $\{x_i, \sigma_i\}_N$ as

$$m_N(x) = \Sigma k(x, x_i)t_i, \qquad \sigma_N(x)^2 = \frac{1}{\beta} + \vec{\phi}(x)^T \mathrm{S_N} \vec{\phi}(x),$$

$$k(x, x') \equiv \beta\phi(x)\mathrm{S}_N\phi(x'), \qquad \mathrm{S_N}^{-1} = \alpha \mathrm{I} + \beta \Phi^T \Phi \,.$$

$$\mathbf{\Phi} = \begin{pmatrix} \phi_0(\mathbf{x}_1) & \phi_1(\mathbf{x}_1) & \cdots & \phi_{M-1}(\mathbf{x}_1) \\ \phi_0(\mathbf{x}_2) & \phi_1(\mathbf{x}_2) & \cdots & \phi_{M-1}(\mathbf{x}_2) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_0(\mathbf{x}_N) & \phi_1(\mathbf{x}_N) & \cdots & \phi_{M-1}(\mathbf{x}_N) \end{pmatrix}.$$

※$\alpha$: parameter of prior distribution of w
$\beta$: precision parameter of the measurement itself

# Detail of Algorithm: Bayesian Optimization with GPR

1. Measurement of data as "initial values"
   for N parameters, measurement of 2N + 1 data point : default ± $\sigma_i$ ( i = 1, 2, ⋯, N)

2. Create a model of the distribution of the objective function based measured data using GPR
   In Bayesian estimation, objective function has distribution for any x, i.e.,

$$p(t|x) = \frac{1}{\sqrt{2\pi}\sigma(x)} \exp\left( -\frac{(x - m(x))^2}{2\sigma(x)^2} \right)$$

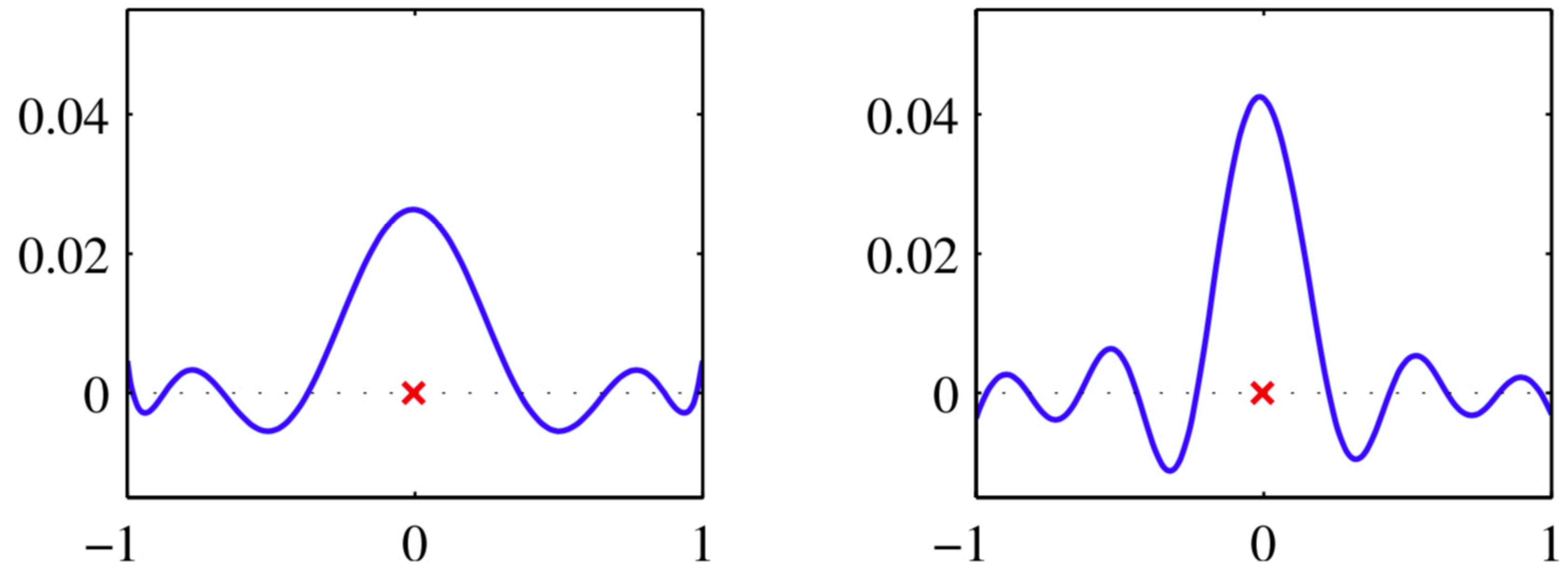
m(x) and $\sigma$(x) are calculated with $\{x_i, \sigma_i\}_N$ as

$$m_N(x) = \underline{\Sigma k(x, x_i)} t_i, \qquad \sigma_N(x)^2 = \frac{1}{\beta} + \vec{\phi}(x)^T S_N \vec{\phi}(x),$$

kernel function:
"Indicator of the strength of correlation between $x_i$ and x"

C. M. Bishop, Pattern Recognition and Machine Learning, Springer (2006)

**Figure 3.11** Examples of equivalent kernels $k(x, x')$ for $x = 0$ plotted as a function of $x'$, corresponding (left) to the polynomial basis functions and (right) to the sigmoidal basis functions shown in Figure 3.1. Note that these are localized functions of $x'$ even though the corresponding basis functions are nonlocal.



# Detail of Algorithm: Bayesian Optimization with GPR

1. Measurement of data as "initial values"
   for N parameters, measurement of 2N + 1 data point : default ± $\sigma_i$ ( i = 1, 2, ⋯, N)

2. Create a model of the distribution of the objective function based measured data using GPR
   In Bayesian estimation, objective function has distribution for any x, i.e.,

$$p(t|x) = \frac{1}{\sqrt{2\pi}\sigma(x)} \exp\left(-\frac{(x-m(x))^2}{2\sigma(x)^2}\right)$$

m(x) and $\sigma$(x) are calculated with $\{x_i, \sigma_i\}_N$ as

$$m_N(x) = \Sigma k(x, x_i) t_i, \qquad \sigma_N(x)^2 = \underbrace{\frac{1}{\beta}}_{\text{Intrinsic error of the measurement}} + \underbrace{\overrightarrow{\phi}(x)^T \mathrm{S_N} \overrightarrow{\phi}(x)}_{\text{Ambiguity from the estimation}},$$

# Detail of Algorithm: Bayesian Optimization with GPR

1. Measurement of data as “initial values”
   for N parameters, measurement of 2N + 1 data point : default ± $\sigma_i$ ( i = 1, 2, ⋯, N)

2. Create a model of the distribution of the objective function based measured data using GPR

3. Calculate EI (Expected Improvement) for all x based on the created model
   Expected improvement is defined as

$$\mathbb{E}_{\alpha}(max\{f(x,\alpha) - f(\hat{x}),0\}) \equiv \int P(\alpha)max\{f(x,\alpha) - f(\hat{x}),0\}d\alpha\,.$$

   f(x, $\alpha$): objective function $\alpha$: error indicators P($\alpha$) : Probability function of $\alpha$
   This value becomes large when the information is insufficient.

# Detail of Algorithm: Bayesian Optimization with GPR

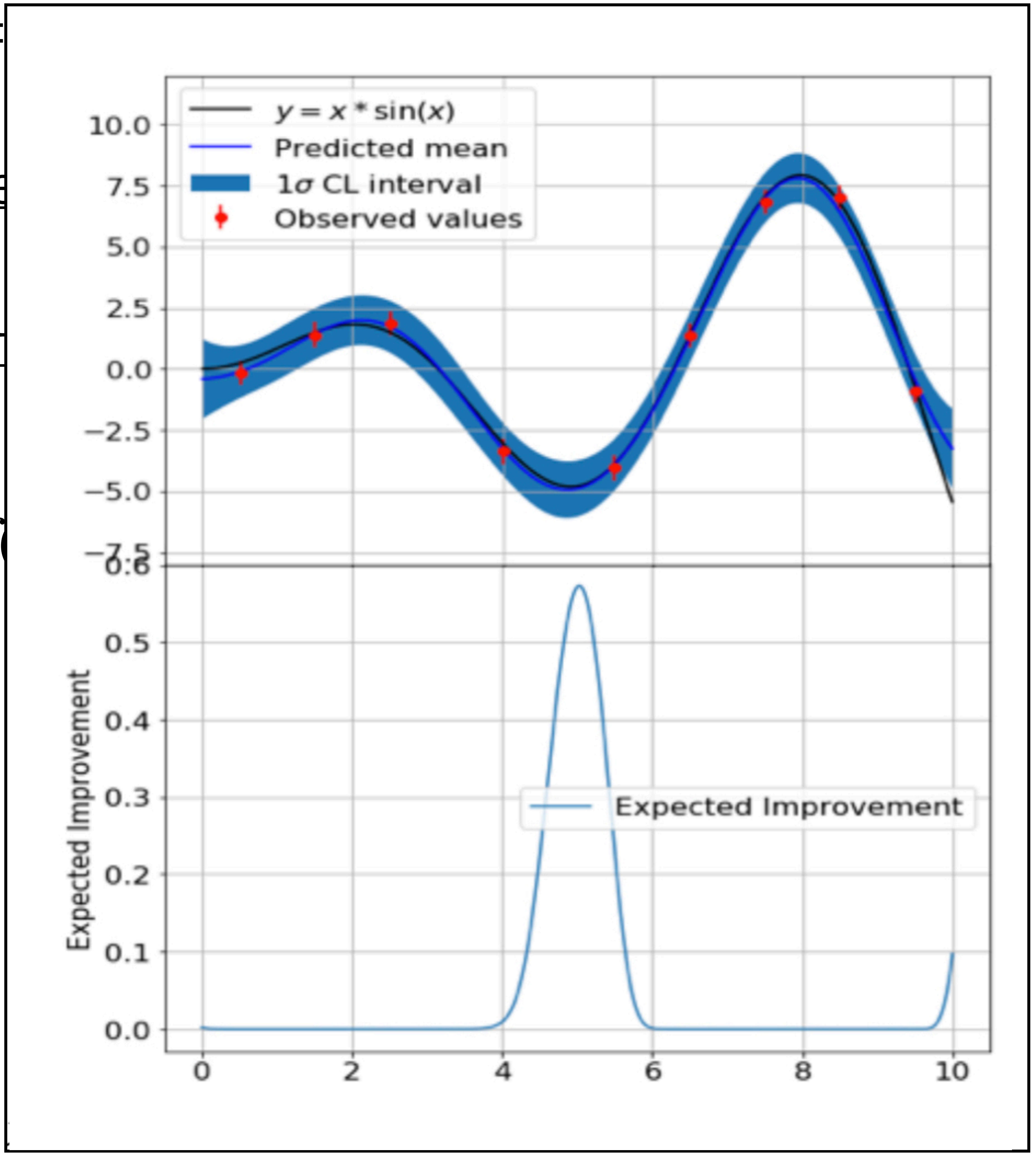
1. Measurement of data as “initial values”
   for N parameters, measurement of 2N + 1 data point : default ±

2. Create a model of the distribution of the objective function base

3. Calculate EI (Expected Improvement) for all x based on the creat
   Expected improvement is defined as

$$\mathbb{E}_\alpha(max\{f(x,\alpha)-f(\hat{x}),0\}) \equiv \int P(\alpha)max\{f(x,\alpha)-f($$

   f(x, $\alpha$): objective function $\alpha$: error indicators P($\alpha$) : Probability
   This value becomes large when the information is insufficient.




E. Iwai et al, Proc. of PASJ 2021 WEOB02

# Detail of Algorithm: Bayesian Optimization with GPR

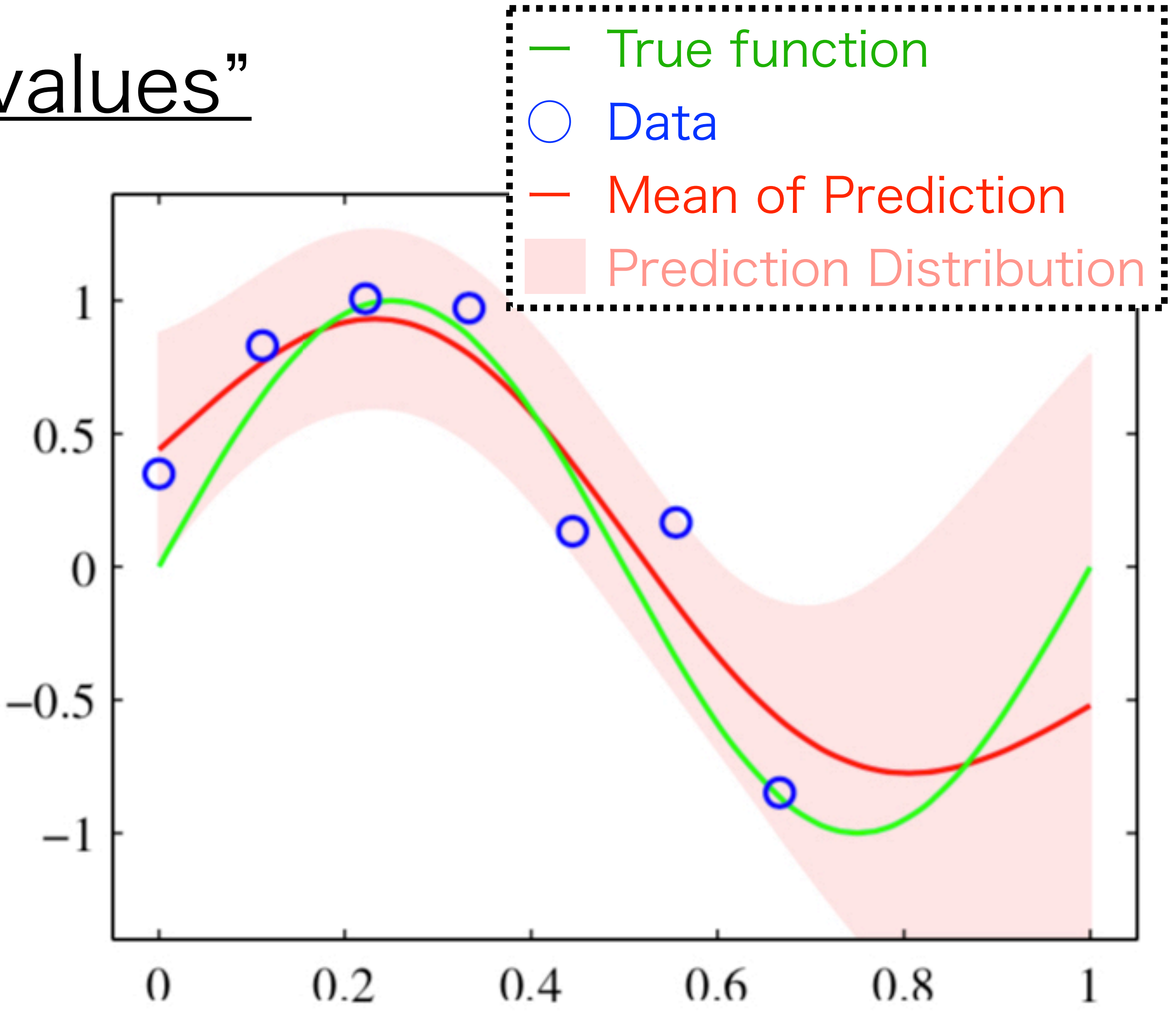
1. Measurement of data as "initial values"
   for N parameters, measurement

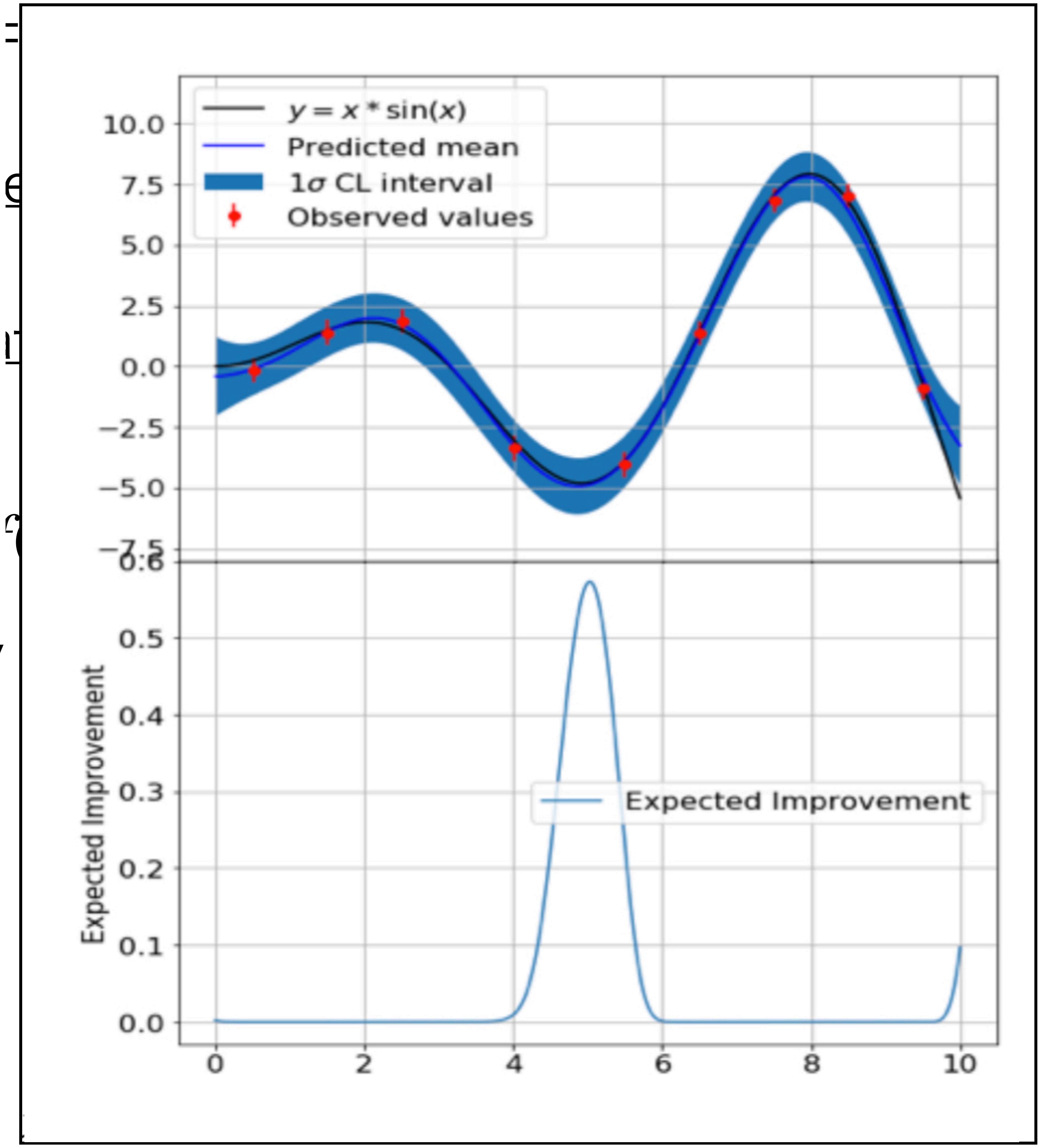
2. Create a model of the distributic

3. Calculate EI (Expected Improven
   Expected improvement is define

$$\mathbb{E}_{\alpha}(max\{f(x,\alpha)-f($$

f(x, $\alpha$): objective function $\alpha$: error indicators P($\alpha$) : Probability
This value becomes large when the information is insufficient.



# Detail of Algorithm: Bayesian Optimization with GPR

1. Measurement of data as "initial values"
   for N parameters, measurement of 2N + 1 data point : default ± $\sigma_i$ ( i = 1, 2, ⋯, N)

2. Create a model of the distribution of the objective function based measured data using GPR

3. Calculate EI (Expected Improvement) for all x based on the created model
   Expected improvement is defined as

$$\mathbb{E}_{\alpha}(max\{f(x,\alpha)-f(\hat{x}),0\}) \equiv \int P(\alpha)max\{f(x,\alpha)-f(\hat{x}),0\}d\alpha\,.$$

   f(x, $\alpha$): objective function $\alpha$: error indicators P($\alpha$) : Probability function of $\alpha$
   This value becomes large when the information is insufficient.

4. Apply a new parameter with maximum EI to the real system and measure the data
   to search the point with the maximum EI, we use the library BoTorch /GPyTorch

5. Repeat 2 ~ 4 and search the best point

# Development History of GPR Optimization System@RIBF

## – 1st attempt an automatic optics optimization by ML

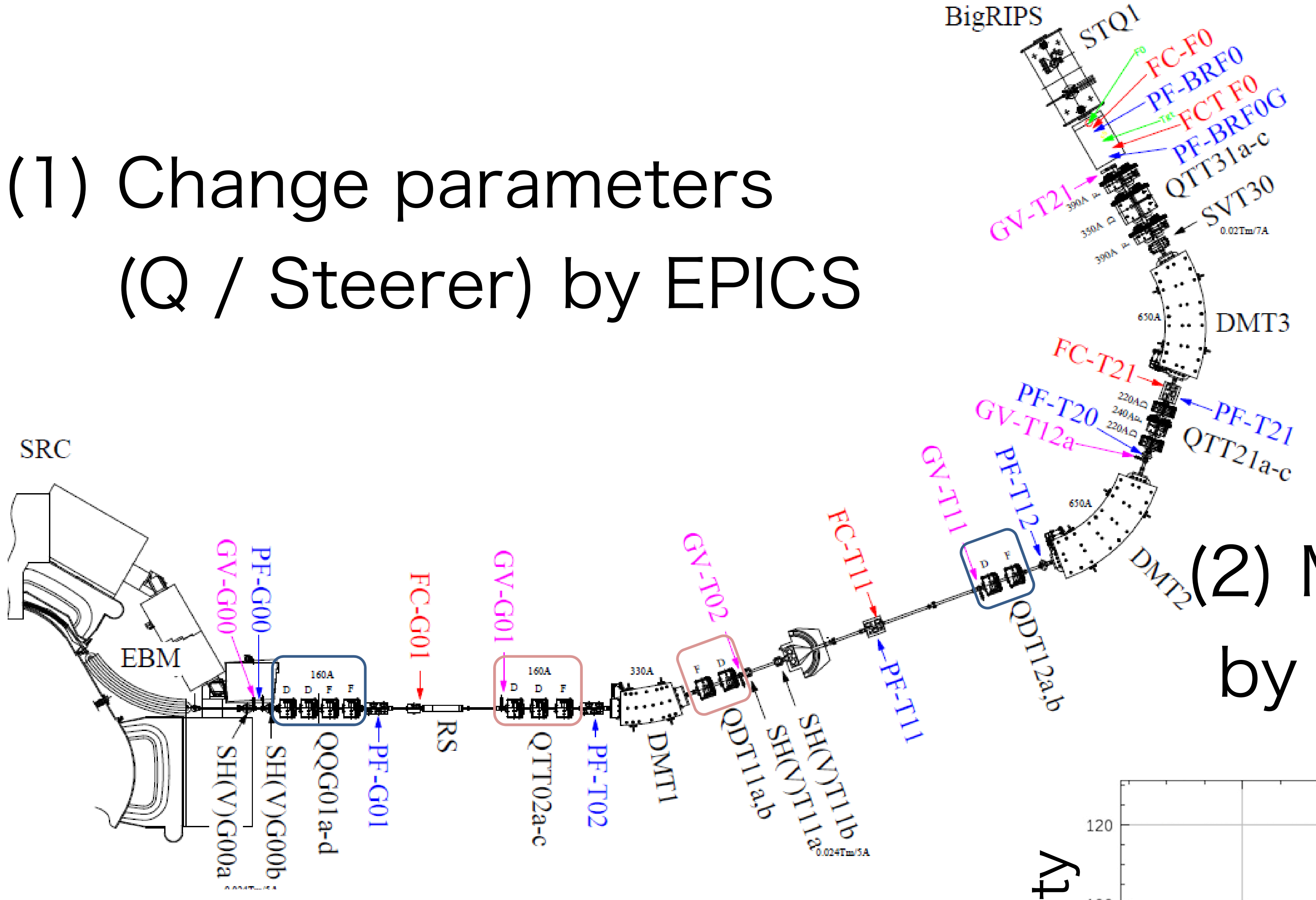
・simultaneous optimization of transmission and beam spot

## – Development of indicators for high intensity beams

・secondary beam with charge conversion

# Automatic Optics Optimization ①: Test with Low Intensity Beam

(1) Change parameters
(Q / Steerer) by EPICS

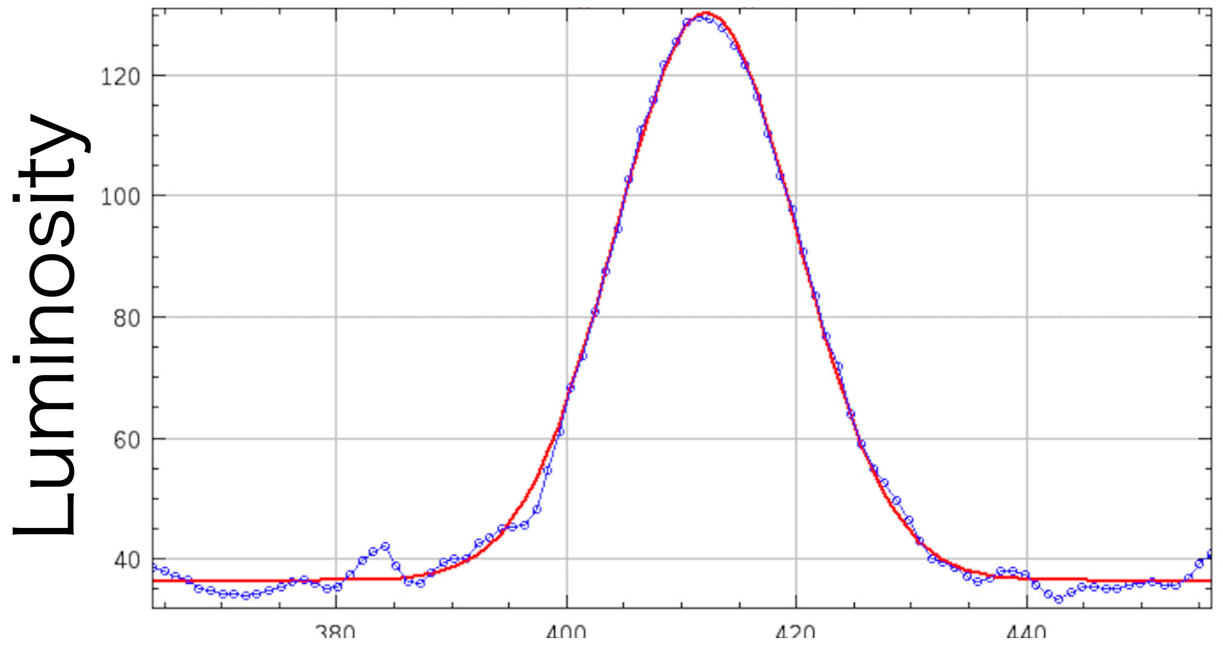


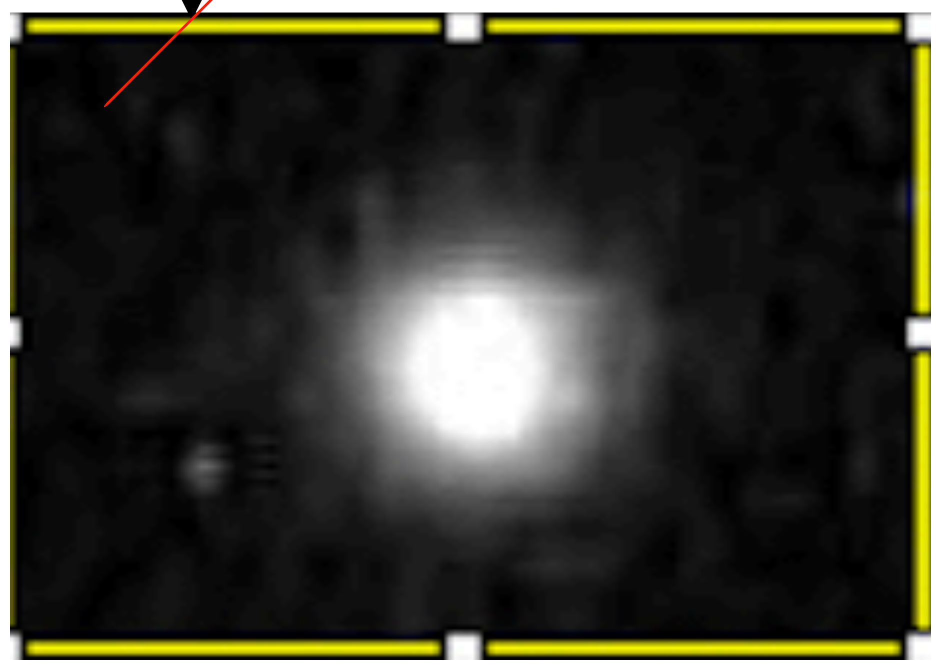

$U^{86+}$ 345 MeV/u
~ 0.001 enA
2.0 e+7 cps

(3) Calculate objective function
Create model by GPR
Choose next param.

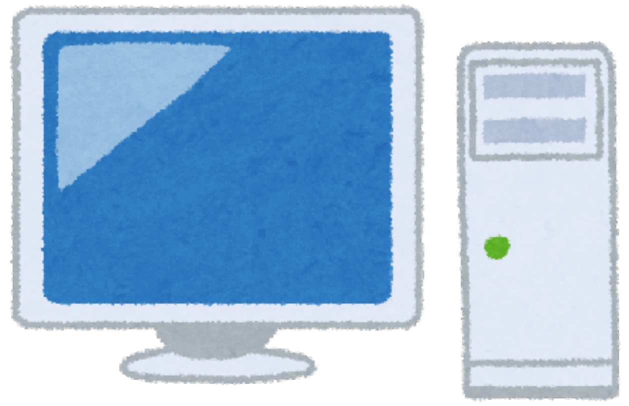
**Gaussian Process Regression (Baysian Opimization)**

(2) Measure beam spot
by fluorescent viewer



Viewer image

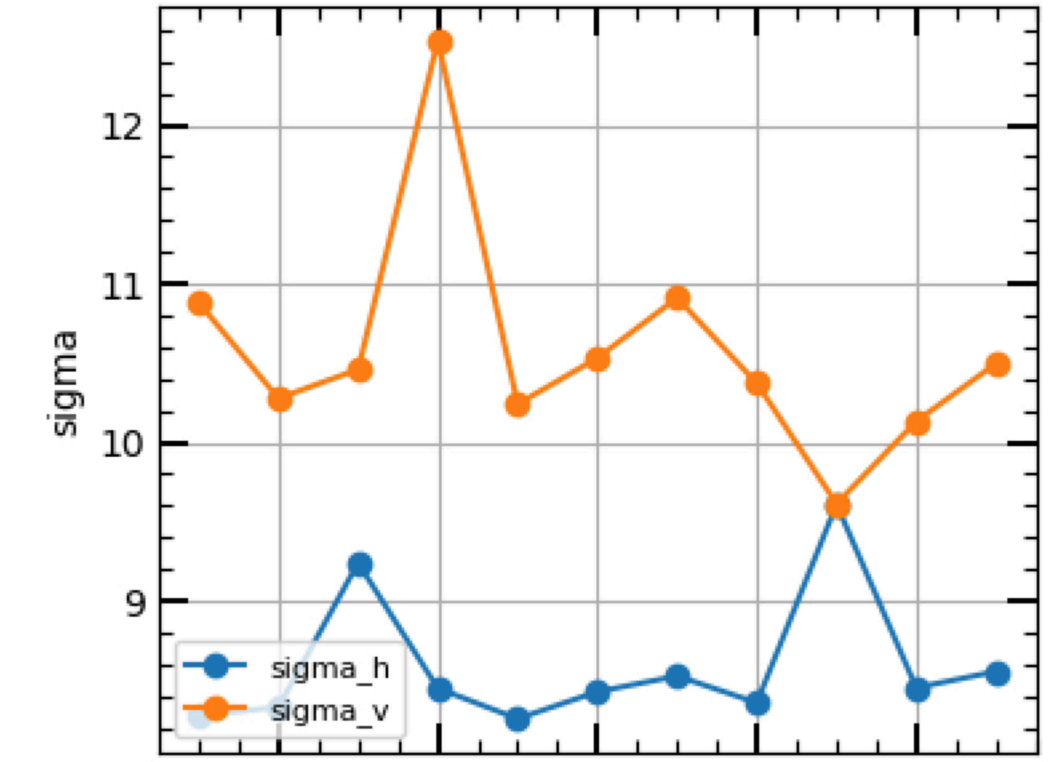


**Intensity : Integral of image luminosity**
**Spot : Image width**

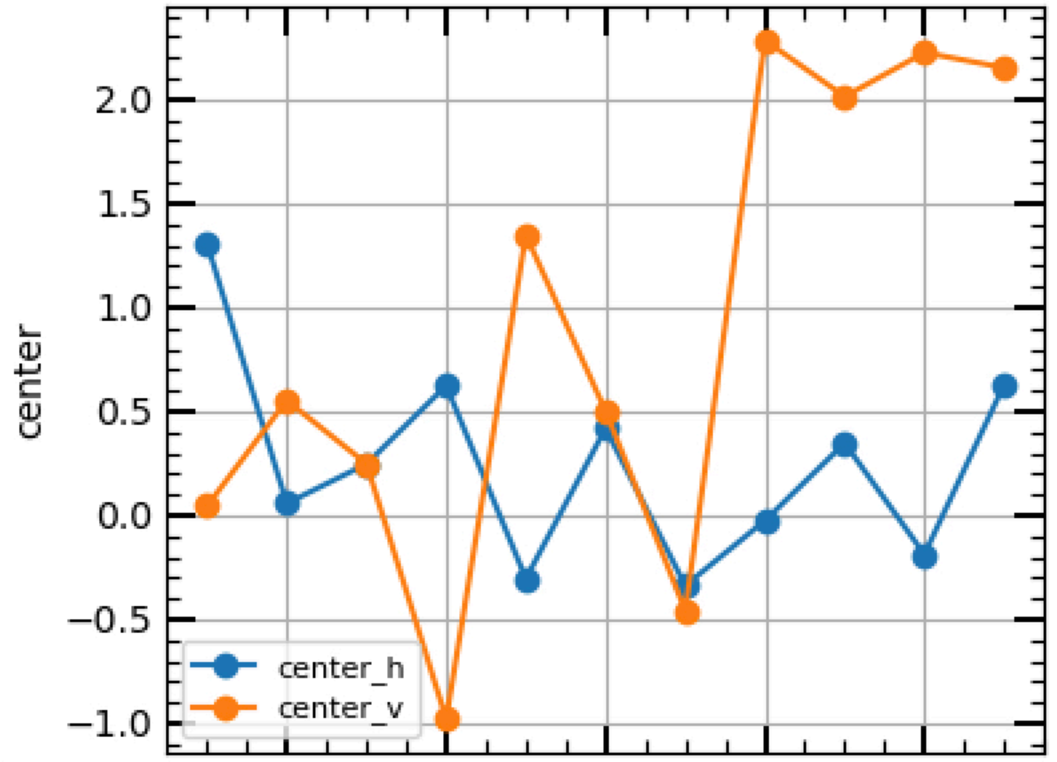
Repeat (1) — (3)
to optimize optics

# 1st Exp. : Auto Tuning with Low Intensity Beam

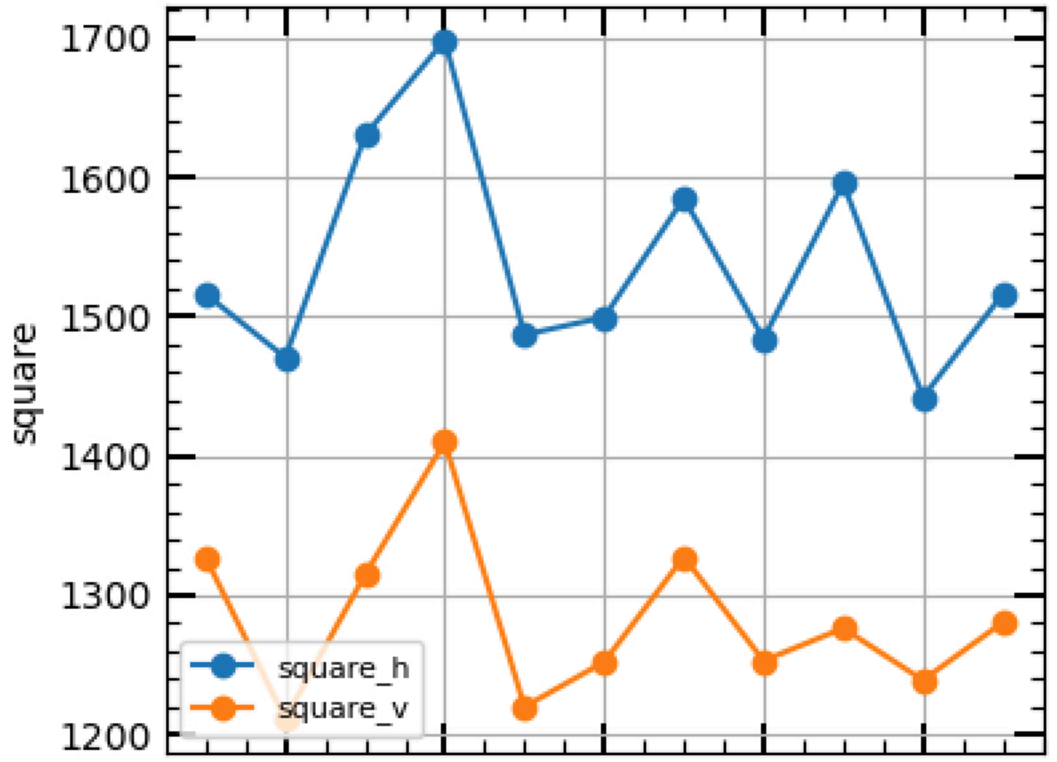
Demonstration test at 2020 Oct. 21:00 ~ 9:00



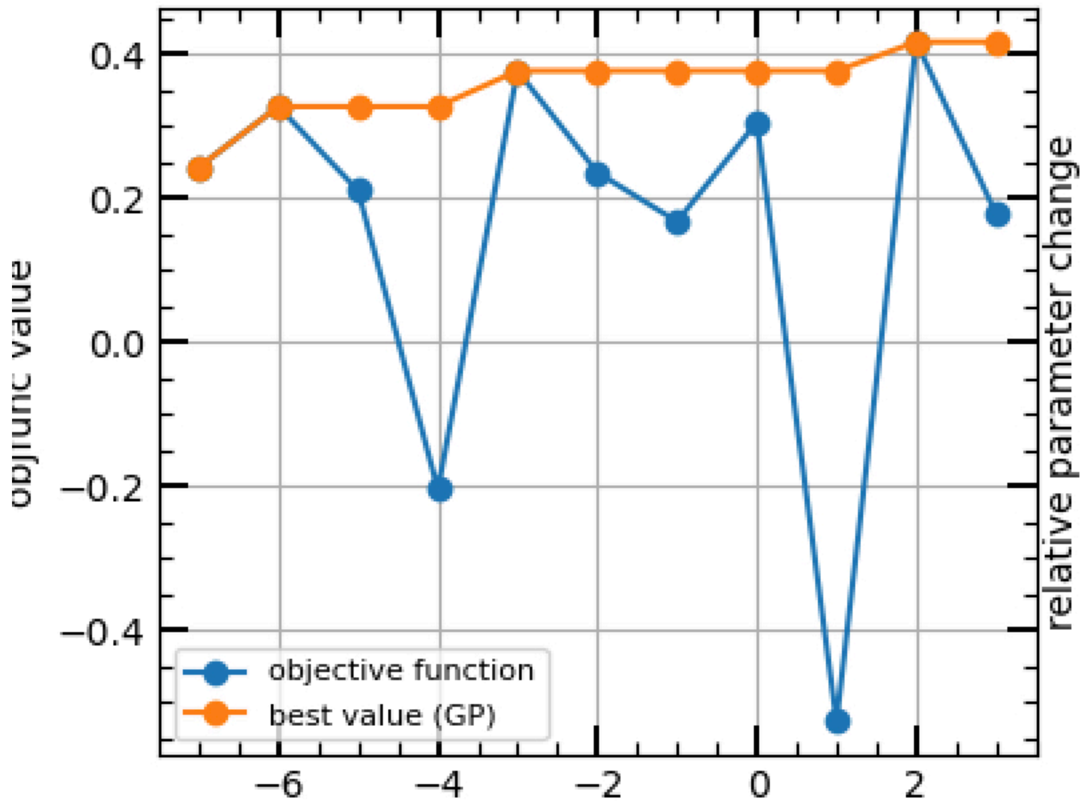

σ h, v vs Epoch



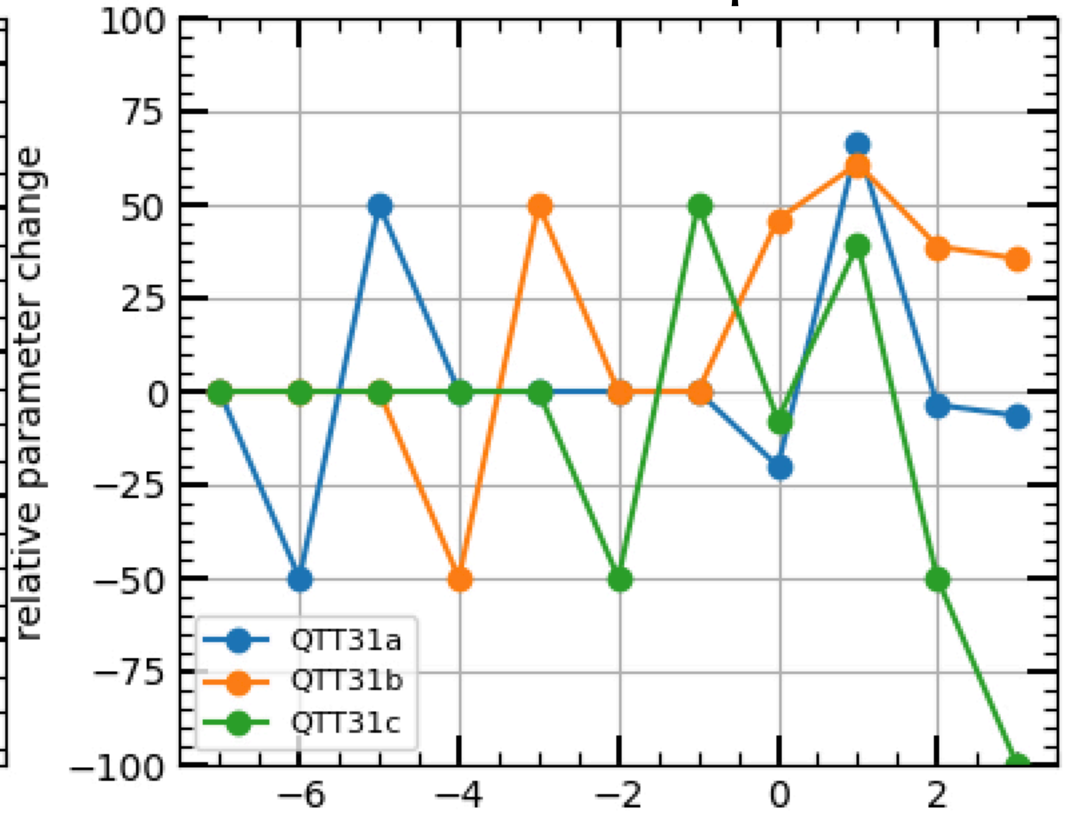

Pos h, v vs Epoch



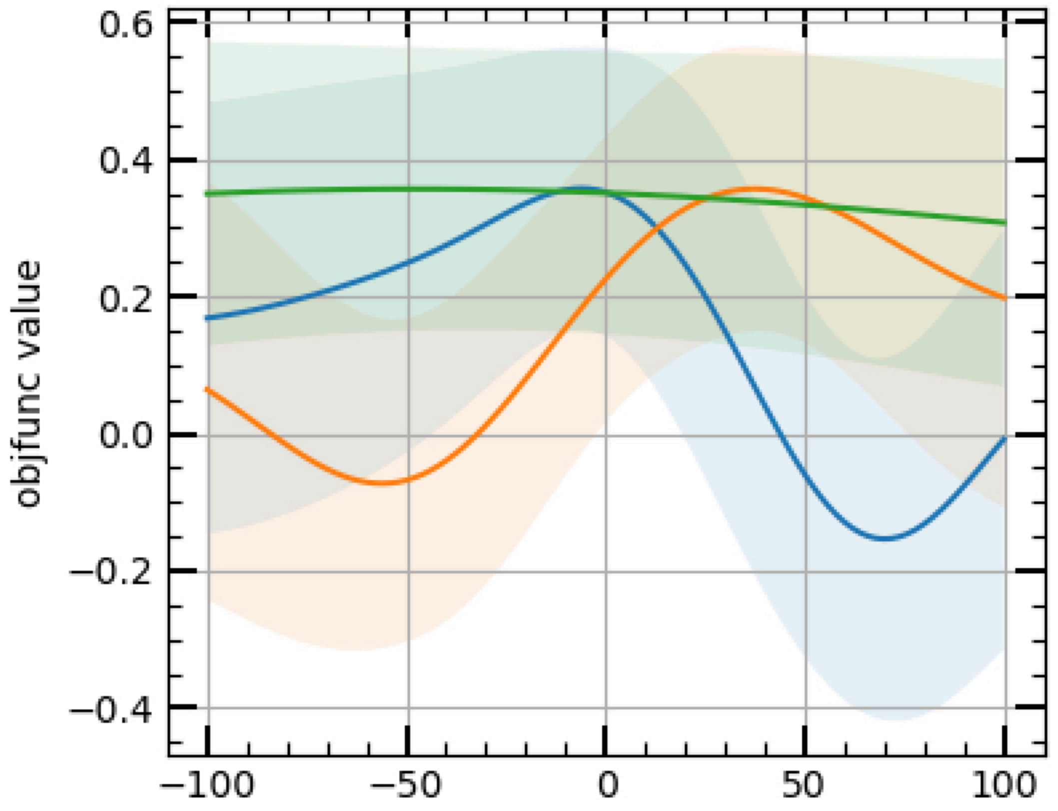

S h, v vs Epoch




Objective function vs Epoch




Current of Q vs Epoch




Objectives vs Current of Q

Initial: manually optimized params.

Goal : good transmission

small beam spot

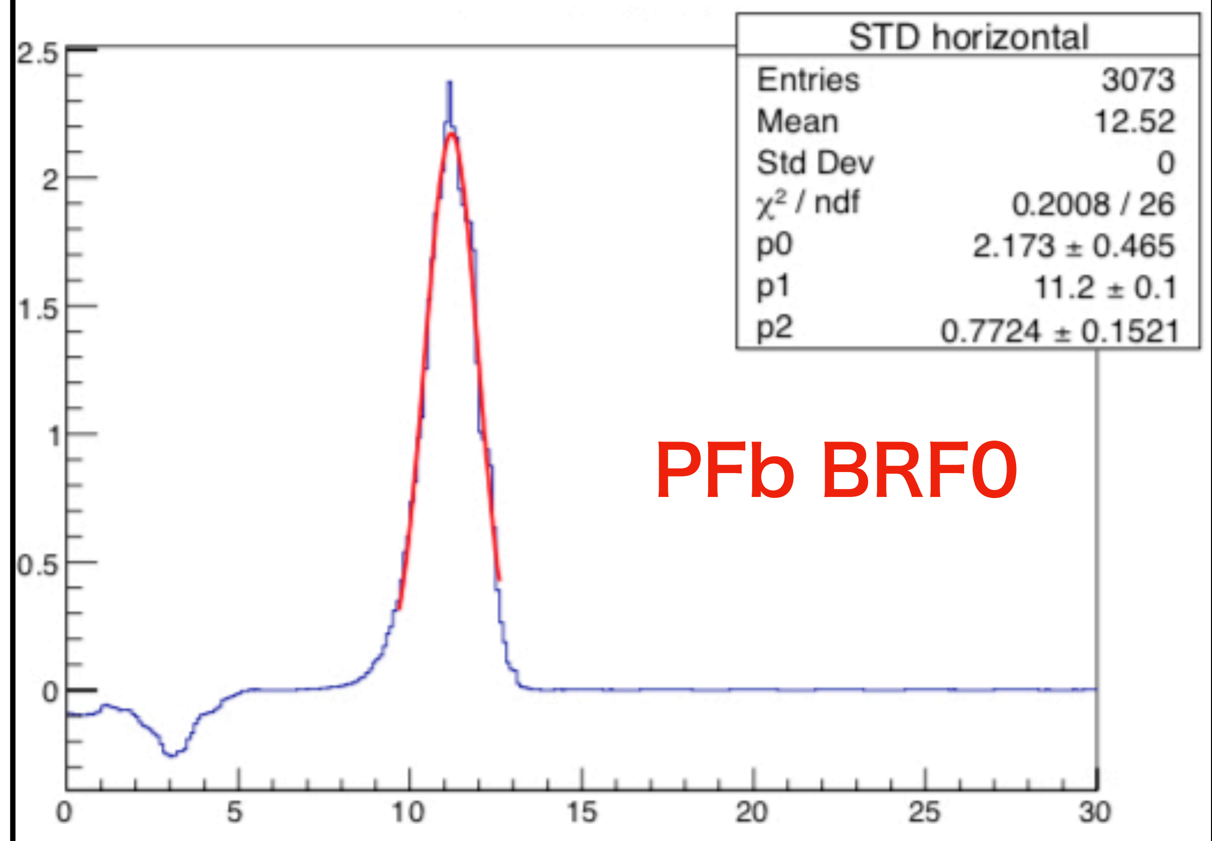
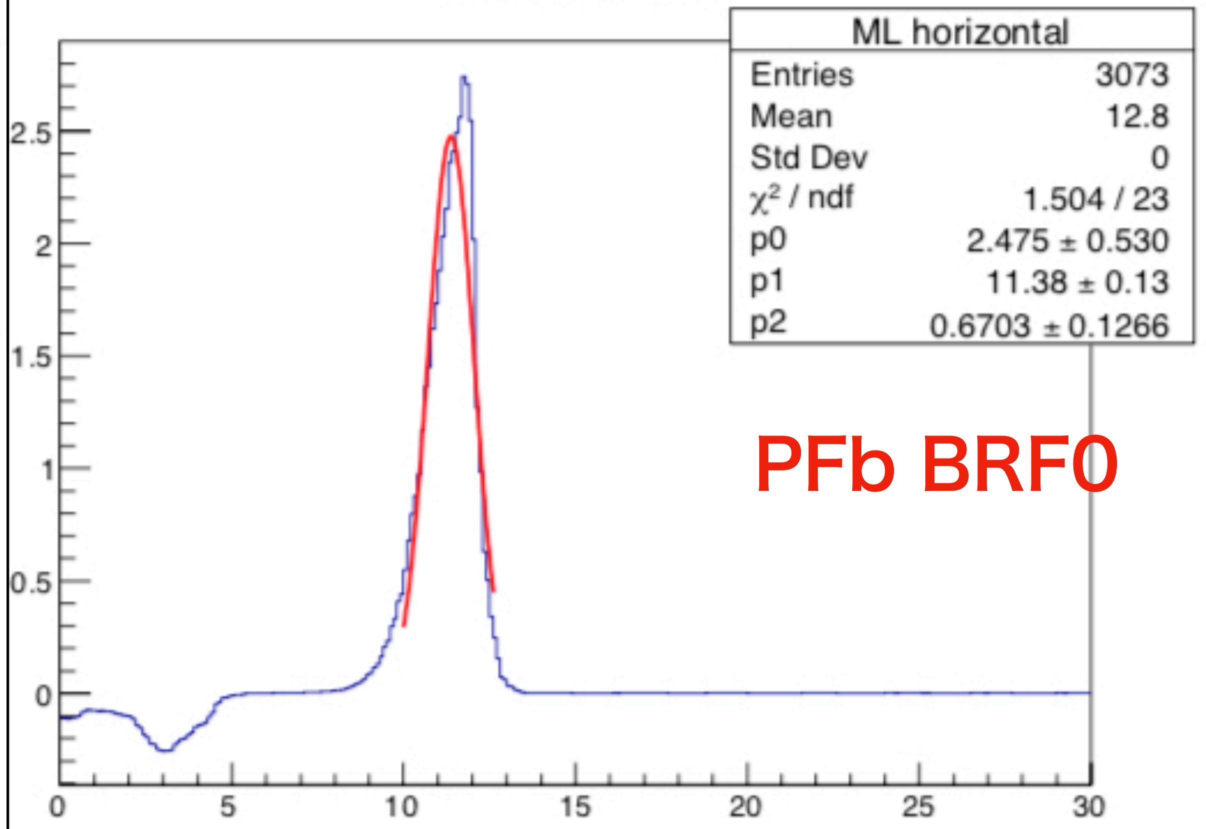
- 3 ~ 7 params. (Quadrupoles)
- 1 epoch ~ 1 s
- 1 try ~ 5 min (~300 epoch)
- ~ 0.001 enA

try optimization with several conditions

# 1st Exp. : Auto Tuning with Low Intensity Beam

Demonstration test at 2020 Oct. 21:00 ~ 9:00

Compare the result of manual optimization / manual + ML optimization
using high intensity beams and wire scanner / Faraday cups

| | Manual Optimization | Manual + ML Optimization |
|---|---|---|
| $FC_{up}$ [eμA] | 7.20 | 7.8 |
| Beam Dump [eμA] | 7.25 | 8.0 |
| Ratio (BD/$FC_{up}$)※ | 1.01 | 1.03 |
| Wire Scanner の像 |  |  |

- **Transmission ↗ 2%**
- **Beam width ↘ 13%**

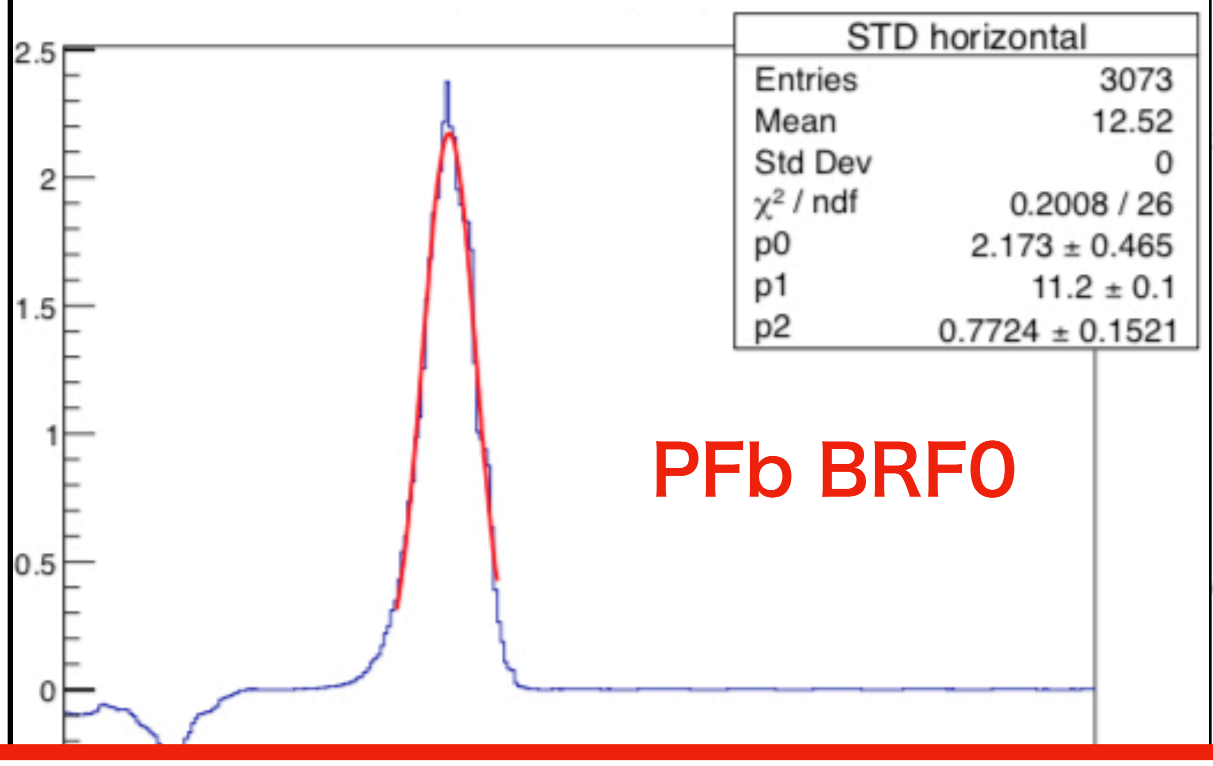
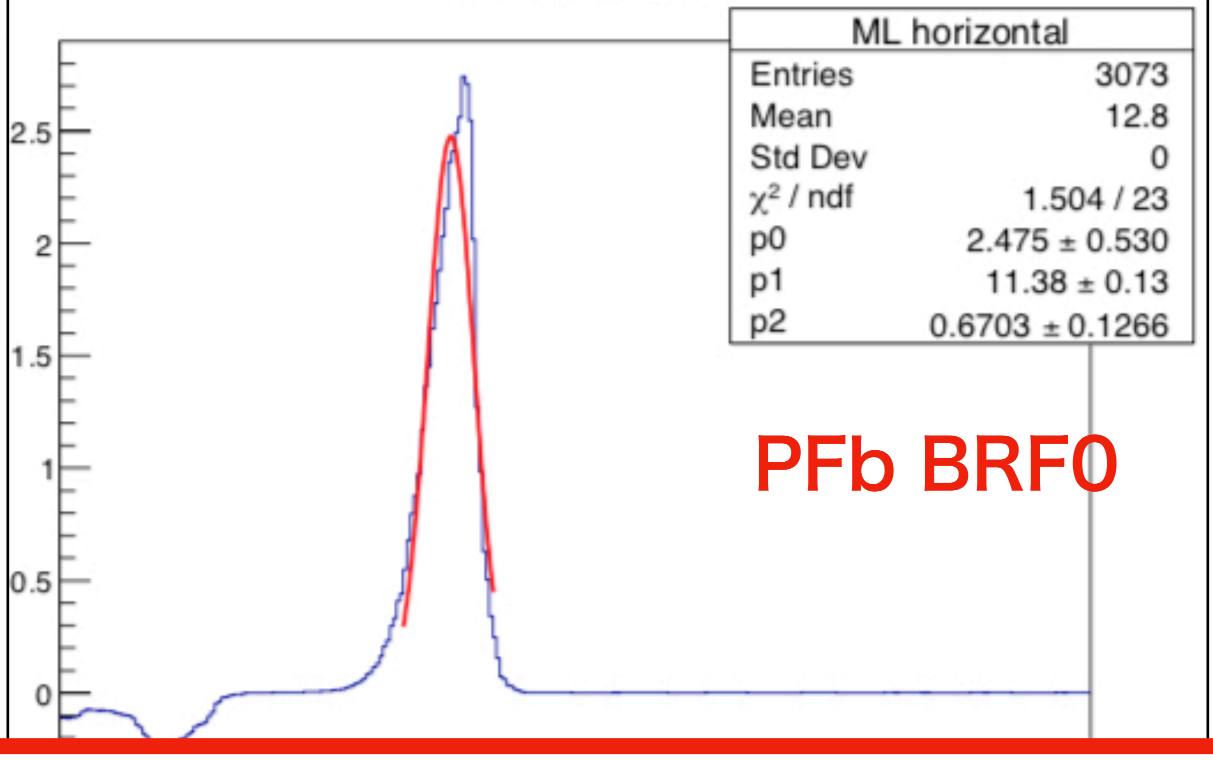
**Significant improvements in the 1st test**

※ Beam dump does not have good suppressor

# 1st Exp. : Auto Tuning with Low Intensity Beam

Demonstration test at 2020 Oct. 21:00 ~ 9:00

Compare the result of manual optimization / manual + ML optimization
using high intensity beams and wire scanner / Faraday cups

| | Manual Optimization | Manual + ML Optimization |
|---|---|---|
| $FC_{up}$ [eμA] | 7.20 | 7.8 |
| Beam Dump [eμA] | 7.25 | 8.0 |
| Ratio (BD/$FC_{up}$)※ | 1.01 | 1.03 |
| Wire Scanner の像 |  |  |

- Transmission ↗ 2%
- Beam width ↘ 13%

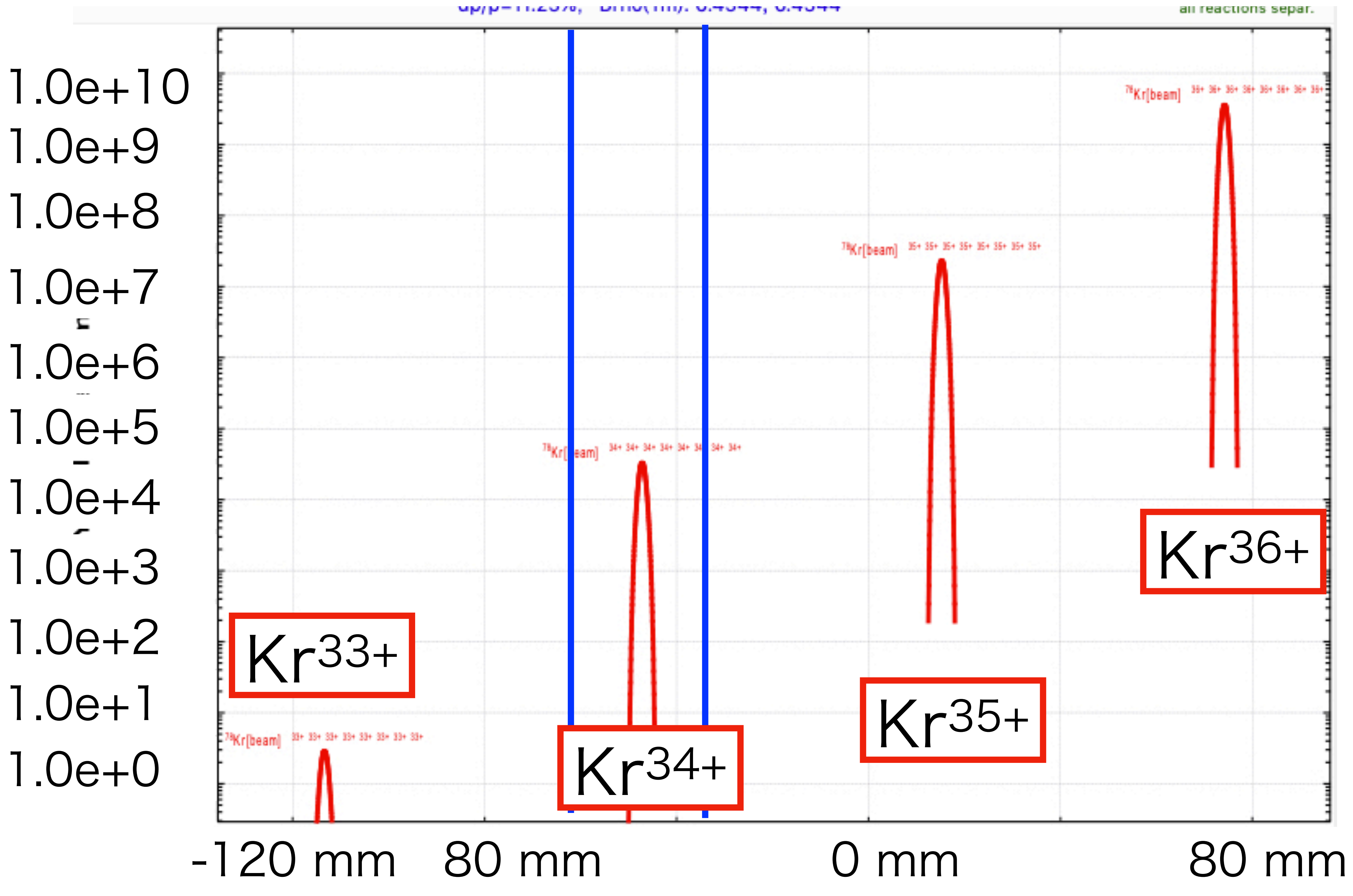
**Significant improvements in the 1st test**

※ Beam dump does not have good suppressor

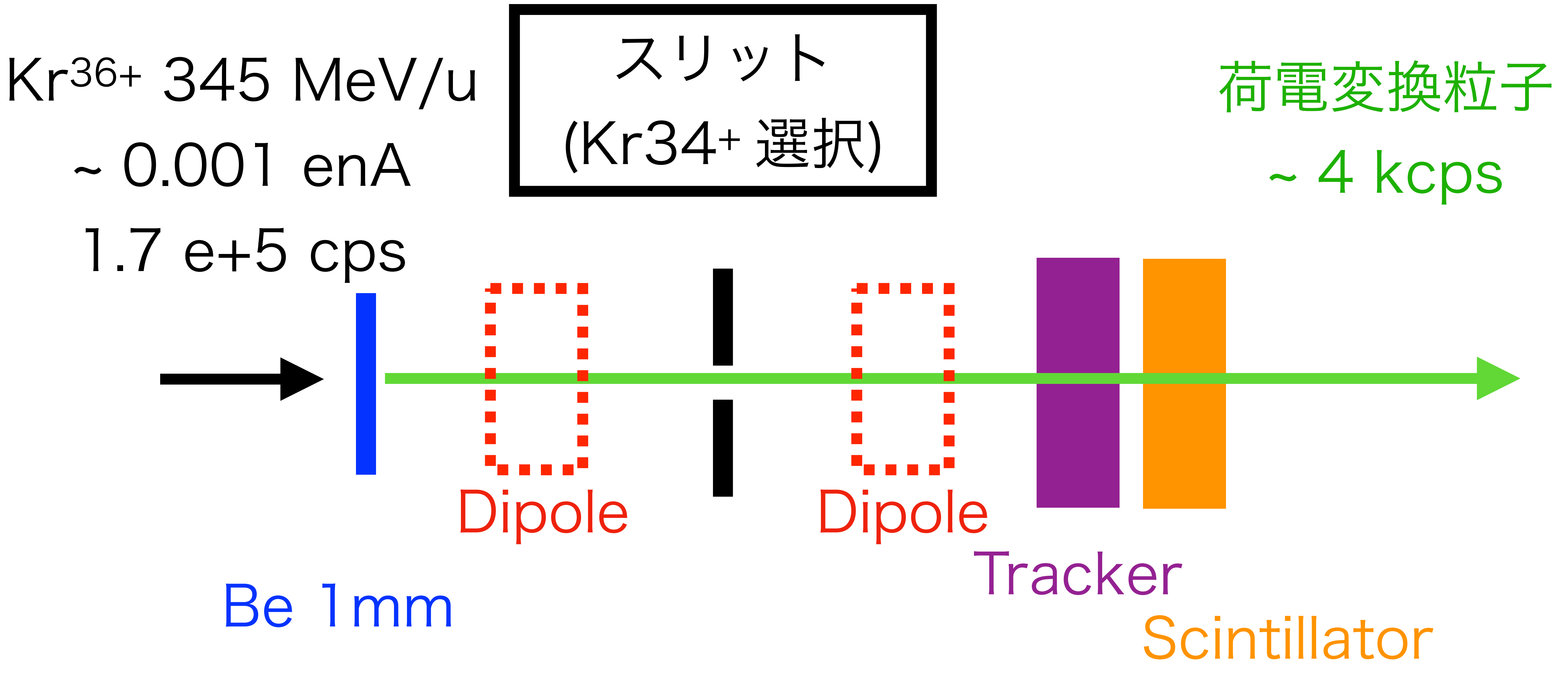
◯ Algorithm works as expected
◯ Optics is improved with low intensity beam (limit by fluorescent viewer: 0.001 enA)
△ Effective algorithm

# “高強度”一次ビームで荷電変換粒子の測定による光学系最適化

一次標的下流**スリット位置**での荷電粒子分布




PPAC (飛跡検出器) / シンチレータ
で荷電変換粒子を測定



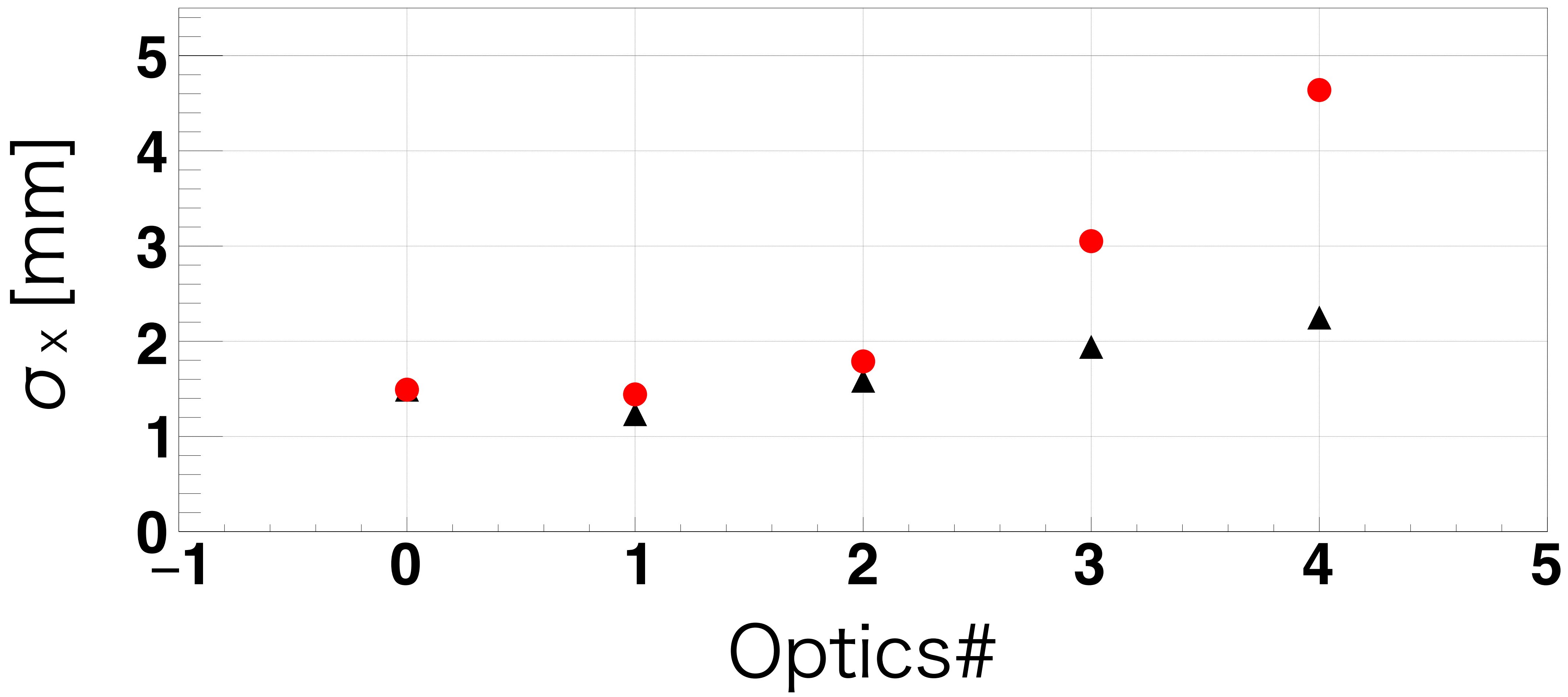

Be 1 mm の標的によって**低強度化**された
荷電変換粒子を下流検出器で測定

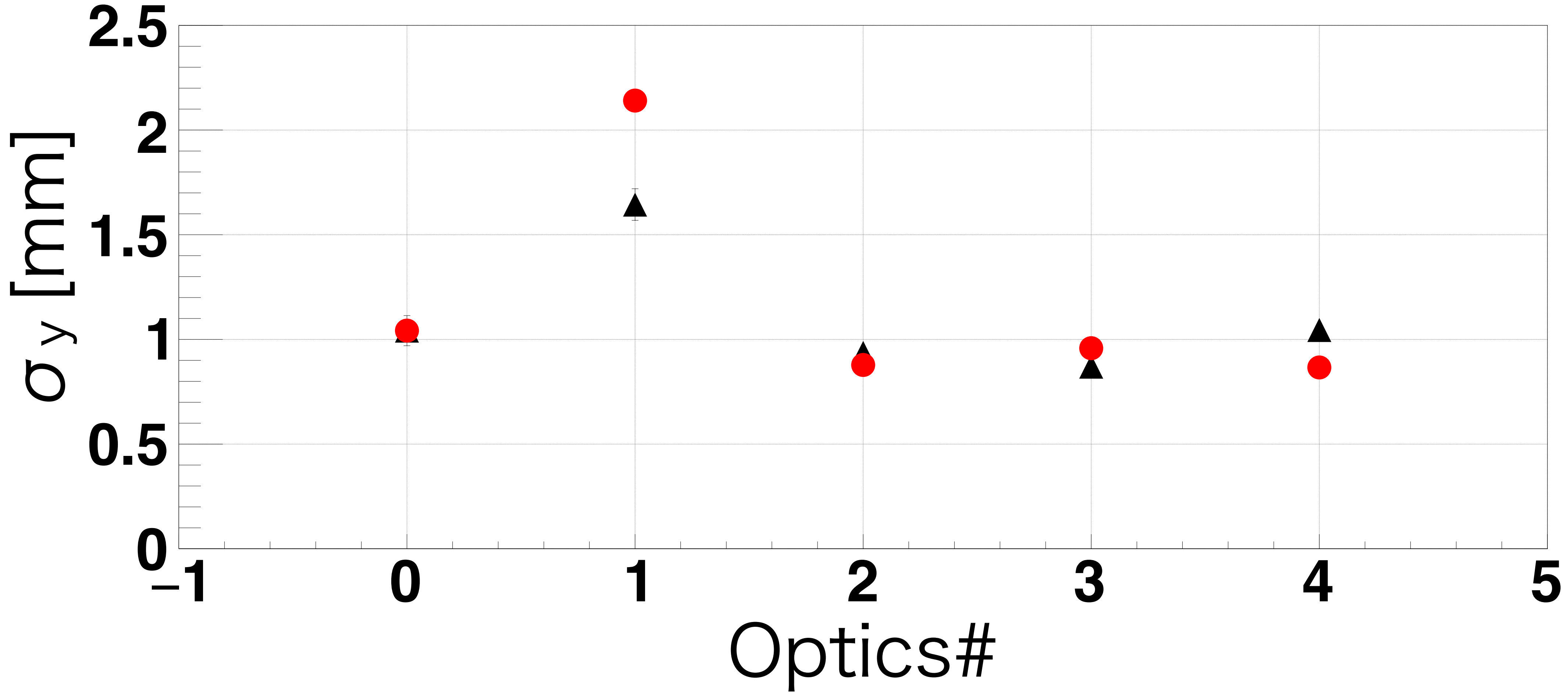
**ビーム強度：標的下流のSci. (荷電変換粒子)**
**ビーム幅　：標的下流の飛跡検出器 (荷電変換粒子)**

| 検出粒子 | 一次ビームの強度 | 下流検出器でのビーム強度 |
|---|---|---|
| $Kr^{35+}$ | 0.001 enA | 1 kcps |
| $Kr^{34+}$ | 1 enA | 2 kcps |
| $Kr^{33+}$ | (300 enA) | 1 kcps |

# A. Compare Beam Spot measured by Viewer and Tracker

Change optics and compare
・fluorescent viewer image of primary beam ($Kr^{36+}$)
・position distribution of secondary beam ($Kr^{34+}$) tracked by PPAC (gas detector)




**Red : Viewer**
**Black: Gas tracker**
**※ Viewer and gas tracker is calibrated by optics 0 data.**

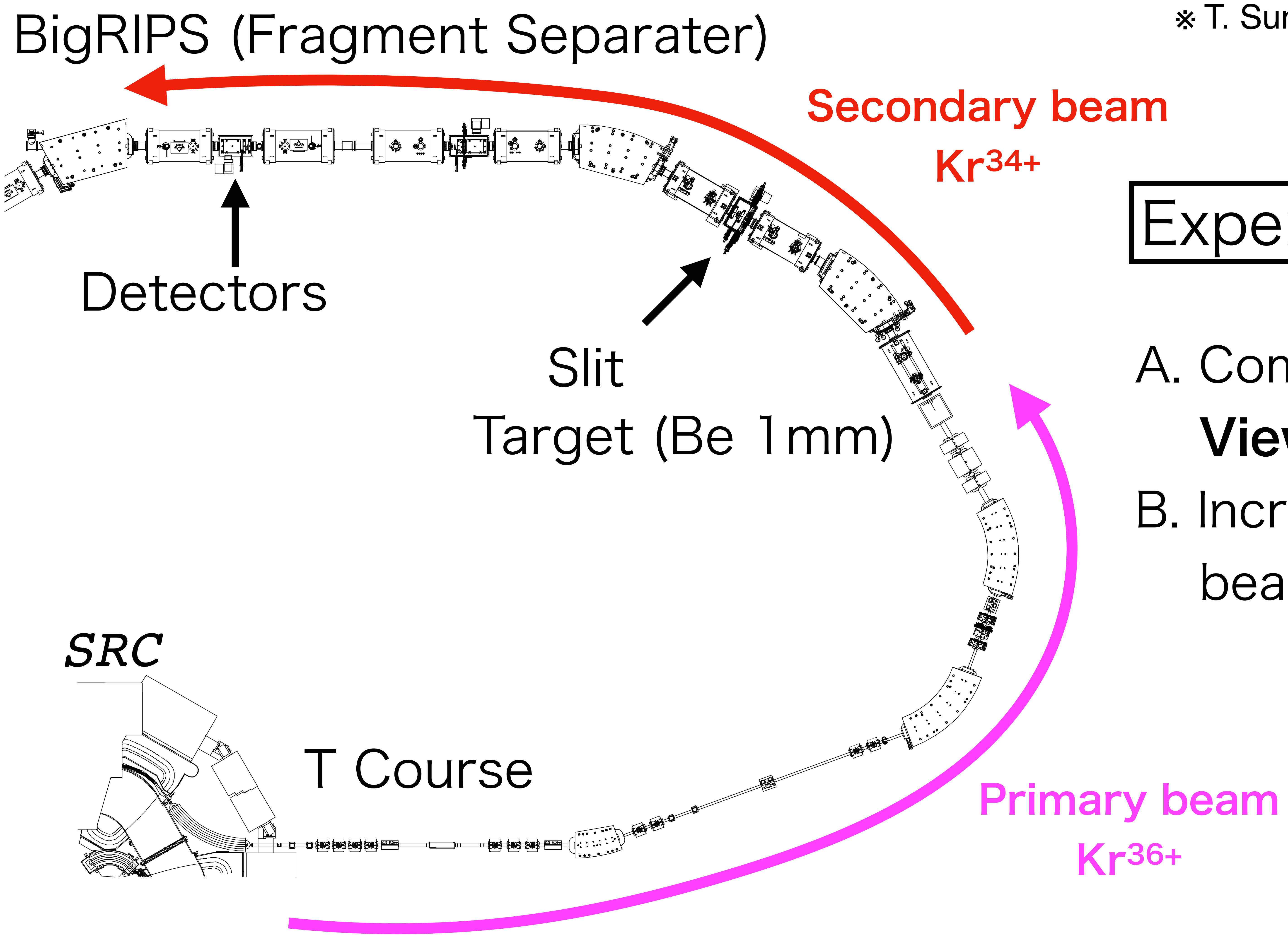
Data of viewer and tracker
→ (qualitatively) consistent
※ When the spot become wide,
non linearity may not be negligible

**Tracking distribution of secondary beam is good probe for the primary beam spot!!**

# Auto Tuning with "High Intensity" Beam in 2022 May

Real time tracker system is realized and connected to EPICS※
→ test with ideal situation with Be target and tracker / scintillator

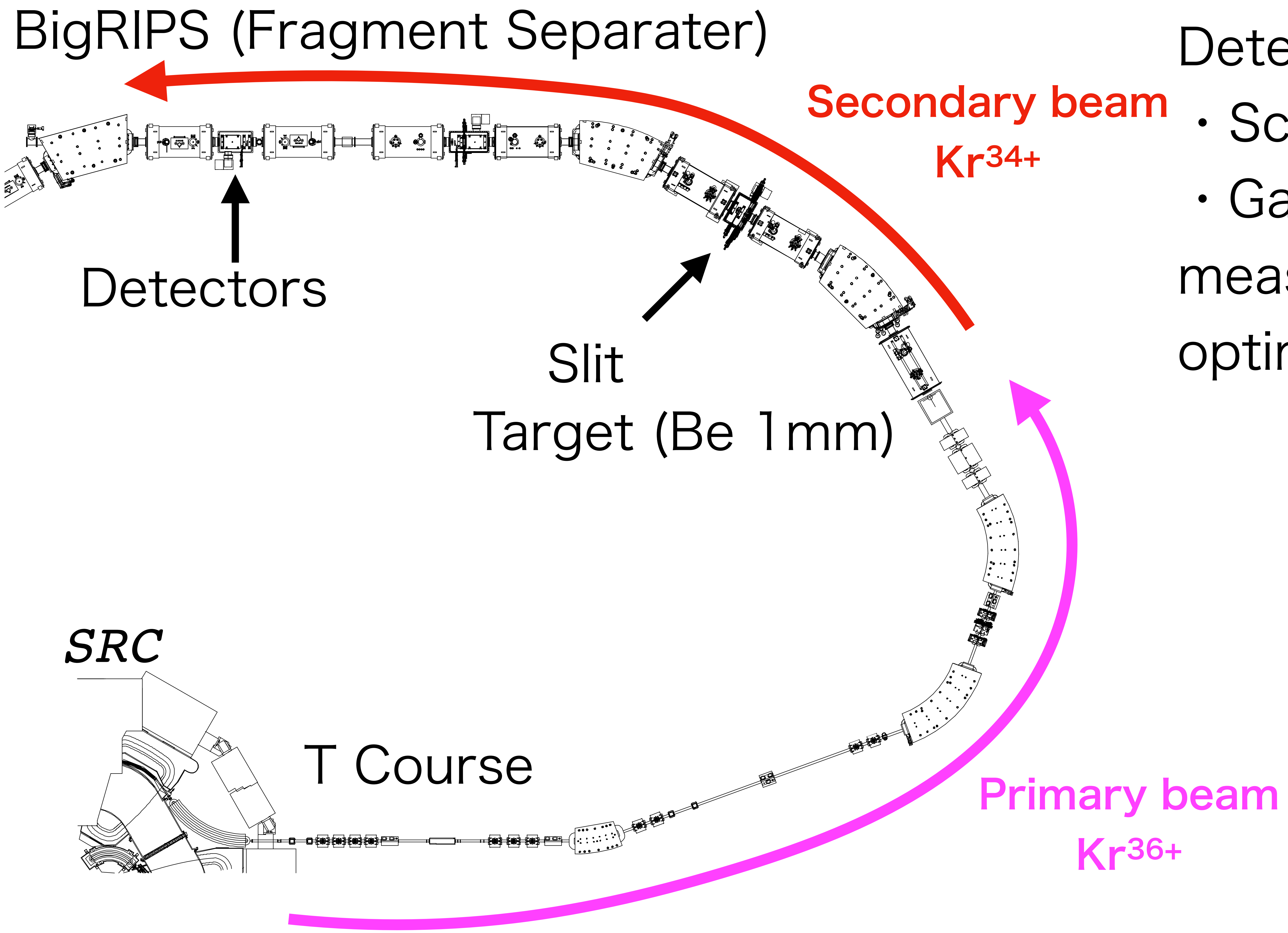
※ T. Sumikama *et al*, RIKEN Accel. Prog. Rep 54, 82 (2021)




Experiment items

A. Compare beam spot measured by **Viewer** and **Tracker**
B. Increase the beam intensity and optimize beam optics using Tracker / Scintillator

# B. Auto Tuning with “High Intensity“ Primary Beam




Detector in downstream

・Scintillator (for Transmission)

・Gas tracker (for spot)

measure $Kr^{34+}$ **10 kcps** / 0.0001 enA

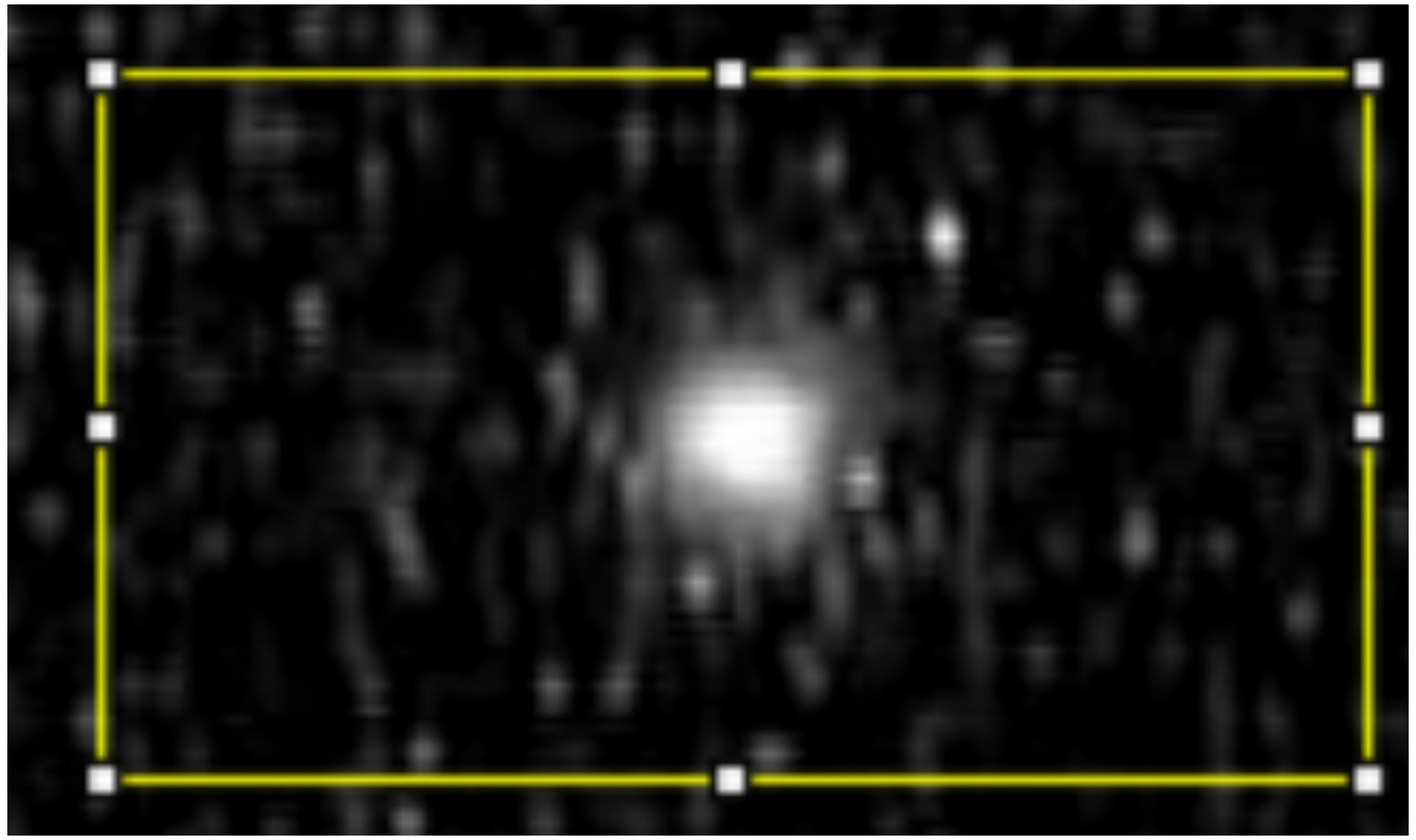
optimize **26 enA** primary beam $Kr^{36+}$

# A. Compare Beam Spot measured by Viewer and Tracker

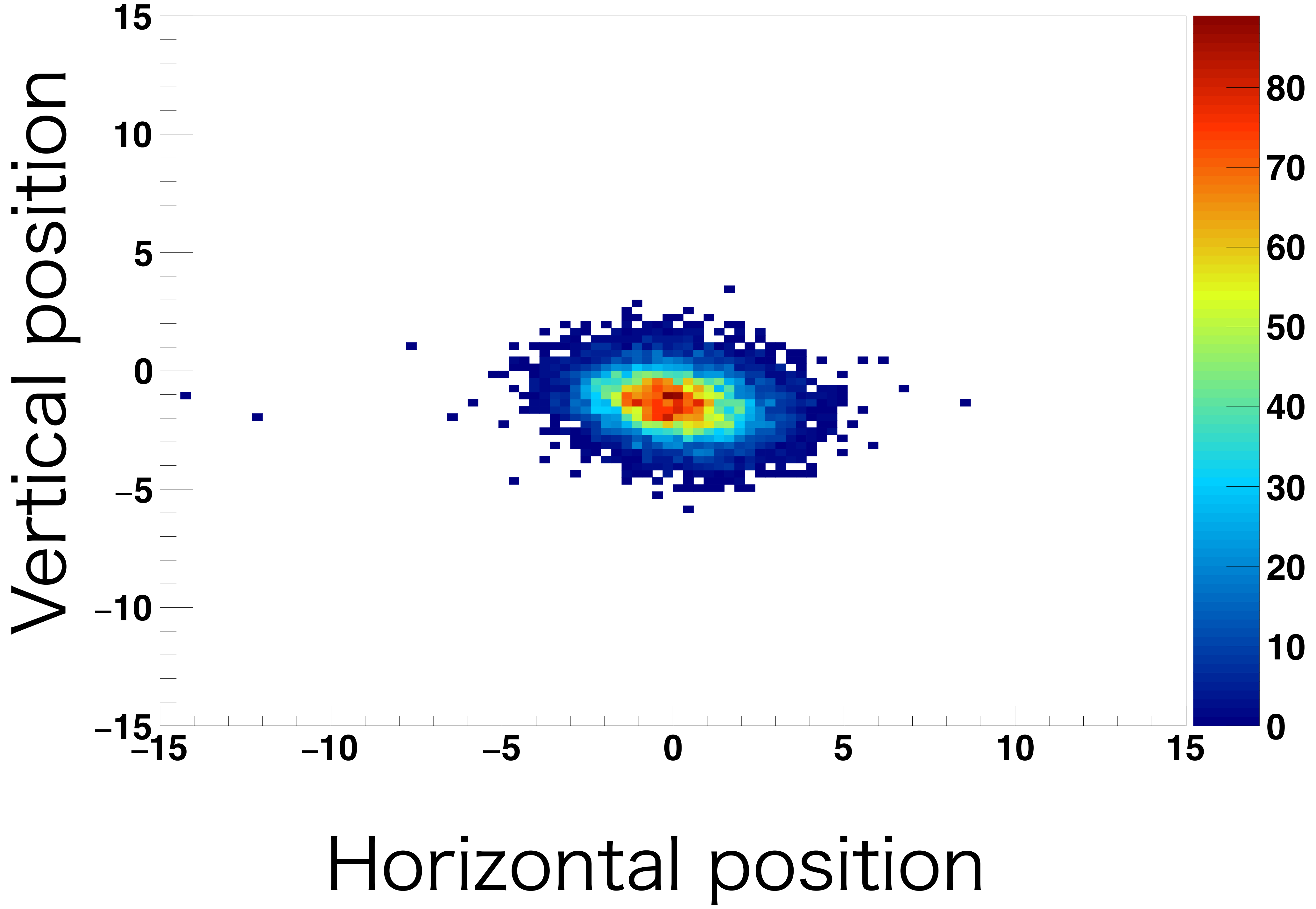
Change optics and compare

- fluorescent viewer image of primary beam ($Kr^{36+}$)
- position distribution of secondary beam ($Kr^{34+}$) tracked by PPAC (gas detector)

Fluorescent viewer ($Kr^{36+}$)



Gas tracker ($Kr^{34+}$)




※ In this measurement, Fluorescent viewer was removed and Be 1mm was inserted.

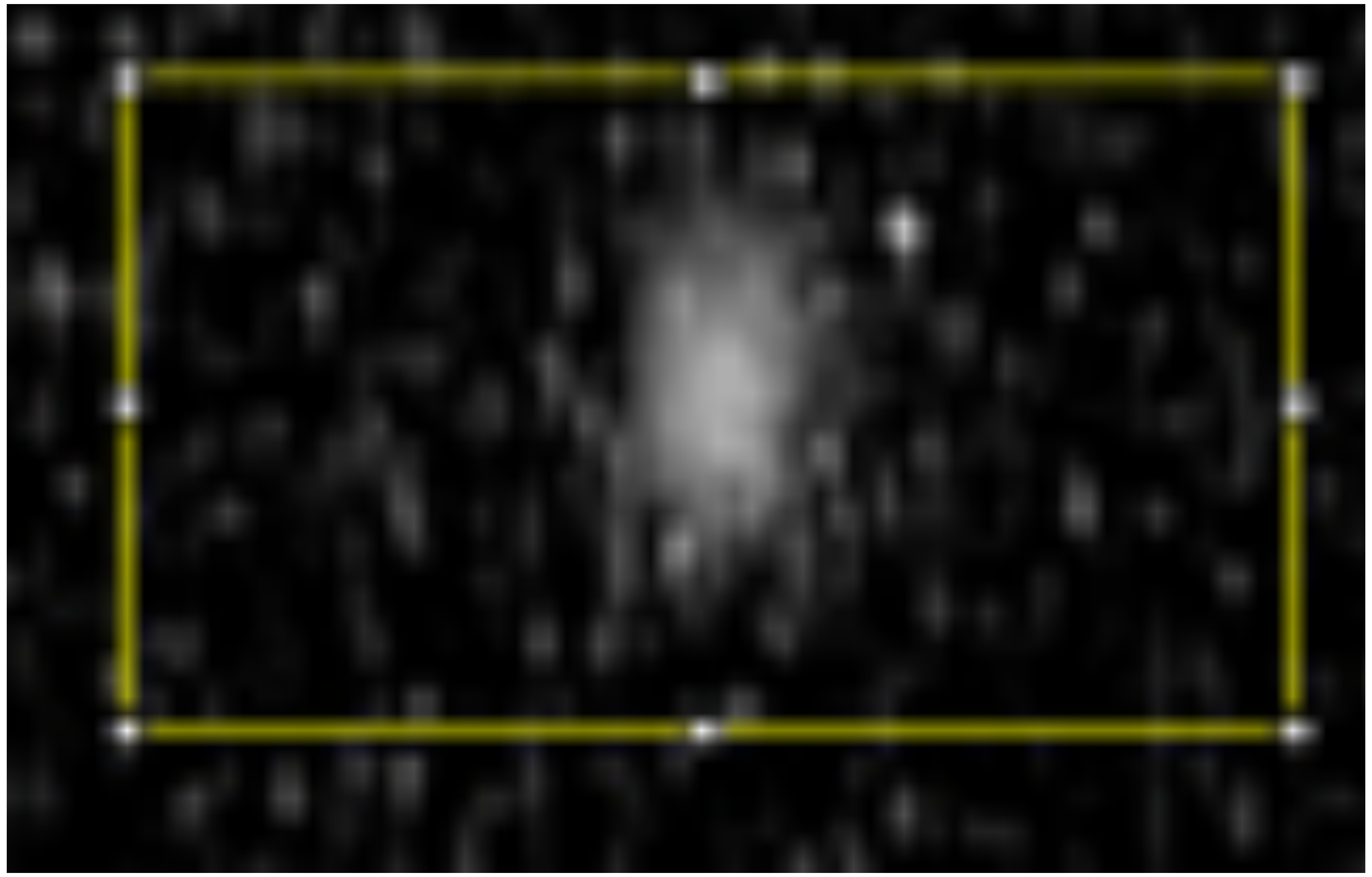
# A. Compare Beam Spot measured by Viewer and Tracker

Change optics and compare

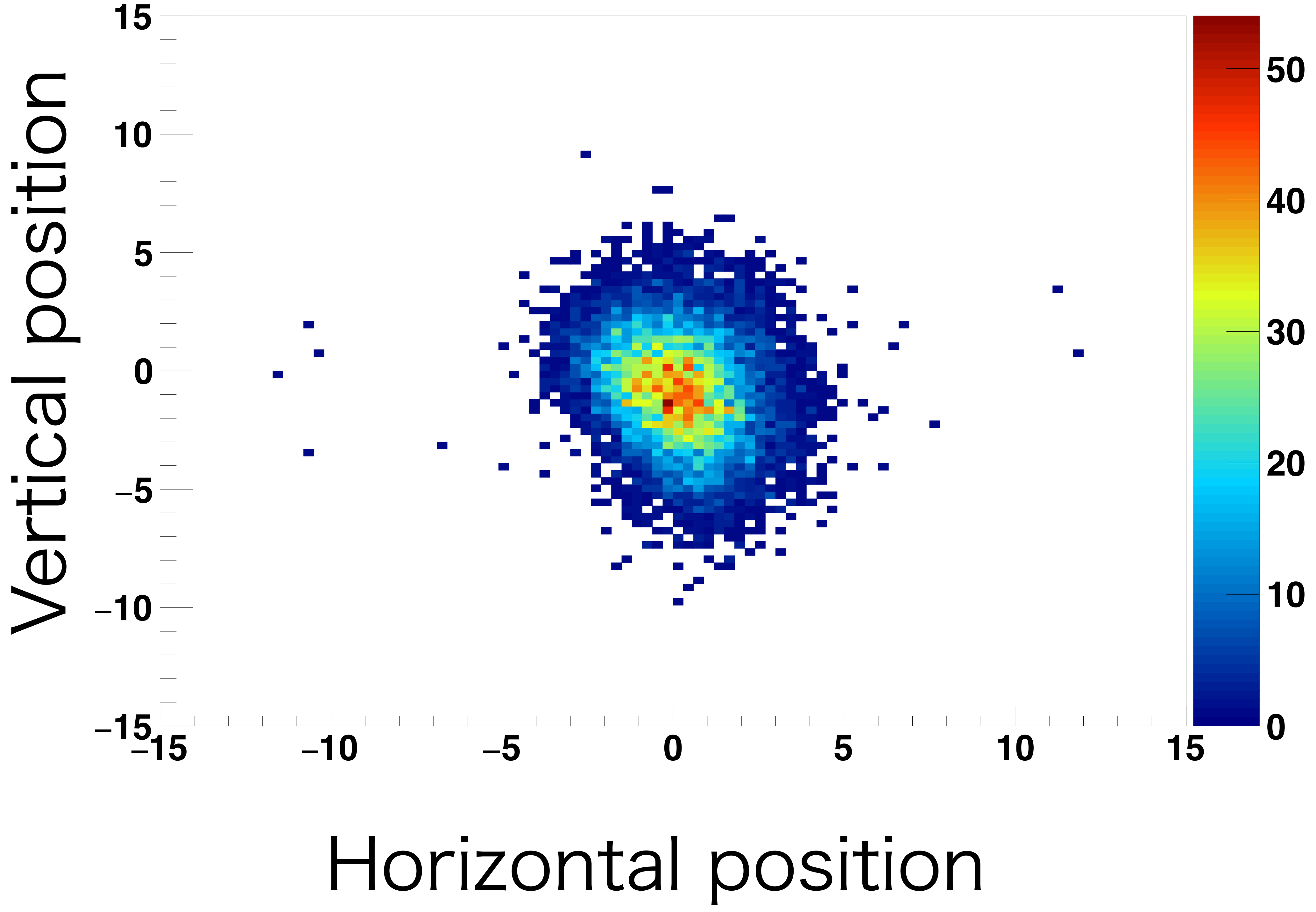
・fluorescent viewer image of primary beam ($Kr^{36+}$)

・position distribution of secondary beam ($Kr^{34+}$) tracked by PPAC (gas detector)

Fluorescent viewer ($Kr^{36+}$)



Gas tracker ($Kr^{34+}$)

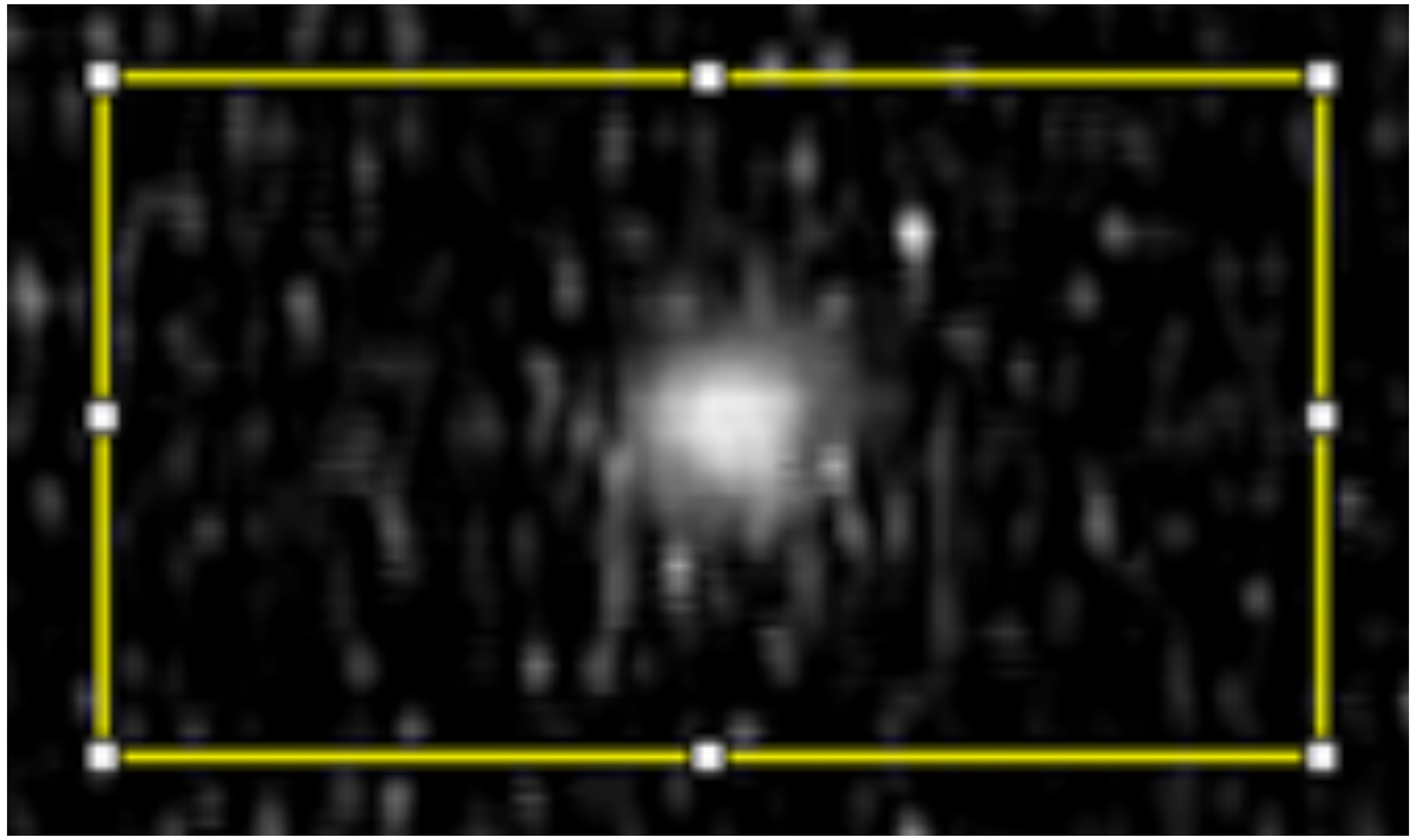



※ In this measurement, Fluorescent viewer was removed and Be 1mm was inserted.

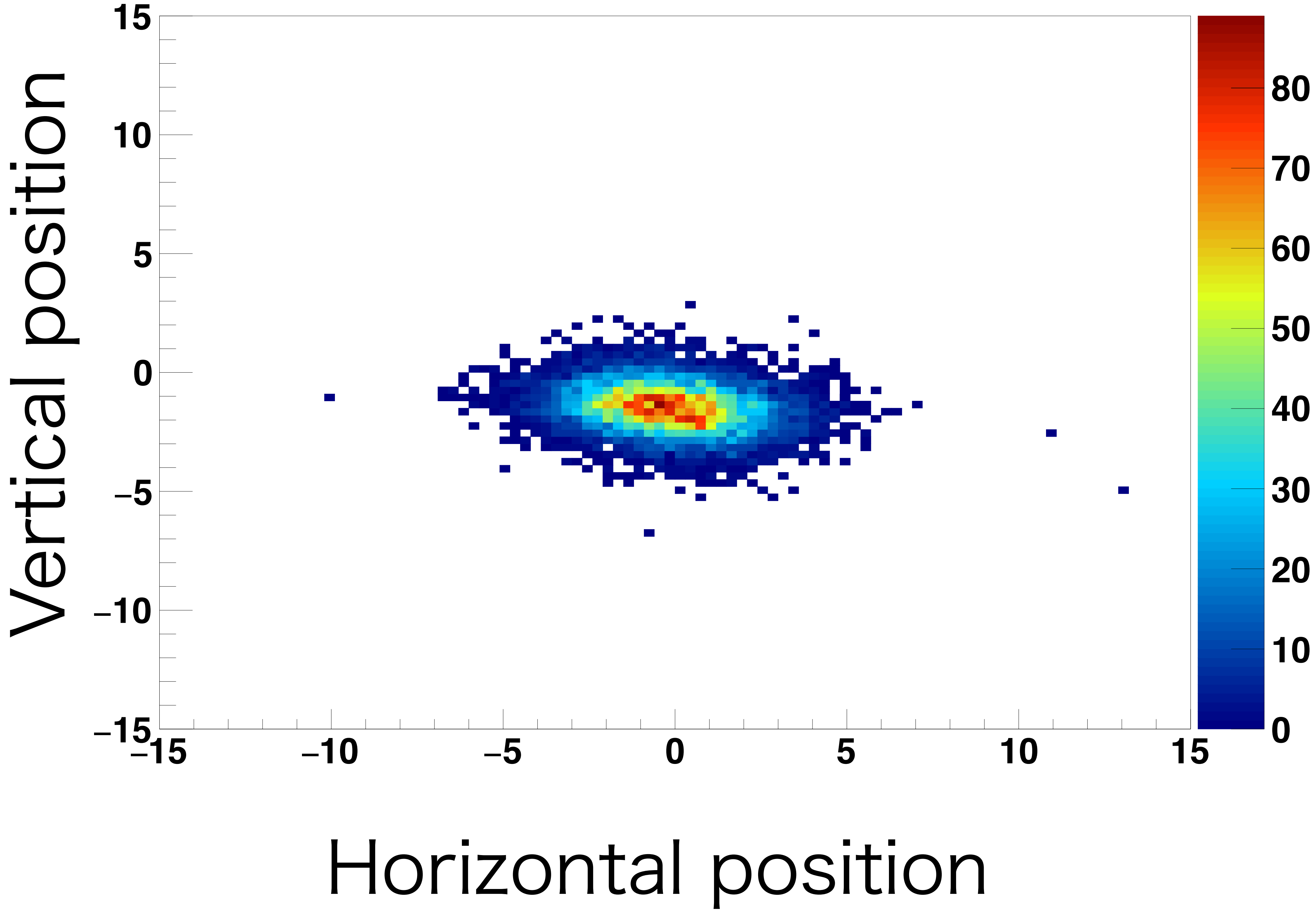
# A. Compare Beam Spot measured by Viewer and Tracker

Change optics and compare
・fluorescent viewer image of primary beam ($Kr^{36+}$)
・position distribution of secondary beam ($Kr^{34+}$) tracked by PPAC (gas detector)

Fluorescent viewer ($Kr^{36+}$)



Gas tracker ($Kr^{34+}$)




※ In this measurement, Fluorescent viewer was removed and Be 1mm was inserted.

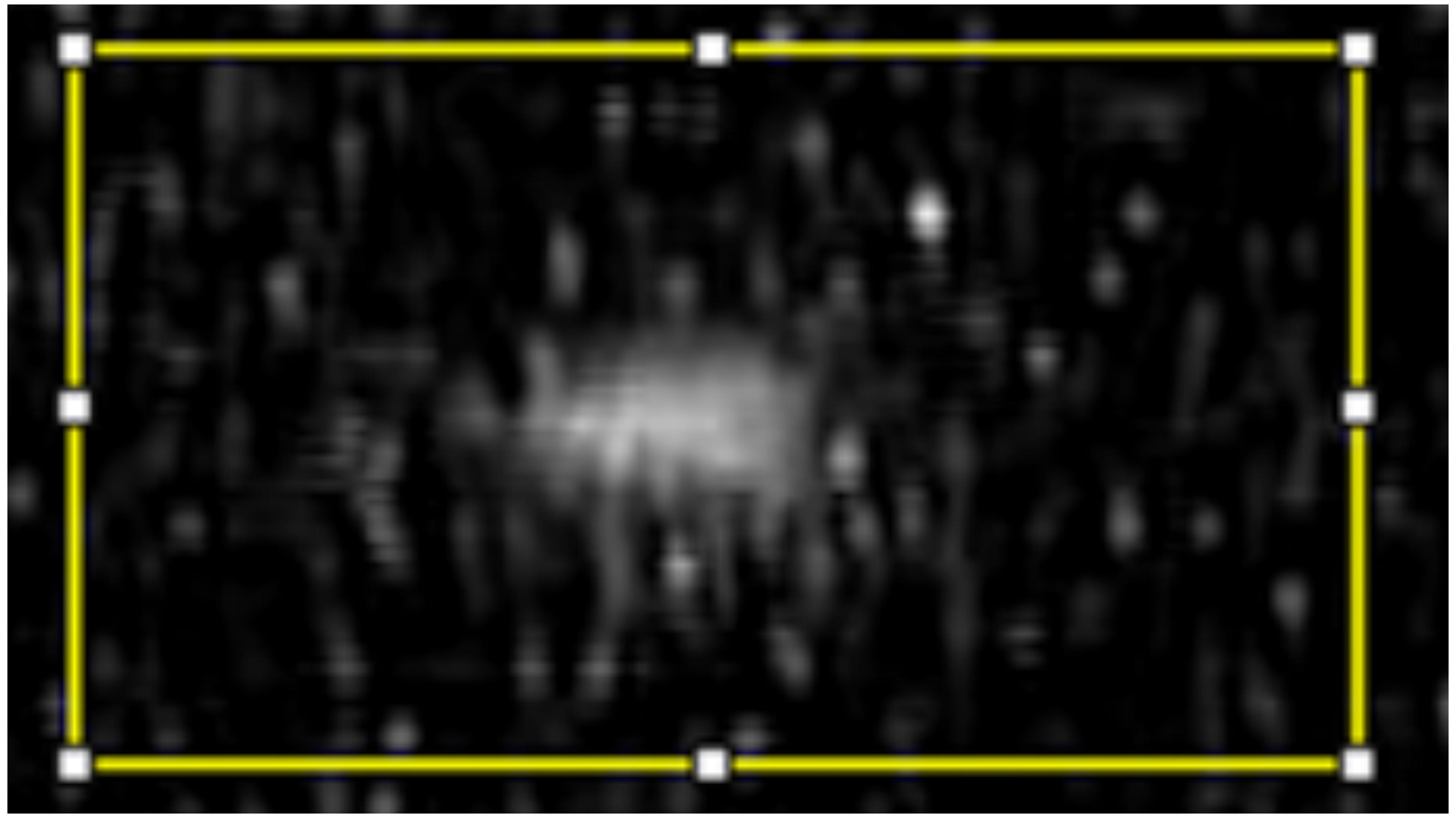
# A. Compare Beam Spot measured by Viewer and Tracker

Change optics and compare

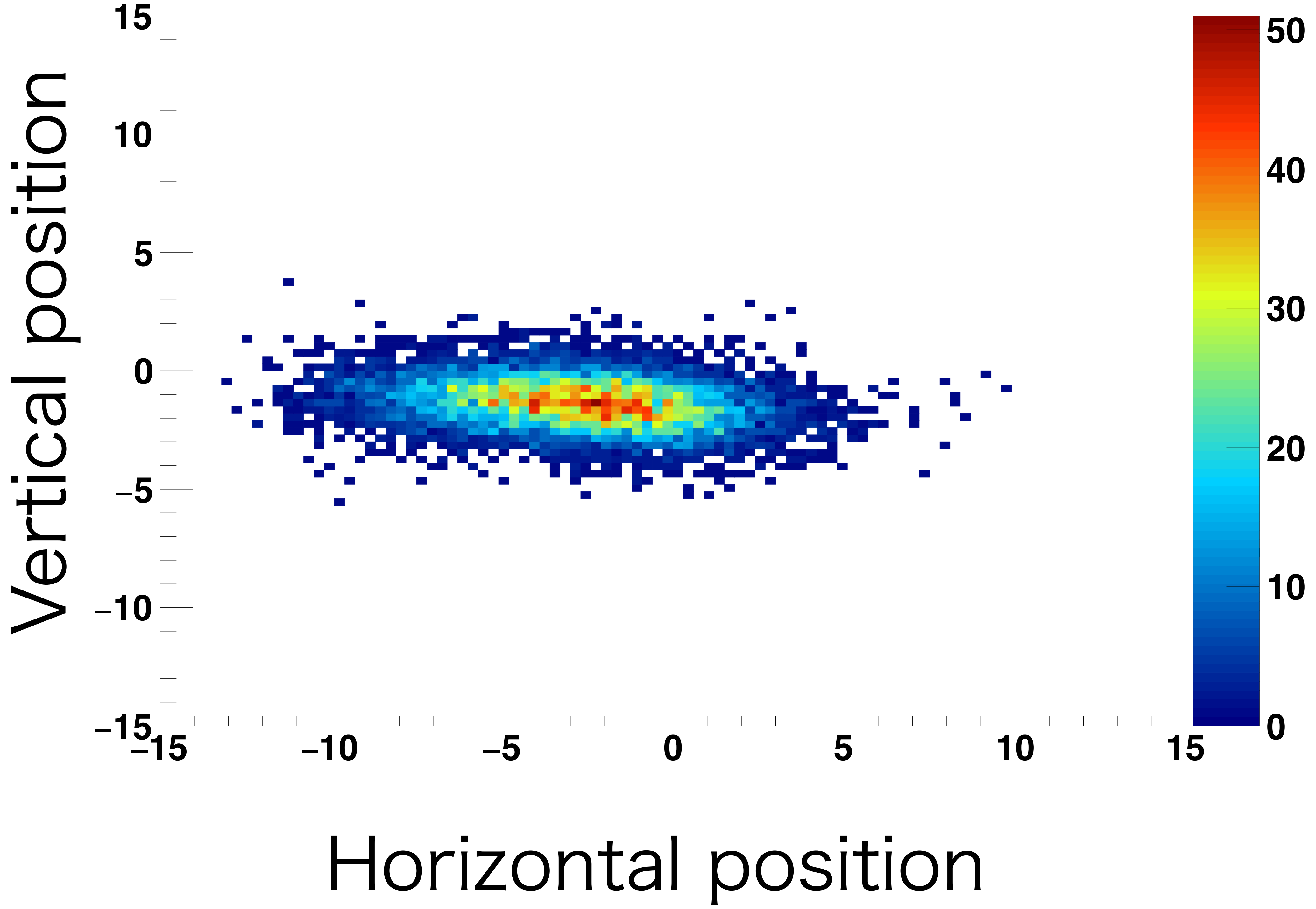
・fluorescent viewer image of primary beam ($Kr^{36+}$)

・position distribution of secondary beam ($Kr^{34+}$) tracked by PPAC (gas detector)

Fluorescent viewer ($Kr^{36+}$)



Gas tracker ($Kr^{34+}$)

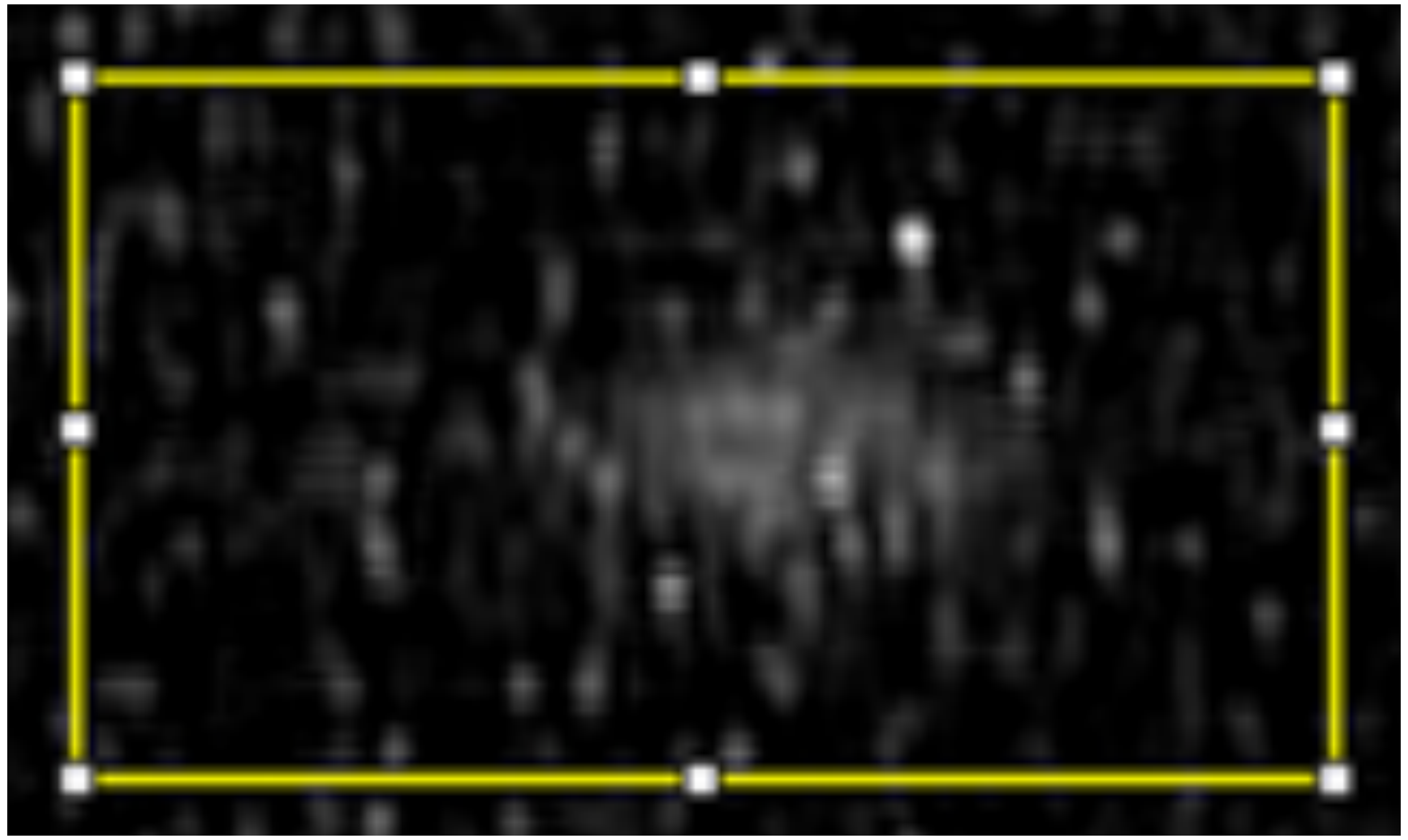



※ In this measurement, Fluorescent viewer was removed and Be 1mm was inserted.

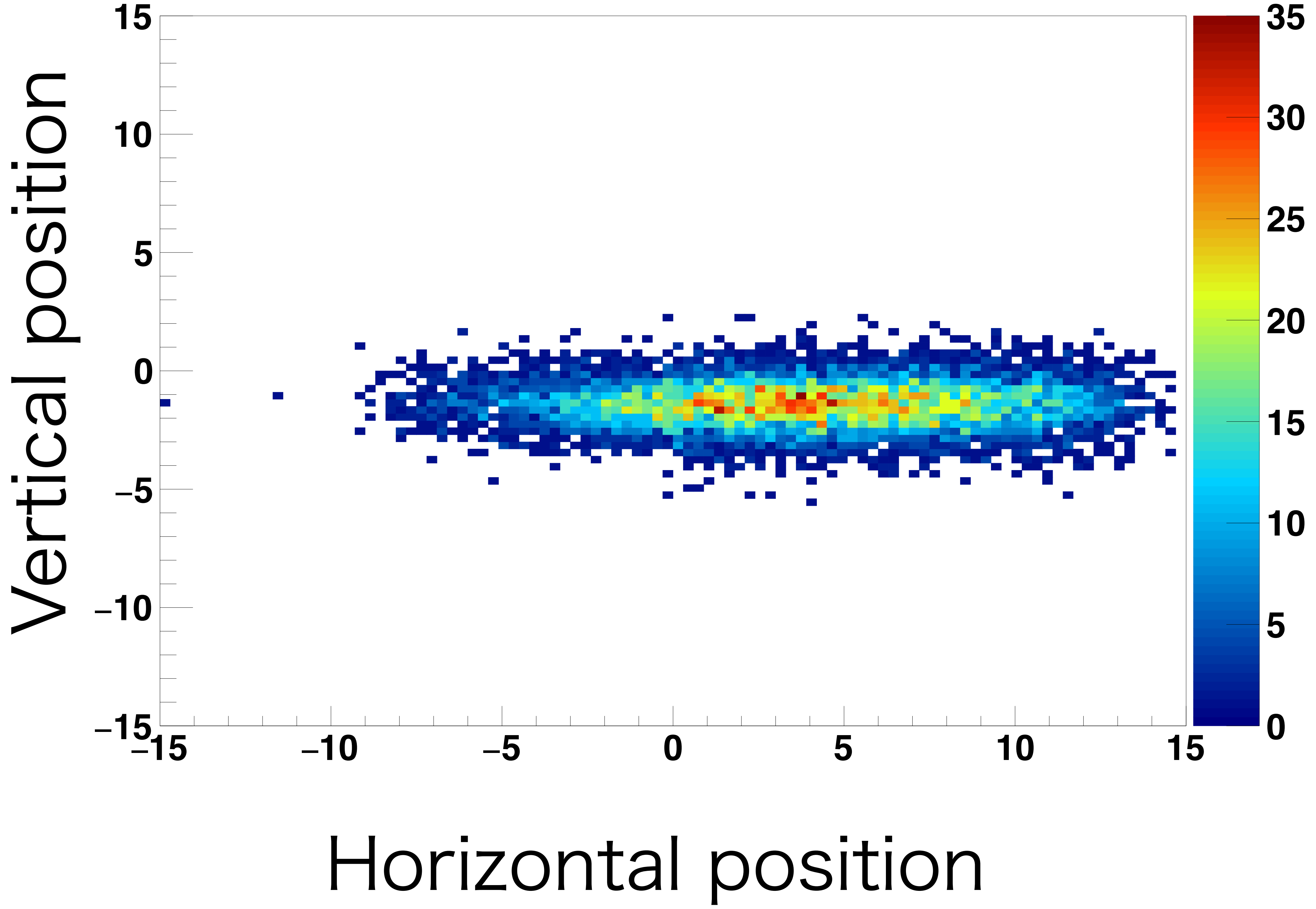
# A. Compare Beam Spot measured by Viewer and Tracker

Change optics and compare
・fluorescent viewer image of primary beam ($Kr^{36+}$)
・position distribution of secondary beam ($Kr^{34+}$) tracked by PPAC (gas detector)

Fluorescent viewer ($Kr^{36+}$)



Gas tracker ($Kr^{34+}$)

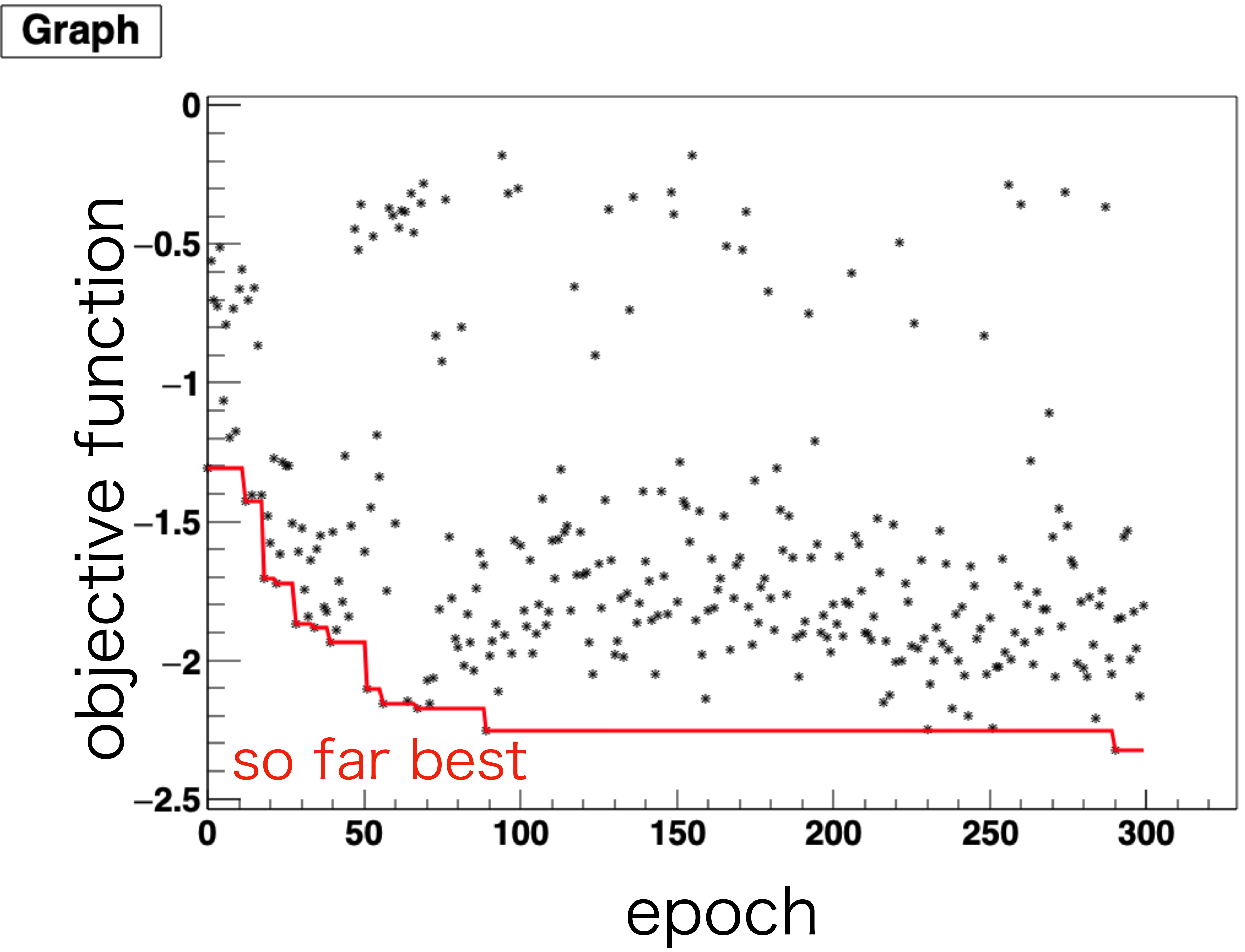


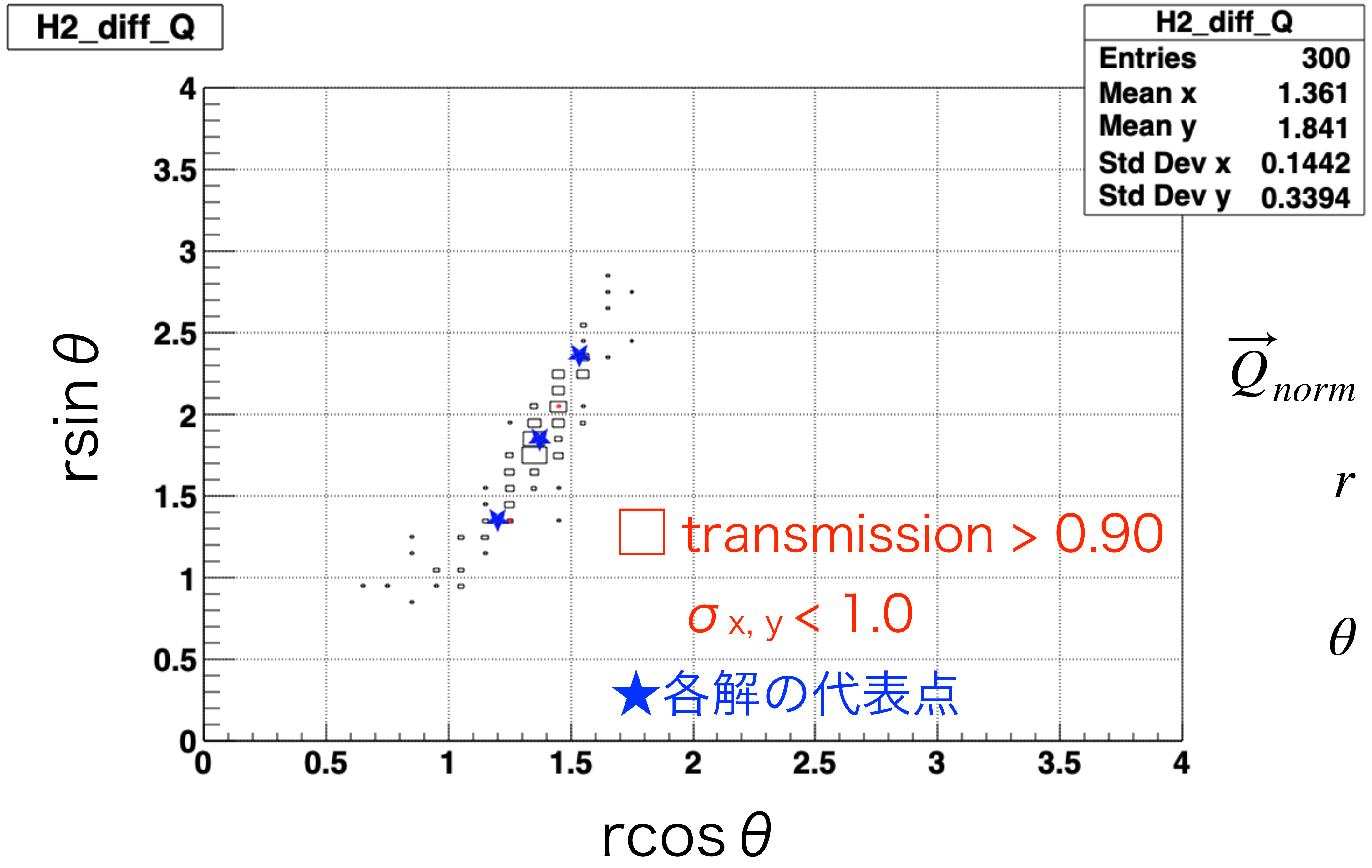

※ In this measurement, Fluorescent viewer was removed and Be 1mm was inserted.

# 各 ハイパーパラメータでのシミュレーション結果の比較

LCB における β の違い (objective function: $f_{sum}$/ phase ellipse C)

Lower Confidence Bound (LCB) / β = 100.0




$$\vec{Q}_{norm} \equiv \{Q_i/Q_i^0\}$$

$$r \equiv |\vec{Q}_{norm}|/|\vec{Q^0}_{norm}|$$

$$\theta \equiv \arccos \frac{\vec{Q}_{norm} \cdot \vec{Q^0}_{norm}}{|\vec{Q}_{norm}||\vec{Q^0}_{norm}|}$$

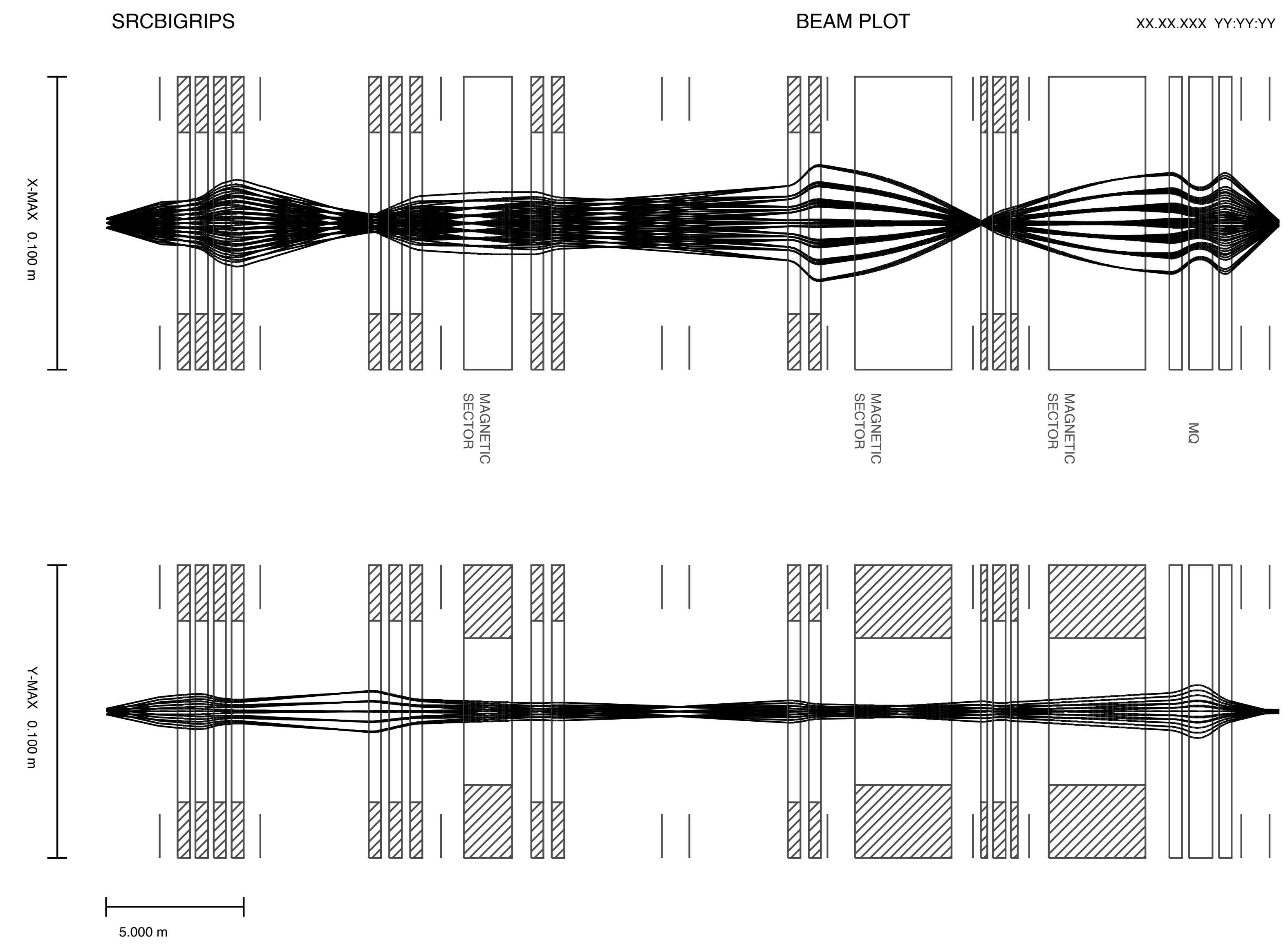
| | epoch | obj. | transmission | σ x | σ y |
|---|---|---|---|---|---|
| 最初に閾値を超えたもの | 51 | -2.46 | 0.966 | 0.59 | 0.92 |
| best | 290 | -2.55 | 0.960 | 1.05 | 0.31 |

# 各 ハイパーパラメータでのシミュレーション結果の比較

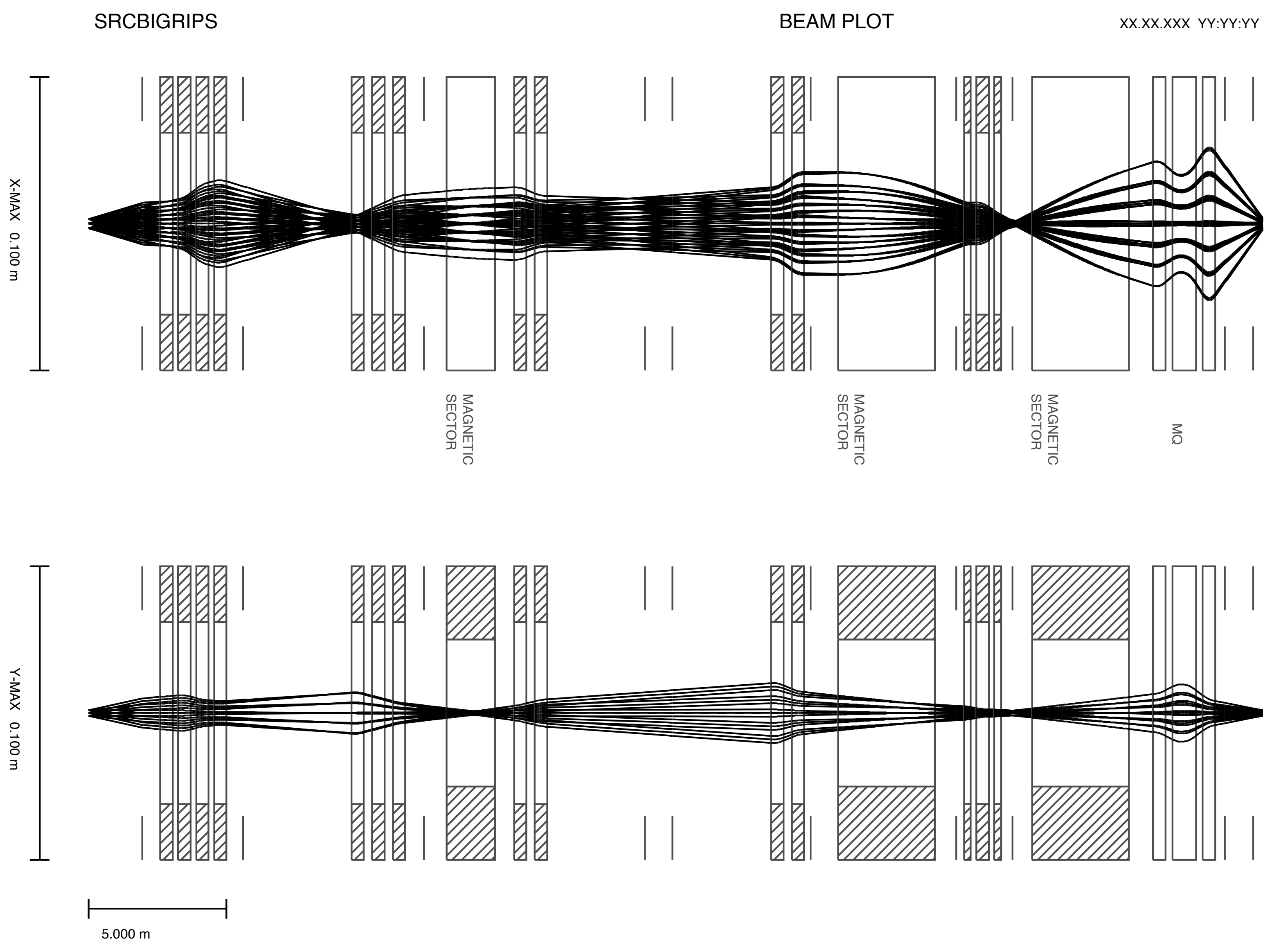
LCB における $\beta$ の違い (objective function: $f_{sum}$/ phase ellipse C)

Lower Confidence Bound (LCB) / $\beta$ = 100.0

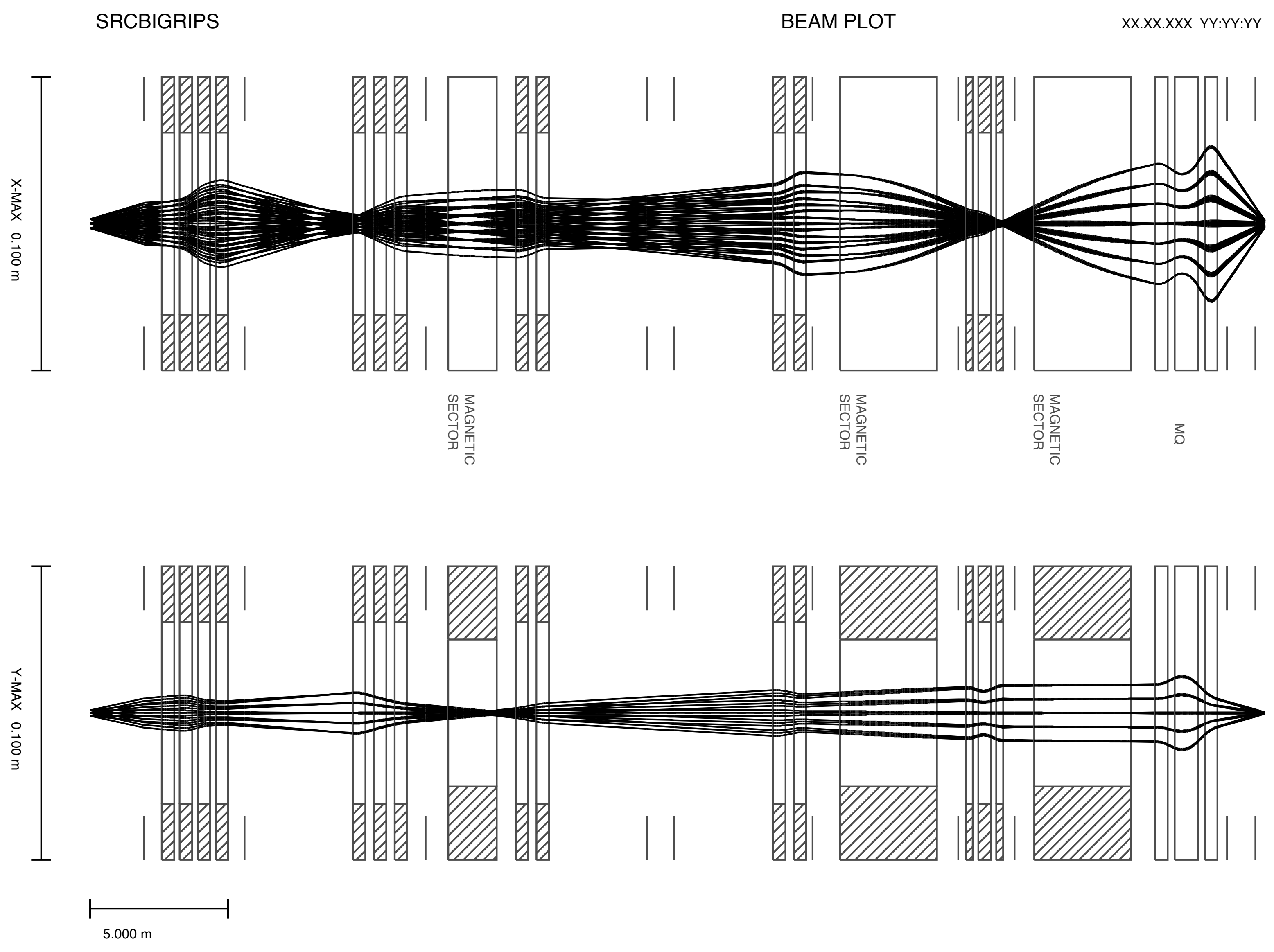
代表点 1 (89 epoch)




代表点3 (243 epoch)




代表点2 (290 epoch)




~ 最大 40 A 程度電流が変化している。