Muon Flavor Violation in Two Higgs Doublet Model with Extra Yukawa Couplings

Wei-Shu Hou and Girish Kumar

Department of Physics, National Taiwan University, Taipei 10617, Taiwan

(Dated: August 20, 2020)

The new round of experiments, MEG II, COMET/Mu2e and Mu3e, would soon start to push the $\mu \to e\gamma$, $\mu N \to eN$ conversion and $\mu \to 3\epsilon$ frontier, while Belle II would probe $\tau \to \mu\gamma$ and $\tau \to 3\mu$. In the general two Higgs doublet model with extra Yukawa couplings, we show that all these processes probe the lepton flavor violating (LFV) dipole transition that arises from the two loop mechanism, with scalar-induced contact terms subdominant. This is because existing data suggest the extra Yukawa couplings $\rho_{\mu e}, \rho_{e e} \lesssim \lambda_\epsilon$, while $\rho_{\mu \tau}, \rho_{e \tau} \lesssim \lambda_\tau$ and $\rho_{\mu \mu} \lesssim \lambda$, with $\lambda_i$ the usual Yukawa coupling of the Standard Model (SM), where $\rho_{\mu\mu\mu\tau}$ and $\rho_{\mu\epsilon\mu\epsilon}$ enter the $\mu\gamma$ and $\tau\mu\gamma$ two loop amplitudes, respectively. With the $B_s \to \mu\mu$ decay rate basically consistent with SM expectation, together with the $B_s$ mixing constraint, we show that $B_s \to \tau\tau$ would also be consistent with SM, while $B_s \to \tau\mu$ and $B \to K\tau\mu$ decays would be out of reach of projected sensitivities, in strong contrast with some models motivated by the $B$ anomalies.

I. INTRODUCTION

The study of muon properties is practically the oldest subject of particle physics, but remains at the forefront of current research. The MEG bound $^1$ on muon flavor violating (μFV) $\mu \to e\gamma$ decay rate at 90% C.L. is

$$B(\mu \to e\gamma) < 4.2 \times 10^{-13}, \quad \text{(MEG, 2016)} \quad (1)$$

while a rather dated result of SINDRUM gives $^2$

$$B(\mu \to 3\epsilon) < 1.0 \times 10^{-12}, \quad \text{(SINDRUM, 1988)} \quad (2)$$

for $\mu^+ \to e^+e^-e^+$ search. A third type of μFV search studies $\mu \to e$ conversion on nuclei. Normalized to the muon capture rate, SINDRUM II finds $^3$

$$R_{\mu e} < 7 \times 10^{-13}, \quad \text{(SINDRUM II, 2006)} \quad (3)$$

for $\mu \to e$ conversion on gold.

With schedules delayed by the current world pandemic, MEG II $^4$ will push the $\mu \to e\gamma$ bound down to $\sim 6 \times 10^{-14}$ with three years of data taking. A new experiment to search for $\mu^+ \to e^+e^-e^+$, Mu3e $^5$, plans to reach down to $5 \times 10^{-15}$ with three years of running, and is limited mostly by the muon beam intensity. Projected intensity improvements $^6$ by up to two orders of magnitude seem feasible, hence Mu3e can eventually reach down to $10^{-16}$ in sensitivity. In contrast, to improve $\mu \to e\gamma$ sensitivity beyond MEG II, innovations are needed for background suppression.

In terms of projected improvements, $\mu \to e$ conversion i.e. $\mu N \to eN$ is perhaps the most promising. SINDRUM II operated at the limits of power consumption, so new developments $^7$ are based on the idea $^8$ of using special solenoids for pion capture, muon transport, as well as detection, which significantly improves muon intensity. Phase I of COMET $^9$ aims for $R_{\mu e} < 7 \times 10^{-15}$, eventually reaching down to $10^{-17}$ for Phase II. Similar to COMET Phase II in design, Mu2e $^10$ aims at $2.6 \times 10^{-17}$ sensitivity. Both experiments can be improved further. For example, ongoing $^6$ PRISM/PRIME $^11$ developments aim at bringing the limit eventually down to a staggering $10^{-19}$. Although the primary objective for $\mu N \to eN$ is contact interactions, it also probes $^{12}$ the dipole interaction, and can be in place to probe $\mu \to e\gamma$ if the associated backgrounds of the latter cannot be brought under control at high muon intensity.

The current bounds and projected sensitivities on μFV processes are summarized in Table 1. The impressive bounds for the muon reflect seven decades of studies. We also list the corresponding processes for $\tau$, i.e. $\tau \to \mu\gamma$ and $\tau \to 3\mu$, where the current bounds are from B factories $^{13, 14}$, and expectations $^{15}$ are for Belle II with 50 ab$^{-1}$ in the coming decade. LHCb can $^{16}$ cross check the Belle II result on $\tau \to 3\mu$ after Upgrade II, i.e. at the High Luminosity LHC (HL-LHC). The heaviness of $\tau$ hence its later discovery, smaller production cross section plus difficulty in detection underlie the weaker search limits. However, its heavy mass and third generation nature offers a different window on New Physics, or equivalently beyond the Standard Model (BSM) physics.

We studied $^{17}$ $\tau \to \mu\gamma$ decay previously in conjunction with $\tau \to \mu\mu$, where $h$ is the 125 GeV boson discovered in 2012 $^{18}$. The context was the two Higgs doublet model (2HDM) with extra Yukawa couplings, which was called general 2HDM ($g$2HDM). The $h$ boson picks up the extra $\rho_{\tau\mu}$ Yukawa coupling from the CP-even exotic Higgs boson $H$ via $h-H$ mixing. Given that this mixing angle, $c_{\tau\mu}$, is known to be small (the alignment phenomenon $^{19}$), or that $h$ so closely resembles the SM Higgs boson $^{18}$, only a weak constraint is placed on $\rho_{\tau\mu}$. Together with the extra top Yukawa coupling $\rho_{t\mu}$, the $\rho_{\tau\mu}$ coupling induces $\tau \to \mu\gamma$ decay via the two-loop mechanism $^{20}$. Taking $\rho_{t\mu} \sim \lambda_\tau \simeq 1$, the strength of the top Yukawa coupling of SM, it was shown that Belle II can probe the $\rho_{\tau\mu} \lesssim \lambda_\tau \simeq 0.010$ parameter space.

Taking $\rho_{t\mu}$ at $\mathcal{O}(\lambda_\tau)$ and $\rho_{\tau\mu} \lesssim \lambda_\tau$ together, they correspond to $^{17}$

$$\rho_{fj}^I \lesssim \lambda_\tau^I, \quad (j \neq 1) \quad (4)$$

with $\rho_{f1}^I \ll \lambda_\tau^I$ expected. As we will see, this relation does not hold for down-type quarks because of tight constraints from ($K$ and $B$ meson physics. The probe of
The paper is organized as follows. In the next Section, we discuss $\mu \rightarrow e\gamma$ in g2HDM, which is pretty much parallel to what we have done for $\tau \rightarrow \mu\gamma$ [17]. We show that the $\mu \rightarrow e\gamma$ process probes the $\rho_{\mu e}\rho_{tt}$ product in g2HDM, as well as $c_e\rho_{ee}$ where $c_e$ is the $h-H$ mixing angle. In Sec. III we cover the $\mu \rightarrow 3e$ and $\mu N \rightarrow eN$ processes, as well as $\tau \rightarrow 3\mu$. We show that the g2HDM effects are very suppressed at tree level, and that all these processes eventually pick up the $\mu e\gamma$ or $\tau\mu\gamma$ dipole couplings. In Sec. IV we contrast the projections of g2HDM with the PS3 model [27], motivated by the $B$ anomalies, covering rare $B$ decays such as $B_q \rightarrow \tau\tau$, $\tau\mu$, $B \rightarrow K^{(*)}\tau\tau$, $K^{(*)}\tau\mu$, and $\tau \rightarrow \mu\gamma$ as well. We also mention $B \rightarrow \mu\nu$, $\tau\nu$ decays, where g2HDM could actually reveal [28] itself. We briefly touch upon muon EDM and $g-2$, before offering our conclusion in Sec. V.

II. THE $\mu \rightarrow e\gamma$ PROCESS

MEG II [4] has genuine discovery potential in g2HDM with extra Yukawa couplings.

We have studied [17] $\tau \rightarrow \mu\gamma$ decay previously and showed that $\rho_{\tau\mu} \simeq \lambda_\tau \simeq 0.010$ (part of Eq. (4)) can be probed by Belle II as it pushes down to $\mathcal{O}(10^{-9})$ [15]. The $\mu \rightarrow e\gamma$ process is the template for $\tau \rightarrow \mu\gamma$ decay, for which the two loop mechanism (see Fig. 1) of Ref. [20] was originally written in g2HDM (called Model III [29] at that time) that possesses extra Yukawa couplings.

Our emphasis is on phenomenological discussion, so we take Ref. [17] as template and do not recount details of the g2HDM here. The formulas used in Ref. [17], besides originating from Ref. [20], have also been checked against those of Ref. [30], although one should use caution with this reference, as it was written in a time when there was a hint for $h \rightarrow \tau\mu$ from CMS, which has subsequently disappeared [18]. What should be emphasized is that, in g2HDM, the exotic Higgs bosons $H$, $A$ (CP-odd) and $H^+$ would naturally populate the 300–600 GeV range, but which we have surprisingly little knowledge of.

In g2HDM, flavor changing neutral Higgs (FCNH) couplings are controlled [29] by the mass-mixing hierarchy, hence the one loop diagram, Fig. 1(left), is expected to be highly suppressed [20] by multiple chirality flips. Using the one loop formula of Ref. [17] with simple change

| $\mu$FV process | Current bound | Future sensitivity |
|------------------|---------------|--------------------|
| $\mu \rightarrow e\gamma$ | $4.2 \times 10^{-13}$ (MEG [4]) | $6 \times 10^{-14}$ (MEG II [4]) |
| $\mu \rightarrow 3e$ | $1.0 \times 10^{-12}$ (SINDRUM [2]) | $\sim 10^{-15}$–$10^{-16}$ (Mu3e [5]) |
| $\mu N \rightarrow eN$ | $7 \times 10^{-13}$ (SINDRUM II [3]) | $\sim 10^{-15}$–$10^{-17}$ (COMET [9]) |
| $\tau \rightarrow \mu\gamma$ | $4.4 \times 10^{-8}$ (BaBar [13]) | $\sim 10^{-9}$ (Belle II [15]) |
| $\tau \rightarrow 3\mu$ | $2.1 \times 10^{-8}$ (Belle [14]) | $3.3 \times 10^{-10}$ (Belle II [15]) |

TABLE I. Summary of current experimental bounds and future sensitivities of $\mu$FV processes.
of indices, we assume $\rho_{\mu e} \rho_{\mu e}$ from an intermediate muon in the loop is negligible compared with $\rho_{\mu e}^2 \rho_{\mu e}$ from an intermediate $\tau$, which is even more so the case for an intermediate $e$. We illustrate this “one loop benchmark” in Fig. [2] for $\rho_{\mu e} = \rho_{\mu e} = \lambda_\tau$ and $\rho_{\mu e} = \lambda_e$, and for $m_A = m_H + 200$ GeV (or with $H \leftrightarrow A$ interchanged). The effect by itself is out of reach for any time to come, unless $A, H$ are very light. In fact, for $m_A = m_H \in (300, 500)$ GeV, due to a cancellation mechanism, the MEG or the future MEG II bounds would allow $\rho_{\mu e}^2 \rho_{\mu e}$ at $O(10^3)$ times larger than $\lambda_e \lambda_e$, which is very accommodating. For nondegenerate $m_H = 300$ GeV, $m_A = 500$ GeV, we find $\rho_{\mu e}^2 \rho_{\mu e} / \lambda_e \lambda_e \lesssim 17$ by MEG can be improved to 6.6 with MEG II, with the results similar for flipping $H \leftrightarrow A$.

It is the two loop mechanism [20] that is of interest for g2HDM, where the $\rho_{\mu e}$ coupling induces $\mu \rightarrow e \gamma$ decay by inserting the $\phi \rightarrow \gamma V^*$ vertex ($\phi = h, H, A$, see Fig. 1(center) and 1(right)) related to the $h \rightarrow \gamma \gamma$ process, with $V = Z$ subdominant. Following Ref. [17] for $\tau \rightarrow \mu \gamma$, we define two BSM benchmarks for illustrating two loop effects. Taking the extra top Yukawa coupling $\rho_{\mu t} \approx 1$ while setting $c_\gamma = 0$, one maximizes the $H, A$ effect but decouples the $h$ boson. This “BSM benchmark” is illustrated in Fig. [2], where $\rho_{\mu e} = \rho_{\mu \mu} \approx 0.3 \lambda_e$ is taken to satisfy the current MEG bound of Eq. (1) at $m_H$ or $m_A = 200$ GeV. The MEG II experiment will continue to probe $\rho_{\mu e}$ down to lower values.

A second benchmark illustrates the effect of the SM-like $h$ boson, where we take $\rho_{\mu t} = 0$ to decouple the exotic $H, A$ scalars, but take $c_\gamma = 0.2$ as a large value that may still be allowed. This “$h$ benchmark” is also plotted in Fig. [2] giving $B(\mu \rightarrow e \gamma) \approx 10^{-14}$ for $\rho_{\mu e} = \rho_{\mu \mu} \approx 0.3 \lambda_e$, which appears out of reach for MEG II. Depending on whether $c_\gamma$ is smaller or larger than 0.2, the rate would drop further or become larger, although a $c_\gamma$ value larger than 0.2 may not be plausible. But the rate scales only with the product of $c_\gamma^2 \rho_{\mu e}^2$, and if $\rho_{\mu t}$ truly vanishes, a $\rho_{\mu e}$ value larger than 0.3$\lambda_e$ is allowed.

We note that, unlike the $\tau \rightarrow \mu \gamma$ case where $h \rightarrow \tau \mu$ [18] provides a constraint [17] on $c_\gamma \rho_{\mu e}$, no realistic constraint on $c_\gamma \rho_{\mu e}$ can be extracted from $h \rightarrow \mu e$ search [18] for our purpose, as $\mu \rightarrow e \gamma$ already constrains $\rho_{\mu e}$ to be so small. On the other hand, the value of $\rho_{\mu t}$ is not known at present, except that any finite value may suffice [21] for EWBG. For instance, in trying to account for the strong bound on electron EDM by ACME [23], the smaller $|\rho_{\mu t}| \approx 0.1$ was chosen in Ref. [22] to ease the tension. While $\rho_{\mu t}$ at $O(1)$ is not strictly ruled out, we stress that $\mu \rightarrow e \gamma$ probes the $\rho_{\mu e} \rho_{\mu e}$ product, hence we do not really know whether we are probing $\rho_{\mu e}$ for the BSM benchmark below the strength of $\lambda_e$ yet. Thus, for example, if $\rho_{\mu t} = 0$ and EWBG is through the $\rho_{\mu e}$ mechanism [21], then the MEG bound of Eq. (1) only requires $\rho_{\mu e} = \rho_{\mu \mu} \lesssim 1.9 \lambda_e$ for our $h$ benchmark, and MEG II could probe down to 0.7$\lambda_e$. Both values are still in accord with Eq. (3), but we note that if $c_\gamma$ is lower than the value of 0.2 used, which seems likely, then the allowed $\rho_{\mu e}$ range would rise.

As a passing remark, we expect $\tau \rightarrow e \gamma$ to be much

FIG. 1. One-loop, two-loop fermion and two-loop $W$ diagrams for $\mu \rightarrow e \gamma$.  

![Diagram](https://via.placeholder.com/150)

FIG. 2. Comparison of benchmark scenarios for $\mu \rightarrow e \gamma$ as function of scalar masses. For one loop red dashed curves, lower (upper) curve is for $m_A = m_H + 100$ (200) GeV, and flipping $H \leftrightarrow A$ is not much different. For the two-loop BSM benchmark, black curve is for degenerate $m_H = m_A$, red (blue) curves show variation in $m_H (m_A)$ with $m_A (m_H)$ heavier by 100, 200 GeV, where satisfying the MEG bound [1] at the low 200 GeV fixes $\rho_{\mu e} = \rho_{\mu \mu} \approx 0.3 \lambda_e$. Holding this value fixed, the two-loop $h$ benchmark is the green dashed horizontal line, which lies below the MEG II [3] sensitivity. See text for further discussion.
suppressed compared with $\tau \rightarrow \mu \gamma$ in g2HDM, as $\rho_{\tau\mu}$ is expected to be much smaller than $\rho_{\tau\mu}$.

### III. OTHER $\mu$FV PROCESSES

#### A. $\mu \rightarrow 3e$ and $\tau \rightarrow \mu\gamma$, $3\mu$

As Mu3e would start soon to finally probe below the old SINDRUM bound of $10^{-12}$, Eq. (2), we estimate the $\mu \rightarrow 3e$ rate. We find, consistent with Ref. [31], the simple tree level formula for $\mu \rightarrow 3e$,

$$B(\mu \rightarrow 3e) = \frac{1}{32} \left[ 2 \sum \frac{y_{\phi\mu \mu} y_{\phi\mu \mu}}{m_{\phi}^2} + 2 \left| \sum \frac{y_{\phi\mu \mu} y_{\phi\mu \mu}}{m_{\phi}^2} \right|^2 + \sum \frac{y_{\phi\mu \mu} y_{\phi\mu \mu}}{m_{\phi}^2} \right],$$

(7)

where we ignore extra Yukawa coupling corrections to the muon decay rate $\Gamma_{\mu}$ [28]. $y_{\phiij}$ are Yukawa couplings for $\phi = h, H, A$ that can be read off from Eq. (3) of Ref. [17], and $m_{\phi}$ are scalar masses normalized to $v$.

In view that 200 GeV may be too aggressive for the lowest possible exotic scalar mass, we take for illustration the relatively conservative $m_{\phi}$ = 300 GeV. We define our benchmark further as follows: We take, somewhat arbitrarily, $c_\gamma = 0.05$ for the effect of $h, H, A$ that can be read off from Eq. (3) of Ref. [17], and $m_{\phi}$ are scalar masses normalized to $v$.

We find $\Gamma_{\mu}$ is saturated by the $m_{\phi}$ $\rightarrow 3e$ dipole again dominates at tree level, which is far out of experimental reach. But the $\mu e\gamma$ dipole coupling can generate $\mu \rightarrow 3e$ [32],

$$B(\mu \rightarrow 3e) = \frac{1}{32} \left[ 2 \sum \frac{y_{\phi\mu \mu} y_{\phi\mu \mu}}{m_{\phi}^2} + 2 \left| \sum \frac{y_{\phi\mu \mu} y_{\phi\mu \mu}}{m_{\phi}^2} \right|^2 + \sum \frac{y_{\phi\mu \mu} y_{\phi\mu \mu}}{m_{\phi}^2} \right],$$

(8)

and we find $\Gamma_{\mu}$ is saturated by the $m_{\phi}$ $\rightarrow 3e$ dipole $\approx 2.6 \times 10^{-15}$ for our benchmark. Though out of reach of Mu3e in early phase, it should be detectable with muon intensity upgrades, where the experiment should be able to confirm the $\mu \rightarrow e\gamma \rightarrow 3e$ nature.

For $\tau$, our benchmark gives $B(\tau \rightarrow \mu\gamma) \approx 3.1 \times 10^{-9}$, which is an order of magnitude below current B factory bound, but reachable by Belle II. Using analogous formulas as above, we find $B(\tau \rightarrow 3\mu)$ is still out of Belle II reach. However, if Belle II discovers $\tau \rightarrow \mu\gamma$ in early data, i.e. above $10^{-8}$, which is certainly possible [17] in g2HDM, it would imply $\tau \rightarrow 3\mu$ at $10^{-10}$ or above, which can be probed by the fixed-target experiment, TauFV [33], that is being planned. Also arising from the $\tau e\gamma$ dipole, $\tau \rightarrow e^+e^-\gamma$ would be slightly higher. But, suppressed by $\rho_{\tau\mu}$, the $\tau \rightarrow e^-\mu^+\mu^-$ process is expected to be far below the $\tau \rightarrow 3\mu$ contact process in g2HDM, while $\tau \rightarrow e^-\mu^+\mu^-$ would be suppressed by the $\tau \rightarrow e\gamma$ dipole transition.

#### B. $\mu N \rightarrow eN$ Conversion

With two competing experiments, COMET and Mu2e, prospects for pushing $\mu \rightarrow e$ conversion during the next decade or more is exceptionally bright, as the current limit [3] of $R_{\mu e} < 7 \times 10^{-11}$, Eq. (5), is expected to improve by $\sim 3$-4 orders of magnitude [27, 30].

The relevant effective Lagrangian is given by [34, 35]

$$C_{\text{eff}} = m_\mu \left( C_{\tau T}^\rho c_{\alpha}c_{\beta}L_{\mu} + C_{\tau T}^\rho c_{\alpha}c_{\beta}R_{\mu} \right) F_{\alpha \beta} + \left( C_{\rho q}^{SR} c_{L\mu} + C_{q \rho q}^{SL} e_{R\mu} \right) m_\mu m_q q Q,$$

(9)

where $C_{\rho q}^{L,R}$ correspond to the $\mu e\gamma$ dipole, while $C_{q \rho q}^{SL(R)}$ are coefficients to contact terms generated by scalar exchange. There are no current-current interactions at tree level in g2HDM. One computes the conversion rate $\Gamma_{\mu \rightarrow e}$ and normalize to the muon capture rate to get $R_{\mu e}$. The conversion rate is given by

$$\Gamma_{\mu \rightarrow e} = m_\mu \left( \frac{1}{2} C_{\tau T}^{L(R)} D + 2 m_p m_p C_{p \rho q}^{SL(R)} S_{p \rho} + p \rightarrow n \right)^2,$$

(10)

where the $L$ and $R$ effects add in quadrature, and $S_{p \rho}$ accounts for lepton-nucleus overlap. For gold, we use $D = 0.189$, $S_{p \mu} = 0.0614$, and $S_{p \mu} = 0.0918$. In Eq. (10),

$$C_{\rho q}^{SL(R)} = \sum C_{q \rho q}^{SR}(f_{p}),$$

(11)

relates to nucleon matrix elements, $f_{p}^{p,n}$, that account for the quark content of the proton, where we use $f_{u}^{p} = 0.024$, $f_{d}^{p} = 0.033$, $f_{s}^{p} = 0.043$ [38]. For heavy quarks, we follow Ref. [37] and use $f_{Q}^{p,n} = (2/27)(1 - f_{p}^{p,n} - f_{d}^{p,n} - f_{s}^{p,n})$ [39] for $Q = c, b, t$.

In g2HDM, the tree level contribution can be written in terms of Wilson coefficients [35] for the contact terms induced by scalar $\phi = h, H, A$ boson exchange,

$$C_{q \rho q}^{SL} = (2/v^4) \sum y_{\phi \rho q} \text{Re} \bar{y}_{\phi \rho q} / m_{\phi}^2,$$

(12)

where $y_{\phi \rho q}$ ($y_{\phi \rho q}$) is normalized to $\lambda_{\mu}$ ($\lambda_{\rho}$), and one flips $y_{\rho \phi q} \rightarrow y_{\phi \rho q}^* \rightarrow C_{q \rho q}^{SR}$. The dipole $C_{\tau T}^{L(R)}$ contributions are related to $\mu \rightarrow e\gamma$, i.e. $C_{\tau T}^{L(R)} = \sqrt{\theta_{\tau \mu \gamma}} A_{L,R}$, where $A_{L,R}$ contribute to $B(\mu \rightarrow e\gamma)$ (see Ref. [17] for $B(\tau \rightarrow \mu\gamma)$ formulas). The $\mu e\gamma$ dipole again dominates $\mu N \rightarrow eN$ conversion, with contact terms subdominant. For our benchmark, we obtain the conversion ratio $R_{\mu e}^{\text{contact}} \approx 2.4 \times 10^{-16}$ for gold as an example, while $R_{\mu e}^{\text{dipole}} \approx 1.6 \times 10^{-15}$. Here we have used $\rho_{\tau\mu} = \lambda_{q}$ for all quarks, except $\rho_{\tau\mu} \approx 0.4$ as inferred from MEG bound with our benchmark. We note that contact terms are relatively important in $\mu \rightarrow e$ conversion compared to $\mu \rightarrow 3\mu$ process. These values can be probed at COMET and Mu2e. In fact, these experiments are poised to overtake MEG II in probing $\mu \rightarrow e\gamma$ in g2HDM. Furthermore, if observed, together with the knowledge of nuclear matrix elements, one can use several different nuclei to probe and extract the effect of the contact term(s) in Eq. (9).
We see that the extra $\rho_{\mu e}$ and $\rho_{ee}$ couplings of g2HDM hide very well so far from muon probes. It is with the help of extra $\rho_{\mu t}$ coupling via the two loop mechanism for $\mu \to e\gamma$ decay that MEG constrains $\rho_{\mu e} \lesssim \lambda_e$ (see Eq. (6)). MEG II would continue this program, but the $\mu N \to eN$ experiments, COMET and Mu2e, would become competitive when $10^{-15}$ sensitivity is reached. Mu2e can confirm the dipole nature once $\mu \to 3e$ is also observed with high muon intensity upgrades. Likewise, $\tau \to \mu \gamma$ would probe $\rho_{\mu t}$ modulo $\rho_{tt}$, but the $\tau \to 3\mu$ process seems out of reach for Belle II (hence LHCb) if g2HDM holds, even if Belle II quickly observes $\tau \to \mu \gamma$. Thus, while there remains hope for discovery, $\mu$FV physics look “sanitized” within g2HDM that possesses these extra $\rho_{\mu t}$ (and $\rho_{tt}$) Yukawa couplings, which bears witness to the long history of muon research.

IV. CONTRAST: MUON, OR BOLD

In this section, we contrast the “sanitized” muon front of the previous sections with what we dub the “bold” BSM front inspired by B anomalies. We refer to Ref. [24] for a discussion of all the current B anomalies, including cautionary notes on the experimental results. Extending from $\mu$FV, we discuss BSM effects in (semi-)leptonic B decays, be it BSM enhancement of $B_q \to \tau\tau$, or the purely BSM decays $B_q \to \tau\mu$, $B \to K\tau\mu$. We also touch upon the $B_q \to \mu\mu$ and $B \to \mu\nu, \tau\nu$ decays, which already appear to be SM-like in rate.

A. BSM-enhanced: $B_q \to \tau\tau$ Modes

The “BaBar anomaly” in $B \to D^{(*)}\tau\nu$ [18, 24] suggests a large tree level BSM effect interfering with the SM $b \to c\tau\nu$ amplitude. Based on general arguments, it was pointed out [40] that such a large effect should be accompanied by similar effects in $b \to s\tau\tau$. Note that, because of the difficult $\tau^+\tau^-$ signature, the experimental bounds [19] are rather poor. Projecting from the BaBar anomaly, Ref. [40] suggested that $B(B_6 \to \tau\tau) \sim 5 \times 10^{-4}$ (or larger) is possible, to be compared with $7.7 \times 10^{-7}$ in SM [11]. Similarly, $B(B \to K^{(*)}\tau\tau) \sim 10^{-4}$ is projected. The theory suggestion was in part stimulated by the LHCb search [42], based on 3 fb$^{-1}$ Run 1 data, setting the 90% C.L. bound of

$$B(B_6 \to \tau\tau) < 5.2 \times 10^{-3}, \quad \text{(LHCb, 2017)} \quad (13)$$

which is an order of magnitude higher than the theory suggestion. Likewise, the only limit on 3-body search, $B(B^+ \to K^+\tau^+\tau^-) < 2.3 \times 10^{-3}$ from BaBar [43], is also poor. One suffers from lack of mass reconstruction capability, and only at the HL-LHC after LHCb Upgrade II [16] can the sensitivity reach $\sim 5 \times 10^{-4}$, touching the upper reaches of projected enhancement [40]. Belle II plans to take some $\Upsilon(5S)$ data early on, and projects the reach of $\sim 8.1 \times 10^{-4}$ [19]. As the environment is clean, Belle II would likely take more $\Upsilon(5S)$ data if the BaBar anomaly is confirmed. For $B \to K^{(*)}\tau\tau$, the Belle II sensitivity of $\sim 2 \times 10^{-5}$ [15] should be able to probe the range of interest at $O(10^{-4})$.

We list the current limits and future prospects for the $B_q \to \tau\tau$ and $B \to K^{(*)}\tau\tau$ modes in Table II.

B. Purely BSM: $B_q \to \tau\mu$ and $B \to K\tau\mu$ Modes

The B anomalies suggest lepton universality violation (LUV), such as $B \to D^{(*)}\tau\nu$ vs $B \to D^{(*)}\mu\nu$, or $B \to K^{(*)}\mu\mu$ vs $B \to K^{(*)}\ee$. It was suggested [44] on general grounds the possibility of accompanying lepton flavor violation (LFV), giving rise to interesting decays such as $B_q \to \ell\ell'$ and $B \to K\ell\ell'$ for $\ell \neq \ell'$. As the B anomalies persisted, serious model building went underway, and we take the so-called PS$^3$ model [25] as the standard bearer for ambitious UV-complete models (which we term “bold”). To handle severe low energy constraints and focus on the third generation, the Pati-Salam (PS) model [45] comes in three copies. The presence of lepto-quarks (LQ) in the Pati-Salam model induce the decays such as $B_q \to \tau\mu$ and $B \to K\tau\mu$, where detailed phenomenology was given in Ref. [26].

These are striking signatures! Before long, with 3 fb$^{-1}$ Run 1 data, LHCb sets [46] the 90% C.L. limit of

$$B(B_6 \to \tau\mu) < 3.4 \times 10^{-5}, \quad \text{(LHCb, 2019)} \quad (14)$$

which contrasts with the poor performance of Eq. (13) for $B_6 \to \tau\tau$. This limit practically ruled out the entire $B(B_6 \to \tau\mu)$ range projected by Ref. [24], forcing model builders to introduce [27] right-handed LQ interaction as tune parameters. In so doing, $B_6 \to \tau\tau$ and $B \to K\tau\mu$ decays get enhanced [27], which is in accordance with Ref. [40]. It would be interesting to see the full 9 fb$^{-1}$ Run 1 + 2 result for $B_6 \to \tau\mu, \tau\tau$ modes. Perhaps because the analysis of Ref. [46] was still underway when the LHCb Upgrade II document [16] was being prepared, we cannot find the sensitivity projections of $B_6 \to \tau\mu$ for full LHCb Upgrade II data (and neither for Belle II), hence we state this explicitly in Table II.

BaBar has searched [47] for the companion $B \to K\tau\mu$ mode. Using full hadronic tag to reconstruct the other charged B hence with full kinematic control, by measuring $K^+$ and $\mu^-$, one projects into the $m_\tau$ window without reconstructing the $\tau$. The result at 90% C.L. is [17]

$$B(B^+ \to K^+\tau^+\mu^-) < 2.8 \times 10^{-5}, \quad \text{(BaBar, 2012)} \quad (15)$$

$$< 3.9 \times 10^{-5}, \quad \text{(LHCb, 2020)} \quad (16)$$

for the better measured charge combination, and Eq. (16) is the recent LHCb measurement [48] with full 9 fb$^{-1}$ Run 1 + 2 data. We first note that Belle has not performed this measurement so far, despite having more data than BaBar. The second point to stress is that, although the LHCb result may not appear competitive at first sight,
they exploit $B^+ \rightarrow D^* \rightarrow B^+ K^-$ decay and use the $K^-$ to tag [39] the $B^+$ for full kinematic control, putting LHCb in the game for the $B^+ \rightarrow K^+ \tau^+ \mu^-$ pursuit, and making things more interesting for the Belle II era.

LHCb also places the best bounds [50] for $B(B_s \rightarrow \mu \nu) < 5.4 \times 10^{-9}$ and $B(B_d \rightarrow \mu \nu) < 1.0 \times 10^{-9}$, as well as $B(B^+ \rightarrow K^+ \mu^+ e^-) < 6.4 \times 10^{-9}$ [51]. The current limits and future prospects for the $B_q \rightarrow \tau \mu$ and $B \rightarrow K^{(*) \tau \mu}$ modes are listed in Table II. The $\mu$ modes are also listed, but aside from the comment given in Ref. [44], it is not easy from the model building point of view to make projections that are experimentally accessible.

C. SM-like: $B_q \rightarrow \mu \mu$ and $B \rightarrow \tau \nu, \mu \nu$ Modes

It is useful to recall that $B_q \rightarrow \mu \mu$ was a front runner in the 2000’s as possibly greatly enhanced, but a few years into LHC running, the $B_{s,d} \rightarrow \mu \mu$ decays became consistent with SM: the PDG values [18] are $B(B_s \rightarrow \mu \mu) = (3.0 \pm 0.4) \times 10^{-9}$ and $B(B^0 \rightarrow \mu \mu) = (1.1^{+1.4}_{-1.3}) \times 10^{-10}$, compared with the SM expectation [52] of $B(B_s \rightarrow \mu \mu) = (3.66 \pm 0.14) \times 10^{-9}$ and $B(B^0 \rightarrow \mu \mu) = (1.03 \pm 0.05) \times 10^{-10}$. We note that ATLAS, CMS and LHCb have recently combined [53] their 2011–2016 data to give $B(B_s \rightarrow \mu \mu) = (2.69^{+0.35}_{-0.33}) \times 10^{-9}$ and $B(B^0 \rightarrow \mu \mu) < 1.6 \times 10^{-10}$ at 90% C.L. A discrepancy for $B_q \rightarrow \mu \mu$ does not seem to be due to the effect of negative central value from ATLAS. We will use the PDG result (see Table II), which is consistent with SM, and thus the two-sided bound, i.e. $B(B \rightarrow \mu \bar{\nu})$ could be above or below the nominal SM value [28] of $3.9 \times 10^{-7}$, and the $R_{B_s}^{\mu \bar{\nu}}$ ratio provides a good probe of g2HDM for Belle II in the next few years.

We reiterate that, though $B_q \rightarrow \mu \mu$ are loop processes while $B \rightarrow \tau \nu, \mu \nu$ are at tree level, and the measured values still have to settle, none are in disagreement with SM expectation, which put constraints on BSM models inspired by $B$ anomalies, as well as g2HDM. The current status and future prospects are listed in Table II.

### Table II: Summary of current experimental data on $B$ decays considered in our analysis. All upper bounds are at 90% C.L., and Phase II for LHCb stands for HL-LHC running after Upgrade II.

| Decay Mode | Current bound | Future sensitivity |
|------------|---------------|--------------------|
| $B_s \rightarrow \tau \tau$ | $5.2 \times 10^{-3} (LHCb [42])$ | $\sim 8 \times 10^{-3}$ (Belle II, 5 ab$^{-1}$ [15]) |
| $B_d \rightarrow \tau \tau$ | $1.6 \times 10^{-3} (LHCb [42])$ | $\sim 5 \times 10^{-4}$ (LHCb Phase II [16]) |
| $B \rightarrow K \tau \tau$ | $2.3 \times 10^{-3} (BaBar [43])$ | $\sim 1 \times 10^{-4}$ (Belle II [15]) |
| $B_s \rightarrow \mu \mu$ | $5.4 \times 10^{-9} (LHCb [50])$ | $\sim 2 \times 10^{-5}$ (Belle II [15]) |
| $B_d \rightarrow \mu \mu$ | $1.2 \times 10^{-8} (LHCb [46])$ | [Not yet publicized] |
| $B \rightarrow K \tau \mu$ | $2.8 \times 10^{-5} (BaBar [47])$ | $1.3 \times 10^{-6}$ (Belle II [15]) |
| $B_s \rightarrow \mu \nu$ | $3.0 \pm 0.4 \times 10^{-9}$ (PDG [18]) | $3 \times 10^{-6}$ (LHCb Phase II [16]) |
| $B_d \rightarrow \mu \nu$ | $1.0 \times 10^{-9} (LHCb [50])$ | $\sim 3 \times 10^{-6}$ (Belle II [15]) |
| $B \rightarrow \tau \nu$ | $5.3 \pm 2.2 \times 10^{-7}$ (Belle [58]) | [LHCb competitive] |
| $B \rightarrow \mu \nu$ | $(1.1 \pm 0.2) \times 10^{-4}$ (PDG [18]) | $\sim 5\%$ (Belle II [15]) |

TABLE II. Summary of current experimental data on $B$ decays considered in our analysis. All upper bounds are at 90% C.L., and Phase II for LHCb stands for HL-LHC running after Upgrade II.

where we add the statistical and systematic errors in quadrature, treating as Gaussian. Eq. (17) is consistent with SM, but gives a two-sided bound, i.e. $B(B \rightarrow \mu \bar{\nu})$ could be above or below the nominal SM value [28] of $3.9 \times 10^{-7}$, and the $R_{B_s}^{\mu \bar{\nu}}$ ratio provides a good probe of g2HDM for Belle II in the next few years.
D. Contrasting g2HDM with “Boldness”

Having presented the status of various (semi-)leptonic rare $B$ decays, where some striking projections arise from models motivated by $B$ anomalies, we turn to contrasting with g2HDM, the projections of which conform better with the more “sanitized” tradition of muon physics.

1. From $\mu$FV to PS$^3$

The purely leptonic $\mu$FV processes discussed previously, such as $\mu \rightarrow e\gamma$ in Sec. II, and $\mu \rightarrow 3e$, $\tau \rightarrow \mu\gamma$, $\tau \rightarrow 3\mu$ and $\mu N \rightarrow eN$ in Sec. III, are illustrated in Fig. 3. That is, the current bounds and future sensitivities listed in Table I are plotted as blue solid and orange dotted circles, respectively. None are so far observed, so the current MEG bound on $\mu \rightarrow e\gamma$ is also marked by a downward red $\downarrow$ for the g2HDM projection, where, for sake of illustration, we have set up a benchmark consistent with Eqs. (4) and (6) and with small $h$-$H$ mixing. As the scalar-induced contact effect is rather small, the dipole $\mu \rightarrow 3e$ transition is also marked by a downward red $\downarrow$. However, though subdominant, the scalar-induced contact effect for $\mu N \rightarrow eN$ is not negligible, and the downward red $\downarrow$ shows the combined dipole plus contact effect, which is destructive. The sign of interference, however, could be easily flipped, so the actual possibilities are considerably broader. The $\tau \rightarrow \mu\gamma$ rate with this benchmark is also illustrated, which falls toward the lower range of Belle II reach, while we predict that $\tau \rightarrow 3\mu$ is out of reach in g2HDM.

Likewise, the current bounds and future sensitivities for (semi-)leptonic rare $B$ decays discussed in Sec. III.A and III.B are also plotted in Fig. 3. Of interest here is some two-sided projections, as they stand at present, for the striking signatures arising from PS$^3$ [27]:

\begin{align}
10^{-4} \lesssim B(B_s \rightarrow \tau\tau) &\lesssim 4.5 \times 10^{-3}, \\
10^{-6} \lesssim B(B_s \rightarrow \tau\mu) &\lesssim 6 \times 10^{-5}, \\
10^{-9} \lesssim B(\tau \rightarrow \mu\gamma) &\lesssim 8 \times 10^{-8},
\end{align}

while $B(B \rightarrow K\tau\mu)$ scales down from $B(B_s \rightarrow \tau\mu)$ by a factor of $\sim 9$, and for $B(B \rightarrow K\tau\tau)$ vs $B(B_s \rightarrow \tau\tau)$ the factor is $\sim 13$. We do not show the $B(\tau \rightarrow \mu\phi)$ mode [27] as it seems out of Belle II reach. These ranges are shown in Fig. 3 as grey shaded bands, where existing bounds for $B_s \rightarrow \tau\mu$ and $\tau \rightarrow \mu\gamma$ cut into the upper ranges of PS$^3$ projections, and are the points of our comparison with g2HDM expectations. As noted, the future sensitivity for $B_s \rightarrow \tau\mu$ is not quite known at present.
We note further that, with $\tau \to \mu^+$ generated by LQ in the loop, there is an anti-correlation with $B(B_s \to \tau \mu)$ within the PS$^3$ scenario \cite{22}: if the limit on $B_d \to \tau \mu$ is pushed further down with 9 fb$^{-1}$ full Run 1 + 2 data, then $B(\tau \to \mu \gamma)$ will move up and become closer to the current limit, and would be a boon to Belle II in the model scenario. Likewise, pushing down on $\tau \to \mu \gamma$ would imply an increased lower bound for $B_s \to \tau \mu$, $\tau \tau$ in PS$^3$. These bounds and (anti-)correlations allow the PS$^3$ model to “provide a smoking-gun signature for this framework . . . or could lead us to rule it out” \cite{22}.

The $B_q \to \mu \mu$ and $B \to \mu \nu$, $\tau \nu$ processes discussed in Sec. III.C are plotted differently in Fig. 3, as they are now mostly found to be consistent with SM expectations (marked as red $\star$). The measured $B_s \to \mu \mu$ rate, shown as the narrow green shaded band, covers the SM expectation but appears slightly on the low side. Likewise, $B \to \tau \nu$ is also measured to be consistent with SM, which Belle II would continue to probe. For $B_d \to \mu \nu$, we plot the more conservative upper limit from PDG, while the latest Belle update on $B \to \mu \nu$ gives a two-sided bound, which is illustrated by the broad green shaded band that covers the SM expectation. The PS$^3$ model shies away from processes that involve only muons, but $B \to \tau \nu$ does provide \cite{22} some constraint.

2. The $bq\ell\ell'$ Processes in g2HDM

The rare $B$ decay processes of interest (we only quote results for $B \to \ell \nu$) are in the form of $bq\ell\ell'$ 4-fermi interactions. Thus, the extra Yukawa couplings that enter on the quark side are $\rho_{bs}, \rho_{bd}$ at tree level, and $\rho_{bs}$ for $\ell (\ell') = \tau, \mu, e$ on the charged lepton side. For the latter, we continue to use our benchmark values $\rho_{\tau \tau}, \rho_{\mu \mu} = \lambda_{e} \approx 0.010$ (Eq. (4)), and $\rho_{\mu e}, \rho_{e e} = \lambda_{e} \approx 0.0000029$ (Eq. (6)). The issue is that, for $\ell = \ell'$, SM loop effects seem affirmed by experiment, while for $\ell \neq \ell'$, there is no SM loop effect, and one would need the leptonic FCNH couplings in g2HDM to act. In the following, we will use tree level approach to $B_q \to \mu \mu$ to infer $B_q \to \ell \ell'$ for $\ell \neq \ell'$ case, while using loop corrections for $B_q \to \ell \ell'$ to discuss $B_q \to \tau \tau$. In each case, the corresponding $B_q$ mixing constraints are taken into account.

It is well known that the measured $B_q$ mixings can be accounted for quite well by SM loop effects. For example, the operator $O_1 = (\bar{s}_a \gamma^\mu Lb_a)(\bar{s}_\beta \gamma^\mu Lb_\beta)$ for $B_s$ mixing has coefficient $(G_{\text{F}} m_W V_{ts}^2/\pi^2) S_0(x_t)$, with $x_t = m_t^2/m_b^2$, and $S_0(x_t) \approx 2.35$ from SM box diagram, and one just replaces $s \to d$ for $B_d$ mixing. In g2HDM, $\rho_{bd} (q = s, d)$ enters $B_d$ mixing at tree level, hence stringent constraints are implied.

The NP effects in $B_s$ mixings can be parameterized by defining $C_{B_s}^{\rho_{bs} \Phi_{bs}} = \langle B_q | H_{\text{full}}^{\rho_{bs}} | B_q \rangle / \langle B_q | H_{\text{SM}}^{\Phi_{bs}} | B_q \rangle$. Using the 2018 NP fit performed by UTfit \cite{59}, one finds

$$C_{B_s} = 1.110 \pm 0.090, \quad \Phi_{B_s} = (0.42 \pm 0.89)^\circ,$$

$$C_{B_d} = 1.05 \pm 0.11, \quad \Phi_{B_d} = (2.0 \pm 1.8)^\circ.$$

For sake of illustration and to reduce the number of parameters, we will treat extra Yukawas as real and assume that adding the g2HDM effect, $C_{B_q}$ and $\Phi_{B_q}$ stay within 2$\sigma$ ranges of Eq. (21).

In g2HDM, the leading effect comes from the operator $O_4 = (\bar{s}_a Lb_a)(\bar{s}_\beta Rb_\beta)$ at tree level, which constrains the product $p_{bs} \rho_{bs}^*$, while the operators $O_2 = (\bar{s}_a Lb_a)(\bar{s}_\beta Lb_\beta)$ and $O'_2 = (\bar{s}_a Rb_a)(\bar{s}_\beta Rb_\beta)$ constrain individual couplings $\rho_{bs}$, $\rho_{bs}$ but are less constraining. Furthermore, the coefficients of $O_1^0$ suffer cancellation between $H$ and $A$ contributions. Assuming $O_4$ dominance, one has the coefficient $C_4 = -y_{\rho b} y_{\rho b} / m_\rho^2$ where $\phi$ is summed over $h, H, A$, and we take $c_\gamma = 0.05$ and $m_H = m_A = 300$ GeV as before. Taking renormalization group evolution into account \cite{61}, using bag factors from Ref. \cite{62} and decay constants from Ref. \cite{63}, we find $|\rho_{bs} \rho_{bs}^*| \lesssim 0.021 \lambda_0^2$. In similar vein, we obtain $|\rho_{bs} \rho_{bd}^*| \lesssim (0.0046 \lambda_0)^2$, where we take $\lambda_0 \approx 0.16$. Assuming reality, we adopt $\rho_{bs} \approx \rho_{bs} = 0.021 \lambda_0 \sim 0.0034$, and $\rho_{bd} \approx \rho_{bd} = 0.0046 \lambda_0 \sim 0.00074$, respectively.

With $\rho_{bs} \rho_{bd}$ and $\rho_{bs} \rho_{bs}$ so small, one may expect $B_d \to \ell \ell'$ modes would be SM-like in g2HDM, which is the case for $B_s \to \mu \mu$, and to some extent $B_d \to \mu \mu$ as well: the measured strengths are indeed SM-like. At tree level, we find that $B_s \to \mu \mu$ gives stringent constraints on $\rho_{bs}(\rho_{bd})$, and can be on a par with those from $B_s$ mixing constraints. For example, for our benchmark of $c_\gamma = 0.05, \rho_{bs} = \lambda_{s} \sim 0.00661$, and $m_H = m_A = 300$ GeV, the 2$\sigma$ range of $B(B_s \to \mu \mu)$ gives the bound on $\rho_{bs} \rho_{bd} \lesssim 0.019 \lambda_0 \sim 0.00030$, which is slightly more stringent than $\rho_{bs}$ mixing. On the other hand, due to poorer measurement of $B_d \to \mu \mu$ so far, bounds on $\rho_{bd}(\rho_{bd})$ from $B_d \to \mu \mu$ are weaker than $B_d$ mixing. Thus, by the fact that $B_s \to \mu \mu$ rates are already SM-like in g2HDM, we expect $B_q \to \tau \tau$ to be not so different from SM expectations if tree contributions prevail.

With $\rho_{sb} \rho_{sb}$ and $\rho_{sb} \rho_{sb}$ suppressed, one has to take up-type extra Yukawa couplings into account, which contribute to $B_q$ mixings and $B_q \to \ell \ell'$ at one loop order. The leading contributions to $B_q$ mixings come from the same box diagrams as SM, but with either one $W^+$ or both replaced by $H^+$, which also generates $O_1$. Considering the effect of $\rho_{tt}$ only, we obtain $\Delta C_3^{WH} = y_t V_{ts}^* V_{tb}^2 |\rho_{tt}|^2 g(y, y_t)/128 \pi^2 m_t^2$, where $y = M_{H}^2 / m_t^4$ for the WH box correction, and $\Delta C_1^{WH} = -V_{ts}^2 V_{tb}^2 |\rho_{tt}|^2 f(y, y_t)/128 \pi^2 m_t^2$ for the HH box correction. Here, $H$ stands as shorthand for $H^+$, and the loop functions $f$ and $g$ are given in the Appendix.

Considering this one loop contribution by itself gives a constraint on the $\rho_{tt} m_{H^+}$ plane. For example, for a 300 GeV charged Higgs, we find $|\rho_{tt}| \lesssim 0.8$, and similar bound from $B_d$ mixing as well. However, we caution inclusion of additional up-type Yukawa couplings can induce cancellation effects, thereby weakening the constraint. Most notably, with $\rho_{ct}$ as small as $O(10^{-2}),$ one can relax $\rho_{tt}$ to $\sim 1$. As stated, we avoid cancellations and discuss tree and loop contributions separately. The same treatment is applied to rare $B$ decays, and we
continue to assume \( \rho_{q6} = \rho_{u6} \) and take them as real.

\( B_q \to \mu \mu \) can also receive significant contribution through one loop diagrams, where the leading effect is from Z penguins with \( H^+ \) and top in the loop. This is a lepton flavor universal contribution and modifies the coefficient of \( O_{10} = (\gamma^\alpha \gamma^\beta L)(\bar{\ell}_\gamma \gamma_\ell \ell) \). We find \[64\] the \( \rho_\ell \) correction \( \Delta C_{10}^{H^+} = [\rho_{\ell}]^{2} h(x_{\ell})/16 \pi a_e \), where the loop function \( h \) is given in the Appendix. The other loop diagrams are suppressed in the small \( \rho_{34} \) approximation and/or by extra lepton \( \rho' \) Yukawa couplings (such as in box diagrams). Similar to \( B_s \) mixing, \( \Delta C_{10}^{H^+} \) puts a constraint on the \( \rho_{\ell \mu} m_{\ell \mu}^2 \) plane. For \( m_H = m_A = 300 \text{ GeV} \), we obtain \( \rho_{\ell \mu} < 0.4 \) for 2\( \sigma \) range of \( \mathcal{B}(B_s \to \mu \mu) \), which is more stringent than \( B_s \) mixing. However, as already noted, the bound weakens if one includes other extra Yukawa couplings such as \( \rho_{\ell \tau} \), which receives \( |V_{cs} / V_{ts}| \) enhancement. In our numerical analysis, we therefore keep the tree level and one loop discussions separate, and only comment on cancellation effects later. Since LFV decays such as \( B_s \to \ell \ell' \) for \( \ell \neq \ell' \) arise at tree level in g2HDM, we give tree level upper reaches with \( \rho_{ab} \) and \( \rho_{bs} \) satisfying 2\( \sigma \) range of \( B_s \) mixing and \( B_s \to \mu \mu \).

The effective Hamiltonian for flavor violating \( B_s \to \tau \mu \) and \( B \to K \tau \mu \) decays is of the form \[65\]

\[
\mathcal{H} = -(C_S O_S + C_P O_P + C_S' O_S' + C_P' O_P'),
\]

where \( O_S = (sRb)(\bar{\ell} \ell'), \quad O_P = (sRb)(\bar{\ell} \gamma_\ell \ell') \), and \( O_S', O_P' \) are obtained by exchanging \( L \leftrightarrow R \). Although \( C_S \) and \( C_P \) vanish for \( \ell = \ell' \) in SM, tree level exchange of scalar bosons in the g2HDM lead to

\[
C_{S,P}' = \frac{m_{\phi} v_{\phi} (y_{\phi} v_{\phi}^* y_{\phi} v_{\phi}^{'})}{2 m_{\phi}},
\]

with \( \phi \) summed over \( h, H \) and \( A \), and \( C_{S,P}' \) is obtained from \( C_{S,P} \) by changing \( y_{\phi \phi} \to y_{\phi \phi b} \).

For \( B_s \to \ell \ell' \) decay, we use \[65\]

\[
\mathcal{B}(B_s \to \ell \ell') \sim f_{B_s}^2 m_{B_s} \frac{\lambda^2 (m_{B_s}, m_{\ell}, m_{\ell'})}{32 \pi (m_b + m_s)^2} \frac{\Gamma_{\text{heavy}}^{\text{dec}}}{{\Delta C_S}^2 + {\Delta C_P}^2} \times \left[ (m_{B_s}^2 - m_+^2) |\Delta C_S|^2 + (m_{B_s}^2 - m_-^2) |\Delta C_P|^2 \right],
\]

where \( \lambda(a,b,c) = [a^2 - (b-c)^2][a^2 - (b+c)^2] \), \( \Gamma_{\text{heavy}}^{\text{dec}} \) is the decay width of the heavy \( B_s \) state, \( m_\pm = m_\ell \pm m_{\ell'} \), and \( \Delta C_i = C_i - C_i^0 \). With our benchmark of \( c_\tau = 0.05 \), \( m_H = m_A = 300 \text{ GeV} \) and leptonic couplings, and the allowed range of \( \rho_{b\bar{b}b} \) extracted from flavor conserving \( B_q \to \mu \mu \) (and in conjunction with bounds from \( B_s \) mixing), the projections of various LFV \( B \) decays in g2HDM are given in Fig. 3 as red \( \Phi \). Analogously, for \( B \to K \ell \ell' \), we use \[65\]

\[
d\mathcal{B}(B \to K \ell \ell')/dq^2 = N_K \frac{\varphi_i |C_i|^2 + C_i^0|^2}{N_{\mathcal{S},P}}
\]

where \( \varphi_i \) is a function of \( B \to K \) form factors and \( N_K \) a normalization factor. Both are \( q^2 \) dependent, and explicit expressions can be found in Ref. \[65\].

\[3. \] **Comparing g2HDM with PS^3**

Let us now make the comparison of the spectacular PS^3 projections with the modesty of g2HDM.

We have taken a simplified approach of treating \( B_s \to \mu \mu \) and \( B_s \) mixing either at tree level, or at one loop level, but not both simultaneously. Either way, the fact that \( B_s \to \mu \mu \) is already consistent with SM expectation implies \( B_s \to \tau \tau \) in g2HDM should also be SM-like, which is more so if loop is dominant. This is in contrast with the sizable enhancement projected in PS^3 (grey shaded band in Fig. 3), which can be probed by LHCb Upgrade II, or dedicated runs by Belle II on \( \Upsilon(5S) \). For g2HDM, some enhancement (or suppression) of \( B_s \to \tau \tau \) is possible, given that tree effect is controlled by \( \rho_{\tau \tau} \) which is at \( O(\alpha) \), while tree effect for \( B_s \to \mu \mu \) is controlled by \( \rho_{\mu \mu} \) which is at \( O(\alpha^2) \). But these order of magnitude estimates suggest that bridging the two orders of magnitude gap is unlikely, and g2HDM should be distinguishable from PS^3. In any case, measurement of \( B_s \to \tau \tau \) is a challenge, while prospects for \( B_d \to \tau \tau \) at Belle II remains to be seen.

More promising for PS^3-type of models would be \( B_s \to \tau \mu \), which can saturate the current bound, and the discovery, perhaps even with Run 1 + 2 data of LHCb, would be truly spectacular. Projections for g2HDM, however, appears quite out of reach, as it is three orders of magnitude below the lower reach of the PS^3 projection. But our previous caution applies, that an order of magnitude enhancement is not impossible, though it would still be far out of reach. In addition, if one allows cancellation between tree and loop effects in both \( B_s \to \mu \mu \) and \( B_s \) mixing, it is not impossible that \( \rho_{b\bar{b}b} \) can be larger than our suggested values, resulting in possible further enhancement of \( B_s \to \tau \mu \). The challenge is with experiment. As we noted in Table II, the projected sensitivities, be it for LHCb or Belle, are not known publicly.

At this point, we remind the reader of the “seesaw” between \( B_s \to \tau \mu \) and \( \tau \to \mu \gamma \) within PS^3 \[27\]. Depending on analysis prowess and/or data accumulation speed, either measurement could be improved substantially in the next couple of years. If one limit is pushed down, then the prospect for the other would rise in PS^3. In contrast, for g2HDM, while there is discovery potential for \( \tau \to \mu \gamma \), one does not expect \( B_s \to \tau \mu \) to be observed any time soon. The situation for the \( B \to K \tau \mu \) mode is similar, where the projected sensitivity is again not yet clear, and we have given the number for Belle II in Table II, which barely starts to touch the PS^3 range. The situation for \( B_d \to \tau \mu \) in g2HDM would correlate with the outcome of \( B_d \to \mu \mu \) measurement, while the PS^3 model does not provide predictions. Neither models foresee \( B_q \to \mu \mu \) and \( B \to K \mu \mu \) modes to be observable. Our projections for g2HDM are given in Fig. 3.

As we have also listed in Fig 3, \( B \to \mu \bar{\nu} \) provides a unique probe \[28\] of g2HDM, while \( B \to \pi \nu \) again appears SM-like already. These are charged \( B \) decays, in contrast to neutral \( B \) decays for \( B_q \to \ell \ell' \). As a re-
minder for purely leptonic µFV processes, the \( \mu \to e \gamma \), \( \mu N \to e N \) and \( \tau \to \mu \gamma \) processes have discovery potential, all basically probing the \( \mu e \gamma \) and \( \tau \mu \gamma \) dipoles in g2HDM, though the \( \mu N \to e N \) process can pick up contact effects. In contrast, \( \mu \to 3e \) and \( \tau \to 3\mu \) would be higher order effects of the respective dipole transitions. We mention in passing that muon \( g-2 \) would not be affected in g2HDM, while muon EDM, \( d_\mu \), would likely scale by \( m_\mu/m_e \sim 200 \), and \( |d_\mu| \lesssim 2 \times 10^{-27} \) e cm seems, unlike electron EDM \( d_e \), far out of experimental reach.

V. DISCUSSION AND CONCLUSION

There are good reasons to take g2HDM, the general two Higgs doublet model with extra Yukawa couplings, very seriously. By discovering the \( h \) boson and finding that it closely resembles the SM Higgs boson, we now have one weak scalar doublet. Whether by Gell-Mann’s Totalitarian Principle [66] or the Principle of Plentitude [67], with the existence of one scalar doublet, there should be a second doublet, and by the same argument, extra Yukawa couplings. To declare [68] Natural Flavor Conservation (NFC) and forbid extra Yukawa couplings, or using a \( Z_2 \) symmetry to forbid extra Yukawa couplings, or using a \( Z_2 \) symmetry to implement it, are not only not natural, but quite ad hoc or artificial. Had supersymmetry (SUSY) emerged at the LHC, it would have given credence to 2HDM-II, a type of 2HDM with \( Z_2 \) symmetry to forbid extra Yukawa couplings. But the lack of evidence for SUSY so far [63] suggests that the SUSY scale is considerably above \( v \), the electroweak symmetry breaking scale.

With three types of charged fermions, each coming in three generations, and that the extra Yukawa couplings are naturally complex, one has 54 new Yukawa couplings, which may appear excessive. There are also 7 new Higgs parameters, which include the \( h-H \) mixing parameter \( c_v \), and the exotic Higgs masses \( m_H \), \( m_A \) and \( m_{H^\pm} \). But the increment of 54 new flavor parameters is on top of the existing plentitude of 13 within SM, while the structure built-in by Nature seems to have helped “obscure” the presence of the extra Higgs sector parameters: as we have stated, \( m_H \), \( m_A \) and \( m_{H^\pm} \) in g2HDM naturally populate the 300–600 GeV range. The latter follows if one takes [12] the principle that all dimensionless parameters in the Higgs potential are \( O(1) \) in strength, with \( v \) as the only scale parameter. It is curious to note that, with \( \rho_H \) naturally \( O(1) \) because it is a cousin to \( \lambda \), it may help keep \( c_v \) small [63]. So the alignment phenomenon may be emergent, while \( \rho_H \) could drive EWBG quite effectively. At any rate, and as we have emphasized, the flavor parameter structure seems to have hidden itself rather well from our view, obscuring also the extra Higgs bosons, which we know so little about.

The flavor structure was first revealed in the 1970s through the fermion mass hierarchy, although the existence of three generations triggered Ref. [68]. But then the mixing hierarchy of \( |V_{ub}|^2 \ll |V_{ub}|^2 \ll |V_{us}|^2 \) came as a surprise in the early 1980s, which led to the Cheng-Shen ansatz [70], suggesting that NFC may be too strong an assumption. Unknown back then was Nature’s further design of alignment, which suppressed FCNH coupling effects of the light, SM-like \( h \) boson. As we stressed in the Introduction, at this point one may find fault in the near diagonal nature of the \( \rho \) Yukawa matrix: Why would Nature turn off the FCNH effects precisely in the sector that we have the best access to? It is a mystery. But Nature has her mysterious ways, and as an experimental science we can only probe further.

In summary, the extra Yukawa couplings of g2HDM has the built-in mass-mixing hierarchy protection as exemplified by Eqs. (4) and (6), plus near diagonal \( \rho \) Yukawa matrix and alignment. The \( \mu \to e \gamma \) and \( \tau \to \mu \gamma \) processes probe \( \mu e \mu \) and \( \tau \mu \rho \) via the two loop mechanism, and generate \( \mu \to 3e \) and \( \tau \to 3\mu \) at higher order. The \( \mu N \to e N \) process probes the combined effect of dipole plus contact terms, and by nature of the process and experimental prowess, one might disentangle the two effects. As a second theme, we do not expect LUV or LFV effects to be observed soon in (semi-)leptonic rare \( B \) decay for g2HDM. This is in contrast with the UV-complete PS\(^3\) model that is the epitome of the recent \( B \) anomalies, where the modes to watch are \( B_s \to \tau \mu, B \to K \tau \mu, \) and to a lesser extent, \( B_s \to \tau \tau, B \to K \tau \tau; \) discovering only \( \tau \to \mu \gamma \) does not distinguish between the two scenarios. For g2HDM, besides the aforementioned \( \mu \)FV processes, \( B \to \mu \nu \) may be the mode to watch, which probes \( \rho_{\tau \mu} \rho_{\tau \nu} \).

Acknowledgments We thank Jack Chen, Gino Isidori, Matt Rudolph and Sheldon Stone for communications. This research is supported by MOST 106-2112-M-002-015-MY3, 108-2811-M-002-626 of Taiwan, and NTU 109L104019.

Appendix A: Loop functions

The loop functions for \( B_q \) mixing and \( B_q \to \ell \ell \) are [64]

\[
\begin{align*}
    f(a) &= -\frac{1}{2} + a \left( \frac{2a \log a}{(a-1)^2} \right), \\
    g(a, b) &= \frac{1}{(a-1)^2} \left[ \frac{3a^2 \log(a)}{a-1} + \frac{(b-4a)(b-a)}{b-1} \right. \\
    &\left. + \frac{(-4a^2 + 3ab^2 + 2ab - b^2) \log(b)}{(b-1)^2} \right], \\
    h(a) &= \frac{-a}{a-1} + \frac{a \log a}{(a-1)^2}.
\end{align*}
\]

[1] A.M. Baldini et al. [MEG], Eur. Phys. J. C 76, 434 (2016).
[2] U. Bellgardt et al. [SINDRUM], Nucl. Phys. B 299, 1 (1988).
